

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Instituto de Biociências
Programa de Pós-Graduação em Genética e Biologia Molecular

Disseminação do HIV-1 subtipo B nas Américas

Dennis Maletich Junqueira

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Mestre.

Orientadora: Dr.^a Patrícia Ashton Prolla

Co-orientadora: Dr.^a Sabrina Esteves de Matos Almeida

Porto Alegre, março de 2011

.....
Este trabalho foi parcialmente financiado e desenvolvido nos laboratórios do Centro de Desenvolvimento Científico e Tecnológico (CDCT), na Fundação Estadual de Produção e Pesquisa em Saúde (FEPPS), contando com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

.....

“Somos arquivos ambulantes de
uma sabedoria ancestral.”

Helena Cronin

AGRADECIMENTOS:

.....

Agradeço a todos aqueles que, de alguma forma, ajudaram na minha formação acadêmica e que compreenderam o quão importante toda essa jornada de aprendizado foi para mim. Todas as discussões, as formulações, as leituras, as noites sem dormir, o nervosismo e a insegurança fizeram parte do processo e valeram à pena! O meu obrigado à Ana, Dadá e Paulo, que com muita raiva (rabdovirus) me abriram os olhos para o mundo da virologia. À Rubiazinha, meu exemplo de dedicação, que me inseriu nas pesquisas com HIV e, além de tudo, ajudou, ensinou, ouviu, aconselhou e me fez crescer não só profissionalmente, mas também como pessoa. À Marish, por apoiar, escutar e discutir com naturalidade as minhas diárias idéias de projetos sem fundamentos. Obrigado à minha grande chefinha, Sabrina, por formar, acreditar, defender e sempre confiar no nosso grupo de pesquisa. A gente ainda vai longe! Obrigado à Lúcia e a todos os pesquisadores que fazem do CDCT um lugar tão especial, em particular ao Leo, à Regina e à Raka. Recentemente, agradeço à Patrícia por embarcar nessa jornada e encarar tudo com muita sensibilidade, profissionalismo e atenção.

Os caminhos vão se definindo, as coisas vão acontecendo e muitas pessoas acabam entrando e participando de tudo isso, às vezes não diretamente, mas não menos fundamentais. Agradeço a Silma e ao Tiago, meus amigos de infância. À Lu Matrix por toda amizade, apoio, carinho e incentivo. Aos meus amigos de faculdade Ju, Mauro, Gabi e Jorge, especialmente ao Bruninho e ao Rodrigo Sam, meus fiéis e sábios escudeiros. Agradeço à Lu e especialmente à Vanessinha, que esteve ali desde o começo, torcendo e acompanhando cada passo. Agradeço ao Marcelo, uma peça-chave, por toda paciência, fortaleza, atenção e carinho. Às minhas irmãs, fundamentais em tudo, que sempre entenderam e fizeram de tudo para ajudar, apoiar e estão sempre ali para o que der e vier. Ao meu sobrinho, meu exemplo de ser-humano, por toda a honestidade e sabedoria. Finalmente agradeço profundamente aos meus pais por toda educação e cuidado e que, às vezes, mesmo sem entender o motivo de tanto estudo, tanto envolvimento e tanta dedicação estiveram sempre ao meu lado. Amo vocês!

SUMÁRIO

.....

06	ABREVIATURAS
07	RESUMO
08	ABSTRACT
09	CAPÍTULO 1: INTRODUÇÃO
10	1.1. AIDS
12	1.2. O vírus
16	1.3. Replicação
19	1.4. Surgimento
24	1.5. Diversidade
29	1.6. Subtipo B
32	1.7. Filogenias
37	1.8. ML
39	1.9. IB
43	CAPÍTULO 2: OBJETIVOS
44	CAPÍTULO 3: MANUSCRITO
45	Abstract
46	Introduction
47	Methods
49	Results
51	Discussion
56	Acknowledgements
56	References
59	Legends
60	Figures
62	Tables
63	Supporting Information
66	CAPÍTULO 4: DISCUSSÃO
70	REFERÊNCIAS

ABREVIATURAS

.....

- AIDS** Acquired Immunodeficiency Syndrome
(Síndrome da Imunodeficiência Adquirida)
- AMOVA** Analysis of Molecular Variation
(Análise de Variação Molecular)
- BSP** Bayesian Skyline Plot
- CRF** Circulating Recombinant Form
(Forma Recombinante Circulante)
- DNA** Desoxyribonucleic Acid
(Ácido desoxirribonucléico)
- HIV** Human Immunodeficiency Vírus
(Vírus da Imunodeficiência Humana)
- HIV-1B** Vírus da Imunodeficiência Humana
Tipo 1 grupo M subtipo B
- IB** Inferência Bayesiana
- MCMC** Markov Chain Monte Carlo
(Cadeia de Markov Monte Carlo)
- MDS** Multidimensional Scaling
- ML** Maximum Likelihood
(Máxima verossimilhança)
- RNA** Ribonucleic Acid
(Ácido ribonucléico)
- SIV** Simian Immunodeficiency Vírus
(Vírus da Imunodeficiência Símia)
- URF** Unique Recombinant Form
(Forma Única Circulante)

RESUMO:

.....

A infecção pelo vírus da imunodeficiência humana (HIV) foi largamente disseminada pela população humana a partir da década de 80. Diversos estudos retrocedem o surgimento da epidemia do HIV a vírus infectantes de símios africanos de vida livre e ao conseqüente contato entre estas espécies e humanos. As rápidas taxas evolutivas e a associação às inter-relações humanas foram suficientes para a disseminação e a diversificação das formas virais. Desta forma, a expansão do HIV-1 pelo mundo estabeleceu uma pandemia formada por diferentes variantes virais. O subtipo B foi a primeira forma detectada do vírus e atualmente está presente em um grande número de países, sendo responsável por aproximadamente 10% das infecções por HIV no mundo todo.

Fora da África, o subtipo B emergiu na região do Caribe, se disseminou pelos países vizinhos e atingiu os Estados Unidos, onde estabeleceu a pandemia. O traçado geográfico das rotas de disseminação nas Américas a partir do Haiti é, até o momento, difuso ou ausente. O presente trabalho visa investigar a história evolutiva do subtipo B nas Américas e, através de filogenias e análises estatísticas, examinar o papel da América do Sul no estabelecimento da epidemia.

Os resultados obtidos sugerem uma participação primária das populações da América do Sul na disseminação do subtipo B. A estruturação populacional e as filogenias corroboram uma ligação epidemiológica importante entre os países do Caribe e os da América do Sul, visto que um clado agrupando isolados sul-americanos emerge diretamente do clado caribenho. Além deste, um grande grupo, formado por sequências das três Américas, denominado cluster pandêmico, compartilha um ancestral comum com o clado da América do Sul. Possíveis explicações para estes resultados envolvem a co-dispersão para as Américas a partir do Caribe e também a intermediação da América do Sul entre o Caribe e os Estados Unidos no estabelecimento da pandemia. Dados históricos de migrações populacionais a partir do Caribe reforçam estes cenários.

Os dados apresentados revelam uma nova perspectiva a respeito da epidemia do HIV-1 subtipo B. Além disso, destacam a utilidade de uma abordagem populacional para o entendimento da dispersão da epidemia, contribuindo para a avaliação entre a diversidade viral e a movimentação de populações humanas.

ABSTRACT:

.....

Infections with human immunodeficiency virus (HIV) have been widely disseminated by the human population from the 1980's. Several studies point the onset of the epidemic to virus infecting free-living African apes through blood contact between these species and humans. The fast evolutionary rates and the human relationships were sufficient for the spreading and diversification of the epidemic. Thus, the spread of HIV-1 established a worldwide pandemic formed by different viral forms. Among them, subtype B was the first form detected and is currently present in a large number of countries, accounting for approximately 10% of HIV infections worldwide.

Out of Africa, subtype B likely emerged in Haiti in the Caribbean region, reached neighboring countries and the United States, where the pandemic was established. The geographical layout of the routes of spread within the Americas, so far, is diffuse or absent. This study aimed to investigate the evolutionary history of subtype B in the Americas and, through phylogenetic and statistical analyses sought to examine the role of South America in the pandemic.

Our main findings suggest a primary involvement of the populations of South America in the spread of subtype B. The genetic structure of the viral population and the phylogenetic analyses supported an epidemiological link among the countries of the Caribbean and South American ones since a clade grouping isolates from South America emerges directly from the Caribbean. Besides this, a large group, formed by sequences of the three Americas, called pandemic cluster, share a common ancestor with the clade of South America. The co-dispersion to the Americas from the Caribbean and the intermediation of South America between the Caribbean and North America epidemics are explanations to support our results.

The data presented here outline a new perspective on the HIV-1 subtype B epidemic. Furthermore, we highlight the usefulness of a population-based approach to understanding the spread of the epidemic, contributing to the assessment between the viral diversity and the movement of human populations.

Capítulo 1: Introdução

.....

Após três décadas da descoberta do vírus da imunodeficiência humana (HIV) e do seu desenfreado alastramento pelos diferentes grupos populacionais (Hemelaar et al., 2006), refletindo em mudanças nos contextos sociais, econômicos e culturais, o mundo se deparou recentemente com uma reversão da pandemia. Apesar de aproximadamente 33 milhões de pessoas conviverem com o vírus (Figura 1.1), a epidemia, desde 1999, já assumiu uma queda de 19% no número de novos casos. O número de mortes relacionadas à conseqüente Síndrome da Imunodeficiência Adquirida (do inglês *Acquired Immunodeficiency Syndrome* - AIDS), causada pelo vírus, não decresceu, mas se encontra em um patamar estável e, no entanto, a qualidade de vida dos pacientes foi beneficiada pela crescente disponibilidade e adesão à terapia antirretroviral (UNAIDS, 2010).

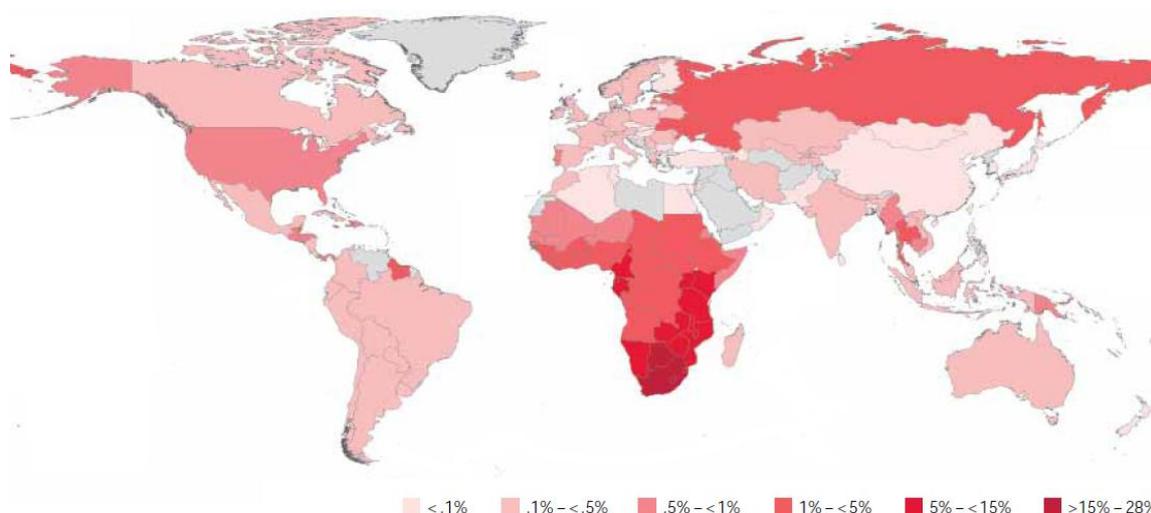


Figura 1.1: Prevalência global de indivíduos vivendo com HIV em 2009. Para as áreas em cinza não há dados disponíveis. Fonte: UNAIDS, 2010

Estudos de origem e disseminação do vírus também demonstram resultados promissores (Paraskevis et al., 2007; Salemi et al., 2008; Oliveira et al., 2010; Gray et al., 2009; Lemey et al., 2003; Guimarães et al., 2009). Ao mesmo tempo em que as altas taxas de mutação e a extrema diversidade genética são um obstáculo para o desenvolvimento de vacinas, drogas anti-vírus e para outras variáveis terapêuticas e prognósticas (Rambaut et al., 2004), a diversidade acumulada ao longo dos anos nos genomas virais possibilita a reconstrução histórica da expansão inicial da pandemia e o entendimento das rotas de disseminação das variedades virais atuais (Vidal et al., 2000; Plenum et al., 1999; Li et al., 1988). Felizmente, a resolução de questões-chave relacionadas à biologia evolutiva do vírus e a interação com o hospedeiro humano em seus processos de migração, sem dúvida alguma, contribuirá fundamentalmente para as áreas que visam beneficiar a vida dos pacientes, amenizar a evolução da doença e eliminar o vírus da população (Rambaut et al., 2004).

1.1. AIDS:

.....

A AIDS foi primeiramente reconhecida nos Estados Unidos em 1981, seguida de um crescimento na incidência de raras doenças oportunistas, como a pneumonia causada por *Pneumocystis carinii*, em homens que fazem sexo com homens que se tornaram imunodeficientes em decorrência da infecção viral (Gottlieb et al., 2006; Ks et al., 2002). A primeira evidência de que a síndrome era causada por um vírus surgiu em 1983, quando partículas virais foram isoladas de pacientes sintomáticos no Instituto Pasteur pela equipe do Dr. Luc Montaigner (Chermann et al., 2007). O vírus, nesse momento, foi denominado LAV, vírus associado à linfadenopatia (Barré-Sinoussi et al., 1983). Na mesma edição da publicação desses resultados, o grupo de Robert Gallo, nos Estados Unidos, relatou o isolamento de uma nova variante de HTLV (vírus T-linfotrófico humano), chamada de HTLV-III, também de pacientes sintomáticos para AIDS (Gallo et al., 1983). Posteriormente, ficou comprovado que o vírus descrito pelo grupo americano era o mesmo isolado pelo grupo francês (Marx et al., 1985). Em 1986, o Comitê Internacional de Taxonomia dos Vírus (ICTV) recomendou a denominação do agente etiológico da AIDS como vírus da imunodeficiência humana – HIV (J. Coffin et

al. 1986). No mesmo ano, a equipe de Luc Montagnier isola um segundo vírus também causador da síndrome da imunodeficiência, denominado de HIV-2 (Clavel et al., 1986).

Diversos fatores biológicos, denominados fatores de risco, são associados à transmissão do HIV; entre eles a concentração do vírus em fluidos biológicos, a integridade e vulnerabilidade da mucosa envolvida, a duração da exposição e a amostra viral transmitida. Os fatores comportamentais, como o número de parceiros sexuais, o uso de preservativos e o compartilhamento de seringas também são diretamente relacionados aos fatores biológicos. As principais formas de transmissão do HIV envolvem as vias sexual, sanguínea e vertical (parto ou aleitamento). Os casos de transmissão ocupacional são raros e giram em torno de apenas 0,3% dos casos de HIV (Schwartz e Nair, 1999; Barré-Sinoussi et al., 1983; Robinson et al., 2002; Sodora e Silvestri, 2008).

O HIV tem acesso ao organismo através das mucosas, mais usualmente, pelo contato sexual. Esta forma de transmissão ocorre pelo contato entre secreções contaminadas pelo vírus e a mucosa retal, genital ou oral. Diversas teorias foram propostas para explicar a dinâmica da infecção viral nas mucosas humanas, mas, até o momento, parece não haver um consenso. A regra geral para estas hipóteses, no entanto, está na infecção primária das células dendríticas, células do sistema imune posicionadas nas camadas abaixo da mucosa. Através destas células, o HIV teria a possibilidade de encontrar e infectar as diversas células do sistema imune, seu alvo celular principal (Hladik e McElrath, 2008; Levy et al., 1990).

A infecção pelo HIV resulta em uma doença crônica e progressiva (Sodora e Silvestri et al., 2008). O curso da infecção é marcado pelo aumento no nível da replicação viral e emergência de amostras mais virulentas. A AIDS é caracterizada por uma gradual depleção de células T CD4+ que, em última análise, levam a uma séria deficiência no estabelecimento de uma resposta imune eficaz contra patógenos (Sodora e Silvestri et al., 2008; Cloyd et al., 2000; Schwartz e Nair, 1999).

1.2. O VÍRUS:

.....

Os retrovírus são vírus envelopados de RNA que se caracterizam pela presença de uma enzima responsável pela retrotranscrição de RNA em DNA, denominada transcriptase reversa (Hoffman et al.,1985). Sete gêneros formam essa família de vírus, incluindo os *Lentivirus*. Este gênero possui características morfológicas e biológicas comuns e é caracteristicamente responsável por doenças crônicas com um longo período de incubação. Geneticamente, estão associados à presença de três genes que codificam proteínas fundamentais para seu ciclo replicativo na célula hospedeira: 5'-*gag-pol-env*-3' (Coffin et al.,1991).

O HIV é um retrovirus, membro do gênero lentivirus (Chang et al., 1993). Seu genoma é composto por duas fitas simples positivas não-complementares de RNA (Schwartz e Nair, 1999) que codificam nove genes encapsulados por, aproximadamente 2000 cópias de uma pequena proteína do capsídeo chamada p24. Pelo menos, sete estruturas secundárias importantes são detectadas no RNA viral para auxiliar a ligação de diversas proteínas (Hofacker et al.,1998). O genoma é ainda envolto por uma proteção lipoprotéica denominada envelope, formado pela membrana plasmática da célula hospedeira onde o vírus foi formado (Cantin et al., 2005). Entre o envelope e o capsídeo, espaço da matriz, pequenas proteínas preenchem e ajudam na sustentação da estrutura (Figura 1.2.1).

O genoma viral é formado por aproximadamente 10.000 nucleotídeos, organizados em nove regiões gênicas que codificam 19 proteínas (Muesing et al.,1985) (Figura 1.2.2). As proteínas codificadas são divididas em três classes: (i) as proteínas estruturais GAG, POL e ENV são responsáveis pela maquinaria de replicação e, certamente, pela estrutura dos vírions; (ii) as proteínas regulatórias TAT e REV são responsáveis pela transcrição de genes virais; e (iii) as proteínas acessórias Nef, Vif, Vpr and Vpu ou Vpx são responsáveis por diversas funções, incluindo o escape do sistema imune (Costin, 1988).

O gene *gag* codifica uma proteína precursora de 55 kilodaltons (p55) responsável pela formação do capsídeo (Bryant e Ratner, 1990). Durante o processo de

maturação viral, a proteína precursora p55 é clivada, pela protease viral, em proteínas menores designadas p17 (proteína da matriz), p24 (proteína formadora do capsídeo cônico), p9 (proteína do nucleocapsídeo) e p6 (proteína do nucleocapsídeo) (Göttlinger et al., 1989).

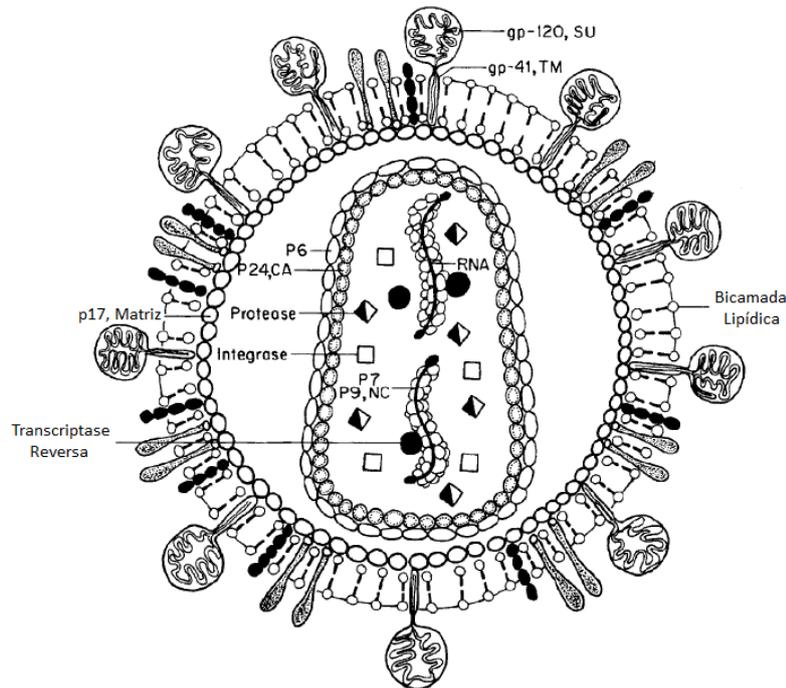


Figura 1.2.1: Estrutura do HIV. Adaptado de Schwartz e Nair (1999).

As proteínas do gene *pol* são sempre traduzidas juntamente com as do gene *gag* (Jacks et al., 1988). Essas proteínas são então produzidas através de uma mudança do quadro de leitura do mRNA. Apenas em 5% dos casos ocorre esta mudança GAG-POL (Parkin et al., 1992). A poliproteína precursora POL é subsequentemente clivada em quatro proteínas fundamentais para a replicação efetiva do vírus: protease (p10), transcriptase reversa (p50), RNase H e integrase (Parkin et al., 1992) (Figura 1.2.2). A proteína p10 produz uma aspartil-protease responsável pela clivagem dos precursores poliprotéicos durante a fase de maturação dos vírions (Ashorn et al., 1990). A transcriptase reversa tem função de polimerase RNA-dependente e DNA-dependente, fundamentais para o processo de transcrição reversa (Zack et al., 1990). A RNase H é responsável pela degradação do RNA viral após a transcrição reversa e, devido a falhas da protease na clivagem, está em 50% dos casos associada à transcriptase reversa. Por último, a integrase é a proteína responsável pela inserção do DNA retrotranscrito no genoma do hospedeiro (Bushman et al., 1990).

