

FUNDAÇÃO OSWALDO CRUZ
ESCOLA NACIONAL DE SAÚDE PÚBLICA
ENSP

Análise Molecular da Epidemia pelo Variante GWGR do
subtipo B do HIV-1 no Brasil

Tese de Doutorado

Nome: Mônica Edelenyi Pinto

Orientadores: Claudio José Struchiner
Claudia Augusta de Moraes Russo

Rio de Janeiro
- 08/10/2003 -

Aos meus filhos, Clara e Bruno.

Agradecimentos

Ao Fernando, pelo carinho e pela paciência nos momentos difíceis.

Aos meus pais, por terem me conduzido até aqui.

Aos meus orientadores, dedico a eles o texto do Paracelso.

Aos meus amigos, eles sabem porque.

E a todos que, de alguma forma, contribuíram para que esta tese chegasse ao fim, muito obrigado.

“Quem nada conhece, nada ama.

Quem nada pode fazer, nada compreende.

Quem nada compreende, nada vale.

Mas quem compreende também ama, observa, vê...

Quanto mais conhecimento houver inerente numa coisa tanto maior o amor...

Aquele que pensa que todos os frutos amadurecem ao mesmo tempo,

como as cerejas, nada sabe a respeito das uvas.”

Paracelso

Resumo

A pandemia de AIDS, causada pelo vírus HIV-1, tem exercido um grande impacto sobre a humanidade com conseqüências devastadoras para muitas populações, especialmente nos países em desenvolvimento. Uma importante característica do HIV-1 é a sua enorme variabilidade genética, observada tanto nas populações virais que infectam um único indivíduo quanto entre os vírus coletados nas diferentes regiões geográficas. Entretanto, apesar desta variabilidade ser intensamente estudada, ainda existem importantes questões a serem respondidas especialmente no que se refere a diferenças nos fenótipos virais. O objetivo da tese de doutorado foi estudar a origem e a expansão da variante GWGR do subtipo B do HIV-1 encontrado principalmente na epidemia brasileira. A abordagem escolhida para o problema foi a construção de uma filogenia com todas as seqüências com o padrão de assinatura GWGR disponíveis no Genbank além de uma amostra das seqüências parentais GPGR, incluindo todas as seqüências brasileiras do subtipo B. A análise filogenética destas seqüências foi feita utilizando-se o modelo de distância de Tamura-3 para a construção de uma árvore de Neighbor-Joining. O suporte estatístico da árvore foi dado pelo teste do ramo interno. A análise filogenética sugere que a epidemia pela variante GWGR possui uma única origem cujo epicentro provavelmente foi o Brasil. A partir deste resultado, os dados epidemiológicos e a análise da topologia da árvore foram comparados para se estudar a dispersão mundial da variante. Apesar da variante GWGR circular em pelo menos 18 países, a sua dispersão foi observada principalmente no Brasil. A hipótese de que esta variante possua alguma vantagem seletiva relacionada a sua dispersão foi analisada. Esta hipótese foi sugerida uma vez que indivíduos infectados pelo sorotipo GWGR possuem um período de sobrevivência sem AIDS superior àqueles contaminados pelo sorotipo GPGR. Mesmo que fatores locais tenham auxiliado a dispersão da variante GWGR no Brasil, não pode ser descartado que a seleção natural atuou sobre este processo.

Abstract

The AIDS pandemics, caused by HIV-1, have exerted a strong impact in humankind, with devastating consequences for several populations especially in the developing countries. A high genetic variability is a remarkable feature of HIV-1. This variability is observed as well as among viral populations that infect a unique host as between viruses obtained from different geographic regions. However, although this variability has been intensely studied, there are still important questions to be answered, specially concerned with possible differences in viral phenotypes. Our work aims at the study of the origin and expansion of the GWGR variant of HIV-1 subtype B found mainly in the Brazilian epidemics. We approach this problem by constructing the virus phylogeny that included all sequences with the GWGR signature pattern available in the data banks along with a sample of the GPGR sequences which included all GPGR subtype B Brazilian sequences. The phylogenetic analyses of such sequences were performed using the Tamura-3 distance model to construct a Neighbor-Joining tree. The statistical support was calculated by the internal branch test. The phylogenetic analyses suggested that the GWGR epidemics have a single origin and that Brazil was probably its epicenter. Epidemiologic data and the analysis of the tree topology were compared in order to study the world wide dispersion of the GWGR variant. Although the GWGR variant circulates in at least 18 countries, its expansion is observed specially in the HIV-1 Brazilian epidemics. The hypothesis that the GWGR viruses have some selective advantage that could have favoured such a dispersion pattern was analyzed. This hypothesis has been raised since individuals with GWGR serotype infection show a longer survival period without AIDS than those infected with GPGR serotype viruses. Even considering that local factors may have contributed to the dispersion of the GWGR variant in Brazil, one cannot discard the role of natural selection on this process.

Sumário

Resumo.....	v
Abstract.....	vi
Lista de Figuras	ix
Lista de Tabelas.....	ix
1. Introdução	1
1.1. O agente etiológico da AIDS	1
1.1.1. Classificação sistemática	2
1.1.2. O genoma do HIV-1	3
1.1.3. A partícula Viral	4
1.1.4. A Variabilidade do HIV-1	8
1.1.4.1. Origem da Variabilidade	8
1.1.4.2. Quasispecies	9
1.1.4.3. Subtipos de HIV-1	11
1.2. Epidemiologia Molecular do HIV-1	13
1.2.2. No mundo	14
1.2.3. No Brasil	17
1.2.4. A variante GWGR	19
2. Objetivos	21
3. Material e Métodos	22
3.1. Amostragem	22
3.1.1. Amostra primária	22
3.1.1.1. Seqüências do mundo	22
3.1.1.2. Seqüências do Brasil	24
3.1.1.3. Seleção das seqüências	25
3.1.2. Amostra Secundária	26
3.1.2.1. Seqüências do mundo	26
3.1.2.2. Seqüências do Brasil	28
3.1.3. Amostra final	32
3.2. Alinhamento das seqüências	34
3.3. Distância genética	38
3.4. Análise filogenética	40
3.5. Teste do ramo interno	42
4. Resultados	44
4.1. Origem da variante GWGR	44
4.1.1. O motivo MGWGR	46
4.1.2. Seqüências GPGR do ramo GWGR	47
4.1.3. Caminho evolutivo	48
4.2. Hipótese de origem múltipla	53
4.2.1. Local de origem	55
4.3. Hipótese da origem única	57
4.3.1. Local de origem	58
4.4. Dispersão da variante GWGR	61
5. Discussão	66
5.1. Seleção Natural	67
5.1.1. Fenótipo SI nas seqüências GPGR e GWGR	71
5.1.2. Por que as diferenças nos fenótipos?	75
5.2. Outros fatores	81
5.2.1. Condições da epidemia	82

5.2.2. Fatores genéticos do hospedeiro	87
5.2.2.1. O co-receptor CCR5	88
5.2.2.2. Sistema MHC	89
6. Conclusões	92
Anexo I	93
Anexo II	102
7. Lista de Referências	103

Lista de Figuras

Figura 1	4
Figura 2	4
Figura 3	5
Figura 4	7
Figura 5	30
Figura 6	35
Figura 7	37
Figura 8	41
Figura 9	45
Figura 10.....	50
Figura 11.....	54
Figura 12.....	57

Lista de Tabelas

Tabela 1	27
Tabela 2	28
Tabela 3	33
Tabela 4	52
Tabela 5	59
Tabela 6	63
Tabela 7	72
Tabela 8	74
Tabela 9	77

1. Introdução

A Síndrome da Imunodeficiência Adquirida, ou AIDS, é uma doença crônica causada pela infecção pelo Vírus da Imunodeficiência Humana (HIV). Ao longo dos últimos 20 anos, desde sua descoberta no início dos anos 80 até os dias atuais, o HIV dispersou por todos os continentes estabelecendo uma pandemia. Durante todo este tempo, a epidemia pela AIDS tem exercido um grande impacto sobre a humanidade com conseqüências devastadoras para muitas populações, especialmente nos países em desenvolvimento.

Segundo as estimativas da Organização Mundial de Saúde (OMS) (<http://www.unaids.org/worldaidsday/2002/press/Epiupdate2002.html>), até o final do ano 2002 existiam mais de 42 milhões de pessoas vivendo com HIV/AIDS em todo o mundo. Ainda naquele ano, foram registradas três milhões de mortes por AIDS e aproximadamente cinco milhões de novas infecções com o HIV.

As estimativas mundiais sobre o número de pessoas infectadas com o HIV apontam a África sub-sahariana como a região mais afetada do planeta, com 29,4 milhões de casos. O segundo lugar é ocupado pela região da Ásia e Pacífico, onde existem 7,2 milhões de pessoas vivendo com o vírus. Os números de casos notificados nas outras regiões são: América Latina, 1,5 milhão; Europa Oriental e Ásia Central, 1,2 milhão; América do Norte, 980.000; Europa Ocidental, 570.000; África do Norte e Oriente Meio, 550.000; Caribe, 440.000; e Austrália e Nova Zelândia, 15.000. Segundo o *Boletim OMS* de 2000, de todo o continente americano, o Brasil tem o segundo número absoluto de pessoas vivendo com o HIV, 540.000, sendo superado somente pelos Estados Unidos.

1.1. O agente etiológico da AIDS

O vírus HIV-1 é o agente etiológico da AIDS. Após o relato dos primeiros casos de AIDS, no início da década de 80, o HIV foi isolado em pacientes com linfadenopatia (hiperplasia dos linfonodos) persistente por três grupos independentes de pesquisadores (Barré-Sinoussi et al., 1983; Gallo et al., 1984; Levy et al., 1984; Popovic et al., 1984). Em 1986, outro tipo de retrovírus associado à imunodeficiência humana foi encontrado em pacientes com AIDS na África Ocidental (Clavel et al., 1986). Este novo vírus foi denominado HIV-2 e os isolados iniciais foram agrupados juntos como HIV-1. Os dois

tipos de HIV (HIV-1 e HIV-2) possuem organização genômica, morfologia e tropismo celular semelhantes, porém diferem quanto à transmissibilidade, à virulência e à distribuição geográfica: enquanto os casos de infecção pelo HIV-2 permanecem quase restritos à África Ocidental, o HIV-1 dispersou-se por todos os continentes e é o responsável pela pandemia de AIDS.

Como será visto a seguir, o HIV-1 é um retrovírus complexo, da subfamília dos lentivírus. Nesta seção da introdução serão apresentadas algumas características do vírus HIV-1 de relevância para a tese.

1.1.1. Classificação sistemática

Os retrovírus são vírus cujo material genético é apresentado em duas fitas de RNA simples de polaridade positiva. Eles são assim denominados devido à sua habilidade em fazer a transcrição reversa do RNA em DNA. Na natureza, estão amplamente distribuídos, infectando vertebrados. Antes da difusão das técnicas de seqüenciamento, os retrovírus eram classificados de acordo com critérios morfológicos determinados por microscopia eletrônica e segundo o seu comportamento no organismo do hospedeiro, ou seja, pela natureza e consequência de sua ação sobre as células infectadas e por sua especificidade de hospedeiro.

Assim, a partir de seu comportamento biológico, os retrovírus foram divididos em oncovírus, lentivírus e espumavírus. Após informações contidas nas seqüências de seus genomas se tornarem disponíveis, os retrovírus passaram a ser classificados em quatro grupos, com base nas suas relações evolutivas. Segundo esta nova classificação, um destes grupos corresponde à família dos lentivírus.

Os lentivírus são assim chamados por possuírem um longo período de incubação e causarem infecção crônica em seus hospedeiros, como ocorre com o HIV-1. A família inclui, além dos HIV-1 e 2, os vírus da imunodeficiência de símios (SIV), o vírus da anemia infecciosa de eqüinos (EIAV) e o vírus visna, que infecta os carneiros (Doolittle & Feng, 1992). Os HIVs e os SIVs formam a família dos lentivírus de primatas (Doolittle & Feng, 1992; Doolittle et al., 1990).

1.1.2. O genoma do HIV-1

De modo geral, os genomas de todos os retrovírus são compostos por quatro segmentos básicos, que são as regiões LTR (*long terminal repeats*) e os genes *gag*, *pol* e *env*. Contudo, como dito anteriormente, os lentivírus são retrovírus complexos, o que significa que o seu genoma contém, além dos elementos citados, um conjunto de genes acessórios e reguladores que controlam a replicação do vírus, ou seja, os genes *vif*, *vpr*, *tat*, *rev* e *nef*. A estrutura básica, e provavelmente ancestral, do genoma dos lentivírus de primatas é formada, nesta ordem, pelos segmentos LTR-*gag-pol-vif-vpr-tat-rev-env-nef*-LTR (Doolittle & Feng, 1992; Doolittle et al., 1990).

As regiões LTR estão localizadas nas duas extremidades do genoma dos retrovírus. Estes elementos contêm regiões importantes para a regulação da transcrição dos provírus (moléculas de DNA virais produzidas por transcrição reversa e integradas nas células do hospedeiro). O gene *gag* codifica as proteínas que compõem o capsídeo viral, ou seja, que formam a sua estrutura protéica. O gene *pol*, por sua vez, codifica as enzimas transcriptase reversa, protease e integrase. Estas enzimas são produzidas como uma poliproteína Gag-pol precursora que posteriormente é processada pela protease viral.

O gene *env* codifica glicoproteínas que, na partícula viral madura, estão inseridas na membrana lipídica, também chamada de envelope, que recobre o envoltório protéico do vírus. As glicoproteínas do envelope desempenham um papel crucial no processo de infecção, pois é através delas que o vírus interage com os receptores localizados na superfície das células do hospedeiro e penetra nestas células.

O HIV-1 contém os genes reguladores *vif*, *vpr*, *tat*, *rev* e *nef* encontrados em todos os lentivírus. Além destes, este vírus também possui os genes *vif*, *vpu* e *vpx*. De modo geral, estes genes controlam a replicação do HIV-1, seja ativando a transcrição do seu genoma (gene *tat*), promovendo a exportação do RNA mensageiro viral do núcleo para o citoplasma (gene *rev*) ou potencializando a liberação do vírus para fora da célula infectada (*vpu*), por exemplo (Kuiken et al., 2000a). A estrutura do genoma do HIV-1 é apresentada na figura 1.

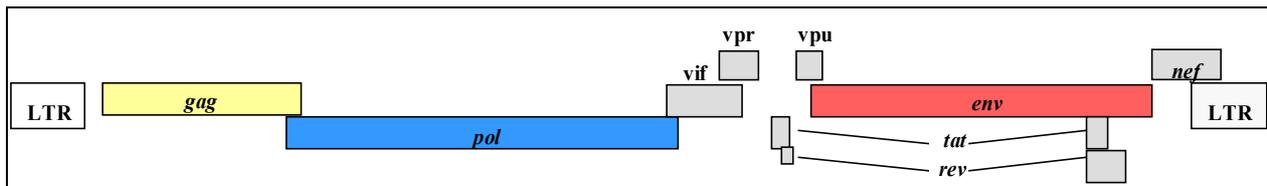


Figura 1. Estrutura do genoma do HIV-1

1.1.3. A partícula Viral

Um esquema da partícula do HIV-1 é apresentado na figura 2.

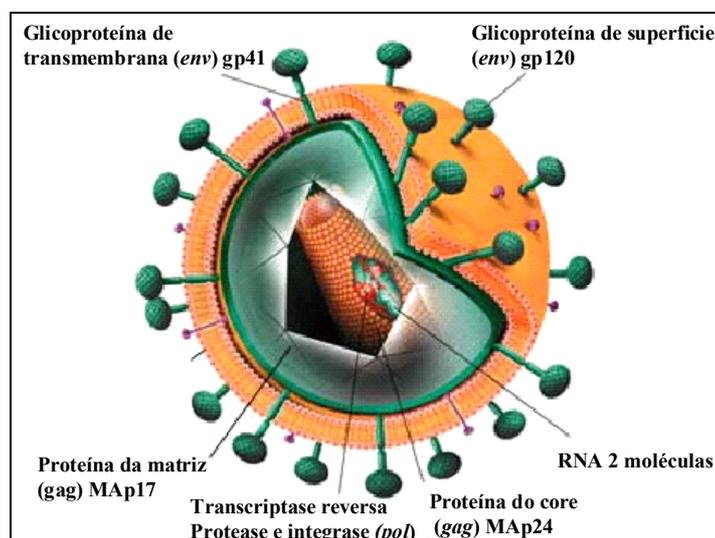


Figura 2. Estrutura da partícula viral do HIV-1

O gene *gag* codifica uma proteína precursora de 55kDa a qual, durante a replicação do vírus, é clivada pela protease viral, dando origem às proteínas p24, p17, p7 e p6 (figura 2). A proteína p17 compõe a matriz do vírus e é encontrada logo abaixo do seu envelope. A proteína p24 forma o nucleocapsídeo, no interior do qual estão localizadas as duas fitas de RNA. As proteínas p7 e p6 estão associadas ao material genético do vírus.

No HIV-1, o gene *env* codifica uma glicoproteína precursora de 160kDa, chamada de gp160. Esta proteína é processada por enzimas celulares em um complexo

não-covalente inserido no envelope da partícula viral madura. Os componentes do complexo são as glicoproteínas gp41 e a gp120. A inserção do complexo no envelope se dá através da gp41 (glicoproteína transmembrana), enquanto a gp120, ancorada na primeira, é exposta ao meio externo.

Como para outros vírus, as glicoproteínas do envelope representam as estruturas mais externas da partícula viral e são, assim, o principal alvo da resposta imune do hospedeiro. Conseqüentemente, estas moléculas e, naturalmente, o gene *env* que as codifica, se caracterizam por uma grande variabilidade. Contudo, esta diversidade não está homogeneamente distribuída ao longo do gene e de seu produto, a glicoproteína precursora gp160.

Na realidade, a partir da análise de diversas moléculas de gp160, esta glicoproteína foi subdividida em regiões constantes e em regiões de grande variabilidade. As regiões constantes receberam os nomes C1, C2, C3, C4, C5 e C6, enquanto as regiões hipervariáveis foram designadas V1, V2, V3, V4 e V5. Os parâmetros utilizados para a determinação destas regiões foram a comparação das seqüências de aminoácidos de diversas proteínas (estrutura primária), a análise da conformação das diversas regiões da molécula gp160 (estruturas secundária e terciária) e a localização dos seus sítios de glicosilação (Modrow et al., 1987; Starcich et al., 1986; Willey et al., 1986).

A distribuição dos segmentos constantes e variáveis na gp160 é apresentada na figura 3.

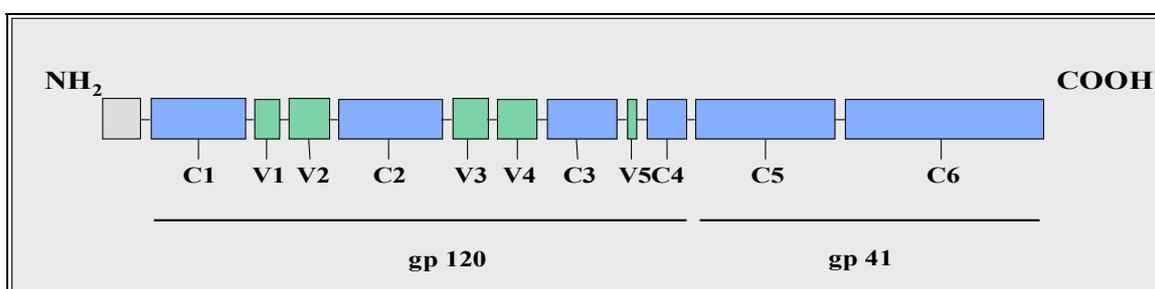


Figura 3 – Regiões constantes (C) e variáveis (V) da glicoproteína gp160 do HIV-1. Como pode ser observado na figura, todas as regiões variáveis estão localizadas na porção N-terminal da molécula que corresponde à glicoproteína de superfície gp120. A porção C-terminal, por outro lado, é formada pelos segmentos C5 e C6 e corresponde à glicoproteína transmembrana gp41

Como pode ser observado, todas as regiões variáveis estão localizadas próximo à porção N-terminal da molécula, onde se localiza a glicoproteína gp120 (Modrow et al., 1987; Starcich et al., 1986; Willey et al., 1986). Uma destas regiões hipervariáveis, a região V3, representa uma estrutura especialmente importante para a biologia do vírus e para a sua interação com os hospedeiros humanos. Além disto, como a região V3 será objeto de estudo da tese, sua estrutura e importância funcional serão resumidamente abordadas a seguir.

A estrutura da região V3 é composta por 33 a 35 resíduos de aminoácidos. Nas suas duas extremidades são encontrados dois resíduos de cisteína invariáveis que interagem entre si, formando uma ponte dissulfeto. A interação entre estes dois resíduos de cisteína confere à região a conformação de uma alça, conhecida como alça V3 (Javaherian et al., 1989; Putney et al., 1986; Rusche et al., 1988).

Apesar da composição da alça V3 ser extremamente heterogênea entre diferentes cepas do HIV-1, esta estrutura possui alguns elementos extremamente conservados. Além dos dois resíduos de cisteína invariáveis, a parte central da alça V3, formada pela sequência GPG (glicina-prolina-glicina), também é bastante conservada entre as variantes do vírus (Catasti et al., 1996; Catasti et al., 1995). Desta forma, a variabilidade da região é especialmente observada nos dois segmentos adjacentes aos resíduos de cisteína (Tian & Chen, 2002). A conformação da alça V3 é apresentada na figura 4.

A região V3 contém o principal epítipo de neutralização do vírus, constituído pelos resíduos dos aminoácidos IPMGPGR(Q)AFY (Javaherian et al., 1989; Putney et al., 1986; Rusche et al., 1988). Como pode ser visto na figura 4, este epítipo inclui o motivo protéico conservado (sublinhado) encontrado no topo da alça V3. Além disto, a região contém também epítipos capazes de induzir a resposta imune celular mediada por linfócitos T citotóxicos (Casement et al., 1995). De um modo geral, os elementos efetores da resposta imune específica são os anticorpos e as células T citotóxicas, capazes de reconhecer, respectivamente, antígenos solúveis e apresentados pelas moléculas de histocompatibilidade; os epítipos são, por sua vez, os sítios dos antígenos com os quais os anticorpos e os receptores das células T interagem. Os epítipos localizados em antígenos de patógenos ou de toxinas e que são reconhecidos por anticorpos são conhecidos como epítipos de neutralização.

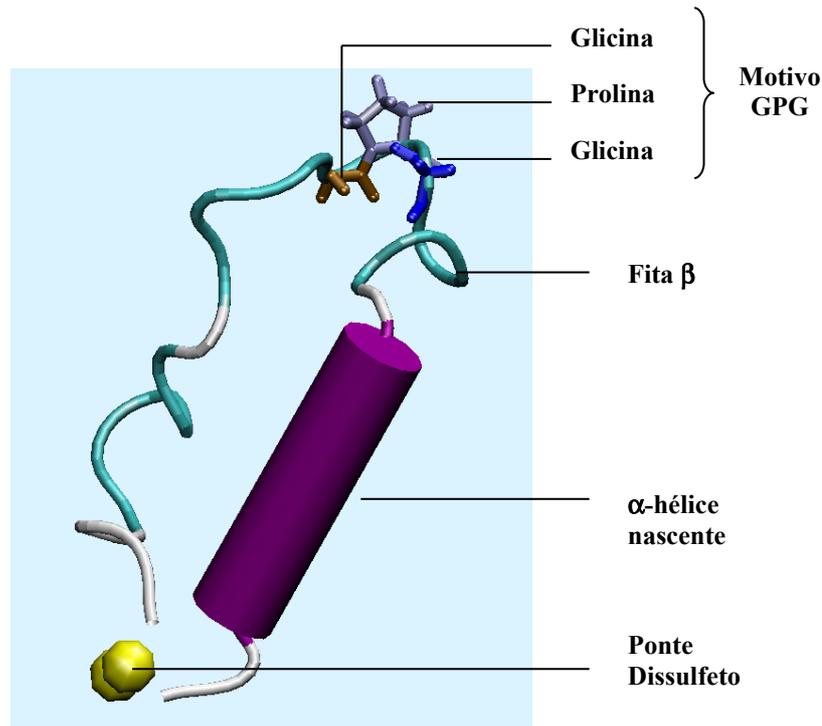


Figura 4– Conformação da alça V3 do HIV-1. Nas duas extremidades da alça estão os dois resíduos de cisteína que interagem entre si para formar a ponte dissulfeto. No topo da alça podem ser vistos os resíduos de glicina, prolina e glicina que formam o motivo protéico GPG, extremamente conservado entre as seqüências do vírus. A conformação da alça V3 possui alguns elementos estruturais secundários bem definidos: uma volta no topo, onde se localiza o motivo GPG, regiões de fitas β adjacentes à volta nas regiões N e C-terminais e uma hélice nascente na região C-terminal.

Além de sua imunogenicidade, a região V3 está envolvida: 1) no tropismo do HIV-1, uma vez que participa da interação do vírus com os co-receptores celulares CCR5 e CXCR4 nas diferentes fases da infecção (van Rij et al, 2000); 2) na penetração do vírus nas células do hospedeiro (Fouchier & Schuitemaker, 1996); e 3) na aquisição do fenótipo indutor de sincício (fenótipo SI), observado em vírus obtidos de pacientes nos estágios tardios da imunodeficiência (Fouchier & Schuitemaker, 1996). Este tema - os fenótipos do HIV-1 no curso da imunodeficiência - por ser especialmente relevante para a tese, será abordado no item 5.1.1, Discussão.

Devido à sua imunogenicidade e à sua importância funcional, existe um grande interesse dos grupos de pesquisa acerca da estrutura e da diversidade da região V3. Isto

se reflete no acúmulo de grande número de seqüências da região disponíveis nos bancos de dados e na ampla literatura que existe sobre o tema.

1.1.4. A Variabilidade do HIV-1

Uma característica importante dos vírus RNA, como os retrovírus, é a sua grande variabilidade. Esta variabilidade é observada tanto nas seqüências de seus genomas (variabilidade genética) quanto nas proteínas que compõem as estruturas desses vírus (variabilidade antigênica). Durante as últimas décadas, vários trabalhos mostraram que estes vírus RNA compartilham algumas características, como uma grande mutabilidade. Além do HIV-1, outros exemplos de vírus RNA patogênicos para o ser humano são o vírus influenza e o vírus da hepatite C (Domingo et al, 1997; Overbaugh & Bangham, 2001; Sala & Wain-Hobson, 1999).

Em conseqüência de sua enorme mutabilidade, os vírus RNA são capazes de responder rapidamente às alterações do meio, como por exemplo à resposta imune do hospedeiro e à administração de drogas antivirais. Além disto, o aparecimento contínuo de genomas mutantes dificulta substancialmente o desenvolvimento de vacinas efetivas contra os mesmos (Novitsky et al., 2002).

1.1.4.1. Origem da Variabilidade

A base molecular da extrema variabilidade dos vírus RNA é o fato de suas enzimas RNA polimerases e transcriptases reversas (encontradas nos retrovírus) não possuírem a propriedade de correção das moléculas de ácidos nucleicos nascentes durante o processo de replicação viral, propriedade esta comum às DNA-polimerases de outros organismos. Esta ausência de correção das RNA-polimerases e transcriptases reversas resulta em uma taxa de erro na incorporação de nucleotídeos de aproximadamente 10^{-3} a 10^{-5} por sítio, por ciclo de replicação (a taxa de erro das DNA polimerases está em torno de 10^{-8} substituições por sítio por ciclo de replicação). No HIV-1, por exemplo, cujo genoma possui cerca de 10 kb (kilobases), a cada ciclo de replicação, seu genoma adquire em média uma substituição de nucleotídeos (Domingo et al., 1997; Meyerhans & Vartanian, 1999; Overbaugh & Bangham, 2001).

Além da enorme probabilidade de mutação por geração, a complexidade genética da população viral em um hospedeiro infectado também será determinada pelo número de ciclos de replicação que ocorre. No caso do HIV-1, a cada dia são produzidas em média 10^{10} partículas virais (Markowitz et al., 2003). Assim, os vírus que possuem taxas elevadas e contínuas de replicação, como o HIV-1, desenvolvem um grau extraordinário de diversidade em suas seqüências (Domingo et al., 1997; Overbaugh & Bangham, 2001). Ainda em relação à infecção pelo HIV-1, o longo curso da infecção, somado aos erros na incorporação de nucleotídeos e à sua elevada taxa de replicação, também contribui para a geração de uma quantidade enorme de variantes virais.

Entre os retrovírus, além dos erros na incorporação de nucleotídeos mediados pela transcriptase reversa (TR) e das suas elevadas taxas de replicação, um outro processo contribui para a enorme diversidade observada em seus genomas. Durante a transcrição reversa do RNA viral, a TR pode “pular” de uma fita de RNA para outra (como foi visto, o genoma do HIV-1 é composto por duas fitas de RNA). Este processo é conhecido como recombinação homóloga. O resultado da recombinação é a síntese de uma fita de DNA viral contendo segmentos dos dois RNA iniciais (Hu & Temin, 1990).

Contudo, para que o processo de recombinação contribua substancialmente para a geração de diversidade, é necessário que as duas fitas de RNA encontradas em uma partícula viral sejam distintas uma da outra, ou seja, é necessário que anteriormente tenha ocorrido a produção de genomas virais “heterozigotos”. Por sua vez, para que estes genomas “heterozigotos” sejam formados, é preciso que em algum momento uma única célula do hospedeiro tenha sido simultaneamente infectada por duas variantes do vírus (Hu & Temin, 1990). Este processo - a co-infecção de um mesmo indivíduo por duas variantes virais - foi demonstrado em casos de infecção pelo HIV-1 (Janini et al., 1998). Além dessa evidência epidemiológica, a literatura mostra que uma grande quantidade das cepas de HIV-1 circulantes possui genomas recombinantes (Holguín et al., 2000; Liitsola et al., 2000; Vicente et al., 2000; Yirrell et al., 2002).

1.1.4.2. Quasispecies

Como consequência da elevada taxa de erro das RNA polimerases e das TRs, uma grande quantidade de genomas mutantes são produzidos durante a replicação dos vírus RNA. Deste modo, as populações de vírus RNA que se replicam em um

hospedeiro infectado são espectros extremamente dinâmicos de mutantes conhecidos como quasispecies.

Como resultado da imensa complexidade destas populações virais, as quasispecies são capazes de responder rápida e eficientemente às perturbações do meio ambiente onde se replicam. A eficiência desta resposta se deve ao fato dos vírus oferecerem um grande espectro de mutantes sobre o qual a seleção natural pode atuar. Quando existem alterações no meio ambiente, por exemplo, pela administração de drogas antivirais e pelo desenvolvimento da resposta imune adaptativa, a presença de um ou mais mutantes mais aptos a replicar neste novo ambiente faz com que a população viral derivada desses mutantes ganhe aptidão (*fitness*) e aumente suas chances de sobreviver.

Desta forma, a base da adaptabilidade dos vírus RNA é a amplificação seletiva de um subgrupo de genomas mais aptos a se replicar em um meio ambiente alterado (Domingo et al., 1997). No caso do HIV-1, a pressão seletiva exercida pelo sistema imune permite aos vírus se adaptar no sentido de encontrarem novas células-alvo no hospedeiro e, assim, manterem uma infecção persistente (Overbaugh & Bangham, 2001).

Na infecção pelo HIV-1, apesar da grande diversidade continuamente gerada na população de vírus durante o curso da imunodeficiência, quando estes vírus são transmitidos a um novo hospedeiro susceptível, existe um significativo estreitamento desta diversidade. Em consequência, as populações de vírus recuperadas de infecções primárias são mais homogêneas e possuem uma distribuição mais estreita de quasispecies virais do que as populações encontradas nos indivíduos doadores (Overbaugh & Bangham, 2001).

Apesar do estreitamento da variabilidade ser observado durante a transmissão do HIV-1 por qualquer via, as quantidades de vírus às quais os indivíduos receptores estão expostos pode variar, dependendo do modo de infecção. De modo geral, as transmissões pelas vias sexual e vertical envolvem quantidades pequenas de vírus e de células infectadas, enquanto grandes cargas de vírus podem ser transmitidas quando há contato direto com o sangue, como ocorre nas transmissões através do compartilhamento de seringas e durante as transfusões de sangue (Quiñones & Arts, 2001).

Existem diversos fatores seletivos que impõem este estreitamento da diversidade, dentre os quais podem ser citados: 1) fatores do hospedeiro, como a

imunidade inata; 2) a densidade de células-alvo no local da infecção, 3) o número de partículas virais transmitidas; e 4) a estrutura das quasispecies do hospedeiro transmissor. De qualquer modo, o estreitamento da população de quasispecies durante o processo de transmissão do vírus atua como um evento fundador e, em consequência, a progênie viral produzida no início da infecção resulta da expansão de uma ou poucas partículas virais (Quiñones & Arts, 2001).

Durante o processo de transmissão a um novo hospedeiro e na posterior expansão de um clone particular de HIV-1 ocorre um processo de seleção natural, de modo que os vírus encontrados nos estágios iniciais da infecção possuem características distintas daqueles observados nos estágios mais tardios. De um modo geral, os isolados virais de indivíduos nos estágios iniciais da infecção pelo HIV-1 apresentam baixas taxas de replicação, não possuem capacidade de induzir a formação de sincícios *in vitro* (fenótipo NSI), replicam mais eficientemente em macrófagos do que em células T transformadas e utilizam preferencialmente o co-receptor CCR5 para infectar a célula (Fouchier & Schuitemaker, 1996; Quiñones & Arts, 2001).

Por outro lado, em estágios avançados da imunodeficiência, a maioria das cepas virais apresenta altas taxas de replicação, replicam-se com maior eficiência em células T e em aproximadamente 50% dos indivíduos os vírus adquirem a capacidade de induzir sincícios (fenótipo SI). Além disto, necessitam do co-receptor CXCR4 para infectar novas células. Apesar das variantes SI/CXCR4 dominarem as populações de quasispecies nas fases mais avançadas da imunodeficiência, as variantes NSI/R5 são tipicamente transmitidas para os hospedeiros durante o processo de infecção (Fouchier & Schuitemaker 1996; Quiñones & Arts, 2001).

1.1.4.3. Subtipos de HIV-1

Além da estrutura de quasispecies encontrada nos hospedeiros infectados, uma outra consequência natural da grande variabilidade do HIV-1 é a grande diversidade observada entre os vírus que circulam na epidemia. Assim, na tentativa de organizar esta diversidade, até 1992 as variantes do vírus eram identificadas como “americanas” (*American-like*) ou “africanas” (*African-like*), com base em seu local de origem.

Com o isolamento e a análise de novos espécimes, esta separação tornou-se inadequada e foi proposta uma classificação em subtipos de seqüências. No momento, 9 subtipos “puros” são reconhecidos e nomeados A, B, C, D, F, G, H, J e K. Junto com os

subtipos, 15 formas recombinantes circulantes (CRF) são encontradas nos diferentes locais afetados pela epidemia, sendo as mais comuns a CRF02_AG e a CRF01_AE (Osmanov et al., 2002). Os subtipos “puros” e os CRFs formam o grupo M (*major*). Amostras muito divergentes, que não se agrupam com nenhuma dessas variantes, são coletivamente conhecidas como grupo O (*out-group*) (Gürtler et al., 1994). Além dos grupos M e O, um novo grupo foi proposto, N (*new*), formado por um pequeno número de amostras descobertas nos Camarões (Simon et al., 1998).

Para evitar erros relativos a genomas recombinantes, o método de escolha para classificação do HIV-1 em subtipos é a análise filogenética do genoma completo do vírus. Esta abordagem, contudo, por ser demorada e ter um alto custo, não pode ser aplicada em larga escala, e deve ser integrada a inquéritos de epidemiologia molecular que incluam técnicas mais acessíveis (Workshop Report from the European Commission DG XII, 1997). Uma alternativa utilizada com frequência é o seqüenciamento e a análise de pequenos fragmentos do genoma, especialmente dos genes *gag* ou *env*. Contudo, mesmo esta última abordagem pode ser inacessível, por exemplo, para grupos de pesquisa de países em desenvolvimento, ou simplesmente inaplicáveis a estudos envolvendo um grande número de amostras.

Desta forma, diversas técnicas alternativas foram desenvolvidas, sendo a mais utilizada o ensaio de mobilidade de heteroduplexes (*Heteroduplex Mobility Assay*) ou HMA (Delwart et al., 1993) que possui quase 100% de correspondência com as classificações obtidas por análise filogenética (Bachmann et al., 1994). A classificação do HIV-1 em subtipos e subgrupos e o desenvolvimento de técnicas acessíveis para o estudo de sua diversidade resultaram na publicação de um grande número de trabalhos, descrevendo a presença das variantes de HIV-1 nas diversas regiões geográficas ou subpopulações.

Apesar de sua importância, os estudos de subtipagem possuem limitações metodológicas. Em primeiro lugar, na sua quase totalidade, os trabalhos utilizam amostras obtidas por conveniência e não representativas da população total. Em segundo lugar, a maioria dos estudos emprega técnicas de subtipagem que analisam somente um fragmento do genoma (geralmente a alça V3 do gene *env*). Devido à crescente participação de formas recombinantes na epidemia mundial, quando se quer determinar a qual subtipo um vírus pertence qualquer técnica que não seja a análise filogenética de todo seu genoma pode resultar em erros de classificação, uma vez que segmentos diferentes de um mesmo vírus podem ser agrupados em subtipos distintos.

Desta forma, sempre que as classificações em subtipos forem estabelecidas de fragmentos do genoma do vírus, os resultados devem indicar qual segmento genômico foi utilizado na análise (Peeters, M., 2000).

1.2. Epidemiologia Molecular do HIV-1

A epidemiologia molecular de doenças infecciosas tem como objetivo recuperar as informações contidas nos genomas dos microorganismos e utilizá-las no estudo das epidemias. Diversos exemplos encontrados na literatura ilustram as múltiplas possibilidades de utilização de técnicas como a Reação em Cadeia da Polimerase (PCR) e a determinação e análise de seqüências de ácidos nucleicos ou aminoácidos, seja na investigação de novas doenças, na identificação de patógenos ou no estudo de formas pouco comuns de transmissão e desenvolvimento de doenças conhecidas (McDade & Anderson, 1996).

Além disto, trabalhos recentes propõem modelos estatísticos que utilizam as informações acumuladas nos genomas dos patógenos para a reconstrução da dinâmica da dispersão dos microorganismos nas populações de hospedeiros (Holmes et al., 1995; Pybus et al., 2000). A compreensão da história da interação destas duas populações – patógenos e hospedeiros – permite fazer predições sobre os caminhos futuros das epidemias. Um exemplo da parceria entre a epidemiologia clássica e a biologia molecular é o estudo da epidemia da AIDS e de seu agente etiológico, o HIV.

Uma ferramenta amplamente utilizada para o estudo da epidemiologia molecular do HIV-1 é a organização das variantes em subgrupos, subtipos e CRFs. Além disto, dentro de uma região geográfica mais restrita, a análise de padrões de mutação observados em cepas de um mesmo subtipo também pode ser útil para a reconstrução da trajetória do vírus entre indivíduos infectados por uma mesma via de transmissão (Kuiken et al., 2000b). Na seção 1.2.2, a distribuição dos subtipos de HIV-1 na epidemia mundial será rapidamente apresentada junto com um exemplo da análise de padrões de mutação na epidemiologia molecular do HIV-1.

1.2.2. No mundo

A observação da distribuição geográfica dos subtipos e subgrupos de HIV-1 permite que se façam algumas considerações. De modo geral, a epidemia nos países industrializados é diferente da epidemia nos países em desenvolvimento. Enquanto nos primeiros, durante a década de 80 e início dos anos 90, a epidemia foi dominada pelo subtipo B, na maioria dos países em desenvolvimento, subtipos não-B sempre foram os mais frequentes. Uma exceção é a América Latina, onde o subtipo B domina a epidemia desde o seu início. Segundo, apesar da grande diversidade do HIV-1, os subtipos A-D e as formas recombinantes CRF01_AE e CRF02_AG são os mais comuns no mundo. Dentre os mais frequentes, destaca-se o subtipo C, estimado como o responsável pela metade das infecções conhecidas (Essex, 1999), além de ser o que mais se expande na pandemia atualmente (Osmanov et al., 2002). Esta observação pode ser atribuída, pelo menos em parte, ao fato do subtipo C circular em países com elevada incidência de novas infecções, como Etiópia e Índia (Osmanov et al., 2002).

Contudo, os dados sobre a distribuição dos subtipos nas diversas populações e regiões geográficas afetadas pela epidemia não são homogêneos. Enquanto faltam informações sobre os variantes que circulam em regiões com grande número de indivíduos infectados (por exemplo, Moçambique, Angola e República Dominicana) existem dados na literatura sobre os vírus que circulam em países como Suécia, Dinamarca e Finlândia, todos com baixa prevalência de infecção. Isto mostra que o conhecimento que se tem da diversidade do HIV-1 é incompleto, abrindo caminho para a existência de variantes genética e antigenicamente distintas dos vírus descritos até o momento.

Claramente, a maior diversidade de seqüências de HIV-1 é observada no continente africano. Na África sub-sahariana estão presentes todos os variantes do vírus (incluindo os grupos O e N), sendo que a maior variedade de subtipos do grupo M é encontrada na África Central (Camarões, Gabão, República Democrática do Congo e República da África Central) (Mboudjeka et al., 1999; Vidal et al., 2000). A presença de todos os subtipos do grupo M e de diversas formas recombinantes na África Central corrobora a hipótese de que a introdução e irradiação inicial do HIV-1 na espécie humana ocorreu nesta região (Vidal et al., 2000). Contudo, apesar da grande variabilidade observada entre os isolados de HIV-1 africanos, em quase toda a África sub-sahariana a epidemia é dominada pelos subtipos A (incluindo a forma recombinante CRF02_AG) e C (Janssens & Nkengasong, 1997; Osmanov et al., 2002; Workshop Report from the European Commission DG XII, 1997).

Por outro lado, as epidemias do continente americano e da Europa Ocidental foram fundadas pelo subtipo B e, durante mais de uma década, este foi o único subtipo encontrado nestas regiões. Durante a década de 90, outros subtipos foram progressivamente introduzidos nestes dois continentes, em parte através do contato sexual com imigrantes africanos e asiáticos (Machuca et al., 2001; Op de Coul et al., 2001b; Russel et al., 2000). Por exemplo, nos Estados Unidos, além do subtipo B, também circulam os subtipos A, D, F e vírus CRF-01 AE (Brodine et al., 1995; Gao et al., 1994; Weidle et al., 2000). No Brasil e no restante da América Latina, apesar de haver diferenças regionais, além do subtipo B também são encontrados vírus A, C, D e F (Bongertz et al., 2000; Hierholzer et al., 2002; Velarde-Dunois et al., 2000). Na Europa Ocidental quase todos os subtipos são encontrados, mas as variantes não-B infectam principalmente imigrantes, infectados em seus países de origem (Alaeus et al., 1997; Arnold et al., 1995; Böni et al., 1999; Clewley et al., 1996; Engelstad, 1996; Holguín et al., 1999).

Como dito anteriormente, os métodos para a subtipagem não são a única ferramenta disponível para os estudos da epidemiologia molecular do HIV-1. Além da subtipagem, a análise de padrões de mutação específicos em determinadas regiões do genoma tem sido utilizada para se reconstruir a dispersão de cepas do vírus dentro de grupos sob risco para a infecção pelo HIV-1. Talvez o melhor exemplo desta abordagem seja a história dos vírus do subtipo B que circulam dentro de diferentes grupos de risco na Europa.

Durante a década de 90, diversos estudos mostraram que as cepas do subtipo B que fundaram a epidemia entre os usuários de drogas da Holanda, da Alemanha, da Suíça e do Reino Unido possuem mutações nos genes *env*, *vpr*, *vpu* e na região do gene *pol* que codifica a enzima transcriptase reversa (RT) que as diferencia dos vírus encontrados nos homossexuais dos mesmos países (Kuiken et al., 2000b). Dentre estas mutações, a mais conservada entre os diferentes isolados é uma substituição sinônima no segundo códon para a glicina (G) no motivo protéico do topo da alça V3 (gene *env*).

Na maioria dos isolados do subtipo B descritos na epidemia mundial, o códon para a segunda glicina no topo da alça V3 é GGG (Kuiken et al., 2000a). Contudo, para a maior parte dos usuários de drogas dos países descritos anteriormente, o códon para a glicina neste sítio é GGC. Por outro lado, esta substituição sinônima é raramente encontrada entre os homossexuais infectados da mesma região. Já nos países do sudoeste da Europa, como França e Espanha, não há segregação de vírus GGC e não-

GGC entre indivíduos destas duas populações sob risco (Lukashov et al., 1996). Desta forma, a assinatura GGC no códon para a glicina, encontrada entre os usuários de drogas da Holanda, da Alemanha, da Suíça e do Reino Unido, foi utilizada como um marcador para se reconstruir a dispersão do HIV-1 dentro da população para outras regiões geográficas (Op de Coul et al., 2001a).

Nos países da Europa Oriental e na Ásia Central a epidemia pelo HIV-1 é mais recente. No momento, esta região experimenta um grande crescimento no número de casos de infecção, especialmente associado ao uso de drogas injetáveis. Nestes países, de modo geral, os casos adquiridos por transmissão homossexual são atribuídos principalmente ao subtipo B, enquanto as infecções na população heterossexual estão relacionadas a diversos subtipos do HIV-1, como A, C, D, F e G (Bobkov et al., 1996; Lukashov et al., 1995; Mayer et al., 1998).

