MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
INSTITUTO OSWALDO CRUZ

Doutorado em Biologia Parasitária

AVALIAÇÃO DA PREVALÊNCIA DE MUTAÇÕES TRANSMITIDAS DE RESISTÊNCIA AOS ANTIRRETROVIRAIS E DA HISTÓRIA EVOLUTIVA DE CLADOS RAROS DO HIV-1 NO RIO DE JANEIRO EM UMA POPULAÇÃO DE GESTANTES ANTES DO INÍCIO DA TERAPIA PARA PREVENÇÃO DA TRANSMISSÃO VERTICAL

EDSON OLIVEIRA DELATORRE

Rio de Janeiro
Outubro de 2015

Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Parasitária

EDSON OLIVEIRA DELATORRE

AVALIAÇÃO DA PREVALÊNCIA DE MUTAÇÕES TRANSMITIDAS DE RESISTÊNCIA AOS ANTIRRETROVIRAIS E DA HISTÓRIA EVOLUTIVA DE CLADOS RAROS DO HIV-1 NO RIO DE JANEIRO EM UMA POPULAÇÃO DE GESTANTES ANTES DO INÍCIO DA TERAPIA PARA PREVENÇÃO DA TRANSMISSÃO VERTICAL

Tese apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Doutor em Biologia Parasitária.

Orientador: Prof. Dra. Mariza Gonçalves Morgado

RIO DE JANEIRO
Outubro de 2015

Ficha catalográfica elaborada pela
Biblioteca de Ciências Biomédicas/ ICICT / FIOCRUZ - RJ

D341 Delatorre, Edson Oliveira

Avaliação da prevalência de mutações transmitidas de resistência aos antirretrovirais e da história evolutiva de clados raros do HIV-1 no Rio de Janeiro em uma população de gestantes antes do início da terapia para prevenção da transmissão vertical / Edson Oliveira Delatorre. – Rio de Janeiro, 2015.
xiv, 223 f. : il. ; 30 cm.

Tese (Doutorado) – Instituto Oswaldo Cruz, Pós-Graduação em Biologia Parasitária, 2015.
Bibliografia: f. 139-154

1. HIV. 2. Gestantes. 3. Diversidade viral. 4. Resistência transmitida. 5. Filogenia. 6. Filogeografia. I. Título.

CDD 616.9792

Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Parasitária

AUTOR: EDSON OLIVEIRA DELATORRE

AVALIAÇÃO DA PREVALÊNCIA DE MUTAÇÕES TRANSMITIDAS DE RESISTÊNCIA AOS ANTIRRETROVIRAIS E DA HISTÓRIA EVOLUTIVA DE CLADOS RAROS DO HIV-1 NO RIO DE JANEIRO EM UMA POPULAÇÃO DE GESTANTES ANTES DO INÍCIO DA TERAPIA PARA PREVENÇÃO DA TRANSMISSÃO VERTICAL

ORIENTADOR: **Prof. Dra. Mariza Gonçalves Morgado**

Aprovada em: 29/10/2015

EXAMINADORES:

Prof. Dr. Thiago Moreno Lopes e Souza - Presidente (Instituto Oswaldo Cruz)
Prof. Dr. Marcelo Alves Soares (Universidade Federal do Rio de Janeiro/Instituto Nacional de Câncer)
Prof. Dra. Caroline Pereira Bittencourt Passaes (Institut Pasteur)
Prof. Dr. Walter de Araujo Eyer Silva (Universidade Federal do Estado do Rio de Janeiro)
Prof. Dr. Eduardo de Mello Volotão (Instituto Oswaldo Cruz)

Rio de Janeiro, 29 de outubro de 2015

Ata da defesa de tese de doutorado em Biologia Parasitária de **Edson Oliveira Delatorre**, sob orientação da Drª. Mariza Gonçalves Morgado. Ao vigésimo nono dia do mês de outubro de dois mil e quinze, realizou-se às treze horas, Auditório da Vice Direção de Ensino do INI, o exame da tese de doutorado intitulada: **"Avaliação da prevalência de mutações transmitidas de resistência aos antirretrovirais e da história evolutiva de clados raros do HIV-1 no Rio de Janeiro em uma população de gestantes antes do início da terapia para prevenção da transmissão vertical"** no programa de Pós-graduação em Biologia Parasitária do Instituto Oswaldo Cruz, como parte dos requisitos para obtenção do título de Doutor em Ciências - área de concentração: Genética e Bioquímica, na linha de pesquisa: Variabilidade Genética de Parasita, Vetores e Hospedeiros.. A banca examinadora foi constituída pelos Professores: Dr. Thiago Moreno Lopes e Souza - IOC/FIOCRUZ (Presidente), Dr. Marcelo Alves Soares– INCA/RJ, Drª. Caroline Pereira Bittencourt Passaes - Institut Pasteur/França e como suplentes: Dr. Walter de Araujo Eyer Silva - UNIRIO/RJ e Dr. Eduardo de Mello Volotão - IOC/FIOCRUZ. Após arguir o candidato e considerando que o mesmo demonstrou capacidade no trato do tema escolhido e sistematização da apresentação dos dados, a banca examinadora pronunciou-se pela APROVAÇÃO da defesa da tese de doutorado. De acordo com o regulamento do Curso de Pós-Graduação em Biologia Parasitária do Instituto Oswaldo Cruz, a outorga do título de Doutor em Ciências está condicionada à emissão de documento comprobatório de conclusão do curso. Uma vez encerrado o exame, o Coordenador do Programa, Dr. Rafael Maciel de Freitas, assinou a presente ata tomando ciência da decisão dos membros da banca examinadora. Rio de Janeiro, 29 de outubro de 2015.

Dr. Thiago Moreno Lopes e Souza (Presidente da Banca):

Dr. Marcelo Alves Soares(Membro da Banca):

Drª. Caroline Pereira Bittencourt Passaes (Membro da Banca):

Dr. Rafael Maciel de Freitas (Coordenador do Programa):

Dedico este trabalho a todos os indivíduos infectados pelo HIV. Que os resultados de nossos esforços gerem benefícios para suas lutas.

AGRADECIMENTOS

À Dra. Mariza Gonçalves Morgado, por ter aceitado me orientar durante o doutoramento. Foram anos muito importantes para minha formação intelectual e profissional.

Ao Dr. Gonzalo Bello, por toda sua ajuda durante meu doutorado, pelas valiosas discussões técnico-científicas e ensinamentos. Exemplo de capacidade, competência e profissionalismo.

Às Dras. Monick Guimarães e Caroline Passaes, fundamentais para meu ingresso e encaminhamento no laboratório, me apresentando todo este mundo novo envolvendo a pesquisa sobre o HIV/aids. Vocês foram fundamentais para meu progresso neste mundo acadêmico e por isso, serei eternamente grato.

À toda equipe do Laboratório de AIDS e Imunologia Molecular do Instituo Oswaldo Cruz (IOC), pelo apoio e pelas discussões científicas e sobre os experimentos da minha tese, fundamentais para minha formação acadêmica.