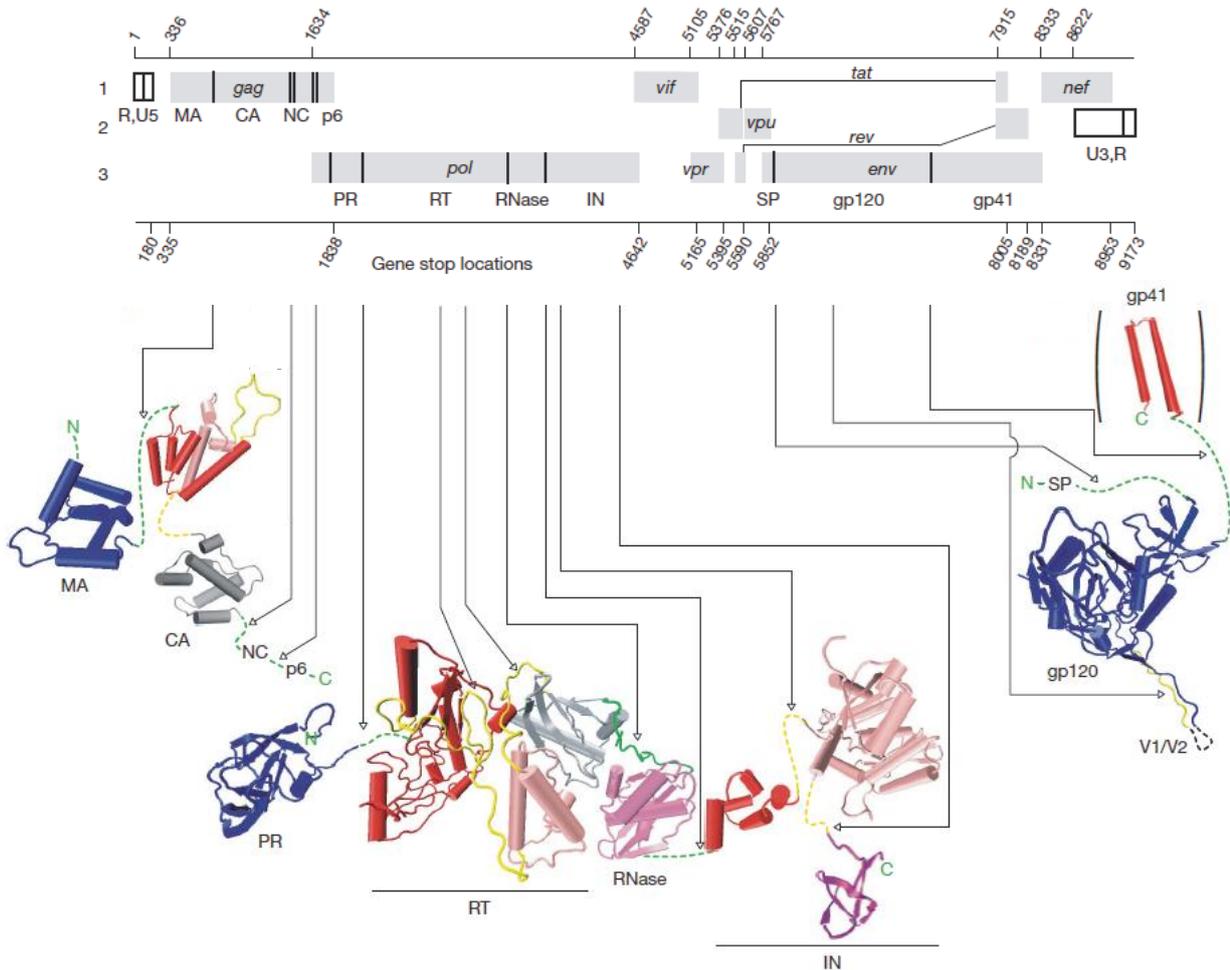


Figura 1.2.2: Organização genômica e relação à estrutura protéica do HIV. Adaptada de Watts et al. (2009).

O gene *env* codifica apenas uma proteína, a gp 160. Depois de sintetizada, esta poliproteína (p160) é levada ao complexo de Golgi para adição de carboidratos (gp 160) (Costin et al.,2007). Subseqüentemente, é clivada por proteases celulares, originando duas proteínas, gp 120 e gp 41 para HIV-1, ou gp 125 e gp 36 para HIV-2. A gp 41 se associa com a gp 120 e, formando trímeros, são levadas à membrana plasmática da célula hospedeira de maneira que a gp 41, proteína transmembrana, ancore a gp 120 voltada para o lado externo (Costin et al.,2007). Estas duas glicoproteínas em conjunto serão parte do envelope viral e têm papel fundamental na interação e fusão do vírus com a célula hospedeira para que se inicie o ciclo replicativo infeccioso (Chan et al.,1997).

A proteína regulatória TAT é um transativador transcricional essencial para a replicação do HIV (Ruben et al.,1989). Se liga, juntamente com proteínas da célula hospedeira, a uma região no RNA viral designada elemento de resposta a transativação,

localizada na região 5' do genoma viral. Age principalmente para promover o alongamento da transcrição do HIV (Southgate e Green, 1991).

A exportação nuclear dos mRNAs virais é criticamente dependente da atividade da proteína REV, um ligante de RNA importante para a transcrição efetiva dos genes do vírus. REV se liga a uma estrutura secundária do RNA viral e auxilia na exportação dos RNAs que constituirão o genoma dos novos vírus. Na ausência da proteína REV a maquinaria do spliceossomo da célula hospedeira atua nos transcritos virais e produz apenas os mRNAs das proteínas regulatórias do vírus mas não RNA genômico (Felber et al.,1990).

Em adição aos genes *gag*, *pol*, *env*, *tat* e *rev* contidos em todos os retrovírus, o HIV contém quatro genes adicionais denominados genes acessórios (Costin,1988). As proteínas codificadas por esses genes, apesar de acessórias, são fundamentais, pois representam fatores de virulência importantes para infecção das células. A proteína Nef é a primeira a ser detectada após a infecção em uma célula (Kim et al.,1989) e sua expressão parece ter diversas consequências, incluindo a influência na modulação da expressão de CD4 na superfície, a ativação deficiente das células de células T e a facilitação da infecção por HIV (Kim et al.,1989). A proteína Vif é incorporada aos vírions do HIV e ajuda na infectividade do vírus. Recentemente, foi descoberto que esta proteína é capaz de bloquear a função da proteína celular APOBEC-3G, proteína com função de desaminar o heteroduplex de DNA:RNA no citoplasma (Wiegand et al.,2004). Vpr também é uma proteína incorporada ao vírion e participa no direcionamento do complexo de pré-integração ao núcleo da célula infectada, sendo capaz de interagir com proteínas dos poros nucleares (Vodicka et al., 1998). Vpu é encontrado apenas no HIV-1 tendo a função de garantir o brotamento do vírus da membrana plasmática de célula hospedeira. No HIV-2, o gene codificante de Vpu é substituído pelo gene *vpx*, que é estruturalmente similar ao gene *vpr* e também apresenta funções semelhantes (Mahalingam et al., 2001).

1.3. REPLICAÇÃO:

.....

O termo tropismo viral se refere ao tipo de células alvo de um determinado vírus. Esta “preferência” é fundamentalmente determinada pela interação dos receptores virais com os receptores celulares localizados na face externa da membrana plasmática. O HIV é capaz de infectar diferentes tipos de células humanas, principalmente aquelas relacionadas com o sistema imune, interagindo com proteínas do tipo CD4. As moléculas de CD4 são glicoproteínas que desempenham função de amplificação de sinal ou sinalização e, desta forma, participam ativamente para efetividade da resposta imune. Excepcionalmente, o HIV tipo 2 apresenta linhagens que são capazes de infectar células humanas independente da presença de CD4 (Reeves et al., 1999).

O tropismo do HIV-1 por diferentes células humanas não é explicado apenas pela ligação às moléculas de CD4 (Berger et al., 1999). Além de receptores, é necessária a interação com outras moléculas para entrada do vírus. Os receptores de quimiocinas CXCR4 e CCR5, membros dos receptores acoplados à proteína G, foram identificados como principais co-receptores para infecção do HIV-1 (Berger et al., 1999). Algumas linhagens do vírus exibem alta eficiência na infecção contínua de células T CD4+ primárias, mas não de macrófagos primários ou monócitos, e geralmente induzem a formação de sincícios. Esta linhagem utiliza predominantemente o co-receptor CXCR4 para entrada na célula hospedeira (Schwartz e Nair, 1999; Coakley et al., 2005). Em contraste, outra linhagem do vírus infecta monócitos, macrófagos e células T primárias com maior eficiência do que células T transformadas, utilizando co-receptores do tipo CCR5 (Coakley et al., 2005). Estas linhagens foram respectivamente denominadas T-trópicas, ou indutoras de sincício, e M-trópicas, ou não indutoras de sincício (Berger et al., 1999). Estas diferenças no tropismo são provavelmente uma consequência das diferenças na sequência de nucleotídeos na região *V3 loop*, subunidade da gp 120, responsável pela interação com os co-receptores (Coakley et al., 2005; Schwartz e Nair, 1999).

Partículas virais que utilizam o CCR5 como co-receptor (R5) são predominantes no curso inicial da infecção pelo HIV-1 e, provavelmente, são a maioria significativa de vírus que são transmitidos de um indivíduo a outro (Brumme et al., 2005; Moyle et al.,

2005). Estes processos de seleção de partículas R5 não são bem compreendidos até o momento, mas acredita-se que na transmissão sexual do vírus os espermatozoides carreguem apenas vírus R5, pois eles possuem receptores CCR5 e não CXCR4 (Muciaccia et al., 2005). Além do mais, os macrófagos, células importantes para o início da infecção pelo HIV, são altamente susceptíveis ao vírus R5 (Schwartz e Nair, 1999). Durante o curso da infecção, ocorre uma adaptação viral para utilização de co-receptores CXCR4 ao invés de CCR5 (Koot et al., 1996). Linhagens que possuem tropismo para os dois co-receptores são, desta forma, consideradas uma forma viral de transição (Coakley et al., 2005).

O ciclo de infecção e replicação do HIV pode ser didaticamente dividido em seis fases: (i) interação e entrada; (ii) desmantelamento; (iii) transcrição reversa; (iv) integração do provírus (v) síntese de proteínas e montagem e (vi) brotamento (Fanales-Belasio et al., 2010) (Figura 1.3.1).

Inicialmente, o processo de infecção da célula hospedeira pelo HIV envolve a interação entre a gp120 e as moléculas receptoras e co-receptoras da célula, CD4 e CCR5 ou CXCR4. No momento do contato, célula CD4+ e vírus, ocorre a interação entre gp120 e CD4, levando a uma complexa mudança conformacional das proteínas do envelope viral a fim de expor o sítio específico de ligação ao co-receptor. A dupla ligação da gp120 em ambos receptores celulares permite uma interação estável com a membrana plasmática da célula, e por consequência levam a mudanças na porção extra-envelope da gp41. Esta segunda mudança aproxima o envelope da membrana plasmática e permite a fusão e consequente entrada do capsídeo viral na célula hospedeira (Schwartz e Nair, 1999; Fanales-Belasio et al., 2010).

Duas hipóteses sugerem diferentes momentos para o desmantelamento do capsídeo (Arhel et al., 2010). Na primeira, o capsídeo é desmantelado próximo à membrana, logo após a fusão com o envelope viral, e será necessário para a formação do complexo de transcrição reversa (Bukrinsky et al., 2004). O segundo modelo propõe que o capsídeo permanece intacto durante a fusão do vírus e será desmantelado ao longo do caminho até o núcleo, enquanto ocorre a transcrição reversa (Warrillow et al., 2009). Neste modelo, o desmantelamento ocorre em resposta a mudanças no ambiente celular.

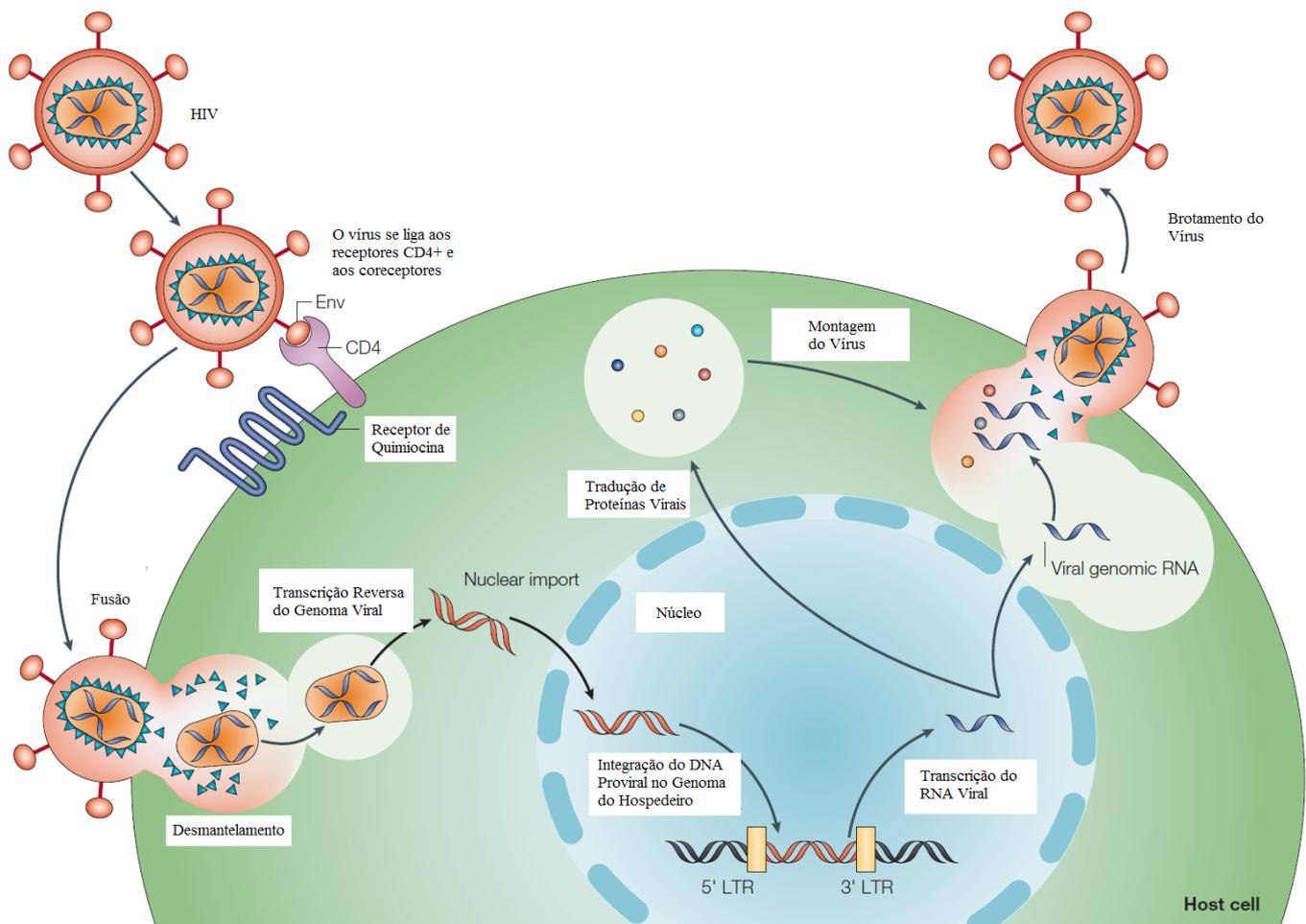


Figura 1.3.1: Aspectos principais do ciclo replicativo do HIV. Adaptado de Rambaut e colaboradores (2004).

A retrotranscrição do genoma viral envolve inicialmente a síntese de uma fita negativa de DNA a partir do RNA pela enzima viral transcriptase reversa. O híbrido DNA/RNA ativa a degradação da fita de RNA, pela porção da RNase H. Duas regiões do RNA viral resistem à degradação realizada pela RNase H e servem como *primers* para síntese da fita complementar de DNA. O produto final da transcrição reversa é uma molécula de DNA dupla-fita (Charneau et al., 1994) que será conduzida ao núcleo para posterior integração no genoma humano (Fanales-Belasio et al., 2010).

O DNA viral dupla-fita recém sintetizado é acompanhado pela Integrase até o núcleo da célula hospedeira (Fanales-Belasio et al., 2010). No núcleo, o genoma viral é inserido randomicamente no genoma da célula hospedeira através da integrase. Esta proteína cliva nucleotídeos das extremidades 3' do DNA viral dupla-fita, criando duas extremidades adesivas e une o DNA viral ao DNA do hospedeiro (Schwartz e Nair, 1999).

Os processos de transcrição iniciais dos genes provirais resultam na síntese de proteínas virais regulatórias como Tat e Rev. Tat se liga ao RNA nascente da transcrição e estimula a formação de transcritos maiores. Rev facilita a transcrição de longos transcritos de RNA e a expressão de genes estruturais e enzimáticos, além de inibir a produção de proteínas regulatórias, promovendo assim a formação de partículas virais. RNAs codificantes virais migram para o citoplasma, onde as proteínas dos novos vírions serão sintetizadas. Próximo a superfície interna da membrana plasmática da célula hospedeira, dois RNAs do vírus se associam às enzimas virais, enquanto proteínas do capsídeo se associam a estes formando o nucleocapsídeo. A partir daí, iniciará o brotamento, processo de saída da partícula viral da célula (Fanales-Belasio et al., 2010; Schwartz e Nair, 1999; Levy et al., 1990) (Figura 1.4.1).

O brotamento do HIV é um evento de fissão no qual a membrana plasmática da célula hospedeira é estruturalmente afetada, criando uma pequena membrana ao redor do capsídeo, denominada o envelope viral (Pornillos et al., 2002). Durante este processo, o vírus pode incorporar proteínas de membrana da célula hospedeira (Fanales-Belasio et al., 2010). Após o brotamento da partícula viral, processos proteolíticos do capsídeo levam a um rearranjo estrutural do vírion e geram uma partícula infectante madura pronta para infectar outra célula (Schwartz e Nair, 1999).

1.4. SURGIMENTO:

.....

O início da história da infecção do vírus da imunodeficiência em humanos é um exemplo catastrófico da relação entre diferentes espécies (Santiago et al., 2005; Nau et al., 2009; Yamaguchi et al., 2000). As condições precárias de vida e a cultura de certos povos habitantes do continente africano colaboraram sobretudo com a origem, expansão e difusão do novo vírus pela população humana (Perrin et al., 2003). Aproximadamente um século depois do surgimento do HIV (Worobey et al., 2008), oito mil pessoas morrem diariamente vítimas da infecção e o alastramento continua sem qualquer menção de ser controlado ou exterminado (UNAIDS, 2009).

Atualmente, 40 espécies de primatas abrigam infecções pelo vírus da imunodeficiência símia (SIV) (Sharp et al., 2010). Apesar do título de causador da imunodeficiência, o perfil patológico dessas infecções permanece desconhecido para a maioria das espécies. Para alguns indivíduos, no entanto, é relatado um quadro patológico semelhante ao da infecção pelo HIV em humanos (Trivedi et al., 2010). Embora espécie-específicos, a transmissão destes vírus entre diferentes espécies de símios já foi observada tanto em cativeiro, quanto em animais de vida livre. A frequência desses eventos e o impacto para a evolução do vírus e para o desenvolvimento da patogenia ainda estão sendo estudados (Hahn et al., 2000). Pesquisas recentes, baseadas em métodos moleculares, demonstram que os SIVs estão presentes na África, junto com seus hospedeiros, há mais de 30.000 anos (Worobey et al., 2010). Sendo assim, as diversas espécies símias representam um antigo e extenso reservatório de vírus com o potencial para infectar e alastrar uma nova epidemia em uma nova espécie (Hahn et al., 2000).

O surgimento de duas linhagens de HIV, HIV-1 e HIV-2, é o resultado de, no mínimo, doze transmissões zoonóticas independentes do vírus símio para humanos (Wertheim e Worobey, 2009; Nau et al., 2009). Particularmente, as duas linhagens humanas não apresentam relação genética próxima, porém estão relacionadas com diferentes linhagens de SIVs, indicando diferentes origens na região da África (Wolfe et al., 2004) (Figura 1.4.1). Análises evolutivas e filogenéticas fornecem claras evidências a respeito da provável origem e das datas de expansão da epidemia e podem desta forma, ajudar a desvendar os fatores genéticos que levaram à diversidade molecular destes vírus (Sharp et al., 2001).

A origem do HIV-2 é atribuída ao contato entre humanos e macacos da espécie *Cercocebus torquatus atys* (*sooty mangabey*), distribuídos na região do oeste da África, se estendendo do Senegal à Costa do Marfim (Lemey et al., 2003; Schim e Aaby, 1999; Yamaguchi et al., 2000). Cinco linhas de evidência foram usadas para conclusão da via zoonótica do vírus tipo dois (Gao et al., 1994; Hahn et al., 2000): (a) similaridade da organização genômica; (b) relação filogenética; (c) prevalência no hospedeiro natural; (d) coincidência geográfica e (e) rota de transmissão plausível. Devido à alta variabilidade do HIV-2 e seu comportamento filogenético quando relacionado com sequências do SIV_{sm} (SIV infectante dos símios *sooty mangabey*), são atribuídas diversas passagens mangabey-humanos (Sharp et al., 2000). Cada passagem pode ser

categorizada como um subtipo epidêmico, classificados de A à H (Santiago et al., 2005). Dados de soroprevalência sugerem que estas passagens interespecíes tenham ocorrido na região da Guiné-Bissau, por volta de 1940 (Lemey et al., 2003), enquanto que a difusão efetiva do vírus por populações humanas só foi ocorrer 20 à 30 anos depois (Poulsen et al., 2000). Atualmente o HIV-2 circula moderadamente em países do oeste da África, principalmente Senegal, Guiné, Gambia e Cabo Verde (Schim e Aaby, 1999). É raro fora deste continente, no entanto, já foi descrito em outras partes da África, na Europa, Américas e Ásia (Evans et al., 1991; Soriano et al., 1996; Poulsen et al., 2000; Schim et al., 1999; Grant and Kevin, 2001; Rambaut et al., 2004).