Ainda nessa última região, em meados da década de 90, ocorreu uma explosão de casos de infecção entre usuários de drogas injetáveis causada por vírus do subtipo A extremamente homogêneos, sugerindo que a epidemia tenha se iniciado a partir de um único evento de exposição (Bobkov et al., 1997; Ferdats et al., 1999; Lukashov et al., 1998). Em seguida, vírus recombinantes AB, dos quais o segmento A parental foi fornecido pelos vírus do subtipo A que circulavam na mesma população, também se dispersaram rapidamente entre os consumidores de drogas injetáveis (Bobkov et al., 1998; Liitsola et al., 1998; Liitsola et al., 2000).

Um caso especial é a epidemia da Romênia. Neste país, os primeiros casos de infecção pelo HIV-1 foram detectados no final dos anos 80, entre crianças infectadas através de agulhas e seringas contaminadas com vírus do subtipo F (Apetrei et al., 1997). Posteriormente, o HIV-1 foi identificado em outras populações sob risco, como crianças infectadas por transmissão vertical e adultos que adquiriram o vírus por contato sexual. Contudo, até o fim da década de 90, o subtipo F se mantinha como o principal responsável pela epidemia em adultos e crianças na Romênia, apesar de outras variantes, como os subtipos A, B, C, D e o recombinante CRF_01_AE, também terem sido identificados no país (Apetrei et al., 1998).

No sul e sudeste da Ásia, os dois principais focos da epidemia, Tailândia e Índia, têm histórias de epidemias que se iniciaram por vias de infecção distintas associadas a diferentes subtipos virais (Essex, 1999). Nos dois países, até meados da década de 90, a maioria dos usuários de drogas injetáveis (UDI) estava infectada por vírus do subtipo B,

enquanto o recombinante CRF_01 AE (Tailândia) e o subtipo C (Índia) eram responsáveis pela maioria dos casos ocorridos por transmissão heterossexual (Essex, 1999; Voevodin et al., 1996). Nos últimos anos, na Tailândia, observa-se o aumento da prevalência de infecções pelos vírus CRF_01 AE entre os usuários de drogas (Kitayaporn et al., 1998; Limpakarnjanarat et al., 1998). Por outro lado, na Índia, atualmente o subtipo C é o responsável pela maioria das infecções, independentemente da via de infecção. Ainda na região, o terceiro maior número de casos é registrado em Miamar (Kusagawa et al., 1998).

A análise molecular das seqüências do subtipo B da Tailândia, de Miamar e da Malásia mostrou que, em sua maioria, estas seqüências possuem uma mutação no topo da alça V3 (motivo GPGQ) que as distingue das seqüências prototípicas do subtipo B (motivo GPGR). Além disto, nas árvores filogenéticas, estas seqüências se agrupam separadamente das seqüências B americanas e européias. Desta forma, estas variantes asiáticas são denominadas subtipo B' (Brown et al., 1996; Kusagawa et al., 1998; Limpakarnjanarat et al., 1998; Mastro et al., 1997; Subbarao et al., 2000).

Na China, o uso de drogas injetáveis é o principal fator de risco para a infecção pelo HIV-1, e a maioria das infecções está associada ao subtipo C (Novitsky et al., 2002). No Japão, até 1990, 80% dos casos de AIDS relatados haviam ocorrido em hemofílicos e o subtipo B predominava no país (Kitsutani et al., 1998). Na década de 90, cresceu o número de casos de infecção e de AIDS atribuídos ao contato sexual (Ichikawa et al., 1997) e novos subtipos foram progressivamente identificados no país, como o A, o C e o recombinante CRF_01 AE (Chijiwa et al., 1999; Kitsutani et al., 1998).

1.2.3. No Brasil

Os primeiros casos de AIDS no Brasil datam de 1980 (MS, 1995) e ocorreram na região sudeste do país. Durante a primeira metade da década de 80, a epidemia esteve restrita às duas maiores regiões metropolitanas, São Paulo e Rio de Janeiro mas, desde então, vem espalhando-se para as outras regiões brasileiras. Atualmente, há registros de casos de AIDS em todos os estados do país, embora o maior número de casos ainda esteja concentrado nas regiões mais desenvolvidas economicamente, Sul e Sudeste.

Além da heterogeneidade na distribuição geográfica dos casos de HIV/AIDS, a epidemia no Brasil é composta por sub-epidemias distintas, com diversas vias de transmissão e disseminadas em épocas diferentes. A participação das diversas categorias de exposição na epidemia brasileira mudou ao longo do tempo (MS, 2002). Por exemplo, durante a primeira década (1980 a 1990), a maior parcela dos casos de AIDS registrados no país estava associada à transmissão homossexual (32%), seguida do uso de drogas injetáveis (18,5%). Naquele período, os casos atribuídos à exposição por contato hetero e bissexual representaram, respectivamente, 16,4% e 15,4% do total.

Durante os anos 90 ocorreu uma diminuição progressiva da proporção de casos entre homo e bissexuais (13,0% e 8,8% em 2000, respectivamente) e uma disseminação da infecção dentro do grupo de heterossexuais (43,5% em 2000). Além disso, ao longo das duas décadas de epidemia, houve uma redução progressiva da proporção de casos de AIDS entre indivíduos do sexo masculino e feminino: em 1990 a razão de sexo era 6,5:1 passando para 2,5:1 em 2002 (MS, 2002).

Quanto aos casos de AIDS em indivíduos infectados por via parenteral (usuários de drogas injetáveis ou hemotransfundidos) observa-se uma diminuição progressiva na sua proporção. Na primeira década (1980-1990) os usuários de drogas contribuíram com 15,1% dos casos e os indivíduos infectados por transfusões de sangue com 2,8%. No final de 2002 as proporções de casos nestas duas populações eram de, respectivamente, 2,6% e 0% (apenas 3 casos de transfundidos foram notificados neste ano). O controle das infecções através de transfusões de sangue se deve à introdução dos testes sorológicos para a detecção de anticorpos específicos contra o vírus, em 1985.

Apesar do primeiro caso de AIDS conhecido no Brasil ter ocorrido em 1980, o vírus HIV-1 foi isolado pela primeira vez no país somente em 1987. Esta primeira variante brasileira foi encontrada em um paciente com AIDS que havia recebido transfusão de sangue dois anos antes (Galvão-Castro et al., 1987). Desde então, foram desenvolvidos diversos estudos visando à identificação e caracterização molecular de vírus brasileiros. Com a classificação dos variantes de HIV-1 em subtipos de seqüências observou-se que a grande maioria das cepas brasileiras agrupava-se com as seqüências prototípicas do subtipo B.

Atualmente sabe-se que pelo menos cinco subtipos circulam no país, A, B, C, D e F, além de diversas formas virais recombinantes B/F (Gao et al., 1996; Morgado et al., 1994; Sabino et al., 1994), B/C (Guimarães et al., 2002) e B/D (Tanuri et al., 1999).

Contudo, o subtipo B permanece associado à grande maioria dos casos de infecção no país (Bongertz et al, 2000; Casseb et al., 1998; Couto-Fernandez et al., 1999; Covas et al., 1998; Morgado et al., 1998b; Pinto et al., 1998; Tanuri et al., 1999; Vicente et al., 2000), seguido do subtipo F (Couto-Fernandez et al., 1999; Pinto et al., 1998; Tanuri et al. 1999; Vicente et al., 2000). Recentemente tem sido observado um crescimento no número de infecções associadas ao subtipo C na região sul do país (Guimarães et al., 2002). Por outro lado, as infecções pelos vírus dos subtipos A e D parecem restritos a um pequeno número de casos (Guimarães et al., 2002; Morgado et al., 1998a; Stefasi et al., 2000).

1.2.4. A variante GWGR

Dentre as variantes virais que compõem a epidemia brasileira classificadas como subtipo B, uma grande proporção possui um padrão de assinatura particular no topo da alça V3 do gene *env*. Nestes vírus, o aminoácido prolina, encontrado na segunda posição do motivo protéico GPGX, extremamente conservado nos diferentes subtipos que circulam na epidemia mundial (Kuiken et al., 2000a), está substituído pelo triptofano gerando o padrão de assinatura GWGR. Nas seqüências prototípicas do subtipo B, o motivo protéico encontrado é o GPGR.

Na literatura, a variante GWGR é citada como subtipo B-Br, referência clara à sua presença no Brasil, ou subtipo B'', para diferenciá-lo do variante B' que circula na Tailândia e em outros países da Ásia, como já mencionado. Contudo, nas variantes asiáticas, a mutação que os caracteriza ocorreu no quarto aminoácido do motivo protéico e seu padrão de assinatura é GPGQ (Naganawa et al., 1997).

O primeiro relato sobre a existência de vírus com o padrão de assinatura GWGR foi feito em um estudo do Japão publicado em 1991 (Hattori et al., 1991). Naquele trabalho, foram analisadas amostras de 28 pacientes dos quais três estavam infectados por vírus GWGR. O ano de soro-conversão não é citado, mas os três estavam no estágio II da imunodeficiência segundo os critérios de classificação dos CDC (infecção assintomática). Os dados disponíveis no artigo referem que os três pacientes haviam adquirido a infecção por via sexual, um dos quais na América do Sul.

No ano de 1993, a variante GWGR foi descrita no Brasil por Potts e colaboradores (Potts et al., 1993). No trabalho, os autores mostraram que a variante

GWGR estava presente em 35% das seqüências do subtipo B analisadas (Potts et al., 1993). A seguir, os trabalhos desenvolvidos no Brasil para a caracterização molecular de amostras de HIV-1 mostraram que uma parcela importante dos casos de infecção pelo subtipo B no país era atribuída a vírus com a mutação P→W. Por exemplo, na cidade do Rio de Janeiro, Tanuri e colaboradores encontraram a variante GWGR em 37,5% de doadores de sangue infectados pelo subtipo B (Tanuri et al., 1999), enquanto em outro trabalho com amostras da mesma cidade a proporção foi de 41% (Morgado et al., 1998b). No Estado de São Paulo a variante GWGR estava presente em aproximadamente 50% dos casos de infecção por vírus do subtipo B (Casseb et al., 1998; Covas et al., 1998). Quanto aos outros estados do Brasil, a variante também foi identificada na Bahia (Potts et al., 1993) e em Goiás (Stefasi et al., 2000).

Em outros países, além do Brasil e do Japão, os vírus GWGR também foram encontrados em amostras da Argentina (Campodonico et al., 1996) e da Bolívia (Velarde-Dunois et al., 2000). Contudo, apesar da variante GWGR ter sido descrita em diversas regiões, sua origem e relevância para a pandemia do HIV-1 não são conhecidas. Uma possível explicação para a sua presença nos diversos países onde é encontrada é que seus representantes regionais tenham se originado por mutações convergentes na região da alça V3. Alternativamente, os vírus GWGR podem ter surgido uma única vez no contexto da epidemia e então se dispersado para as outras regiões.

2. Objetivos

O objetivo principal da tese de doutorado foi estudar a origem evolutiva e a dispersão da variante GWGR do subtipo B do HIV-1 encontrado no Brasil e em outros países do mundo. Para que tal objetivo fosse alcançado, as seguintes etapas precisaram ser cumpridas (objetivos secundários):

- 1) Conhecer a distribuição mundial da variante GWGR.
- 2) Conhecer a diversidade do HIV-1 no Brasil para se compreender a participação da variante GWGR na epidemia brasileira.
- 3) Analisar filogeneticamente todas as seqüências GWGR disponíveis no mundo, além de uma amostra representativa das seqüências GPGR.
- 4) Traçar o curso da epidemia GWGR no mundo.
- 5) Examinar os possíveis motivos para a disseminação da variante GWGR.

3. Material e Métodos

3.1. Amostragem

O objetivo principal da tese de doutorado foi estudar a origem evolutiva do variante GWGR do subtipo B encontrado na epidemia pelo HIV-1. Desta forma, o primeiro passo para o desenvolvimento do trabalho foi a busca e sistematização do conjunto de seqüências da região V3 com os motivos GWGR e GPGR que seriam utilizadas nas análises subseqüentes. Para isto, todas as seqüências GWGR e GPGR obtidas das diversas regiões geográficas e disponíveis no banco de seqüências eletrônico, ou Genbank, foram identificadas e selecionadas segundo critérios que serão expostos mais adiante neste texto (<http://www.ncbi.nlm.nih.gov>).

Também para o desenvolvimento do trabalho, era necessário conhecer a participação do variante GWGR na epidemia brasileira. Assim, foi preciso que todas as seqüências da região V3 do Brasil disponíveis no banco de dados e classificadas como subtipo B, independentemente do motivo protéico existente no topo da alça V3, fossem identificadas e analisadas.

Uma vez que as estratégias empregadas para a identificação das seqüências GWGR e GPGR mundiais e das seqüências brasileiras foram diferentes, estas serão expostas separadamente ao longo do texto.

3.1.1. Amostra primária

3.1.1.1. Seqüências do mundo

A primeira etapa do trabalho foi a identificação de todas as seqüências com o padrão de assinatura GWGR na região V3 e de todas as seqüências com o motivo parental GPGR na mesma região disponíveis no Genbank. Para isto, a estratégia utilizada foi o desenvolvimento de um *script* na linguagem Perl (Schwartz & Phoenix, 2001) capaz de reconhecer os dois padrões de assinatura nas seqüências traduzidas do gene (seqüências de ENV) sempre que estes motivos estivessem localizados entre dois resíduos de cisteína. A imposição dos resíduos de cisteína como delimitadores no *script* foi feita para aumentar a especificidade da busca. Como dito na introdução da tese, a região V3 forma uma alça determinada por pontes dissulfeto entre resíduos de cisteína extremamente conservados.

Por outro lado, uma vez que substituições dos aminoácidos glicina, glicina e arginina encontrados, respectivamente, na primeira, na terceira e na quarta posição dos tetrapeptídeos (GXGR) são descritas na literatura, a possibilidade de tais mutações existirem nas seqüências depositadas no banco de dados foi considerada no desenho do programa. Desta forma, foram criadas quatro variações para cada padrão de assinatura. O *script* básico e suas variações estão descritos a seguir:

Padrão de assinatura GPGR	Padrão de assinatura GWGR
CX(n)GPGRX(n)C	CX(n)GWGRX(n)C
CX(n)XPGRX(n)C	CX(n)XWGRX(n)C
CX(n)GPXRX(n)C	CX(n)GWXRX(n)C
CX(n)GPGXX(n)C	CX(n)GWGXX(n)C

No dia 20 de julho de 2001, todas as seqüências depositadas no Genbank foram retiradas no formato FASTA (Higgins et al, 1996) e o arquivo foi rastreado com o *script* descrito acima. A busca das seqüências no banco de dados mostrou que, das 26.374 seqüências disponíveis na data citada, 18.097 (68,6%) tinham o aminoácido prolina na segunda posição dos tetrapeptídeos permitidos no *script* enquanto que em 206 seqüências (0,8%) este sítio era ocupado por um resíduo de triptofano.

A partir das seqüências GPGR e GWGR identificadas no Genbank, uma amostra primária foi definida e analisada. Esta amostra primária incluiu todas as 206 seqüências GWGR, além de uma amostra sistemática de 400 seqüências com a assinatura GPGR, retirada do conjunto inicial das 18.097 seqüências encontradas no Genbank. Para que esta amostra fosse obtida, os números de acesso, únicos para cada seqüência depositada no banco de dados, foram ordenados de forma crescente e, a cada 45 seqüências, uma foi retirada (por exemplo, a primeira, a seguir a 46^a, a 91^a, e assim por diante). Esta amostra sistemática foi necessária uma vez que a análise de todas as seqüências GPGR, segundo os critérios que serão definidos mais adiante, seria extremamente trabalhosa e demorada.

Para todas as seqüências da amostra primária:

- 1) Todas as seqüências de nucleotídeos correspondentes foram selecionadas do Genbank.
- 2) Os dados referentes à classificação das seqüências em subtipos foram obtidos diretamente dos bancos de dados (Genbank e PubMed) ou dos artigos científicos onde as seqüências são referidas.
- 3) Foram utilizados os critérios de exclusão que serão posteriormente definidos no item 3.1.1.3 para a análise das seqüências.

3.1.1.2. Seqüências do Brasil

A segunda etapa do trabalho foi a análise da diversidade na região do gene *env* entre as seqüências do Brasil. A inclusão de todas as seqüências brasileiras nesta etapa foi feita para se melhor estudar a participação da variante GWGR na epidemia brasileira. Uma vez que um grande número de infecções no país é atribuído à variante GWGR, era necessário conhecer a proporção destes vírus entre todas as seqüências brasileiras de HIV-1 (região V3) depositadas no Genbank. Uma outra razão para a inclusão de todas as seqüências brasileiras no estudo é que, na literatura, além dos vírus GPGR e GWGR, outras variantes do subtipo B também são descritas no Brasil. Desta forma, desejou-se conhecer a extensão da diversidade dos vírus do subtipo B no país para se estudar as relações evolutivas entre estas outras variantes e os vírus GPGR e GWGR.

Para a identificação de todas as seqüências do gene *env* do Brasil disponíveis no Genbank foi utilizada uma estratégia diferente da empregada para as seqüências GWGR e GPGR mundiais. Neste caso, as seqüências foram identificadas diretamente com o auxílio das palavras-chave HIV-1, Brasil e *env*.

Os dados referentes à classificação das seqüências em subtipos, a região (estado ou cidade) de origem da seqüência, a data de coleta da amostra e a provável via de infecção do paciente foram obtidos diretamente do Genbank ou dos artigos correspondentes.

Em julho de 2001 havia 234 seqüências contendo a região V3 do gene *env* do Brasil depositadas no Genbank. Estas seqüências, definidas como a amostra primária

brasileira, foram analisadas segundo os critérios de exclusão definidos no item 3.1.1.3.

3.1.1.3. Seleção das seqüências

As seqüências de aminoácidos e de nucleotídeos com os motivos GPGR e GWGR da amostra primária do mundo e do Brasil foram analisadas segundo os critérios definidos a seguir:

1) Foram excluídas todas as seqüências de subtipos não-B do grupo M ou dos grupos O e N. Na ausência de informações quanto à classificação em subtipos, esta foi feita por análise filogenética. Para isto, seqüências prototípicas de todos os subtipos do grupo M e dos grupos O e N disponíveis no Laboratório Nacional de Los Alamos (<http://hiv-web.lanl.gov>) foram utilizadas como padrão. Uma seqüência de SIVcpz (*Simian Immunodeficiency Virus* de chimpanzés) foi escolhida como grupo externo. As seqüências com subtipo desconhecido, as seqüências prototípicas e o grupo externo foram alinhadas pelo programa ClustalW (Higgins et al., 1996; <http://www.ddbj.nig.ac.jp/htmls/E-mail/clustalw-e.html>) <http://www.ddbj.nig.ac.jp/htmls/E-mail/clustalw-e.html>. O alinhamento foi ajustado manualmente. Uma árvore de Neighbor-Joining foi construída com as seqüências alinhadas usando-se o modelo de distância de Kimura-2 parâmetros. A consistência dos ramos das árvores filogenéticas foi determinada pelo teste de *bootstrap* (Efron et al, 1996).

Tipicamente, as seqüências de diferentes subtipos formam ramos separados na árvore filogenética. Além disto, a distância genética entre os diferentes subtipos foi estimada entre 0,162 e 0,190. Dentro dos subtipos, a distância varia entre 0,054 e 0,155 (Triques et al., 2000). Desta forma, a classificação das seqüências em subtipos foi feita a partir de sua posição dentro dos ramos formados pelas seqüências prototípicas na árvore filogenética. As distâncias genéticas entre as seqüências que estavam sendo analisadas e as seqüências prototípicas com as quais estas se agruparam também foram calculadas para confirmar a classificação.

2) Todas as seqüências que continham fragmentos de outros subtipos (formas recombinantes) foram excluídas.

3) No caso de existir mais de uma seqüência obtida de um único indivíduo (clones), somente uma foi incluída.

4) Quando: a) mais de um dos tetrapeptídeos definidos pelo *script* foram localizados em uma única seqüência ou b) o peptídeo encontrado continha mutações na primeira, na terceira ou na quarta posição, ou seja, quando os respectivos aminoácidos não eram glicina (G), glicina e arginina (R), as seqüências de aminoácidos foram alinhadas com seqüências prototípicas do subtipo B (<http://hiv-web.lanl.gov>). O alinhamento foi feito com o programa ClustalW e ajustado manualmente. O tetrapeptídeo presente na crista da alça V3 do gene *env*, em cada seqüência alinhada, foi identificado analisando-se os sítios homólogos nas seqüências prototípicas. As seqüências que não continham algum dos tetrapeptídeos definidos pelo *script* descrito no item 3.1.1.1 foram excluídas.

5) Foram excluídos todos os pseudogenes, todas as seqüências com o motivo GPGR com menos de 50 aminoácidos (150 nucleotídeos) e as seqüências que continham múltiplos sítios de nucleotídeos sem informações definidas (*missing data*).

3.1.2. Amostra Secundária

3.1.2.1. Seqüências do mundo

A análise da amostra primária das seqüências GPGR e GWGR, feita segundo os critérios definidos no item 3.1.1.3, mostrou que, das 400 seqüências GPGR iniciais, 137 seqüências preenchiam tais critérios e foram então excluídas; dentre as 206 seqüências GWGR, 125 foram retiradas da análise da mesma forma. As 263 seqüências GPGR e as 81 seqüências GWGR restantes foram definidas como amostra secundária. A distribuição geográfica das seqüências GPGR e GWGR da amostra secundária está apresentada na tabela 1.

A tabela 1 mostra que a maioria das seqüências GPGR incluídas na amostra secundária foi obtida nos Estados Unidos, na Holanda, na França, na Itália e no Reino Unido, países onde o subtipo B é responsável pela maioria das infecções pelo HIV-1.

Tabela 1 Distribuição geográfica das seqüências do gene *env* do HIV-1 do subtipo B com os motivos GPGR e GWGR encontradas no Genbank

País	GPGR	GWGR	País	GPGR	GWGR
Alemanha	-	1	Haiti	2	-
Argentina	1	3	Havái	1	-
Austrália	8	-	Holanda	31	2
Barbados	1	-	Ilhas Reunidas	-	1
Bélgica	1	-	Indonésia	1	-
Bolívia	-	2	Itália	20	-
Brasil	-	54	Japão	11	1
Canadá	1	-	Lituânia	1	1
China	1	1	Malásia	1	-
Chipre	1	-	México	1	1
Coréia	3	-	Paraguai	-	1
Dinamarca	2	1	Reino Unido	14	-
Djibuti	1	-	República Checa	-	2
Espanha	4	1	Rússia	3	-
Estados Unidos	115	1	Suécia	8	-
Estônia	1	-	Suíça	4	1
Etiópia	1	-	Tailândia	4	-
Filipinas	-	3			
França	20	4	Total	263	81

Contudo, somente duas seqüências dos HIV circulantes nas epidemias regionais da América Latina (Argentina e México) foram incluídas na amostra secundária. Deve ser ressaltado que nos países desta região existe um grande número de pessoas vivendo com HIV/AIDS e a maior parcela dos casos são atribuídos ao subtipo B (Bobkov et al., 1998; Fernandez-Medina et al., 1999; Halani et al., 2001; Velarde-Dunois et al., 2000). A razão para esta distorção é o fato de que, na verdade, o banco de dados representa uma amostra viciada da diversidade do HIV-1, ditada por interesses científicos relacionados às características particulares de cada epidemia ou pelos recursos econômicos oferecidos para a pesquisa da infecção em cada país.

A distribuição geográfica das seqüências GWGR (tabela 1) mostra que a variante participa da epidemia em pelo menos 18 países. No continente americano, além do Brasil, a variante também está na Argentina, na Bolívia, no México, no Paraguai e

nos Estados Unidos. Na Europa, vírus B com a assinatura GWGR circulam na Holanda, França, Suíça, Alemanha, Dinamarca, Espanha, República Checa e Lituânia. Dentre os países asiáticos, a variante foi detectada na China, no Japão, nas Filipinas e nas Ilhas Reunidas.

Contudo, como pode ser observado na tabela 1, o maior número de seqüências GWGR foi identificado no Brasil: das 81 seqüências GWGR da amostra secundária, 54 são brasileiras. Estes resultados sugerem que, apesar dos vírus GWGR serem encontrados em três continentes, o Brasil representa o principal foco da epidemia por esta variante.

3.1.2.2. Seqüências do Brasil

Do grupo inicial de 234 seqüências brasileiras, 55 seqüências foram excluídas por não preencherem os critérios de seleção definidos no item 3.1.1.3. A distribuição geográfica das 179 seqüências brasileiras restantes está apresentada na tabela 2 segundo a sua classificação em subtipos.

Tabela 2. Número de casos de AIDS registrados em cada estado até o ano de 2002 e distribuição das seqüências do gene *env* do HIV-1 segundo subtipos e formas recombinantes obtidas de amostras do Brasil.

Estado	Nº de Casos	Subtipos e Formas Recombinantes* (FR)							Total
		Subtipo B**			outros				
		B-GPGR	B-GWGR	B-outros	C	D	F	FR	
São Paulo (SP)	117.993	6	8	-	2	-	2	1	19
Rio de Janeiro (RJ)	36.462	31	37	14	1	1	5	12	101
Rio Grande do Sul (RS)	21.611	-	-	-	1	-	-	-	1
Minas Gerais (MG)	15.438	6	1	-	-	-	3	-	10
Paraná (PR)	11.504	-	-	1	-	-	-	-	1
Bahia (BA)	5.423	9	1	1	-	-	1	-	12
Amazonas (AM)	1.532	1	-	-	-	-	1	4	6
Desconhecido	-	11	7	8	2	-	-	1	29
Total		64	54	24	6	1	12	18	179

* Todas as formas recombinantes foram agrupadas juntas independentemente dos subtipos parentais.

** O subtipo B segundo o padrão de assinatura encontrado no topo da alça V3.

Nessa tabela, os vírus do subtipo B foram separados em três grupos, B-GPGR, B-GWGR e B-outros. Neste último grupo foram incluídas todas as variantes do subtipo

B que apresentavam outros motivos protéicos no topo da alça V3 (por exemplo, GFGR, GLGR, GGGR, etc) encontradas no Genbank.

Como “formas recombinantes” (FR) foram agrupadas na tabela todas as seqüências de *env* obtidas de vírus dos quais mais de um segmento do genoma havia sido seqüenciado e onde a análise destes fragmentos havia mostrado que pertenciam a diferentes subtipos. Desta forma, todos os vírus recombinantes foram agrupados juntos, sem que padrões específicos de recombinação fossem considerados. Estes vírus recombinantes foram excluídos das análises subseqüentes, mesmo quando todo o segmento do gene *env* (ou da região V3) era classificado como subtipo B.

Como pode ser observado na tabela 2, a maioria das seqüências brasileiras encontradas no Genbank, 142 (79,3%), foi classificada como subtipo B. Dentre as seqüências do subtipo B, 64 apresentavam o motivo GPGR no topo da alça V3 enquanto em 54 seqüências o motivo GWGR foi encontrado.

Uma vez que nenhum trabalho para a caracterização da diversidade do HIV-1 no Brasil objetivou a identificação específica de uma das duas variantes do subtipo B em detrimento da outra, não é provável que tenha havido vício para a obtenção das seqüências GWGR ou GPGR. Desta forma, o número semelhante de seqüências GWGR e GPGR do Brasil encontradas no Genbank comprova as observações feitas por diversos autores de que as duas variantes circulam em proporções semelhantes na epidemia brasileira (Casseb et al., 1998; Covas et al., 1998; Michael-Hendry et al., 1996; Tanuri et al., 1999).

Ainda dentro do subtipo B, 24 seqüências apresentaram outros motivos protéicos no topo da alça V3. Estas variantes serão estudadas separadamente no próximo capítulo: “Resultados”.

Além do subtipo B, também foram encontradas no Genbank seqüências brasileiras classificadas como subtipos F (6,7%, n=12), C (3,6%, n=6), D (0,5%, n=1) e formas recombinantes B/F e B/D (10,0%, n=18). Como descrito na literatura (Tanuri et al., 1999; Bongertz et al., 2000), o subtipo F é o segundo mais comumente encontrado no Brasil. Quanto ao subtipo C, apesar de sua participação crescente na epidemia da região Sul do Brasil (Guimarães et al., 2002), até a data da busca, havia somente seis seqüências deste subtipo depositadas no Genbank, das quais duas não possuíam referência quanto à sua origem geográfica.

Também como descrito na literatura, os subtipos A e D parecem representar variantes raras no Brasil (Guimarães et al., 2002; Morgado et al., 1998). Somente uma seqüência do subtipo D tinha sido depositada no Genbank até o momento da busca e nenhuma seqüência A foi encontrada.

Para uma melhor caracterização da epidemia de HIV-1 no Brasil, o número de casos acumulados de AIDS para cada estado até o final de 2002 (MS, 2002) também foi incluído na tabela 2. Além disto, a figura 5 mostra a distribuição dos casos de AIDS e das seqüências do gene *env* nas diferentes regiões do país.

Como pode ser visto na tabela 2 e na figura 5, as seqüências brasileiras foram obtidas nos estados do Rio de Janeiro (n=101), São Paulo (n=19), Bahia (n=12), Minas Gerais (n=10), Amazonas (n=6), Paraná (n=1) e Rio Grande do Sul (n=1). Das 29 seqüências restantes não foi possível descobrir sua origem geográfica.

Até julho de 2001, não havia seqüências do gene *env* relacionadas aos outros estados do Brasil depositadas no Genbank. O fato de somente sete estados estarem representados na figura 5 e na tabela 2 mostra que, na realidade, o conhecimento da diversidade das cepas de HIV-1 que circulam no Brasil é pequeno e direcionado para os vírus encontrados na região sudeste do país.

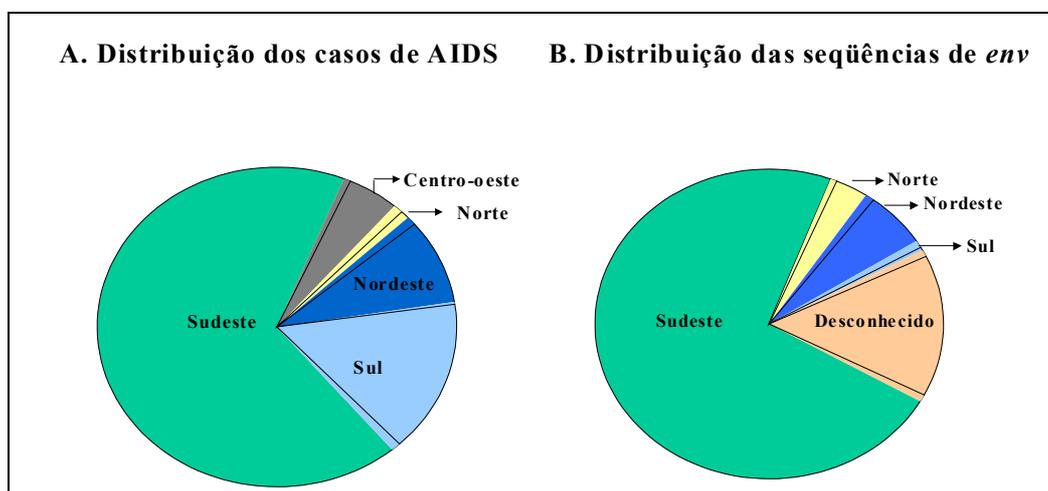


Figura 5. Distribuição dos casos de AIDS (A) e das seqüências do gene *env* (B) nas regiões do Brasil

A região sudeste tem o maior número de casos acumulados de AIDS do Brasil, 67,3% do total (figura 5A), e representa a maior proporção das seqüências brasileiras do

gene *env* depositadas do Genbank (72,6%) (figura 5B). Entretanto, mesmo quando comparamos as duas maiores cidades brasileiras, Rio de Janeiro e São Paulo (ambas localizadas na região sudeste), a distribuição das seqüências não corresponde à magnitude das epidemias locais: enquanto São Paulo concentra 45,5% dos casos, somente 10,6% das seqüências brasileiras representam sua epidemia (Bongertz et al., 2000; Couto-Fernandez et al., 1994; Covas et al., 1998; Gao et al., 1996; Gao et al., 1998; Morgado et al., 1994; Sabino et al., 1996; Sabino et al., 1994). Por outro lado, 56,4% das seqüências de todos os subtipos vieram do Rio de Janeiro (Bongertz et al., 2000; Couto-Fernandez et al., 1994; Gao et al., 1998; Janini et al., 1998; Morgado et al., 1998a; Morgado et al., 1994; Ramos et al., 1999; Sabino et al., 1994; Tanuri et al., 1999), onde foram reportados 14,9% dos casos de AIDS.

Ainda na região sudeste, 5,6% das seqüências disponíveis foram colhidas em Minas Gerais, com 5,7% dos casos. Nenhuma seqüência do Espírito Santo foi encontrada no Genbank.

Pouco se sabe sobre as variantes do vírus que circulam nas demais regiões do Brasil. Duas seqüências representam a epidemia na região sul, uma classificada como subtipo B, do Paraná (Couto-Fernandez et al., 1994), e outra como subtipo C, do Rio Grande do Sul (Gao et al., 1996), estados com, respectivamente, 4,5% e 7,5% dos casos notificados. A região norte contribuiu com seis seqüências do Amazonas, 3,4% do total, das quais quatro são formas recombinantes (Vicente et al., 2000). Menos de 2% dos casos de AIDS foram descritos na região e 0,6% no estado. O estado da Bahia, com 2,3% dos casos de AIDS, é o único da região nordeste a ter seqüências no banco de dados (Couto-Fernandez et al., 1994; Couto-Fernandez et al., 1999; Potts et al., 1993); na região nordeste foram notificados 9,4% dos casos. Nenhuma seqüência do gene *env* da região centro-oeste do Brasil estava depositada no Genbank na data da busca.

Como visto acima, as informações sobre a diversidade do HIV-1 no Brasil estão limitadas a poucas regiões geográficas. Esta observação é importante, uma vez que a variabilidade do vírus é um desafio para o desenvolvimento de vacinas contra o HIV-1 (Osmanov et al. 2002).

Devido à relevância da diversidade do HIV-1, em 1993 o Ministério da Saúde criou no Brasil a “Rede Nacional para o Isolamento e Caracterização do HIV”. Os principais objetivos da Rede Nacional são: 1) desenvolver um sistema para a vigilância contínua da variabilidade entre os isolados de HIV-1 que circulam nas diferentes

regiões geográficas do país, 2) conhecer as propriedades genéticas e antigênicas das cepas relevantes para a epidemia brasileira e 3) participar dos esforços internacionais para a caracterização da variabilidade do HIV-1 (Galvão-Castro et al., 1996). Esta rede nacional foi implementada nos moldes da “Rede para o Isolamento e Caracterização do HIV”, criada pela Organização Mundial de Saúde (OMS) também no início da década de 90 (Who Network for HIV Isolation and Characterization, 1994).

Contudo, das 179 seqüências brasileiras encontradas no Genbank durante o desenvolvimento do trabalho de tese, somente 24 (13%) haviam sido geradas pela Rede Nacional (Bongert et al., 2000). As demais seqüências resultaram de trabalhos desenvolvidos por grupos regionais de pesquisadores.

3.1.3. Amostra final

A partir da amostra secundária das seqüências GPGR e GWGR encontradas na epidemia mundial e das seqüências brasileiras, um subgrupo foi escolhido para a análise filogenética. Este subgrupo foi definido como amostra final (tabela 3).

A amostra final incluiu:

a) quando disponível, duas seqüências com o motivo GPGR de cada país onde a variante foi encontrada; no caso de mais de duas seqüências de um mesmo país estarem disponíveis, foram escolhidas as seqüências mais longas e, então, as mais antigas, segundo a data de coleta da amostra.

b) todas as seqüências GWGR do mundo.

c) 55 seqüências do Brasil classificadas como subtipo B com o motivo GPGR no topo da alça V3. Como pode ser visto, nove seqüências brasileiras com o motivo GPGR foram excluídas da análise. A razão para isto é que tais seqüências continham menos de 150 nucleotídeos ou apresentavam múltiplos sítios de nucleotídeos sem informações precisas (*missing data*).

Tabela 3 Distribuição geográfica das seqüências de env do subtipo B com os motivos GPGR e GWGR incluídas na árvore filogenética. A Tabela apresenta o ano de identificação da amostra mais antiga incluída na análise.

País	GPGR Ano (n)	GWGR Ano (n)	País	GPGR Ano (n)	GWGR Ano (n)
Alemanha	-	? (1)	Haiti	1992 (2)	-
Argentina	1991 (1)	1992 (3)	Havaí	1995 (1)	-
Austrália	1992 (2)	-	Holanda	1987 (2)	1991 (2)
Barbados	1992 (1)	-	Ilhas Reunidas	-	1996 (1)
Bélgica	1998 (1)	-	Indonésia	1993 (1)	-
Bolívia	-	1996 (2)	Itália	? (2)	-
Brasil	1988 (55)	1988 (54)	Japão	1991 (2)	1997 (1)
Canadá	? (1)	-	Lituânia	1996 (1)	1992 (1)
China	1991 (1)	1996 (1)	Malásia	1992 (1)	-
Chipre	1994 (1)	-	México	1998 (1)	1998 (1)
Coréia	1989 (2)	-	Paraguai	-	1995 (1)
Dinamarca	1996 (2)	1996 (1)	Reino Unido	1990 (1)	-
Djibuti	1988 (1)	-	República Checa	-	1997 (2)
Espanha	1999 (2)	1993 (1)	Rússia	1997 (2)	-
Estados Unidos	1983 (2)	1999 (1)	Suécia	1992 (2)	-
Estônia	1999 (1)	-	Suíça	1991 (2)	1992 (1)
Etiópia	? (1)	-	Tailândia	1994 (2)	-
Filipinas	-	1991 (3)			
França	? (2)	1993 (4)	Total	98	81

Além das seqüências GPGR e GWGR, a amostra final incluiu também o fragmento correspondente à alça V3 da amostra ZR59 obtida da República Democrática do Congo (antigo Zaire). Este fragmento representa a seqüência mais antiga de HIV-1, obtida de uma amostra colhida em 1959. Segundo os seus autores, esta seqüência se localiza próximo ao nó ancestral dos subtipos B, D e F (Zhu et al., 1998) na árvore filogenética. Desta forma, o fragmento ZR59 foi incluído, pois se esperava que representasse a seqüência ancestral de todas as variantes do subtipo B.

Contudo, para que fosse possível estudar a direção da evolução das seqüências GPGR e GWGR ao longo do tempo sem precisar pressupor a seqüência ZR59 como original, escolheu-se construir uma árvore filogenética com raiz, utilizando-se para isto um grupo externo de referência. O grupo externo de referência é um taxon (unidade taxonômica, que no trabalho de tese são seqüências de DNA) para o qual há informações prévias indicando que ele divergiu antes daqueles incluídos na análise filogenética. Neste caso, para grupo externo, foi escolhida uma seqüência prototípica do

subtipo A. Este subtipo foi incluído como grupo externo devido à clara distinção entre os subtipos A e B nas análises filogenéticas. A seqüência escolhida foi uma das seqüências de referência disponíveis no Laboratório Nacional de Los Alamos (<http://hiv-web.lanl.gov>) (número de acesso AF069670).

3.2. Alinhamento das seqüências

Para que a reconstrução filogenética das seqüências de DNA selecionadas no trabalho pudesse ser feita, era necessário que estas seqüências fossem alinhadas.

O objetivo do alinhamento é mover seqüências de tamanhos diferentes para se obter o maior número possível de pares de sítios homólogos de nucleotídeos. De um modo geral, se as seqüências homólogas tivessem todas o mesmo tamanho, elas poderiam ser diretamente comparadas, visando a identificação dos sítios onde ocorreram substituições de nucleotídeos. Contudo, uma vez que, freqüentemente, as seqüências possuem números diferentes de nucleotídeos, a sua comparação envolve a identificação dos locais onde devem ter ocorrido inserções e deleções desde que divergiram de um ancestral comum. Desta forma, um alinhamento representa uma hipótese específica sobre a evolução das seqüências em questão (Li, 1997).

Durante o processo de alinhamento, tanto as inserções como as deleções introduzem intervalos nas seqüências de DNA. Como muitas vezes é impossível saber se houve a inserção de nucleotídeos em um grupo de seqüências ou se ocorreu deleção nas outras, os intervalos inseridos durante o alinhamento são chamados de “indels” (*insertions/deletions*).

Quando a divergência entre as seqüências é pequena, as posições dos intervalos podem ser facilmente inferidas por inspeção visual. Contudo, quando a divergência entre as seqüências é grande ou quando muitas seqüências devem ser alinhadas simultaneamente, o seu alinhamento necessita ser feito por programas de computador. Apesar de diversos métodos terem sido desenvolvidos para o alinhamento de duas ou mais seqüências, o princípio geral do alinhamento é o mesmo e será descrito a seguir (Nei & Kumar, 2000a).

Para simplificar, vamos considerar inicialmente que somente duas seqüências precisam ser alinhadas. Uma forma simples de se fazer isto é através de uma comparação bidimensional, utilizando-se uma matriz de pontos (figura 6).

	A	T	G	A	G	C	A	A	T
A	
C						.			
G			.		.				
A	
C						.			
T		.							.
A	
T		.							.

Figura 6. Exemplo de matriz de pontos usada no alinhamento de seqüências.

Nesta matriz, os pontos são colocados quando os nucleotídeos nas duas seqüências são idênticos. Isto significa que, se as duas seqüências forem iguais, uma linha diagonal de pontos será traçada na matriz. No caso das seqüências diferirem devido à presença de um intervalo em uma delas, a linha diagonal irá “saltar” no meio da linha, o que permitirá a identificação do intervalo (Nei & Kumar, 2000a).

Assim, um alinhamento entre duas seqüências é definido como uma seqüência ordenada de pares de nucleotídeos. Quando não há inserções ou deleções, cada par contém um nucleotídeo de cada seqüência. Por outro lado, se um intervalo foi introduzido em uma das seqüências, o par conterá um nucleotídeo e um elemento nulo (-).

Uma vez que, além dos intervalos, as seqüências também diferem devido a substituições de nucleotídeos, muitas vezes a identificação dos intervalos é uma tarefa difícil. Além disto, freqüentemente é necessário o alinhamento de mais de duas seqüências. Por esta razão diversos métodos matemáticos e programas computacionais foram propostos para a construção de alinhamentos de múltiplas seqüências. Dentre os

métodos matemáticos, o mais popular utiliza um índice de similaridade para a comparação de cada par de seqüências. Segundo este método, o alinhamento escolhido é aquele que maximiza o índice de similaridade.

Para o cálculo do índice de similaridade, considera-se o número de pares com elementos iguais (α), o número de pares com elementos diferentes (β) e o número de intervalos (γ) introduzidos durante cada possível alinhamento, independentemente do tamanho dos intervalos. A partir destes parâmetros a distância entre as duas seqüências pode então ser medida por:

$$E = \text{Min}(w_1\beta + w_2\gamma)$$

onde $\text{Min}(\cdot)$ representa o valor mínimo de $w_1\beta + w_2\gamma$ entre todos os alinhamentos possíveis, w_1 é a penalidade atribuída a pares diferentes e w_2 é a penalidade para um intervalo.

Contudo, a premissa de que a penalidade para um intervalo (w_2) é a mesma, independentemente de seu tamanho é irreal. Assim, é comum pressupor que a penalidade de um intervalo é uma função de seu comprimento. Do mesmo modo, como os pares com diferentes nucleotídeos podem ocorrer devido a transições ou transversões, diferentes penalidades (w_1) também podem ser atribuídas às duas situações. Com isto, a escolha do melhor alinhamento depende muito dos valores relativos de w_1 e w_2 .

Quando se escolhe um valor para w_1 maior do que para w_2 , o alinhamento com o maior índice de similaridade terá muitos intervalos e poucos pares de nucleotídeos diferentes. Contudo, uma vez que, na natureza, as inserções e deleções ocorrem com menos freqüência do que as substituições de nucleotídeos, o alinhamento escolhido nesta situação poderá ser irreal. Desta forma, é aconselhável atribuir um valor superior a w_2 do que a w_1 (Nei & Kumar, 2000a).

O alinhamento das seqüências incluídas na análise filogenética foi feito com o programa ClustalW (www.ddbj.nig.ac.jp/htmls/E-mail/clustalw-e.html), que utiliza o índice de similaridade descrito. O ClustalW é um programa computacional desenvolvido para o alinhamento múltiplo progressivo de um conjunto de seqüências de nucleotídeos ou de aminoácidos com diferentes graus de diversidade.

O alinhamento múltiplo progressivo parte de suposições prévias sobre as relações filogenéticas das seqüências determinadas a partir da sua ordem de ramificação em uma árvore filogenética inicial (Higgins et al, 1996; <http://bimas.dcert.nih.gov/clustalw/clustalw.html>). O algoritmo básico do alinhamento múltiplo consiste de três estágios principais: inicialmente, todas as seqüências são alinhadas par a par para o cálculo de uma matriz de similaridade, que considera o número de resíduos idênticos entre as duas seqüências, uma penalidade para cada intervalo introduzido e outra penalidade para cada par de nucleotídeos diferentes, como já mencionado.

A seguir, árvores de Neighbor-Joining (NJ) são calculadas a partir da matriz de similaridade com o objetivo de guiar o processo final de alinhamento. Finalmente, grupos de seqüências obtidos pelas árvores de NJ são progressivamente alinhados de acordo com a ordem de ramificação das árvores-guia. Neste processo final, os pares de seqüências com as menores distâncias são alinhados primeiro, e o alinhamento de seqüências cuja relação é mais distanciada é feito progressivamente, para grupos cada vez maiores. Quando um intervalo é introduzido, ele é colocado em todas as seqüências do grupo e os intervalos introduzidos nos alinhamentos prévios são mantidos fixos (Higgins et al, 1996; <http://bimas.dcert.nih.gov/clustalw/clustalw.html>). Desta forma, no algoritmo do alinhamento múltiplo progressivo, a ordem na qual as seqüências são alinhadas é crucial (Li, 1997).