Agradeço também à toda equipe do Departamento de Doenças Sexualmente Transmissíveis do Hospital Geral de Nova Iguaçu (HGNI), sem os quais o recrutamento de pacientes não seria possível.

Agradeço à Plataforma de Sequenciamento de DNA do IOC (DNA-RPT01A, PDTIS/FIOCRUZ), vinculada ao Laboratório de Genômica Funcional e Bioinformática do IOC, sua coordenadora Aline S. Moreira e toda sua equipe sempre prestativa no auxílio para obtenção das sequências nucleotídicas.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES e Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ pelas bolsas de doutorado concedidas.

A Deus, meus pais e familiares por todo o apoio e força para enfrentar os desafios.

À minha esposa, amiga e companheira Suwellen, por toda motivação e incentivo para a conclusão desta tese.

Gostaria de agradecer especialmente aos voluntários que contribuíram com o indispensável, seu material biológico, sem o qual esta tese não seria possível.

“Em algum lugar, algo incrível
espera para ser descoberto.”
Carl Sagan

Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

AVALIAÇÃO DA PREVALÊNCIA DE MUTAÇÕES TRANSMITIDAS DE RESISTÊNCIA AOS ANTIRRETROVIRAIS E DA HISTÓRIA EVOLUTIVA DE CLADOS RAROS DO HIV-1 NO RIO DE JANEIRO EM UMA POPULAÇÃO DE GESTANTES ANTES DO INÍCIO DA TERAPIA PARA PREVENÇÃO DA TRANSMISSÃO VERTICAL

RESUMO

TESE DE DOUTORADO EM BIOLOGIA PARASITÁRIA

Edson Oliveira Delatorre

A epidemia global do HIV está evoluindo em direção à um aumento nas taxas de mutações de resistência e diversidade molecular. Estudos prévios conduzidos na região metropolitana do estado do Rio de Janeiro (RJ) encontraram prevalências moderadas de mutações transmitidas de resistência às drogas antirretrovirais (MTRD) do HIV-1 e uma alta diversidade molecular. Neste trabalho, avaliamos a prevalência das MTRD e a história evolutiva de clados raros do HIV-1 no RJ em uma população de gestantes virgens de terapia. Foram observadas mudanças na tendência das MTRD com as maiores taxas (de inibidores nucleosídicos da transcriptase reversa para inibidores da protease) e um aumento na prevalência de subtipos não-B quando comparados a um estudo anterior conduzido no mesmo local cinco anos antes. As MTRD encontradas podem afetar o desfecho virológico dos regimes antirretrovirais padrão adotados para prevenção da transmissão vertical do HIV-1. Além disso, o nível de MTRD encontrado nesta população de gestantes é considerado alto e justifica a importância da realização da genotipagem pré-tratamento do HIV para determinação de resistência. Através das análises evolutivas dos clados do HIV-1 raramente encontrados no RJ, verificou-se a circulação de linhagens do subtipo C, CRF02_AG e CRF45_cpx provavelmente importadas do continente africano. A maioria das sequências brasileiras do subtipo C formaram um clado monofilético, que provavelmente se originou no Burundi e foi introduzido no estado do Paraná por volta de 1975, disseminando-se rapidamente para outras regiões brasileiras, incluindo o RJ. Cinco introduções adicionais de cepas do subtipo C provavelmente originadas em países da África Oriental, Meridional e Central foram detectadas no estado do RJ. As sequências brasileiras do CRF02_AG distribuíram-se em cinco linhagens distintas, provavelmente originadas na África Ocidental, provavelmente Gana, Senegal e Nigéria. Dois clados monofiléticos compostos somente por sequências fluminenses foram identificados, com datas de origem estimadas por volta de 1985. Dentre as 10 amostras recombinantes com fragmentos do subtipo K do HIV-1 isoladas no RJ e que tiveram seus genomas parciais ou quase totais sequenciados, seis mostraram um padrão similar ao CRF45_cpx ao longo de todo o genoma. As restantes foram classificadas como recombinantes de segunda-geração entre o CRF45_cpx e os subtipos B e F1, prevalentes na epidemia brasileira. Todas as sequências CRF45_cpx brasileiras, exceto uma, formaram um clado monofilético, que parece ser o resultado de um único evento de introdução ocorrido por volta de 1984. Esta linhagem se disseminou pelos estados do RJ, São Paulo e Minas Gerais e é relacionada a sequências da Argentina, Itália e Bélgica. Estas análises filogenéticas indicam um influxo contínuo de linhagens do HIV-1 de origem africana no Brasil, que foram introduzidas e/ou tem se disseminado no RJ nos últimos 30-40 anos, demonstrando a importância deste estado para a introdução de novas variantes do HIV na epidemia brasileira.

Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

EVALUATION OF THE PREVALENCE OF TRANSMITTED DRUG RESISTANCE MUTATIONS AND THE EVOLUTIONARY HISTORY OF RARE HIV-1 CLADES IN RIO DE JANEIRO IN A POPULATION OF PREGNANT WOMEN BEFORE THE THERAPY FOR VERTICAL TRANSMISSION PREVENTION

ABSTRACT

PHD THESIS IN PARASITE BIOLOGY

Edson Oliveira Delatorre

The global HIV epidemic is evolving towards an increase in the rates of resistance mutations and molecular diversity. Previous studies conducted in the metropolitan region of Rio de Janeiro (RJ) state have found moderate prevalences of transmitted drug resistance mutations (TDRM) and high molecular diversity of HIV-1. In this work, we evaluated the prevalence of TDRM and the evolutionary history of rare HIV-1 clades in RJ found in an antiretroviral-naive pregnant women population. A change in the trend of the TDRM with higher rates (from nucleoside reverse transcriptase inhibitors to protease inhibitors) and an increase in the prevalence of non-B subtypes were observed when compared to a previous study conducted at the same location five years earlier. The TDRM level found in this population can affect the virological outcome of standard antiretroviral regimens to prevent HIV-1 vertical transmission. The TDRM level found in this pregnant women population is considered high and reinforce the relevance of the pretreatment genotyping of HIV for resistance determination. Through evolutionary analysis of HIV-1 clades rarely found in RJ, the circulation of HIV lineages of subtype C, CRF02_AG and CRF45_cpx probably imported from Africa were observed. Most Brazilian sequences of subtype C branched into a single monophyletic clade, which probably originated in Burundi and was introduced in Paraná state around 1975, spreading rapidly to other regions, including RJ. Five additional introductions of subtype C strains probably originated in Eastern, Central and Southern African countries were detected in the RJ. The Brazilian CRF02_AG sequences were distributed in five distinct lineages, which probably originated in West Africa, probably Ghana, Senegal and Nigeria. Two monophyletic clades formed only by sequences from RJ were identified, and their origin dates were estimated at around 1985. Among the 10 samples with HIV-1 subtype K recombinant fragments isolated in RJ who had their partial or near-full length genomes sequenced, six showed a CRF45_cpx-like pattern throughout the genome. The remaining ones were classified as second-generation recombinants between CRF45_cpx and subtypes B and F1, prevalent in the Brazilian epidemic. All Brazilian CRF45_cpx sequences, except one, formed a monophyletic clade that seems to be the result of a single introduction event occurred around 1984. This lineage has spread through the states of RJ, São Paulo and Minas Gerais and is related to sequences from Argentina, Italy and Belgium. These phylogenetic analyzes indicate a continuous influx of HIV-1 African strains in Brazil, which were introduced and/or has spread in Rio de Janeiro in the last 30-40 years, demonstrating the importance of this state to the introduction of new HIV variants in the Brazilian epidemic.