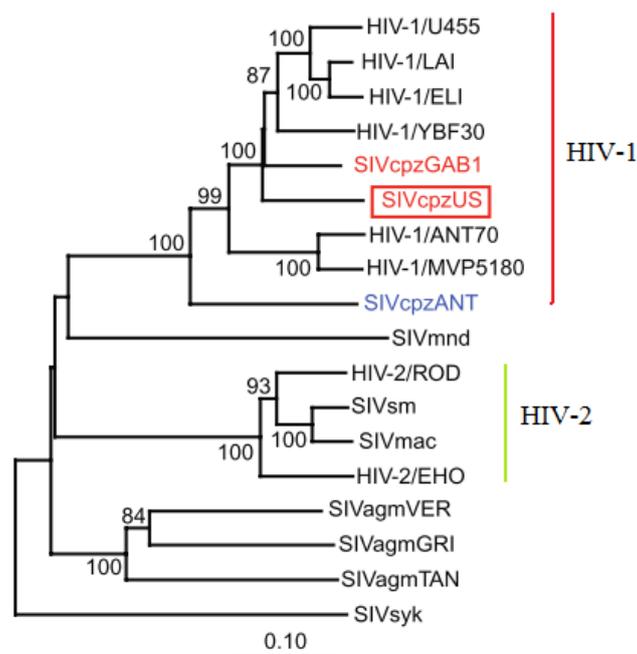


Figura 1.4.1: Relações filogenéticas entre os isolados de HIV e SIV. Figura adaptada de Gao e colaboradores (1999).

Assim como o HIV-2, o vírus do tipo 1 se originou através do contato interespecíes na região da África. Análises filogenéticas, relacionando SIVs de diferentes espécies e o vírus humano, indicam uma relação, ainda que distante, entre o HIV-1 e o SIV infectante de chimpanzés da espécie *Pan troglodytes troglodytes* (SIVcpzPtt) (Bailes et al., 2003). Essa distância filogenética tem notável relação com os processos de recombinação entre diferentes linhagens virais ocorridos na população símia. Provavelmente, os chimpanzés selvagens eram infectados simultaneamente por duas formas do vírus, adquiridos quando da caça e morte de pequenos macacos de outras espécies. Os descendentes atuais desses SIVs são encontrados em macacos *red-*

capped mangabeys (*Cercocebus torquatus*, SIV_{rcm}) e macacos *greater spot-nosed* (*Cercopithecus nictitans*, SIV_{gsn}) (Bailes et al., 2003). A co-infecção gerou oportunidade para o processo de recombinação gênica nas células dos hospedeiros e, assim, geraram um novo vírus capaz de infectar outros símios e mais significativamente, capaz de infectar hospedeiros humanos e causar AIDS. Segundo Bailes e colaboradores (2003) este seria o processo evolutivo que governou a origem e disseminação do SIV em chimpanzés. A partir destes hospedeiros o vírus parece ter cruzado a barreira entre espécies múltiplas vezes e em quatro ocasiões foi capaz de se estabelecer no organismo humano em níveis suficientes para se tornar transmissível. Essas efetivas passagens geraram quatro grupos filogeneticamente relacionados de HIV-1, denominados M (*major*), N (*new*), O (*outlier*) e P (Gao et al., 1999; Nau et al., 2009) (Figura 1.4.2).

O grupo M é a forma viral do HIV-1 mais disseminada no mundo (Hemelaar et al., 2006). Estudos filogenéticos indicam que o grupo M surgiu na região do sudeste de Camarões (Keele et al., 2006) por volta de 1908 (1884-1924) (Worobey et al., 2008) e teve um provável epicentro de disseminação em humanos na região de Kinshasa (República Democrática do Congo, ex-Zaire) (Vidal et al., 2000). Por efeitos genéticos, durante a disseminação, o grupo M diversificou-se em nove subtipos filogeneticamente relacionados, denominados como A, B, C, D, F, G, H, J e K. O espalhamento local e global destes subtipos está intrinsecamente relacionado com o movimento de populações humanas (Perrin et al., 2003). Na época de estabelecimento da epidemia na África, eram relatadas altas taxas de migrações de regiões rurais para centros urbanos, o que resultou no aumento do número de cidades. Por conta das migrações, houve uma mudança econômica e social nas novas cidades africanas, desmembrando famílias rurais, causando a perda de valores culturais, instaurando a pobreza e prostituição o que, em última análise, favoreceu a propagação de doenças sexualmente transmissíveis como o HIV (Perrin et al., 2003).

O grupo N provavelmente surgiu a partir da transmissão de um SIV_{cpz} para a população humana na região centro-sul da República dos Camarões (Keele et al., 2006). O início da epidemia é estimado para o ano de 1963 (1948-1977) (Wertheim e Worobey, 2009). Curiosamente, os vírus humanos do grupo N são descendentes de recombinações entre diferentes SIVs de chimpanzés ocorridas ainda quando o vírus infectava estes símios algum tempo antes da transmissão para humanos. Estas conclusões surgiram a partir do estudo de chimpanzés de vida livre que apresentavam

(SIV_{gor}), evidenciando uma possível transmissão gorilas-humanos. O vírus já foi isolado em um segundo paciente na República de Camarões, confirmando que o grupo P está circulando ativamente na população humana (Vallari et al., 2010). Estes dados revelam que assim como chimpanzés e Sooty mangabeys, gorilas também são prováveis reservatórios de vírus capazes de infectar humanos (Nau et al., 2009).

Diversas teorias tentam explicar as transmissões virais interespecíes do HIV-1 e HIV-2 (Wolfe et al., 2004; Marcia et al., 2005; Sharp et al., 2001). Dentre estas, a teoria natural de transferência, ou teoria do caçador é largamente aceita por diversos pesquisadores e explica de maneira plausível o surgimento do vírus em humanos (Sharp et al., 2001). Segundo esta, os vírus, HIV-1 e HIV-2, foram transmitidos de símios para humanos quando caçadores ou vendedores de carcaças de macacos foram mordidos ou expostos ao contato com sangue dos animais infectados. A exposição resultante ao sangue ou a outros fluidos corporais dos animais contaminados pode ter levado a uma contaminação humana por SIVs capazes de infectar e se multiplicar nas células do novo hospedeiro (Sharp et al., 2001). Um estudo recente demonstra que infecções por SIV em humanos não são eventos raros na África Central: 2,3% da população geral da República de Camarões apresentam sororreatividade para antígenos de SIV e 7,8% da população das vilas em que há consumo de carne de macaco também apresentam reatividade (Kalish et al., 2005). Desta forma, é possível que diferentes formas de SIV tenham sido transmitidas aos humanos antes mesmo das transmissões que levaram à ocorrência da atual pandemia, no entanto, tenham ocorrido em vilas isoladas e, pela falta de hospedeiros susceptíveis, acabaram por desaparecer (Rambaut et al., 2004).

1.5. DIVERSIDADE:

.....

Desde sua descoberta em meados de 1981, o HIV tem evoluído e se disseminado de diferentes formas em diferentes regiões do mundo (Ramos et al., 1999; Bello et al., 2006; Salemi et al., 2008; Gray et al., 2009; Hemelaar et al., 2006; Lemey et al., 2006; Rambaut et al., 2004) . Através da ocorrência de transmissões símios-humanos, o vírus emergiu naturalmente diversificado em uma ampla variedade de tipos e grupos. Aliado

a isso, a expansão na população humana tem sido acompanhada por uma das mais altas taxas evolutivas descritas para um patógeno humano (Rambaut et al., 2004; Ariën et al., 2007). Atualmente 33,3 milhões de pessoas estão infectadas no mundo inteiro (UNAIDS, 2010) abrigando variadas formas genéticas do vírus e possibilitando a emergência de novos padrões moleculares (Dalai et al., 2009; Hemelaar et al., 2006; Santoro-Lopes et al., 2000). O alto grau de variabilidade genética do HIV pode ter importantes implicações para a patogênese, transmissão, diagnóstico e para o desenvolvimento de vacinas (Geretti et al., 2006; Hemelaar et al., 2006).

As altas taxas evolutivas descritas para o HIV (Rambaut et al., 2004) são fundamentalmente consequência do processo de transcrição reversa. A infidelidade à complementaridade na adição de nucleotídeos durante o mecanismo de conversão do RNA em DNA que ocorre na transcrição reversa gera processos de inserções, deleções, duplicações ou erros de pareamento de nucleotídeos nos novos genomas provirais (Spira et al., 2003). A frequência destes processos evolutivos é mantida devido a falta de uma região de exonuclease 5'-3', responsável pela correção dos erros de incorporação (Preston et al., 1988). Devido à alta frequência dos erros, essa dinâmica garante que cada novo vírus tenha praticamente um genoma único e diferente de seu parental. A infidelidade da transcriptase reversa, o surgimento de aproximadamente 10^{10} - 10^{12} diferentes vírions por dia dentro de um mesmo hospedeiro, o grande número de indivíduos infectados e a persistência da infecção formam um conjunto de fatores ideais para oportunizar a geração e expansão da diversidade viral atual (Ho e Huang, 2002; Nelson et al., 2003; Perelson et al., 1996; Preston et al., 1988; Quinones-Mateu e Arts, 1999; Walker e Korber, 2001).

Além dos mecanismos intrínsecos à biologia viral, os processos de recombinação genômica homóloga também governam a plasticidade molecular do HIV (Hemelaar et al., 2006). A recombinação é gerada em células hospedeiras infectadas por duas ou mais variantes virais (grupos ou subtipos), resultando da capacidade da transcriptase reversa trocar o *template* de RNA no momento da transcrição reversa. Dessa forma, durante a transcrição reversa, a enzima pode alternar de uma fita de RNA para outra, produzindo uma fita de DNA viral mosaica. O cDNA retrotranscrito é composto por partes de uma variante e por partes de uma ou outras variantes dos vírus co-infectantes (Nájera et al., 2002) (Figura 1.5.1). Dois tipos de formas recombinantes são descritos envolvendo o HIV, denominados de formas recombinantes circulantes

(CRF) e formas recombinantes únicas (URF). A diferença entre os dois termos está associada aos padrões epidemiológicos de surgimento e disseminação das formas mosaicas. Se o recombinante, caracterizado por seqüenciamento completo do genoma viral, é identificado em três indivíduos epidemiologicamente não ligados, as formas são designadas CRF. Já, genomas com padrões de recombinação únicos, sem “valor” epidemiológico, são denominados URFs (Peeters e Sharp, 2000). Atualmente são descritos 50 CRFs, incluindo 49 para HIV-1 e apenas um para HIV-2, além de diversas formas URF (<http://hiv-web.lanl.gov/CRFs/CRFs.html>).

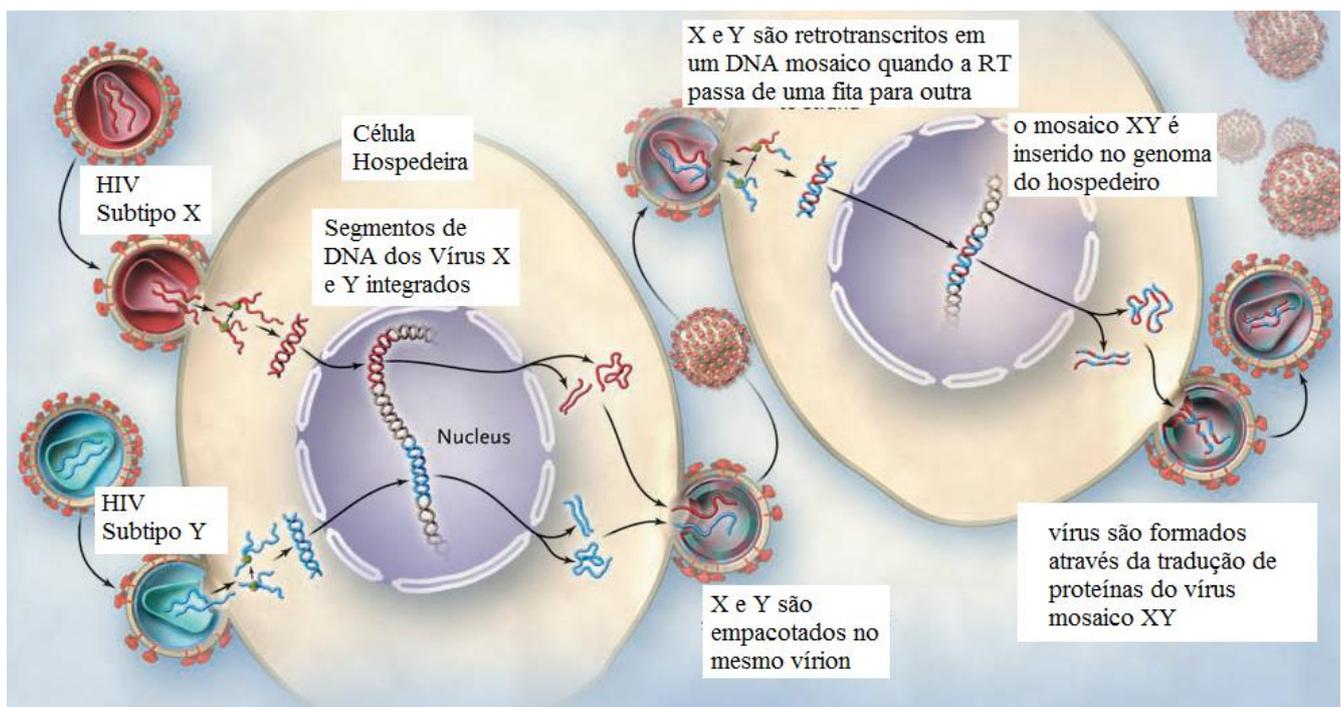


Figura 1.5.1: Evolução da diversidade do HIV durante o ciclo replicativo e a conseqüente formação de vírus recombinantes. Adaptado de Taylor e colaboradores (2008).

Desde a descoberta do HIV como agente infeccioso, cientistas de todo o mundo mobilizam forças para entender os fatores de origem, transmissão e expansão do vírus nos hospedeiros humanos. Nesse sentido, estudos sobre a diversidade molecular das formas virais em circulação e o impacto dessas nos diversos aspectos relacionados à infecção estão continuamente sendo investigados e atualizados (Bello et al., 2008; Carrion et al., 2003; Hemelaar et al., 2006; Kandathil et al., 2005; Rambaut et al., 2004; Geretti et al., 2006). Atualmente, a pandemia é entendida e compartimentalizada primariamente através das relações filogenéticas entre os isolados de HIV nas mais diversas partes do mundo.

A epidemiologia molecular tem se mostrado uma ferramenta valiosa para o estudo das epidemias de doenças infecciosas, uma vez que permite que as informações moleculares acumuladas nos genomas dos microorganismos sejam utilizadas para reconstruir e prever o seu curso (McDade e Anderson, 1996; Pinto e José, 2006). A permanente circulação geográfica dos grupos, subtipos e CRFs permite que estas variantes sejam empregadas como marcadores da epidemia e que a caracterização dos vírus circulantes em cada país seja útil aos programas de vigilância epidemiológica locais (Pinto e José, 2006).

Em uma escala mundial, enquanto os outros grupos são raros, o HIV-1 Grupo M é o tipo mais freqüente na população humana, representando aproximadamente 95% dos casos de HIV (Hemelaar et al., 2006) (Figura 1.5.2). As distâncias genéticas de aminoácidos na região do *env* entre os nove subtipos deste grupo chegam a 35%, enquanto na região *gag* essa diferença gira em torno de 15%. Dentro dos subtipos A e F, devido a uma variação genética mais acentuada, existe a formação de diferentes grupos geneticamente relacionados denominados sub-subtipos. Dentro do subtipo A, quatro sub-subtipos são reconhecidos: A1, A2, A3 e A4, enquanto que para o subtipo F dois cladados são descritos (F1 e F2) (Taylor e Hammer, 2008; Thomson et al., 2002).

A África, por ser o local de origem do HIV, é o continente com a maior diversidade de grupos e subtipos (Figura 1.5.2). O subtipo C, responsável por metade das infecções por HIV no mundo, é predominante no sul e no leste do continente africano (Hemelaar et al., 2006), além de ter grande prevalência na Índia. Infecções pelos subtipos A, B, D, F, G, H, J, K e ainda pelo CRF01_AE e CRF02_AG são também importantes epidemiologicamente (Taylor e Hammer, 2008).

A epidemiologia molecular do grupo M nas Américas não é tão complexa como aquela instaurada na África (Russell et al., 2000; Mehta et al., 2010; Nadai et al., 2009). Atualmente, há a circulação predominante do subtipo B na região, no entanto, na América do Sul, o subtipo F e recombinantes BF também possuem grande valor epidemiológico. Outros subtipos e formas recombinantes como CRF12_BF, URFs BF (Liebert et al., 2002; Thomson et al., 2002; Bello et al., 2010; Leal et al., 2007), subtipo A (Aulicino et al., 2005) e o CRF02_AG (Carrion et al., 2003) já foram também notificados. Recentemente, o subtipo C tem se disseminado pela região sul da América do Sul e propiciado o surgimento de recombinantes BC (Bello et al., 2008; Fontella et

al., 2008; Brígido et al., 2007). A América Central possui também uma epidemia baseada na circulação do subtipo B, além de abrigar menor quantidade dos subtipos C, D, F, G, H e J (Hemelaar et al., 2006).

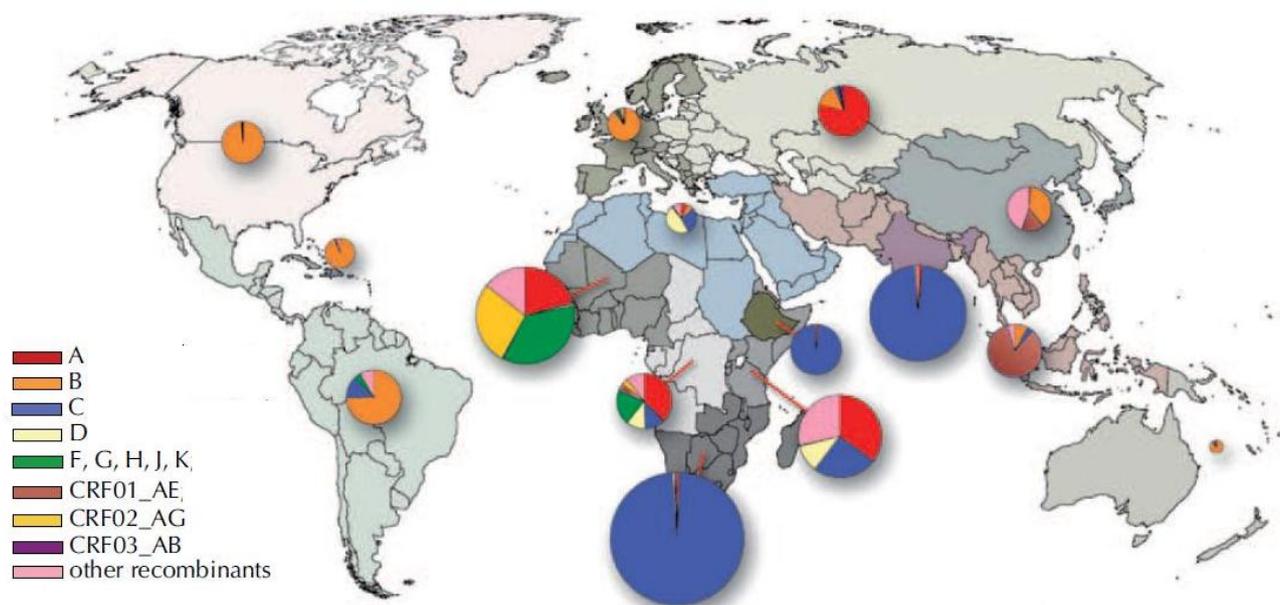


Figura 1.5.2: Distribuição Regional dos subtipos do HIV-1 grupo M. Adaptado de Hemelaar e colaboradores (2006).

As epidemias da Europa Oriental e da Europa Ocidental parecem diferir e abrigar diferentes subtipos de HIV-1. A epidemia da Europa Ocidental, assim como as Américas e a Oceania, é composta principalmente pelo subtipo B (88% das infecções). O restante das infecções é dividido pelos subtipos C, A, G e CRF02_AG. Na Europa Oriental e na Ásia Central, 79% das infecções são causadas pelo subtipo A e apenas 15% pelo subtipo B. Além disso, esta região é a única no mundo onde o CRF03_AB parece circular significativamente. Finalmente, a Ásia parece ser dominada epidemiologicamente pelo CRF01_AE e em menores quantidades circulam o subtipo B e os recombinantes CRF01_AE, CRF07_BC e CRF08_BC (Hemelaar et al., 2006).

A relação entre os subtipos do HIV-1, a transmissibilidade e a progressão da doença são tópicos de contínuas discussões. Dados recentes, envolvendo transmissão do HIV de mãe para filho durante a gravidez, mostraram que o subtipo C foi mais frequentemente transmitido que o subtipo B (Renjifo et al., 2004). Mulheres grávidas infectadas com o subtipo C no Quênia mostraram-se mais propensas a liberar células vaginais infectadas do que aquelas portadoras do subtipo B, implicando em uma transmissão sexual mais provável pelo subtipo C (John-Stewart et al., 2005). No

Senegal, 54 profissionais do sexo foram avaliadas e os resultados mostraram que aquelas que estavam infectadas pelos subtipos B, C ou D tiveram uma progressão da doença mais rápida que aquelas infectadas pelo subtipo A (Kanki et al., 1985). Além disso, diversas observações sugerem que o CRF02_AG tem um potencial patogênico aumentado e que esta característica facilitaria sua dispersão. No entanto, a maioria dos estudos nesta área é caracterizada por um baixo número de pacientes, curto acompanhamento, controles controversos, comparações não específicas (subtipos VS. não-B) e um grande número de fatores de confusão como variação imunológica, grupos étnicos, nutrição e co-infecção (Geretti et al., 2006).

1.6. SUBTIPO B:

.....