Apesar do algoritmo do alinhamento múltiplo progressivo ser rápido e fornecer resultados razoáveis, não há garantia de que o melhor alinhamento foi encontrado (Nei & Kumar, 2000a). Desta forma, os alinhamentos construídos pelos programas de computador que empregam este algoritmo, como o ClustalW, devem ser sempre inspecionados e freqüentemente precisam ser ajustados manualmente.

No caso das seqüências do trabalho de tese, após a inspeção visual do alinhamento produzido pelo programa ClustalW, este alinhamento inicial foi manualmente ajustado para se obter um produto final melhor. Para isto, uma das seqüências incluídas no alinhamento foi escolhida como “guia”, para a qual o alinhamento da seqüência de aminoácidos correspondente também estava disponível.

Inicialmente, os intervalos introduzidos nesta seqüência “guia” durante o alinhamento feito pelo programa ClustalW foram reposicionados considerando-se: 1) a formação dos códons para os aminoácidos encontrados na seqüência traduzida, disponível no artigo, para que esta se mantivesse na sua fase aberta de leitura e 2) a

existência de intervalos no alinhamento original desta seqüência. A seguir, os intervalos das demais seqüências foram reajustados, considerando-se aqueles da seqüência “guia”.

Durante o ajuste do alinhamento, sempre que havia dúvidas sobre a localização de um intervalo, as seqüências de nucleotídeos eram traduzidas pelo programa MEGA2 (<http://www.megasoftware.net>; Nei & Kumar, 2000). Uma razão para este procedimento é que o alinhamento das seqüências de aminoácidos geralmente produz resultados melhores do que o alinhamento de nucleotídeos. Isto ocorre porque os aminoácidos são mais conservados na natureza, uma vez que muitas mutações nas seqüências de nucleotídeos são sinônimas, ou seja, não resultam na substituição de aminoácidos. Além disto, existem 20 aminoácidos e somente quatro nucleotídeos, o que faz com que a probabilidade de que dois sítios não homólogos sejam idênticos pelo acaso seja menor entre as seqüências de aminoácidos do que entre as seqüências de nucleotídeos (Nei & Kumar, 2000a).

Uma outra razão para a tradução das seqüências durante o processo de ajuste do alinhamento é a identificação de códons de terminação introduzidos no alinhamento inicial. Obviamente, isto só se aplica no caso de seqüências codificadoras de proteínas. Desta forma, quando o alinhamento de uma seqüência codificadora contém códons de terminação em seu interior, isto significa que os intervalos foram mal posicionados e que o alinhamento deve ser revisto.

No caso do trabalho de tese, a região do genoma estudada codifica uma estrutura do envelope viral e todos os pseudogenes (genes com códons de terminação) foram previamente excluídos. Desta forma, a presença de códons de terminação introduzidos durante o alinhamento inicial auxiliou o seu ajuste. O alinhamento final das seqüências é apresentado no Anexo I.

3.3. Distância genética

O programa MEGA2 (<http://www.megasoftware.net>; Nei & Kumar, 2000a) foi utilizado para a análise das seqüências alinhadas. No grupo das 180 seqüências do subtipo B, 285 sítios foram analisados para o cálculo de suas distâncias evolutivas. Estas distâncias foram computadas ignorando-se os intervalos, introduzidos durante o alinhamento, que envolviam apenas cada par de seqüências comparadas, ou seja, as deleções foram consideradas par a par. Devido à presença de diversos intervalos nas seqüências incluídas na análise, para somente 72 sítios de nucleotídeos havia informações em todas as seqüências. Desta forma, se somente estes sítios tivessem sido

utilizados para o cálculo das distâncias genéticas (deleção completa), as informações evolutivas contidas na maioria dos sítios seriam perdidas.

As seqüências alinhadas foram analisadas para a escolha do modelo de distância que seria utilizado na análise filogenética. O primeiro passo foi o cálculo da distância p , que representa a proporção de sítios com nucleotídeos diferentes entre todos os nucleotídeos analisados ($p = n_d/n$, onde n_d = número de sítios diferentes e n = número de sítios estudados). A distância p média entre todas as seqüências foi 0,139, com erro padrão (SE) de 0,009. Como pode ser visto na figura 7, a distância p calculada foi inferior a 0,2, o que poderia sugerir o uso desta distância na construção da árvore.

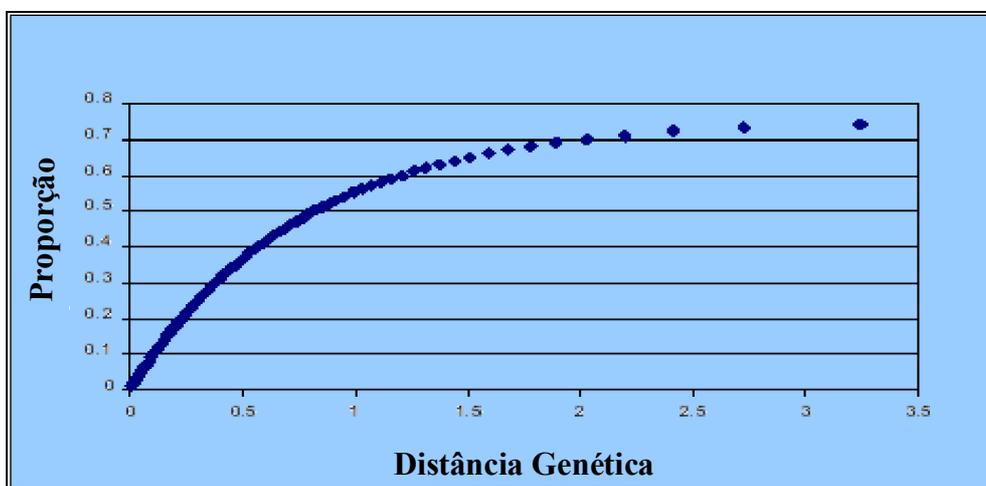


Figura 7. Relação entre distância p estimada (eixo x) e proporção real de sítios com substituições de nucleotídeos (eixo y)

Contudo, deve-se levar em conta que o segmento do gene *env* escolhido para a análise possui regiões hipervariáveis, nas quais múltiplas substituições podem ter ocorrido. Assim, o valor de p encontrado, mesmo sendo relativamente pequeno, pode refletir uma média entre regiões variáveis e conservadas, caracterizando uma subestimativa do valor verdadeiro. Neste caso, a distância p não é uma boa escolha para o cálculo das distâncias genéticas entre as seqüências (Nei & Kumar, 2000a).

Uma vez que a distância p não deve ser aplicada às seqüências em questão, torna-se necessário escolher o modelo de distância genética mais adequado. O modelo mais simples que corrige para múltiplas substituições em sítios de nucleotídeos é o modelo de Jukes-Cantor (Nei & Kumar, 2000a). Contudo, este modelo não é a melhor escolha quando existem vícios em direção às transições ou desvios na composição de

nucleotídeos. Desta forma, o passo seguinte foi calcular a proporção entre transições e transversões nas seqüências (R) e o seu conteúdo de nucleotídeos (conteúdo de GC).

A razão encontrada entre as transições e as transversões (R) foi 1,540, ou seja, superior ao esperado (0,5) caso não houvesse vício em direção às transições. Em relação à composição de nucleotídeos das seqüências, as proporções médias foram 21,6%, para timina (T), 14,6% para citosina (C), 46,0% para adenina (A) e 17,8% para guanina (G). O conteúdo médio de G+C foi 32,4%. Como pode ser visto, observou-se um vício em direção à incorporação de adenina; este vício já havia sido descrito na literatura para as seqüências de HIV-1 (Kypr & Mrázek, 1987).

A existência de desvios no valor de R e no conteúdo de nucleotídeos, quando não corrigidos por um modelo de distância adequado, pode resultar em estimativas erradas das distâncias genéticas entre as seqüências (Nei & Kumar, 2000a). No caso das seqüências incluídas no trabalho, os valores de R e de G+C indicam que as mutações acumuladas nestas seqüências eram viciadas, tanto em direção às transições como em relação à probabilidade de substituição entre as bases. Desta forma, o modelo de distância de Tamura 3-parâmetros foi escolhido para o cálculo das distâncias genéticas, uma vez que corrige para os dois vícios observados (Nei & Kumar, 2000a; Tamura, 1992).

3.4. Análise filogenética

O método de Neighbor-joining (NJ) escolhido para a construção da árvore filogenética é um dos métodos de distância desenvolvidos para a reconstrução de filogenias. De modo geral, nos métodos de distância, a árvore filogenética é construída considerando as relações entre valores de distâncias evolutivas computadas para todos os pares de seqüências incluídos na análise. Estas distâncias evolutivas são calculadas segundo o modelo de distância escolhido como o mais adequado para cada análise (no caso da tese, o modelo de Tamura-3). Após estas distâncias serem calculadas para cada par de seqüências, os valores encontrados são dispostos em uma matriz de distância que é utilizada para a construção da árvore.

O método de NJ utiliza o princípio de evolução mínima, segundo o qual a melhor árvore é aquela cuja topologia possui a menor soma total de ramos internos (S). No método de NJ, um algoritmo heurístico é utilizado para a construção da árvore, e este algoritmo é denominado agrupamento de vizinhos. Desta forma, um conceito

importante no método de NJ é o conceito de “seqüências vizinhas”, que são definidas como duas seqüências que são conectadas por um único nó. A partir deste conceito, é possível definir a topologia de uma árvore através da união sucessiva de “seqüências vizinhas”.

A construção de uma árvore de NJ começa com a produção de uma árvore em estrela, que é construída sob a premissa de que não há qualquer agrupamento entre as seqüências (figura 8A).

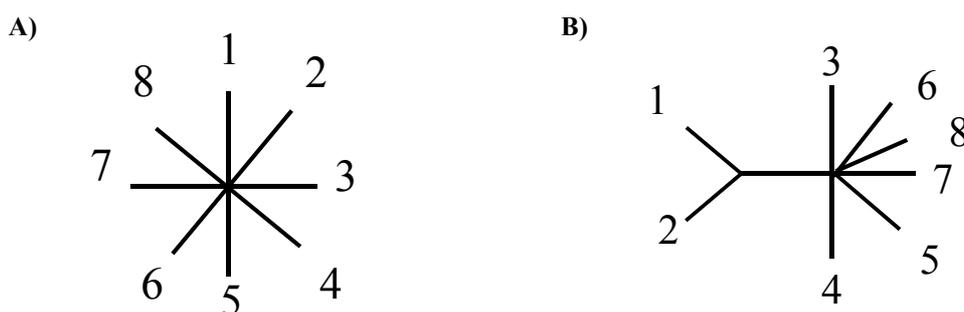


Figura 8. Duas etapas do algoritmo para a construção de árvores de NJ

Nesta árvore em estrela, os tamanhos dos ramos internos são estimados e a soma de todos os ramos é computada. A seguir, todos os pares de seqüências são agrupados como possíveis vizinhas e a soma do tamanho dos ramos para cada par é calculada. Então o par de seqüências com a menor soma de tamanho de ramos é escolhido e estas seqüências são consideradas como “vizinhas” (figura 8B). Uma vez que um par de seqüências vizinhas é identificado, elas são combinadas como um taxon composto e é criado um nó que conecta estas duas seqüências.

O passo seguinte é calcular a distância entre o novo nó e as seqüências restantes. A partir deste processo, uma nova matriz de distância é produzida, a partir da qual uma nova matriz de tamanhos de ramos é calculada. Nesta última, para o par previamente identificado de seqüências, será considerado o valor médio dos tamanhos dos ramos internos. A seguir, para se encontrar um novo par de seqüências vizinhas, escolhe-se aquele com a menor soma de ramos totais. Este processo é repetido sucessivamente até que todas as seqüências sejam agrupadas em uma única árvore. Esta árvore final é a árvore de NJ.

Dentre os métodos disponíveis para a construção de árvores filogenéticas, o

método de NJ é uma boa escolha quando se deseja analisar um grande número de seqüências (Nei & Kumar, 2000b). Neste trabalho de tese, 181 seqüências foram incluídas na análise. A razão para isto é que o tempo computacional despendido por este método é pequeno, apesar de ser um método extremamente eficiente na recuperação de árvores verdadeiras em estudos de simulação (Nei& Kumar, 2000c).

Além do curto tempo computacional e a eficiência em recuperar a árvore correta, uma outra vantagem do método de NJ é a existência de métodos estatísticos bem estabelecidos para se testar as árvores produzidas. (Estes métodos estatísticos também são aplicados aos outros métodos de distância). Dois exemplos destes métodos estatísticos são o *bootstrap* e o teste do ramo interno. No trabalho de tese, o teste do ramo interno foi escolhido para se testar a consistência da árvore construída com as seqüências GPGR e GWGR. Os princípios deste teste e a razão para a sua escolha serão expostos a seguir.

3.5. Teste do ramo interno

O teste do ramo interno, desenvolvido para árvores construídas pelos métodos de distância, como as árvores de NJ, examina a confiabilidade de cada ramo interno de uma determinada topologia. O princípio subjacente a este teste é que, se um determinado ramo tem comprimento não significativamente superior a zero, então este ramo pode representar uma ramificação incorreta da árvore filogenética e não deve ser considerado.

Assim, para a execução deste teste, em primeiro lugar, são calculadas as estimativas não viciadas das distâncias evolutivas entre todos os pares de seqüências incluídas na análise. As medidas encontradas são dispostas em uma matriz de distância. A partir desta matriz, estima-se a soma dos comprimentos dos ramos da árvore. A seguir, a probabilidade de confiança de cada ramo interno é estimada, testando-se a hipótese nula de que este ramo tenha comprimento zero. Para se testar esta hipótese utiliza-se o teste Z , pressupondo-se que os tamanhos dos ramos calculados têm distribuição normal. A hipótese nula é rejeitada com um nível de confiança α se $|Z| > Z_{\alpha/2}$.

$\alpha/2$.

Apesar do teste de *bootstrap* ser amplamente utilizado para se testar a confiabilidade de árvores filogenéticas construídas com seqüência de HIV-1, neste

trabalho de tese o teste do ramo interno foi escolhido para o mesmo fim. A razão para isto é que o teste de *bootstrap*, quando aplicado a árvores construídas por métodos de distância, tende a oferecer estimativas mais conservadoras da confiança dos tamanhos de ramos do que o teste do ramo interno. Além disto, foi sugerido que, para árvores de NJ, o teste do ramo interno pode oferecer um suporte estatístico melhor do que o teste de *bootstrap* (Russo, 1997).

4. Resultados

4.1. Origem da variante GWGR

Uma árvore filogenética foi construída com o método de NJ para se estudar a origem evolutiva da variante GWGR (figura 9 e anexo II). A análise da topologia da árvore mostra que todas as seqüências GWGR se agrupam em um único ramo interno, com probabilidade de confiança de 87%, independentemente de sua origem geográfica. Contudo, além das seqüências GWGR, oito seqüências GPGR também se agruparam dentro deste ramo. Apesar desse grupo incluir algumas seqüências GPGR, para simplificar, no restante do texto, ele será referido como o ramo GWGR.

A partir da distribuição das seqüências GWGR na árvore filogenética, duas hipóteses podem ser propostas para se explicar a origem evolutiva da variante. A primeira hipótese é a de que os vírus GWGR tenham surgido em diversos momentos no curso da epidemia. A segunda hipótese é que todas as seqüências GWGR evoluíram a partir de um único ancestral comum. Estas duas hipóteses serão analisadas separadamente nas próximas seções. Contudo, antes de discuti-las, algumas considerações serão feitas a seguir com o objetivo de facilitar o fluxo do texto.

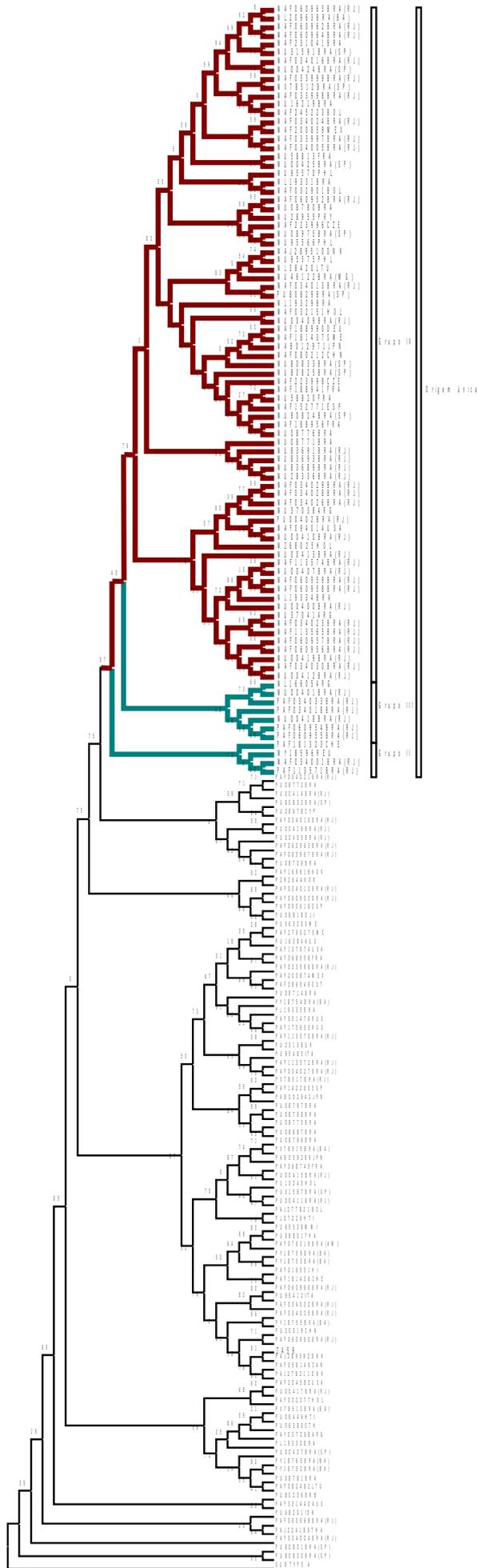


Figura 9 – Árvore de neighbor-joining para todas as seqüências analisadas na tese. Os nomes das seqüências das variantes GWGR são iniciados por W e os das variantes GPGR são iniciados por P. Os valores acima dos ramos internos denotam a probabilidade de confiança do ramo ser positivo (CP). ZR59 é a seqüência mais antiga conhecida de HIV, isolada de uma amostra do antigo Zaire colhida em 1959. O grupo externo é um representante do subtipo A.

4.1.1. O motivo MGWGR

Uma característica do ramo GWGR é a presença de um resíduo de metionina (M) precedendo o motivo GWGR, ao invés da isoleucina (I) encontrada na maioria das seqüências do subtipo B (Kuiken et al., 2000a). A presença desta metionina nas seqüências GWGR já havia sido descrita por alguns autores (Covas et al., 1998; Morgado et al., 1994). No grupo de seqüências analisadas, a metionina foi encontrada em 51 das 81 (63%) seqüências com o motivo GWGR. Nas demais seqüências GWGR, a isoleucina estava presente em somente cinco (6,2%), enquanto uma (1,2%), duas (2,5%) e 22 (27,2%) seqüências possuíam, respectivamente, treonina, valina e leucina nesse sítio.

No grupo GPGR, a isoleucina estava presente em 79 seqüências - 81%. No restante das seqüências GPGR os aminoácidos encontrados foram fenilalanina (n=1, 1%), valina (n=1, 1%), metionina (n=6, 6,1%) e leucina (n=10, 10,2%). Em uma das seqüências GPGR não foi possível definir o aminoácido deste sítio, uma vez que faltava a identificação de um dos nucleotídeos do códon correspondente. Dentre as oito seqüências com o motivo GPGR localizadas dentro do ramo GWGR, três possuíam a substituição da isoleucina pela metionina, enquanto em duas seqüências a isoleucina estava preservada. As outras três seqüências tinham leucina ou fenilalanina nesse sítio.

Considerando-se que o códon mais freqüente para a isoleucina no sítio adjacente ao motivo protéico é ATA e que a metionina é somente codificada por ATG, a predominância de seqüências GWGR com a substituição I→M é interessante, uma vez que esta mutação ocorre na direção oposta à esperada devido ao vício para a incorporação de adenina que ocorre nos genomas do HIV-1 (Kypr & Mrázek, 1987). A substituição da isoleucina pela metionina no sítio adjacente ao motivo protéico será referida mais adiante no texto, quando as hipóteses para a origem da variante GWGR forem analisadas.

4.1.2. Seqüências GPGR do ramo GWGR

A presença das oito seqüências GPGR dentro do ramo GWGR pode ser explicada, alternativamente, por três razões:

1) A localização das seqüências GPGR dentro do ramo GWGR pode ter ocorrido por erro de reconstrução filogenética. Contudo, a probabilidade de 87% de confiança do ramo GWGR calculada pelo teste do ramo interno fala contra esta possibilidade. Apesar do valor do teste não ser significativo no nível de 5%, como habitualmente é considerado, ainda assim é extremamente alto se comparado dentre as filogenias de vírus.

2) A segunda explicação para a inclusão de seqüências com o motivo GPGR dentro do ramo GWGR é que estas seqüências tenham evoluído por mutação reversa. Para que este processo ocorresse, seria necessário que os três nucleotídeos do códon TGG para o triptofano fossem substituídos. Isto porque, na maioria das seqüências de HIV-1 disponíveis, a prolina encontrada na segunda posição do tetrapeptídeo está codificada por CCA (Kuiken et al., 2000a). As etapas evolutivas que envolvem a substituição da prolina pelo triptofano e, por conseguinte, os passos que envolvem a reversão para a característica original serão apresentados na próxima seção.

3) A terceira possibilidade é que as seqüências GPGR incluídas dentro do ramo GWGR possam, na verdade, representar formas recombinantes intrasubtípicas GPGR-GWGR. Neste caso, o motivo protéico no topo da alça V3 teria sido herdado de uma seqüência GPGR, enquanto o restante do segmento de DNA analisado seria derivado de vírus GWGR.

De modo geral, os vírus recombinantes são identificados quando se caracteriza diferentes partes do genoma, seja através da análise filogenética de mais de um segmento de DNA ou por técnicas moleculares mais simples, como o HMA (*heteroduplex mobility assay*). Como dito na introdução do texto, a recombinação entre genes e intra-genes de vírus dos diversos subtipos do grupo M e de vírus do grupo O é amplamente descrita na literatura (Belda et al., 1998; Liitsola et al., 2000; Peeters, 2000). Nas análises filogenéticas, a suspeita de que uma determinada seqüência contenha segmentos de diferentes subtipos se dá quando esta se agrupa fora dos ramos dos subtipos conhecidos ou dentro de um ramo, mas separada das demais seqüências.

Apesar da recombinação também poder ocorrer entre seqüências de um único subtipo, a grande similaridade entre estas seqüências faz com que os métodos normalmente utilizados para se detectar genomas recombinantes sejam inadequados para a sua identificação (Peeters, 2000).

A distribuição das oito seqüências GPGR dentro do ramo GWGR não seguiu um padrão homogêneo. Destas seqüências, seis se localizaram nos ramos internos mais basais do ramo GWGR enquanto duas seqüências se posicionaram no interior deste ramo. Por esta razão, as possibilidades das seis primeiras seqüências terem se originado por mutação reversa ou sua posição resultar de erro de reconstrução filogenética serão discutidas mais adiante nesse capítulo, quando forem apresentadas as duas hipóteses para a origem da variante GWGR (itens 4.2 e 4.3).

Em relação às outras duas seqüências GPGR (números de acesso U80829 e U00402), a sua posição no interior do ramo GWGR reforça a hipótese que, neste caso, não houve erro de reconstrução filogenética. Além disto, as duas seqüências se agruparam com seqüências GWGR em ramos internos com, respectivamente, 81% e 70% de probabilidades de confiança, o que também reforça essa argumentação.

Por outro lado, nessas duas seqüências, o códon utilizado para a prolina foi CCG, ao invés do códon CCA encontrado na maioria das seqüências de HIV-1 (Kuiken et al., 2000a). Como será visto a seguir, na análise do caminho evolutivo, este códon seria um intermediário na substituição da prolina pelo triptofano, codificado por TGG. Além destas duas seqüências, em somente outras seis seqüências GPGR incluídas na análise filogenética o códon CCG foi encontrado. Desta forma, para estas duas seqüências, a explicação mais provável para a sua localização dentro do ramo GWGR é a de que tenham evoluído por mutação reversa.

4.1.3. Caminho evolutivo

Como dito anteriormente, duas hipóteses podem explicar a origem da variante GWGR, ou seja, estes vírus possuiriam uma única origem ou teriam surgido em mais de um momento no curso da epidemia (origem múltipla). Contudo, independentemente do número de vezes que a variante surgiu, as mutações necessárias para a substituição da prolina pelo triptofano no motivo protéico são essencialmente as mesmas. Desta forma, nesta seção, um caminho evolutivo para a origem do variante GWGR será proposto.

Uma vez que o foco da tese é o motivo protéico GWGR, a análise do caminho evolutivo ficará restrita às substituições de nucleotídeos necessárias para a substituição do resíduo de prolina pelo triptofano nas seqüências protéicas. Entretanto, como todas as seqüências GWGR se agruparam em um único ramo da árvore com alta probabilidade de confiança, outras substituições de nucleotídeos (e possivelmente de aminoácidos) naturalmente se acumularam ao longo do segmento de DNA analisado para que esta distribuição ocorresse.

Para que o caminho evolutivo fosse conhecido, seria necessário ter em mãos seqüências representativas de todas as variantes do subtipo B que surgiram no curso da epidemia. Isto porque substituições múltiplas e reversas podem ter ocorrido nos três sítios do códon em questão durante a evolução da variante GWGR e, sem que todas as seqüências estejam disponíveis, essas mutações passaram despercebidas.

Assim, com este imenso número de seqüências, poder-se-ia, então, construir uma árvore filogenética e identificar aquelas que possuísem as mutações intermediárias à fixação do triptofano. Como esta análise é impossível de ser feita, o caminho evolutivo proposto no trabalho de tese é o mais parcimonioso, ou seja, é o caminho evolutivo que necessita do menor número de eventos de substituição no códon para a prolina para que este aminoácido seja substituído pelo triptofano. Esse caminho é representado na figura 10.

Como pode ser visto na figura 10, a substituição do resíduo de prolina pelo triptofano envolve a substituição dos três nucleotídeos do códon. Isto porque, apesar da degeneração do terceiro sítio para a prolina no código genético (CCN), na maior parte das seqüências de HIV-1 descritas, o códon usado para este aminoácido na posição é CCA (Kuiken et al., 2000a). O códon CCA para a prolina é encontrado inclusive no fragmento ZR59, considerado um ancestral próximo ao subtipo B (Zhu et al., 1998), e nas seqüências do SIV_{cpz} (Kuiken et al., 2000a). Por outro lado, o único códon disponível para o triptofano no código genético universal é TGG.

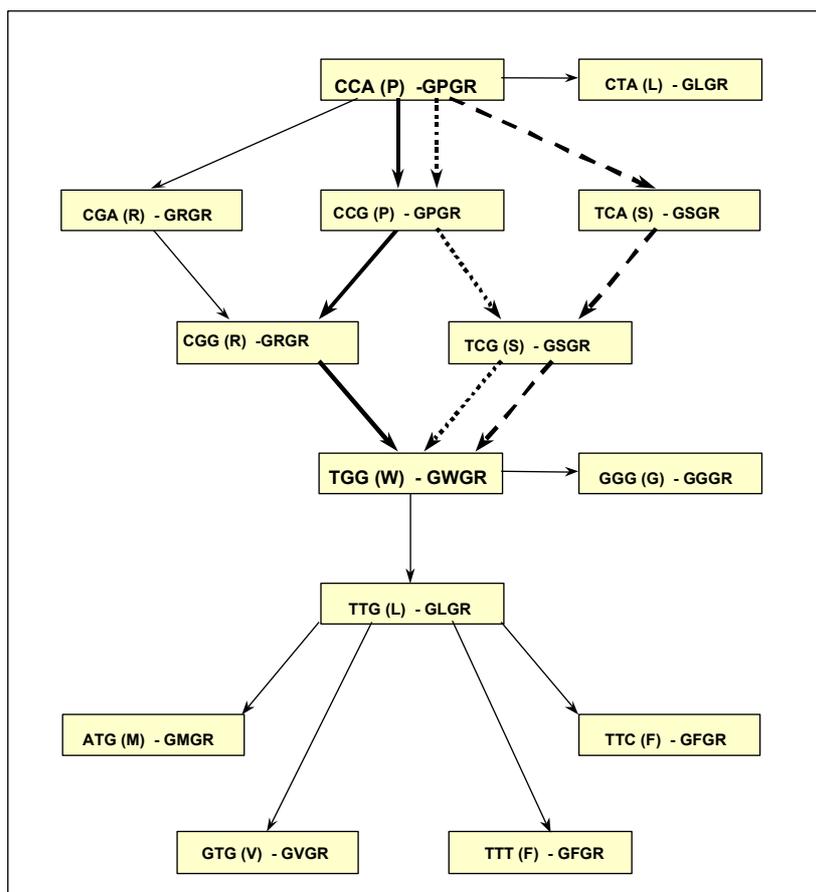


Figura 10. Origem evolutiva da variante GWGR

Desta forma, o caminho evolutivo mais parcimonioso (figura 10) envolve uma substituição de nucleotídeo sinônima, ou seja, do códon CCA para CCG, e duas substituições de nucleotídeos não-sinônimas, ou seja, do códon CCG para TCG (que codifica para a serina) ou para CGG (que codifica para a arginina). (As substituições de nucleotídeos sinônimas não resultam na substituição de aminoácidos nas seqüências protéicas enquanto as substituições não-sinônimas sim). Segundo esse caminho, as seqüências GPGR nas quais a prolina está codificada por CCG e as seqüências com os padrões de assinatura GRGR e GSGR poderiam representar etapas intermediárias na evolução da variante GWGR.

Para se estudar a participação das seqüências com os motivos GRGR e GSGR na epidemia do HIV-1, as seqüências com os dois padrões de assinatura depositadas no Genbank foram identificadas. Nesta busca foi empregada a mesma estratégia descrita

para as seqüências GPGR e GWGR (item 3.1.1.1). Naturalmente, neste caso, P e W, no *script* original, foram substituídos por R e S.

A busca das seqüências GRGR e GSGR no Genbank resultou em 183 seqüências com a arginina na segunda posição do tetrapeptídeo e 332 seqüências com a serina nesta posição. A seguir, as seqüências GRGR e GSGR foram analisadas de acordo com os critérios definidos no item 3.1.1.3. Após esta análise, do grupo inicial, somente nove seqüências com o motivo GRGR e 17 seqüências GSGR foram selecionadas. A distribuição geográfica e o ano de coleta das seqüências GRGR e GSGR do subtipo B estão resumidos na tabela 4.

A identificação das seqüências GRGR e GSGR depositadas no Genbank sugere que não há uma disseminação de vírus do subtipo B com estes padrões de assinatura na epidemia mundial de HIV-1: apesar destes vírus terem sido encontrados em diversos países, o número de seqüências disponíveis de cada um deles foi pequeno quando comparado ao número de seqüências GPGR locais disponíveis. Uma possível explicação para este resultado é o vício de amostragem existente no banco de dados. Assim, seqüências com os motivos GRGR e GSGR ancestrais podem circular em alguma região geográfica sem que jamais tenham sido identificadas.

Uma outra possibilidade é que vírus com os padrões intermediários GRGR e GSGR tenham circulado principalmente durante uma etapa muito inicial da epidemia sem ter deixado muitos descendentes. Uma vez que no banco de dados há poucas seqüências representativas dos primeiros anos da epidemia (tabela 1), se os vírus GRGR e GSGR circularam especialmente nesta época sua presença pode simplesmente não ter sido registrada.

Uma evidência a favor desta última hipótese é o fato de que, na literatura, o registro mais antigo da variante GWGR vem da identificação de soros colhidos no Brasil em 1983 que reagem a peptídeos sintéticos GWGR (Michael-Hendry et al., 1996). Isto significa que, se a variante GWGR circula na epidemia desde o início da década de 80, os seus intermediários, naturalmente, surgiram em algum momento antes desta época.

Os dados referentes à distribuição geográfica, ano de coleta da amostra e códon utilizado para os aminoácidos arginina e serina das seqüências GRGR e GSGR também são apresentados na tabela 4. Como pode ser observado, as seqüências com o padrão de assinatura GRGR foram identificadas nos seguintes países: Bélgica, Brasil, Cuba,

Espanha, Estados Unidos, França, Hungria e Uruguai. Dentre estas seqüências, somente aquelas de Cuba, do Brasil, dos Estados Unidos e do Uruguai utilizavam o códon CGG, enquanto nas demais seqüências o códon para a arginina é o CGA

Tabela 4. Distribuição geográfica das seqüências do gene env com códons para arginina (R) ou aerina (S) na segunda posição do motivo protéico na crista da alça V3

Arginina			
No acesso	País	Ano	Códon
AJ228162	Bélgica	?	CGA
AF113573	Brasil	1994	CGG
AF034015	Brasil	1996	CGG
U48855	Cuba	1994	CGG
AY017449	Espanha	1999	CGA
U96608	EUA	?	CGG
AF125615	França	1997	CGA
AJ245464	Hungria	?	CGA
U66418	Uruguai	1995	CGG
Serina			
AY010876	Austrália	1995	TCA
U20030	China	1991	TCA
AJ278210	Dinamarca	?	TCA
AF203287	EUA	1982	TCG
AF204643	EUA	1983	TCA
U36018	EUA	1986	TCA
AF372599	EUA	?	TCA
AF256375	EUA	?	TCA
AF138327	EUA	1990	TCG
AF137668	EUA	1994	TCA
U30001	EUA	1985	TCA
Z47528	Holanda	1994	TCA
U75174	Itália	?	TCA
U75042	Itália	1996	TCA
U95398	Itália	1997	TCA
AB002949	Japão	1994	TCA
AF236278	Reino Unido	?	TCA

Por outro lado, o motivo GSGR foi encontrado em seqüências da Austrália, China, Dinamarca, Estados Unidos, Holanda, Itália, Inglaterra e Japão. Dentre estas, o códon TCG foi encontrado em somente duas seqüências dos Estados Unidos enquanto, nas demais seqüências, a serina estava codificada por TCA.

Como pode ser visto, os únicos países onde a variante GWGR e as seqüências com os padrões intermediários no caminho evolutivo proposto foram encontradas são o Brasil e os Estados Unidos.

4.2. Hipótese de origem múltipla

Nesta seção, será estudada a hipótese de que os eventos que deram origem às seqüências GWGR tenham ocorrido em mais de uma ocasião no curso da epidemia do HIV-1. Considerando-se esta hipótese, diversos cenários poderiam ser propostos para explicar a origem das seqüências GWGR. A razão para isto, como no caso do caminho evolutivo proposto anteriormente (item 4.1.3), é a possibilidade de substituições múltiplas e reversas poderem ter ocorrido no curso da evolução das variantes. Contudo, somente um dos possíveis cenários será considerado no trabalho. Também como no caso do caminho evolutivo, o cenário escolhido foi aquele que incluía o menor número de eventos de mutação necessários para explicar a origem da variante GWGR e a presença das seqüências GPGR no ramo GWGR. Esse cenário é apresentado na figura 11.

No cenário da figura 11, os eventos de mutação que originaram os vírus GWGR estão representados em vermelho, enquanto as mutações reversas necessárias para explicar o agrupamento das seqüências GPGR com os vírus GWGR na árvore filogenética estão representadas em azul. Para facilitar o fluxo do texto, as seqüências que serão citadas foram numeradas de 1 a 13.

Ainda nessa figura, todas as seqüências GPGR posicionadas fora do ramo GWGR estão reunidas no grupo I. Nos grupos II e III estão as seqüências que formaram, respectivamente, o primeiro e segundo ramos a divergir das seqüências GPGR. As seqüências localizadas nestes ramos estão numeradas de 1 a 4 (grupo II) e de 5 a 11 (grupo III) da figura 9. Finalmente, no grupo IV, está a maioria das seqüências GWGR e as duas seqüências GPGR localizadas no interior do ramo GWGR (números de acesso U80829 e U00402), numeradas como 12 e 13.

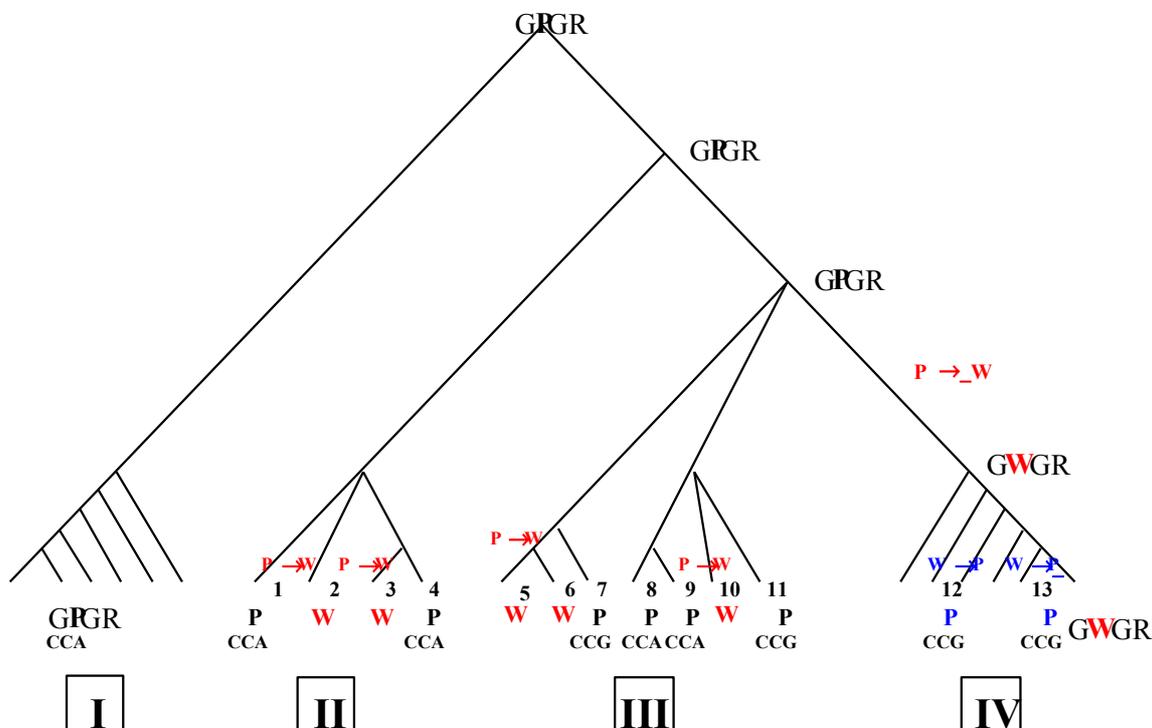


Figura 11. Diagrama ilustrando o cenário proposto para a hipótese de origem múltipla da variante GWGR.

Nesse cenário, cinco eventos de mutação nas seqüências protéicas são necessários para explicar a substituição da prolina pelo triptofano. Estes eventos, representados na figura, deram origem às seqüências 2, 3, 5 e 6, 10 e às seqüências do grupo IV. Além destes, outros dois eventos de mutações reversas precisariam ter ocorrido para justificar a inclusão das seqüências GPGR 12 e 13 dentro do ramo GWGR.

Como pode ser visto na figura 11, o grupo II inclui duas seqüências GPGR (seqüências 1 e 4) e duas seqüências GWGR (seqüências 2 e 3) originadas, cada uma, por um evento de mutação P→W. Uma vez que todas as seqüências incluídas neste grupo possuem a substituição da isoleucina pela metionina antes do motivo protéico, a seqüência que deu origem a todo o grupo possuía o motivo protéico MGPGR.

No grupo III estão incluídas sete seqüências, sendo três com o motivo GWGR (seqüências 5 e 6 e 10) e quatro seqüências GPGR (seqüências 7, 8, 9 e 11). Neste grupo, as seqüências 5 e 6 descenderam de um único evento de mutação P→W enquanto a seqüência 10 resultaria de um outro evento. Da mesma forma que no grupo II, a seqüência ancestral ao grupo possuiu o motivo protéico GPGR no topo da alça V3. Contudo, neste caso, o aminoácido encontrado antes do tetrapeptídeo era uma isoleucina (motivo IGPGR).

Finalmente, todas as seqüências GWGR do grupo IV resultariam do quinto evento de mutação P→W e, neste caso, o motivo protéico encontrado na seqüência ancestral seria MGWGR. Quanto às duas seqüências GPGR incluídas no grupo (números de acesso U00402 e U80829), a explicação mais provável é que estas tenham evoluído por mutação reversa. A seqüência U00402 possui metionina antes do motivo protéico e se agrupou com outras duas seqüências GWGR (também com metionina) em um ramo interno com 69% de probabilidade de confiança.

Já a seqüência U80829 se agrupou internamente com uma seqüência GWGR em um ramo com 82% de confiança, e as duas seqüências formaram um ramo mais interno com outras quatro seqüências GWGR (suporte de 83%). Para este último ramo, o motivo parental mais provável é o IGWGR, uma vez que uma de suas seqüências possui a isoleucina antes do tetrapeptídeos enquanto nas demais seqüências o aminoácido encontrado na posição é a leucina (códon CTA e CTG).

No cenário proposto, além dos sete eventos de mutação no códon para o aminoácido na segunda posição dos tetrapeptídeos, é necessário que a substituição da isoleucina pela metionina tenha ocorrido em dois momentos. Contudo, tal mutação envolve a substituição da adenina (A) da terceira posição do códon ATA (isoleucina) por guanina para gerar o códon ATG (metionina). Como já exposto no texto, nos genomas do HIV-1 existe um vício para a incorporação de adenina, ao menos parcialmente atribuído à hipermutação G→A (Martínez et al., 1994; Vartanian et al., 1994). Uma vez que estes eventos de mutação seguem a direção contrária ao vício, é pouco provável que tenham ocorrido duas vezes.

4.2.1. Local de origem

A posição mais basal do ramo GWGR é um bom indicador da origem da

variante. Considerando a origem múltipla, na árvore filogenética, o primeiro grupo a divergir dos ramos com as seqüências GPGR (grupo II da figura 11) é formado por duas seqüências GWGR, uma do Brasil e outra das Ilhas Reunidas, e por duas seqüências GPGR. A seqüência GWGR brasileira incluída nesse grupo foi coletada em 1996 de um doador de sangue sobre o qual não havia informações disponíveis (Tanuri et al., 1999).

Quanto à seqüência GWGR das Ilhas Reunidas, esta foi obtida de uma amostra colhida em 1996 de um indivíduo que se infectou com o HIV-1 através de uma transfusão de sangue recebida na França (Brenzel-Pesce et al., 1999). Apesar de a data da transfusão do paciente não estar disponível, dois dados sugerem que tenha sido administrada ao menos uma década antes da coleta da amostra. Primeiro, quando a amostra foi colhida o paciente já tinha desenvolvido AIDS. Apesar da taxa de progressão da imunodeficiência, na população geral, variar muito entre os pacientes, o tempo mediano de incubação sem AIDS está em torno de 8 a 10 anos. Em segundo lugar, o risco para a transmissão do vírus através de transfusão de sangue diminuiu substancialmente após a introdução dos testes sorológicos específicos para o HIV, em 1985. Na França, foi estimado que o risco de infecção era de aproximadamente 1 por 150,000 unidades de sangue (por definição, uma unidade de sangue se refere ao volume obtido de um único doador) em 1990 (Msellati et al., 1990).

No segundo ramo a divergir das seqüências GPGR (grupo III) estão três seqüências GWGR e quatro seqüências com o motivo GPGR. Uma das variantes GWGR foi encontrada na Argentina, em uma amostra colhida entre 1991 e 1992. Contudo, os dados referentes ao paciente não estavam disponíveis no trabalho (Carrillo et al., 1992). As outras duas seqüências GWGR são do Brasil, de 1990, uma delas de um homem bissexual e a outra de uma mulher que adquiriu a infecção por contato sexual (Morgado et al., 1994).

Desta forma, as seqüências GWGR localizadas nos primeiros ramos internos do grupo GWGR vieram do Brasil, da Argentina e da França, o que sugere estes países como locais prováveis para a origem da variante GWGR. Contudo, como visto anteriormente, na análise do caminho evolutivo, seqüências com os padrões intermediários não foram encontradas na França ou na Argentina. Além disto, não há indícios na literatura de que a variante GWGR contribua substancialmente para a epidemia dos dois países. Por outro lado, duas seqüências GRGR foram identificadas no Brasil e 67% das seqüências GWGR são brasileiras.

Além do que foi exposto, mesmo que a variante GWGR tenha surgido mais de uma vez no curso da epidemia pelo HIV-1 e em locais diferentes, a distribuição das seqüências GWGR na árvore filogenética sugere que a epidemia por esta variante tem uma única origem no Brasil. A razão para essa hipótese é que, na árvore filogenética (figura 9), a grande maioria (94%) das seqüências com este padrão de assinatura se agrupam em um único ramo interno cujas posições basais são ocupadas por seqüências brasileiras. No cenário proposto este ramo está representado no grupo IV.

4.3. Hipótese da origem única

Segundo a hipótese da origem única da variante GWGR, todos os vírus com este padrão de assinatura que participam da epidemia pelo HIV-1 teriam se originado de um único ancestral comum. Assim sendo, a dispersão mundial do variante resultaria da sua irradiação a partir de uma única região geográfica. O cenário proposto para explicar esta hipótese está apresentado na figura 12.