ÍNDICE

ÍNDICE DE FIGURAS

LISTA DE SIGLAS E ABREVIATURAS

APOBEC3	do inglês, *apolipoprotein mRNA editing enzyme, catalytic polypeptide-like*
CDC	Centro de Controle e Prevenção de Doenças dos Estados Unidos da América, do inglês *Centers for Disease Control and Prevention*
cDNA	DNA complementar
CI	Costa do Marfim
CPI	complexo pré-integração
CRFs	formas recombinnates circulantes, do inglês *circulating recombinant forms*
DNA	ácido desoxirribonucléico
env	gene do envelope, do inglês *envelope*
gag	gene dos antígenos grupo específicos, do inglês g*roup-specific antigens*
GALT	tecidos linfóides associados ao intestino, do inglês *gut-associated lymphoid tissues*
GH	Gana
GM	Gâmbia
GW	Guiné Bissau
HIV	vírus da imunodeficiência humana
HSH	homens que fazem sexo com homens
HTLV	vírus T-linfotrópico humano
ICTV	Comitê Internacional de Taxonomia de Vírus, do inglês *International Committee on Taxonomy of Viruses*
INI	inibidor da integrase
INNTR	inibidor não-nucleosídeo da transcriptase reversa
INTR	inibidor nucleosídeo da transcriptase reversa
IP	inibidor da protease
LAV	vírus associado à linfadenopatia
LR	Libéria
MS	Ministério da Saúde
OMS	Organização Mundial da Saúde
pol	gene da polimerase, do inglês *polymerase*
PR	protease

RNA	ácido ribonucléico
RT	transcriptase reversa
SAMHD1	do inglês, *SAM domain and HD domain-containing protein 1*
SIV	vírus da imunodeficiência símia
SIV*cpz*	vírus da imunodeficiência símia infectando chimpanzés
SIV*cpzPtt*	vírus da imunodeficiência símia infectando chimpanzés da subespécie *Pan troglodytes troglodytes*
SIV*gor*	vírus da imunodeficiência símia infectando gorilas
SIV*smm*	vírus da imunodeficiência símia infectando macacos mangabey
SL	Serra Leoa
SN	Senegal
TARV	terapia antirretroviral combinada de alta potência
MTRD	mutações transmitidas de resistência às drogas
TV	transmissão vertical do HIV
UNAIDS	do inglês, *Joint United Nations Programme on HIV/AIDS*
URFs	formas recombinantes únicas, do inglês *unique recombinant forms*

1 INTRODUÇÃO

1.1 A descoberta da Síndrome da Imunodeficiência Adquirida - aids

Em 1981, os Centros de Controle e Prevenção de Doenças dos Estados Unidos da América (CDC, do inglês *Centers for Disease Control and Prevention*) foram alertados sobre um grupo de homossexuais masculinos que apresentavam sintomas de pneumonia causada por *Pneumocystis jirovecii* (anteriormente denominado *Pneumocystis carinii*), um fungo oportunista raro conhecido por infectar pacientes imunocomprometidos (CDC 1981; Gottlieb et al. 1981), e um câncer raro de pele conhecido como sarcoma de Kaposi (Hymes et al. 1981). Estes relatos foram associados a uma provável síndrome que causava imunossupressão e linfadenopatia generalizada, e que começou a ser encontrada também em outros grupos, como usuários de drogas intravenosas (Masur 1982; Quagliarello 1982), hemofílicos (CDC 1982a), beneficiários de transfusão sanguínea (CDC 1982b), crianças (CDC 1982c), parceiras sexuais femininas de homens infectados (CDC 1983a; Masur 1982), prisioneiros (CDC 1983b), haitianos (CDC 1982d) e indivíduos da região central da África (Clumeck et al. 1983).

Considerados em conjunto, as características clínicas destes diferentes casos forneceram evidências claras de que esta síndrome recém descoberta seria causada por um agente infeccioso com transmissibilidade sanguínea, sexual e vertical. O CDC então montou uma força de trabalho para estudar a origem destes surtos, que ainda era desconhecida. Em 1982, a doença foi denominada de Síndrome da Imunodeficiência Adquirida (aids) (CDC 1982e).

Dois anos após os primeiros relatos do aparecimento da aids, dois grupos de pesquisa, liderados por Robert Gallo e Luc Montagnier declararam independentemente que um novo retrovírus poderia ser o agente etiológico da doença (Barre-Sinoussi et al. 1983; Gallo et al. 1984; Popovic et al. 1984). Gallo defendia que este vírus se tratava de um novo isolado do vírus T-linfotrópico humano (HTLV), previamente descrito por seu grupo, nomeando-o como HTLV-III (Popovic et al. 1984). O vírus isolado pelo grupo de Montagnier foi denominado como vírus associado à linfadenopatia (LAV) (Barre-Sinoussi et al. 1983). Posteriormente, verificou-se que ambos eram isolados do mesmo vírus (Ratner et al. 1985), e em 1986 o Comitê Internacional de Taxonomia de Vírus (ICTV, do inglês

International Committee on Taxonomy of Viruses) renomeou o vírus causador da aids como Vírus da Imunodeficiência Humana (HIV) (Coffin et al. 1986).

Em 1985, dois anos após o isolamento do HIV, outro vírus que causava os mesmos sintomas da aids foi identificado em trabalhadoras do sexo de Senegal (Barin et al. 1985). Posteriormente este novo vírus também foi encontrado em dois indivíduos de Guiné-Bissau e Cabo Verde (Clavel et al. 1986). Estes novos casos correspondiam a um novo tipo de HIV, o que levou a uma modificação da nomenclatura para que a diferenciação entre eles pudesse ser feita. Os primeiros isolados virais obtidos pelos grupos de Gallo e Montaigner passaram a ser classificados como HIV do tipo 1 (HIV-1), enquanto os novos isolados provenientes da região Ocidental da África como HIV do tipo 2 (HIV-2) (Barin et al. 1985; Clavel et al. 1986).

O HIV-1 circula mundialmente e é o responsável pela pandemia de aids, enquanto o HIV-2 está restrito à África Ocidental. Entretanto, casos esporádicos do HIV-2 já foram descritos em diversos países (Campbell-Yesufu e Gandhi 2011), inclusive no Brasil (Pieniazek et al. 1991). Além das diferenças na distribuição, também existem evidências de que a transmissão do HIV-2 é menos eficiente quando comparada ao HIV-1 e possuindo um período entre a infecção inicial e o aparecimento da doença mais longo (Campbell-Yesufu e Gandhi 2011; Marlink et al. 1994).