O HIV-1 grupo M subtipo B é a forma genética viral mais disseminada fora da África. Está presente nas Américas, na Europa Ocidental, na Europa Oriental, Ásia Central, Ásia Oriental, Oceania e norte da África. O número de infecções causadas por este subtipo chega a 10% de todas as infecções por HIV no mundo (Figura 1.6.1) (Hemelaar et al., 2006).

Historicamente, foi o primeiro subtipo a ser detectado no mundo. Os primeiros casos de AIDS, relatados em um grupo de homens que fazem sexo com homens nas cidades de Nova York e Los Angeles, foi uma consequência da infecção pelo subtipo B (Hymes et al., 1981). Atualmente, é a forma circulante na maioria dos países desenvolvidos, caracteristicamente relacionado ao comportamento homossexual e ao uso de drogas injetáveis (Chalmet et al., 2010; Cohen et al., 1995). Devido a essa prevalência nestes países, pioneiros em pesquisas de ponta, foi o subtipo modelo usado para o desenvolvimento das primeiras drogas antirretrovirais (Geretti et al., 2009; Kosakovsky e Smith, 2009).

A disseminação do subtipo B, após o surgimento na África, ocorreu através da introdução do vírus no Haiti por volta de 1966. Traçados históricos revelam que esta introdução está relacionada com a independência do Congo, ocorrida em 1960, e o

conseqüente retorno de Haitianos ao Caribe. Após disseminação pelo Haiti, o vírus foi provavelmente introduzido em países vizinhos no Caribe e levado aos Estados Unidos em torno do ano de 1969 (Gilbert et al., 2007).

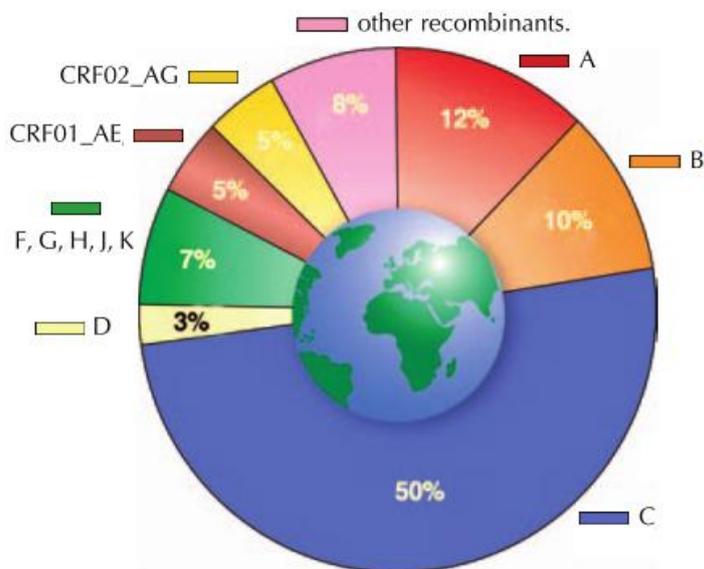


Figura 1.6.1: Distribuição global de subtipos e formas recombinantes. Os números de infecções causadas pelos subtipos do HIV-1 e formas recombinantes é mostrado como proporções do número global de casos de HIV. Figura adaptada de Hemelaar e colaboradores. (2006)

A relação entre Haiti e Estados Unidos na epidemia do subtipo B foi anteriormente uma questão de especulação (Pape et al., 2008; Farmer, 1992). Após os inúmeros relatos de Haitianos que viviam nos Estados Unidos manifestando sintomas relacionados à AIDS (Barry et al., 1984), em 1983, o CDC (*Center for Disease Control*), órgão do governo responsável pelo controle de doenças nos Estados Unidos, classificou esta população como um efetivo grupo de risco. A este grupo, designado por “4H”, também pertenciam as profissionais do sexo (*Hookers*), homossexuais (*Homosexuals*) e usuários de drogas injetáveis (*Heroin addicts*) (Pitchenik, 1983). Esta medida criou uma situação de estigmatização e discriminação da população Haitiana que perdurou mesmo depois da retratação do CDC em 1985 (Greco, 1983). A partir de então, hipotetizou-se uma transferência da epidemia do Haiti para os Estados Unidos. Diversos pesquisadores (na época e ainda hoje) negam esta transferência e especulam que a origem da epidemia, na verdade, se estabeleceu de forma contrária (Farmer, 1992; Pape et al., 2008), ou seja, que a introdução do vírus no Haiti diretamente dos Estados Unidos teria ocorrido como conseqüência do turismo sexual. Esta hipótese é baseada em

testes retrospectivos que demonstram a ausência de casos de AIDS no Haiti antes de 1979 (Liautaud et al., 1983; Farmer, 1992; Pape et al., 2008).

Recentemente, amostras de sangue periférico de pacientes haitianos que viviam nos Estados Unidos na década de 80 foram recuperadas e através de análises moleculares, os pesquisadores identificaram a presença de HIV em cinco dessas amostras. Utilizando análises filogenéticas os autores puderam mostrar um sinal genético ancestral para estes isolados e sugerir que a epidemia dos Estados Unidos é derivada do Haiti com 99,79% de probabilidade. Ainda, através das relações evolutivas, os autores sugeriram uma expansão da epidemia do subtipo B para o mundo através dos Estados Unidos (Gilbert et al., 2007).

Dados do início da epidemia mostram que a Europa também foi atingida inicialmente pelo subtipo B. As características da transmissão, devido ao contato Europa-Estados Unidos, eram semelhantes àquele apresentado nos países da América do Norte, estando principalmente relacionado ao contato homossexual (Melbye et al., 1984; Casado et al., 2000). Ao contrário das Américas, onde a introdução do vírus foi provavelmente monofilética, a epidemia da Europa é derivada de diferentes introduções por diferentes rotas (Paraskevis et al., 2009).

Para outras regiões do mundo, como América do Sul, Caribe, África e Ásia, a epidemia do subtipo B é estudada com foco em prevalência (Nadai et al., 2009; Castro et al., 2003; Thorne et al., 2010; Esteves et al., 2009; Liebert et al., 2005; Liebert et al., 2002). Dados de origem, evolução e disseminação nessas regiões ainda são pontuais ou inexistentes (Bello et al., 2006; Deng et al., 2008; Li et al., 2010).

Além da extensa disseminação, o subtipo B exibe certo grau de variabilidade genética em árvores filogenéticas (Leal e VillaNova, 2010). Curiosamente, isolados de Trinidad e Tobago, Tailândia, China e Coreia são notáveis pela sua relação genética com isolados de outros países (Gilbert et al., 2007; Leal e VillaNova, 2010; Nadai et al., 2009). Além destes, isolados que apresentam um motivo GWGR na região V3 *loop* do envelope já foram observados em diversos países, como China, França, Filipinas, República Tcheca e Cuba (Leal e VillaNova, 2010). No entanto, altas frequências destes isolados são vistas apenas no Brasil (Di Mascio et al., 2004; Diaz et al., 2008; Santoro-lobes et al., 2000). A total variabilidade genética dentro do subtipo B e sua real

importância para aspectos envolvendo a infecção ainda não são completamente entendidos.

1.7. FILOGENIAS:

.....

O conteúdo genético dos vírus em circulação contém informações que podem remontar a história demográfica da população de onde o vírus foi amostrado (Harvey et al., 1996). A construção de filogenias, ou árvores filogenéticas, permite representar as estimativas genéticas a respeito da relação entre táxons e sua descendência dentro de um determinado tempo evolutivo (Posada et al., 2002). Como resultado, a inferência de parâmetros demográficos a partir de dados genéticos se tornou uma importante ferramenta em diversas áreas do conhecimento, incluindo a genética estatística, antropologia, biologia da conservação, epidemiologia e virologia (Harvey et al., 1996).

A epidemia do HIV difere de outros patógenos infecto-contagiosos pela variabilidade molecular extrema e pela cronicidade da infecção, dado que indivíduos mesmo não sintomáticos são ainda hospedeiros para evolução e disseminação do vírus. Desta forma, a abordagem epidemiológica dada ao estudo de origem, disseminação e diversidade da epidemia do HIV teve que tomar diferentes rumos a partir da epidemiologia convencional e partir para métodos que envolvessem um caráter evolutivo (Pilcher et al., 2008). A partir de então, com a abundância de dados genéticos, diversos trabalhos vem construindo o cenário atual de entendimento da difusão do HIV pelo mundo através de análises filogenéticas e filogeográficas (Gilbert et al., 2007; Gray et al., 2009; Salemi et al., 2008; Paraskevis et al., 2007; Robbins et al., 2003; Bello et al., 2008; Tee et al., 2009; Plenum et al., 1999).

Partindo do simples pressuposto que dez sequências divergem de seu ancestral comum ao longo do tempo, é possível assumir que o número de diferenças nucleotídicas entre elas irá se alterar. A construção de uma árvore filogenética a partir destas sequências parece uma tarefa simples: apenas contar as diferenças entre as sequências e agrupar aquelas mais próximas. A simplicidade de tal algoritmo subestima a complexidade dos problemas da inferência filogenética. Como a taxa de mudança dos nucleotídeos não é constante ao longo do tempo, a simples contagem das diferenças não

é necessariamente uma indicação confiável de quanto elas divergiram. Acima de tudo as filogenias devem levar em conta a incerteza dos processos evolutivos (Holder e Lewis, 2003).

A inferência filogenética é um processo de inferência estatística que definirá a melhor relação histórico-evolutiva entre os indivíduos analisados, denominados táxons. As relações evolutivas são baseadas na relação entre táxons ancestrais e táxons atuais. No entanto, para a maioria dos organismos, geralmente os táxons ancestrais não podem ser amostrados, pois estão restritos ao passado. No caso do HIV, estão restritos a um ciclo replicativo anterior dentro de um mesmo hospedeiro. Sendo assim, os métodos filogenéticos avaliarão estas relações e, com base no modelo evolutivo assumido, elegerão árvores que se adequam melhor aos dados (Miyaki et al., 2001).

As árvores filogenéticas podem ser representadas de duas maneiras estabelecidas conforme a relação com um grupo sabidamente ancestral: árvores não-enraizadas e árvores enraizadas. A primeira forma não aponta onde está a espécie ancestral de todo o grupo, mostrando apenas as relações de topologia (a forma da árvore) entre os táxons. Já as árvores enraizadas permitem o estabelecimento da direção das modificações evolutivas. O grupo externo, responsável pelo enraizamento, determina os estados primitivos de certo caráter dentro daquele grupo de organismos (Miyaki et al., 2001). O número exato de possibilidades de relações evolutivas entre táxons, estabelecendo árvores filogenéticas enraizadas, é dado pela fórmula (Felsenstein, 1978):

$$N = \frac{(2n - 3)!}{2^{n-2}(n - 2)!}$$

onde n corresponde ao número de táxons. Conforme se aumenta o n, maior o número de possíveis árvores filogenéticas e maior o tempo computacional para processamento dos dados. Sendo assim, um n de apenas 10 táxons pode formar até 34.459.425 árvores enraizadas. Dentro desse grupo de árvores é ainda necessário escolher àquela que melhor explica a evolução dos táxons analisados.

A reconstrução filogenética é um complexo processo que envolve uma série de passos necessários para preparação dos dados para análise. O alinhamento é o primeiro passo nesse estudo evolutivo. A metodologia filogenética é baseada no pressuposto que os caracteres são homólogos, ou seja, os nucleotídeos observados em determinada

posição são derivados da mesma posição no genoma do organismo ancestral comum das sequências analisadas. Atualmente, diversos *softwares* computacionais realizam automaticamente este processo (Harrison, 2006).

Depois que o alinhamento foi proposto, diversos métodos podem ser usados para estimar as filogenias das sequências sob estudo. Todos os métodos usualmente utilizados atualmente levam em conta dois critérios: a especificação do modelo de busca para achar as árvores e a especificação de um critério-ótimo necessário para avaliar a árvore proposta (*optimality criteria*) (Posada et al., 2002).

Idealmente, para um dado conjunto de sequências, todas as árvores formadas deveriam ser analisadas para garantir a escolha da melhor árvore como resultado final. No entanto, dependendo do número de táxons a ser avaliados, a análise de todas as árvores é exaustiva e geralmente impraticável computacionalmente (Crandall e Posada, 2002). Como no exemplo demonstrado anteriormente, com apenas 10 táxons, aproximadamente 34,5 milhões de árvores serão analisadas em busca da melhor. Desta forma, para obtenção de uma árvore filogenética, faz-se necessária a escolha de algoritmos que vasculhem o maior número de árvores em busca da melhor. Dois grupos de algoritmos são destacados: os algoritmos exatos, que consideram todas as árvores possíveis (busca exaustiva e *branch-and-bound*) e os algoritmos heurísticos, que procuram a árvore ótima dentro de um conjunto de árvores, gastando menos tempo computacional (Miyaki, Russo, and Pereira 2001). Atualmente, devido ao tempo e à exigência computacional, os métodos heurísticos são preferidos aos exatos. Como desvantagem, repetidos processos de procura em um mesmo conjunto de sequências pode levar a resultados diferentes, dependendo da árvore que é construída inicialmente pelo algoritmo (Crandall e Posada, 2002). As árvores iniciais também sofrem a influência de algoritmos para serem montadas, entre eles: *neighbour-joining*, *stepwise addition* e *star decomposition*. O *neighbour-joining*, algoritmo mais popular atualmente, converte as diferenças entre os táxons em distâncias simples (contando o número de diferenças entre duas sequências) e constrói uma árvore inicial (Holder and P. O. Lewis 2003).

O critério-ótimo é por definição o critério utilizado para verificar a adequação dos dados a uma determinada topologia associada aos tamanhos de ramos – em termos do esperado número de mudanças por sítio desde a derivação do ancestral (Edwards,

1996). O principal critério usado em filogenias de HIV é a máxima verossimilhança (ML) (Crandall e Posada, 2002), no entanto, recentemente, a estatística Bayesiana, fundamentada também na máxima verossimilhança, invadiu a construção de filogenias (Holder e Lewis, 2003; Liao et al., 2009; Gilbert et al., 2007; Keng et al., 2009; Fontella et al., 2008). Devido aos processos evolutivos que governam as modificações do genoma do HIV (Holder e Lewis, 2003), o dois métodos parecem se adequar as análises do vírus (Lemey et al., 2003; Dalai et al., 2009; Bailes et al., 2003; Fontella et al., 2008; Bello et al., 2007) (Tabela 2.1)

Tabela 1.7.1: Comparação de métodos filogenéticos. Adaptado de Holder e colaboradores (2003).

Método	Vantagens	Desvantagens	Software
Máxima Verossimilhança	A probabilidade captura totalmente o sinal dos dados para construção de uma filogenia	Pode ser muito lenta (depende do poder de procura das árvores e dos recursos computacionais)	PAUP PAML phyML
Estatística Bayesiana	Tem grande ligação com a máxima verossimilhança; pode ser mais rápida para encontrar o suporte estatístico dos ramos	Os parâmetros para as “distribuições anteriores” devem ser especificados; pode ser difícil especificar quando a cadeias convergiram	Mr. Bayes BEAST

O calculo do “valor-ótimo” de uma árvore filogenética, dado um determinado critério, requer um modelo que indique a probabilidade de substituição de um nucleotídeo por outro, denominado de modelo de substituição (Crandall e Posada, 2002). A taxa de evolução das sequências não é constante ao longo do tempo, portanto, a quantidade de diferenças genéticas entre as sequências não necessariamente indica há quanto tempo elas divergiram. Além disso, existem outros fatores presentes na história do organismo que pode levar a alterações no ritmo de divergência das sequências. Os modelos de substituição são abstrações ou simplificações da real evolução dos genomas, mas incluem as características mais importantes desses processos (Penny et al., 1994). Como modelos, existem certas constâncias nos mais diversos métodos:

1. As substituições são descritas por um processo de Markov homogêneo, onde a probabilidade de substituição de um nucleotídeo X pelo Y não depende do estado prévio do nucleotídeo X;

2. Os eventos de substituição são independentes ao longo dos sítios das sequências nucleotídicas;
3. Os eventos de substituição são reversíveis. No entanto a probabilidade dos eventos permanece a mesma;
4. As taxas de mudança são as mesmas para todos os sítios, ou seja, todos os sítios evoluem nas mesmas taxas.
5. A composição de nucleotídeos é estacionária, ou seja, a frequência esperada das bases nitrogenadas não muda em diferentes partes da árvore.

Os modelos de substituição são representados por matrizes 4 x 4, onde cada elemento representa as taxas instantâneas de mudança de um nucleotídeo X (linha) para o Y (coluna) (Posada et al., 2002). Análises típicas de substituição incluem o GTR (generalised time-reversible) (Tavaré, 1986), modelo que requer seis parâmetros para taxa de substituição e também quatro parâmetros para a frequência das bases. Os diversos modelos como um todo também consideram o parâmetro r , que define a taxa de mudança de um par de nucleotídeos, e o parâmetro π , que reflete as frequências dos nucleotídeos, conforme abaixo (adaptada de Posada et al., 2002):

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix}
 & \text{A} & & \text{C} & & \text{G} & & \text{T} \\
 \text{A} & -r_2\pi_C - r_4\pi_G - r_6\pi_T & & r_2\pi_C & & r_4\pi_G & & r_6\pi_T \\
 & r_1\pi_A & & -r_1\pi_A - r_8\pi_G - r_{10}\pi_T & & r_8\pi_G & & r_{10}\pi_T \\
 \text{C} & r_3\pi_A & & r_7\pi_C & & -r_3\pi_A - r_7\pi_C - r_{12}\pi_T & & r_{12}\pi_T \\
 & r_5\pi_A & & r_9\pi_C & & r_{11}\pi_G & & -r_5\pi_A - r_9\pi_C - r_{11}\pi_G \\
 \text{T} & & & & & & & & \text{G}
 \end{pmatrix}$$

Uma importante extensão desses modelos de substituição incorpora a possibilidade de variação nas taxas evolutivas entre os sítios, permitindo ao modelo mais realismo (Yang, 1993). Desta forma, para cada sítio no DNA será atribuída uma probabilidade de evolução a uma taxa contida em um intervalo discreto de probabilidades. O método mais comum de heterogeneidade de taxas evolutivas para o estudo do HIV atualmente é modelado através da distribuição gama (Γ). Essa distribuição considera um número específico de taxas de evolução para os sítios do DNA. A aplicabilidade deste modelo nas inferências filogenéticas é facilitada pela simplicidade do método, já que apenas um único parâmetro (alfa, α) controla a forma da distribuição gama (Figura 1.7.1). Quando $\alpha < 1$, existe um grande número de taxas de evolução entre os sítios das sequências em análise, ou seja, quanto maior α , menor a

heterogeneidade. Algumas vezes, uma proporção de sítios invariáveis (I), no qual uma determinada proporção de sítios é assumida como incapaz de sofrer substituição também pode ser usada para modelar a heterogeneidade entre os sítios (Posada et al., 2002).

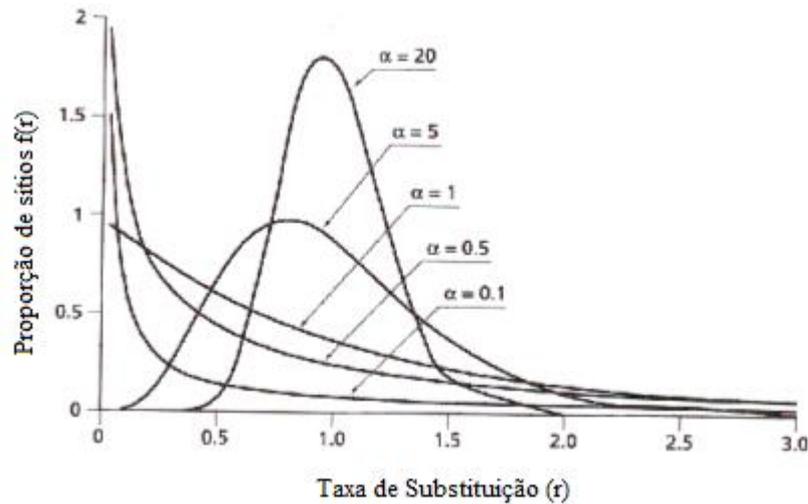


Figura 1.7.1: Diferentes formas da distribuição gama conforme o valor do parâmetro α .

Após a escolha do modelo de substituição e a construção da filogenia, é necessário verificar o grau de confiança das relações estabelecidas na árvore. Tipicamente, a confiança das relações entre os ramos é acessada utilizando o *bootstrap*. Recentemente, abordagens Bayesianas associadas à inferência filogenética têm fornecido meios para estimação simultânea da topologia e das medidas de incerteza de cada ramo, fornecendo, ao final das análises, as probabilidades posteriores (Holder e Lewis, 2003).

1.8. ML:

.....

As análises de Máxima Verossimilhança (ML) reconstróem de maneira acurada a relação entre seqüências homólogas que se distanciaram ao longo do tempo. Neste método de avaliação das árvores construídas, a hipótese (topologia, comprimento dos ramos e modelo de substituição) é avaliada pela capacidade de prever os dados (seqüências de nucleotídeos) observados - verossimilhança (Holder e Lewis, 2003), ou

seja, a verossimilhança de uma árvore é proporcional à probabilidade de observar o dado dada a árvore. A sequência de nucleotídeos é tomada como dados, pois a probabilidade de observar os nucleotídeos em um determinado sítio da sequência de DNA depende da topologia da árvore, do tamanho dos ramos que sustentam os táxons e do modelo evolutivo associado à substituição de nucleotídeos. Além disso, segundo este método, o número de substituições em um intervalo de tempo de duração T , segue uma distribuição de Poisson com parâmetro λT . O parâmetro λ é derivado de uma distribuição exponencial que estima o tempo entre uma substituição e outra. Finalmente, a probabilidade de observar os dados em um sítio particular é a soma de todos os possíveis nucleotídeos que poderiam ser observados nos nós internos da árvore. O número das possíveis combinações de nucleotídeos rapidamente se torna muito grande para ser enumerados (Huelsenbeck e Bollback, 2001). No entanto, através do algoritmo criado por Felsenstein esses cálculos podem ser realizados de uma maneira computacionalmente eficiente (Felsenstein, 1981). Os números infinitesimais resultantes desses processos são submetidos aos cálculos de logaritmo para torná-los “processáveis” nos atuais computadores (Huelsenbeck e Bollback, 2001).