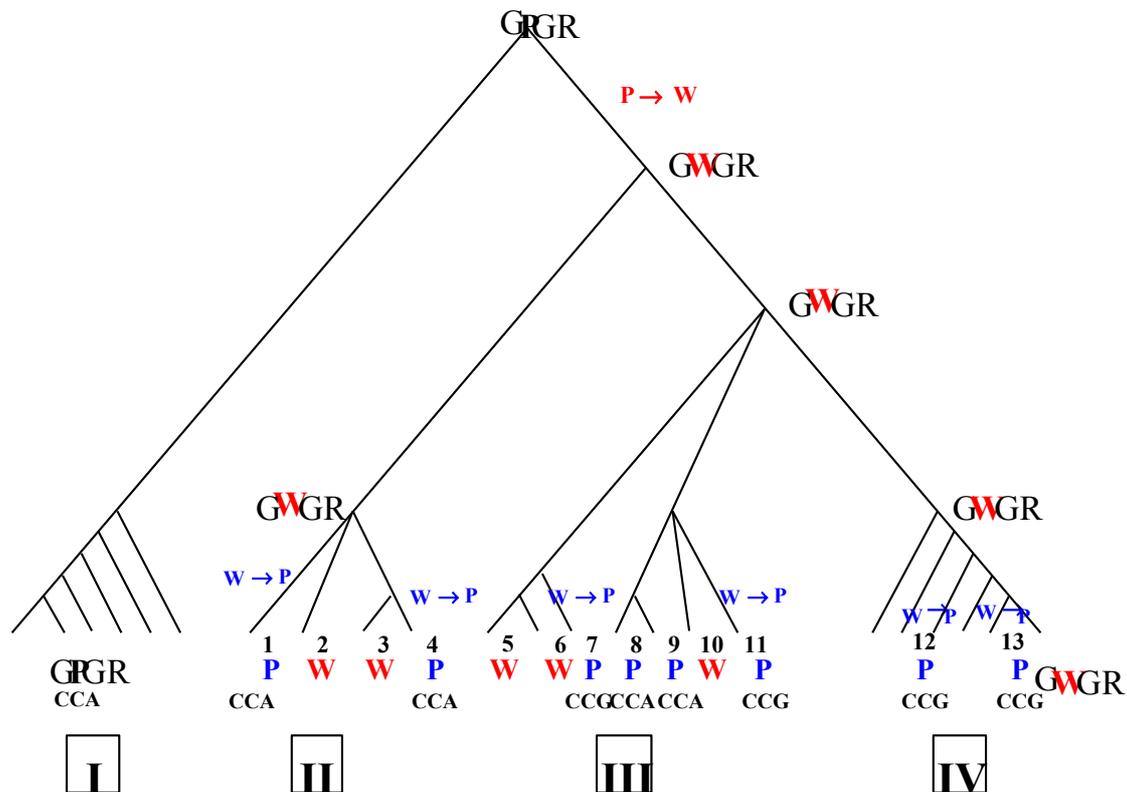


Figura 12. Diagrama ilustrando o cenário proposto para a hipótese de origem única da variante GWGR.

Na figura 12, da mesma forma que na figura anterior, o evento de mutação que deu origem às seqüências GWGR está representado em vermelho enquanto as mutações reversas necessárias ao cenário estão marcadas em azul. Também como na figura 11, as seqüências que serão discutidas foram numeradas de 1 a 13 e reunidas nos grupos I, II, III e IV segundo os mesmos critérios.

No cenário da figura 12, a substituição da prolina pelo triptofano ocorreu antes da divergência de todas as seqüências que formam o ramo GWGR na árvore filogenética, da mesma forma que a substituição da isoleucina pela metionina que antecede o motivo protéico do topo da alça V3. Assim, as seqüências ancestrais aos grupos II, III e IV possuem o motivo protéico MGWGR.

Ainda nesse cenário, seis eventos de mutação reversa são necessários para explicar a inclusão das seqüências GPGR no ramo GWGR. Dois destes eventos teriam ocorrido dentro do grupo II e teriam originado as seqüências 1 e 4. No grupo III, outros dois eventos de mutação são necessários para explicar a inclusão das seqüências GPGR no grupo, e tais eventos originariam as seqüências 7 e as seqüências 8, 9 e 11. Contudo, nas seqüências do grupo III, o aminoácido que antecede o motivo protéico é a isoleucina, mas a presença deste aminoácido pode ser facilmente atribuída à substituição da guanina do códon ATG (metionina) pela adenina do códon ATA devido ao vício na incorporação de adenina.

Os outros dois eventos de mutação reversa resultaram nas seqüências 12 e 13 do grupo IV. Contudo, a posição destas seqüências na árvore filogenética já foi discutida anteriormente.

4.3.1. Local de origem

Como dito anteriormente, as seqüências GWGR incluídas nos ramos basais do ramo GWGR foram obtidas no Brasil, Argentina e França. Desta forma, segundo a hipótese da origem única, a variante GWGR surgiu em um dos três países. Uma vez que a expansão da variante é observada principalmente no Brasil, e três das cinco seqüências incluídas nos ramos basais são brasileiras, este país pode ser apontado como o local mais provável para a origem da variante GWGR.

Para melhor se estudar esta possibilidade, foi feita uma análise com todas as seqüências brasileiras do subtipo B depositadas no Genbank. Para isto, as seqüências

brasileiras que apresentavam outros padrões de assinatura além do GPGR, do GWGR e do GRGR foram incluídas no caminho evolutivo discutido anteriormente (figura 10). Além disto, os motivos protéicos encontrados no topo da alça V3 e os códons para os aminoácidos que ocupam a segunda posição dos tetrapeptídeos de todas as seqüências brasileiras foram analisados e estão resumidos na tabela 5.

Tabela 5. Distribuição geográfica das seqüências do subtipo B segundo os motivos na região V3 do gene *env* de amostras identificadas no Brasil.

Motivo/Codom	Estados com amostras							Total	
	Rio de Janeiro (RJ)	São Paulo (SP)	Minas Gerais (MG)	Paraná (PR)	Amazonas (AM)	Bahia (BA)	Não conhecido		
GPGR									
CCA	Ano	1988	1987	1995	-	1997	1988	1992	
	n	23	5	5	-	1	8	11	53
CCG	Ano	1990	1993	?	-	-	1995	-	
	n	6	1	1	-	-	1	-	9
CC?	Ano	1996	-	-	-	-	-	-	
	n	2	-	-	-	-	-	-	2
GWGR									
TGG	Ano	1990	1988	?	-	-	1993	1992	
	n	37	8	1	-	-	1	7	54
GRGR									
CGG	Ano	1994	-	-	-	-	-	1996	
	n	1	-	-	-	-	-	1	2
GLGR									
CTA	Ano	1992	-	-	-	-	-	-	
	n	2	-	-	-	-	-	-	2
TTG	Ano	1992	-	-	-	-	-	1993	
	n	3	-	-	-	-	-	2	5
GGGR									
GGG	Ano	1990	-	-	-	-	-	-	
	n	1	-	-	-	-	-	-	1
GMGR									
ATG	Ano	-	-	-	-	-	1988	1993	
	n	-	-	-	-	-	1	1	2
GVGR									
GTG	Ano	1992	-	-	-	-	-	1993	
	n	1	-	-	-	-	-	1	2
GFGR									
TTT	Ano	1992	-	-	1988	-	-	1995	
	n	6	-	-	1	-	-	2	9
TTC	Ano	-	-	-	-	-	-	1993	

Como pode ser visto na tabela 5 e na figura 10, no Brasil, além das variantes GPGR, GWGR e GRGR, também circulam vírus do subtipo B com os motivos protéicos GLGR, GGGR, GMGR, GVGR e GFGR. A análise dos códons para os aminoácidos encontrados nestes motivos, respectivamente, leucina, glicina, metionina, valina e fenilalanina, mostra que as mudanças evolutivas necessárias para a fixação destes aminoácidos podem ser mais bem explicadas a partir do códon TGG para o triptofano do que a partir do códon CCA para a prolina, uma vez que, no primeiro caso, envolvem um menor número de substituições de nucleotídeos.

As datas de coleta das amostras das quais as seqüências brasileiras foram obtidas também foram incluídas na tabela 5. Como pode ser visto, os vírus do subtipo B com motivos atípicos circulam no Brasil pelo menos desde o final dos anos 80. Estes resultados confirmam que a variante GWGR tem uma participação importante na epidemia brasileira desde a década de 80.

Em conjunto, a análise da árvore filogenética, a distribuição geográfica das variantes intermediárias propostas no caminho evolutivo, o fato da dispersão da variante GWGR ter sido observada principalmente dentro do Brasil, os indícios sorológicos de que a variante circula neste país desde o início da década de 80 e os padrões de assinatura encontrados entre as seqüências brasileiras sugerem que o Brasil seja o local de origem da epidemia pela variante GWGR.

É interessante notar que nenhuma seqüência africana com o motivo GWGR nem com os padrões de assinatura intermediários foi encontrada no Genbank. Na verdade, o único indício da presença dos vírus GWGR no continente africano vem do relato de viagens feitas ao Dibuti por um marinheiro francês infectado pela variante (Lasky et al., 1997). Este fato chama a atenção, uma vez que a transmissão do HIV-1 para a espécie humana e a irradiação das variantes que deram origem aos subtipos do grupo M, inclusive o subtipo B, teria ocorrido na África Central (Vidal et al., 2000). Desta forma, a ausência de seqüências GWGR africanas mostra que, apesar do local de origem dos vírus parentais GPGR ser a África, as mutações necessárias para a formação da variante GWGR teriam acontecido fora do continente africano, após o subtipo B ter sido trazido para o Ocidente.

4.4. Dispersão do variante GWGR

Como discutido até o momento, os resultados do trabalho sugerem que a epidemia pela variante GWGR se originou no Brasil. Nesta seção, uma rota para a dispersão mundial desta variante, considerando-se o Brasil como ponto de partida, será analisada. Para isto, os dados epidemiológicos, quando disponíveis, dos indivíduos dos quais as seqüências GWGR foram obtidas foram retirados diretamente do Genbank ou dos artigos científicos originais. Estes dados incluíram a provável via de infecção pelo HIV-1, o país e o ano de coleta da amostra (ou da infecção) e o relato de viagens ao exterior ou de contato sexual com estrangeiros (tabela 6). Estes dados foram utilizados, junto com a análise da árvore filogenética, para se propor uma rota para a dispersão da variante GWGR.

Como mostrado na tabela 1, na América Latina, além do Brasil, a variante GWGR também circula na Argentina (Campodonico et al., 1996; Carrillo et al., 1992), no Paraguai (Cabello et al., 1995), na Bolívia (Velarde-Dunois et al., 2000) e no México (Rivera-Morales et al., 2001). Na Argentina, três seqüências GWGR foram identificadas e sobre duas destas seqüências havia dados epidemiológicos disponíveis na literatura (Campodonico et al., 1996). Destas duas seqüências, uma foi encontrada em um paciente heterossexual e se agrupou com seqüências brasileiras e com a seqüência GWGR encontrada nos Estados Unidos (ramo interno com 79% de suporte). A outra seqüência da Argentina foi obtida de um usuário de drogas injetáveis que se infectou em 1987, o que indica que a variante circula no país pelo menos desde aquele ano. Na árvore filogenética, esta última seqüência também se agrupou com seqüências do Brasil em um ramo interno com 94% de suporte estatístico. Sobre a terceira seqüência da Argentina não havia dados epidemiológicos disponíveis, mas esta seqüência também se agrupou com seqüências brasileiras.

A seqüência GWGR do Paraguai foi identificada em um paciente do sexo masculino com sintomas relacionados à AIDS, mas a sua categoria de exposição não foi especificada no artigo (Cabello et al., 1995). Na árvore, esta seqüência se agrupou com outras seqüências do Brasil, uma da República Checa e uma das Filipinas (ramo interno com 85% de suporte). A variante GWGR foi encontrada no México em um paciente homossexual cujo diagnóstico de infecção pelo HIV-1 foi feito em 1996 (Rivera-Morales et al., 2001); na árvore filogenética a seqüência se agrupou com seqüências do Rio de Janeiro em um ramo interno com 75% de suporte. As duas seqüências da Bolívia

foram identificadas em heterossexuais que tiveram o diagnóstico da infecção em 1992 e se agruparam com seqüências brasileiras (Velarde-Dunois et al., 2000). No artigo original onde estas duas seqüências bolivianas são citadas, ambas se agruparam com seqüências do Brasil incluídas na análise (Velarde-Dunois et al., 2000).

Em resumo, todas as seqüências GWGR encontradas nos demais países da América Latina formaram ramos com seqüências GWGR brasileiras, o que sugere que a introdução da variante nesses países ocorreu diretamente do Brasil. De fato, alguns dados da literatura sugerem uma ligação entre a epidemia de HIV-1 no Brasil e as epidemias dos países da América Latina. Por exemplo, o padrão de distribuição dos subtipos no Brasil, na Argentina (Campodonico et al., 1996) e na Bolívia (Velarde-Dunois et al., 2000) é semelhante, com o predomínio de vírus do subtipo B, seguidos dos vírus F. Além disso, no Brasil e na Argentina, junto com vírus do subtipo B, também circulam diversas formas recombinantes B/F (Thomson et al., 2000) tendo inclusive sido demonstrado que o segmento F dos recombinantes argentinos são evolutivamente próximos aos vírus F do Brasil (Thomson et al., 2002). Finalmente, devido à proximidade geográfica entre o Brasil e o Paraguai, foi sugerido que a epidemia deste último país está relacionada à epidemia brasileira (Cabello et al., 1995).

Fora da América Latina, o motivo GWGR foi identificado em 20 seqüências GWGR estrangeiras. Destas 20 seqüências, 13 se agruparam em um único ramo interno com 62% de suporte estatístico junto com seqüências brasileiras. Para facilitar o fluxo do texto, este ramo será referido como ramo estrangeiro. No ramo estrangeiro estão incluídas seqüências da Dinamarca, das Filipinas, da Lituânia, da Holanda, da Alemanha, da Suíça, do Japão, da China, da França, da República Checa e da Espanha.

A origem dos vírus GWGR que circulam na China (Chen et al., 1999), na Alemanha (Casado et al., 2000b), na Espanha (Casado et al., 2000a), na Suíça (Stoeckli et al., 2000) e na Lituânia (Lukashov et al., 1995) não pode ser traçada a partir dos dados epidemiológicos disponíveis. A introdução do variante GWGR na Dinamarca também não é clara, mas é possível que tenha ocorrido diretamente do Brasil. No artigo científico que discute a questão, uma das seqüências do subtipo B encontradas na Dinamarca foi obtida de um paciente que se infectou no Brasil, porém os autores não a identificam em relação às demais seqüências dinamarquesas (Machuca et al., 2001). De qualquer forma, na árvore filogenética, a seqüência dinamarquesa se agrupou com seqüências das Filipinas, da Lituânia e do Brasil (ramo interno com 83% de suporte estatístico).

Tabela 6. Dados epidemiológicos das seqüências GWGR encontradas fora do Brasil.

Nº de acesso	Seqüência		Dados epidemiológicos		
	País de origem	Ano da coleta	Via de infecção	Local de origem	Outras informações
AF188990	Alemanha	?	Homossexual	?	
U37038	Argentina	1995	Heterossexual	Brasil	Soroconversão em 1993
U37041	Argentina	1995	UDI	Brasil	Soroconversão em 1987
L16605	Argentina	1991	?	Brasil	
AF245223	Bolívia	1996	Heterossexual	Brasil	Diagnóstico sorológico em 1992
AF032901	Bolívia	1996	Heterossexual	Brasil	Diagnóstico sorológico em 1992
AF080212	China	1996	?	?	Doador Inapto
AJ289510	Dinamarca	1996	?	Brasil ?	Um dos pacientes estudados no artigo original se infectou no Brasil
AF152771	Espanha	1993	?	?	
AF094014	EUA	1999	?	Brasil	
U95570	Filipinas	1992	Homossexual	?	
U95566	Filipinas	1991	Heterossexual	Brasil ?	Viagem ao Extremo Oriente, à África e à América do Sul
U95575	Filipinas	1993	Heterossexual	Brasil	Viagem à Europa, ao Brasil e às Ilhas Malvinas
U58813	França	1993	Heterossexual	Dibuti	Vigem ao Dibuti
U58820	França	1995	Heterossexual	França	Vigem ao Dibuti
AF188941	França	?	Homossexual	França	
AF188956	França	?	Homossexual	?	
AF032151	Holanda	1991	?	Brasil	Imigrante da Turquia
Z68025	Holanda	1991	Homossexual	?	Soroconversão em 1991
Y18596	Ilhas Reunidas	1996	Transfusão	França	Transfusão recebida na França
AB012971	Japão	1997	Homossexual	?	
L38420	Lituânia	1992	Homossexual	?	
AF200859	México	1998	Homossexual	Brasil	Diagnóstico sorológico em 1996
U28955	Paraguai	1995	?	Brasil	
AF223996	R. Checa	1997	Homossexual	?	
AF223998	R. Checa	1997	?	?	
AF181437	Suíça	1992	Homossexual	?	Diagnóstico sorológico em 1992

A presença da variante GWGR no Japão é conhecida desde 1991. No artigo original é referido que os vírus GWGR foram encontrados em pacientes infectados por contato sexual, dos quais um havia adquirido a infecção na América do Sul (Hattori et al., 1991). Contudo, estas seqüências não estão disponíveis no Genbank e não puderam ser incluídas na análise filogenética. Por outro lado, a seqüência GWGR japonesa incluída na análise foi obtida de um homossexual do sexo masculino que negava viagens ao estrangeiro e esta se agrupou dentro do ramo estrangeiro junto com seqüências da China, da Alemanha e da Suíça (Kitsutani et al., 1998).

Ainda no ramo estrangeiro estão incluídas: uma das três seqüências GWGR identificadas nas Filipinas; uma das duas seqüências GWGR holandesas, uma das duas seqüências da República Checa e duas das quatro seqüências francesas. A seguir, a introdução da variante nestes países será estudada separadamente.

Dentre as quatro seqüências GWGR obtidas na França, duas foram encontradas em amostras colhidas de homossexuais masculinos (Casado et al., 2000b). Por outro lado, as outras duas seqüências francesas foram derivadas de amostras colhidas de marinheiros franceses que supostamente se infectaram no Dibuti entre 1993 e 1995 (Lasky et al., 1997). Na árvore filogenética, as duas seqüências francesas obtidas de homossexuais e a seqüência de um dos marinheiros se agruparam no ramo estrangeiro, dentro de um ramo interno com 80% de suporte estatístico junto com seqüências do Brasil, da República Checa e da Espanha. Ao contrário do que é descrito no artigo científico original, a similaridade entre a seqüência do marinheiro e as outras duas seqüências francesas sugere que, na verdade, este se infectou na França antes de embarcar para o estrangeiro, e não no Dibuti (Lasky et al., 1997). Naturalmente, não se pode descartar a possibilidade de a infecção ter ocorrido na República Checa ou na Espanha. Por outro lado, a seqüência do outro marinheiro francês se agrupou com uma seqüência brasileira e fora do ramo estrangeiro (ramo interno com 96% de suporte estatístico).

A presença dos vírus GWGR na República Checa (Reinis et al., 2002), na Holanda (Op de Coul et al., 2001b) e nas Filipinas (Paladin et al., 1998) parece resultar de múltiplas introduções da variante nestes países. Na árvore filogenética, as duas seqüências checas se agruparam em ramos distintos, uma das quais dentro do ramo estrangeiro junto com seqüências do Brasil, França e Espanha (ramo interno com 80% de suporte). A outra seqüência checa se localizou fora do ramo estrangeiro formando um ramo interno com 85% de suporte com seqüências do Brasil, Paraguai e Filipinas.

Quanto às Filipinas, somente uma seqüência se agrupou no ramo estrangeiro, junto com seqüências do Brasil, Dinamarca e Lituânia (83% de suporte). De fato, o paciente do qual foi obtida a amostra tem história de viagens à Europa, às Ilhas Malvinas e ao Brasil. As outras duas seqüências das Filipinas se posicionaram fora deste ramo, uma das quais foi encontrada em um paciente que relatava viagens ao Extremo Oriente, à África e à América do Sul e se agrupou com uma seqüência do Brasil. A terceira seqüência foi obtida de um homossexual masculino e se agrupou com uma seqüência do Brasil e outra da Bolívia.

Duas seqüências GWGR foram encontradas na Holanda. Na árvore filogenética, uma destas seqüências se agrupou com uma seqüência do Brasil (99% de suporte) e foi encontrada em um paciente heterossexual (Op de Coul et al., 2001b). A outra seqüência foi obtida de um homossexual e se agrupou fora do ramo estrangeiro (Kuiken et al., 1996).

Não foi possível resgatar qualquer informação relacionada à seqüência dos Estados Unidos, porém esta se agrupou fora do ramo estrangeiro, junto com seqüências brasileiras na árvore filogenética (Zhang et al., 1999).

Do que foi exposto acima, os dados epidemiológicos disponíveis e a análise da distribuição das seqüências na árvore filogenética sugerem que a introdução da variante GWGR na Argentina, no Paraguai, no México, na Bolívia, nos Estados Unidos, na França, nas Filipinas e na Holanda tenha ocorrido diretamente do Brasil. Uma outra observação que pode ser feita se refere ao fato de a maioria das seqüências GWGR encontradas fora da América Latina (65%) ter se localizado em um único ramo interno. De fato, o agrupamento destas seqüências pode sugerir que a variante tenha sido exportada do Brasil para algum dos países representados neste ramo e, posteriormente, teria dispersado para as outras regiões. Dos países que têm seqüências incluídas no ramo estrangeiro, aquele que tem o indício mais antigo da presença da variante GWGR na epidemia local é a França. Este indício vem do fato do paciente das Ilhas Reunidas com vírus GWGR ter se infectado na França, provavelmente na primeira metade da década de 80.

5. Discussão

Os resultados da tese sugeriram que a epidemia pela variante GWGR do subtipo B do HIV-1 tem uma única origem e que o Brasil foi o seu epicentro. Contudo, apesar da variante ter sido encontrada em 18 países, sua dispersão foi observada principalmente no Brasil. Por que isto ocorreu?

Antes de discutirmos a questão, vamos fazer algumas considerações. Em primeiro lugar, um fato que chama a atenção é que, enquanto a variante GWGR provavelmente se disseminou na epidemia do HIV-1 a partir de um único vírus ancestral, os vírus parentais GPGR circulam no Ocidente desde o início da década de 80 e sua presença nos países onde o subtipo B é majoritário resulta de múltiplas re-introduções desses vírus. Este fato, além de ser descrito na literatura (Kuiken et al., 2000b; Op de Coul et al., 2001b), foi também demonstrado pela distribuição das seqüências GPGR e GWGR na árvore filogenética discutida no trabalho de tese. Nessa árvore, as seqüências com os dois padrões de assinatura do Brasil, por exemplo, se agruparam com seqüências de diversos países (figura 9).

Naturalmente, o agrupamento de vírus GPGR brasileiros com seqüências de outros países na árvore filogenética não indica a direção da migração destes vírus, ou seja, não indica se o Brasil exportou ou recebeu cepas com este motivo protéico. De qualquer forma, uma vez que as epidemias regionais pelo HIV-1 não são isoladas, a migração dos vírus GPGR no Brasil ocorreu obviamente em ambas as direções. Deve ser lembrado que, nos países da América Latina que têm fronteira com o Brasil, o subtipo B é o responsável pela grande maioria das infecções pelo HIV-1 (Cabello et al., 1995; Campodonico et al., 1996; Carrillo et al., 1992; Rivera-Morales et al., 2001; Velarde-Dunois et al., 2000).

Desta forma, no Brasil, poder-se-ia esperar que, ao longo dos últimos 20 anos, a proporção de infecções pela variante GWGR diminuísse enquanto o número de casos atribuídos aos vírus GPGR aumentasse progressivamente devido às suas múltiplas re-introduções. De fato, a comprovada introdução no país de outros subtipos, como A, C, D e F (Bongertz et al., 2000; Guimarães et al., 2002; Morgado et al., 1998), após a explosão da epidemia pelo subtipo B, mostra que o HIV-1 teve oportunidade de entrar no país em diversas ocasiões.

Contudo, a diminuição progressiva do número de infecções pela variante GWGR

no Brasil não parece ter ocorrido. No registro mais antigo da presença da variante no país, estes vírus foram encontrados em proporções semelhantes aos vírus GPGR (Michael-Hendry et al., 1996) e, nos trabalhos subsequentes, proporções semelhantes das duas variantes do subtipo B também foram encontradas no Brasil (Couto-Fernandez et al., 1999; Covas et al., 1998). Além disto, no trabalho de tese, quando todas as seqüências brasileiras depositadas no Genbank foram recuperadas, as proporções de seqüências GPGR e GWGR do subtipo B disponíveis foram, respectivamente, 45% e 38%. Deve-se levar em conta, entretanto, que todos os estudos desenvolvidos para a caracterização molecular (e antigênica, feitas nos estudos de sorotipagem) dos vírus brasileiros não utilizaram amostragens representativas da população infectada, o que limita a comparação da magnitude da epidemia pelas duas variantes do subtipo B, GPGR e GWGR.

Mesmo considerando a limitação de amostragem apontada no parágrafo anterior, nesta seção do trabalho de tese, no item 5.1, “Seleção Natural”, será discutida a hipótese de que os vírus GWGR possuam alguma vantagem seletiva em relação aos seus parentais que possa ter favorecido a sua manutenção na epidemia brasileira em proporções semelhantes às dos vírus GPGR, apesar da primeira ter provavelmente uma única origem e os últimos terem sido repetidamente introduzidos no país.

Uma alternativa para a hipótese de que a disseminação da variante GWGR no Brasil tenha sido favorecida por alguma vantagem seletiva que estes vírus possuam em relação aos parentais GPGR é que a sua dispersão tenha ocorrido devido a fatores não relacionados à aptidão do vírus. Dentre estes fatores os mais importantes seriam: 1) as condições locais da epidemia brasileira no momento em que a variante GWGR foi introduzida no país e 2) a composição genética da população brasileira. Estas possibilidades são abordadas na seção: 5.2, “Outros fatores”.

5.1. Seleção Natural

Para que a seleção natural tenha atuado sobre a dispersão da variante GWGR na epidemia brasileira é necessário considerar, em primeiro lugar, que os vírus GWGR e GPGR possuam alguma diferença em seus fenótipos. Em segundo lugar, se a diferença no fenótipo realmente existe, é necessário que esta confira à variante GWGR uma chance maior de se perpetuar na epidemia do que a chance da variante GPGR. Isto significa que, por alguma razão, a variante GWGR seja mais apta para “sobreviver” e

“deixar descendentes” na população humana do que os vírus GPGR. Contudo, que vantagem seletiva seria esta?

De um modo geral, para os vírus, a aptidão (*fitness*) é um parâmetro que define o seu grau de adaptação em um ambiente biológico (Domingo et al., 1997), ou seja, nos seus hospedeiros. Quanto aos vírus RNA, como o HIV-1, sua aptidão se expressa sob duas formas: 1) por sua capacidade de ultrapassar as barreiras do sistema imune inato e, assim, infectar novos hospedeiros, ou seja, por sua transmissibilidade; e 2) após penetrar no organismo do hospedeiro, por sua capacidade de escapar do sistema imune adaptativo e manter uma infecção persistente através da produção contínua de novas partículas virais (Quiñones & Arts, 2001).

Na literatura, há alguns anos, foi sugerido que os vírus dos subtipos B, C e o recombinante CRF_01 AE diferiam quanto à transmissibilidade (Essex, 1999; Workshop Report from the European Commission DG XII, 1997). Isto porque, principalmente na Tailândia, os vírus do subtipo B estão especialmente associados a infecções por uso de drogas injetáveis enquanto os vírus CRF_01 AE são mais freqüentemente encontrados nos indivíduos infectados por contato sexual. Uma vez que uma diferença semelhante foi encontrada entre os subtipos B e C na África do Sul (van Harmelen et al., 1997), alguns trabalhos *in vitro* foram conduzidos no sentido de estudar a eficiência destes vírus em infectar diferentes células em cultura (Dittmar et al., 1997; Karita et al., 1997; Pope et al., 1997; Soto-Ramirez et al., 1996).

Desta forma, foi inicialmente mostrado que os vírus C e CRF_01 AE eram mais eficientes em infectar células de Langerhans do que vírus do subtipo B. Uma vez que estas células são encontradas em grandes quantidades na vagina, no cérvix e na glândula, e, via de regra, estão ausentes no reto, os autores sugeriram que o tropismo celular observado *in vitro* fosse um indicador da eficiência de transmissão por contato heterossexual (Soto-Ramirez et al., 1996). Contudo, trabalhos subseqüentes não confirmaram estes resultados (Dittmar et al., 1997; Karita et al., 1997; Pope et al., 1997). Assim, se existem diferenças na transmissibilidade de variantes do HIV-1, estas diferenças não foram ainda demonstradas.

No caso das duas variantes do subtipo B, GPGR e GWGR, o tipo de desenho de estudo utilizado (a abordagem ideal para o problema seria o desenvolvimento de um estudo prospectivo) e as informações disponíveis dos pacientes infectados não permitem que se faça qualquer análise sobre a existência de diferenças no seu fenótipo que

interferiram na sua transmissibilidade. Estas limitações, associadas à ausência de evidências na literatura quanto a diferenças na transmissibilidade dessas duas variantes, impossibilita a abordagem comparativa de sua aptidão em relação ao fenótipo transmissibilidade.

Contudo, nos últimos anos, a possibilidade dos subtipos de HIV-1 estarem associados às diferenças no tempo de sobrevida sem AIDS vem sendo explorada por alguns autores. Uma limitação dos trabalhos desenvolvidos, contudo, é que em sua maioria estes foram estudos seccionais e que envolviam um número reduzido de indivíduos (Kaleebu et al., 2001; Neilson et al., 1999), enquanto a melhor abordagem para se responder à questão seria o estabelecimento de coortes de casos incidentes em populações onde mais de um subtipo circulam (Workshop Report from the European Commission DG XII, 1997). O único estudo de seguimento, comparando progressão da imunodeficiência em indivíduos infectados com diferentes subtipos de HIV1, sugere que o subtipo A está associado a um período de sobrevida sem AIDS mais longo do que os subtipos C, D e G (Kanki et al., 1999).

Porém como uma menor taxa de progressão para a AIDS poderia conferir uma maior aptidão aos vírus que causam estas infecções?

De uma forma geral, o modo como a virulência de qualquer microorganismo patogênico (sendo virulência definida como o potencial – probabilidade ou taxa - de um patógeno produzir morbidade ou mortalidade em hospedeiros infectados) está relacionada com a sua transmissibilidade é importante, uma vez que a transmissão determina o sucesso evolutivo do microorganismo. Classicamente, a premissa subjacente aos estudos sobre a evolução de patógenos ditava que a seleção natural favoreceria a evolução em direção à coexistência benigna entre o hospedeiro e o parasita, ou seja, que a seleção natural favoreceria cepas com reduzida virulência. A lógica por trás deste conceito é que hospedeiros vivos e móveis podem transmitir os organismos que os infectam de forma mais eficiente do que os hospedeiros debilitados (Ewald, 1996; Lipsitch & Moxon, 1997).

Contudo, estudos recentes sobre a evolução da virulência enfocam que os patógenos estão sujeitos a uma “barganha” (*trade off*) entre os benefícios competitivos que ganham com a maior exploração de seus hospedeiros e o preço que pagam por qualquer efeito da doença por eles causada que reduza o contato infeccioso entre os hospedeiros infectados e os susceptíveis. A premissa básica da teoria da “barganha” é que maior virulência está correlacionada com maior propagação do patógeno. Nos

modelos propostos segundo essa teoria, a seleção natural pode favorecer mesmo as cepas com níveis intermediários ou elevados de virulência, desde que essas cepas mais virulentas possuam alguma vantagem evolutiva capaz de compensar a menor oportunidade de transmissão que resulta da morte ou da debilidade dos hospedeiros. Um exemplo seria as altas taxas de replicação no organismo dos hospedeiros observadas em alguns parasitas, como o HIV. As altas taxas de replicação têm sido relacionadas à maior transmissibilidade e virulência; neste caso, a maior virulência seria compensada pela persistência dos patógenos nos hospedeiros devido ao escape de sua depuração pelo sistema imune (Ewald, 1996; Lipsitch & Moxon, 1997).

Segundo Ewald (Ewald, 1996), no caso das doenças de transmissão sexual, um dos fatores dos quais depende esta “barganha evolutiva” é o comportamento sexual do hospedeiro. Nesse caso, infecções de curta duração seriam ineficientes para a maioria dos patógenos a menos que os seus hospedeiros mudassem de parceiros a cada poucos dias, antes que os microorganismos fossem eliminados pelo sistema imune ou causassem a morte do hospedeiro, por exemplo. Desta forma, os patógenos sexualmente transmissíveis precisam permanecer infecciosos por períodos de tempo que atravessam mais do que uma única mudança de parceiro sexual.

No caso do HIV-1, a infecção é normalmente detectada quando o paciente apresenta alguma manifestação clínica associada ao vírus ou quando teve contato sexual com um parceiro infectado (ou sob risco para a infecção), é usuário de drogas injetáveis ou recebeu transfusão de sangue. Assim, se o paciente estiver infectado com uma cepa menos patogênica do que as cepas circulantes, a infecção deste paciente demorará mais tempo para ser detectada do que as infecções causadas pelas demais cepas, uma vez que este paciente levará mais tempo para apresentar-se doente (deve-se, contudo, levar em conta que o tempo de progressão para a AIDS é extremamente variável). Além disso, durante o tempo em que o paciente permanecer assintomático, ele poderá manter o seu comportamento social de antes da infecção, como o seu comportamento sexual habitual ou o uso de drogas injetáveis, por exemplo.

Em relação às duas variantes do subtipo B, GPGR e GWGR, Santoro-Lopez et al e Casseb et al, no Brasil, mostraram que o tempo de sobrevivência sem AIDS para os pacientes infectados pelo sorotipo GWGR é superior ao tempo de sobrevivência dos pacientes infectados pelo sorotipo GPGR (Casseb et al., 2002; Santoro-Lopes et al., 2000). Nesse caso, uma menor taxa de progressão para a AIDS poderia significar, para a variante GWGR, uma maior chance de infectar novos hospedeiros antes que sua presença fosse

descoberta do que vírus mais agressivos que circulam no restante da população; a maior chance de deixar descendentes é um dos fatores sobre os quais a seleção natural pode atuar. Em outras palavras, a menor virulência seria favorecida e compensada por uma maior chance da variante ser transmitida para novos hospedeiros e, assim, se perpetuar na população.

Desta forma, decidiu-se investigar se as duas variantes do subtipo B possuem algum marcador que indique a existência de diferenças na sua patogenia. A discussão dos resultados encontrados é apresentada a seguir.

5.1.1 Fenótipo SI nas seqüências GPGR e GWGR

Como dito na introdução da tese, os isolados de HIV-1 recuperados nos diferentes estágios da imunodeficiência têm fenótipos distintos. Resumidamente, os vírus que circulam no hospedeiro humano nos primeiros anos da infecção geralmente se caracterizam por possuírem baixas taxas de replicação, por não induzirem a formação de sincícios *in vitro* (fenótipo NSI), por se replicarem com mais eficiência em macrófagos do que em células T transformadas e por utilizarem preferencialmente o co-receptor CCR5. Por outro lado, os vírus isolados de pacientes nas fases mais avançadas da imunodeficiência apresentam altas taxas de replicação, se replicam com maior eficiência em células T e utilizam o co-receptor CXCR4. Além disto, em aproximadamente 50% dos indivíduos nos estágios tardios da infecção, uma parcela da população de quasispecies adquire a capacidade de formar sincícios (fenótipo SI) (Fouchier & Schuitemaker, 1996; Quiñones-Mateu & Arts, 2001).

Essas mudanças no fenótipo dos isolados de HIV-1 observadas durante a progressão da imunodeficiência parecem envolver diversas estruturas do HIV-1. Contudo, a associação mais marcante envolve a capacidade de induzir a formação de sincícios e o domínio da região V3 do gene *env*. Isto porque a mudança para o fenótipo SI que ocorre nos estágios finais da imunodeficiência está relacionada à substituição de dois aminoácidos na alça V3 localizados nas posições 306 e 320 do ENV. Assim, no início da infecção, na maioria dos isolados, estes sítios são ocupados, respectivamente, por resíduos de serina (S) e de ácido glutâmico (E), que são substituídos por resíduos de arginina (R, posição 306) e arginina ou lisina (R ou K, posição 320) nos vírus SI. Em conseqüência à substituição destes dois aminoácidos ocorre um aumento na carga molecular da região V3, uma vez que a lisina e a arginina possuem cargas positivas (Fouchier & Schuitemaker, 1996).

Para se estudar se as seqüências GPGR e GWGR incluídas no trabalho de tese possuem proporções diferentes de mutações associadas à progressão para a AIDS nos sítios 306 e 320 da alça V3, os aminoácidos encontrados nestes sítios nos dois grupos de seqüências foram analisados. A tabela 7 apresenta as seqüências com os motivos GPGR e GWGR, nas quais pelo menos uma das substituições de aminoácidos associadas à indução de sincício foi encontrada.

Tabela 7 Seqüências com os motivos GPGR e GWGR com substituição de aminoácidos nas posições 302 e 320 da alça V3

GPGR			GWGR		
Identificação	país	ano	Identificação	país	ano
X78513	Brasil	1988	U08975	Brasil	1992
U08446	Haiti	1992	U37041	Argentina	1995
AY037268	Argentina	1984	U28955	Paraguai	1995
X78515	Brasil	1988			
X78517	Brasil	1988			
U08772	Brasil	1992			
AF113570	Brasil	1996			
AF200874	México	1998			
AF034027	Brasil	1996			
U95485	Itália	?			
AF137974	EUA	1983			
Y18760	Brasil	1995			
AF368558	França	?			
AB059299	Japão	?			
U00427	Brasil	1992			
U20019	China	1991			
AF076319	Brasil	1997			
AF142260	Espanha	2000			

Como pode ser observado na tabela 7, das 98 seqüências GPGR 18 (18%) possuem pelo menos uma das substituições no topo da alça V3 associadas ao fenótipo SI. Por outro lado, no grupo GWGR (n=81) a proporção foi de três seqüências em um total de 81 (3,7%).

Uma questão que poderia ser apontada em relação aos resultados apresentados na discussão é que a menor proporção de mutações nas posições 306 e 320 da alça V3

no grupo GWGR seria uma decorrência natural do fato das seqüências GWGR se agruparem em um único ramo com 87% de probabilidade de confiança na árvore filogenética, ou seja, devido ao fato destas seqüências representarem um grupo mais homogêneo de vírus do que a variante GPGR. De fato, a distância média entre as seqüências GWGR, calculada pelo modelo de Tamura-3, foi 0,135, inferior à obtida dentro do grupo GPGR, que foi de 0,154. Contudo, a diferença entre estes dois valores de distância foi pequena e provavelmente insuficiente para justificar os resultados na seção de discussão.

Uma outra questão é que, uma vez que a epidemia pelos vírus GPGR no mundo é anterior à epidemia pela variante GWGR, as seqüências GPGR incluídas na análise podem ter sido obtidas de pacientes com um tempo de infecção mais longo do que aqueles infectados com a variante GWGR. Assim, os pacientes infectados com vírus GPGR teriam uma maior probabilidade de portarem cepas com o fenótipo SI do que os pacientes infectados com a variante GWGR simplesmente por terem uma infecção pelo HIV-1 mais antiga. Contudo, deve ser reforçado que o tempo de progressão para AIDS é muito variável entre os pacientes infectados, o que limita este tipo de inferência.

Ao contrário do que ocorre no resto do mundo, há indícios de que as duas variantes do subtipo B circulam no Brasil desde o início da epidemia brasileira (Michael-Hendry et al., 1996). Neste caso, as seqüências das duas variantes no Brasil podem ter tido chances semelhantes de terem sido obtidas de amostras colhidas de pacientes em estágios tardios (ou precoces) da infecção. Além disto, as seqüências brasileiras GPGR e GWGR depositadas no Genbank foram obtidas dos mesmos trabalhos e todas foram incluídas na análise. Considerando-se que, no Brasil, as epidemias pelas duas variantes são contemporâneas, as proporções de seqüências GPGR e GWGR brasileiras com mutações na alça V3 associadas ao fenótipo SI foram calculadas. Esta análise mostrou que, no conjunto de amostras brasileiras, as mutações associadas ao fenótipo SI estão presentes em nove das 55 seqüências GPGR (16,4%) e em somente uma das 54 seqüências GWGR (1,8%) (tabela 7).

Para se estudar se os resultados obtidos com as seqüências brasileiras poderiam ser viciados pelo ano de coleta das amostras das quais as seqüências foram derivadas, a distribuição temporal destas seqüências foi considerada e é apresentada na tabela 8.

Tabela 8 Distribuição das seqüências GPGR e GWGR segundo ano de coleta da amostra.

Ano	GPGR			GWGR		
	Todos	Brasil	Mundo	Todos	Brasil	Mundo
1983	2	-	2	-	-	-
1984	1	-	1	-	-	-
1987	3	2	1	-	-	-
1988	4	3	1	1	1	-
1989	1	-	1	-	-	-
1990	3	2	1	7	7	-
1991	5	1	4	3	-	3
1992	20	14	6	12	10	2
1993	6	5	1	9	6	3
1994	1	-	1	5	5	-
1995	15	14	1	11	8	3
1996	18	13	5	18	12	6
1997	3	1	2	4	1	3
1998	2	-	2	1	-	1
1999	2	-	2	-	-	-
2000	1	1	1	-	-	-
ND	11	-	11	10	4	6
Total	98	55	43	81	54	27
Média	1993	1993	1992	1994	1994	1994

Como pode ser visto na tabela, a maioria das seqüências brasileiras provém de amostras colhidas entre os anos de 1992 a 1996. Neste período, foram obtidas 84% (46/55) das amostras GPGR brasileiras e 82% (41/50) das amostras GWGR. Além disto, os trabalhos brasileiros nos quais estas seqüências foram geradas foram desenvolvidos com o objetivo de se conhecer a diversidade do HIV-1 no Brasil, sem que tivessem empregado alguma abordagem que favorecesse a identificação de alguma das duas variantes, o que torna improvável a ocorrência de vício em direção à identificação de vírus GPGR ou GWGR em momentos específicos da epidemia. Assim, o fato de a maioria das seqüências dos dois grupos ter sido encontrada em amostras colhidas no mesmo período sugere que a maior proporção de mutações associadas ao fenótipo SI das seqüências GPGR brasileiras não deve ser atribuída a diferenças nas datas de coleta.

Quando a média dos anos de coleta foi calculada para cada grupo, foi possível observar que as seqüências GWGR brasileiras são um pouco mais recentes do que as seqüências GPGR (o ano médio para o grupo GPGR foi 1993 e para o grupo GWGR foi 1994). Esta diferença entre as médias nos anos de coleta poderia ser atribuída ao fato de os pacientes infectados pelos vírus GWGR terem um período de infecção assintomática mais longo do que os pacientes GPGR. Entretanto, uma vez que todas as seqüências brasileiras foram obtidas de amostras colhidas por conveniência, esta diferença pode simplesmente ser consequência de erros de amostragem.

A menor proporção de seqüências GWGR com mutações associadas à formação de sincícios em relação às seqüências GPGR sugere que a evolução da variante GWGR em direção ao fenótipo SI seja mais lenta do que a evolução da variante GPGR. Mesmo à luz das limitações apontadas anteriormente, esses resultados poderiam explicar o fato de os pacientes infectados com o sorotipo GWGR possuírem um período de sobrevivência maior sem AIDS do que aqueles infectados pelo sorotipo GPGR, como já observado (Casseb et al., 2002; Santoro-Lopes et al., 2000).

Se os indivíduos infectados pelos vírus GWGR experimentam um período de infecção assintomática mais longo do que aqueles que portam a variante parental GPGR, isto poderia significar, para a variante GWGR, uma maior oportunidade de ser transmitida para novos hospedeiros, ou seja, de deixar descendentes, frente aos vírus GPGR. Esta diferença no fenótipo dos dois vírus, então, conferiria à variante GWGR uma aptidão para se perpetuar na epidemia superior à dos seus parentais. Por sua vez, a diferença na aptidão relativa das duas variantes explicaria o fato dos vírus GWGR serem encontrados na epidemia brasileira em proporções semelhantes aos vírus GPGR, mesmo tendo se originado de um único ancestral comum.

Por outro lado, o fato de os pacientes infectados pela variante GWGR terem uma sobrevivência maior do que aqueles infectados pelos vírus GPGR significa que os primeiros permanecem mais tempo na população e tiveram maiores chances de serem amostrados nos trabalhos desenvolvidos no Brasil. Este fato pode ter influenciado o número de seqüências GWGR encontradas no Genbank, todas obtidas de estudos seccionais, na maioria dos quais não havia referência quanto à época de soroconversão dos pacientes.

5.1.2. Por que as diferenças nos fenótipos?

Agora, o que determinaria este fenótipo vantajoso para a variante GWGR?

Por definição, esses vírus foram identificados devido à substituição da prolina por um resíduo de triptofano no topo da alça V3. Nesse contexto é necessário considerar se: 1) o triptofano encontrado no motivo protéico no topo da alça V3 é essencial para as diferenças observadas nos fenótipos das duas variantes ou 2) se simplesmente representa um marcador para estes vírus. Neste último caso, outras mutações na região do gene *env* também estariam envolvidas nas diferenças fenotípicas associadas à progressão da imunodeficiência.

Comparando-se somente a região V3 dos dois grupos de seqüências, pode-se observar que as seqüências consenso das variantes GPGR e GWGR (tabela 9) possuem, com uma única exceção, os mesmos aminoácidos em cada sítio analisado (além de, naturalmente, a substituição da prolina pelo triptofano). Tal exceção ocorre na posição 309, adjacente ao motivo protéico do topo da alça V3. Assim, enquanto na maioria das seqüências GPGR este sítio é ocupado por um resíduo de isoleucina (I), na maior parte das seqüências GWGR a isoleucina está substituída por metionina (M). A presença deste resíduo de metionina adjacente ao motivo protéico já havia sido mencionada durante a discussão da origem da variante GWGR.