O primeiro caso de aids no Brasil foi registrado em 1980 na cidade de São Paulo, sendo classificado corretamente somente dois anos depois (Parker et al. 1994). Em 1987, pesquisadores do Instituto Oswaldo Cruz isolaram o HIV-1 pela primeira vez na América Latina, a partir de amostra biológica de um paciente do Rio de Janeiro infectado por transfusão sanguínea (Galvão-Castro et al. 1987).

Durante a década de 1980, já estava evidente de que o HIV-1 já tinha se disseminado por diversos países de diferentes continentes, o que indicava que a transmissão do vírus estava ocorrendo sem ser notada pelos sistemas de vigilância em saúde. Existem diversos fatores que podem ter contribuído para a disseminação inicial do HIV, a maioria dos quais se intensificou a partir da segunda metade do século XX.

Viagens nacionais e internacionais exerceram um papel primordial para a disseminação inicial do HIV. No continente Africano, que atualmente é aceito como o local de origem do HIV (Gao et al. 1999; Lemey et al. 2003), a disseminação do vírus provavelmente seguiu por rotas de caminhoneiros, espalhando-se entre cidades do

interior do continente (Hudson 1996). As fronteiras relativamente abertas do continente em conjunção com corredores de mobilidade com alta acessibilidade [definida como a facilidade com a qual um grupo de pessoas se move entre diferentes regiões (Yoshida e Deichmann 2009)] permitiram que em poucos anos o HIV estivesse presente em praticamente toda a África Subsaariana (Clumeck et al. 1983; Van de Perre et al. 1984; Piot et al. 1984) e por conta dos movimentos migratórios internacionais, que muitas vezes seguem laços coloniais (Faria et al. 2012), fizeram com que o HIV chegasse à Europa (Bryceson et al. 1988; Frøland et al. 1988) e Américas (CDC 1982d; Galvão-Castro et al. 1987; Gottlieb et al. 1981).

Outro fator que também poderia ter contribuído para a disseminação inicial do HIV-1 no continente africano seriam os conflitos armados envolvendo os movimentos de libertação, impulsionados após a Segunda Guerra Mundial. Além de contribuir para a disseminação do vírus nos países em conflito, os soldados também ajudaram na exportação do HIV para outros países no momento de retorno para os seus países de origem, como foi o caso dos soldados cubanos que serviram em Angola durante as décadas de 1970-1980 (Torres-Anjel 1992a).

Diversos fatores contribuíram para a aceleração da dispersão do HIV-1 durante as décadas de 1970-1980. Além dos movimentos migratórios, sejam eles por motivos militares, turísticos ou de trabalho, a transmissão do HIV para certos grupos mais vulneráveis, como homens que fazem sexo com homens (HSH), beneficiários de transfusão de sangue e usuários drogas injetáveis também tiveram um grande impacto na disseminação da epidemia (CDC 2001; Torres-Anjel 1992b).

Os primeiros relatos da aids foram definidos como uma “doença adquirida através de contatos sexuais” (CDC 2001). Dentre os diversos casos de transmissão sexual relatados durante o início da década de 1980, a transmissão entre HSH correspondia a uma grande parcela, embora transmissões heterossexuais também tenham sido encontradas. Desde então, os HSH foram definidos como um grupo vulnerável à infecção pelo HIV, constituindo até hoje um grupo predominantemente mais afetado, principalmente em áreas que possuem uma epidemia do HIV concentrada, como no caso de alguns países da Europa, Estados Unidos da América e Brasil (UNAIDS 2014).

Nesta mesma época, a transfusão de sangue se tornou rotina da prática médica, e na ausência de testes para detecção do HIV, a utilização de sangue contaminado pelo vírus acabou ocorrendo em diversos locais do mundo, aumentando as chances da infecção pelo HIV naqueles indivíduos que entraram em

contato com o mesmo e/ou seus derivados (CDC 1982a; CDC 1982b). Isto foi o que ocorreu com muitos hemofílicos ao se beneficiarem do fator de coagulação VIII produzido a partir do sangue combinado de centenas de doadores, no qual, a presença de somente um doador portador do HIV poderia contaminar todo o lote.

O aumento da disponibilidade de drogas injetáveis como a heroína acabou gerando um comportamento de risco para a transmissão do HIV: o compartilhamento de seringas não esterilizadas, que potencializou a transferência do vírus entre os usuários de drogas injetáveis. Estes constituem até hoje um dos grupos mais vulneráveis à transmissão do HIV (Harm Reduction International 2014).

Todas as formas de disseminação que contribuíram para a dispersão inicial do HIV entre os diferentes países ainda estão atuando até os dias de hoje, modificando sua prevalência nas diferentes populações (De Cock et al. 2012) e aumentando a diversidade genética viral local, graças à incorporação de novas linhagens do HIV, o que pode ter algum impacto na eficiência do tratamento ou para o desenvolvimento de vacinas (Aldrich e Hemelaar 2012).

1.2 A pandemia do HIV/aids

Atualmente a aids é definida como uma doença crônica e progressiva causada pelo HIV de tipos 1 e 2. A doença caracteriza-se por uma supressão profunda da imunidade mediada por células devido ao aumento gradual do nível de replicação viral, resultando numa depleção progressiva de células T $CD4^+$ que, em última análise, favorece o aparecimento de infecções oportunistas, neoplasias secundárias e doenças neurológicas (Douek 2003; Fauci et al. 1996).

A infecção pelo HIV causa um espectro de sintomas clínicos caracterizando estágios de progressão bem documentados, que se estendem desde o período de soroconversão, ou fase aguda, até a fase de aids. Entre estas duas fases existe um período de latência clínica (ou período assintomático), onde o HIV continua se replicando sem gerar sintomas (Douek 2003). A duração deste período assintomático, porém, é extremamente variável e distintos perfis de progressão podem ser observados entre os indivíduos infectados pelo HIV (Deeks e Walker 2007). Em indivíduos não tratados, estima-se que o tempo médio entre o contágio e o aparecimento da doença esteja em torno de dez anos (Giesecke et al. 1990; Poorolajal et al. 2015) (Figura 1). Nos últimos 20 anos, com a introdução da terapia antirretroviral combinada de alta potência (TARV) observou-se um impacto

expressivo na redução da mortalidade e morbidade em decorrência da aids. Este aspecto será apresentado mais em detalhes ao longo deste trabalho.