O ML estima os parâmetros através da procura por árvores onde a combinação de todos os parâmetros maximize o valor da função de probabilidade. Na problemática das filogenias, a maioria dos computadores segue os seguintes passos para encontrar a árvore com maior verossimilhança (Huelsenbeck e Bollback, 2001):

1. Uma árvore com determinada topologia é construída;
2. Os parâmetros para esta árvore são modificados até ter a verossimilhança aumentada;
3. O valor de máxima verossimilhança para esta árvore é armazenado;
4. Outras topologias são construídas e seus parâmetros também são avaliados;
5. Finalmente, a árvore que possuir o valor de máxima verossimilhança é a melhor estimativa evolutiva para o dado conjunto de sequências e modelo evolutivo.

Embora estes processos simplifiquem os verdadeiros processos que governam a evolução de uma sequência, eles parecem ser relativamente robustos para estimar as relações evolutivas entre os táxons. Além disso, as estimativas de confiabilidade, ou

suporte estatístico, para os diversos ramos clados de uma árvore podem ser acessadas, como já exposto anteriormente, através do método de *bootstrapping* (Crandall e Posada, 2002).

1.9. IB:

.....

A inferência Bayesiana (IB) é baseada nas probabilidades posteriores de um parâmetro, ou seja, a probabilidade de um parâmetro condicionado à observação dos dados. Nas análises de Máxima verossimilhança, pelo contrário, a melhor árvore maximiza a probabilidade de observar os dados, dados os parâmetros. As probabilidades posteriores podem ser calculadas utilizando a fórmula de Bayes (Huelsenbeck e Bollback, 2001), conforme segue:

$$\Pr(\text{árvore}|\text{seqs}) = \frac{\Pr(\text{árvore}) \Pr(\text{seqs}|\text{árvore})}{\Pr(\text{seqs})}$$

O termo $\Pr(\text{árvore}|\text{seqs})$ é a probabilidade posterior e é dado pela probabilidade da hipótese (árvore) dados os dados (alinhamento de sequências). O termo $\Pr(\text{seqs}|\text{árvore})$ é o mesmo para a função de ML e o multiplicador $\Pr(\text{árvore})$ é a probabilidade incondicional da hipótese, também chamado de probabilidade anterior (*prior probabilities*). Através da multiplicação da probabilidade anterior pela verossimilhança (Huelsenbeck e Bollback, 2001; Holder e Lewis, 2003), encontramos a probabilidade posterior. O denominador $\Pr(\text{seqs})$ é uma integração sobre todas as possibilidades de topologias, tamanhos de ramo e valores para os parâmetros do modelo evolutivo, o que garante que a soma da probabilidade posterior para todos eles seja 1.

As probabilidades anteriores de diferentes hipóteses são especificadas pelo pesquisador. Na maioria dos aplicativos que lidam com IB, existem distribuições uniformes associadas às probabilidades anteriores que assumem que todos os valores possíveis são dados pela mesma probabilidade. Desta forma, as diferenças nas resultantes probabilidades posteriores são atribuídas à diferenças na verossimilhança. Apesar disso, as árvores finais de uma análise de ML e uma gerada por IB podem ter

diferenças, pois estes dois métodos tratam alguns parâmetros de maneiras diferentes (Holder e Lewis, 2003).

Como as funções de probabilidade associadas aos cálculos da verossimilhança são muito complexas para serem calculados rapidamente, os métodos de inferência filogenética Bayesiana utilizam a Cadeia de Markov Monte Carlo (MCMC) para aproximar as distribuições probabilísticas em uma grande variedade de contextos (Holder e Lewis, 2003). Desta forma, o tempo computacional exigido é menor que no método de ML. A MCMC realiza uma série de processos na forma de uma conceitual cadeia:

1. Uma determinada topologia X é construída;
2. A probabilidade posterior relativa a esta árvore é calculada;
3. Perturbações aleatórias em alguns parâmetros são realizadas no presente estado da filogenia e geram uma filogenia Y;
4. A probabilidade posterior é novamente calculada para este novo estado;
5. Se essa filogenia Y tem uma probabilidade posterior maior do que a razão das probabilidades posteriores (da árvore X e da árvore Y), essa árvore será continuará na cadeia. Caso contrário, a árvore do passo X continuará;
6. Supondo que a árvore Y teve valores melhores, ela sofrerá uma nova perturbação aleatória nos parâmetros. Isso gerará a árvore W;
7. Todas estas árvores amostradas são armazenadas para posterior comparação.

Repetindo este processo milhões de vezes, uma longa cadeia no “espaço de parâmetros” é criada. Como, na maioria das vezes, ocorre a rejeição de árvores com menor probabilidade do que a árvore no estado atual, a cadeia tende a amostrar filogenias semelhantes. Se as novas propostas de filogenias são muito próximas ao estado atual, a cadeia deverá “correr” por muitos passos. Estimar o quanto a cadeia deve percorrer para amostrar um suficiente número de árvores de todas as árvores possíveis para as sequências dadas (espaço de árvores) é um fator fundamental para obter bons resultados em uma análise Bayesiana. A figura 2.2 exemplifica o andamento da MCMC para amostrar as árvores em um espaço de árvores. Supondo que os quadrados em a, b e c representem todas as filogenias possíveis para um conjunto de sequências e que os

pontos pretos sejam as filogenias que vão sendo amostradas com o desenvolvimento da MCMC, vemos que, ao final do processo (Figura 2.1 c), amostrou-se um grande número de filogenias representativo do espaço de árvores total (Lewis et al., 2001).



Figura 2.2: Espaço de possíveis árvores analisados pela Cadeia de Markov Monte Carlo. (a) um milhão de passos percorridos pela cadeia ; (b) Trinta milhões de passos percorridos pela cadeia e (c) Cem milhões de passos percorridos pela cadeia. Retirado de Lewis e colaboradores (2001).

Os métodos Bayesianos de inferência filogenética ainda têm a vantagem de aplicar modelos que envolvem relógios moleculares. As distâncias genéticas, depois de “tratadas” pelos modelos de substituição, não tem qualquer significado sozinhas quando se deseja estimar, por exemplo, a idade do ancestral comum mais recente de dois táxons. Esta e outras questões podem ser avaliadas quando aplicamos uma medida de tempo nas inferências, a fim de calibrar as taxas evolutivas. Sequenciamentos de amostras isoladas em diferentes épocas podem fornecer a calibração adequada para inferências temporais, pois se assume uma taxa evolutiva constante ao longo de um tempo t para todos os ramos de uma filogenia. Esse modelo de evolução constante é conhecido como relógio molecular estrito (Kumar, 2005; Bromham e Penny, 2003). As taxas evolutivas dependem de diversos fatores e podem variar, nem sempre seguindo a constância proposta por este modelo. Recentemente, um novo modelo para relógios moleculares foi proposto, sendo denominado de relógio molecular relaxado. Segundo esse, as taxas de evolução podem variar ao longo da árvore para diferentes grupos e não são correlacionadas, ou seja, grupos evolutivamente próximos não necessariamente terão taxas de evolução semelhantes (relógio molecular relaxado não correlacionado) (Drummond et al., 2006).

Adicionalmente, complexos modelos de dinâmica populacional podem ser analisados sob uma perspectiva Bayesiana. Quando o conjunto de sequências submetido às análises são isolados de uma população homogênea, os parâmetros de história

demográfica podem ser usados para modelar as mudanças populacionais ao longo do tempo. Modelos não paramétricos, como o Bayesian Skyline Plot, e modelos paramétricos, que podem assumir um número populacional constante, um crescimento exponencial da população ou ainda um crescimento logístico estão incluídos entre os modelos disponíveis para análise Bayesiana (Drummond e Rambaut, 2007).

Capítulo 2: Objetivos

.....

Este trabalho teve como principal objetivo identificar o padrão de disseminação do HIV-1 subtipo B pelas Américas.

Objetivos específicos:

1. Descrever a possível rota de entrada do subtipo B nas Américas;
2. Estimar o ano de entrada do subtipo nas Américas;
3. Entender o envolvimento da América do Sul no estabelecimento da epidemia;
4. Traçar possíveis rotas de disseminação entre os diversos países componentes da América do Sul, América Central e América do Norte.
5. Contribuir para o delineamento do cenário epidemiológico atual de disseminação do HIV-1.

Capítulo 3: Manuscrito

.....

Reviewing the history of HIV-1: Spread of subtype B in the Americas

Dennis Maletich Junqueira^{1,2}, Rubia Marília de Medeiros^{1,2}, Maria Cristina Cotta Matte^{1,2}, Leonardo Augusto Luvison Araújo², José Artur Bogo Chies¹, Patrícia Ashton-Prolla^{1,3}, Sabrina Esteves de Matos Almeida²

¹ Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil

² Centro de Desenvolvimento Científico e Tecnológico (CDCT), Fundação Estadual de Produção e Pesquisa em saúde (FEPPS), Porto Alegre, Brazil

³ Laboratório de Medicina Genômica, Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, Brazil

Key words: HIV, subtype B, epidemic, Americas, phylogeography

Running title: Dissemination of HIV-1 subtype B in the Americas

Abstract:

The diffusion of HIV-1 subtype B (HIV-1B) is a reflection of the movement of human populations tied to social, political and geographical issues. The initial dissemination of the virus outside of Africa seems to have a passive involvement of Caribbean populations in the spreading to the United States; however the exact pathways for the full implementation of the pandemic in the Americas are still uncertain. Here we propose a geographical scenario to the dissemination of HIV-1B in the Americas applying phylogenetic analyses and genetical statistics on 313 available sequences of *pol* gene from 27 countries. Maximum likelihood and Bayesian inference were used to explore phylogenetic relationships between HIV-1B sequences and F_{ST} estimations were applied to infer the genetic structuring in the viral population. We found that the initial dissemination and consequent spreading of subtype B in the Americas occurred through a single introduction in the Caribbean around 1964 (1950-1967). Phylogenetic trees presented evidences for several events of primary outbreaks in countries from South America directly seeded by the Caribbean epidemic. Cuba provides an exception where its epidemic seems to be a result of a South American introduction. One clade comprising isolates from different countries emerged in the most derived branches reflecting the intense circulation of the virus in the American continent. Statistical analyses supported the genetic compartmentalization among Americas, revealing a close relationship between the South American and Caribbean epidemics. These findings highlight the complex patterns of human migration and the intrinsic dissemination of the HIV epidemic.

Introduction:

The intense displacement of human populations is reflected in the current diffusion and expansion of the HIV epidemic around the world (1-3). Population bottlenecks, recombination, genetic drift and founder effects are characteristics associated with the viral dissemination within these human populations and define the variability and deployment of the HIV / AIDS pandemic (4, 5). A single event of transmission in an unaffected area may result in the rapid spread of a unique viral form within a group with defined risk behavior (6, 7) and therefore be responsible for the establishment of the epidemic in that area (8, 9) . However, in consequence of the rapid evolution of HIV and its global diffusion, the exact pathways of dissemination are often uncertain.

The emergence of HIV-1 resulted from the cross-species transmission of simian immunodeficiency viruses from chimpanzees to humans in West-Central Africa at the beginning of the twentieth century (10). Group M, responsible for the vast majority of HIV infections worldwide, initially spread throughout Africa and, by the action of several genetic forces, has diversified into different subtypes (11-14). The spreading of these variants in the human population was not noticed for nearly eight decades. The first records of infection date from 1981, in American patients infected with subtype B viruses presenting clinical symptoms of what nowadays is known as AIDS (15, 16).

The spread of subtype B out of Africa occurred initially through a single introduction in Haiti in the 1960s and is probably associated with the return of Haitian professionals from work missions in the Congo (5). After expansion in Caribbean, current evidence points to the dissemination of the virus from there directly to North America. The subsequent transmission and spread of the virus in the United States allowed the epidemic to grow and expand to other parts of the world (5, 17-19). Nowadays, subtype B occupies an important position in the epidemiological profile of various countries in Europe, Asia and Africa and is also the only subtype circulating in several countries of the Americas (19-23).

The dissemination of an infectious disease reflects complex interactions between the infectious agent, host and the environment (7). Thus, a strategy widely used in epidemiological research to identify pathways of dissemination of an infectious agent, combines the analysis of sociodemographic evidence as a complement to phylogenetic

data (3, 5, 24, 25). In a recent study, Gilbert et al. (2007) demonstrates the geographic sequence for the emergence of the pandemic caused by HIV-1 subtype B out of Africa (starting in the Haiti) by taking 117 sequences from 19 countries (5). However, South America is poorly represented in this work by only nine sequences from four countries. Considering that Caribbean has close economical, historical, social, and even geographic relationships with several countries in South America (26-28), it is possible to assume that HIV-1B transmission could have occurred directly from Caribbean to South America. In the present study we further investigated the evolutionary history of subtype B in the Americas using a set that encompasses 313 available sequences of *pol* gene from 27 countries and, through phylogenetics, examined the role of South America in the establishment of the HIV pandemic.

Material and Methods:

Dataset Selection:

Around 6000 HIV-1 subtype B sequences of protease and partial reverse transcriptase segments of the *pol* gene (2253-3233 relative to HXB2) were selected from the Los Alamos HIV Sequence Database (<http://www.hiv.lanl.gov/>) and Genbank. To ensure the selection of high quality data this set of sequences should meet the following criteria: (a) samples isolated from patients living in the Americas; (b) country of origin clearly established; (c) only one sequence per patient; (d) no report of intersubtype recombination; (e) no evidence of hypermutation and (f) no occurrence of premature stop codons, frameshift mutations or ambiguities saturation. In order to understand the spread of the virus by as many countries without losing quality of results all sequences were examined for evidence of inter-subtype recombination. We selected sequences with minimum confidence threshold of 0.95 pure subtype B in a window size of 200 using RIP tool at Los Alamos HIV Sequence Database. Moreover, the dataset was reevaluated using additional reference sequences by obtaining a neighbor-joining phylogeny in order to guarantee the selection of non-inter-subtype recombinants. Even though all HIV sequences currently described can be considered intrasubtype recombinants at some level, these evolutionary events are likely insignificant in the context of origin and geographical grouping (29, 30).

Sequence alignments were created using MUSCLE (31) and manually edited for optimization. Additionally, four sequences of subtype D and two of subtype C were

selected as outgroup. Three different alignments were built for this study, and all of them are available upon request.

Phylogenetic Reconstruction:

The reconstruction of phylogenetic trees was performed by maximum likelihood (ML) using a set of 313 subtype B sequences that met our quality criteria (SI Table 1). The ML analysis was conducted with the program phyML (32) under the GTR model of nucleotide substitution with proportion of invariable sites and substitution rate heterogeneity (GTR+G+I). The nearest neighbor interchange was used for heuristic tree searches. Support for internal nodes was obtained via parametric bootstrap using 1000 replicates.

A Bayesian Markov Chain Monte Carlo approach as implemented in BEAST v1.5.4 (33) was performed in a set of 263 sequences (supporting information (SI) Table 1) in order to reconstruct the phylogenetic tree and estimate the date of the most recent common ancestor of the epidemic in Americas. The evolutionary history was inferred with a Bayesian Skyline (BSP) coalescent tree prior, under an uncorrelated lognormal relaxed clock, GTR + I + G model of nucleotide substitution. Three independent runs of 300 million steps sampled every 30000 generations were performed and the effective sample size was evaluated in TRACER (34). The maximum sum of clade credibilities tree was selected from the posterior tree distribution.

Genetic Diversity:

Population genetic structure of HIV subtype B among countries of North America, Central America, South America and Caribbean were quantified using estimates of F statistics (35). A set of 308 sequences were built, excluding countries represented by only one sequence and changing any ambiguity by “N”. Estimations were calculated using an analysis of molecular variance (AMOVA) (36) in Arlequin Ver 3.5.1.2 (37) under the Kimura two parameter model and 10.000 randomizations. Invariable sites were included and sites with gaps/missing data were considered. A nonmetric multidimensional scaling (MDS) plot was obtained using SPSS v.8.

Results:

Phylogenetic Analysis:

We inferred a Maximum likelihood phylogeny to investigate the role of South America in the HIV subtype B epidemic. A set of 313 public sequences of protease and partial reverse transcriptase genes from 27 countries representing North America, South America and Caribbean (including two countries of the Central America) were carefully selected. Additionally, four African strains of subtype D and two of subtype C were used as outgroup. This approach supported the relationship between sequences from Caribbean and South America. Notably, sequences from Colombia, Venezuela, Brazil, Suriname e Guyana were intermingled with sequences from Trinidad e Tobago, Dominican Republic and Haiti.

To further explore these relationships Bayesian trees were inferred using 263 sequences (SI Table 1), representing North America (n=71), Central America and the Caribbean (n=88) and South America (n=104) under uncorrelated relaxed clock and with a Bayesian Skyline coalescent tree prior. The effective sample size (ESS) was calculated by combining the output of the three runs for each model, excluding 10% steps as a burn-in for each chain. Bayesian MCMC independent runs converged for similar values and all parameter estimates showed ESS values more than 200.

Bayesian genealogies pointed to topologies similar to ML trees, supporting an ancient clade including isolates from different countries of the Caribbean (Figure 1 and SI Figure 1). Sequences from Haiti, Dominican Republic, Trinidad and Tobago, Santa Lucia and St. Vincent were nested together in these deep branches. Interestingly, eight isolates from South American countries, including Suriname, Guyana, Brazil, Colombia and Venezuela and one from the United States were intermingled within this clade. Aside from two clusters of Trinidad and Tobago isolates, no other country showed compartmentalized grouping of sequences. The finding that 52% of the HIV-1B Caribbean sequences could be traced back to a unique most recent common ancestor suggests a single major introduction of HIV-1B from Africa followed by local epidemic spread (SI Table 2). Using an evolutionary time-scale spanning 26 years Bayesian analysis indicates that the HIV-1B epidemic in the Caribbean countries evolved from a common ancestor introduced around 1964 (1950-1967).

The Caribbean clade is followed by three pandemic clusters intermingling sequences from North America, Central America, South America and the Caribbean. The small cluster (pandemic clade A, Figure 1) positioned basal to the pandemic clades is composed of six isolates, including three of Venezuela and one from Colombia indicating that the dispersion of subtype B out of the Caribbean seeded the epidemic of nearby countries. Those sequences occupying more derived positions are grouped into two clusters (pandemic clade B and pandemic clade C) sharing a common ancestor with the pandemic clade A. The pandemic clade B is mainly composed of South American isolates (48%, Si Table 2), including 41% of the total sequences from South American countries. Among this clade those sequences positioned nearer the tree root were also obtained from South American countries suggesting an origin of this clade in that region. Brazilian isolates form a monophyletic cluster and it seems to represent the main evolutionary history of the pandemic in that country since only four Brazilian sequences grouped outside this clade. Similarly twelve isolates from Cuba grouped together with sequences from South America indicating that the Cuban epidemic may have arise in the South American countries.

Within the pandemic cluster B, in a recent time in the evolutionary scale, the grouping of South American sequences with North American isolates, mainly the United States, could indicate an ongoing viral gene flow between the two Americas in both directions. The existence of several independent South American clades was also an indication of separate transmission networks, originated from different introductions of the virus at different times.

Genetic Diversity Analysis:

We used a statistical genetics framework to understand the relationship between the epidemics of HIV-1 subtype B in North, Central and South America. Differences in the degree of genetic diversity among continents evidenced by Φ_{ST} estimations provide further evidence for genetic structuring in the subtype B population of the Americas (Table 1 and Figure 2). The highest level of molecular variation among continents suggested a separation of Central America and North America (Φ_{ST} : 0.04373, $p < 0.00001$). Despite being structured, F_{ST} estimates indicated a closer relationship between Central America and South America (Φ_{ST} : 0.03022, $p < 0.00001$) (Table 1).

In addition, we constructed a multidimensional scaling plot based on F_{ST} from 22 countries (Figure 2). South American countries grouped among a cluster of North America and a cluster primarily composed of countries from Central America (Figure 2). It is interesting to note that South American countries, with the exception of Guyana and Suriname, are more tightly clustered than the isolates from Central America or North America. Guyana and Suriname grouped together with Trinidad and Tobago in the “Central America” cluster. Mexico and the United States seem to be genetically related to the epidemic of Brazil and other South American countries. Finally, Canada, despite being positioned away from other countries of North America still maintains a closer relationship with the United States.

Discussion:

Emergence and dispersion pattern of HIV-1 subtype B from the epicenter of Africa is a major event in the history of the epidemic and a major public health issue. The pandemic has been the focus of several research groups from different areas of human knowledge since it was the first subtype to be isolated in industrialized nations and the first to spread among mobile populations (17, 38-40). In the field of phylogeography, it is subject of a continuing debate. Several suppositions were made about the temporal and geographical patterns, including its origin and diffusion (5, 41-44). Gilbert et al. (2007) fueled the discussions tracing back the beginning of the epidemic out of Africa to 1966 (1962-1970) arising most likely from Haiti in a monophyletic cluster (5). The authors suggested that Haiti was the key conduit for the introduction of subtype B in the United States before it reaches a global dissemination. Our results add another piece to this epidemiological puzzle providing evidence that the spread of HIV-1B in the early 1960s in the Americas is not as unidirectional as initially suggested. Phylogenetic and statistical approaches employed here point to an important participation of South American countries in the transmission and evolution of subtype B epidemic in the Americas.