Na literatura, é descrito que a prolina na segunda posição do motivo protéico é um aminoácido extremamente conservado (Kuiken et al., 2000a), sendo encontrado em todos os subtipos do HIV-1 e nas seqüências dos vírus da imunodeficiência de chimpanzés, o SIV_{cpz} (Kuiken et al., 2000a). Esta extrema conservação da prolina mostra que, neste caso, existe uma forte pressão seletiva contra a substituição deste aminoácido. Esta seleção a favor da prolina indica que a sua conservação deve ser importante para a conformação e função biológica da região V3.

Contudo, as implicações das substituições da prolina pelo triptofano e, da mesma forma, da isoleucina pela metionina, observadas nas seqüências GWGR para a conformação da alça V3, ainda não foram investigadas. Uma possibilidade é que, uma vez que o triptofano e a metionina possuem características moleculares e propriedades físico-químicas diferentes da prolina e da isoleucina, respectivamente, as duas mutações sejam necessárias para que a conformação da região V3 dos vírus GWGR seja mantida.

Tabela 9- Seqüências consenso construídas para todo o conjunto de seqüências GWGR e GPGR incluídas na análise filogenética e a proporção de sítios conservados em cada grupo. As setas indicam as posições 306 e 320 onde ocorrem as mutações associadas ao fenótipo SI.

		Sítios de aminoácidos na alça V3																																		
		296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330
P		C	T	R	P	N	N	N	T	R	K	S	I	H	I	G	P	G	R	A	F	Y	A	T	G	E	I	I	G	D	I	R	Q	A	H	C
%		99	89	100	98	82	93	94	97	96	80	66	97	69	81	94	100	100	79	90	76	87	63	97	84	40	93	87	99	81	99	99	83	98	85	100
W		C	T	R	P	N	N	N	T	R	K	S	I	H	M	G	W	G	R	A	F	Y	A	T	G	E	I	I	G	D	I	R	Q	A	H	C
%		100	90	100	99	64	98	95	100	96	93	76	98	94	63	95	100	99	100	75	84	96	75	90	98	65	98	86	100	76	99	98	88	99	84	99
												^											^													

Assim, apesar da extrema conservação da prolina, a expansão de vírus com o padrão de assinatura GWGR na epidemia indica, dentro de uma visão mais conservadora, que o aminoácido triptofano também pode ser tolerado nesse sítio da região V3 sem que a sua presença torne estes vírus inaptos para sobreviver no hospedeiro humano. Isto porque, se este fosse o caso, essa variante teria sido eliminada pela seleção purificadora, o que não ocorreu. Ao invés disto, em relação à variante GWGR, esta não somente se estabeleceu na epidemia brasileira, como parece possuir um fenótipo mais vantajoso em relação aos vírus GPGR.

Um fato que deve ser discutido com os resultados encontrados no trabalho de tese é a recente disseminação de vírus do subtipo C na pandemia do HIV-1 (Osmanov et al., 2002). Nos últimos anos, tem sido observado que o subtipo C é, dentre todas as variantes do HIV-1, o responsável pelo maior número de casos incidentes de infecções. Alguns exemplos mostram que, em regiões onde anteriormente a epidemia era dominada por outros subtipos, como Índia, China e África sub-sahariana, atualmente a maioria das infecções é atribuída ao subtipo C (Osmanov et al., 2002). No Brasil, apesar do subtipo B ser o mais prevalente, os casos associados ao subtipo C têm aumentado, especialmente na região sul do país (Guimarães et al., 2002).

Recentemente, diversos estudos têm sido conduzidos para avaliar os fatores subjacentes à disseminação do subtipo C. Neste caso, o topo da alça V3 não estaria relacionado a este fato, uma vez que as seqüências do subtipo C possuem o motivo protéico GPGQ no topo da alça, sendo este o padrão de assinatura mais comum entre as seqüências dos subtipos não-B de HIV-1 (Kuiken et al., 2000a).

Contudo, é descrito na literatura que os vírus do subtipo C não evoluem, durante o curso da infecção, para o fenótipo SI e nem utilizam, nas fases tardias, o co-receptor CXCR4, como é normalmente observado para os vírus dos outros subtipos (Treurnicht et al., 2002). Associa-se a estas diferenças a observação de que as mutações para aminoácidos básicos que ocorrem normalmente nas posições 306 e 320 da alça V3 durante a progressão da imunodeficiência estão ausentes nos vírus do subtipo C (Ping et al., 1999; Treurnicht et al., 2002). Comparando-se estas observações com os resultados do presente trabalho, apesar de as mutações associadas ao fenótipo SI terem sido encontradas nas seqüências GWGR, estas mutações foram menos freqüentes nesse grupo do que entre as seqüências GPGR.

Além dos vírus do subtipo C não adquirirem o fenótipo indutor de sincício, Ball e colaboradores mostraram que vírus dos subtipos B e C diferem em relação à sua capacidade de se replicar em células de diversas linhagens. Mais especificamente, os

autores mostraram que os isolados do subtipo C se replicam, *in vitro*, de forma menos eficiente em células mononucleares e em linfócitos T CD4⁺ - os principais alvos do vírus durante a infecção pelo HIV - do que os vírus do subtipo B. Por outro lado, os dois subtipos demonstraram possuir a mesma aptidão para se replicar em células de Langerhans, as quais são abundantes no trato genital e participam da infecção primária através do contato sexual (Ball et al., 2003).

A partir dos seus resultados, os autores sugerem que os vírus do subtipo C seriam tão eficientes quanto os vírus B com relação à sua transmissão por contato sexual. Contudo, após infectarem um novo hospedeiro, os vírus C, por se replicarem com menos eficiência em linfócitos e macrófagos, determinariam uma progressão mais lenta da imunodeficiência do que os vírus do subtipo B. Assim, a menor eficiência de replicação conferiria a estes vírus, na verdade, uma aptidão maior dentro da população humana. Ainda segundo os autores, seria essa a explicação para a rápida expansão do subtipo C observada na pandemia pelo HIV-1 (Ball et al., 2003).

Comparando o modelo proposto para a expansão mundial do subtipo C (Ball et al., 2003) com o modelo sugerido no presente trabalho para explicar a dispersão da variante GWGR, verificamos que em ambos os casos os vírus em questão se caracterizam pela ausência (subtipo C) ou por um menor número (variante GWGR) de mutações associadas à progressão da imunodeficiência. Nos dois casos, os fenótipos associados à progressão mais lenta para a AIDS confeririam a ambos uma maior oportunidade de serem transmitidos para novos hospedeiros. Desta forma, estes vírus seriam mais aptos a deixar descendentes na população humana do que outras variantes do HIV-1.

Em relação ao subtipo C, sua capacidade de gerar novas infecções parece ser superior à dos demais subtipos do grupo M uma vez que, como já foi exposto no texto, este subtipo é o responsável pelo maior número de casos incidentes de infecção pelo HIV-1 que ocorreram nos últimos anos (Osmanov et al., 2002). No caso da variante GWGR, uma capacidade para gerar novas infecções maior do que a dos vírus parentais GPGR poderia explicar a sua manutenção na epidemia brasileira ao longo de duas décadas, apesar das múltiplas introduções dos vírus GPGR no Brasil.

Contudo, se a diferença na aptidão relativa das duas variante do subtipo B realmente existe, ela não deve ser muito acentuada, uma vez que, se este fosse o caso, a epidemia GWGR teria provavelmente uma maior magnitude nos demais países onde a variante circula e, no Brasil, seria majoritária ou mesmo única. Ainda assim, se a

hipótese proposta for verdadeira, deve-se esperar que, nos próximos anos, a proporção de infecções pela variante GWGR aumente nos demais países onde foi encontrada.

Um contra-exemplo para a hipótese proposta na discussão da tese é a epidemia pelo HIV-2, associada a uma menor transmissibilidade e patogenicidade deste vírus quando comparada à epidemia pelo HIV-1.

Na África Ocidental, onde o maior número de casos de infecção pelo HIV-2 é encontrado, as principais vias de infecção são o contato heterossexual e o contato com sangue contaminado. Contudo, a eficiência na transmissão sexual do HIV-2 parece ser de cinco a nove vezes inferior à do HIV-1. Uma outra diferença quanto à transmissibilidade dos dois vírus é que o HIV-2 é raramente transmitido verticalmente, enquanto um terço das gestantes soro-positivas para o HIV-1 e não tratadas com anti-retrovirais transmitem a infecção para seus filhos (Bock & Markovitz, 2001; Reeves & Doms, 2002).

Em relação à patogenicidade, apesar do HIV-2 também causar AIDS, a progressão da imunodeficiência associada a este vírus é normalmente mais lenta do que a observada nas infecções pelo HIV-1 (Bock & Markovitz, 2001; Reeves & Doms, 2002). Desta forma, poder-se-ia argumentar que, analogamente ao que foi proposto para o subtipo C e a variante GWGR, também neste caso o HIV-2 teria uma maior chance de ser transmitido para novos hospedeiros do que as demais variantes do HIV-1.

Entretanto, de um modo geral, uma importante diferença entre os dois tipos de HIV é que as infecções causadas pelo HIV-2 estão associadas a uma menor taxa de replicação e, conseqüentemente, a níveis de carga viral mais baixos do que as infecções pelo HIV-1. Os níveis mais baixos de partículas virais circulante, por sua vez, parecem relacionados a diferenças na regulação da transcrição deste vírus, quando comparada ao HIV-1. Esta diferença entre os dois vírus parece ser um fator importante para as diferenças na patogenicidade e transmissibilidade observada entre eles, uma vez que os níveis de carga viral estão relacionados à eficiência de transmissão e ao desenvolvimento de AIDS (Bock & Markovitz, 2001; Reeves & Doms, 2002).

Em relação à variante GWGR, apesar de os pacientes infectados com estes vírus no Brasil progredirem mais lentamente para AIDS do que aqueles infectados com vírus GPGR, não foram observadas diferenças nos níveis de cargas virais entre os dois grupos (Santoro-Lopes et al., 2000). Quanto às infecções pelo subtipo C, a menor eficiência na replicação desses vírus em células T CD4-positivas e em células mononucleares *in vitro* poderia resultar em níveis inferiores de carga viral. Contudo, um estudo seccional

mostrou que mulheres infectadas por vírus C possuíam níveis de carga viral mais elevados do que aquelas com vírus dos subtipos A, D ou G. Entretanto, a natureza deste estudo não permitiu que se discriminasse entre as possibilidades das pacientes com vírus C estarem infectadas há mais tempo do que as demais ou terem um tempo de progressão menor para AIDS (Neilson et al., 1999). O único estudo de seguimento comparando a progressão da imunodeficiência entre pacientes infectados por diferentes subtipos de HIV-1 (inclusive o C) não incluiu os níveis de cargas virais entre as variáveis analisadas (Kanki et al., 1999).

Desta forma, se as infecções pelo subtipo C e por vírus B GWGR estão associadas à progressão mais lenta da imunodeficiência (o que seria seletivamente vantajoso nos modelos propostos), esta menor patogenicidade não parece estar relacionada aos níveis de cargas virais observados nas infecções por esses vírus. Quais fatores relacionados à biologia dos vírus são subjacentes à progressão mais lenta para AIDS, se isto realmente ocorre, devem ser ainda determinados.

Considerando-se essas observações, o fato das infecções pelo HIV-2 cursarem normalmente com níveis inferiores de carga viral conferiria a estes vírus uma menor aptidão para estabelecer epidemias do que o HIV-1 devido à sua menor transmissibilidade, apesar dessas infecções estarem associadas a uma progressão mais lenta da imunodeficiência. Como já foi dito, as infecções por este tipo de vírus permanecem restritas a poucas regiões geográficas. Por outro lado, no caso do subtipo C e da variante GWGR, a ausência de evidências de que estes vírus sejam menos transmissíveis do que as outras variantes do HIV-1 associada a uma possível menor patogenia capacitaria estes vírus a se perpetuarem na epidemia mundial, no caso do subtipo C, e na brasileira, no caso da variante GWGR.

5.2. Outros fatores

A pandemia pelo HIV-1 é extremamente complexa, e cada país experimenta epidemias regionais com características particulares. Para tal complexidade contribuem, por exemplo: 1) fatores do hospedeiro, como a habilidade do sistema imune em responder ao vírus, mediados especialmente pelos genes HLA (*human lymphocyte antigen*); 2) as condições sócio-econômicas de cada país, o que pode ser evidenciado pelo grande número de casos de infecção nas regiões economicamente menos favorecidas; 3) o fluxo de indivíduos entre os diversos países onde o HIV-1 circula, seja

devido às migrações, às suas relações comerciais ou ao turismo e 4) a história da introdução do HIV-1 em cada país, ou seja, o momento da sua introdução no contexto da epidemia mundial, o(s) grupo(s) de indivíduos infectados e as inter-relações entre estes grupos.

A dispersão do HIV-1 em cada região geográfica é profundamente influenciada pelos fatores citados. Assim sendo, mesmo que existam diferenças na aptidão relativa das duas variantes do subtipo B, é muito provável que as características particulares de cada epidemia também tenham contribuído para a heterogeneidade observada na dispersão mundial dos vírus GWGR. Por outro lado, se as duas variantes possuem a mesma aptidão para se perpetuar na população humana, então o grande número de infecções causadas pelos vírus GWGR encontrado no Brasil deve ser atribuído a condições específicas da epidemia brasileira.

Neste ponto da tese, vamos discutir a hipótese da dispersão da variante GWGR não ter ocorrido especificamente pelo fato destes vírus representarem variantes mais aptas a deixar descendentes na epidemia do que os seus parentais GPGR. Dois fatores serão abordados, por serem possivelmente os mais relevantes: em primeiro lugar, as condições das diversas epidemias regionais no momento da identificação da primeira sequência GWGR em cada país e, a seguir, a composição genética das populações de hospedeiros humanos.

5.2.1. Condições das epidemias

Para que um patógeno tenha sucesso no estabelecimento de uma epidemia, é necessário que, após a sua introdução em um hospedeiro susceptível, o número médio de infecções secundárias produzidas por ele em uma população onde todos os indivíduos são susceptíveis (ou número reprodutivo básico, R_0) seja superior a 1 (May et al., 2001). Brevemente, em doenças sexualmente transmissíveis (DST), como o HIV, R_0 depende do tempo de duração da infecciosidade, da taxa epidemiologicamente efetiva de troca de parceiros e da probabilidade de transmissão do patógeno. A probabilidade de transmissão, por sua vez, apesar de estar relacionada com o tempo de duração da infecciosidade, também possui componentes comportamentais, como o uso de preservativos, circuncisão e presença de lesões na mucosa genital associadas a outras DST, as quais estão relacionadas com o acesso a tratamento médico, por exemplo (May et al., 2001).

Assim, apesar do sucesso em estabelecer epidemias para um patógeno como o HIV-1 depender de uma complexa interação entre fatores biológicos e comportamentais, o momento histórico da introdução do vírus em uma população é importante, uma vez que influenciará a chance que esse patógeno terá de encontrar novos hospedeiros susceptíveis. Considerando-se que as variantes GPGR e GWGR possuem a mesma infecciosidade (ou seja, a mesma aptidão para deixar descendentes) e que o comportamento sexual médio das diversas populações onde a variante GWGR foi semeada não difere substancialmente, dos componentes de R_0 citados para as doenças infecciosas aquele que poderia estar associado à dispersão heterogênea desses vírus seria a sua probabilidade de transmissão.

A observação da distribuição geográfica dos subtipos de HIV-1 na epidemia mundial mostra que, na maioria dos casos, o subtipo mais prevalente em cada país é aquele que fundou a epidemia local. Um exemplo que pode ser citado se refere à história epidemiológica do próprio subtipo B o qual, apesar de ter sido sempre minoritário na África sub-sahariana, foi levado para o Ocidente provavelmente no final da década de 70 e teve a oportunidade de fundar as epidemias da América e Europa Ocidental. Assim, nas duas regiões, mesmo após mais de duas décadas, o subtipo B permanece o responsável pela grande maioria das infecções pelo HIV-1. Um outro exemplo da fundação de uma epidemia por um subtipo globalmente minoritário é a introdução do subtipo F na Romênia, no final da década de 80, onde até o momento responde pelo maior número de casos de infecção pelo HIV.

Além de diferentes subtipos terem fundado epidemias regionais, em alguns casos, indivíduos sob riscos específicos para a infecção portam subtipos distintos. Quanto a isto, o melhor exemplo é o que ocorre na Tailândia, onde introduções independentes fundaram as epidemias entre os usuários de drogas injetáveis e heterossexuais, infectados, em sua maioria, pelo subtipo B e o CRF_01 AE, respectivamente. Ainda hoje, apesar da segregação das variantes nestes grupos já não ser tão marcante como nos primeiros anos da epidemia, a maioria dos usuários de drogas está infectada pelo subtipo B enquanto o CRF_01 AE é majoritário entre os pacientes heterossexuais.

Apesar dos registros da presença da variante GWGR na pandemia pelo HIV-1 serem incompletos, com a exceção do Brasil, não há indícios de que estes vírus tenham sido semeados nos diversos países onde são encontrados antes da década de 90 (tabela 1). Por outro lado, no Brasil, a presença de anticorpos reativos a peptídeos com o

motivo GWGR em soros colhidos em 1983 (Michael-Hendry et al., 1996) mostra que estes vírus circulam no país desde o início da década de 80. Considerando-se que nem o momento preciso do aparecimento da variante no contexto da pandemia nem o de sua introdução em cada país são conhecidos, alguns cenários para explicar a distribuição geográfica da variante GWGR serão propostos e discutidos a seguir.

Cenário 1: Neste cenário, a variante GWGR teria surgido antes que o HIV-1 tivesse sido semeado no Ocidente, ou seja, na África Central, e, a partir desta região, iniciou as epidemias da Europa e América junto com os parentais GPGR do subtipo B. Neste caso, a introdução das duas variantes, GPGR e GWGR, nos diferentes países da Europa e América poderia ter sido simultânea. Contudo, pelo menos dois argumentos falam contra esta possibilidade: 1) a variante GWGR não foi encontrada na África e 2) se a variante houvesse participado da fundação das epidemias americana e da Europa Ocidental, seria esperado que a proporção de casos de infecção por vírus GWGR nestas regiões fosse semelhante à relatada para os GPGR, como ocorre no Brasil, o que não parece ser verdade. Apesar do retrato que se tem da distribuição mundial das variantes do HIV-1 ser extremamente incompleto, pelo menos para países como Estados Unidos e Holanda isto não se aplica devido ao grande número de relatos na literatura sobre os vírus que circulam nestas duas regiões, refletido na grande quantidade de seqüências locais disponíveis no Genbank. Por exemplo, na amostra sistemática de 263 seqüências GPGR analisadas na tese, 43,7% foram obtidas nos Estados Unidos e 11,8% na Holanda; juntos, os dois países contribuíram com mais de 50% das seqüências GPGR.

Cenário 2: A variante GWGR teria surgido fora da África, logo após a dispersão dos vírus B no Ocidente, ou seja, no final da década de 70 e início da de 80. Neste caso, o local de origem não teria sido o Brasil, mas sim outro país da Europa ou América. A partir deste país, então, a variante teria sido dispersada pelo mundo Ocidental ainda no início dos anos 80. Dois fatos suportam esta hipótese, sendo eles: o registro sorológico da variante no Brasil, em 1983, e o indício de sua presença na França, também na primeira metade da década de 80, sugerido na tese a partir do caso de infecção do paciente das Ilhas Reunidas, adquirida através de uma transfusão de sangue recebida na França. Contudo, se a variante houvesse surgido fora do Brasil durante os primeiros anos da epidemia, isto significaria que teria participado do estabelecimento da epidemia em seu local de origem, em um momento em que não havia medidas de controle para se conter a dispersão do HIV-1 e virtualmente todos os indivíduos eram susceptíveis ao vírus. Segundo os dados disponíveis, em nenhum outro local, exceto o Brasil, a variante

parece contribuir significativamente para a epidemia do HIV-1.

Cenário 3: A variante teria surgido no Brasil, logo após a introdução dos vírus GPGR no país, e foi eventualmente introduzida em outras regiões. Quando se considera unicamente o fato da variante ter se expandido principalmente no Brasil, esta parece ser a hipótese mais provável. Contudo, algumas questões relacionadas à sua distribuição mundial permanecem por ser esclarecidas. Se a proporção de casos de infecção por vírus GWGR fora do Brasil é realmente pequena, então a sua entrada nos demais países deve ter se dado em condições desfavoráveis à sua dispersão. A relevância desta observação se baseia no argumento de que, em diversas regiões onde a variante é encontrada, como Argentina, Lituânia e República Checa, a epidemia pelo HIV-1 é atribuída a mais de um subtipo do HIV-1, o que significa que, após a fundação destas epidemias por subtipos específicos, outras variantes tiveram a oportunidade de entrar e se expandir nesses países onde, entretanto, há poucos registros dos vírus GWGR.

Nesse caso, a possibilidade de irradiação mundial da variante GWGR a partir do Brasil ter ocorrido imediatamente após a sua origem não parece ser um fato provável. Uma alternativa para esta hipótese é que a dispersão mundial dos vírus GWGR tenha ocorrido a partir de eventos esporádicos de infecção, no decorrer das duas décadas seguintes. Neste caso, a variante poderia ter sido importada diretamente do Brasil ou então ter entrado inicialmente em alguns países e, a seguir, ter sido levada para outras regiões. Novamente, para que este argumento fosse válido, também seria necessário que os indivíduos responsáveis pela dispersão do vírus para cada país, pelo menos na maioria dos casos, não tivessem o comportamento de risco ao qual a expansão do HIV-1 fosse atribuída no momento.

Se a introdução da variante GWGR na Europa Ocidental, onde a epidemia teve início no começo dos anos 80, ocorreu realmente no final da década de 90 (há registros da variante em países da região a partir dos primeiros anos da década), deve ter se dado em poucas ocasiões. De outro modo, seria de se esperar um maior número de registros de casos de infecção por vírus GWGR. Em virtualmente todos os países da Europa Ocidental são descritos casos de infecção por múltiplos subtipos não-B, inicialmente registrados em imigrantes que portam vírus freqüentes em seus países de origem ou em nativos que se contaminaram em outros lugares e retornaram ao país. Posteriormente, também na maioria dos casos, estes outros subtipos foram identificados na população autóctone. Assim, se somente um ou poucos casos de transmissão da variante GWGR ocorreram nestes países, e já quando havia medidas de controle da epidemia, então a sua

baixa freqüência na Europa Ocidental poderia ser atribuída à pouca oportunidade que a variante teve para se expandir.

Na Europa Oriental e Ásia, a explosão da epidemia foi mais tardia, no início dos anos 90, e o uso de drogas injetáveis desempenhou (e ainda desempenha) um papel fundamental na dispersão do HIV-1. O fato de seqüências GWGR terem sido encontradas em países da região no início dessa década poderia sugerir que a variante teve a oportunidade de se expandir na região junto com os vírus fundadores. Contudo, a infecção dos pacientes das duas regiões que portam vírus GWGR parece ter ocorrido através do contato sexual, e não pelo uso de drogas. Naturalmente, mesmo que o uso de drogas seja o principal risco para a aquisição do vírus na Europa Oriental e Ásia, um grande número de infecções por via sexual também ocorre, o que significa que a variante, teoricamente, poderia ter se estabelecido no local. Novamente, a explicação mais plausível para o pequeno número de seqüências GWGR encontradas na Europa Oriental e Ásia parece ser a pouca oportunidade que a variante deve ter tido de se disseminar.

De um modo geral, os pequenos números de infecções pela variante GWGR em toda a Europa e Ásia encontrados na literatura podem ser uma decorrência do fato da variante ter sempre circulado nos dois continentes em pequenos números. Se a variante não foi introduzida em qualquer país da região no momento específico em que a epidemia estava expandindo nas diversas populações sob risco, então é possível se considerar que tenha permanecido restrita a um pequeno número de indivíduos.

Contudo, no cenário 3 proposto, o número restrito de registros da variante GWGR que mais merece atenção é o observado na América Latina. Como já foi oportunamente mencionado nesse trabalho, existe um grande intercâmbio entre as epidemias dos diversos países latino-americanos, e o Brasil parece ter desempenhado um papel importante na disseminação do HIV-1 no sub-contidente. A epidemia da Argentina, por exemplo, está relacionada à brasileira, fato este tornado evidente com o compartilhamento de vírus do subtipo F, e de recombinantes BF, entre os dois países. Contudo, somente três seqüências GWGR foram identificadas na Argentina. Em relação aos demais países, o pequeno número de seqüências de HIV-1 disponíveis de cada um deles pode explicar a baixa freqüência de seqüências GWGR (por ocasião da coleta de dados para a tese, havia somente um trabalho da Bolívia e um do Paraguai referindo a diversidade local do HIV-1). Contudo essa explicação não parece se aplicar à Argentina, uma vez que existem diversos estudos sobre a diversidade do HIV-1 no país.

Assim, no caso da América Latina, a pouca oportunidade de dispersão atribuída a um pequeno número de introduções em momentos desfavoráveis para o vírus, apesar de passível de ter ocorrido, não é uma explicação que possa ser imediatamente aceita. Contudo, os poucos dados disponíveis não permitem que se analise a questão de forma mais detalhada.

5.2.2. Fatores genéticos do hospedeiro

Paralelamente à possibilidade da grande dispersão da variante GWGR na epidemia brasileira ter sido propiciada pelo momento em que foi introduzida no país (como proposto no cenário 3 do item 5.2.1), também deve ser considerada a hipótese da composição genética da população brasileira ter influenciado esta ocorrência.

A diversidade genética do hospedeiro desempenha um papel fundamental no destino de indivíduos expostos a uma grande quantidade de microorganismos patogênicos. Em alguns casos, os mecanismos específicos através dos quais genótipos do hospedeiro conferem resistência a infecções são conhecidos (como o genótipo CCR5- Δ 32/CCR5- Δ 32 nas infecções pelo HIV-1, Anexo II). Entretanto, de um modo geral, a resistência a infecções parece envolver um conjunto complexo de efeitos genéticos dos hospedeiros, e seu mecanismo se torna ainda mais complexo quando se considera a diversidade dos patógenos e fatores do meio que influenciam a interação hospedeiro-parasita (Carrington et al., 2001).

Em relação ao HIV, nos últimos anos, tem-se acumulado um grande número de informações na literatura sugerindo a participação de diversos componentes da resposta imune natural e adquirida na progressão da imunodeficiência e, em um grau muito menor, na resistência à infecção. Dentre estes fatores estão componentes do complexo principal de histocompatibilidade (*major histocompatibility complex*, MHC) - como os genes que codificam as proteínas das classes I e II do HLA (*human lymphocyte antigen*) e o fator de necrose tumoral (TNF) -, os genes da família dos receptores para β -quimiocinas (CCR), o gene que codifica o ligante do receptor CCR5, RANTES, a interleucina 10 (IL10) e a lecitina que se liga à manose (MLB) (Carrington et al., 2001; Carrington & O'Brien, 2003; Roger, 1998).

Contudo, em diversos casos, os resultados encontrados por diferentes autores são contraditórios e diversos fatores parecem contribuir para esta inconsistência, como o reduzido número de indivíduos incluídos em muitos estudos, a extrema heterogeneidade

do curso clínico da progressão para a AIDS e a dificuldade em se quantificar precisamente a exposição ao vírus (Carrington & O'Brien, 2003). Nesse contexto, dentre os fatores citados na literatura, aqueles cuja associação com os dois processos, infecção e patogenia, é mais consistente são a presença de determinados alelos da classe I do HLA e os genes da família dos receptores para β -quimiocinas (CCR) (Carrington et al., 2001; Carrington & O'Brien, 2003; Roger, 1998). A relevância da diversidade destes dois componentes da resposta imune na susceptibilidade à infecção pelo HIV-1 e na progressão da imunodeficiência será discutida a seguir.

5.2.2.1. O co-receptor CCR5

Diversos receptores para quimiocinas servem como co-fatores (co-receptores) para a entrada do HIV-1 nas células humanas (Roger, 1998). Um destes co-receptores é a molécula CCR5, utilizada preferencialmente pelo vírus nas fases iniciais da infecção. Indivíduos homozigotos para o alelo mutante CCR5- Δ 32 do receptor são virtualmente resistentes à infecção pelo HIV-1. Por outro lado, indivíduos que possuem o genótipo heterozigoto CCR5+/CCR5- Δ 32 progridem mais lentamente para a AIDS em relação aos que são homozigotos para o alelo selvagem (Anexo II) (Carrington et al., 2001). Nas fases mais tardias da imunodeficiência, como dito anteriormente no texto, os isolados virais usam, na maioria das vezes, outro co-receptor para infectar novas células, o CXCR4.

Agora, no caso de um paciente heterozigoto para o gene CCR5 (genótipo CCR5+/CCR5- Δ 32) estar infectado com uma cepa incapaz de utilizar o co-receptor CXCR4, como ocorre com os vírus do subtipo C, qual seria o efeito resultante da interação dos dois fatores, ou seja, a expressão reduzida de co-receptores CCR5 e a utilização exclusiva deste co-receptor pelos vírus? (Naturalmente, não se pode excluir a possibilidade de a variante escolher um receptor alternativo para penetrar nas células).

À primeira vista, a resposta mais simples seria que, nesse caso, a progressão para a AIDS se apresentaria ainda mais lenta nesses pacientes, uma vez que o efeito da expressão reduzida de CCR5 na superfície das células estaria presente durante todo o curso da imunodeficiência. Para que tal efeito fosse notado, entretanto, seria necessário que a variante CCR5-exclusiva estivesse dispersa em populações onde a frequência de hospedeiros heterozigotos para o alelo CCR5- Δ 32 fosse significativamente diferente, como por exemplo no Norte da Europa, onde as maiores frequências são descritas, e na África ou Ásia, onde o alelo mutante é raramente encontrado (Gonzalez et al., 2001). Para se testar esta hipótese, uma alternativa seria comparar a progressão para a AIDS

em indivíduos infectados pelo subtipo C nestes dois grupos de populações, uma vez que este subtipo utiliza somente o co-receptor CCR5. Contudo, este tipo de estudo deveria, necessariamente, incluir um grande número de pacientes, uma vez que diversos outros fatores, geneticamente determinados ou não (como o uso de anti-retrovirais, por exemplo), além de, naturalmente, a frequência do alelo mutante, têm uma participação importante na determinação do curso clínico da imunodeficiência.

Em relação à variante GWGR, para que a frequência do alelo CCR5- Δ 32 houvesse influenciado a sua dispersão mundial, seria necessário que, em primeiro lugar, a população brasileira apresentasse uma frequência de CCR5- Δ 32 significativamente diferente da encontrada no restante das populações. Na população mundial, a frequência estimada do alelo CCR5- Δ 32 foi 0,08 em populações de origem européia (Gonzalez et al., 2001), variando de acordo com a região, sendo maior nos países do norte da Europa (Libert et al, 1998); 0,03 em afro-americanos; menor ou igual a 0,01 na Ásia; e nula na África (Gonzalez et al., 2001). No Brasil, dois estudos encontraram frequências de 0,035 (Passos & Picanco, 1998) e 0,013 e 0,044 (Chies & Hutz, 2003) em indivíduos de origem africana e européia, respectivamente. As frequências do alelo nos dois grupos brasileiros estudados não parecem diferir daquelas encontradas no restante do mundo.

Em segundo lugar, seria necessário que a variante possuísse um fenótipo distinto em relação ao uso de co-receptores, como ocorre com o subtipo C. Até o momento, entretanto, nenhum trabalho investigou esta possibilidade. Se estas duas premissas fossem preenchidas, então a progressão mais lenta para a AIDS observada nos pacientes brasileiros infectados pelo sorotipo GWGR poderia ser atribuída à interação das duas condições, ou seja, o uso diferencial de co-receptores e constituição genética da população brasileira. Neste caso, porém, a seleção natural não poderia ser excluída do processo. Ao invés disto, o efeito seletivo positivo de uma menor patogenia, como discutido na seção anterior, seria o resultado da interação vírus-hospedeiro, e não somente do fenótipo do vírus, como foi proposto no item 5.1.

5.2.2.2 Sistema MHC

O segundo componente da resposta imune que desempenha um importante papel na diversidade da resposta do hospedeiro à infecção pelo HIV é o sistema HLA classe I, que pertence ao sistema MHC. Os genes HLA (além das moléculas de classe I, o MHC também codifica para moléculas HLA de classes II e III) representam o complexo genético mais polimórfico no ser humano e codificam proteínas fundamentais para as respostas imunes adquiridas. Diversas associações do HLA com doenças infecciosas

têm sido sugeridas na literatura. Entretanto, estas associações são difíceis de serem identificadas, possivelmente porque nestas doenças existe um amplo complexo de epítomos antigênicos, derivados das estruturas dos parasitas, envolvidos na sua patogênese. Estudos conclusivos sobre a associação de alelos do HLA com doenças infecciosas específicas requerem grandes amostras, estratificação adequada da composição étnica dos indivíduos, informações clínicas acuradas e o uso de modelos que consideram outros efeitos genéticos conhecidos sobre as doenças (Carrington et al., 2001; Carrington & O'Brien, 2003).

A hipótese de que os indivíduos heterozigotos nos loci do HLA, por sua capacidade de apresentarem uma maior variedade de peptídeos ao sistema imune, poderiam desenvolver uma resposta mais eficiente contra o HIV-1 do que os homozigotos foi testada e confirmada por alguns autores. Nesses estudos foi encontrado que a heterozigose nos três loci do HLA classe I (A, B e C) confere uma resistência relativa à progressão para AIDS (Carrington et al., 2001; Carrington & O'Brien, 2003).

Além do efeito da heterozigose, uma grande quantidade de trabalhos descreveu associações genéticas entre os alótipos das classes I e II do HLA e diversos aspectos da infecção pelo HIV, mas muitos se basearam em pequenas amostras e os seus resultados não foram posteriormente confirmados por outros estudos. Um dos fatores que podem contribuir para as incongruências observadas entre esses estudos é o fato de a distribuição dos alelos diferir de uma população para outra e, desta forma, um alelo pode não ser associado em uma população simplesmente porque é muito raro para ser detectado (Roger, 1998).

Entretanto, uma grande quantidade de informações, obtidas em estudos com coortes independentes, apontam para o envolvimento de determinados alelos de classe I na proteção ou susceptibilidade à progressão para a AIDS (Carrington et al., 2001; Carrington & O'Brien, 2003). Destes, as associações mais consistentes envolvem o alelo HLA-B*35, cuja presença resulta em uma maior susceptibilidade à progressão para AIDS, e os alelos HLA-B*27 e B*57, que conferem um efeito protetor à progressão da imunodeficiência (Carrington et al., 2001; Carrington & O'Brien, 2003).

Contudo, os estudos desenvolvidos até o momento não referem efeito protetor ou susceptibilidade em relação a variantes específicas do HIV-1. Em relação ao fato de a dispersão da variante GWGR ter sido observada principalmente no Brasil, para que os alelos do sistema HLA, especialmente de classe I, pelas razões já expostas, tivessem contribuído para isto seria necessário, como no caso do alelo mutante CCR5- Δ 32, que a sua frequência na população brasileira seguisse um padrão distinto dos observados nas

demais regiões onde a variante foi encontrada. A heterozigose, que pode ser elevada na nossa população devido à sua grande miscigenação, não deve estar envolvida, uma vez que a maior capacidade de apresentar antígenos observada nos pacientes heterozigotos deveria, naturalmente, favorecer a resposta imune às duas variantes, e não somente a uma delas.

Até o momento, poucos trabalhos foram publicados no Brasil sobre a frequência dos alelos do HLA na população brasileira. Destes, a maioria se refere aos alelos de classe II (Louzada-Junior et al., 2001) ou estão restritos a pequenas populações locais (Braun-Prado et al., 2000) ou às populações indígenas (Erlich et al., 1997). Nenhum trabalho, até o momento, abordou a frequência no Brasil dos alelos associados à progressão mais lenta da imunodeficiência.

Assim, uma abordagem possível para se futuramente estudar a participação da composição genética brasileira nos loci HLA classe I em relação à progressão diferencial para a AIDS nos pacientes infectados com as variantes GWGR e GPGR seria a análise das frequências dos três alelos, HLA-B*35, B*27 e B*57, dentro de grupos infectados com os dois variantes. Naturalmente, isto só explicaria a expansão de vírus GWGR na população brasileira, no caso desta expansão estar relacionada aos genes HLA, se a frequência dos alelos protetores (B*27 e B*57) no grupo GWGR brasileiro fosse superior ao encontrado na população mundial e no grupo de pacientes GPGR do Brasil.

6. Conclusões

No trabalho de tese, foi sugerido que a variante GWGR que circula na epidemia pelo HIV-1 tem uma única origem e que a sua dispersão mundial provavelmente ocorreu a partir do Brasil. Contudo, apesar destes vírus terem sido encontrados em 18 países, a sua expansão foi observada principalmente nesse país. Uma vez que já havia sido demonstrado que os pacientes infectados com o sorotipo GWGR têm um tempo de sobrevivência mais longo sem AIDS do que aqueles contaminados pelo sorotipo GPGR, decidiu-se investigar se a expansão da variante na epidemia brasileira poderia ser atribuída à seleção natural. Os resultados encontrados mostraram que as seqüências brasileiras com o motivo GWGR possuem uma proporção menor de mutações associadas à progressão da imunodeficiência do que as seqüências com o padrão de assinatura GPGR. Desta forma, foi sugerido que a progressão mais lenta da infecção observada entre os pacientes infectados pelo subtipo GWGR poderia conferir a estes vírus uma maior oportunidade de serem transmitidos para novos hospedeiros, ou seja, de deixar descendentes. Entretanto, mesmo que a variante GWGR possua um fenótipo seletivamente vantajoso sobre os vírus parentais GPGR, o curso da dispersão mundial desta variante provavelmente foi influenciado pelas condições locais de cada epidemia.

Anexo I Alinhamento das seqüências do gene env incluídas na análise filogenética.

ZA59 -----CAGCTTAACGAATCTGTA---GAAATTAATTGTATA
PU56380ETH -----GTACAGCTGAAAGACTCTGTA---GAAATTAATTGTACA
PX78513BRA (BA) TCTGAGAAATTTACGGACAATGCTAAAACCAATAATAGTACAGCTGAGAGAACTCTGTA---GTAATTAATTGTCTCA
PU08446HTI TCTGAAAATTTTCACAAACAATGCTAAAACCAATAATAATACAGCTGAAGGAATCTGTA---GTAATTAATTGTACA
PAY037268ARG TCTGCTAACCTCTCAGATAATACTAAAACCAATAATAGTACAGCTGAATGAATCTATA---GCAATTAATTGTACA
PAJ277821BEL TCTGAAAATTTTCAGTGACAATGCTAAAACCAATAATAGTACAGTTGAATTAATCTGTA---GAAATTAATTGTACA
PU08796BRA TCTGAAAATTTTCATGAACAATGCTAAAACCAATAATAGTACAGCTGAATGAATCTGTA---GAAATTAATTGTACA
PX78515BRA (BA) TCTGACAATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAATGTATCTGTA---GAAATTAATTGTACA
PU08797BRA TCTGAAAATTTTCACAAACAATGCTAAAACCAATAATAGTGCAGCTAAAGGACCCCTGTA---GACATTAATTGTACA
PU08798BRA TCTGAAGATCTCACAACAATGCTAAAACCAATAATAGTGCAGCTAAAGGACCCCTGTA---GACATTAATTGTACA
PX78517BRA (RJ) TCTGAAAATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAATAAACCTGTA---GAAATTAATTGTACA
PU08775BRA -----AATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTAAATGAAACTGTA---CAAATCAATTGTACA
PU65538MWI -----AATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAATGAAGCTGTA---GAAATTAATTGTACA
PU08801THA TCTAGCAATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAATGAATCTGTA---GAAATTAATTGTACA
PAF034021BRA (RJ) -----AATTTTCACAGACAATGCTAAAACCAATAATAGTACAACCTGAATGAGTCTGTA---ACAATTAATTGTACA
PU08772BRA -----AATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAACGAATCTGTA---GTAATTAATCCACA
PL07226HTI TCTGAGAAATTTTCACGACAATGCTAAAACCAATAATAATACAGCTGAATGCATCTGTA---GAAATTAATTGTACA
PAF113570BRA (RJ) -----AATGCTAAAACCAATAATAGTACAGCTGAACGTTCTCTGTA---GAAATTAATTGTACA
PU23138USA TCTGACAACCTCAGGAACAATGCTAAAACCAATAATAGTACAGCTGAACGCATCTGTA---GAAATTAATTGTACA
SUBTYPE_A TCTGAAAATATCTCAGACAATGCCAAAACCAATAATAGTACAACCTACCGAGCCCTGTA---ACAATTAATTGTACC
PAF200874MEX TCTAAAATTTCTCGGACAATACTAAAACCAATAATAGTGCAGCTGAAAGACCCCTGTA---AAAATTAATTGTACA
PAF169616KOR TCTGAAAATTTTCACAAACAATGCTAAAACCAATAATAGTACAGCTGAACGCATCTGTA---GAAATTAATTGTACA
PAF034010BRA (RJ) -----AATTTTCACAAACAATGCTAAAACCAATAATAGTACAGTTGAACGAATCTATA---GAAATCAATTGTACA
PAF034012BRA (RJ) -----AATTTTCACAATAATGCTAAAACCAATAATAGTACAGCTGAATAAATCTATA---ACAATTAATTGTACA
WAF113565BRA (RJ) -----AATTTTACTAACAATGCTAAAACCAATAATAGTACAGCTGAACGAACCTGTA---GAAATTAATTGTACA
WAF034030BRA (RJ) -----AATTTTACAACAATGCTAAAACCAATAATAGTACAACCTGAACAAAACCTGTA---CAAATTAATTGTACA
WY18596REU TCTGAAAATTTTCACAAACAATGCTAAAACCAATAATAGTACAGCTGAACAAAACCTGTA---CAAATTAATTGTACA
WAF080212CHN -----ATTTTCACTAACAATGCTAAAACCAATAATAGTACATCTGAACGAATCTGTA---GAAATTAATTGTACA
WAF113574BRA (RJ) -----AATGTTAAGACCAATAATAGTACAGCTGAACCAATCGATA---GAAATTAATTGTACA
WAF231041BRA -----TGATA
WU31591BRA (SP) -ACTACGTGTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAATAAACCTGTG---GCAATTAATTGTACA
WAF060965BRA (RJ) -----CCATAATAGTACAGCTGAACGAATCTGTA---CCAATCCATTGTACA
WAF188990DEU -----ACCATAATAGTACAGCTGAACGCATCTATA---GAAATTAATTGTACA
WAF034029BRA (RJ) -----AATATCNCGGACAATGCTAAAACCAATAANTACNGCTGAACCAANTCTGTA---GTAATTAATTGTACA
WAF034028BRA (RJ) -----AATATCAGGACAATGCTAAAACCAATAATAGTACAGCTGAACCAATCTGTA---GTAATTAATTGTACA
WAF034016BRA (RJ) -----AATCTCAGACAATACTAAAACCAATAATAGTACACCTGAACGAATCTGTA---GTAATTAATTGTACA
WAF094014USA TCTAAAATTTTCACAAACAATGCTAAAACCAATAATAGTACAGCTGAATGAATCTGTA---GTAATTAATTGTACA
WAF034026BRA (RJ) -----AATTTTCACAAACAATGCTAAAACCAATAATAGTACAGCTGAACACATCTGTA---GTAATTAATTGTACA
WAF060952BRA (RJ) -----TAATAGTACAGCTGAATGCATCTGTA---GAAATTAATTGTACA
WU08780BRA -----AATTTCCACGACAATGCTAAAACCAATAATAGTACAGCTGAATGATACTATA---GAGATTAATTGTACA
WAF034024BRA (RJ) -----AATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAACAAACCTATA---AGAATTAATTGTACA
WAF033997BRA (RJ) -----AATTTTCACGACAATGCTAAAACCAATAATAGTACANCTGAACNAATCTATA---GCANTTANTTGTACA
WU08975BRA (SP) TCCGAAAATTTCTCGGACAATGCTAAAACCAATAATAGTACAGCTGAACGAACCTGTA---AAAATTAATTGTATA
TCTGAAAATTTTCACGACAATGCTAAAACCAATAATAATACAGCTGAATAAATCTGTA---AACATTACTTGTATG
WAF034005BRA (RJ) -----AATCTCAGGACAATGCTAAAACCAATAATAGTACAGCTGAAGAAATCTGTA---GCAATTAATTGTACA
WAF033999BRA (RJ) -----AATCTCAGACAATACTAAAACCAATAATAGTACAGCTGAACCAATCTGTA---GTAATTAATTGTACA
WAF034001BRA (RJ) -----AATCTCACAACAATGCTAAAACCAATAATAGTACAGCTGAATGAAACTGTA---CAAATTAATTGTACA
PAF113571BRA (RJ) -----AATCTCACAACAATGCTAAAACCAATAATAGTACAGCTGAACGAAACTGTA---CTACTTAATTGTACA
WU83693BRA (RJ) -----AATCTCAGACAATACTAAAACCAATAATAGTACAGCTGAAGAAGCTGTA---GAAATTAATTGTACA
WU83689BRA (RJ) -----AATCTCAGACAATACTAAAACCAATAATAGTACAGCTGAAGAAGCTGTA---GAAATTAATTGTACA
WU28336BRA (RJ) TCTGAGAAATTTTCACGACAATACTAAAACCAATAATAGTACAGCTGAAGAAGCTGTA---GAAATTAATTGTACA
WU83691BRA (RJ) -----AATCTCAGACAATACTAAAACCAATAATAGTACAGCTGAAGAAGCTGTA---GAAATTAATTGTACA
WU08771BRA -----AATATCAGGACAATAAGAAAACCAATAATAGTACAGCTGAATGAAACTGTA---GAAATTAATTGTACA
WL19331BRA -----AATATCAGACAGCTGCTAAAACCAATAATAGTACAGCTGAATGAAACTGTA---GAAATTAATTGTACA
PAF034033BRA (RJ) -----AATTTTCACAACAATGCTAAAACCAATAATAGTGCAGCTAAAACGAATCTGTA---CAAATTAATTGTACA
PAF034018BRA (RJ) -----AATTTTACAACAATGCTAAAACCAATAATAGTACAGCTGAACGAATCTGTA---GTAATTAATTGTACA
PAF113572BRA (RJ) -----AATTTCTCGGACAATGCTAAAACCAATAATAGTACAGCTGAAGAAGCTGTA---GAAATTAATTGTACA
PAF034027BRA (RJ) -----AATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAAGAAGCTGTA---GAAATTAATTGTACA
PU95485ITA TCTGAAAATTTTCACGACAATGCTAAAACCAATAATAGTACATCTGAATGAATCTGTA---GAAATTAATTGTGCA
PY18759BRA (BA) -----AATTTCACTGACAATGCTAAAACCAATAATAGTACAGCTGAATGAAACTGTA---GAAATTAATTGTACA
PY18753BRA (BA) -----AATTTCACTGACAATGCTAAAACCAATAATAGTACAGCTGAATGAAACTGTA---GAAATTAATTGTACA
PU56320SWE TCTGAAAATTTTCAGACAATGCTAAAACCAATAATAGTACAGCTGAATGAAACTGTA---AAAATTAATTGTATA
PU68201IDN TCTGAAAATTTCTCGGACAATGCTAAAACCAATAATAGTACAGCTGACAGTACCTGTA---CAAATTAATTGTACA
PAF034004BRA (RJ) -----AATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAAGAAGCTGTA---GAAATTAATTGTACA
PAF137974USA TCTGAAAATTTCTCAGACAATGCTAAAACCAATAATAGTACAGCTGAAGAAGCCCTGTA---AACATTACTTGTACA
PY18754BRA (BA) -----AATTTCACTGACAATACTAAAACCAATAATAGTACAGCTGAAGAAGCCCTGTA---AAAATTAATTGTACA
PAF033996BRA (RJ) -----AATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAACGAATCTGTA---ACAATTAATTGTACA
PY18760BRA (BA) -----AATTTTCACGACAATACTAAAACCAATAATAGTACAGCTGAATGTACCTGTA---GAAATTAATTGTACA
PY18752BRA (BA) -----AATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAAGAAGCAATGTA---GAAATTAATTGTACA
PAF368558FRA TCTGAAAATATCAGACAATGCTAAAACCAATAATAGTACAGCTGAAGAAGCTGTA---GAGATTAATTGTACA
PAF034025BRA (RJ) -----AATTTTCACGACAATGCTAAAACCAATAATAATACAGCTGAATGATCAGTA---GAAATTAATTGTACA
PY18755BRA (BA) -----AATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAATGATCCTGTA---GAAATTAATTGTACA
PU13245HOL TCTGCCAATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTAAAATGAATCTGTA---GTAATTAATTGTACA
PAF022277HOL TCTGAAAATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTAAAATGAATCTGTA---GAAATTAATTGTACA
PAF204580USA TCTGAAAATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTAAAACCAACCTGTA---AAAATTAATTGTATA
PAB059299JPN TCTGAAAATTTTCACGACAATGCTAAAACCAATAATAGTACAGTTGAAAGAAGCTGTA---AAAATTAATTGTACA
PU08781BRA -----AATTTTCACGACAATACTAAAACCAATAATAGTACAGCTAAAACGAACCTGTA---GAAATTAATTGTACA
PU16384AUS TCTGACAATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAAGAAGCTGTA---AATATTACTTCTGTA
PAF279027SWE -----GAAATTTCTCGGACAATGCTAAAACCAATAATAGTACAATTGAGAGAAATTTGTA---GAAATTTATTGTATA
PL19335BRA -----AATTTTCACGACAATACTAAAACCAATAATAATACAGCTGAAGAAGCAATGTA---GAAATTAATTGTACA
PL19332BRA -----AATTTTCACAGTCAATACTAAAACCAATAATAGTACAGCTGAATGAATCTGTA---CAAATTAATTGTACA
WU37041ARG TCTGAGAAATTTTCACGACAATGCTAAAACCAATAATAGTACAGCTGAACAAATCTGTA---GCAATTAATTGTACA
WU37038ARG TCTGAAAATTTTCACAAACAATGCTAAAACCAATAATAGTACAGCTGAACGAATCTGTA---GTAATTAATTGTACA
WAF060962BRA (RJ) -----AGTACAGCTGAACGAATCTGTG---GTAATTAATTGTACA
WAF060959BRA (RJ) -----TAGTACAGCTGAACGAATCTGTA---GTAATTAATTGTACA