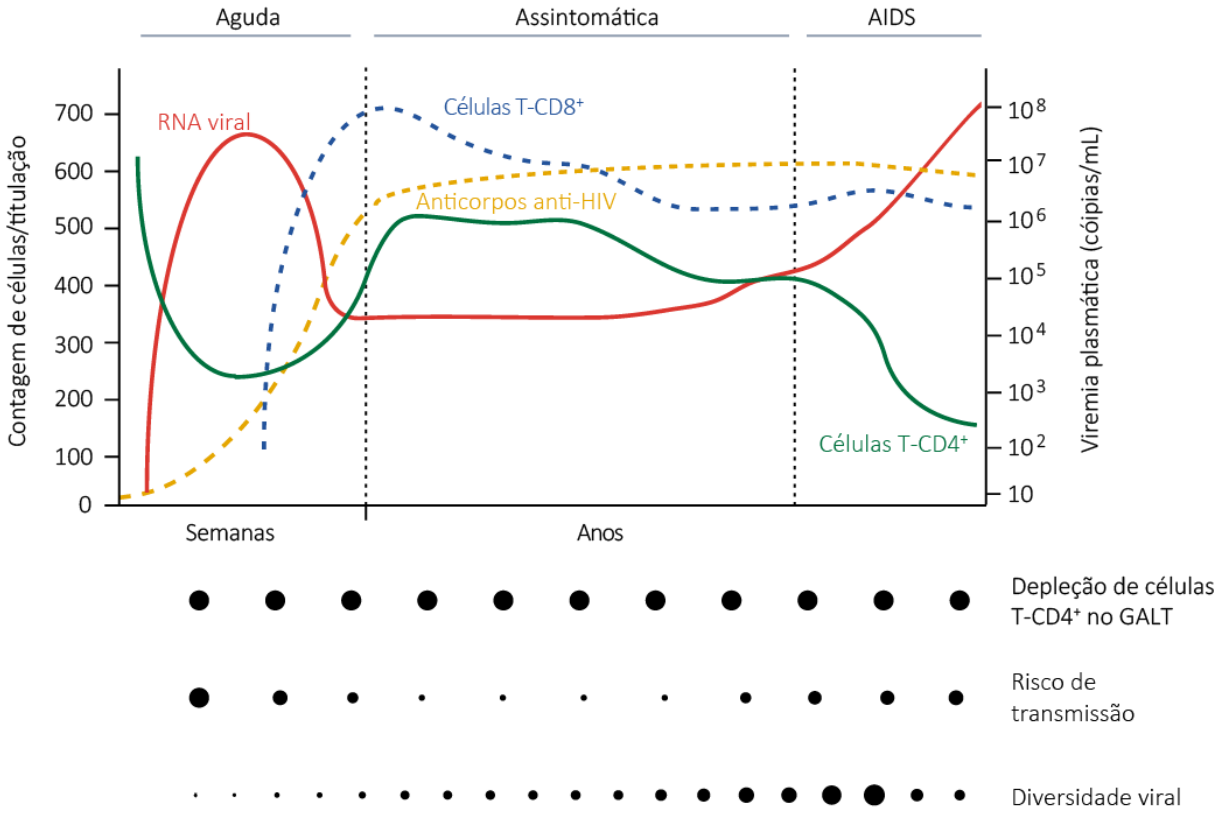


Figura 1 – História natural da infecção pelo HIV. Mudanças dinâmicas na viremia plasmática, os números de células T CD4⁺ e CD8⁺ e na titulação de anticorpos anti-HIV durante a progressão da doença estão indicados em função do tempo. Cada linha do gráfico representa uma das características indicadas de acordo com as cores. Diferenças nas taxas de depleção de células T-CD4⁺ no GALT, risco de transmissão e diversidade viral estão representados na forma de círculos pretos, cujo diâmetro é diretamente proporcional aos valores observados. GALT - tecidos linfóides associados ao intestino, do inglês *gut-associated lymphoid tissues*. Adaptado de Simon et al. (2006).

No final de 2014, foi estimado pelo Programa Conjunto das Nações Unidas sobre HIV/AIDS (UNAIDS, do inglês *Joint United Nations Programme on HIV/AIDS*,) e a Organização Mundial da Saúde (OMS) a existência de 36,9 milhões de pessoas infectadas pelo HIV mundialmente, sendo que 2,0 milhões de pessoas foram infectadas e 1,2 milhão morreram naquele ano (UNAIDS 2014).

A prevalência mundial da infecção pelo HIV-1 parece ter se estabilizado em 0,8% (0,7-0,9%) na última década (Figura 2). Entretanto, o total de pessoas infectadas pelo HIV vem aumentando, uma vez que ocorreu uma diminuição da taxa de mortalidade em decorrência do aumento do acesso à TARV. Em 2015, o número de pessoas recebendo TARV chegou a 15 milhões, correspondendo a mais de 40%

do total de pessoas infectadas, um passo importante para se reduzir a mortalidade associada à aids e as taxas de transmissão, metas estabelecidas pela UNAIDS para se obter o fim da epidemia de AIDS até 2030 (UNAIDS 2015a). Estas metas, conhecidas como “90-90-90”, consistem em ter 90% das pessoas com HIV diagnosticadas; deste grupo, 90% seguindo o tratamento; e, dentre as pessoas tratadas, 90% com carga viral indetectável até 2020. Quando esta meta tríplice for alcançada, pelo menos 73% de todas as pessoas vivendo com HIV no mundo terão supressão viral, o que através de modelos matemáticos, permitirá a extinção da epidemia de aids até 2030, gerando benefícios tanto para a saúde quanto para a economia (UNAIDS 2015b).

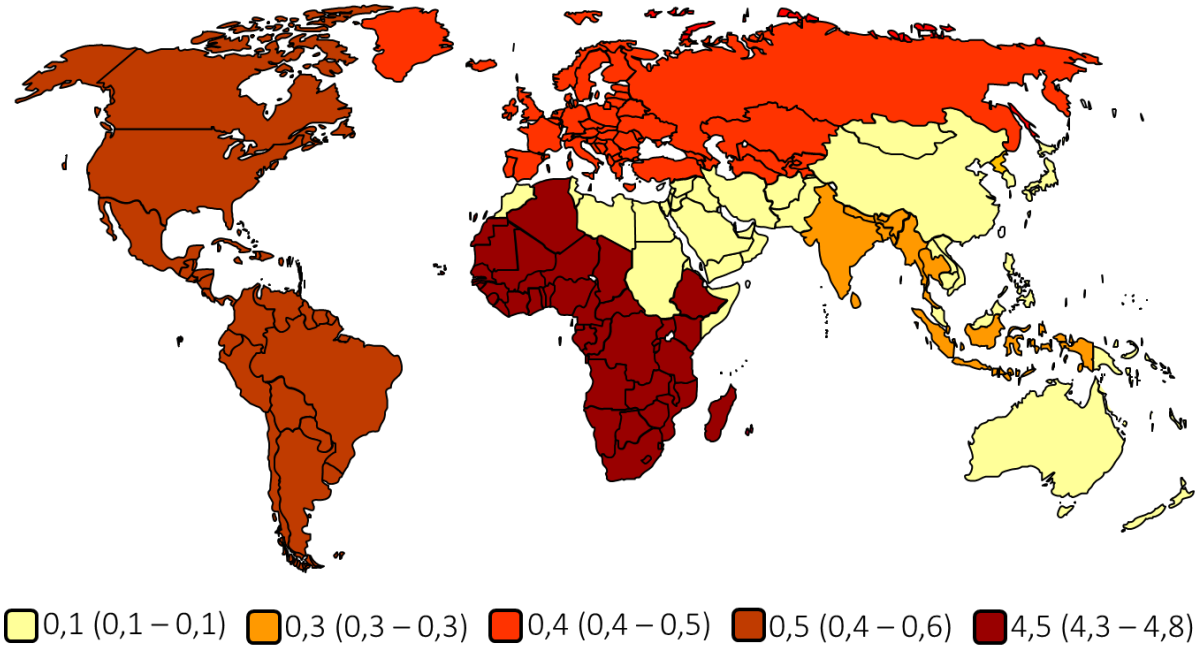


Figura 2 – Distribuição da prevalência da HIV em adultos (15 – 49 anos). Mapa representativo construído de acordo com as regiões analisadas pela OMS. A cor de cada região corresponde à prevalência (%) de acordo com a legenda abaixo da figura. A prevalência global do HIV no ano de 2014 foi de 0,8% (0,7 – 0,9). Adaptado a partir de dados da OMS (http://gamapserver.who.int/).