The Caribbean countries had undoubtedly an important role in the HIV-1 subtype B epidemic. On the *pol* gene phylogeography, a consistent clade of isolates from the Caribbean fell on deep branches suggesting that these countries were the first populations to encounter HIV-1 in the Americas. However we could not state Haiti, as previously suggested (5), or another country as the predecessor of the dissemination due

to the absence of sequences from the early introduction. Together, within the Caribbean clade, Trinidad and Tobago seems to have its epidemic primarily derived from two or more effective HIV-1 introductions, as in all trees obtained most of the sequences of this country form two distinct clades (Figure 1). This result adds to previous studies that used *gag*, *env* and, even the nearly complete genome and found a monophyletic epidemic in that country (5, 19, 45). Additionally, we provide evidence for an introduction of the virus at the beginning of the epidemic in the northern countries of South America, such as Suriname, Guyana, Brazil, Colombia and Venezuela. Due to the presence of only one isolate from the United States in the Caribbean clade our results do not suggest a significant event of direct transmission of the HIV-1B in the Caribbean to the United States early in the epidemic. The direct introduction of the virus in the United States would generate a genetic signature and when dealing with appropriate viral genetic markers the phylogenetic pattern expected under such model, should effectively group a significant number of sequences from the United States dating to the beginning of the epidemic nested within the basal Caribbean strains. Our evolutionary analyses of HIV-1B agreed about the timing events of introduction in Americas (5) estimating the date of the MRCA to 1964 (1950-1967). The clustering of Caribbean strains together in deep branches could be the result of founder effects from one or few introductions in that region at 1960s.

Although the low probability, both methods Bayesian and ML analyses evidenced a cluster derived from the Caribbean clade grouping isolates collected at different times from South American countries (pandemic clade A). This cluster occupies the basal position of the pandemic clades within subtype B (Figure 1) providing evidence for two epidemiological scenarios: (a) the direct introduction of HIV-1B in South America and the seeding of a secondary outbreak in the United States or; (b) the concomitant spread from the Caribbean to South and North America.

Supporting the scenario *a*, the beginning of HIV epidemic in Americas coincides with a boom in oil producing experienced by Venezuela (26, 27). Deepest changes in the economical context implemented policies to attract immigrant people from Colombia, Ecuador, Peru, Cuba, Trinidad and Tobago, Dominican Republic and the United States (26, 27). The population of migrants from other Latin American countries tripled in Venezuela's censuses between 1970 and 1980 (26). It would not be surprising if the circumstances surrounding this movement of people from the Caribbean could

also introduce and disseminate the HIV epidemic in South America. The pandemic clade A and the pandemic clade B (Figure 1) Furthermore, reinforcing our hypothesis, Leal *et al.* (2010) using 66 HIV-1 subtype B near full length genome (8160-bp) of worldwide isolates found that the epidemic in Brazil, a south American country, share a common ancestor with a “North American-European cluster” while Haitian strains occupy the deepest positions (45). The most recent common ancestor for the subtype B epidemic in Brazil (1967 and 95% high posterior probabilities between 1946- 1984) estimated by Bello *et al.* (2007) is only a few years more recent than our estimations to the beginning of the epidemic in Americas. Together, these evidences suggest a link between the Caribbean epidemic and the direct introduction of HIV-1B in the South America.

Yet according to the scenario *a*, after the spreading in countries of South America the virus were introduced in North America. Emigrations from Mexico to the United States have been the largest migratory movement on the planet (26). In this way, Mexico could represent the “entrance door” of the epidemic to the United States. Interestingly, of ten phylogenetic relationships encompassing Mexican sequences through the whole ML tree, eight are linked with South American strains (80%). Bayesian trees included only two Mexican sequences due to the information about the sampling date (SI Table 1).

Finally, the bi-directional spread of HIV-1B from the Caribbean as suggested by the scenario *b* may also have taken part in the implementation of the epidemic in the Americas since two distinct clusters, including the pandemic cluster B, arise from that clade in the Bayesian trees. The historical factors involved in scenario *a* could also be applied to support the scenario *b*. In addition, the historical registers of AIDS cases among Haitian individuals in the 1970s (46) at the United States would be an indication of the entry pathway of the epidemic in North America from the Caribbean. After all, one could also assume a pandemic in the Americas constituted from the co-existence of the two scenarios. In this case, South American countries might have an epidemic originated from virus circulating in the Caribbean countries and from the pandemic expansion at the United States.

The epidemic in Cuba deserves special attention since in addition to the extraordinary genetic diversity and the lower prevalence of HIV in the Americas (47,

48); the country seems to have also a remarkable subtype B epidemic. One clade encompassing 12 Cuban sequences is clearly related to samples of South American countries, suggesting an alternative epidemiological link in the Caribbean region. Such relationship credits the hypothesis that the expansion of the epidemic from the Caribbean countries is not especially directed to the United States as previously suggested (5). After all, historically, South American countries played an important role in the definition of the politics in Cuba (49).

Besides the phylogenetic analyses the degree of differentiation between continent compartments was calculated using AMOVA. Human migrations are a confusing factor in the already complex HIV dissemination. The exchange of people among countries can mix viral subpopulations and mask the historical processes that create patterns of genetic variation and the geographic signatures beyond an epidemic (7). Despite the high rate of migration in the Americas the data presented here demonstrate a significant degree of population structure between continents (Table 1 and Figure 2). Such structuring can have a source in the selection effects of the host's immune system, in the recombination processes and in the founder effects (19); however there is a lack of evidence for selection (19) and all sequences used here met the criteria of no intersubtype genetic mixing. Thus, founder effects that acted in the viral population in the beginning of the epidemic and are still detectable 40 years later seem to better explain the genetic compartmentalization.

The Φ_{ST} estimates testing the viral structuring of the Caribbean, North America and South America revealed that the Caribbean epidemic has a closer genetic relationship to that established in South America (Figure 2), suggesting a epidemiological link between the two Americas. Yet, Φ_{ST} estimates from South American and North American epidemics point to the circulation of genetically related viruses. This low differentiation among regions indicate a substantial similarity among the epidemics in North America and Central America and support the grouping of the pandemic clades mixing sequences of different countries from these regions. Furthermore, in order to gather information on the relationship among countries we conducted a comparison among the Americas including those as subpopulations. The relationship among countries is show in the Figure 2 and explicit some of the results found in our trees. Again, it seems that Central America is genetically related to South American countries than North America. Moreover, aside from Guyana and Suriname

which have epidemics related to that of Trinidad and Tobago, all countries of South America are closely related suggesting an intricate epidemic for that continent. Finally, the role of Mexico in the HIV-1B dispersion throughout the America could be rethought since our phylogenetic analyses and genetic statistics point to a connection among United States, Mexico and South American countries.

We recognize that the *pol* gene may not offer enough genetic variation to ensure a strong phylogenetic signal and thus obtain a high statistical support for the branches of the trees. However, our analysis was able to recover the dating and geographic tracing of previous studies using different genes such as *env* and *gag* (5, 19, 45). Likewise, it was recently demonstrated that this gene is useful to the identification of transmissions events by phylogenetic means (50). Together, *pol* provided the larger set comprising dated sequences sampled in the wider possible number of countries in the Americas. The recovery of ancestral sequences from American countries, especially South American ones, would undoubtedly better trace the spread of the HIV-1B strengthening the occurrence of a scenario or another. Our work do not contradicts previous studies, in fact, through statistical genetics and phylogenetic analyses, our results add one more piece to the history of HIV-1 subtype B in the Americas, revealing that part of the epidemic in South America is directly derived from the Caribbean. We support a scenario that begins with the introduction and spread of the virus locally in the Caribbean region followed by the dispersion of the virus in the northern South America, establishing an epidemic genetically similar to that one in the Caribbean. Due to several migration waves from the various countries of Latin America toward the United States (26), the epidemiological link between South America and North America could be easily established. However, a direct link between the Caribbean and North America also contributed to the dissemination of HIV-1B since historical registers point to this connection (51, 52). The data presented here show a new perspective for addressing the epidemic of HIV-1 subtype B in the Americas. In addition, this work highlights the usefulness of methods of population genetics in understanding the evolution and spread of the epidemic, contributing primarily to the evaluation of interactions between the virus and the migration processes governing the diffusion of human populations.

Acknowledgments:

We thank to Centro Nacional de Supercomputação (CESUP-RS) for allowing access to its computational resources and to the Núcleo de Bioinformática do Laboratório de Imunogenética UFRGS, especially Dinler Antunes to provide us some help in the technical aspects. We are very grateful to Sidia Callegari-Jacques for the contribution in the population genetics approach, to Luciana Tovo Rodrigues for support in the AMOVA estimates, and to Vanessa Rodrigues Paixão-Cortês for useful comments on the manuscript.

References:

1. Osseo-Asare AD (2007) The African Aids Epidemic: A History. *Social History of Medicine* 20:401-402.
2. Salathé M, Jones JH (2010) Dynamics and Control of Diseases in Networks with Community Structure. *PLoS Computational Biology* 6:11.
3. Gray RR et al. (2009) Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *AIDS* 23:F9-F17.
4. Rambaut A, Posada D, Crandall K a, Holmes EC (2004) The causes and consequences of HIV evolution. *Nature reviews. Genetics* 5:52-61.
5. Gilbert MTP et al. (2007) The emergence of HIV/AIDS in the Americas and beyond. *Proceedings of the National Academy of Sciences of the United States of America* 104:18566-70.
6. Hué S, Pillay D, Clewley JP, Pybus OG (2005) Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proceedings of the National Academy of Sciences of the United States of America* 102:4425-9.
7. Perrin L, Kaiser L, Yerly S, Ag CRF (2003) Travel and the spread of HIV-1 genetic variants. *The Lancet* 3:22-27.
8. Guimarães ML et al. (2009) Close phylogenetic relationship between Angolan and Romanian HIV-1 subtype F1 isolates. *Retrovirology* 1958:1-11.
9. Bello G et al. (2008) Origin and evolutionary history of HIV-1 subtype C in Brazil. *AIDS* 22:1993-2000.
10. Worobey M et al. (2008) Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* 455:661-664.
11. Vidal N et al. (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* 74:10498-507.

12. Sharp PM, Hahn BH (2010) The evolution of HIV-1 and the origin of AIDS. *Philosophical Transactions of the Royal Society B* 365:2487-2494.
13. Preston BD, Poiesz BJ, Loeb LA (1988) Fidelity of HIV-1 reverse transcriptase. *Science* 242:1168-1171.
14. Wei X et al. (1995) Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 373:117-122.
15. Gallo RC et al. (1983) Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* 220:865-867.
16. Barré-Sinoussi F et al. (1983) Isolation of a T-lymphotropic retrovirus from patient at risk for AIDS. *Science* 220:868-870.
17. Paraskevis D et al. (2009) Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. *Retrovirology* 6:49.
18. Russell KL et al. (2000) Emerging genetic diversity of HIV-1 in South America. *AIDS* 14:1785-91.
19. Nadai Y et al. (2009) HIV-1 epidemic in the Caribbean is dominated by subtype B. *PloS one* 4:e4814.
20. Hemelaar J, Gouws E, Ghys PD (2006) Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *Aids*:13-23.
21. Castro E et al. (2003) Molecular epidemiology of HIV-1 in Venezuela: high prevalence of HIV-1 subtype B and identification of a B/F recombinant infection. *Journal of acquired immune deficiency syndromes (1999)* 32:338-44.
22. Pinto ME, Schrago CG, Miranda a B, Russo C a M (2008) A molecular study on the evolution of a subtype B variant frequently found in Brazil. *Genetics and molecular research* 7:1031-44.
23. Hierholzer J et al. (2002) Molecular Epidemiology of HIV Type 1 in Ecuador, Peru, Bolivia, Uruguay, and Argentina. *AIDS research and human retroviruses* 18:1339-50.
24. Dalai SC et al. (2009) Evolution and molecular epidemiology of subtype C HIV-1 in Zimbabwe. *AIDS* 23:2523-32.
25. Oliveira T de, Pillay D, Gifford RJ (2010) The HIV-1 subtype C epidemic in South America is linked to the United Kingdom. *PloS one* 5:e9311.
26. Pellegrino A (2000) Trends in international migration in Latin America and the Caribbean. *International Social Science Journal* 52,165:395-408.
27. Durand J (2009) Processes of Migration in Latin America and the Caribbean (1950-2008) Human Development Research Papers, Human Development Report Office, United Nations Development Programme.
28. Solimano A, World Bank. Development Research Group. Macroeconomics and Growth. (2001) *International migration and the global economic order : an overview* (World Bank, Development Research Group, Macroeconomics and Growth)
29. Rousseau CM et al. (2007) Extensive intrasubtype recombination in South African human immunodeficiency virus type 1 subtype C infections. *Journal of virology* 81:4492-500.

30. Lemey P et al. (2004) The molecular population genetics of HIV-1 group O. *Genetics* 167:1059-68.
31. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792-1797.
32. Guindon S, Delsuc F, Dufayard J-F, Gascuel O (2009) Estimating maximum likelihood phylogenies with PhyML. *Methods In Molecular Biology* 537:113-137.
33. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology* 7:214.
34. Rambaut A, Drummond AJ (2003) Tracer [computer program]. Available at: <http://beast.bio.ed.ac.uk/tracer/>.
35. Wright S (1951) The genetical structure of populations. *Annals of Eugenics* 15:323-354.
36. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* 491:479-491.
37. Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* 1:47-50.
38. Parekh B et al. (1999) Impact of HIV type 1 subtype variation on viral RNA quantitation. *AIDS research and human retroviruses* 15:133-42.
39. Hu DJ, Buvé a, Baggs J, Groen G van der, Dondero TJ (1999) What role does HIV-1 subtype play in transmission and pathogenesis? An epidemiological perspective. *AIDS* 13:873-81.
40. Chalmet K et al. (2010) Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. *BMC infectious diseases* 10:262.
41. Robbins KE et al. (2003) U.S. Human Immunodeficiency Virus Type 1 Epidemic : Date of Origin , Population History , and Characterization of Early Strains. *Journal of Virology* 77:6359-6366.
42. Holmes EC (2007) When HIV spread afar. *Proceedings of the National Academy of Sciences USA* 104:18351-18352.
43. Lukashov VV, Goudsmit J (2002) Recent Evolutionary History of Human Immunodeficiency Virus Type 1 Subtype B: Reconstruction of Epidemic Onset Based on Sequence Distances to the Common Ancestor. *Journal of Molecular Evolution* 54:680-691.
44. Korber B et al. (2000) Timing the ancestor of the HIV-1 pandemic strains. *Science* 288:1789-1796.
45. Leal É, VillaNova F (2010) Diversity of HIV-1 Subtype B: Implications to the Origin of BF Recombinants. *Plos one* 5(7) e11833.
46. Pape J, Johnson WD Jr (1993) AIDS in Haiti: 1982-1992. *Clinical Infectious Diseases* 17 suppl 2: S341-5.
47. Cuevas MT et al. (2002) High HIV-1 genetic diversity in Cuba. *AIDS* 16:1643-53.
48. Hiv T, Hiv G (2000) Report on the Global HIV/AIDS UNAIDS, 2000.

49. Bethell L, Zoumaras T (1993) Cuba: a short history. *History Reviews Of New Books* 22:117–117.
50. Clewley JP, Cane PA, Pillay D (2004) HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. *Aids* 18(5):719-28.
51. Pitchenik AE, et al. (1983) Opportunistic infections and Kaposi's sarcoma among Haitians: Evidence of a new acquired immunodeficiency state. *Annals of Internal Medicine* 98:277–284.
52. Liautaud B, Laroche C, Duvivier J, Pean-Guichard C (1983) Kaposi's sarcoma in Haiti: unknown reservoir or a recent appearance? *Annales de dermatologie et de venerologie* 110:213-219.

Legends:

Figure 1. Majority-rule Bayesian consensus tree of 268 HIV-1 subtype B *pol* sequences isolated from 25 countries in the Americas. Out-grouped are the subtype C and D. Branches are colored due to sample origin. Orange branches include isolates from Central America (including the Caribbean); green branches include South American isolates and blue branches, isolates from North America. Branches are not drawn to scale. Posterior probabilities are shown for key nodes. Tips of the tree contain isolate information regarding the subtype, country, year of isolation and GenBank accession number. Abbreviations of countries are as follows: AR, Argentina; AG, Antigua and Barbuda; BH, Bahamas; BR, Brazil; BW, Botswana; CA, Canada; CD, Democratic Congo; CM, Cameroon; CO, Colombia; CU, Cuba; DO, Dominican Republic; EC, Ecuador; ET, Ethiopia; GD, Grenada; GY, Guyana; HT, Haiti; JM, Jamaica; LC, Santa Lucia; SR, Suriname; TT, Trinidad and Tobago, TZ, Tanzania; UG, Uganda; US, United States; VC, St. Vincent; VE, Venezuela.

Figure 2. Synthetic map illustrating the distribution and geographic origin of strains isolated in the Americas and the genetical structuring among continents and countries. **(a)** Countries of sample isolation are colored-divided by geopolitical continents comprising South America, Central America (including Caribbean) and North America. Isolates from the gray-colored countries were not included in this study. Countries located in Central America are represented by numbers: (1) Bahamas; (2) Puerto Rico; (3) Antigua and Barbuda; (4) Santa Lucia; (5) St. Vincent; (6) Grenada; (7) Jamaica; (8) Honduras and (9) Panama. Red dotted lines represent F_{ST} estimates between continents. South America is genetically intermediary related to Central America and North America. **(b)** Non-metric multidimensional scaling plot of the F_{ST} estimates among 22 South American countries. Dimension 1 separates the 308 isolates by country.

SI Figure 1. Part of the 50% majority rule consensus tree from the Bayesian MCMC (BEAST). This is the full version of the collapsed group in the Figure 1 showing the evolutionary relationships among the sequences in the pandemic clade. Orange branches include isolates from Central America (including the Caribbean); green branches include South American isolates and blue branches, isolates from North America. Branches are not drawn to scale. Abbreviations of countries are as follows: AR, Argentina; AG, Antigua and Barbuda; BH, Bahamas; BR, Brazil; BW, Botswana; CA, Canada; CD, Democratic Congo; CM, Cameroon; CO, Colombia; CU, Cuba; DO, Dominican Republic; EC, Ecuador; ET, Ethiopia; GD, Grenada; GY, Guyana; HT, Haiti; JM, Jamaica; LC, Santa Lucia; SR, Suriname; TT, Trinidad and Tobago, TZ, Tanzania; UG, Uganda; US, United States; VC, St. Vincent; VE, Venezuela.



Tables:

Table 1. Analysis of molecular variance (AMOVA) of HIV-1 subtype B isolates from North, Central and South Americas.

Division (number of populations tested)	Source of Variation	Fixation index
North America (3) <i>vs.</i> Central America (11) <i>vs.</i> South America (8)	Among divisions	Φ_{ct} : 0.00663
	Among populations within divisions	Φ_{sc} : 0.06225*
	within populations	Φ_{st} : 0.06847*
North America <i>vs.</i> Central America <i>vs.</i> South America	Among divisions	Φ_{st} : 0.02457*

* $p < 0.05$

Supporting Information (SI):

Table 1. Geographic origin, time spanning and the number of sequences used to trace back the dissemination of HIV-1 subtype B in the Americas.

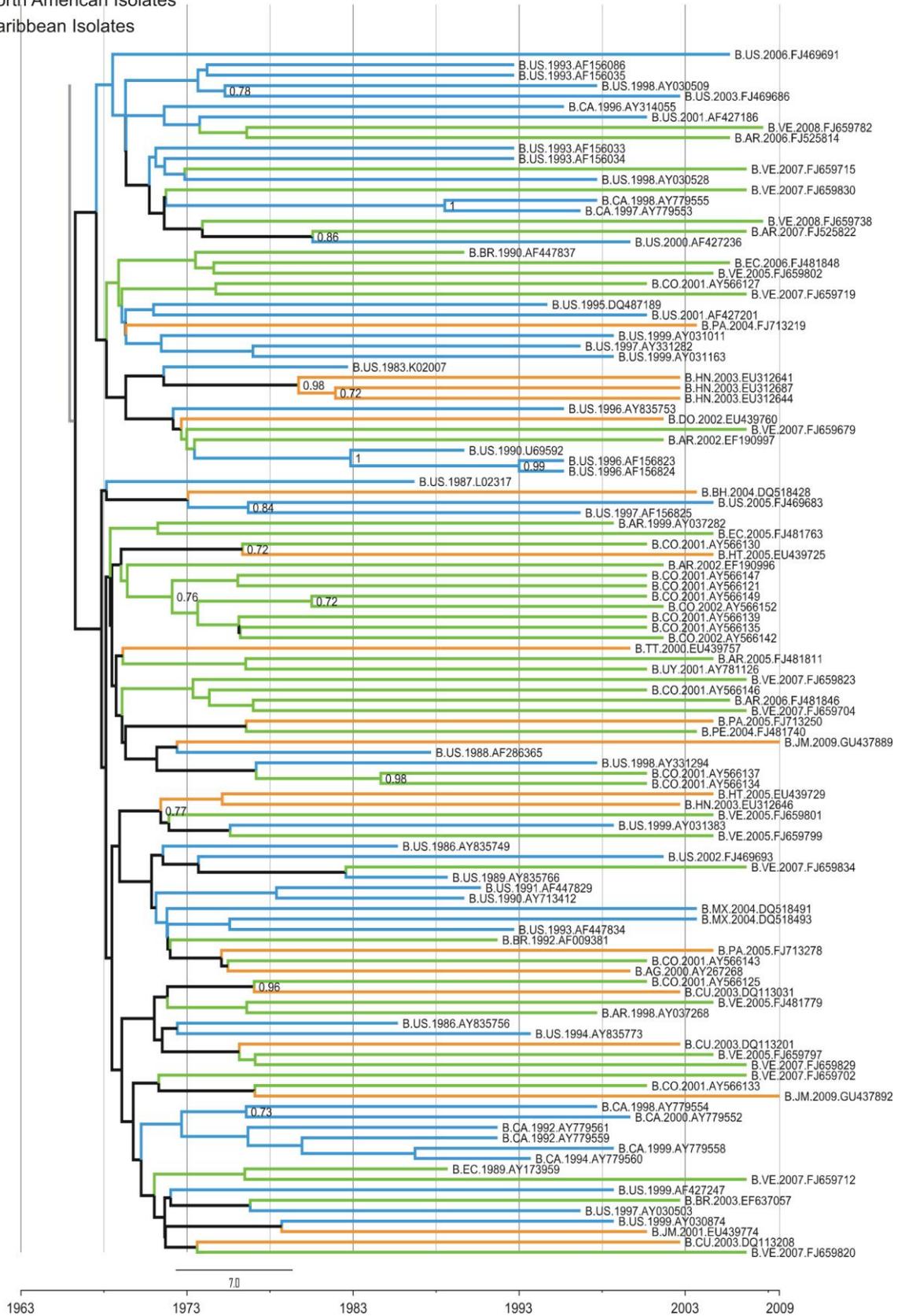
	Country	Region	Number of Sequences		Total	Time frame
			dated	undated		
1.	Antigua and Barbuda	Caribbean	3	2	5	2000
2.	Argentina	South America	9	4	13	1998-2007
3.	Bahamas	Caribbean	4	-	4	2004-2005
4.	Bolivia	South America	1	-	1	1999
5.	Brazil	South America	22	-	22	1990-2005
6.	Canada	North America	9	-	9	1992-2000
7.	Chile	South America	-	25	25	-
8.	Colombia	South America	27	-	27	2001-2006
9.	Cuba	Caribbean	19	-	19	1999-2003
10.	Dominican Republic	Caribbean	12	-	12	2000-2005
11.	Ecuador	South America	5	-	5	1989-2006
12.	Grenada	Caribbean	3	-	3	2000
13.	Guyana	South America	1	-	1	2000
14.	Honduras	Central America	4	-	4	2003
15.	Haiti	Caribbean	8	-	8	2004-2005
16.	Jamaica	Caribbean	7	-	7	2001-2009
17.	Santa Lucia	Caribbean	1	-	1	2000
18.	Mexico	North America	2	8	10	2004
19.	Panama	Central America	4	-	4	2004-2005
20.	Peru	South America	1	-	1	2004
21.	Puerto Rico	Caribbean	-	1	1	-
22.	Suriname	South America	4	-	4	2000
23.	Trinidad and Tobago	Caribbean	19	-	19	2000-2001
24.	United States	North America	60	-	60	1983-2006
25.	Uruguay	South America	2	-	2	2001
26.	St. Vincent	Caribbean	4	-	4	2000
27.	Venezuela	South America	32	10	42	2004-2008
	-	-	263	50	313	1983-2009

Table 2. Number of sequences grouped within each clade of the HIV-1 subtype B Bayesian phylogeny and the percentage of the sequences by the sampled region considering 71 sequences from North America, 08 sequences from Central America, 80 sequences from Caribbean and 104 from South America.