WAF060958BRA (RJ) -----TAAACCATAATTGTACAGCTGAACGAATCTGTA---GTAATTAATTGTACA
WU16219BRA TCAGAAAATTCACAGACAATGCTAAAACCATATAATAGTACAGCTGAACACATCTATA---GAAATTAATTGTACA
WU00400BRA (RJ) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAACCAATCTGTA---GTAATTAATTGTACA
WU00418BRA (RJ) TCTGAAAATTCACAAAACATGCTAAAACCATATAATAGTACAGCTAAAACGAATCTGTA---GAAATTAATTGTACA
PAF060954BRA (RJ) -----AATTAATTGTACA
PU00417BRA (RJ) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTAAAACCAATCTGTA---GTAATTAATTGTACA
PU00427BRA (SP) TCTGAAAATTCACAGACAATGCTAAAACCATATAATAGTACAGCTGAACACACCTGTA---GAAATTTATTGTACA
WL19334BRA -----AATTTACAGACAATGCTAAAACCATATAATAGTACAGCTGAACCAATCTGTA---GTAATTAATTGTACA
WU00419BRA (RJ) -----ATACGTGAA---GGAAAGTAATTGTACA
WU00412BRA (RJ) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAACGAATCTGTA---GAAATTAATTGTACA
WL16605ARG -----TGTACA
WU00401BRA (RJ) TCTGAGAATTCACGACAATGCTAAGATCATAATAGTACACCTGAACGAACTGTA---GAAATTAATTGTACA
PU00414BRA (RJ) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAACGAACTGTA---GAAATTAATTGTACA
WU00413BRA (RJ) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAACGAATCTGTA---GAAATTAATTGTACA
WAF060957BRA (RJ) -----AGAAATTAATTGTACA
WAF032151HOL TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAATCTGTA---GAAATTAATTGTACA
WU00409BRA (RJ) TCTGAAAATTCACGGATAATGCTAAAACCATATAATAGTACAGCTGAATGAATCTGTA---GAAATTAATTGTACA
WAF152771ESP -----ACATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
WX78512BRA (SP) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAACGAACTGTA---GTAATTAATTGTACA
WAF188941FRA -----ACATAATAGTACAGCTGAACGAACTGTA---GAAATTAATTGTACA
PAF060955BRA (RJ) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTAAAACCAATCTGTA---GTAATTAATTGTACA
WAF060964BRA (RJ) -----GCTAAAACCATATAATAGTACAGCTGAATGGATCAGAT---GCAATTAATTGTACA
WAF034023BRA (RJ) -----AATTTTACAAAACATGCTAAAACCATATAATAGTACAGCTGAACGAATCTGTA---GTAATTAATTGTACA
WL19329BRA -----AATATCACAACAATGCTAAAACCATATAATAGTACAGCTGAACGAATCTGTA---GAAATTAAGTGTACA
PU80831BRA (SP) -----TCTCGGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTATA
PU80830BRA (SP) -----TCTCGGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTATA
WU28955PRY -----AATTTCTCGGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTG---GAAATTAATTGTACA
WAF188956FRA -----ACATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
WU08776BRA -----AATTTACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
WU80824BRA (SP) -----TCACGAACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
WU80825BRA (SP) -----TCACGAACAATGCTAAAACCATATAATAGTACAGCTGAAAGAACTGTA---GAAATTAATTGTACA
WAF033998BRA (RJ) -----AATCTCACGACAATGCTAAAACCATATAATAGTACAGCTAAAACGAATCTGTA---GAANTTAATTGTACA
WU80833BRA (SP) -----TCACGAACAATGCTAAAACCATATAATAGTACAGCTGAACGAACTGTA---GAAATTAATTGTACA
WAF034013BRA (RJ) -----AATTTACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
PU80829BRA (SP) -----TCACGAACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
PU80832BRA (SP) -----TCACGAACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
PAF060966BRA (RJ) -----TGCTAAAACCATATAATAGTACAGCTGCAATGAAGCTGTA---GAAATTAATTGTACA
PAF034002BRA (RJ) -----AATTTCTCGAACAATGCTAAAACCATATAATAGTACAGCTGAATGCATCTGTA---GAAATTAATTGTACA
PU95412ITA TCCGCCAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGCATCTGTA---GAAATTAATTGTATA
PAF060967BRA (RJ) -----GCATAATAGTACAGCTGAATGAATCTGTA---GTAATTAATTGTACA
PU08709BRA TCCGCCAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAATCTGTA---GAAATTAATTGTACA
PAF060963BRA (RJ) -----ACATAATAGTACAGCTAAAACGAATCTGTA---GTAATTAATTGTACA
PU00416BRA (RJ) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAATCTGTA---GCAATTAATTGTACA
PAF368745FRA TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAAAGATGCTGTA---GAAATTAATTGTACA
PU00415BRA (RJ) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
PU80236BRB TCTAAAATTCACAGACAATGCTAAAACCATATAATAGTACAGCTGAAACAAGCTGTA---AAAATTAATTGTACA
PAF058149CAN -----ATAATAGTACAGCTAAGCGAGTCTGTA---GAATTTACTTGTACA
TCTGAAAATTCCTCGGATAATGCTAAAACCATATAATAGTACAGCTGACAGAACTGTA---GAAATTAATTGTACA
PAJ278211DNK TCTGCCAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAATCTGTA---GTAATTAATTGTACA
PAJ289592DNK TCTGACAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAATCTGTA---GCAATTAACTGTACA
PU20019CHN TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAATCTGTA---GCAATTAATTGTACA
PAF286549EST TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAATCTGTA---GTAATTAATTGTACA
PU31587BRA (SP) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAATCTGTA---GTAATTAATTGTACA
PU00411BRA (RJ) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAATCTGTA---GTAATTAATTGTACA
PU08687BRA TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTAAAAGAACTGTA---GAAATTAATTGTACA
PAF051476RUS TCTGCAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAAAGAACTGTA---GAAATTAATTGTACA
PAF170653RUS TCTGCAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAAAGAACTGTA---GAAATTAATTGTACA
PU08714BRA TCTGCAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
PAF076319BRA (AM) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
PU00405BRA (RJ) TCCGCAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAATCTGTA---GTAATTAATTGTACA
WZ68025HOL TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAAAGAACTGTA---GTAATTAATTGTACA
PAF181323CHE TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTAAAATGAACTGTA---GAAATTAATTGTACA
WAF223996CZE TCTGAAAATTCACAAAATGCTAAAACCATATAATAGTACAGCTGAAATGAACTGTA---GAGATTAATTGTATA
WAF223998CZE TCTGAAAATTCACAAAACATGCTAAAACCATATAATAGTACAGCTGAAATGAACTGTA---GAAATTAATTGTACA
WAB012971JPN TCTGAAAATTCACAAAACATGCTAAAACCATATAATAGTACAGCTGAACGAACTGTA---GAAATTAATTGTACA
WAF181437SWE TCTGAAAATTCACAAAACATGCTAAAACCATATAATAGTACAGCTGAAATGAACTGTA---GAAATTAATTGTACA
PU00402BRA (RJ) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAACGAATCTGTA---GTAATTAATTGTACA
WL20963BRA (BA) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
WU00424BRA (SP) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAACGAATCTGTA---GTAATTAATTGTACA
WU58813FRA TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAACGAACTGTA---GAAATTAATTGTACA
WU00425BRA (SP) TCTAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
WU58820FRA TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAACGAACTGTA---GAAATTAATTGTACA
WU46122BRA (MG) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAAATGAACTGTA---GAAATTAATTGTACA
WU00410BRA (RJ) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GTAATTAATTGTACA
WAJ289510DNK TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---AAAATTAATTGTACA
WL38420LUT TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAACGAACTGTA---GAAATTAATTGTACA
WU00407BRA (RJ) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAACCAATCTGTA---GAAATTAATTGTACA
PAF016551HI TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
PZ92644KOR TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
PAF090618ESP T-TGAAAATTCACAAAACATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
PAF181436CHE TCTGAAAATTCACAAAACATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
PU58818DJJ TCGAAAATTCACAAAACATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GTAATTAATTGTACA
PAF331440AUS TCTGACAATTCCTCAGACAATGCTAAAAGTATAATAGTACAGCTGACGAACTGTA---AAAATTAATTGTGCA
PAF060968BRA (RJ) TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACATCTTAATGAACTGTA---GAAATTAATTGTACA
PAF060960BRA (RJ) TCTGCAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACA
PAB002942JPN TCTGAAAATTCACAAAACATGCTAAAACCATATAATAGTACAGCTGAAAGAACTGTA---GAAATTAATTGTACA
PAF082482LUT TCTAGCAATTCCTCAGACAATGCTAAAACCATATAATAGTACAGCTGAAAGAACTGTA---GTGATTAATTGTACA
WAF060953BRA (RJ) TCTGAAAATTCACAAAACATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GCAATTAATTGTACA
WAF032901BOL TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGATGAACTGTA---GAAATTAATTGTACA
WAF245223BOL TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAACCAATCTGTA---GAAATTAATTGTACA
WU95570PHL TCTGAAAATTCACGACAATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTACT
WU95566PHL TCTGAAAATTCACAAAACATGCTAAAACCATATAATAGTACAGCTGAATGAACTGTA---GAAATTAATTGTATA

WU95575PHL TCTGAAAATTTTCACGAACAATGCTAAAACCATATAGTACAGCTGAATGCAACTGTA---GAAATTAATTGTACA
WAF060956BRA (RJ) TCTGAAAATTTTCACGAACAATGCTAAAACCATATAATAGTACAGCTGAATGCAACTGTA---GAAATTAATTGTACA
PAJ224186THA TCTGAAAATATACACGCAATGCTAAAACCATATAATAGTACATCTGAACCAATCTGTA---CCGATGAATTGTATA
PU28678CYP TCTGCAATTTTCACGAACAATGCTAAAACCATATAATAGTACAGCTGAACAAAACCTGTA---GAAATTAATTGTACA
PAF142260ESP TCTGAAAATTTTCACGCAATGCTAAAACCATATAATAGTACAGCTGAACAAATCTATT---GAAATTAATTGTACA

ZA59 AGACCAACCAACAATACA---AGACAAAGTACACATATAGGACCAGGGCAA---GCATTTTATACAACAGGAGCA
PU56380ETH AGACCCAGCAACAATACA---AGAAAAGTATAACCATGGGCCAGGGAGA---GTATATTATACAACAGGACAA
PX78513BRA (BA) AGACCAACCAACAATACA---AGAAAAGGATGACTCTAGGACCAGGGAGA---GTATATTATACAACAGGAGAA
PU08446HTI AGACCAATAACAATACA---AGGAAAAGGATATCTATAGGACCAGGGAGA---GTATGGTATACAACAGGACAA
PAY037268ARG AGACCTGGCAACCGTACA---AGAAGAGGATACATCTAGGACCAGGGAGA---GTATATTATACAACAGGAGAA
PAJ277821BEL AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGAGG---GCAATTTTATACAACAGGAGAA
PU08796BRA AGACCAACCAACAATATACAAGAAAGGATATAATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
PX78515BRA (BA) AGACCAACCAACAAGAAA---AGAAGAGGATACATATAGGACCAGGGAGA---GCATTTTATGGAACAGAA---
PU08797BRA AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
PU08798BRA AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
PX78517BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
PU08775BRA AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
PU65538MWI AGACCAACCAACAATACA---AGAAAAGGATACATCTAGGACCAGGGCAA---GCATGGTATACAACAGGACAA
PU08801THA AGACCAACCAACAATACA---AGAAAAGGATACATCTAGGACCAGGGCGA---GCATGGTATACAACAGGACAA
PAF034021BRA (RJ) AGACCCAGCAACAATACA---AGAAGAGGATACATATAGGACCAGGGAGA---GCATGGTATACAACAGGAA
PU08772BRA AGACCCAGCAACAATACA---AGAAAAGGATACATCTAGGACCAGGGGGT---TCATTTTATACAACAGGAGAA
PL07226HTI AGACCAATAACAATACA---AGAAAAGGATATCCCTATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
PAF113570BRA (RJ) AGACCAATAACAATACA---AGGAAACGGATATCTATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
PU23138USA AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
SUBTYPE_A AGACCTAGCAACAATACA---AGAACAGGATACATCTAGGACCAGGACAA---GCATTTTATACAACAGGAGAA
PAF200874MEX AGATCTTACAGCTATACC---AGAAGACGATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
PAF169616KOR AGACCAACCAACAATACA---AGAAAAGGATATAAGGATAGGACCAGGGAGC---GCATTTTATACAACAGGAGAA
PAF034010BRA (RJ) AGACCAACCAACAATACA---ATAAAAGGATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
PAF034012BRA (RJ) AGACCAATAACAATACA---AGAAAAGGGGATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
WAF113565BRA (RJ) AGACCCAGCAACAATACA---AGGAAAAGTATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF034030BRA (RJ) AGACCCAGCAACAATACA---AGAAGAGGATACATCTAGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WY18596REU AGACCCAGCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---TCATTTTATACAACAGGAGAA
WAF080212CHN AGACCAACCAACAATACA---AGGAAAAGTATACATATGAGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF113574BRA (RJ) AGACCAATAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF231041BRA AGACCAACCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WU31591BRA (SP) AGACCAACCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF060965BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATATCCATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF188990DEU AGACCAATAACAATACA---AGAAAAGGATACATCTGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF034029BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF034028BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATATCCATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF034016BRA (RJ) AGACCAACCAACAATACA---AGGAAAAGTATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF094014USA AGACTTAACAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF034026BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF060952BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATACATCTAGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WU08780BRA AGACCAACCAACAATACA---AGAAAAGGATACATCTAGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF034024BRA (RJ) AGACCCAGCAACAATACA---AGAAAAGGATACATCTGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF033997BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---ACATTTTATACAACAGGAGAA
WU08975BRA (SP) AGACCAACCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---ACATTTTATACAACAGGAGAA
WAF200859MEX AGACCAACCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF034005BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---TCATTTTATACAACAGGAGAA
WAF033999BRA (RJ) AGACCAACCAACAATACA---AGAAGAGGATACATATGGGCTGGGGGAGA---GCATTTTATACAACAGGAGAA
WAF034001BRA (RJ) AGACCTAACAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
PAF113571BRA (RJ) AGACCCAGCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---ACATTTTATACAACAGGAGAA
WU83693BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WU83699BRA (RJ) AGACCCAGCAACAATACA---AGAAAAGGATACATCTGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WU28336BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WU83691BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WU08771BRA AGACCAACCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WL19331BRA AGACCCAGCAATAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---ACATTTTATACAACAGGAGAA
PAF034033BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGGAGC---GCAATGTATACAACAGGAGAA
PAF034018BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGGAGA---GCATGGTATACAACAGGAGAA
PAF113572BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGAGC---GCAATTTATACAACAGGAGAA
PAF034027BRA (RJ) AGACCAACCAACAATACA---AGCAAGGATACATATAGGACCAGGGGAGA---GCAATTTATACAACAGGAGAA
PY95485ITA AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGGAGA---ACATTTTATACAACAGGAGAA
PY18759BRA (BA) AGACCAACCAACAATACA---AGAAGAGGATACAGATAGGACCAGGCAAA---GCATTTTATACAACAGGAGAA
PY18753BRA (BA) AGACCAACCAACAATACA---AGAAGAGGATACATATAGGACCAGGCAGA---GCATTTTATACAACAGGAGAA
PU56320SWE AGACCAACCAACAATACA---AGAAGAGGATACATATAGGACCAGGGGAGA---GCATTTTATACAACAGGAGAA
PU68201IDN AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGGAGA---GCATTTTATACAACAGGAGAA
PAF034004BRA (RJ) AGACCAACCAATAACT---AGAAAAGGATACATATAGGACCAGGGGAGA---GCATTTTATACAACAGGGGAA
PAF137974USA AGACCAACCAACAATACA---AGAAAAGGATACATATAGCACCAGGGGAGA---GCATTTTATACAACAGGAGAA
PY18754BRA (BA) AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGGAGA---GCATTTTATACAACAGGAGAA
PAF033996BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGGAGA---GCATTTTATACAACAGGAGAA
PY18760BRA (BA) AGACCAACCAACAATACA---AGCAAAAGTATAAATATAGGACCAGGGGAGA---GCATGGTATACAACAGGAGAA
PY18752BRA (BA) AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGGAGA---GCATTTTATACAACAGGAGAA
PAF368558FRA AGACCAACCAACAATACA---AGAAGAGGATACATATAGGACCAGGGGAGA---GCATGGTATACAACAGGAGAA
PAF034025BRA (RJ) AGACCAACCAACAATACA---AGAAAAGGATATCCATATGGGACCAGGGGAA---GCATTTTATACAACAGGAGAA
PY18755BRA (BA) AGACCAACCAACAATACA---AGAAAAGGATATAAATATAGGACCAGGGGAGA---GCATTTTATACAACAGGGGAA
PU13245HOL AGACCAACCAACAATACA---AGAAGAGGATATAAATATAGGACCAGGGGAGA---GCATTTTATACAACAGGAGAA
PAF022277HOL AGACCAACCAACAATACA---AGAAAAGGATATCTCTATAGGACCAGGGGAGA---GCATTTTATACAACAGGAGAA
PAF204580USA AGACCAACCAACAATACA---AGAAAAGGATATAACTATAGGACCAGGGGAGA---GCATTTTATACAACAGGAGAA
PAB059299JPN AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGGAGA---GCATTTTATACAACAGGAGAA
PU08781BRA AGACCAACCAACAATACA---AGAAAAGGATATAACCATAGGACCAGGGGAGA---GCATTTTATACAACAGGAGAA
PU16384AUS AGACCAACCAACAATACA---AGAAGAGGATACATATAGGACCAGGGGAGA---GCATTTTATACAACAGGAGAA
PAF279027SWE AGACCTAACAACAATACA---AGAAGAGGATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
PL19335BRA AGACCAACCAACAATACA---AGAAAAGGATACATATAGGACCAGGGGAGA---ACATTTTATACAACAGGAGAA
PL19332BRA AGACCAACCAACAATACA---AGAAAAGGATATCCCTATAGGACCAGGGGAGA---GCATTTTATACAACAGGAGAA
WU37041ARG AGACCAACCAACAATACA---AGCAAAAGGATACATATGGGATGGGGGAGA---ACATTTTATACAACAGGAGAA
WU37038ARG AGACCAACCAACAATACA---AGAAAAGGATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA

WAF060962BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACCTATGGGATGGGGGAGA---ACATTATATGCAACAGGAGAA
WAF060959BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGA---GCATTTTATACAATGGAGAA
WAF060958BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGA---GCATTTTATACAATGGAGAA
WU16219BRA AGACCCAACAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGA---ACATTTTATGCAACAGGAGAA
WU00400BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACACATAGGATGGGGGAGA---GCAATTTATGCAACAGGAGCA
WU00418BRA (RJ) AGGCCCCAACCAATACA---AGAAAAGTATACATTTAGGATGGGGGAGA---GCATTTTATACAACAGGACAG
PAF060954BRA (RJ) AGGCCCCAACCAATACA---AGAAAAGTATACATTTCCGGCCAGGAAGT---GCAATGTATGCAACAGGACAA
PU00417BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACCTATAGCACCAGGGAGC---GCATGGTTTCAACAGGAGAA
PU00427BRA (SP) AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
WL19334BRA AGACCCAACAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGA---GCATTTTATACAACCGGAGAA
WU00419BRA (RJ) AGGCCCCAACCAATACA---AGAAAAGTATACATTTAGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WU00412BRA (RJ) AGGCCCCAACCAATACA---AGAAAAGTATACATTTAGGATGGGGGAGA---GCATTTTATGCAACAGGAGCT
WL16605ARG AGACCCAACAACAATACA---AGAAAAGTATACATTTAGGATGGGGGAGA---GCATTTTATGCAACAGGAGAG
WU00401BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACATGTAGGATGGGGGAGA---GCATTTGATACAACAGGAGAG
PU00414BRA (RJ) AGACCCAGCAACAATACA---AGAAAAGTATACATTTAGGACCAGGGGGA---GCATTTTATACAACAGGAGAA
WU00413BRA (RJ) AGACCCAGCAACAATACA---AGAAAAGTATACATTTAGGATGGGGGAGA---GCATTTTATGCAACAGGAGAA
WAF060957BRA (RJ) AGGCCCCAACCAATACA---AGAAAAGTATACAGATAGGATGGGGGAGA---GCATTTTATGCAACAGGAGAA
WAF032151HOL AGACCCAACAACAATACA---AGAAAAGTATACATTTAGGATGGGGGAGA---GCATTTATATGCAACAGGAGAA
WU00409BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACATATAGGATGGGGGAGA---GCATTTTATGCAACAGGAGAA
WAF152771ESP AGACCCAACAACAATACA---AGAAAAGTATCCACATGGGATGGGGAAGA---GCATTTTATGCAACAGGAGAA
WX78512BRA (SP) AGACCCAACAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGA---ACATTTTATGCAACAGGAGAA
WAF188941FRA AGACCCAGCAACAATACA---AGAAAAGTATACATATGGGATGGGGAAGA---GCATTTTATGCAACAGGAGAA
PAF060955BRA (RJ) AGGCCCCAACCAATACA---AGAAAAGTATACATTTAGGACCAGGAAGA---GCATTTTATGCAACAGGAGAA
WAF060964BRA (RJ) AGACCCAGCAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGA---ACATTTTATACAACAGGAGAA
WAF034023BRA (RJ) AGCCCAAGCAACAATACA---AGAAAAGTATACACTTAGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WL19329BRA AGACCCCTTCAACTACA---AGAAAAGTATACATACAGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
PU80831BRA (SP) AGACCTTAAACAATACA---AGAAAAGTATACATATAGGACCAGGGAAA---GCATTTTATGCAACAGGAGAA
PU80830BRA (SP) AGACCTTAAACAATACA---AGAAAAGTATACATATGGGACCAGGGAAA---GCATTTTATGCAACAGGAGAA
WU28955PRY AGACCCAACAACAATACA---AGAAAAGTATACATCTAGGATGGGGGAGA---ACATTTTATACAACAGAAAAA
WAF188956FRA AGACCCGGCAACAATACA---AGAAAAGTATACATCTAGGATGGGGAAGA---TCACCTTATGCAACAGGAGAC
WU08776BRA AGACCCGGCAACAATACA---AGAAAAGTATACATATGGGATGGGGAAGA---GCATTTTATGCAACAGGAGAA
WU08024BRA (SP) AGACCCAGCAACAATACA---AGAAAAGTATACCTTAGGATGGGGAAGA---GCATTTTATGCAACAGGAGAA
WU80825BRA (SP) AGACCCAACAACAATACA---AGAAAAGTATACATTTAGGATGGGGAAGA---GCATTTTATGCAACAGGAGAA
WAF033998BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
WU80833BRA (SP) AGACCCAACAACAATACA---AGAAAAGTATACATCTAGGATGGGGGAGA---GCATTTTATGCAACAGGAGAA
WAF034013BRA (RJ) AGACCCGGCAACAATACA---AGAAAAGTATACATCTAGGATGGGGGAGA---GCATTTTATGCAACAGGAGAA
WU08029BRA (SP) AGACCCGGCAACAATACA---AGAAAAGTATACCTCTGGGACCAGGGAGA---GCATTTTATGCAACAGGAGAA
PU80832BRA (SP) AGACCCAACAACAATACA---AGAAAAGTATACATTTAGGACCAGGGAGA---GCATTTTATGCAACAGGAGAA
PAF060966BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGCAGA---GCATTTTATGCAACA---GAC
WAF034002BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGGAGA---GCATTTTATGCAACAGGAGAA
PU95412ITA AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGCAGA---GCATTTTATGCAACAGGAGAA
PAF060967BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATAACTATAGGACCAGGCAGA---GCATTTTATACAACAGGAGAA
WU80709BRA AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGCAAA---GCATTTTATGCAACAGGAGAA
PAF060963BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACCTATAGGACCAGGGAGA---GCATTTTATGCAACAGGAGAC
PU00416BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACGTATAGGACCAGGGAGC---GCATTTTATGCAACAGGAGAC
PAF368745FRA AGACCCAGCAACAATACA---AGAAAAGTATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGCC
PU00415BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCCTGGGAGA---GCATGGTTTCAAGAGGAGCC
PU80236BRB AGACCCAACAACAATACA---AGAAAAGTATACATATAGCACCAGGGAGA---GCATTTTATGCAACAGGAGAA
PAF058149CAN AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
PAJ278211DNK AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
PAJ289592DNK AGACCCAACAACAATACA---AGAAAAGTATACGTCTNGGACCAGGGGCA---GCATTTTATACAACA---GAA
PU20019CHN AGACCCAACAACAATACA---AGAAAAGTATCTATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
PAF286549EST AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGGAGAGGACATTTTATGCAACA---GAT
PU31587BRA (SP) AGACCCAGCAACAATACA---AGAAAAGTATACACATAGGACCAGGGAGA---GCATTTTATGCAACAGGAGAA
PU00411BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAA
PU08687BRA AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGGAGA---GCATTTTATGCAACAGGAGAC
PAF051476RUS AGACCCAACAACAATACA---AGAAAAGTATTCATATAGGACCAGGGAGA---GCATTTTATGCAACAGGAGAC
PAF170653RUS AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGGAGA---GCATTTTATGCAACAGGAGAC
PU08714BRA AGACCCAGCAACAATACA---AGAAAAGTATACATATAGGACCAGGGAGA---GCATTTTATGCAACAGGAGAC
PAF076319BRA (AM) AGACCCACCACAGTACA---AGAAAAGTATACATATAGGACCAGGGGCA---GCATGGTATGCAACAGGAGAA
WU00405BRA (RJ) AGACCCGGCAACAATACA---AGAAAAGTATACATATAGGACCAGGCAGA---GCATTTTATACAACAGGAGAC
WZ68025HOL AGACCCAACAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGA---GCATTTTATGCAACAGGAGAA
PAF181323CHE AGACCCAACAACAATACA---AGAAAAGTATACATATGGGACCAGGGAGA---GCATTTTGGGCAACAGGAGAA
WAF223996CZE AGACCCAACAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGA---ACATTTTATGCAATGGGAGAA
WAF223998CZE AGACCCAGCAACAATACA---AGAAAAGTATACATATGGGATGGGGAAGA---ACATTTTATGCAACAGGAGAA
WAB012971JPN AGACCCAACAACAATACA---AGAAAAGTATATATATGATGGGGGAGA---GCAATGTATGCAACAGGAGAC
PAF181437SWE AGACCCAACAACAATACA---AGAAAAGTATACATATGGGATGGGGAAGA---GCATTTTATGCAACAGGAGAA
PU00402BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACATATGGGACCAGGGAGG---GCATTTTATACAACAGGAGAA
WL20963BRA (BA) AGACCCAACAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGG---ACATTTTATGCAACAGGAGAA
WU00424BRA (SP) AGACCCAGCAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGA---GCATTTTATGCAACAGGAGAA
WU58813FRA AGACCCAGCAACAATACA---AGAAAAGTATACATATGGGATGGGGAAGA---GCATTTTATACAACAGGAGCT
WU00425BRA (SP) AGACCCAACAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGT
WU58820FRA AGACCCAGCAACAATACA---AGAAAAGTATACACATGGGATGGGGAAGA---GCATTTTATACAACAGGAGCT
WU46122BRA (MG) AGACCC---AACAATACA---AGAAAAGTATACATCTAGGATGGGGGAGA---GCATTTTATGCAACAGGAGAA
WU00410BRA (RJ) AGGCCCCAATAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGTGAA
WAJ289510DNK AGACCCAACANAATAACA---AGAAAAGTATACATCTAGGATGGGGGAGA---GCATTTTATGCAACAGGAGAA
WL38420LTU AGACCCAACAACAATACA---AGAAAAGTATACATATAGGATGGGGGAGA---GCAGTTTATGCAACAGGAGAA
WU00407BRA (RJ) AGACCCAACAACCAGACA---AGAAAAGTATACATATGGGATGGGGGAGA---GCATTTTATACAACAGGAGAA
PAF016551HI AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGCAGA---GCATTTTATACAACAGGAGAC
PZ92644KOR AGACCCCAACAACAATACA---AGAAAAGTATACGTATAGGACCAGGGAGC---ACATTTTATGCAACAGGAGCT
PAF090618ESP AGACCAACAACAATACA---AGAAAAGTATATCTATAGGACCAGGGAGT---GCATTTTATACAACAGGAGAC
PAF181436CHE AGACCCAATAATAACA---AGAAAAGTATACCTATAGGACCAGGGAGA---GCATTTTATACAACAGGAGAC
PU58818DJI AGACCCAGCAATAACA---AGAAAAGTATACATATAGGACCAGGGAGA---GCATTTTATGCAACAGGAGAA
PAF331440AUS AGACCCAACAACAATACA---AGAAAAGTATACATATAGCACCAGGGAGA---GCATTTTATGCAACAGGAGAA
PAF060968BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACAGTTAGGACCAGGAGCA---GCATTTTATGCAACAGGAGAA
PAF060960BRA (RJ) AGACCCAGCAACAATACA---AGAAAAGTATACATATAGGACCAGGGAGT---GCATTTTATGCAACAGGAGAA
PAB02942JPN AGACCCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGGAAA---GCATTTTATGCAACAGGAGAA
PAF082482LTU AGACCCAACAACAATACA---AGAAAAGTATAAATATAGCACCAGGGAGA---GCAATTTTATGCAACAGGAGAA
PAF060953BRA (RJ) AGACCCAACAACAATACA---AGAAAAGTATACATCTAGGACCAGGGGAGA---ACATTTTATGCAACAGGAGAA
WAF032901BOL AGACCCAACAACAATACA---AGAAAAGTATACATATGGGATGGGGGAGA---ACACTTTTATGTAACAGGACAA
WAF245223BOL AGACCCAACAACAATACA---AGAAAAGTATACATATGAGGTGGGGGAGG---GACTTTTATGCAACAGGAGAC

WU95570PHL AGACCCAGCAACAATACA---AGAAAAGTATACATGTAGGATGGGGGAGA---TCATTATATGCAACAGGACAA
WU95566PHL AGACCCAGCAACAATACA---AGAAAAGTATACATGTAGGATGGGGGAGA---ACATTTTATGCAACAGGAGAA
WU95575PHL AGACCCAGCAACAATACA---AGAAAAGTATACATGTAGGATGGGGGAGA---GCATTTTATGCAACAGGAGAA
WAF060956BRA (RJ) AGGCCCCAGCAACAATACA---AGAAAAGTATACAGATAGGATGGGGGAGA---GCATTTTATGCAACAGGAGAA
PAJ224186THA AGCCCAACAACAATACA---AGAAAAGTATACATATAACACCAGGGGAGA---GCATTTATATACAACAGAC---
PU28678CYP AGACCAACAACAATACA---AGAAAAGTATACATATAGGACCAGGGGAGA---GCATTTTATGCAACAGGAGAC
PAF142260ESP AGACCAACAACAATACA---AGCGAAAAGTATACATATAGGACCAGGGGAAA---GCATTTTATGCAACAAAAAGA

ZA59 ATAATAGGAGATATAAGACAAGCACATTGTAACATTAATAGAGCAAAA-----
PU56380ETH ATAATAGGAGATATAAGAAAAGCACATTGCAACATTAGTAGTACAAAA---TGGAAAT---AACGCTTTAGGACAG
PX78513BRA (BA) ATAGTAGGAGATATAAGAAAAGCACATTGTAACATTAGTAGAACAAAA---TGGAAAT---AACACCTTTAGAAAAG
PU08446HTI ATAATAGGAAATATAAGAAAAGCACATTGTAACCTTAGTAGCACAGCA---TGGAAAT---AACACCTTTAGGACAG
PAY037268ARG ATAATGGGAGATATAAGAAAAGCACATTGTAACCTTAGTAAAGCAGAC---TGGAAAT---AACACTGTAAGAAAAG
PAJ277821BEL ATAATAGGAGATATAAGACAAGCACATTGCAACCTTAATAGAGCACAA---TGGGAT---GACACTTTAAAAATG
PU08796BRA ATAAAAGGAAATATAAGA---CCATATTGTAACCTTAGTAGTAAAAA---TGGAAAT---AACACTTTAAAAACAG
PX78515BRA (BA) ATAACAGGAGATATAAGACAAGCACATTGTAACATTAGTAGAGCAAAA---TGGAAAT---AACACTTTAAAAACAG
PU08797BRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGCAGAGCACAA---TGGAAAT---GACACTTTAAGCAAG
PU08798BRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGCAGAGCACAA---TGGAAAT---GACACTTTAAGCAAG
PX78517BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGCAACCTTAGCAGAGCACTA---TGGAAAT---GACACTTTAAAAACAG
PU08775BRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGCAGAGCACAA---TGGAAAT---AACACTTTAAAAACAG
PU65538MWI ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGCACAAAA---TGGAAAT---AACACTTTAAAAACAG
PU08801THA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGCACAAAA---TGGAAAT---AACACTTTAAGACAG
PAF034012BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAACAGCA---TGGGAA---AAAACCTTTACAAAGG
PU08772BRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGGACAAAA---TGGGAA---AACACTTTAAAAACAG
PL07226HTI ATAATAGGAGATATAAGACAAGCATATTGTAACCTTAATAGAACACAA---TGGGAT---GGCACTTTAAAAACAG
PAF113570BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAAGACAA---TGGAAAT---GAACTTTAGATCG
PU23138USA ATAATAGGGGATATAAGACAAGCACATTGTAACATTAGTGGGAAAAA---TGGAAAT---AACACTTTAAGCAG
SUBTYPE A ATAACAGGGGATATAAGACAAGCACATTGTAACCTTAGTAGTAAAGCAGCA---TGGAAAT---AAAACCTTACAAAGC
PAF200874MEX ATAGAAGGAGATATAAGACAAGCATATTGTAACCTTAGTAGAGACAGCA---TGGAAAT---AAAACCTTTAAGACAG
PAF169616KOR ATAATAGGAGATATAAGACAAGCACATTGTAACATTAGTGCAGCAAAA---TGGGAA---AACACTTTAAGACAG
PAF034010BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGCAACCTTAGTAGAACAAAA---TGGGAA---AACACGTTAAAAACAG
PAF034012BRA (RJ) ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTGCATCAGAT---TGGAAAT---AAAACCTTTAAGACAG
WAF113565BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAATAAAAACAAAA---TGGAAAT---GAACTTTAAGACAG
WAF034030BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAACACGA---TGGGAA---GAACTTTAAGACAG
WY18596REU ATAACAGGGGACATAAGACAAGCACATTGTAACCTTAGTAGAAGACAGAA---TGGAAAT---AACACTTTAAGACAG
WAF080212CHN ATAGTAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGACCAAAA---TGGGAG---AAAACGTTAAGACAG
WAF113574BRA (RJ) AGGATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAGAAGCAAGA---TGGGAA---AACACTTTAAAAACAG
WAF231041BRA ATAATAGGAAATATAAGACAAGCACATTG-----
WU31591BRA (SP) ATAATAGGAAAAATAAGACAAGCACATTGTAACCTTAGTAGAGCAGAA---TGGAAAT---AACACTTTAAAAACAA
WAF060965BRA (RJ) ATAATAGGAGATATAAGACAAGCATATTGTAACCTTAACGCAACAGCA---ATGGAA---AAAACCTTTAAAC---
WAF188990DEU ATAATAGGAAATATAAGACAAGCATATTGTAACCTTAACGCAACAGCAAAA---TGGAAAT---AATACT-----
WAF034029BRA (RJ) ATAATAGGAGATATAAGACAAGCATATTGTAACATTAGTATAACAGAT---TGGAAAT---AAAACCTTTAAGCAG
WAF034028BRA (RJ) ATAATAGGAGATATAAGACAAGCATCTTGTAAACATTAGTATAACAGAT---TGGAAAT---AAAACCTTTAAGCAG
WAF034016BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAACAGCA---TGGAAAT---AACACTTTAAGACAG
WAF094014USA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAAGAGAA---TGGAAA---AATACTTTAAAAAAG
WAF034026BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAACAGAT---TGGAAAT---GAGACTTTAAGGAAG
WAF060952BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAAAGT-----
WU08780BRA ATAGTAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAGACATA---TGGGAA---AAAACCTTTAAAAAAG
WAF034024BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAACAGAA---TGGAAAT---AACACTTTAAGAAAAG
WAF033997BRA (RJ) ATAGTAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGCAG---TGGGAA---NCCACTTTAGGACGG
WU08975BRA (SP) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAAGTAGCA---TGGAAAT---AGAACTTTAAGAAAAG
WAF200859MEX ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTGGAGCAGAT---TGGAAAT---AACACTTTAGCAGAG
WAF034005BRA (RJ) ATAGTAGGAGATATAAGACAAGCACATTGTAACCTTAGTTCATGTC---TGGAAAT---AACACTTTAAGGAGAG
WAF033999BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAAGCAGAC---TGGGAA---AACACTTTAAGAACGG
WAF034001BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGGAGAAAA---TGGAAAT---AAAACCTTTAAAAACAG
PAF113571BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGCAGCAAAA---TGGAAAT---AACACTTTAAGAACAG
WU83693BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGCAAAA---TGGGAC---AACACTTTAAGAACAG
WU83689BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGCTAAA---TGGAAAT---AACACTTTAAGAACAG
WU28336BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGCAAAA---TGGAAAT---AACACTTTAAGAACAG
WU83691BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGCAAAA---TGGAAAT---AACACTTTAAGAACAG
WU08771BRA ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAGAGCAAAA---TGGAAAT---AACACTTTAAGAACAG
WL19331BRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGGTGTA---TGGAAAT---AACACTTTAAGAACAG
PAF034033BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGGAGGAA---TGGGAA---AAAACCTTTAAGAACGG
PAF034018BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGGAGAGACAAA---TGGGAA---AAAGCTTTAAGAACAG
PAF113572BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGTAAA---TGGAAAT---AATACTTTAAGAACAG
PAF034027BRA (RJ) ATAATAGGAGATATAAGACAAGCATATTGTAACATTAGTAAAGAAAA---TGGAAAT---AACACTTTAATATGAG
PU95485ITA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAGCAAAA---TGGAAAT---AAAACCTTTAAGAACAG
PY18759BRA (BA) ATAATAGGAGATATAAGACAAGCATATTGTAACCTTAGTAGAGCAAAA---TGGAAAT---AACACTTTAAGAACAG
PY18753BRA (BA) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGCAAAA---TGGAAAT---AGCACCTTTAAGAACAG
PU56320SWE ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTGAAGAAAA---TGGAAAT---AACACTTTAAGAACAG
PU68201IDN ATAATAGGGGATATAAGAAAAGCACATTGTAACCTTAGTAGGACAGAA---TGGAAAT---AATACTTTAAAAAAG
PAF034004BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGCAACA---TGGAAAT---AAAACCTTTAAGAACAG
PAF137974USA ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAGAGAAAA---TGGAAAT---AGCACTTTAAGAACAG
PY18754BRA (BA) ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAGAGTAGAA---TGGAAAT---AACACTTTAAGAACAG
PAF033996BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAGCAGAT---TGGAAAT---AACACTTTAAGAACAA
PY18760BRA (BA) ATAATAGGAGATATAAGACAAGCATATTGTAACCTTAGTAAAGCAAAA---TGGAAAT---AACACTTTAAGAACAG
PY18752BRA (BA) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGCAAAAACAAAA---TGGAAAT---AACACTTTAAGAACAG
PAF368558FRA ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAAAGCAAAA---TGGAAAT---AACACTTTAAGAACAG
PAF034025BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAAGCAAAA---TGGAAAT---AACACTTTAAGAACAG
PY18755BRA (BA) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGCAAAA---TGGAAAT---AACACTTTAAGAACAG
PU13245HOL ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAGACAAAA---TGGCAT---AACACTTTAAGAACAG
PAF022277HOL ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGCAGAA---TGGAAAT---AATACTTTAAGAACAG
PAF204580USA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTGAAGCACAA---TGGAAAT---GACACTTTAAGAACAG
PAB059299JPN ATACAAGGAGATATAAGACAAGCACATTGTAACCTTAGTGAAGCACAA---TGGAAAT---AACACTTTAAGAACAG
PU08781BRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAGCAAAA---TGGGAA---AGCACTTTAAGAACAG
PU16384AUS ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGAACAG---TGGCAT---AAAGCTTTAAGAACAG
PAF279027SWE ATAGTAGGAGATATAAGACAAGCACATTGTAACCTTAGTGAAGCAAAA---TGGAAAT---AACGCTTTAATGTCAG
PL19335BRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGCAGAA---TGGAAAT---AACACTTTAAGAACAG
PL19332BRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGCACAA---TGGAAAT---AACACTTTAAGAACAG