Globalmente, as mulheres correspondem a pouco mais de 47% do número de pessoas vivendo com HIV/aids, proporção que permanece estável nos últimos anos. Adolescentes e adultos jovens (entre 15-24 anos) correspondem a 34% das novas infecções pelo HIV em adultos mundialmente. Cerca de 220.000 crianças com idade abaixo de 15 anos se infectaram em 2014. Este número representa uma diminuição de 58% desde 2000, graças ao aumento do acesso às intervenções para prevenção da transmissão vertical do HIV. A disponibilidade de medicamentos antirretrovirais para mulheres grávidas com o objetivo de prevenir a transmissão vertical do HIV

aumentou para 73% em 2014. Entretanto a quantidade de crianças vivendo com HIV aumentou de 1,6 milhão em 2001 para 2,6 milhões em 2014, sendo a maioria das novas infecções em países subdesenvolvidos (UNAIDS 2015a; UNAIDS 2014).

Enquanto abriga somente cerca de 10% da população mundial, a África Subsaariana concentra cerca de 70% (25,8 milhões) das infecções pelo HIV em nível mundial e 66% (790 mil) das mortes relacionadas a aids em 2014. Mesmo com o avanço da terapia antirretroviral, a aids permanece como a principal causa de mortalidade na região.

Entre os países que compõem a África Subsaariana existe uma heterogeneidade na severidade da epidemia. A prevalência na maioria dos países da África Ocidental é menor que 2%, enquanto pode chegar a mais de 25% em países da África Meridional. A maior parte da transmissão do HIV na população geral da África Subsaariana ocorre através da via heterossexual, o que explica o grande número de homens e mulheres infectadas, e consequentemente crianças (UNAIDS 2015a; UNAIDS 2014). Entretanto, as transmissões envolvendo HSH são importantes, uma vez que uma proporção significativa destes mantêm relações sexuais paralelas tanto com homens quanto com mulheres. Por causa do estigma e da frequente criminalização deste grupo na África Subsaariana, baixos índices de prevenção de transmissão são alcançados, o que resulta em uma alta prevalência de indivíduos infectados pelo HIV, o que juntamente com o comportamento bissexual contribuem para os altos níveis de propagação do HIV para a população em geral (Beyrer et al. 2010).

A prevalência da infecção pelo HIV nos países que compõem a África Central e Ocidental é geralmente inferior à encontrada nos países da África Oriental e Meridional (UNAIDS 2014). A República Democrática do Congo é um caso interessante, pois mesmo tendo sido demonstrado que o HIV está presente neste país pelo menos desde 1959 (Zhu et al. 1998), a prevalência permaneceu estável em uma taxa relativamente baixa (1,0%). Em muitos países a taxa de incidência de infecções pelo HIV parece ter estabilizado, enquanto declina em outros. Pelo menos em parte, estas reduções regionais da incidência do HIV refletem uma tendência de redução de comportamentos de risco e dos esforços de prevenção da transmissão, além do aumento da disponibilidade da terapia antirretroviral, já que em 2014, 41% do total de pessoas vivendo com o HIV na região estavam recebendo terapia. Isto representa uma mudança notável desde 2000, quando a cobertura do tratamento na África Subsaariana era praticamente nula (UNAIDS 2015a).

No continente Americano, uma das consequências da disponibilidade da TARV foi o aumento do número de pessoas vivendo com HIV na América do Norte, mesmo que a incidência esteja estável em aproximadamente 50.000 novas infecções por ano desde a década de 1990. Estima-se que 1,2 milhão de pessoas estavam infectadas pelo HIV nos EUA em 2012. Os grupos mais afetados são os HSH e usuários de drogas. A região do Caribe possui a maior prevalência do HIV em adultos depois da África Subsaariana, de aproximadamente 1,1% (UNAIDS 2014). Nesta região predomina a transmissão heterossexual, com o grupo de HSH apresentando certa importância em alguns países (Beyrer et al. 2012).

A América Latina possui aproximadamente 1,7 milhão de pessoas vivendo com HIV, compondo uma epidemia diversa, porém com a maior parte das transmissões atribuída a HSH e trabalhadores do sexo. Na maioria dos países desta região, as taxas de prevalência são inferiores a 1%, porém entre grupos específicos, como trabalhadores do sexo e HSH a prevalência é mais alta, atingindo patamares >5% e >20%, respectivamente (Beyrer et al. 2012; Prüss-Ustün et al. 2013; UNAIDS 2014).

1.2.1 A epidemia de HIV/aids no Brasil

O Brasil é responsável por um terço das infecções da América Latina, com uma estimativa de aproximadamente 734 mil pessoas vivendo com HIV/aids em 2014, o que corresponde a uma prevalência nacional menor que 1%. Porém nos grupos populacionais em situação de maior vulnerabilidade, tais como HSH, travestis, profissionais do sexo e usuários de drogas são observadas taxas de prevalência mais altas (Brasil 2014), podendo chegar a 4,9% em mulheres profissionais do sexo (Malta et al. 2010) e 14,2% em HSH (Kerr et al. 2013), o que caracteriza o Brasil como um país com epidemia concentrada em populações que fazem parte dos grupos de maior vulnerabilidade. Na verdade, o mais correto seria considerar a epidemia do HIV no Brasil como um mosaico de epidemias regionais, refletindo a extensão e a diversidade sócio-geográfica do país e sua heterogeneidade regional. Os motivos da alta prevalência do HIV nas populações de maior vulnerabilidade seriam as barreiras significativas no acesso aos serviços de saúde, que têm na sua base o preconceito, o estigma e a discriminação, como também a persistência das práticas de sexo anal desprotegido no grupo de HSH. A principal e mais forte estratégia de prevenção adotada pelo país é o foco das ações nestas populações mais vulneráveis, oferecendo acesso a preservativos, combinado

com intervenções comunitárias, além da oferta de teste anti-HIV para um diagnóstico cada vez mais precoce (Brasil 2008).