Clade (number of sequences)	Region	Number of sequences (% of the clade)	Number of sequences (% of the region)
Caribbean Clade (51 sequences)	North America	1/51 (1.96%)	1/71 (1.41%)
	Central America	0/51 (0%)	0/08 (0%)
	Caribbean	42/51 (82.35%)	42/80 (52.5%)
	South America	8/51 (15.69%)	8/104 (7.69%)
Pandemic Clade A (06 sequences)	North America	1/06 (16.67%)	1/71 (1.41%)
	Central America	0/06 (0%)	0/08 (0%)
	Caribbean	1/06 (16.67%)	1/80 (1.25%)
Pandemic Clade B (90 sequences)	South America	4/06 (66.66%)	4/104 (3.85%)
	North America	21/90 (23.33%)	21/71 (29.58%)
	Central America	1/90 (1.11%)	1/08 (12.5)
	Caribbean	25/90 (27.78%)	25/80 (31.25%)
Pandemic Clade C (116 sequences)	South America	43/90 (47.78%)	43/104 (41.35%)
	North America	48/116 (41.38%)	45/71 (67.6%)
	Central America	7/116 (6%)	7/08 (87.5%)
	Caribbean	12/116 (10.35%)	12/80 (15%)
	South America	49/116 (42.24%)	49/104 (47.11%)

SI Figure 1.

- South American Isolates
- North American Isolates
- Caribbean Isolates



Capítulo 4: Discussão

.....

Os resultados descritos neste trabalho indicam que parte da atual epidemia do subtipo B na América do Sul ocorreu a partir de introduções de vírus circulantes em países do Caribe. A concordância de dois métodos de construção de árvores filogenéticas (ML e IB) em conjunto com os métodos estatísticos de análise da população viral fornece robustez às nossas conclusões destacando uma nova perspectiva para o entendimento da dispersão e estabelecimento da pandemia nas Américas.

Neste sentido, o Caribe tem um papel importante na história epidemiológica do subtipo B. Concordando com trabalhos anteriores, nossos dados sugerem uma única entrada do vírus na região por volta de 1964 (1950-1967). Como esperado, a proximidade geográfica e cultural levaram o vírus a se disseminar por diversos países da região, incluindo Trinidad e Tobago, República Dominicana, Haiti, São Vicente e Santa Lucia. Nossos resultados não reproduzem os achados de trabalhos anteriores que sugerem uma entrada inicial do vírus no Haiti (Gilbert et al., 2007). A indisponibilidade de isolados que datam do início da epidemia de países do Caribe em bancos de dados públicos impossibilitou a utilização de sequências mais antigas que 2004. Os sinais filogenéticos de sequências antigas permitem um melhor ajuste na calibração do relógio molecular utilizado na análise Bayesiana e podem influenciar na organização topológica das filogenias (Bromham e David, 2003).

A circulação do vírus em diversos países do Caribe foi provavelmente suficiente para alimentar uma introdução direta em alguns países do norte da América do Sul, como Suriname, Guiana, Venezuela e Colômbia. Análises clássicas de genética estatística, utilizando as estimativas F , também revelaram esta relação. A presença de um grupo formado por sequências isoladas na América do Sul, emergindo do clado basal do Caribe, antes do *cluster* que compreende ao clado pandêmico, alimenta a

formação de duas hipóteses principais. Primeiramente, o vírus após circular no Caribe, pode ter sido introduzido na América do Sul e depois levado à América do Norte. Após o estabelecimento da epidemia nos Estados Unidos, o vírus pode ter sido re-introduzido na América do Sul. Em uma segunda hipótese, haveria uma co-dispersão para a América do Norte e para a América do Sul a partir da epidemia do Caribe.

Considerando os aspectos geográficos, a hipótese de co-dispersão parece mais provável. O clado formado principalmente por sequências sul americanas, neste caso, não seria um clado basal ao pandêmico, mas sim um clado monofilético de introdução do vírus a partir do Caribe. Ainda para esta hipótese, a saída do vírus do Caribe (e a efetiva expansão) teria acontecido, pelo menos, duas vezes: uma para a América do Sul e outra, provavelmente para a América do Norte. Este cenário também é sustentado pelas análises de AMOVA, que sugerem uma estruturação populacional diferente para a América do Sul e a América do Norte.

Em um trabalho recente, Leal e colaboradores (2010) mostram que a epidemia do Brasil emerge da epidemia corrente no Caribe e compartilha um ancestral comum com os vírus circulantes na América do Norte e na Europa (Figura 4.1). A monofilia e a separação do clado “Estados Unidos – Europa” e Brasil sustenta a hipótese de que a epidemia do subtipo B foi diretamente introduzida na América do Sul a partir do Caribe, sem a intermediação da América do Norte. Além disso, a análise da situação histórica nas Américas, na década de 60, indica a ocorrência de um grande processo migratório de diversos países da América Latina em direção à Venezuela. Dados de pesquisadores venezuelanos, notificam a detecção de três casos de HIV em índios amazonenses, cujo sangue total foi coletado em 1968 (Rodríguez et al. 1985).

O Brasil, em meio a este cenário, parece ter sido um importante ponto de disseminação da epidemia, já que diversos isolados deste país agrupam dentro do clado sul-americano derivado do Caribe. Este clado abriga a maioria dos isolados brasileiros, evidenciando uma possível entrada monofilética no país. Ainda assim, não é possível descartar a introdução do vírus por outras origens, dado que isolados do mesmo país ocupam outras posições da árvore. Curiosamente, a epidemia do subtipo B no Brasil é caracterizada por uma variante denominada B'. Esta variante representa em torno de 40-60% do subtipo B circulante no país e é raro (<1%) em países da América do Norte e da Europa (Bongertz et al. 2000; Potts et al. 1993; Bello, Eyer-Silva, José C Couto-

Fernandez, Monick L Guimarães, Chequer-Fernandez, Teixeira, and Mariza G Morgado 2007b). De forma especulativa, extrema peculiaridade pode ser atribuída a uma ação do efeito fundador após a entrada única do vírus no Brasil a partir do Caribe.

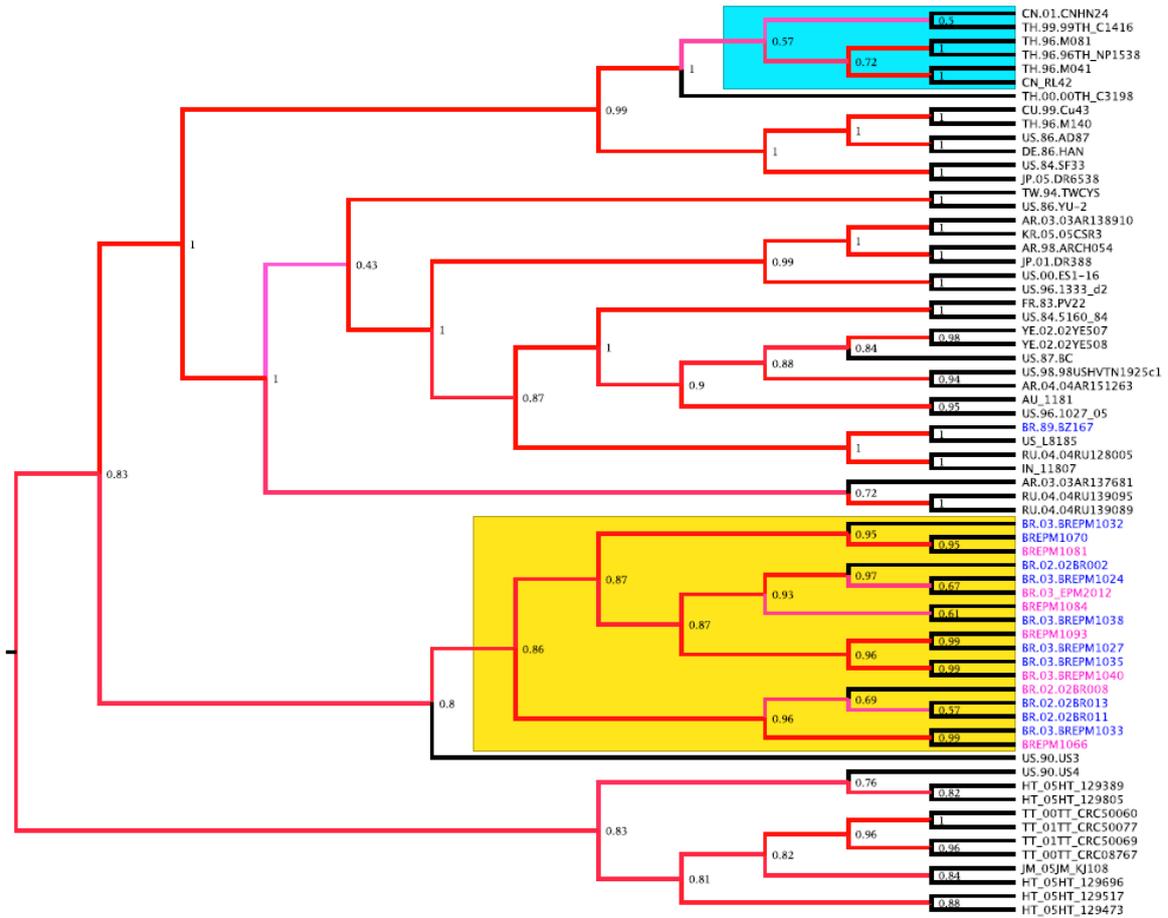


Figura 4.1: Árvore Bayesianas extraída de Leal et al. (2010) utilizando 66 sequências do HIV-1 subtipo B (8160 pares de base) isoladas de diversas partes do mundo. A área amarela marca o clado formado por sequências isoladas no Brasil. O nome dos isolados indica o país de coleta, o ano e a informação a respeito da identificação da sequência.

Nossos achados envolvendo a epidemia do subtipo B em Cuba reiteram a idéia de que o Caribe alimentou a epidemia da América do Sul e revelam que a população da América do Sul também “exportou” o vírus. Nossos resultados são pioneiros, visto que análises sobre origem e disseminação do HIV-1 nas Américas são ainda bastante escassos.

As estimativas de estatística F revelaram uma estruturação populacional importante na epidemia do subtipo B nas Américas e, de certa maneira, corroboram com as análises filogenéticas. Apesar de ser incomum o uso de estatística genética em

trabalhos de disseminação de HIV, estas estimativas foram importantes para avaliar e clarear o entendimento das relações entre as epidemias dos diferentes países. A estruturação populacional foi capaz de separar cada uma das Américas em grupos geneticamente distintos e ainda revelou uma situação genética próxima entre as epidemias da América do Sul e do Caribe. Os três países que compõem a América do Norte provavelmente abrigam epidemias geneticamente distantes àquela do Caribe. Este resultado pode ter sido mascarado pelo grande movimento populacional turístico que ocorre na região, principalmente nos Estados Unidos e Canadá (Pellegrino, 2000).

Trabalhos como esse podem ser de fundamental interesse para o monitoramento da disseminação de variantes virais resistentes a medicamentos antirretrovirais e podem servir como um modelo para o estudo da difusão de outras doenças infecto-contagiosas. Perspectivamente, estudos sobre a “interação” e estruturação entre epidemias de países vizinhos podem ajudar na avaliação da influência das bordas geográficas na dinâmica global da pandemia. Da mesma forma, futuros trabalhos, utilizando sequências do início da epidemia e envolvendo um número maior de países serão necessários para rever o papel da América do Sul e definir o real caminho da disseminação do HIV-1B dentro das Américas.

A difusão em massa do HIV-1 subtipo B em países desenvolvidos marca o início de um importante evento na história da saúde mundial e na história da humanidade. Evidências diversas demonstram que a cooperação entre países e os investimentos em resposta à epidemia do HIV podem gerar claras reduções das alarmantes estatísticas envolvendo morte de pessoas no mundo todo. Além destes, a participação definitiva dos setores de pesquisa contribuiu fundamentalmente para o esclarecimento, tratamento e aumento da expectativa de vida de indivíduos contaminados com o vírus (UNAIDS, 2010).

REFERÊNCIAS:



- Arhel, Nathalie. 2010. "Revisiting HIV-1 uncoating." *Retrovirology* 7(1): 96.
- Ariën, Kevin K, Guido Vanham, and Eric J Arts. 2007. "Is HIV-1 evolving to a less virulent form in humans?" *Nature reviews. Microbiology* 5(2): 141-51.
- Ashorn, P et al. 1990. "An inhibitor of the protease blocks maturation of human and simian immunodeficiency viruses and spread of infection." *Proceedings of the National Academy of Sciences of the United States of America* 9(2): 7472-7476.
- Aulicino, Paula C et al. 2005. "Circulation of novel HIV type 1 A, B/C, and F subtypes in Argentina." *AIDS research and human retroviruses* 21(2): 158-64.
- Bailes, Elizabeth et al. 2003. "Hybrid origin of SIV in chimpanzees." *Science* 300(5626): 1713.
- Barry, M, J Mellors, and F Bia. 1984. "Haiti and the AIDS connection." *Journal Of Chronic Diseases* 37(7): 593-595.
- Barré-Sinoussi, F et al. 1983. "Isolation of a T-lymphotropic retrovirus from patient at risk for AIDS." *Science* 220: 868-870.
- Bello, Gonzalo, Monick L Guimarães, and Mariza G Morgado. 2006. "Evolutionary history of HIV-1 subtype B and F infections in Brazil." *AIDS* 20(5): 763-8.
- Bello, Gonzalo et al. 2007. "Demographic history of HIV-1 subtypes B and F in Brazil." *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 7(2): 263-70.
- Bello, Gonzalo et al. 2008. "Origin and evolutionary history of HIV-1 subtype C in Brazil." *Aids*.
- Bello, Gonzalo et al. 2010. "Phylogenetics of HIV-1 Circulating Recombinant Forms 12 _ BF and 38 _ BF in Argentina and Uruguay." *Retrovirology*: 1-9.
- Berger, EA, Philip M Murphy, and Joshua M Farber. 1999. "Chemokine receptors as HIV-1 coreceptors: Roles in Viral Entry , Tropism , and Disease." *Allergy*: 657-700.
- Bongertz, V et al. 2000. "HIV-1 diversity in Brazil: genetic, biologic, and immunologic characterization of HIV-1 strains in three potential HIV vaccine evaluation sites. Brazilian Network for HIV Isolation and Characterization." *Journal of acquired immune deficiency syndromes (1999)* 23(2): 184-93.
- Bromham, Lindell, and David Penny. 2003. "The modern molecular clock." *Nature Reviews Genetics* 4(3): 216-224.
- Brumme, ZL et al. 2005. "Molecular and clinical epidemiology of CXCR4-using HIV-1 in a large population of antiretroviral-naive individuals." *The Journal of Infectious Diseases* 192(3): 466-474.
- Bryant, M, and L Ratner. 1990. "Myristoylation-dependent replication and assembly of human immunodeficiency virus 1." *Proceedings of the National Academy of Sciences of the United States of America* 87(2): 523-7.

- Brígido, LFM et al. 2007. "HIV type 1 subtype C and CB Pol recombinants prevail at the cities with the highest AIDS prevalence rate in Brazil." *AIDS Research and Human Retroviruses* 23(12): 1579-1586.
- Bukrinsky, Michael. 2004. "A hard way to the nucleus." *Molecular medicine Cambridge* 10(1-6): 1-5.
- Bushman, F D, T Fujiwara, and R Craigie. 1990. "Retroviral DNA integration directed by HIV integration protein in vitro." *Science* 249(4976): 1555-1558.
- Cantin, Réjean, Sylvie Méthot, and Michel J Tremblay. 2005. "Plunder and stowaways: incorporation of cellular proteins by enveloped viruses." *Journal of Virology* 79(11): 6577-6587.
- Carrion, G et al. 2003. "Circulating recombinant form CRF02_AG in South America." *AIDS research and human retroviruses* 19(4): 329-32.
- Casado, C et al. 2000. "Different distribution of HIV type 1 genetic variants in European patients with distinct risk practices." *AIDS Research and Human Retroviruses* 16(3): 299-304.
- Castro, Erika et al. 2003. "Molecular epidemiology of HIV-1 in Venezuela: high prevalence of HIV-1 subtype B and identification of a B/F recombinant infection." *Journal of acquired immune deficiency syndromes (1999)* 32(3): 338-44.
- Chalmet, Kristen et al. 2010. "Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections." *BMC Infectious Diseases* 10(1): 262.
- Chan, D C et al. 1997. "Core structure of gp41 from the HIV envelope glycoprotein." *Cell* 89(2): 263-73.
- Chang, S Y et al. 1993. "The origin of HIV-1 isolate HTLV-IIIB." *Nature* 363(6428): 466-9.
- Charneau, P et al. 1994. "HIV-1 reverse transcription. A termination step at the center of the genome." *Journal of Molecular Biology* 241(5): 651-662.
- Chermann, J C et al. 2007. "Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS)." *Science* 220(4599): 868-871.
- Clavel, F. et al. 1986 " Isolation of a new human retrovirus from West African patients with AIDS " *Science* 233(4761):343-6.
- Cloyd, Miles W, JJ-y Chen, and L Wang. 2000. "How does HIV cause AIDS ? The homing theory." *Molecular Medicine* 6: 4-7.
- Coakley, Eoin, Christos J Petropoulos, and JM Whitcomb. 2005. "Assessing chemokine co-receptor usage in HIV." *Current Opinion in Infectious Diseases* 18(1): 9-15.
- Coffin, J et al. 1986. "What to call the AIDS virus?" *Nature* 321(6065): 10.
- Coffin, JM et al. 1991. "Retroviridae and Their Replication." In Raven Press Ltd., p. 1437-1500.
- Cohen, J. 1995. "Differences in HIV strains may underlie disease patterns." *Science* 270(5233): 30-31.
- Corbet, S et al. 2000. "env sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area." *Journal of virology* 74(1): 529-34.
- Costin, Joshua M. 2007. "Cytopathic mechanisms of HIV-1." *Virology journal* 4: 100.

- Crandall, Keith A, and David Posada. 2002. "Phylogenetic Approaches to Molecular Epidemiology in The molecular epidemiology of Human viruses Thomas Leitner, ed. Kluwer Academic Publishers, Boston.
- Dalai, Sudeb C et al. 2009. "Evolution and molecular epidemiology of subtype C HIV-1 in Zimbabwe." *AIDS* 23(18): 2523-32.
- Deng, X et al. 2008. "The epidemic origin and molecular properties of B': a founder strain of the HIV-1 transmission in Asia." *AIDS* 22(14): 1851-1858.
- Di Mascio, Michele et al. 2004. "Evolution of the V3 envelope domain in proviral sequences and isolates of human immunodeficiency virus type 1 during transition of the viral biological phenotype." *Journal of Virology* 78(19): 10566-10573.
- Diaz, Ricardo Sobhie et al. 2008. "Selective regimes and evolutionary rates of HIV-1 subtype B V3 variants in the Brazilian epidemic." *Virology* 381(2): 184-93.
- Drummond, Alexei J et al. 2006. "Relaxed Phylogenetics and Dating with Confidence" *PLoS Biology* 4(5): e88.
- Drummond, Alexei J, and Andrew Rambaut. 2007. "BEAST: Bayesian evolutionary analysis by sampling trees." *BMC evolutionary biology* 7: 214.
- Edwards, A W F. 1996. "The origin and early development of the method of minimum evolution for the reconstruction of phylogenetic trees." *Systematic Biology* 45(1): 79-91.
- Esteves, A et al. 2000. "Genetic characterization of HIV type 1 and type 2 from Bissau, Guinea-Bissau (West Africa)." *Virus Research* 68(1): 51-61.
- Evans, B G et al. 1991. "HIV-2 in the United Kingdom--a review." *CDR London England Review* 1(2): R19-R23.
- Fanale-Belasio, E. et al. 2010. "HIV virology and pathogenetic mechanisms of infection: a brief overview." *Annali dell'Istituto Superiore di Sanità* 46: 5-14.
- Farmer, Paul. 1992. *AIDS and Accusation: Haiti and the Geography of Blame*. University of California Press.
- Felber, B K, C M Drysdale, and G N Pavlakis. 1990. "Feedback regulation of human immunodeficiency virus type 1 expression by the Rev protein." *Journal of Virology* 64(8): 3734-3741.
- Felsenstein, J. 1978. "Cases in which parsimony or compatibility methods will be positively misleading." *Systematic Zoology* 27(4): 401-410.
- Felsenstein, Joseph. 1981. "Evolutionary Trees from DNA Sequences : A Maximum Likelihood Approach." *Journal of Molecular Evolution*: 368-376.
- Fontella, Rachel, MA Soares, and Carlos G Schrago. 2008. "On the origin of HIV-1 subtype C in South America." *Aids*.
- Gallo, RC et al. 1983. "Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS)." *Science* 220(4599): 865-867.
- Gao, F et al. 1999. "Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*." *Nature* 397(6718): 436-41.