WU37041ARG ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAGAA---TGGAGA---CAAACCTTTAAAAACAG
WU37038ARG ATAATAGGAGATATAAGAAAAGCACATTGTAACCTTAGTAAAACACAC---TGGATA---GACACTTTAAAAAAG
WAF060962BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAATGAAACAAAA---TGGAGA---GACACTTTACAAAAG
WAF060959BRA (RJ) ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
WAF060958BRA (RJ) ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
WU16219BRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---ACCACTTTAGAACAG
WU00400BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAGAACAG
WU00418BRA (RJ) ATAATAGGAGAAATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AAAACTTTAAAAACAG
PAF060954BRA (RJ) ATAATAAG---ATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---CGAACTTTAAAAACAG
PU00417BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AATACTTTAAAAACAA
PU00427BRA (SP) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---GATACCTTTAAACACAG
WL19334BRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAGCAG
WU00419BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAGCAG
WU00412BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAACAAAAACAAAA---TGGGGA---AAAACTTTAGAACAG
WL16605ARG ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAGCAG
WU00401BRA (RJ) ATAATAGGTAATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAATG
PU00414BRA (RJ) ATAATAGGAAATATAAGAAAAGCATTGTAACCTTAATGAAACAAAA---TGGGGA---AATACTTTAAAAACAA
WU00413BRA (RJ) ATAGTGGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
WAF060957BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AAGACTTTAAAAACAG
WAF032151HOL ATAATAGGAGATATAAGACAAGCATATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACGTTAAGACAG
WU00409BRA (RJ) ATAACAGGAGATATAAGAAAAGCATATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAGACAG
WAF152771ESP ATAGTGGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
WX78512BRA (SP) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTCTAAACCAG
WAF188941FRA ATAATAGGAAACATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AAAACTTTAAAAACAG
PAF060955BRA (RJ) ATCATAGGAAACCAAGACAAGCACATTGTAACCTTAGTAAAACAGCANNNTGGGGA---CAAACCTTTACAAAAG
WAF060964BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAATAAGCAGAAA---TGGGGA---AACACTTTAACACAG
WAF034023BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
WL19329BRA ATCATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---GACACTTTAAAAACAG
PU80831BRA (SP) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTA
PU80830BRA (SP) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTA
WU28955PRY ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAGCAG
WAF188956FRA ATAATAGGAGACATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AAGACTTTAAAAACAG
WU08776BRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAA
WU80824BRA (SP) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTA
WU80825BRA (SP) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTA
WAF033998BRA (RJ) NTAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AGCACTTTAAAAACAG
WU80833BRA (SP) ATAGTGGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTA
WAF034013BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAGACAG
PU80829BRA (SP) ATAACAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AGCACTTTA
PU80832BRA (SP) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTA
PAF060966BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAGCAG
PAF034002BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---GACACTTTAAAAACAG
PU95412ITA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---GACACTTTAGAAAAG
PAF060967BRA (RJ) ATAATAGGAAATATAAGACAAGCACATTGTAACCTGAGTAGAGAC-----TGGGGA---AAAAACGCTACGAGAG
PU08709BRA ATAATAGGAAATATAAGACAAGCATATTGTAACCTTAGTAAAACAAAA---TGGGGA---AAAAACGCTACGAGAG
PAF060963BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACCG
PU00416BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---CACACTTTAACACAG
PAF368745FRA ATAACAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---GACACTTTAAAAACAG
PU00415BRA (RJ) ACCACAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACCCCTAAAAACAG
PU80236BRB ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAGACAG
PAF058149CAN ATAATAGGAGATATACGACAGGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
PAJ278211DNK ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAACACCAG
PAJ289592DNK ATAATAGGAGATATCAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAGATGG
PU20019CHN ATAATAGGCGATATACGACAAGGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---ACCACTTTAGACCAG
PAF286549EST ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAGACAG
PU31587BRA (SP) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
PU00411BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAACTG
PU08687BRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAACACAG
PAF051476RUS GTAACAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
PAF170653RUS ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAACACAG
PU08714BRA ATAATAGGAGACATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
PAF076319BRA (AM) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAGACAG
PU00405BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---GACACTTTAAGACAG
WZ68025HOL ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTACAAACAG
PAF181323CHE ATAGTAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
WAF223996CZE ATAGTAGGAAATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AAAACTTTAGAACAT
WAF223998CZE ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACCG
WAB012971JPN ATAGTAGGAGATATAAGAAAAGCATATTGTAACCTTAATAAAAAACAAAA---TGGGGA---AACACTTTAAGACAG
WAF181437SWE ATAATAGGAAATATAAGACAAGCATATTGTAACCTTAATGAAACAAAA---TGGGGA---AACACTTTAAAAACAG
PU00402BRA (RJ) ATTCTAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
WL20963BRA (BA) ATAATAGGAGCTATAAGAC-----TGGGGA---AACACTTTAAAAACAG
WU00424BRA (SP) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAGACAG
WU58813FRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AAAACTTTACAAAGAG
WU00425BRA (SP) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAGACAG
WU58820FRA ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAGACAG
WU46122BRA (MG) ATAATAGGAGATATAAGNACAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAGACAG
WU00410BRA (RJ) ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAGAAAAGAT
WAJ289510DNK ATAATAGGAGATATAAGACAAGCA-ATTGTAACCTTGCTGGGAAAAAC---TGGGGA---AACACTTTAAGACAG
WL38420LTU ATAATAGGAGATATAAGACGAGCATATTGTAACCTTAGTAAAACAAAA---TGGGGA---AGCACTTTAAAAAAG
WU00407BRA (RJ) ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAACACAG
PAF016551HI ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
PZ92644KOR ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAGACAG
PAF090618ESP GTAATAGGAAATATAAGAAATAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AAGACTTTAACACAG
PAF181436CHE ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
PU58818DJI ATAATAGGAAATATAAGACAAGCATATTGTAACCTTAGTAAAACAAAA---TGGGGA---AATGCTTTAAAAACAG
PAF331440AUS ATAATAGGAGATATAAGAAAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
PAF060968BRA (RJ) ATAATAGGAGACATCGAAAAGGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AACACTTTAAAAACAG
PAF060960BRA (RJ) GTAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AATACTTT-----
PAB002942JPN GTAATAGGAGATATAAGAAAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---AGCACTTTAAAAACAG
PAF082482LTU ATAATAGGAAATATAAGACAAGCACATTGTAACCTTAGTAAAACAAAA---TGGGGA---GACACTTTAAATACAG
PAF060953BRA (RJ) ATAATAGGAAATATAAGACTAGCATATTGTAACCTTAGTAAAACAAAA---TGGGGA---AAAACTTTAGACAG

WAF032901BOL ATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGCAAAA---TGGAA---AATACTTTAGAAAA
WAF245223BOL ATAATAGGAGATATAAGACAAGCACATTGTAACCTCAGTAGTAAAAGCCGTGG---AGCAACTTTAACACAG
WU95570PHL ATAATAGGAGATATAAGACAAGCACATTGT-----
WU95566PHL ATAATAGGAAATATAAGACAAGCACATTGT-----
WU95575PHL ATAATAGGAGATATAAGACAAGCACATTGT-----
WAF060956BRA (RJ) ATAATAGGGGATATAAGACAAGCACATTGTAACCTTAGTAGTACACAT---TGGAAC---AAGACTTTAAACAGG
PAJ224186THA ATAATAGGAGATATAAGACGAGCACATTGTAACCTTAGTAAAACAGAT---TGGGGT---GACACTTTAACACGG
PA228678CYP ATAATAGGAGATATAAGACTAGCATATTGTAACCTTAGTGGAGCAGAA---TGGAA---AACACTTTAAACACAG
PAF142260ESP ATAATAGGAGATATAAGAAAAGCATATTGTAATATTAGTAGAGTACAA---TGGAA---GAGACTTTAAACACAG

ZA59
PU56380ETH ATAGTTAAAAAATTAAGA-----GAACAA-----TTTAAG---AATAAAACAATAAC-
PX78513BRA (BA) ATAGTTGGAAAAATTAAGA-----GAACAA-----TTTGGG---AATAAAACAATAGTC
PU08446HTI ATAGTTGAAAAATTAAGA-----GCACAA-----TTTGGG---AATAAAACAATAAGA
PAY037268ARG ATAGCTATTTAAATTAAGA-----GAAAA---TTT---AATAAAACATAGTC
PAJ277821BEL ATAGTGAAAAAATTAAGA-----GAGCAA-----TTTGAG---AATAAAACAATAGTC
PU08796BRA ATAGTTACAAAAATTAAGA-----GAAAA---TTTGGG---AATAAAACAATAGTC
PX78515BRA (BA) GTAGTTACAAAAATTAAGA-----GAACAA-----TTTGGG---AATAAAACAATAGTC
PU08797BRA ATAGTTACAAAAATTAAGA-----GAACAA-----TTTGAG---AATAAAACAATAAAG
PU08798BRA ATAGTTACAAAAATTAAGA-----GAACAA-----CTTGAG---AATAAAACAATAAAG
PX78517BRA (RJ) ATAGTTACAAAAATTAAGA-----GAACAA-----GTTGGG---AATAGCACAAATAGAC
PU08775BRA ATAGTTAAAAAATTAAGA-----GAACAA-----TTTGGG---ATTAGCACAAATAGCC
PU65538MWI ATAGCTGAAAAATTAAGA-----AAACAA-----TTTGGG---AACAAAACAATAGTC
PU08801THA ATAACGAAAAAATTAAGA-----GAACAA-----TTTGGG---AACAAAACAATAGTC
PAF034021BRA (RJ) GTAGCTAGGAAAAATTAAGA-----GAACAA-----TTTGAG---AATAGCACAAATAGTC
PU08772BRA ATAGCTGATAAAATTAAGA-----GAACAA-----TTAAG---AATAAAACAATAGTC
PL07226HTI ATAGTTACAAAAATTAAGA-----GAACAA-----TTTGGGAGGAACAAAACAATAATC
PAF113570BRA (RJ) ACAGTTATAAAATTAAGA-----GAACAA-----TTTGGG---AATAAAACAATAGTC
PU23138USA ATAGTTGAAAAATTAAGA-----GAACAAGAACCAATTTAAG---AATAAAACAATAATC
SUBTYPE_A ATAGTTACACAATTAAGA-----GTATAC-----TGG-----AATAGAACAATAATC
PAF200874MEX ATAGTTACAAAAATTAAGA-----AAACAA-----TTTAAG---AATAAAACAATAGTC
PAF169616KOR ATAGCTATAAAATTAAGA-----GAACAA-----TTTGTG---AATAAAACAATAACC
PAF034010BRA (RJ) ATAGCTATAAAATTAATA-----NNCCAA-----TATGTA---AATAAAACAATAGTC
PAF034012BRA (RJ) ATCTCTGATAAAATTAAGA-----GAACAA-----TTTGGG---AATAAAACAATAGTT
WAF113565BRA (RJ) ATAGCTACAAAAATTAAGA-----GAGCAA-----TTTAGC---AATAAAACAATAATC
WAF034030BRA (RJ) ATAGTTATAAAGTTAAGA-----GAACAA-----TTTGGCAATAATAAAACAATAATC
WY18596REU ATAGTTAGAAAAATTAAGA-----GAACAA-----TTTGGG---AATAAAACAATAATC
WAF080212CHN ACAGTTAGAAAAATTAAGA-----GAACAA-----TTTGGC---AATAAAACAATAATC
WAF113574BRA (RJ) ATAGTTGATAAATTAGGA-----GAACAA-----TTTGGC---AATAAAACAATAGCC
WAF231041BRA
WU31591BRA (SP) ATAGCTATAAGATTAGAG-----AACCAA-----TTTACA---AACATA-----
WAF060965BRA (RJ) -----
WAF188990DEU
WAF034029BRA (RJ) GTAGCTGAAAAAATTAAGA-----GAACAA-----TTTAAG---AATAAAACAATAAAT
WAF034028BRA (RJ) GTAGCTGAAAAAATTAAGA-----GAACAA-----TTTAAG---AATAAAACAATAAAT
WAF034016BRA (RJ) ATAGCTAGAAAAATTACAA-----GAACGA-----TTT---AATAAAACAATATT---C
WAF094014USA ATAGCATATAAAATTTAGG-----GATCAA-----TTTGAG---AATAAAACAATAGCC
WAF034026BRA (RJ) ATAGCTGAAAAAATTAAGA-----GAACAA-----TTT-----AATAAAACAATAGTC
WAF060952BRA (RJ) -----
WU08780BRA ATAGCTATAAAATTAGGA-----GAACAA-----TTTAAG---AACAAAACAATAGCC
WAF034024BRA (RJ) ATAGCTAAAAAATTAACA-----GAACAA-----TTT-----AATAAAACAATAGCC
WAF033997BRA (RJ) NTAGCTGAAAAANTTANAA-----GAACAA-----TTT-----AATAGAACAATAGNC
WU08975BRA (SP) ATAGCTATAAAATTAAGG-----AACCAA-----TTT-----AATTACAACAATAGATA
WAF200859MEX ATAGTCACAAAAATTAAAA-----GAACAA-----TTTAAA---AATAAAACAATAGTC
WAF034005BRA (RJ) ATAGTTACAAAAATTAATAAAGAGAACAA-----TTTAGG---AACAAAACAATAATC
WAF033999BRA (RJ) ATAGCTATAAAATTAAGA-----GAACAA-----TTTCCG---AATAAAACAATTAAC
WAF034001BRA (RJ) ATAGTTGCAAAAATTAAAA-----GAAAA---TTT---AATACTACTACAATAATC
PAF113571BRA (RJ) ATAGTT-----TTAAAATTAATAGAAAA-----TTT-----AATACAACAATAATC
WU83693BRA (RJ) GTAGTTATGAAGTTAAGA-----GAACAC-----TTT-----AATAAAACAATAATC
WU83689BRA (RJ) ATAGTTATGAAGTTAAGA-----GAACAC-----TTT-----AATAAAACAATAATC
WU28336BRA (RJ) ATAGTTATGAAGTTAAGA-----GAACAC-----TTT-----AATAAAACAATAATC
WU83691BRA (RJ) ATAGTTATGAAGTTAAGA-----GAACAC-----TTT-----AATAAAACAATAATC
WU08771BRA ATAGTTATAAAATTAAAA-----GAACAC-----TTT-----AATAAAACAATAATC
WL19331BRA ATAGTTAGAAAAATTAAG-----GAACAC-----TTT-----AATAAAACAATAATC
PAF034033BRA (RJ) ATAGCTGCAAAAATTAAAA-----GAACAA-----TTTGAG---AATAGAACAAATAGCA
PAF034018BRA (RJ) ATAGCTACAAAAATTAAGA-----GAACAA-----TTC-----AATAAAATATAACC
PAF113572BRA (RJ) ATAGTTACAAAAATTAAGA-----GAACAA-----TTTGAG---AATAAAACAATAGTC
PAF034027BRA (RJ) ATAGTTAAAAAATTAAGA-----GAACAA-----TTTGAG---AATAAAACAATAGTC
PU95485ITA ATAGCTATAAAATTAAGA-----GAACAA-----TTT-----AATAAAACAATAGTC
PY18759BRA (BA) ATAGTTGGAAAAATTAAGA-----GAACAA-----TTTGGG---AATAAAACAATAGTC
PY18753BRA (BA) ATAGTTGGAAAAATTAAGA-----GAACAA-----TTTGGG---AATAAAACAATAGTC
PU56320SWE ATAGCTGACAAAATTAAGA-----GAACAA-----TTTGGG---AGTAAAACAATAAGA
PU68201IDN GTAGTTGAAAAATTAGGA-----AAAGAA-----TTTAAG---GATAAAACAATAGCC
PAF034004BRA (RJ) GTAGTTGAAACAGTTAAGA-----AAACAC-----TTTATG---AATAGAACAATAGTC
PAF137974USA ATAGTTGAAAAATTAAGA-----GAAAA---TTTGGGAACAATAACAACAATAGTC
PY18754BRA (BA) ATAGTTGAAAAATTAAGA-----GAACAA-----TTTGGG---AATAAAACAATAGTC
PAF033996BRA (RJ) ATAGTTGGAAAAATTAAGA-----GAACAA-----TTTAAG---AATAAAACAATAGCC
PY18760BRA (BA) ATAGCTAAAAAATTAAG-----GAAAA---TTT-----AATAAAACAATAGTC
PY18752BRA (RJ) GTAGCTAAAAAATTAAGA-----GAAAA---TTTGT---AATAAAACAATAGTC
PAF368558FRA ATAGTTGTCAAATTAAGA-----GAAAA---TTTGGG---AATAAAACAATAATC
PAF034025BRA (RJ) GTGGTTGCAAAAATTAAGA-----GAACAA-----TTTGGGGGAATAAAACAATAATC
PY18755BRA (BA) GTAGTTACAAAAATTAAGA-----GAACAA-----TTTGGG---AATAATACAATAATC
PU13245HOL ATAGTTACAAAAATTAAGA-----GAACAA-----TTTGGGAATAATAAGACAATAGTC
PAF022277HOL ATAGTTAACAAAATTAAGA-----GAAAA---TTTAAG---AATAAAACAATAGTC
PAF204580USA GTAGTTAGAAAAATTAGA-----GAACAA-----TTTGT---AATAAAACAATAGTC
PAB059299JPN ATAGTTATAAAATTAAGA-----GAGCAA-----TTTAGG---AATAAAACAATAGCC
PU08781BRA ATAGTTAAAAATTTACAA-----GACCAA-----TTT-----AATAAAACAATAGTC
PU16384AUS ATAGCTGAAAAATTAAGA-----GAACAA-----TTT---AATAAAACAACAATAATC
PAF279027SWE ATAGCTGAAAAATTAAGA-----GAACAA-----TTT---AAAAATAAAACAATAATC

PL19335BRA ATAGTTAAAAAATTAAGA-----GAACAA-----TAT--AACATTAGAACAATAGTC
 PL19332BRA ATAGCTACAAAGTTAAGA-----GAACAA-----TTT--AAGAATAAAACAATAGTC
 WU37041ARG ATAGAGACAAAAATTAAGA-----GAACAA-----TTT--AATAATAAAACAATAATC
 WU37038ARG GTAGGGGTAATAAATTAAGA-----GAACAA-----TTT--AATAA--AACTATAATA
 WAF060962BRA (RJ) ATAAGCTACAAAAATTAAG-----AGAAC-----
 WAF060959BRA (RJ) GTAGCTATAAAAATTAAGG-----GAACAA-----TAT--AACATAAAACAATAATG
 WAF060958BRA (RJ) GTAGCTATAAAAATTAAGG-----GAACAA-----TAT--AAGAATAAAACAATAATC
 WU16219BRA ATAGCTAAAAAATTAAGA-----GAACAA-----TTT--AATAATAAAACAATAAAA
 WU00400BRA (RJ) ATAGCTAAAAAATTAAGA-----GAACAA-----TTT-----AATAAAACAATAGTC
 WU00418BRA (RJ) ATAGCTATAAAAATTAAGA-----GAACAA-----TTTGGC--AATAAAACAGTAATC
 PAF060954BRA (RJ) ATAGCTATAAAAATTAAGT-----AAACAA-----TTTGGC--AATAAAACAAGT---
 PU00417BRA (RJ) ATAGCTAAAAAATTAAGA-----GAACAA-----TTTGAA--AATAAAACAATACAA
 PU00427BRA (SP) ATAGTTACAGAATTAAGA-----GAACAA-----TTAAG--AATAAAACAATAGTC
 WL19334BRA GTAGCTAAAAAATTAAGA-----GAACAA-----TTT--AATAATAAAACAATAGTC
 WU00419BRA (RJ) -----
 WU00412BRA (RJ) ATAGTTATAAAAATTAAGA-----GAACAA-----TTTGTC--AATAAAACAATAACC
 WL16605ARG -----
 WU00401BRA (RJ) ATAGCTGCAAAAATTAAGA-----GAACAG-----TTT-----AATAAAACAATAAAA
 PU00414BRA (RJ) ATAGCTGAAAAATTAAGA-----GAACAA-----TTT-----AATAAAACAATAGCC
 WU00413BRA (RJ) ATAGCTATAAAAATTAAGA-----GAACAA-----TTTGGG--AATAAAACAATAGAA
 WAF060957BRA (RJ) ATAGCTACAAAATTAAGA-----GAA-----
 WAF032151HOL ATAGCTATAAAAATTAAGG-----GAACAA-----TTAAG--AACAAAACAATAGTC
 WU00409BRA (RJ) ATAGCTATAAAAATTAAGG-----GAACAG-----TTAAG--AACAAAACAATAATC
 WAF152771ESP GTAGCTATAAAAATTAAGG-----GAA-----
 WX78512BRA (SP) ATAGCTATAAAAATTAAGA-----GAACAA-----TTT-----AATAACACAATAAGC
 WAF188941FRA -----
 PAF060955BRA (RJ) GTAGCTATAAAAATTAACA-----GAACAA-----TTCA-----
 WAF060964BRA (RJ) ATAAGGCAAAAATTAAG-----AAAACA-----
 WAF034023BRA (RJ) ATAGTTATAAAAATTAAGG-----GAACAA-----TTTAGA--AATAAAACAACAATC
 WL19329BRA ATAGTTATAAAAATTAAGG-----GAACAA-----TTAAG--AATAAAACAATAGCC
 PU80831BRA (SP) -----
 PU80830BRA (SP) -----
 WU28955PRY ATAATTATAAAAATTAAGA-----GAACAA-----TTTACG--AACAAAACAATAGTC
 WAF188956FRA -----
 WU08776BRA ATAGCTACAAAATTAAGA-----GGACAA-----TTT--CCTAACAAAACAATAGTT
 WU80824BRA (SP) -----
 WU80825BRA (SP) -----
 WAF033998BRA (RJ) ATAGCTATGAAAATTAAGA-----GAACAA-----TTT-----GATAAAACAACAACA
 WU80833BRA (SP) -----
 WAF034013BRA (RJ) ATAGTTACAAAATTAAGA-----GAACAA-----TTT-----AATAAAACAATAATC
 PU80829BRA (SP) -----
 PU80832BRA (SP) -----
 PAF060966BRA (RJ) ATAGTTATAAAAATTAAG-----AGAACAA-----TTTGGGAAT-----
 PAF034002BRA (RJ) ACAGCTAGGAAAATTAAGA-----GAACAA-----TTTGGGAATAATACAACAATCTTC
 PU95412ITA ATAGTTATAAAAATTAAGA-----GAACAA-----TTTGGGAATAAAACAATAATC---
 PAF060967BRA (RJ) -----
 PU08709BRA GTAAGTAGAAAATTAAGCA-----GAGCAC-----TTCCCT--AATAAAACAATAATC
 PAF060963BRA (RJ) GTAGCTATAAAGTT-----
 PU00416BRA (RJ) ATAGGTAGAAAATTAACA-----GAACAA-----TTTAAT--AATAAAACAATAATC
 PAF368745FRA ATAGTTAAAAAATTAAGA-----GAACAA-----TTTGGG--AATAAAACAATAGTC
 PU00415BRA (RJ) ATAGTTATAAAAATTAACA-----GAACAA-----TTTGGG--AATAAAGCAATAATC
 PU80236BRB GTAGCTACAAAATTAAGG-----GAACAA-----TTTGGG--AATAAAACAATAATC
 PAF058149CAN ATAGTGAAAATTAAGG-----GAACAA-----TTTGGG--AATAGTATAACA---
 PAJ278211DNK ATAGTTAAAAAATTAAGA-----GAACAA-----TTTGGG--AGTAATACAACAATC
 PAJ289592DNK GTAGCCAACAATAAAG-----GAACAA-----TTGGGG--AATAATGAAAC---
 PU20019CHN ATAGTTAAAATTTAAGT-----GACCAG-----TTTGGG--AATAATACCCAGTC
 PAF286549EST ATAGTTACAAAATTAAGA-----GAACAA-----TTAAG--AATAAAACAATAGTC
 PU31587BRA (SP) ATAGTTACAAAATTAAGA-----GAACAA-----TTTA--AATAAAACAATAGTC
 PU00411BRA (RJ) ATAGTTACAAAATTAAGA-----GAACAA-----TTTGGG--AATAAAACAATAGCC
 PU08687BRA ATAGTTGAAAATTAAGA-----AAACAA-----TTTGAT--AATAAAACAATAGCC
 PAF051476RUS ATAGTTAGCAAAATTAAGA-----AAACAA-----TTTCGG--GATAAAACAATAGTC
 PAF170653RUS ATAGTTAGCAAAATTAAGA-----GAACAA-----TTTGGG--AATAAAACAATAGTC
 PU08714BRA ATAGTTGAAAATTAAGA-----GAACAA-----TTTGGG--AATAAAACAATAGTC
 PAF076319BRA (AM) ATAGCTGCAAAAATTAAGA-----GAACAA-----TTTGGG--AACAAAACAATAGTG
 PU00405BRA (RJ) ATAGCTATAAAAATTAAGA-----GAACAG-----TTAATGGAACAGAACAATAATC
 WZ68025HOL ATAGCTAGCAAAATTAAGG-----GAACAA-----TTTGAG--AATAAAACAATAGTC
 PAF181323CHE ATAGTTATGAAAATTAAGA-----GAACAA-----TTT--AATAAAACAATAATA
 WAF223996CZE ATAGCTGTAATAAATTAAGA-----GAACAA-----TTT-----AATAAAACAATAGCC
 WAF223998CZE ATAGCTATAAAAATTAAGA-----GAACAA-----TTT-----AATAAAACAATAAGT
 WAB012971JPN GTAGCTGAAAATTAAGAA-----GAACAA-----TTTAAAG--AATAAAACAATATC
 WAF181437SWE GTAGTTAGAAAATTAAGA-----GAACAA-----TTT--AATACTACAACAATAAAT
 PU00402BRA (RJ) -----
 WL20963BRA (BA) -----
 WU00424BRA (SP) ATACGC-----
 WU58813FRA GTAGGTAATAAATTAAC-----GAACAC-----TTCTTC--AATAAAACAATAAGC
 WU00425BRA (SP) -----
 WU58820FRA ATCGCTGTAATAAATTAGAA-----GCACAA-----TTCCAT--AAAACAACAATAGTC
 WU46122BRA (MG) -----
 WU00410BRA (RJ) ATAGCTAGAAAATTAAGA-----GAACAT-----TTTAAG--GAGAAACCAATAGTC
 WAJ289510DNK ATAGCTACAAAATTAAGG-----AAACAC-----TTTGA-----
 WL38420LTU ATAGCTACAAAATTAGGG-----GAACAC-----TTT--AAT-----
 WU00407BRA (RJ) ATAGCTGAAAATTAAGA-----GAACAA-----TTTGGG--AATAAAACAATAATC
 PAF016551HI ATAGTTGAAAATTAAGA-----GAACAA-----TTTGAG--AATAAAACAATAGTT
 PZ92644KOR CTAGTTATAAAAATTAAGA-----GAACAA-----TTTGGG--AATAAAACAATAGCC
 PAF090618ESP ATAAGTGAAAATTAAGA-----AAACAA-----TATGGG--AAT-----
 PAF181436CHE GTAGTTGAAAATTAAGA-----GAACAA-----TTTGGGAATAATAAAACAATAGTC
 PU58818DJI ATAGTTACAAAATTAAGA-----GAACAA-----TTTGGGAATAA--AACATAAAC
 PAF331440AUS -----
 PAF060968BRA (RJ) -----
 PAF060960BRA (RJ) -----
 PAB002942JPN ATGGTTGAAAATTAAGA-----GAACAA-----TTTGAG--AATAAAACAATACAA

PAF082482LTU	ATA-----
PAF060953BRA (RJ)	GTAGCTAGAAAATTAGGA-----GAACAA-----TTT-----AACAAAA-----
WAF032901BOL	ATAGCTACAAAATTAAAG-----GAACAA-----TTT-----AATAAAA-----
WAF245223BOL	GTAGCTAACAAAATTAAAA-----GAACAC-----T-----
WU95570PHL	-----
WU95566PHL	-----
WU95575PHL	-----
WAF060956BRA (RJ)	ATAGCTATAAAAATTAGGA-----GAACAA-----TTT-----AATAAAACAATAATC
PAJ224186THA	GTAGCTGAAAAGTTAAGT-----GCTCAA-----TTTGGG---AATGGTAATATAACC
PU28678CYP	ATAGTTGGCAAATTACTA-----GAAATAGAACAATTTAAG---AATAAAACAATAGTC
PAF142260ESP	ATAGTTGGAAAATTAGA-----GAACAA-----TTTGAAAAT-----

Anexo II O co-receptor CCR5

O co-receptor CCR5 é expresso na superfície de células T ativadas e de memória, monócitos e macrófagos, micróglia e, em menor quantidade, na superfície de células B (Carrington et al., 2001). A descoberta de que esta molécula era utilizada como co-receptor pelo vírus apontou para a influência de seu gene na infecção pelo HIV-1. A seguir, grupos independentes de pesquisadores identificaram um alelo mutante do gene, conhecido como CCR5- Δ 32, que se caracteriza por uma deleção de 32 pares de bases. A perda deste segmento gênico resulta em uma mudança na fase de leitura e na formação prematura de um códon de terminação. Em consequência, as moléculas produzidas são inoperantes e não podem ser expressas na superfície celular (Carrington et al., 2001; Roger, 1998).

O fato do alelo CCR5- Δ 32 não ser expresso na superfície das células tem como consequência óbvia o fato de os indivíduos homozigotos para este alelo serem, na sua quase totalidade, resistentes à infecção pelo HIV-1. Por outro lado, apesar dos heterozigotos (genótipo CCR5+/CCR5- Δ 32) não estarem protegidos contra a infecção, foi observado que estes indivíduos progridem mais lentamente para a AIDS do que aqueles que possuem os dois alelos selvagens (genótipo CCR5+/+). A explicação para isto é que, nas células desses pacientes, ocorre a formação de heterocomplexos CCR5- Δ 32/CCR5, o que diminui a quantidade média de moléculas CCR5 expressas na superfície celular. Desta forma, a replicação *in vitro* de isolados R5 é deficiente nas células obtidas de pacientes heterozigotos o que, *in vivo*, se traduz em níveis inferiores de carga viral (Carrington et al., 2001).

Lista de Referências

- ALAEUS, A.; LEINER, T.; LIDMAN, K. & ALBERT, J., 1997. Most HIV-1 Genetic Subtypes Have Entered Sweden. *AIDS* 11, 199-202.
- APETREI, C.; LOUSSERT-AJAKA, I.; COLLIN, G.; LETOURNEUR, F.; DUCA, M.; SARAGOSTI, S.; SIMON, F. & BRUN-VÉZINET, F., 1997. HIV Type 1 Subtype F Sequences in Romanian Children and Adults. *AIDS Research and Human Retroviruses* 13, 363-365.
- APETREI, C.; NECULA, A.; HOLM-HANSEN, C.; LOUSSERT-AJAKA, I.; PANDREA, I.; COZMEI, C.; STREINU-CERCEL, A.; PASCU, F.-R.; NEGUT, E.; MOLNAR, G.; DUCA, M.; PECEC, M.; BRUN-VÉZINET, F. & SIMON, F., 1998. HIV-1 Diversity in Romania. *AIDS* 12, 1079-1085.
- ARNOLD, C.; BARLOW, K. L.; PARRY, J. V. & CLEWLEY, J. P., 1995. At Least Five HIV-1 Sequences Subtypes (A, B, C, D, A/E) Occur in England. *AIDS Research and Human Retroviruses* 11, 427-429.
- BACHMANN, M. H.; DELWART, E. L.; SHPAER, E. G.; LINGENFELTER, P.; SINGAL, R.; MULLINS, J. I. & THE WHO NETWORK FOR HIV ISOLATION AND CHARACTERIZATION, 1994. Rapid Genetic Characterization of HIV Type 1 Strains From Four World Health Organization-Sponsored Vaccine Evaluation Sites Using a Heteroduplex Mobility Assay. *AIDS Research and Human Retroviruses* 10, 1345-1353.
- BALL, S. C.; ABRAHA, A.; COLLINS, K. R.; MAROZSAN, A. J.; BAIRD, H.; QUIÑONES-MATEU, M. E.; PENN-NICHOLSON, A.; MURRAY, M.; RICHARD, N.; LOBRITZ, M.; ZIMMERMAN, P. A.; KAWAMURA, T.; BLAUVELT, A. & ARTS, E. J., 2003. Comparing the Ex Vivo Fitness of CCR5-Tropic Human Immunodeficiency Virus Type 1 Isolates of Subtypes B and C. *Journal of Virology* 77, 1021-1038.
- BARRÉ-SINOUSSE, F.; CHERMANN, J. C.; REY, F.; NUGEYRE, M. T.; CHAMARET, S.; GRUEST, J.; DAUGUET, C.; AXLER-BLIN, C.; VÉZINET-BRUN, F.; ROUZIOUX, C.; ROZENBAUM, W. & MONTAGNIER, L., 1983. Isolation of a T-Lymphotropic Retrovirus From a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). *Science* 220, 868-871.

BELDA, F. J.; BARLOW, K. L.; MURPHY, G.; PARRY, J. V. & CLEWLEY, J. P., 1998. A Dual Subtype B/E HIV Type 1 Infection With a Novel V3 Loop Crown Motif Among Infections Acquired in Thailand and Imported into England. *AIDS Research and Human Retroviruses* 14, 911-916.

BOBKOV, A.; CHEINGSONG-POPOV, R.; SELIMOVA, L.; KAZENNOVA, E.; KARASYOVA, N.; KRAVCHENKO, A.; LADNAYA, N.; POKROVSKY, V. & WEBER, J. N., 1996. Genetic Heterogeneity of HIV Type 1 in Russia: Identification of H Variants and Relationship With Epidemiological Data. *AIDS Research and Human Retroviruses* 12, 1687-1690.

BOBKOV, A.; CHEINGSONG-POPOV, R.; SELIMOVA, L.; LADNAYA, N.; KAZENNOVA, E.; KRAVCHENKO, A.; FEDOTOV, E.; SAUKHAT, S.; ZVEREV, S.; POKROVSKY, V. & WEBER, J. N., 1997. An HIV Type 1 Epidemic Among Injecting Drug Users in the Former Soviet Union Caused by a Homogeneous Subtype A Strain. *AIDS Research and Human Retroviruses* 13, 1195-1201.

BOBKOV, A.; KAZENNOVA, E.; SELIMOVA, L.; BOBKOVA, M.; KHANINA, T.; LADNAYA, N.; KRAVCHENKO, A.; POKROVSKY, V.; CHEINGSONG-POPOV, R. & WEBER, J. N., 1998. A Sudden Epidemic of HIV Type 1 Among Injecting Drug Users in the Former Soviet Union: Identification of Subtype A, Subtype B, and Novel GagA/EnvB Recombinants. *AIDS Research and Human Retroviruses* 14, 669-676.

BOCK, P. J. & MARKOVITZ, D. M., 2001. Infection with HIV-2. *AIDS* 15(suppl 5), S35-S45.

BONGERTZ, V.; BOU-HABIB, D. C.; BRÍGIDO, L. F. M.; CASEIRO, M.; CHEQUER, P. J. N.; COUTO-FERNANDEZ, J. C.; FERREIRA, P. C.; GALVÃO-CASTRO, B.; GRECO, D.; GUIMARÃES, M. L.; CARVALHO, M. I. L.; MORGADO, M. G.; OLIVEIRA, C. A. F.; OSMANOV, S.; RAMOS, C. A.; ROSSINI, M.; SABINO, E. C.; TANURI, A.; UEDA, U., 2000. HIV-1 Diversity in Brazil: Genetic, Biologic, and Immunologic Characterization of HIV-1 Strains in Three Potential HIV Vaccine Evaluation Sites. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 23, 184-193.

BÖNI, J.; PYRA, H.; GEBHARDT, M.; PERRIN, L.; BÜRGISSE, P.; MATTER, L.; FIERZ, W.; ERB, P.; PIFFARETTI, J.-C.; MINDER, E.; GROB, P.; BURCKHARDT, J. J.; ZWAHLEN, M. & SCHÜPBACH, J., 1999. High Frequency of Non-B Subtypes

in Newly Diagnosed HIV-1 Infections in Switzerland. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 22, 174-179.

Braun-Prado, K., Vieira Mion, A. L., Farah Pereira, N., Culpi, L. & Petzl-Erler, M. L., 2000. HLA class I polymorphism, as characterised by PCR-SSOP, in a Brazilian exogamic population. *Tissue Antigens* 56(5): 417-427.

BRENGEL-PESCE, K.; INNOCENTI-FRANCILLARD, P.; MORAND, P.; MICHAULT, A.; POUBEAU, P.; BARIN, F.; SARAGOSTI, S. & SEIGNEURIN, J.-M., 1999. Serologic and Genetic Characterization of HIV Type 1 Subtype on Reunion Island. *AIDS Research and Human Retroviruses* 15, 787-792.

BRODINE, S. K.; MASCOLA, J. R.; WEISS, P. J.; ITO, S. I.; PORTER, K. R.; ARTENSTEIN, A.; GARLAND, F. C.; MCCUTCHAN, F. E. & BURKE, D. S., 1995. Detection of Diverse HIV-1 Genetic Subtypes in the USA. *The Lancet* 346, 1198-1199.

BROWN, T. M.; ROBBINS, K. E.; SINNIHAH, M.; SARASWATHY, T. S.; LEE, V.; HOOI, L. S.; VIJAYAMALAR, B.; LUO, C.-C.; OU, C.-Y.; RAPIER, J.; SCHOCHTMAN, G. & KALISH, M. L., 1996. HIV Type 1 Subtypes in Malaysia Include B, C, and E. *AIDS Research and Human Retroviruses* 12, 1655-1657.

CABELLO, A.; CABRAL, M.; VERA, M. E.; KIEFER, R.; AZORERO, R. M.; EBERLE, J.; GÜRTLER, L. & VON BRUNN, A., 1995. Analysis of the V3 Loop Sequences From 10 HIV Type 1-Infected AIDS Patients From Paraguay. *AIDS Research and Human Retroviruses* 11, 1135-1137.

CAMPODONICO, M.; JANSSENS, W.; HEYNDRICKX, L.; FRANSEN, K.; LEONAERS, A.; FAY, F. F.; TABORDA, M.; VAN DER GROEN, G. & FAY, O. H., 1996. HIV Type 1 Subtypes in Argentina and Genetic Heterogeneity of the V3 Region. *AIDS Research and Human Retroviruses* 12, 79-81.

CARRILLO, M. G.; PICCARDO, C. & LIBONATTI, O., 1992. Análisis Molecular Del Principal Epítotope De Neutralización ("V3 Loop") Del Virus De La Inmunodeficiencia Humana Tipo 1 De La Argentina. *Revista Argentina de Microbiología* 24, 91-101.

Carrington, M. & O'Brien, S. J., 2003. The influence of HLA Genotype on AIDS. *Annual Reviews of Medicine* 54:535-551.

Carrington, M., Nelson, G. & O'Brien, S. J., 2001. Considering genetics profiles in functional studies of immune responsiveness to HIV-1. *Immunology Letters* 79:131-140.

CASADO, C.; URTASUN, I.; MARTÍN-WALTHER, M. V.; GARCÍA, S.; RODRÍGUEZ, C.; DEL ROMERO, J. & LÓPEZ-GALÍNDEZ, C., 2000a. Genetic Analysis of HIV-1 Samples From Spain. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 23, 68-74.

CASADO, C.; URTASUN, I.; SARAGOSTI, S.; CHAIX, M.-L.; DE ROSSI, A.; CATTELAN, A. M.; DIETRICH, U. & LÓPEZ-GALÍNDEZ, C., 2000b. Different Distribution of HIV Type 1 Genetic Variants in European Patients With Distinct Risk Practices. *AIDS Research and Human Retroviruses* 16, 299-304.

CASEMENT, K. S.; NEHETE, P. N.; ARLINGHAUS, R. B. & SASTRY, K. J., 1995. Cross-Reactive Cytotoxic T Lymphocytes Induced by V3 Loop Synthetic Peptides From Different Strains of Human Immunodeficiency Virus Type 1. *Virology* 211, 261-267.

CASSEB, J.; HONG, M. A.; GONSALEZ, C.; BRÍGIDO, L. F. M.; DUARTE, A. J. S. & MICHAEL-HENDRY, R., 1998. Two Variants of HIV-1 B Serotype Are Transmitted Heterosexually in São Paulo, Brazil. *Brazilian Journal of Medical and Biological Research* 31, 1243-1246.

CATASTI, P.; BRADBURY, E. M. & GUPTA, G., 1996. Structure and Polymorphism of HIV-1 Third Variable Loops. *The Journal of Biological Chemistry* 271, 8236-8242.

CATASTI, P.; FONTENOT, J. D.; BRADBURY, E. M. & GUPTA, G., 1995. Local and Global Structural Properties of the HIV-MN V3 Loop. *The Journal of Biological Chemistry* 270, 2224-2232.

CHEN, J.; YOUNG, N.; SUBBARAO, S.; WARACHIT, P.; SAGUANWONGSE, S.; WONGSHEREE, S.; JAYAVASU, C.; LUO, C.-C. & MASTRO, T. D., 1999. HIV Type 1 Subtypes in Guangxi Province, China, 1996. *AIDS Research and Human Retroviruses* 15, 81-84.

CHIJIWA, K.; ISHIBASHI, T.; KASHIWAGI, S. & MORI, R., 1999. The Distribution of HIV-1 Subtypes in Fukuoka, Japan. *Microbiology and Immunology* 43, 271-278.

CLAVEL, F.; GUÉTARD, D.; BRUN-VÉZINET, F.; CHAMARET, S.; REY, M. A.; SANTOS-FERREIRA, M. O.; LAURENT, A. G.; DAUGUET, C.; KATLAMA, C.;

ROUZIOUX, C.; KLATZMANN, D.; CHAMPALIMAUD, J. L. & MONTAGNIER, L., 1986. Isolation of a New Human Retrovirus From West Africa Patients With AIDS. *Science* 233, 343-346.

CLEWLEY, J. P.; ARNOLD, C.; BARLOW, K. L.; GRANT, P. R. & PARRY, J. V., 1996. Diverse HIV-1 Genetic Subtypes in UK. *The Lancet* 347, 1487.

COUTO-FERNANDEZ, J. C.; JANSSENS, W.; HEYNDRIKX, L.; MOTTE, J.; FRANSEN, K.; PEETERS, M.; DELAPORTE, E.; GALVÃO-CASTRO, B.; PIOT, P. & VAN DER GROEN, G., 1994. Genetic and Antigenic Variability of HIV Type 1 in Brazil. *AIDS Research and Human Retroviruses* 10, 1157-1163.

COUTO-FERNANDEZ, J. C.; MORGADO, M. G.; BONGERTZ, V.; TANURI, A.; ANDRADE, T.; BRITES, C. & GALVÃO-CASTRO, B., 1999. HIV-1 Subtyping in Salvador, Bahia, Brazil: A City With African Sociodemographic Characteristics. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 22, 288-293.

COVAS, D. T.; BÍSCARO, T. A.; KASHIMA, S.; DUARTE, G. & MACHADO, A. A., 1998. High Frequency of the GWG (Pro Trp) Envelope Variant of HIV-1 in Southeast Brazil. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 19, 74-79.

DELWART, E. L.; SHPAER, E. G.; LOUWAGIE, J.; MCCUTCHAN, F. E.; GREZ, M.; RÜBSAMEN-WAIGMANN, H. & MULLINS, J. I., 1993. Genetic Relationships Determined by a DNA Heteroduplex Mobility Assay: Analysis of HIV-1 Env Genes. *Science* 262, 1257-1261.

DITTMAR, M. T.; SIMMONS, G.; HIBBITTS, S.; O'HARE, M.; LOUISIRIOTCHANAKUL, S.; BEDDOWS, S.; WEBER, J.; CLAPHAM, P. R. & WEISS, R. A., 1997. Langerhans Cell Tropism of Human Immunodeficiency Virus Type 1 Subtype A Through F Isolates Derived From Different Transmission Groups. *Journal of Virology* 71, 8008-8013.

DOMINGO, E.; MENÉNDEZ-ARIAS, E. & HOLLAND, J., 1997. RNA Virus Fitness. *Reviews of Medical Virology* 7, 87-96.

DOOLITTLE, R. F. & FENG, D.-F., 1992. Tracing the Origin of Retroviruses. *Current Topics in Microbiology and Immunology* 176, 195-210.

- DOOLITTLE, R. F.; FENG, D. F.; MCCLURE, M. A. & JOHNSON, M. S., 1990. Retrovirus Phylogeny and Evolution. *Current Topics in Microbiology and Immunology* 157, 1-18.
- EFRON, B.; HALLORAN, E. & HOLMES, S., 1996. Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences of United States of America* 93, 13429-13434.
- ENGELSTAD, M., 1996. Genetic Diversity of Env V3 Loop Sequences in HIV Type 1-Infected Homosexuals, Heterosexuals, Intravenous Drug Users, and Hemophilia Patients in Norway: High Prevalence of Subtype B and Detection of Subtype C. *AIDS Research and Human Retroviruses* 12, 1733-1738.
- Erlich, H. A., Mack, S. J., Bergstrom, T. & Gyllensten, U. B., 1997. HLA class II alleles in Amerindians: implications for the evolution of HLA polymorphism and the colonization of the Americas. *Hereditas* 127(1-2): 19-24.
- ESSEX, E., 1999. Human Immunodeficiency Viruses in the Developing World. *Advances in Virus Research* 53, 71-88.
- EWALD, P. W., 1996. Guarding Against the Most Dangerous Emerging Pathogens: Insights from Evolutionary Biology. *Emerging Infectious Diseases* 2(4), 245-257.
- FERDATS, A.; KONICHEVA, V.; DIEVBERNA, I.; LILJA, E. & ALBERT, J., 1999. An HIV Type 1 Subtype A Outbreak Among Injecting Drug Users in Latvia. *AIDS Research and Human Retroviruses* 15, 1487-1490.
- FERNANDEZ-MEDINA, D.; JANSSON, M.; RABINOVICH, R. D.; LIBONATTI, O. & WIGZELL, H., 1999. Identification of Human Immunodeficiency Virus Type 1 Subtypes B and F, B/F Recombinant and Dual Infection With These Subtypes in Argentina. *Scandinavian Journal of Infectious Diseases* 31, 235-242.
- FOUCHIER, R. A. M. & SCHUITEMAKER, H., 1996. Molecular Determinants of Human Immunodeficiency Virus Type 1 Phenotype Variability. *European Journal of Clinical Investigation* 26, 175-185.
- GALLO, R. C.; SALAHUDDIN, S. Z.; POPOVIC, M.; SHEARER, G. M.; KAPLAN, M.; HAYNES, B.; PALKER, T. J.; REDFIELD, R.; OLESKE, J.; SAFAI, B.; WHITE, G.; FOSTER, P. & MARKHAM, P. D., 1984. Frequent Detection and Isolation of

Cytopathic Retroviruses (HTLV-III) From Patients With AIDS and at Risk for AIDS. *Science* 224, 500-503.