Dados do Ministério da Saúde indicam que a epidemia se encontra em um processo de estabilização, ainda que em patamares elevados. A prevalência nacional da infecção pelo HIV na população de 15-49 anos está estável em aproximadamente 0,6% desde 2004, sendo 0,3% em mulheres e 0,7% em homens. O crescimento dos casos de aids na população feminina afeta especialmente as mulheres em idade reprodutiva (Brasil 2014).

Os mais de 750 mil casos de aids notificados até 2014 não se distribuem de forma homogênea entre as regiões do país. O maior número se concentra na região Sudeste (54,4%), seguido pelas regiões Sul (20,0%), Nordeste (14,3%), Centro-Oeste (5,8%) e Norte (5,4%) do Brasil. Em 2013, foram notificados cerca de 39 mil casos de aids no Brasil, valor que vem se mantendo estável nos últimos 5 anos. A taxa de detecção nacional foi de 20,4 casos para cada 100 mil habitantes, sendo a maior encontrada na região Sul (30,5/100.000 habitantes), seguida pela região Norte (26,1), Centro-Oeste (20,3), Sudeste (18,7) e Nordeste (16,0). A incidência nacional reduziu cerca de 10% nos últimos 10 anos, contudo, existem diferenças significativas nas taxas das diferentes regiões brasileiras. Entre 2003 e 2013, a incidência diminuiu 32% na região Sudeste, 9% na região Sul, permaneceu estável na região Centro-Oeste, enquanto nas demais regiões observa-se um aumento, sendo de 62% na região Nordeste e 120% na região Norte (Brasil 2014).

O estado do Rio de Janeiro tem o segundo maior número de casos de aids no Brasil, respondendo por 11% do total de casos notificados no país. A cidade de Rio de Janeiro é a segunda maior cidade do Brasil e sua área metropolitana tem uma população de mais de 10 milhões de habitantes, onde residem cerca de 90% dos indivíduos vivendo com HIV/aids no estado (Brasil 2012a).

Vale ressaltar que o tratamento de pacientes com HIV no Brasil continua sendo considerado uma referência entre países em desenvolvimento, por haver oferta de acesso gratuito aos medicamentos antirretrovirais. Além disso, em dezembro de 2013, o Brasil tornou-se o primeiro país em desenvolvimento e o terceiro do mundo a recomendar o início imediato da TARV para todas as pessoas vivendo com HIV/aids, independentemente da contagem de células T $CD4^+$, considerando a motivação do paciente (Brasil 2013), implementando a terapia não somente para a restauração da saúde do paciente, como também uma medida de saúde pública para o controle da transmissão do HIV.

Quase 400 mil pessoas vivendo com HIV/aids estavam recebendo TARV em 2014, porém ainda existe um número considerável de pessoas que possuem indicação de tratamento, mas não estão sendo tratadas por razões diversas. Desde a introdução da TARV, a taxa de mortalidade causada pela aids foi reduzida de 9,6 mortes anuais por 100 mil habitantes em 1996, para 5,7 em 2013 (Brasil 2014). Outra consequência do acesso à TARV é a redução do número de casos de transmissão vertical do HIV (TV), isto é, quando o HIV é transmitido da mulher grávida para seu bebê, seja durante a gravidez, no parto ou pela amamentação.

1.2.2 Infecção pelo HIV-1 em gestantes e impacto na transmissão vertical

De acordo com o relatório da UNAIDS de 2014, 16 milhões de mulheres vivem com HIV/aids globalmente, o que corresponde a aproximadamente 47% do total de adultos infectados (UNAIDS 2014). No Brasil, embora a epidemia esteja concentrada nas populações de maior vulnerabilidade, nas quais a prevalência da infecção pelo HIV-1 pode atingir proporções maiores a 10%, a principal via de transmissão do HIV quando se consideram os números absolutos é a heterossexual. Esta via vem se estabelecendo desde o início dos anos 1990 e consequentemente levou a um deslocamento do número de casos notificados de aids dos grupos populacionais em situação de maior vulnerabilidade para a população geral, aumentando consideravelmente a frequência de casos em mulheres (Brasil 2014).

A crescimento dos casos de AIDS na população feminina afeta especialmente as mulheres em idade reprodutiva e entre os anos 2000 e 2013, 77 mil casos de infecção pelo HIV-1 em gestantes foram notificados no país, dos quais 41,7% concentraram-se na região Sudeste. As maiores proporções de gestantes infectadas pelo HIV-1 estão concentradas na faixa etária de 20 a 29 anos (Brasil 2014). A taxa de detecção de gestantes HIV positivas no Brasil vem apresentando uma tendência de aumento estatisticamente significativo nos últimos dez anos, passando de 2,0 casos a cada mil nascidos vivos em 2004 para 2,5 casos a cada 1000 nascidos vivos em 2013, entretanto na região Sudeste uma tendência oposta foi observada, apresentando uma queda de 16% quando se comparam a última década. Mesmo com a redução das taxas de detecção do HIV em gestantes na região Sudeste, o estado do Rio de Janeiro ainda apresenta índices superiores à média nacional, com 3,0 casos para cada mil nascidos vivos.

As taxas de TV do HIV-1 sem qualquer intervenção se situam em torno de 20% (Connor et al. 1994) porém, com a utilização de antirretrovirais combinados,

reduzindo a carga viral a valores menores do que 1.000 cópias de RNA/mL e uso de cesariana eletiva, é possível reduzir a TV para níveis inferiores a 2% (World Health Organization 2010). No Brasil, dos 237 de casos de aids notificados em crianças menores de 13 anos em 2013, 99,6% ocorreram por TV, porém o número absoluto de casos tem diminuído acentuadamente ao longo dos últimos anos (Brasil 2014).

De acordo com as recomendações para profilaxia da TV e terapia antirretroviral do Ministério da Saúde, as gestantes portadoras de infecção pelo HIV deverão receber sempre quimioprofilaxia, com associação de três antirretrovirais, de duas classes diferentes, preferencialmente dois inibidores nucleosídicos da transcriptase reversa associado a um inibidor da protease ou, alternativamente, dois inibidores nucleosídicos da transcriptase reversa e um inibidor não-nucleosídico da transcriptase reversa, independentemente da situação virológica, clínica ou imunológica, com o objetivo de prevenir a transmissão vertical (Brasil 2010). A introdução da TARV deve ser precoce, após o primeiro trimestre, entre a 14ª e a 28ª semana de gravidez. A suspensão de todo o esquema ARV após o parto, em gestantes com indicação apenas de profilaxia, deve ser sistemática, com interrupção do inibidor não-nucleosídico da transcriptase reversa duas semanas antes dos outros ARV que compõem o esquema, reduzindo-se assim, o risco de desenvolvimento de mutações de resistência (Brasil 2010).