- Gao, F et al. 1994. "Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology." *Journal of Virology* 68(11): 7433-7447.
- Geretti, Anna Maria. 2006. "HIV-1 subtypes: epidemiology and significance for HIV management." *Current Opinion in Infectious Diseases* 19(1): 1-7.
- Geretti, Anna Maria et al. 2009. "Effect of HIV-1 subtype on virologic and immunologic response to starting highly active antiretroviral therapy." *Clinical Infectious Diseases* 48(9): 1296-1305.
- Gilbert, MTP et al. 2007. "The emergence of HIV/AIDS in the Americas and beyond." *Proceedings of the National Academy of Sciences of the United States of America* 104(47): 18566-70.
- Gottlieb, Michael S. 2006. "Pneumocystis pneumonia--Los Angeles, 1981." *American Journal of Public Health* 96(6): 980-981.
- Grant, Alison D, and Kevin M De Cock. 2001. "ABC of AIDS. HIV infection and AIDS in the developing world." *BMJ British Medical Journal* 322(7300): 1475-1478.
- Gray, RR et al. 2009. "Spatial phylodynamics of HIV-1 epidemic emergence in east Africa." *Aids*.
- Greco, RS. 1983. "Haiti and the stigma of AIDS." *Lancet* 2(8348): 515-516.
- Guimarães, Monick L et al. 2009. "Close phylogenetic relationship between Angolan and Romanian HIV-1 subtype F1 isolates." *Retrovirology* 6: 39.
- Göttlinger, H G, J G Sodroski, and W A Haseltine. 1989. "Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1." *Proceedings of the National Academy of Sciences of the United States of America* 86(15): 5781-5785.
- Hahn, B H et al. 2000. "AIDS as a zoonosis: scientific and public health implications." *Science* 287(5453): 607-14.
- Harvey, P H et al. 1996. *New Uses for New Phylogenies*. eds. P H Harvey et al. Oxford University Press.
- Heeney, J L, A G Dalgleish, and RA Weiss. 2006. "Origins of HIV and the Evolution of Resistance to AIDS." *Science* 313(5786): 462-466.
- Hemelaar, Joris, Eleanor Gouws, and Peter D Ghys. 2006. "Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004." *Aids*: 13-23.
- Hladik, Florian, and M Juliana McElrath. 2008. "Setting the stage: host invasion by HIV." *Nature reviews. Immunology* 8(6): 447-57.
- Ho, David D, and Yaoping Huang. 2002. "The HIV-1 vaccine race." *Cell* 110(2): 135-138.
- Hofacker, I L et al. 1998. "Automatic detection of conserved RNA structure elements in complete RNA virus genomes." *Nucleic Acids Research* 26(16): 3825-3836.
- Hoffman, A D, B Banapour, and J A Levy. 1985. "Characterization of the AIDS-associated retrovirus reverse transcriptase and optimal conditions for its detection in virions." *Virology* 147(2): 326-335.
- Holder, Mark, and PO Lewis. 2003. "Phylogeny estimation: traditional and Bayesian approaches." *Nature reviews. Genetics* 4(4): 275-84.

- Huelsenbeck, J P, and J P Bollback. 2001. "Application of the likelihood function in phylogenetic analysis." In *Handbook of Statistical Genetics*, eds. D Balding, M Bishop, and C Cannings. Wiley, p. 415-443.
- Hymes, K B et al. 1981. "Kaposi's sarcoma in homosexual men-a report of eight cases." *Lancet* 2(8247): 598-600.
- Jacks, T et al. 1988. "Characterization of ribosomal frameshifting in HIV-1 gag-pol expression." *Nature* 331(6153): 280-283.
- Jill Harrison, Jane A Langdale. 2006. "A step by step guide to phylogeny reconstruction." *The Plant Journal* 45(4): 561-572.
- John-Stewart, Grace C et al. 2005. "Subtype C Is associated with increased vaginal shedding of HIV-1." *The Journal of Infectious Diseases* 192(3): 492-496.
- Kalish, Marcia L et al. 2005. "Central African hunters exposed to simian immunodeficiency virus." *Emerging Infectious Diseases* 11(12): 1928-1930.
- Kandathil, A J et al. 2005. "Molecular epidemiology of HIV." *The Indian journal of medical research* 121(4): 333-44.
- Kanki, Phyllis J et al. 1985. "Human Immunodeficiency Virus Type 1 Subtypes Differ in Disease Progression." *The Journal of Infectious Diseases* 179:68-73
- Keele, Brandon F et al. 2006. "Chimpanzee reservoirs of pandemic and nonpandemic HIV-1." *Science* 313(5786): 523-526.
- Keng, Kok et al. 2009. "Estimating the date of origin of an HIV-1 circulating recombinant form." *Virology* 387(1): 229-234.
- Kim, SY et al. 1989. "Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression." *Journal of Virology* 63(9): 3708-3713.
- Koot, M et al. 1996. "Relation between changes in cellular load, evolution of viral phenotype, and the clonal composition of virus populations in the course of human immunodeficiency virus type 1 infection." *The Journal of Infectious Diseases* 173(2): 349-354.
- Kosakovsky Pond, Sergei L, and DM Smith. 2009. "Are all subtypes created equal? The effectiveness of antiretroviral therapy against non-subtype B HIV-1." *Clinical Infectious Diseases* 48(9): 1306-1309.
- Ks, S et al. 2002. "A History of HIV Discovery The Early Years of HIV / AIDS." *Science* 298:1727-1728.
- Kumar, Sudhir. 2005. "Molecular clocks: four decades of evolution." *Nature Reviews Genetics* 6(8): 654-662.
- Leal, É, and Fe VillaNova. 2010. "Diversity of HIV-1 Subtype B : Implications to the Origin of BF Recombinants." *Plos One* 5(7): 1-9.
- Leal, Élcio et al. 2007. "Spread of HIV-1 BF and CB recombinants in South America." *AIDS Research and Human Retroviruses*: 52841-52841.
- Lemey, Philippe et al. 2003. "Tracing the origin and history of the HIV-2 epidemic." *Proceedings of the National Academy of Sciences of the United States of America* 100(11): 6588.

- Lemey, Philippe et al. 2004. "The molecular population genetics of HIV-1 group O." *Genetics* 167(3): 1059-68.
- Lemey, Philippe, Andrew Rambaut, and Oliver G Pybus. 2006. "HIV evolutionary dynamics within and among hosts." *AIDS reviews* 8(3): 125-40.
- Levy, J A. 1990. "The pathogenesis of HIV infection." *Hospital practice Office ed* 25(11): 41-48.
- Lewis, P. 2001. "Phylogenetic systematics turns over a new leaf." *Trends in ecology & evolution (Personal edition)* 16(1): 30-37.
- Li, WH, M Tanimura, and P M Sharp. 1988. "Rates and dates of divergence between AIDS virus nucleotide sequences." *Molecular biology and evolution* 5(4): 313-30.
- Li, Y et al. 2010. "Explosive HIV-1 subtype B' epidemics in Asia driven by geographic and risk group founder events." *Virology* 402(2):223-227.
- Liao, Huanan et al. 2009. "Phylogenetic analysis of the dissemination of HIV-1 CRF01_AE in Vietnam." *Virology*: 5-10.
- Liautaud, B et al. 1983. "Kaposi's sarcoma in Haiti: unknown reservoir or a recent appearance?" *Annales de dermatologie et de venerologie* 110(3): 213-219.
- Liebert, Mary Ann et al. 2002. "Molecular Epidemiology of HIV Type 1 in Ecuador, Peru, Bolivia, Uruguay, and Argentina." *AIDS Research and Human Retroviruses* 18(18): 1339-1350.
- Liebert, Mary Ann et al. 2005. "Circulation of Novel HIV Type 1 A , B / C , and F Subtypes in Argentina." *AIDS Research and Human Retroviruses* 21(2): 158-164.
- Mahalingam, Sundarasamy et al. 2001. "Functional Analysis of the Simian Immunodeficiency Virus Vpx Protein : Identification of Packaging Determinants and a Novel Nuclear Targeting Domain." *Journal of Virology* 75(1): 362-374.
- Marx, JL. 1985. "A virus by any other name . . ." *Science* 227(4693): 1449-1451.
- McDade, J E, and B E Anderson. 1996. "Molecular epidemiology: applications of nucleic acid amplification and sequence analysis." *Epidemiologic Reviews* 18(1): 90-97.
- Mehta, Sanjay R et al. 2010. "The relatedness of HIV epidemics in the United States-Mexico border region." *AIDS research and human retroviruses* 26(12): 1273-7.
- Melbye, M et al. 1984. "Seroepidemiology of HTLV-III antibody in Danish homosexual men: prevalence, transmission, and disease outcome." *British medical journal Clinical research ed* 289(6445): 573-575.
- Miyaki, Cristina Yumi, Cláudia A De Moraes Russo, and Sergio Luiz Pereira. 2001. "Reconstrução filogenética. Introdução e o método da máxima parcimônia." In ed. Sergio Russo Matioli. Editora Holos, p. 97-107.
- Moyle, Graeme J et al. 2005. "Epidemiology and predictive factors for chemokine receptor use in HIV-1 infection." *Journal of Infectious Disease* 191(6): 866-872.
- Muciaccia, Barbara et al. 2005. "Beta-chemokine receptors 5 and 3 are expressed on the head region of human spermatozoon." *The FASEB journal official publication of the Federation of American Societies for Experimental Biology* 19(14): 2048-2050.

- Muesing, M A et al. 1985. "Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus." *Nature* 313(6002): 450-458.
- Nadai, Yuka et al. 2009. "HIV-1 epidemic in the Caribbean is dominated by subtype B." *PloS one* 4(3): e4814.
- Nau, Jean-Yves. 2009. "A new human immunodeficiency virus derived from gorillas." *Revue medicale suisse* 5(215): 1741.
- Nelson, M et al. 2003. *An open-label study of tenofovir in HIV-1 and Hepatitis B virus co-infected individuals.*
- Nájera, Rafael et al. 2002. "Genetic recombination and its role in the development of the HIV-1 pandemic." *AIDS* 16 Suppl 4: S3-16.
- Oliveira, T de, Deenan Pillay, and Robert J Gifford. 2010. "The HIV-1 subtype C epidemic in South America is linked to the United Kingdom." *PloS one* 5(2): e9311.
- Pape, Jean William et al. 2008. "The epidemiology of AIDS in Haiti refutes the claims of Gilbert et al." *Proceedings of the National Academy of Sciences of the United States of America* 105(10): E13.
- Paraskevis, Dimitrios et al. 2007. "Increasing Prevalence of HIV-1 Subtype A in Greece : Estimating Epidemic History and Origin." *The Journal of Infectious Diseases* 196: 1167-1176.
- Paraskevis, Dimitrios, O Pybus, and G Magiorkinis, et al. 2009. "Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach." *Retrovirology* 6: 49.
- Parkin, N T, M Chamorro, and HE Varmus. 1992. "Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo." *Journal of Virology* 66(8): 5147-5151.
- Peeters, M, and P M Sharp. 2000. "Genetic diversity of HIV-1: the moving target." *AIDS* 14 Suppl 3: S129-S140.
- Penny, D et al. 1994. "The role of models in reconstructing evolutionary trees." In Clarendon Press, p. 211-230.
- Perelson, a S et al. 1996. "HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time." *Science* 271(5255): 1582-6.
- Perrin, Luc, Laurent Kaiser, and Sabine Yerly. 2003. "Travel and the spread of HIV-1 genetic variants." *The Lancet infectious diseases* 3(1): 22-7.
- Pilcher, Christopher D, Joseph K Wong, and Satish K Pillai. 2008. "Inferring HIV Transmission Dynamics from Phylogenetic Sequence Relationships." *PLoS Medicine* 5(3): 3-5.
- Pinto, Edelenyi, and Claudio José. 2006. "A diversidade do HIV-1 : uma ferramenta para o estudo da pandemia HIV-1 diversity : a tool for studying the pandemic." *Review Literature And Arts Of The Americas* 22(3): 473-484.
- Pitchenik AE, et al. (1983) Opportunistic infections and Kaposi's sarcoma among Haitians: Evidence of a new acquired immunodeficiency state. *Ann Intern Med* 98:277-284.
- Plantier, JC et al. 2009 "A new human immunodeficiency virus derived from gorillas" *Nature Medicine* 15(8):871-872.

- Plenum, B et al. 1999. "Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*." *Nature* 397:436-441.
- Pornillos, Owen, Jennifer E Garrus, and Wesley I Sundquist. 2002. "Mechanisms of enveloped RNA virus budding." *Trends in cell biology* 12(12): 569-79.
- Posada, David, K Crandall, and D Hillis. 2002. "Phylogenetics of HIV." *Computational and Evolutionary Analysis of HIV Molecular Sequences*: 121–160.
- Potts, K E et al. 1993. "Genetic heterogeneity of the V3 region of the HIV-1 envelope glycoprotein in Brazil. Brazilian Collaborative AIDS Research Group." *AIDS* 7(9): 1191-1197.
- Poulsen, A G et al. 2000. "Risk factors for HIV-2 seropositivity among older people in Guinea-Bissau. A search for the early history of HIV-2 infection." *Scandinavian Journal of Infectious Diseases* 32(2): 169-175.
- Preston, B D, B J Poiesz, and L A Loeb. 1988. "Fidelity of HIV-1 reverse transcriptase." *Science* 242(4882): 1168-1171.
- Quinones-Mateu, M E, and Eric J Arts. 1999. "Recombination in HIV-1: update and implications." *Aids Reviews* 1(2): 89-100.
- Rambaut, Andrew et al. 2004. "The causes and consequences of HIV evolution." *Nature reviews. Genetics* 5(1): 52-61.
- Ramos, A et al. 1999. "Dual and Recombinant Infections : An Integral Part of the HIV-1 Epidemic in Brazil." *Emerging Infectious Diseases* 5(1): 65-74.
- Reeves, Jacqueline D et al. 1999. "Primary human immunodeficiency virus type 2 (HIV-2) isolates infect CD4-negative cells via CCR5 and CXCR4: comparison with HIV-1 and simian immunodeficiency virus and relevance to cell tropism in vivo." *Journal of Virology* 73(9): 7795-7804.
- Renjifo, Boris et al. 2004. "Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D". *AIDS* 18(12):1629-36.
- Robbins, Kenneth E et al. 2003a. "U . S . Human Immunodeficiency Virus Type 1 Epidemic : Date of Origin , Population History , and Characterization of Early Strains." *Journal of Virology* 77(11): 6359-6366.
- Robinson, Harriet L. 2002. "New hope for an AIDS vaccine." *Nature reviews. Immunology* 2(4): 239-50.
- Rodriquez, L et al. 1985. "Antibodies to HTLV-III/LAV among aboriginal Amazonian Indians in Venezuela." *Lance* 2(8464): 1098-1100.
- Ruben, S et al. 1989. "Structural and functional characterization of human immunodeficiency virus tat protein." *Journal of Virology* 107(3): 486-492.
- Russell, KL et al. 2000. "Emerging genetic diversity of HIV-1 in South America." *AIDS* 14(12): 1785-91.
- Salemi, Marco et al. 2008. "High-resolution molecular epidemiology and evolutionary history of HIV-1 subtypes in Albania." *PloS one* 3(1): e1390.
- Santiago, ML et al. 2005. "Simian Immunodeficiency Virus Infection in Free-Ranging Sooty Mangabeys (*Cercocebus atys atys*) from the Tai-Forest , Cote d'Ivoire : Implications for the Origin of Epidemic Human Immunodeficiency Virus Type 2." *Journal of Virology* 79(19): 12515-12527.

- Santoro-lobes, Guilherme et al. 2000. "HIV Disease Progression and V3 Serotypes in Brazil : Is B Different from B-Br ?" *AIDS Research and Human Retroviruses* 16(10): 953-958.
- Schim Van Der Loeff, M F, and P Aaby. 1999. "Towards a better understanding of the epidemiology of HIV-2." *AIDS* 13 Suppl A: S69-S84.
- Schwartz, S a, and M P Nair. 1999. "Current concepts in human immunodeficiency virus infection and AIDS." *Clinical and diagnostic laboratory immunology* 6(3): 295-305.
- Sharp, P M et al. 2000. "Origins and evolution of AIDS viruses: estimating the time-scale." *Biochemical Society transactions* 28(2): 275-82.
- Sharp, P M et al. 2001. "The origins of acquired immune deficiency syndrome viruses: where and when?" *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 356(1410): 867-76.
- Sharp, Paul M, Beatrice H Hahn, and Phil Trans R Soc B. 2010. "The evolution of HIV-1 and the origin of AIDS The evolution of HIV-1 and the origin of AIDS." *Philosophical transactions of the Royal Society of London. Series B Biological Sciences* 20:2487-2494.
- Sodora, Donald L, and Guido Silvestri. 2008b. "Immune activation and AIDS pathogenesis." *AIDS* 22(4): 439-46.
- Soriano, V et al. 1996. "Epidemiology of HIV-2 infection in Spain." *European Journal of Clinical Microbiology and Infectious Diseases* 15(5): 383-388.
- Southgate, C D, and MR Green. 1991. "The HIV-1 Tat protein activates transcription from an upstream DNA-binding site: implications for Tat function." *Genes & Development* 5(12B): 2496-2507.
- Spira, S et al. 2003. "Impact of clade diversity on HIV-1 virulence , antiretroviral drug sensitivity and drug resistance." *Journal of Virology*: 229-240.
- Tavaré, S. 1986. "Some probabilistic and statistical problems in the analysis of DNA sequences" ed. R M Miura. *Lectures on Mathematics in the Life Sciences* 17(17): 57-86.
- Taylor, Barbara S, and Scott M Hammer. 2008. "The challenge of HIV-1 subtype diversity." *The New England journal of medicine* 359(18): 1965-6.
- Tee, Kok Keng et al. 2009. "Estimating the date of origin of an HIV-1 circulating recombinant form." *Virology* 387(1): 229-34.
- Thomson, M M, L Perez-Alvarez, and R Najera. 2002. "Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy." *Lancet Infectious Disease* 2(1473-3099)
- Thomson, Michael M et al. 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." *The Journal of General Virology* 83:107-119.
- Thorne, Claire et al. 2010. "Central Asia: hotspot in the worldwide HIV epidemic." *The Lancet Infectious Diseases* 10(7): 479-488.
- Trivedi, Bijal. 2010. "The primate connection." *Nature* 466(7304): S5.
- UNAIDS Report on the global AIDS Epidemic 2010. Joint United Nations Programme on HIV/AIDS. Disponível em http://www.unaids.org/globalreport/Global_report.htm.

- Vallari, Ana et al. 2010. "Confirmation of Putative HIV-1 Group P in Cameroon." *Journal of Virology* 85(3): 1403-7.
- Van Heuverswyn, Fran et al. 2006. "Human immunodeficiency viruses: SIV infection in wild gorillas." *Nature* 444(7116): 164.
- Vidal, N et al. 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(22): 10498-507.
- Vodicka, Marie A et al. 1998. "HIV-1 Vpr interacts with the nuclear transport pathway to promote macrophage infection." *Genes & Development* 12(2): 175-185.
- Watts, Joseph M et al. 2009. "Architecture and secondary structure of an entire HIV-1 RNA genome." *Nature* 460(6):711-719.
- Walker, B D, and B T Korber. 2001. "Immune control of HIV: the obstacles of HLA and viral diversity." *Nature Immunology* 2(6): 473-475.
- Warrilow, David, Gilda Tachedjian, and David Harrich. 2009. "Maturation of the HIV reverse transcription complex: putting the jigsaw together." *Reviews in Medical Virology* 19(6): 324-337.
- Wertheim, Joel O, and Michael Worobey. 2009a. "Dating the Age of the SIV Lineages That Gave Rise to." *PLoS Computational Biology* 5(5).
- Wiegand, Heather L et al. 2004. "A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins." *The European Molecular Biology Organization Journal* 23(12): 2451-2458.
- Wolfe, Nathan D et al. 2004. "Naturally acquired simian retrovirus infections in central African hunters." *Lancet* 363(9413): 932-937.
- Worobey, Michael et al. 2008. "Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960." *Nature* 455(7213): 661-664.
- Worobey, Michael et al. 2010. "Island biogeography reveals the deep history of SIV." *Science* 329(5998): 1487.
- Yamaguchi, J, SG Devare, and C A Brennan. 2000. "Identification of a new HIV-2 subtype based on phylogenetic analysis of full-length genomic sequence." *AIDS Research and Human Retroviruses* 16(9): 925-930.
- Yang, Z. 1993. "Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites." *Molecular Biology and Evolution* 10(6): 1396-1401.
- Zack, J A et al. 1990. "HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure." *Cell* 61(2): 213-222.