GALVÃO-CASTRO, B.; COUTO-FERNANDEZ, J. C.; MELLO, M. A.; CARVALHO, M. I. L.; CASTELLO-BRANCO, L. R.; BONGERTZ, V.; FERREIRA, P. C. P.; MORGADO, M.; SABINO, E.; TANURI, A. & THE BRAZILIAN NETWORK FOR HIV-1 ISOLATION AND CHARACTERIZATION, 1996. A Nationwide Effort to Systematically Monitor HIV-1 Diversity in Brazil: Preliminary Results. *Memórias do Instituto Oswaldo Cruz* 97, 335-338.

GALVÃO-CASTRO, B.; DOS SANTOS, J. I.; COUTO-FERNANDEZ, J. C.; BONGERTZ, V.; BOU-HABIB, D. C.; SION, F. S.; BARTH, O. M.; PEREIRA, H. & PEREIRA, M. S., 1987. Isolation and Antigenic Characterization of Human Immunodeficiency Virus (HIV) in Brazil. *Memórias do Instituto Oswaldo Cruz* 82, 453-456.

GAO, F.; MORRISON, S. G.; ROBERTSON, D. L.; THORNTON, C. L.; CRAIG, S.; KARLSSON, G.; SODROSKI, J.; MORGADO, M.; GALVÃO-CASTRO, B.; VON BRIESEN, H.; BEDDOWS, S.; WEBER, J.; SHARP, P. M.; SHAW, G. M.; HAHN, B. H. & THE WHO AND NIAID NETWORKS FOR HIV ISOLATION AND CHARACTERIZATION, 1996. Molecular Cloning and Analysis of Functional Envelope Genes From Human Immunodeficiency Virus Type 1 Sequences Subtypes A Through G. *Journal of Virology* 70, 1651-1667.

GAO, F.; ROBERTSON, D. L.; CARRUTHERS, C. D.; MORRISON, S. G.; JIAN, B.; CHEN, Y.; BARRÉ-SINOUSI, F.; GIRARD, M.; SRINIVASAN, A.; ABIMIKU, A. G.; SHAW, G. M.; SHARP, P. M. & HAHN, B. H., 1998. A Comprehensive Panel of Near-Full-Length Clones and Reference Sequences for Non-Subtype B Isolates of Human Immunodeficiency Virus Type 1. *Journal of Virology* 72, 5680-5698.

GAO, F.; YUE, L.; HILL, S. C.; ROBERTSON, D. L.; GRAVES, A. H.; SAAG, M. S.; SHAW, G. M.; SHARP, P. M. & HAHN, B. H., 1994. HIV-1 Sequence Subtype D in the United States. *AIDS Research and Human Retroviruses* 10, 625-627.

Gonzales, E., Dhanda, R., Bamshad, M., Mummidi, S., Geevarghese, R., Catano, G., Anderson, S. A., Walter, E. A., Stephan, K. T., Hammer, M. F., Mangano, A., Sen, L., Clark, R. A., Ahuja, S. S., Dolan, M. J. & Ahuja, S. K. Global survey of genetic variation in CCR5, RANTES, and MIP-1 α : Impact on the epidemiology of the HIV-1

pandemic. *Proceedings of the National Academy of Sciences of United States of America* 98(9): 5199-5204.

GUIMARÃES, M. L.; MOREIRA, A. S.; LOUREIRO, R.; GALVÃO-CASTRO, B., MORGADO, M. G. & THE BRAZILIAN NETWORK FOR HIV-1 ISOLATION AND CHARACTERIZATION, 2002. High Frequency of Recombinant Genomes in HIV Type 1 Samples From Brazilian Southeastern and Southern Regions. *AIDS Research and Human Retroviruses* 18, 1261-1269.

GÜRTLER, L.; HAUSER, P. H.; EBERLE, J.; VON BRUNN, A.; KNAPP, S.; ZEKENG, L.; TSAGUE, J. M. & KAPTUE, L., 1994. A New Subtype of Human Immunodeficiency Virus Type 1 (MVP-5180) From Cameroon. *Journal of Virology* 68, 1581-1585.

HALANI, N.; WANG, B.; GE, Y. C.; GHARPURE, H.; HIRA, S. & SAKSENA, N. K., 2001. Changing Epidemiology of HIV Type 1 Infection in India: Evidence of Subtype B Introduction in Bombay From a Common Source. *AIDS Research and Human Retroviruses* 17, 637-642.

HATTORI, T.; SHIOZAKI, K.; EDA, Y.; TOKIYOSHI, S.; MATSUSHITA, S.; INABA, H.; FUJIMAKI, M.; MEGURO, T.; YAMADA, K.; HONDA, M.; NISHIKAWA, K. & TAKATSUKI, K., 1991. Characteristics of the Principal Neutralizing Determinant of HIV-1 Prevalent in Japan. *AIDS Research and Human Retroviruses* 7, 825-830.

HIERHOLZER, J.; MONTANO, S.; HOELSCHER, M.; NEGRETE, M.; HIERHOLZER, M.; AVILA, M. M.; CARRILLO, M. G.; RUSSI, J. C.; VINOLES, J.; ALAVA, A.; ACOSTA, M. E.; GIANELLA, A.; ANDRADE, R.; SANCHEZ, J. L.; CARRION, G.; SANCHEZ, J. L.; RUSSELL, K.; ROBB, M. L.; BIRX, D.; MCCUTCHAN, F. E. & CARR, J. K., 2002. Molecular Epidemiology of HIV Type 1 in Ecuador, Peru, Bolivia, Uruguay, and Argentina. *AIDS Research and Human Retroviruses* 18, 1339-1350.

HIGGINS, D. G.; THOMPSON, J. D. & GIBSON, T. J., 1996. Using Clustal for Multiple Sequence Alignments. *Methods in Enzymology* 266, 383-402.

HOLGUÍN, A.; RODÉS, B. & SORIANO, V., 2000. Recombinant Human Immunodeficiency Viruses Type 1 Circulating in Spain. *AIDS Research and Human Retroviruses* 16, 505-511.

- HOLGUÍN, A.; RODÉS, B.; DIETRICH, U. & SORIANO, V., 1999. Human Immunodeficiency Viruses Type 1 Subtypes Circulating in Spain. *Journal of Medical Virology* 59, 189-193.
- HOLMES, E. C.; NEE, S.; RAMBAUT, A.; GARNETT, G. P. & HARVEY, P. H., 1995. Revealing the History of Infectious Disease Epidemics Through Phylogenetic Trees. *Philosophical Transactions of the Royal Society* 349, 33-40.
- HU, W.-S. & TEMIN, H. M., 1990. Retroviral Recombination and Reverse Transcription. *Science* 250, 1227-1233.
- ICHIKAWA, S.; ICHIKAWA, S.; KIHARA, M. & YAMAZAKI, S., 1997. Descriptive Epidemiology of HIV/AIDS in Japan, 1985-1994. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 14, S3-S12.
- JANINI, L. M.; TANURI, A.; SCHECHTER, M.; PERALTA, J. M.; VICENTE, A. C. P.; TORRE, N. D.; PIENIAZEK, N. J.; LUO, C.-C.; RAMOS, A.; SORIANO, V.; SCHOCHTMAN, G.; RAYFIELD, M. A. & PIENIAZEK, D., 1998. Horizontal and Vertical Transmission of Human Immunodeficiency Virus Type 1 Dual Infections Caused by Viruses of Subtypes B and C. *Journal of Infectious Diseases* 177, 227-231.
- JANSSENS, W.; BUVÉ, A. & NKENGASONG, J. N., 1997. The Puzzle of HIV-1 Subtypes in Africa. *AIDS* 11, 705-712.
- JAVAHERIAN, K.; LANGLOIS, A. J.; MCDANAL, C.; ROSS, K. L.; ECKLER, L. I.; JELLIS, C. L.; PROFY, A. T.; RUSCHE, J. R.; BOLOGNESI, D. P.; PUTNEY, S. D. & MATTHEWS, T. J., 1989. Principal Neutralizing Domain of the Human Immunodeficiency Virus Type 1 Envelope Protein. *Proceedings of the National Academy of Sciences of United States of America* 86, 6768-6772.
- KALEEBU, P.; ROSS, A.; MORGAN, D.; YIRRELL, D.; ORAM, J.; RUTEMBERWA, A.; LYAGOPA, F.; HAMILTON, L.; BIRYAHWAHO, B. & WHITWORTH, J., 2001. Relationship Between HIV-1 Env Subtypes A and D and Disease Progression in a Rural Ugandan Cohort. *AIDS* 15, 293-299.
- KANKI, P. J.; HAMEL, D. J.; SANKALÉ, J.-L.; HSIEH, C.-C.; THIOR, I.; BARIN, F.; WOODCOCK, S. A.; GUÈYE-NDIAYE, A.; ZHANG, E.; MONTANO, M.; NDOYE, I.; ESSEX, M. E. & MBOUP, S., 1999. Human Immunodeficiency Virus Type 1 Subtypes Differ in Disease Progression. *The Journal of Infectious Diseases* 179, 68-73.

KARITA, E.; NKENGASONG, J. N.; WILLEMS, B.; VANHAM, G.; FRANSEN, K.; HEYNDRIKX, L.; JANSSENS, W.; PIOT, P. & VAN DER GROEN, G., 1997. Macrophage-Tropism of HIV-1 Isolates of Different Genetic Subtypes. *AIDS* 11, 1303-1304.

KITAYAPORN, D.; VANICHSENI, S.; MASTRO, T. D.; RAKTHAM, S.; VANIYAPONGS, T.; DES JARLAIS, D. C.; WASI, C.; YOUNG, N.; SUJARITA, S.; HEYWARD, W. L. & ESPARZA, J., 1998. Infection With HIV-1 Subtypes B and E in Injecting Drug Users Screened for Enrollment Into a Prospective Cohort in Bangkok, Thailand. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 19, 289-295.

KITSUTANI, P. T.; NAGANAWA, S.; SHIINO, T.; MATSUDA, M.; HONDA, M.; YAMADA, K.; TAKI, M. & SUGIURA, W., 1998. HIV Type 1 Subtypes of Nonhemophiliac Patients in Japan. *AIDS Research and Human Retroviruses* 14, 1099-1103.

KUIKEN, C. L.; FOLEY, B.; HAHN, B.; KORBER, B.; MARX, P. A.; MCCUTCHAN, F.; MELLORS, J. W.; MULLINS, J. I.; SODROSKI, J. & WOLINKSY, S., 2000a. *HIV Sequence Compendium 2000*. Los Alamos National Laboratory, Los Alamos: Theoretical Biology and Biophysics Group.

KUIKEN, C.; LUKASHOV, V. V.; BAAN, E.; DEKKER, J.; LEUNISSEN, J. A. M. & GOUDSMIT, J., 1996. Evidence for Limited Within-Person Evolution of the V3 Domain of the HIV-1 Envelope in the Amsterdam Population. *AIDS* 10, 31-37.

KUIKEN, C.; THAKALLAPALLI, R.; ESKILD, A. & DE RONDE, A., 2000b. Genetic Analysis Reveals Epidemiologic Patterns in the Spread of Human Immunodeficiency Virus. *American Journal of Epidemiology* 152, 814-822.

KUSAGAWA, S.; SATO, H.; WATANABE, S.; NOHTOMI, K.; KATO, K.; SHIINO, T.; THWE, M.; OO, K. Y.; LWIN, S.; MRA, R.; KYWE, B.; YAMAZAKI, S. & TAKEBE, Y., 1998. Genetic and Serologic Characterization of HIV Type 1 Prevailing in Myanmar (Burma). *AIDS Research and Human Retroviruses* 14, 1379-1385.

KYPR, J. & MRÁZEK, J., 1987. Unusual Codon Usage of HIV. *Nature* 327, 20.

LASKY, M.; PERRET, J.-L.; PEETERS, M.; BIBOLLET-RUCHE, F.; LIEGEOIS, F.; PATREL, D.; MOLINIER, S.; GRAS, C. & DELAPORTE, E., 1997. Presence of

Multiple Non-B Subtypes and Divergent Subtype B Strains of HIV-1 in Individuals Infected After Overseas Deployment. *AIDS* 11, 43-51.

LEVY, J. A.; HOFFMAN, A. D.; KRAMER, S. M.; LANDIS, J. A. & SHIMABUKURO, J. M., 1984. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science* 225, 840-842

LI, W.-H., 1997. Estimating the Number of Nucleotide Substitutions Between Sequences. In: *Molecular Evolution* (Wen-Hsiung Li), pp. 79-99, Massachusetts: Editoria Sinauer Associates, Inc.

Libert, F., Cochaux, P., Beckman, G., Samson, M., Aksenova, M., Cao, A., Czeizel, A., Claustres, M., de la Rua, C., Ferrari, M., Ferrec, C., Glover, G., Grinde, B., Güran, S., Kucinskas, V., Lavinha, J., Mercier, B., Ogur, G., Peltonen, L., Rosatelli, C., Schwartz, M., Spitsyn, V., Timat, L., Beckman, L., Parmentier, M. & Vassart, G., 1998. The Δ ccr5 mutation conferring protection against HIV-1 in Caucasian population has a single and recent origin in Northeastern Europe. *Human Molecular Genetics* 7(3):399-406.

LIITSOLA, K.; HOLM, K.; BOBKOV, A.; POKROVSKY, V.; SMOLSKAJA, T.; LEINIKKI, P.; OSMANOV, S.; SALMINEN, M. & UNAIDS VIRUS ISOLATION NETWORK, 2000. An AB Recombinant and Its Parental HIV Type 1 Strains in the Area of the Former Soviet Union: Low Requirements for Sequence Identity in Recombination. *AIDS Research and Human Retroviruses* 16, 1047-1053.

LIITSOLA, K.; TASHKINOVA, I.; LAUKKANEN, T.; KOROVINA, G.; SMOLSKAJA, T.; MOMOT, O.; MASHKILLEYSON, N.; CHAPLINSKAS, S.; BRUMMER-KORVENKONTIO, H.; VANHATALO, J.; LEINIKKI, P. & SALMINEN, M., 1998. HIV-1 Genetic Subtype A/B Recombinant Strain Causing an Explosive Epidemic in Injecting Drug Users in Kaliningrad. *AIDS* 12, 1907-1919.

LIMPAKARNJANARAT, K.; UNGCHUSAK, K.; MASTRO, T. D.; YOUNG, N. L.; LIKHITYINGVARA, C.; SANGWONLOY, O.; WENIGER, B. G.; PAU, C.-P. & DONDERO, T. J., 1998. The Epidemiological Evolution of HIV-1 Subtypes B and E Among Heterosexuals and Injecting Drug Users in Thailand, 1992-1997. *AIDS* 12, 1108-1109.

LIPSITCH, M. & MOXON, E. R., 1997. Virulence and transmissibility of pathogens: what is the relationship? *Trends in Microbiology* 5(1), 31-37.

LouzadaJunior, P., Smith, A. G., Hansen, J. A. & Donadi, E. A., 2001. HLA-DRB1 and -DQB1 alleles in the Brazilian population of the northeastern region of the state of Sao Paulo. *Tissue Antigens* 57(2):158-162.

LUKASHOV, V. V.; CORNELISSEN, M. T. E.; GOUDSMIT, J.; PAPUASHVILLI, M. N.; RYTIK, P. H.; KHAITOV, R. M.; KARAMOV, E. V. & DE WOLF, F., 1995. Simultaneous Introduction of Distinct HIV-1 Subtypes into Different Risk Groups in Russia, Byelorussia and Lithuania. *AIDS* 9, 435-439.

LUKASHOV, V. V.; KARAMOV, E. V.; EREMIN, V. F.; TITOV, L. P. & GOUDSMIT, J., 1998. Extreme Founder Effect in an HIV Type 1 Subtype A Epidemic Among Drug Users in Svetlogorsk, Belarus. *AIDS Research and Human Retroviruses* 14, 1299-1303.

LUKASHOV, V. V.; KUIKEN, C.; VLAHOV, D.; COUTINHO, R. A. & GOUDSMIT, J., 1996. Evidence for HIV Type 1 Strains of U.S. Intravenous Drug Users As Founders of AIDS Epidemic Among Intravenous Drug Users in Northern Europe. *AIDS Research and Human Retroviruses* 12, 1179-1183.

MACHUCA, R.; BOGH, M.; SALMINEN, M.; GERSTOFT, J.; KVINESDAL, B.; PEDERSEN, C.; OBEL, N.; NIELSEN, H. & NIELSEN, C., 2001. HIV-1 Subtypes in Denmark. *Scandinavian Journal of Infectious Diseases* 33, 697-701.

MARKOWITZ, M.; LOUIE, M.; HURLEY, A.; SUN, E.; MASCIO, M. D.; PERELSON, A. S. & Ho, D. D. 2003. A Novel Antiviral Intervention Results in More Accurate Assessment of Human Immunodeficiency Virus Type 1 Replication Dynamics and T-Cell Decay In Vivo. *Journal of Virology* 77(8): 5037-5038.

MARTÍNEZ, M.-A., VARTANIAN, J.-P., WAIN-HOBSON, S., 1994. Hypermutagenesis of RNA using human immunodeficiency virus type 1 reverse transcriptase and biased dNTP concentrations *Proceedings of the National Academy of Sciences of United States of America* 91, 11787-11791.

MASTRO, T. D.; KUNANUSONT, C.; DONDERO, T. J. & WASI, C., 1997. Why Do HIV-1 Subtypes Segregate Among Persons With Different Risk Behaviors in South Africa and Thailand? *AIDS* 11, 113-116.

MAY, R. M.; GUPTA, S. & MCLEAN, A. R. 2001. Infectious disease dynamics: What characterizes a successful invader *Philosophical Transactions of the Royal Society* 356(1410), 901-910.

MAYER, V.; APETREI, C.; HABEKOVA, M.; TCHENTSOVA, N.; BRUCKOVA, M.; KLASKALA, W.; BAUM, M. K.; BRUN-VÉZINET, F. & SIMON, F., 1998. HIV-1 Diversity in Heterosexual Population in Slovakia, Ukraine and the Czech Republic. *AIDS* 12, 1106-1108.

MBOUDJEKA, I.; BIKANDOU, B.; ZEKENG, L.; TAKEHISA, J.; HARADA, Y.; YAMAGUCHI-KABATA, Y.; TANIGUCHI, Y.; IDO, E.; KAPTUE, L.; M'PELLE, P.; PARRA, H. J.; IKEDA, M.; HAYAMI, M. & MIURA, T., 1999. Genetic Diversity of HIV-1 Group M From Cameroon and Republic of Congo. *Archives of Virology* 144, 2291-2311.

MCDADE, J. E. & ANDERSON, B. E., 1996. Molecular Epidemiology: Applications of Nucleic Acid Amplification and Sequence Analysis. *Epidemiologic Reviews* 18, 90-97.

MEYERHANS, A. & VARTANIAN, J., 1999. The Fidelity of Cellular and Viral Polymerases and its Manipulation for Hypermutagenesis. In: *Origin and Evolution of Viruses* (E. Domingo, R. Webster & J. Holland, org), pp 87-114, Londres: Editora Academic Press.

MICHAEL-HENDRY, R.; HANSON, C. V.; BONGERTZ, V.; MORGADO, M. G.; DUARTE, A.; CASSEB, J.; BRIGIDO, L. F.; SABINO, E.; DIAZ, R. S. & GALVÃO-CASTRO, B., 1996. Immunoreactivity of Brazilian HIV Isolates With Different V3 Motifs. *Memórias do Instituto Oswaldo Cruz* 91, 347-348.

MODROW, S.; HAHN, B. H.; SHAW, G. M.; GALLO, R. C.; WONG-STAAAL, F. & WOLF, H., 1987. Computer-Assisted Analysis of Envelope Protein Sequences of Seven Human Immunodeficiency Virus Isolates: Prediction of Antigenic Epitopes in Conserved and Variable Regions. *Journal of Virology* 61, 570-578.

MORGADO, M. G.; GUIMARÃES, M. L.; GRIPP, C. B. G.; COSTA, C. I.; NEVES, JR, I.; VELOSO, V. G.; LINHARES-CARVALHO, M. I.; CASTELLO-BRANCO, L. R.; BASTOS, F. I.; KUIKEN, C.; CASTILHO, E. A.; GALVÃO-CASTRO, B.; BONGERTZ, V. & THE EVANDRO CHAGAS HOSPITAL AIDS CLINICAL RESEARCH GROUP, 1998a. Molecular Epidemiology of HIV-1 in Brazil: High

Prevalence of HIV-1 Subtype B and Identification of an HIV-1 Subtype D Infection in the City of Rio De Janeiro, Brazil. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 18 , 488-494.

MORGADO, M. G.; GUIMARÃES, M. L.; NEVES JR, I.; SANTOS, V. G. V.; CARVALHO, M. I. L.; CASTELLO-BRANCO, L. R.; BASTOS, F. I.; CASTILHO, E. A.; GALVÃO-CASTRO, B.; BONGERTZ, V. & THE EVANDRO CHAGAS HOSPITAL AIDS CLINICAL RESEARCH GROUP, 1998b. Molecular Epidemiology of HIV in Brazil: Polymorphism of the Antigenically Distinct HIV-1 B Subtype Strains. *Memórias do Instituto Oswaldo Cruz* 93, 383-386.

MORGADO, M. G.; SABINO, E. C.; SHPAER, E. G.; BONGERTZ, V.; BRIGIDO, L. F.; GUIMARAES, M. D. C.; CASTILHO, E. A.; GALVÃO-CASTRO, B.; MULLINS, J. I.; MICHAEL-HENDRY, R. & MAYER, A., 1994. V3 Region Polymorphisms in HIV-1 From Brazil: Prevalence of Subtype B Strains Divergent From North American/European Prototype and Detection of Subtype F. *AIDS Research and Human Retroviruses* 10, 569-575.

MS (Ministério da Saúde), 1995. *AIDS. Boletim Epidemiológico*, 8.

MS (Ministério da Saúde), 2002. *AIDS. Boletim Epidemiológico*, 16(1).

MSELLATI, P.; DUPON, M.; MORLAT, P.; LACOSTE, D.; PELLEGRIN, J.-C.; DABIS, F. & THE GOUPE D'EPIDÉMIOLOGIE CLINIQUE DU SIDA EN AQUITAINE, 1990. A Cohort Study of 89 HIV-1-Infected Adult Patients Contaminated by Blood Products: Bordeaux 1981-1989. *AIDS* 4, 1105-1109.

NAGANAWA, S.; AYUTTAYA P. I. N.; DUANGCHANDA, S.; AUWANIT, W.; WARACHIT, P.; MIYAMURA, K.; YAMAZAKI, S. & HONDA, M., 1997. A Characteristic Change of Consensus Core Motif in the V3 Region of HIV Type 1 Clade B, but Not in Clade E, in Thailand. *AIDS Research and Human Retroviruses* 13, 271-273.

NEI, M. & KUMAR, S., 2000. *Molecular Evolution and Phylogenetics*. New York: Editora Oxford University Press.

NEI, M. & KUMAR, S., 2000a. Evolutionary Change in DNA Sequences. In: *Molecular Evolution and Phylogenetics* (Masatoshi Nei & Sudhir Kumar). Pp 33-50, New York: Editora Oxford University Press.

NEI, M. & KUMAR, S., 2000b. Phylogenetic Inference: Distance Methods. In: *Molecular Evolution and Phylogenetics* (Masatoshi Nei & Sudhir Kumar). Pp 87-113. New York: Editora Oxford University Press.

NEI, M. & KUMAR, S., 2000c. Accuracies and Statistical Testes of Phylogenetic Trees In: *Molecular Evolution and Phylogenetics* (Masatoshi Nei & Sudhir Kumar). Pp 165-186, New York: Editora Oxford University Press.

NEILSON, J. R.; JOHN, G. C.; CARR, J. K.; LEWIS, P.; KREISS, J. K.; JACKSON, S.; NDUATI, R. W.; MBORI-NGACHA, D.; PANTELEEFF, D. D.; BODRUG, S.; GIACHETTI, C.; BOTT, M. A.; RICHARDSON, B. A.; BWAYO, J.; NDINYA-ACHOLA, J. & OVERBAUGH, J., 1999. Subtypes of Human Immunodeficiency Virus Type 1 and Disease Stage Among Women in Nairobi, Kenya. *Journal of Virology* 73, 4393-4403.

NOVITSKY, V.; SMITH, U. R.; GILBERT, P.; MCLANE, M. F.; CHIGWEDERE, P.; WILLIAMSON, C.; NDUNG'U, T.; KLEIN, I.; CHANG, S. Y.; PETER, T.; THIOR, I.; FOLEY, B. T.; GAOLEKWE, S.; RYBAK, N.; GASEITSIWE, S.; VANNBERG, F.; MARLINK, R.; LEE, T. H. & ESSEX, M., 2002. Human Immunodeficiency Virus Type 1 Subtype C Molecular Phylogeny: Consensus Sequence for an AIDS Vaccine Design? *Journal of Virology* 76, 5435-5451.

OP DE COUL, E. L. M.; COUTINHO, R. A.; VAN DER SCHOOT, A.; VAN DOORNUM, G. J. J.; LUKASHOV, V. V.; GOUDSMIT, J.; CORNELISSEN, M. T. E & DUTCH HIV-1 SUBTYPE SURVEILLANCE, 2001B. The Impact of Immigration on the Env HIV-1 Subtype Distribution Among Heterosexuals in the Netherlands: Influx of Subtype B and Non-B Strains. *AIDS* 15, 2277-2286.

OP DE COUL, E. L. M.; PRINS, M.; CORNELISSEN, M. T. E.; VAN DER SCHOOT, A.; BOUFASSA, F.; BRETTLE, R. P.; AGUADO, I. H.; SCHIFFER, V.; MCMENAMIN, J.; REZZA, G.; ROBERTSON, R.; ZANGERLE, R.; GOUDSMIT, J.; COUTINHO, R. A.; LUKASHOV, V. V. & THE EUROPEAN AND ITALIAN SEROCONVERTER STUDIES, 2001a. Using Phylogenetic Analysis to Trace HIV-1 Migration Among Western European Injecting Drug Users Seroconverting From 1984 to 1997. *AIDS* 15, 257-266.

OSMANOV, S.; PATTOU, C.; WALTER, N.; SCHWARTÄNDER, B.; ESPARZA, J. & THE WHO AND NIAID NETWORKS FOR HIV ISOLATION AND

CHARACTERIZATION, 2002. Estimated Global Distribution and Regional Spread of HIV-1 Genetic Subtypes in the Year 2000. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 29, 184-190.

OVERBAUGH, J. & BANGHAM, C. R. M., 2001. Selection Forces and Constraints on Retroviral Sequence Variation. *Science* 292, 1106-1109.

PALADIN, F. J. E.; MONZON, O. T.; TSUCHIE, H.; APLASCA, M. R. A.; LEARN, G. H. & KURIMURA, T., 1998. Genetic Subtypes of HIV-1 in the Philippines. *AIDS* 12, 291-300.

PEETERS, M., 2000. Recombinant HIV sequences: their role in the global epidemic. In: *HIV Sequence Compendium 2000* (C. L. Kuiken; B. Foley; B. Hahn; B. Korber; P. A. Marx; F. McCutchan; J. M. Mellors; J. I. Mullins; J. Sodroski & S. Wolinsky), Los Alamos National Laboratory, Los Alamos: Theoretical Biology and Biophysics Group.

PING, L.-H.; NELSON, J. A. E.; HOFFMAN, I. F.; SCHOCK, J.; LAMERS, S. L.; GOODMAN, M.; VERNAZZA, P.; KAZEMBE, P.; MAIDA, M.; ZIMBA, D.; GOODENOW, M. M.; ERON JR, J. J.; FISCUS, S. A.; COHEN, M. S. & SWANSTROM, R., 1999. Characterization of V3 Sequence Heterogeneity in Subtype C Human Immunodeficiency Virus Type 1 Isolates From Malawi: Underrepresentation of X4 Variants. *Journal of Virology* 73, 6271-6281.

PINTO, M. E.; TANURI, A. & SCHECHTER, M., 1998. Molecular and Epidemiologic Evidence for the Discontinuous Introduction of Subtypes B and F Into Rio De Janeiro, Brazil. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 19, 310-312.

POPE, M.; HO, D. D.; MOORE, J. P.; WEBER, J. & DITTMAR, M. T., 1997. Different Subtypes of HIV-1 and Cutaneous Dendritic Cells. *Science* 278, 786-788.

POPOVIC, M., SARNGADHARAN, M. G., READ, E. & GALLO, R. C., 1984. Detection, Isolation, and Continuous Production of Cytopathic Retroviruses (HTLV-III) From Patients With AIDS and Pre-AIDS. *Science* 224, 497-500.

POTTS, K. E.; KALISH, M. L.; LOTT, T.; ORLOFF, G.; LUO, C.-C.; BERNARD, M.-A.; ALVES, C. B.; BADARO, R.; SULEIMAN, J.; FERREIRA, O.; SCHOCHTMAN, G.; JOHNSON, W. D.; OU, C.-Y.; HO, J. L. & THE BRAZILIAN COLLABORATIVE

AIDS RESEARCH GROUP, 1993. Genetic Heterogeneity of the V3 Region of the HIV-1 Envelope Glycoprotein in Brazil *AIDS* 7, 1191-1197.

PUTNEY, S. D.; MATTHEWS, T. J.; ROBEY, W. G.; LYNN, D. L.; ROBERT-GUROFF, M.; MUELLER, W. T.; LANGLOIS, A. J.; GHRAYEB, J.; PETTEWAY JR, S. R.; WEINHOLD, K. J.; FISCHINGER, P. J.; WONG-STAAAL, F.; GALLO, R. C. & BOLOGNESI, D. P., 1986. HTLV-III/LAV-Neutralizing Antibodies to an E. Coli-Produced Fragment of the Virus Envelope. *Science* 234, 1392-1395.

PYBUS, O. G.; RAMBAUT, A. & HARVEY, P. H., 2000. An Integrated Framework for the Inference of Viral Population History From Reconstructed Genealogies. *Genetics* 155, 1429-1437.

QUIÑONES-MATEU, M. E. & ARTS, E. J., 2001. Fitness: Implications for Drug Resistance, Disease Progression, and Global Epidemic Evolution. In: *HIV-1 Sequence Compendium 2001*. Los Alamos National Laboratory, Los Alamos: Theoretical Biology and Biophysics Group.

RAMOS, A.; TANURI, A.; SCHECHTER, M.; RAYFIELD, M. A.; HU, D. J.; CABRAL, M. C.; BANDEA, C. I.; BAGGS, J. & PIENAZEK, D., 1999. Dual and Recombinant Infections: An Integral Part of the HIV-1 Epidemic in Brazil. *Emerging Infectious Diseases* 5(1).

REEVES, J. D. & DOMS, R. W., 2002. Human immunodeficiency virus type 2. *Journal of General Virology* 83, 1253-1265.

REINIS, M.; BRUCKOVÁ, M.; GRAHAM, R. R.; VANDASOVÁ, J.; STANKOVÁ, M. & CARR, J. K., 2002. Genetic Subtypes of HIV Type 1 Viruses Circulating in the Czech Republic. *AIDS Research and Human Retroviruses* 17, 1305-1310.

RIVERA-MORALES, L. G.; NOVITSKY, V. A.; TRUJILLO, R.; VALLE-MONTALVO, C. L.; CANO-DOMINGUEZ, C.; RAMOS-JIMENEZ, J.; JIMENEZ-RIOS, E.; FLORES-FLORES, L.; LOPEZ-GUILLEN, P.; GILBERT, P.; VANNBERG, F.; TAMEZ-GUERRA, R.; RODRIGUEZ-PADILLA, C. & ESSEX, M., 2001. The Molecular Epidemiology of HIV Type 1 of Men in Mexico. *AIDS Research and Human Retroviruses* 17, 87-92.

ROGER, M., 1998. Influence of host genes on HIV-1 disease progression. *The FASEB Journal* 12, 625-632.

RUSCHE, J. R.; JAVAHERIAN, K.; MCDANAL, C.; PETRO, J.; LYNN, D. L.; GRIMAILA, R.; LANGLOIS, A.; GALLO, R. C.; ARTHUR, I. O.; FISCHNGER, P. J.; BOLOGNESI, D. P.; PUTNEY, S. D. & MATTHEWS, T. J., 1988. Antibodies That Inhibit Fusion of Human Immunodeficiency Virus-Infected Cells Bind a 24-Amino Acid Sequence of the Viral Envelope, Gp120. *Proceedings of the National Academy of Sciences of United States of America* 85, 3198-3202.

RUSSEL, K. L.; CARCAMO, C.; M.WATTS, D.; SANCHEZ, J.; GOTUZZO, E.; EULER, A.; BLANCO, J. C.; GALEANO, A.; ALAVA, A.; MULLINS, J. I.; HOLMES, K. K. & CARR, J. K., 2000. Emerging Genetic Diversity of HIV-1 in South America. *AIDS* 14, 1785-1791.

RUSSO, C. A. M., 1997. Efficiencies of Different Statistical Testes in Supporting a Known Vertebrate Phylogeny. *Molecular Biology and Evolution* 14, 1078-1080.

SABINO, E. C.; DIAZ, R. S.; BRIGIDO, L. F.; LEARN, G. H.; MULLINS, J. I.; REINGOLD, A. L.; DUARTE, A. J. S.; MAYER, A.; BUSCH, M. P., 1996. Distribution of HIV-1 Subtypes Seen in an AIDS Clinic in Sao Paulo City, Brazil. *AIDS* 10, 1579-1584.

SABINO, E. C.; SHPAER, E. G.; MORGADO, M. G.; KORBER, B.; DIAZ, R. S.; BONGERTZ, V.; CAVALCANTE, S.; GALVÃO-CASTRO, B.; MULLINS, J. I. & MAYER, A., 1994. 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* 68, 6340-6346.

SALA, M. & WAIN-HOBSON, S., 1999. Drift and Conservatism in RNA Virus Evolution: Are They Adapting or Merely Changing? In: *Origin and Evolution of Viruses* (E. Domingo, R. Webster & J. Holland, org), pp 115-140, Londres: Editora Academic Press.

SANTORO-LOPES, G.; HARRISON, L. H.; TAVARES, M. D.; XEXÉO, A.; DOS SANTOS, A. C. E. & SCHECHTER, M., 2000. HIV Disease Progression and V3 Serotypes in Brazil: Is B Different From B-Br? *AIDS Research and Human Retroviruses* 16, 953-958.

SCHWARTZ, R. L. & PHOENIX, T., 2001. *Learning Perl, 3rd Edition*. Sebastoplo, CA: Editora O' Reilly & Associates, Inc

SIMON, F.; MAUCLÈRE, P.; ROQUES, P.; LOUSSERT-AJAKA, I.; MÜLLER-TRUTWIN, M. C.; SARAGOSTI, S.; GEORGES-COURBOT, M. C.; BARRÉ-SINOUSI, F. & BRUN-VÉZINET, F., 1998. Identification of a New Human Immunodeficiency Virus Type 1 Distinct From Group M and Group O. *Nature Medicine* 4, 1032-1037.

SOTO-RAMIREZ, L. E.; RENJIFO, B.; MCLANE, M. F.; MARLINK, R.; O'HARA, C.; SUTTHENT, R.; WASI, C.; VITHAYASAI, P.; VITHAYASAI, V.; APICHARTPIYAKUL, C.; AUEWARAKUL, P.; CRUZ, V. P.; CHUI, D.-S.; OSATHANONDH, R.; MAYER, K.; LEE, T.-H. & ESSEX, M., 1996. HIV-1 Langerhans' Cell Tropism Associated With Heterosexual Transmission of HIV. *Science* 271, 1291-1293.

STARCICH, B. R.; HAHN, B. H.; SHAW, G. M.; MCNEELY, P. D.; MODROW, S.; WOLF, H.; PARKS, E. S.; PARKS, W. P.; JOSEPHS, S. F. & GALLO, R. C., 1986. Identification and Characterization of Conserved and Variable Regions in the Envelope Gene of HTLV-III/LAV, the Retrovirus of AIDS. *Cell* 45, 637-48.

STEFASI, M. M. A.; PEREIRA, G. A. S.; MARTELLI, C. M. T.; SHINDO, N. & GALVÃO-CASTRO, B., 2000. Evidence of HIV-1 Genetic Diversity Among Pregnant Women With AIDS or Infected With HIV-1 in Central Brazil. *Journal of Acquired Immune Deficiency Syndrome and Human Retrovirology* 23, 205-207.

STOECKLI, T. C.; STEFFEN-KLOPFSTEIN, I.; ERB, P.; BROWN, T.; KALISH, M. L. & THE SWISS HIV COHORT STUDY, 2000. Molecular Epidemiology of HIV-1 in Switzerland: Evidence for a Silent Mutation in the C2V3 Region Distinguishing Intravenous Drug Users From Homosexual Men. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 23, 58-67.

SUBBARAO, S.; VANICHSENI, S.; HU, D. J.; KITAYAPORN, D.; CHOOPANYA, K.; RAKTHAM, S.; YOUNG, N.; WASI, C.; SUTTHENT, R.; LUO, C.-C.; RAMOS, A. & MASTRO, T. D., 2000. Genetic Characterization of Incident HIV Type 1 Subtype E and B Strains From a Prospective Cohort of Injecting Drug Users in Bangkok, Thailand. *AIDS Research and Human Retroviruses* 16, 699-707.

TAMURA, K., 1992. Estimation of the Number of Nucleotide Substitutions When There Are Strong Transition-Transversion and G+C-Content Biases. *Molecular Biology and Evolution* 9, 678-687.

TANURI, A.; SWANSON, P.; DEVARE, S.; BERRO, O. J.; SAVEDRA, A.; COSTA, L. J.; TELLES, J. G.; BRINDEIRO, R.; SCHABLE, C.; PIENAZEK, D. & RAYFIELD, M. A., 1999. HIV-1 Subtypes Among Blood Donors From Rio De Janeiro, Brazil. *Journal of Acquired Immune Deficiency Syndromes* 20, 60-66.

TELLES, P. R.; BASTOS, F. I.; GUYDISH, J.; INCIARDI, J. A.; SURRATT, H. L.; PEARL, M. & HEARST, N., 1997. Risk Behavior and HIV Seroprevalence Among Injecting Drug Users in Rio De Janeiro, Brazil. *AIDS* 11, S35-S42.

THOMSON, M. M.; DELGADO, E.; HERRERO, I.; VILLAHERMOSA, M. L.; PARGA, E. V.; CUEVAS, M. T.; CARMONA, R.; MEDRANO, L.; PÉREZ-ÁLVAREZ, L.; CUERVAS, L. & NÁJERA, R., 2002. 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* 83, 107-119.

THOMSON, M.; VILLAHERMOSA, M. L.; PARGA, E. V.; CUEVAS, M. T.; DELGADO, E.; MANJÓN, N.; MEDRANO, L.; PÉREZ-ÁLVAREZ, L.; CONTRERAS, G.; CARRILLO, M. G.; SALOMON, H. & NÁJERA, R., 2000. Widespread Circulation of a B/F Intersubtype Recombinant Form Among HIV-1-Infected Individuals in Buenos Aires, Argentina. *AIDS* 14, 897.

TIAN, H.; LAN, C. & CHEN, Y.-H., 2002. Sequence Variation and Consensus Sequence of V3 Loop on HIV-1 Gp120. *Immunology Letters* 83, 231-233.

TREURNICHT, F. K.; SMITH, T.-L.; ENGELBRECHT, S.; CLAASSEN, M.; ROBSON, B. A.; ZEIER, M.; VAN RENSBURG, E. J., 2002. Genotypic and Phenotypic Analysis of the Env Gene From South African HIV-1 Subtype B and C Isolates. *Journal of Medical Virology* 68, 141-146.

TRIQUES, K.; BOURGEOIS, A.; VIDAL, N.; MPOUDI-NGOLLÉ, E.; MULANGA-KABEYA, C.; NZILAMBI, N.; TORIMIRO, N.; SAMAN, E.; DELAPORTE, E. & PEETERS, M., 2000. Near-Full-Length Genome Sequencing of Divergent African HIV Type 1 Subtype F Viruses Leads to the Identification of a New HIV Type 1 Subtype Designated K. *AIDS Research and Human Retroviruses* 1602, 151.

VAN HARMELEN, J.; WOOD, R.; LAMBRICK, M.; RYBICKI, E. B.; WILLIAMSON, A.-L. & WILLIAMSON, C., 1997. An Association Between HIV-1 Subtypes and Mode of Transmission in Cape Town, South Africa. *AIDS* 11, 81-87.

VAN RIJ, R. P.; BLAAK, H.; VISSER, J. A.; BROUWER, M.; RIENTSMA, R.; BROERSEN, S.; HUSMAN, A.-M. R. & SCHUITEMAKER, H., 2000. Differential Coreceptor Expression Allows for Independent Evolution of Non-Syncytium-Inducing and Syncytium-Inducing HIV-1. *The Journal of Clinical Investigation* 106, 1039-1052.

VARTANIAN, J.-P., MEYERHANS, A., SALA, M., WAIN-HOBSON, S., 1994. GA hypermutation of the human immunodeficiency virus type 1 genome: Evidence for dCTP pool imbalance during reverse transcription *Proceedings of the National Academy of Sciences of United States of America* 91, 3092-3096.

VELARDE-DUNOIS, K. G.; GUIMARÃES, M. L.; FUENTE, C. L. A.; ANDRADE, R.; ARÉVALO, R.; PANTOJA, S.; MARISCAL, R.; SANDOVAL, R.; IRIARTE, F.; CHAMÓN, V.; MELGAR, M. L.; CARVAJAL, R. & MORGADO, M. G., 2000. Molecular Characterization of Human Immunodeficiency Virus Type 1-Infected Individuals From Bolivia Reveals the Presence of Two Distinct Genetic Subtypes B and F. *AIDS Research and Human Retroviruses* 16, 1921-1926.

VICENTE, A. C. P.; OTSUKI, K.; SILVA, N. B.; CASTILHO, M. C.; BARROS, F. S.; PIENAZEK, D.; HU, D. J.; RAYFIELD, M. A.; BRETAS, G. & TANURI, A., 2000. The HIV Epidemics in the Amazon Basin Is Driven by Prototypic and Recombinant HIV-1 Subtypes B and F. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 23, 327-331.

VIDAL, N.; PEETERS, M.; MULANGA-KABEYA, C.; NZILAMBI, N.; ROBERTSON, D.; ILUNGA, W.; SEMA, W.; TSHIMANGA, K.; BONGO, B. & DELAPORTE, E., 2000. Unprecedented Degree of Human Immunodeficiency Virus Type 1 (HIV-1) Group M Genetic Diversity in the Democratic Republic of Congo Suggests That the HIV-1 Pandemic Originated in Central Africa. *Journal of Virology* 74, 10498-10507.

VOEVODIN, A.; CRANDALL, K. A.; SETH, P. & MUFTI, S. A., 1996. HIV Type 1 Subtypes B and C From New Regions of India and Indian and Ethiopian Expatriates in Kuwait. *AIDS Research and Human Retroviruses* 12, 641-643.

WEIDLE, P. J.; GANEA, C. E.; IRWIN, K. L.; PIENAZEK, D.; MCGOWAN, J. P.; OLIVO, N.; RAMOS, A.; SCHABLE, C. A.; LAL, R. B.; HOLMBERG, S. D. & ERNST, J., 2000. Presence of Human Immunodeficiency Virus (HIV) Type 1, Group

M. Non-B Subtypes, Bronx, New York: A Sentinel Site for Monitoring HIV Genetic Diversity in the United States. *The Journal of Infectious Diseases* 181, 470-475.

WHO NETWORK FOR HIV ISOLATION AND CHARACTERIZATION, 1994. HIV Type 1 Variation in World Health Organization-Sponsored Vaccine Evaluation Sites: Genetic Screening, Sequence Analysis, and Preliminary Biological Characterization of Selected Viral Strains. *AIDS Research and Human Retroviruses* 10, 1327-1343.

WILLEY, R. L.; RUTLEDGE, R. A.; DIAS, S.; FOLKS, T.; THEODORE, T.; BUCKLER, C. E. & MARTIN, M. A., 1986. Identification of Conserved and Divergent Domains Within the Envelope Gene of the Acquired Immunodeficiency Syndrome Retrovirus. *Proceedings of the National Academy of Sciences of United States of America* 83, 5038-5042.

WORKSHOP REPORT FROM THE EUROPEAN COMMISSION DG XII, 1997. HIV-1 Subtypes: Implications for Epidemiology, Pathogenicity, Vaccines and Diagnostics. *AIDS* 11, UNAIDS17-UNAIDS36.

YIRRELL, D. Y.; KALEEBU, P.; MORGAN, D.; WATERA, C.; MAGAMBO, B.; LYAGOBA, F. & WHITWORTH, J., 2002. Inter- and Intra-Genetic Intersubtype HIV-1 Recombination in Rural and Semi-Urban Uganda. *AIDS* 16, 279-286.

ZHANG, L.; RAMRATNAM, B.; TENNER-RACZ, K.; HE, Y.; VESANEN, M.; LEWIN, S.; TALAL, A.; RACZ, P.; PERELSON, A. S.; KORBER, B.; MARKOWITZ, M. & HO, D. D., 1999. Quantifying Residual HIV-1 Replication in Patients Receiving Combination Antiretroviral Therapy. *The New England Journal of Medicine* 340, 1605-1613.

ZHU, T.; KORBER, B.; NAHMIAS, A. J.; HOOPER, E.; SHARP, P. M.; & HO, D. D., 1998. An African HIV-1 Sequence From 1959 and Implications for the Origin of the Epidemic. *Nature* 391, 594-597.