Em um estudo pioneiro, conduzido em escala nacional envolvendo 63 serviços de saúde de 20 estados (incluindo o Distrito Federal) a taxa de transmissão vertical do HIV foi estimada em 8.6% em 2000 e 7.1% em 2001, porém foi observado a existência de variação entre as taxas de diferentes regiões (Menezes Succi 2007). Outros estudos, realizados em menor escala, reforçaram a existência deste panorama complexo, com taxas variando de nulas no Rio de Janeiro (Pilotto et al. 2013), entre 2,7%-6% em São Paulo (Gonçalves et al. 2011; Matida et al. 2011), 8,2% na Bahia (Patricio et al. 2015) e próximas a 10% no Espírito Santo e Amazonas (Soeiro et al. 2011; Vieira et al. 2011). Entretanto, a variação das taxas de transmissão vertical observadas entre os diferentes estados pode ser o reflexo da adesão ao acompanhamento pré-natal, do tratamento quimioprofilático com antirretrovirais antes e/ou durante o parto e no neonatal, além da prática do aleitamento materno.

1.3 Classificação e ciclo replicativo do HIV-1

O HIV pertence ao gênero *Lentivirus* da família *Retroviridae*. Membros desta família possuem um genoma que consiste de duas fitas de RNA de polaridade positiva com cerca de 10 Kb e composto por 9 genes. A leitura de cada fita de RNA é organizada em três fases abertas principais e que apresentam sobreposição, constituindo um RNA policistrônico (Figura 3) (Herschhorn e Hizi 2010).

Brevemente, o ciclo de replicação do HIV se inicia com a etapa de adsorção, através do reconhecimento das moléculas de CD4 presentes na superfície de células susceptíveis pelas glicoproteínas de superfície do HIV (gp120), em conjunto com a subsequente interação com um co-receptor das famílias CC ou CXC de receptores de quimiocinas, sendo CCR5 e CXCR4 os mais importantes. Posteriormente, o envelope viral se fusiona com a membrana plasmática da célula mediada pela glicoproteína transmembranar (gp41), que após uma mudança conformacional, expõe as regiões que compõem o peptídeo de fusão (Wilen et al. 2012).

Partículas virais parcialmente desencapsuladas são transportadas em direção ao núcleo enquanto o processo de transcrição reversa do RNA genômico ocorre pela ação da transcriptase reversa. Após a transcrição reversa, o cDNA recém formado se associa a um complexo composto por proteínas virais e celulares denominado de complexo pré-integração (CPI). A enzima viral integrase é uma das que compõe o CPI, e parece mediar a importação desse complexo ao núcleo, auxiliando a integração do cDNA ao DNA cromossômico da célula infectada (Craigie e Bushman 2012; Hu e Hughes 2012a). Nos linfócitos T ativados, as cópias integradas do DNA do HIV (provírus) funcionam como molde para a RNA polimerase II celular, que realiza a síntese do RNA viral. A regulação da síntese do RNA viral é realizada pelo LTR (repetições terminais longas, do inglês *long terminal repeats*). Estas terminações são produzidas em ambas as extremidades do DNA proviral durante o processo de transcrição reversa. O LTR na região 5’ do provírus serve como um promotor da transcrição do genoma viral, enquanto o LTR da região 3’ auxilia a poliadenilação do RNA viral transcrito.

RNAs mensageiros transcritos inalterados ou parcialmente processados são exportados do núcleo para o citoplasma onde ocorrerá a tradução (Karn e Stoltzfus 2012). O precursor gp160 é traduzido no retículo endoplasmático, enquanto as poliproteínas Gag e Gag-Pol são sintetizadas por ribossomos livres no citosol, e ao

se associar com dímeros de RNA genômico, concentram-se na membrana plasmática. Isto resulta na formação das partículas esféricas imaturas contendo as glicoproteínas transmembranar e de superfície. O processamento proteolítico das proteínas Gag e Pol pela PR durante ou imediatamente após a liberação da partícula gera os vírions maduros do HIV (Sundquist e Kräusslich 2012).

Após a formação da partícula viral madura, o HIV pode infectar outras células susceptíveis na forma de partículas livres, ou através da formação de sinapses virológicas mediadas pelo contato célula-a-célula (Sundquist e Kräusslich 2012), tendo esta última forma sido demonstrada como a mais eficiente por facilitar os mecanismos de evasão da resposta imune (Abela et al. 2012) e da ação dos antirretrovirais (Sigal et al. 2011) (Figura 3). O HIV é capaz de replicar-se em uma grande variedade de tipos celulares, como por exemplo monócitos, macrófagos, linfócitos T e linfócitos B, porém é a depleção de linfócitos T CD4 auxiliares, células com um importante papel no sistema imune do hospedeiro, a principal causa da aids.

A região genômica *gag* (antígenos grupo específico, do inglês *group-specific antigens*) codifica as proteínas do capsídeo. Seu precursor é a proteína p55, que é processada pela protease viral durante ou após a liberação da progênie viral em matriz (p17), capsídeo (p24), nuclecapsídeo (p7), proteína p6 e dois peptídeos espaçadores denominados SP1 e SP2. A região genômica *pol* codifica as enzimas virais protease (PR), transcriptase reversa (TR) e integrase, responsáveis respectivamente pela: hidrólise das poliproteínas virais em produtos proteicos funcionais; polimerização do DNA complementar (cDNA) viral a partir do molde de RNA genômico; e integração do cDNA viral no cromossomo celular do hospedeiro (Fanales-Belasio et al. 2010).

A região genômica *env* codifica as proteínas externas do envelope do vírus e tem um papel fundamental na entrada do vírus na célula. As glicoproteínas virais são produzidas como um precursor (gp160) que é processado por uma enzima celular, originando a glicoproteína de superfície (gp120) e a glicoproteína transmembranar (gp41). A gp120 contém os sítios de ligação para o receptor CD4 como também para receptores de quimiocina, que atuam respectivamente como receptores e co-receptores de entrada para o HIV-1. Estas glicoproteínas virais estão sujeitas a uma alta pressão seletiva diversificadora, resultando na seleção de variantes virais com capacidade de evadir o controle do sistema imune (Wilen et al. 2012).

As demais proteínas codificadas pelo HIV (Vif, Vpr, Tat, Rev, Vpu, Vpx e Nef) são os produtos primários da tradução dos RNA mensageiros que sofreram *splicing*. Tat e Rev são fatores regulatórios essenciais para a expressão dos genes do HIV. Vif, Vpr, Vpu, Vpx e Nef são proteínas acessórias produzidas pelos lentivírus de primatas (SIV e HIV) (Karn e Stoltzfus 2012). Elas possuem um papel na replicação viral *in vivo*, uma vez que neutralizam o efeito de fatores de restrição do hospedeiro, intensificam a replicação viral e infectividade dos vírions ou facilitam a evasão viral da resposta imune adaptativa. A proteína Vif é responsável pelo bloqueio da potente função antiviral da proteína APOBEC3, um fator de restrição do hospedeiro que promove um acúmulo de mutações guanina-para-adenina no genoma viral (Malim 2009). Nef é responsável na manipulação da maquinaria celular, ativando as células T e estabelecendo o estado de infecção persistente (Das e Jameel 2005). A proteína Vpr participa na regulação da importação nuclear do complexo de pré-integração do HIV (Craigie e Bushman 2012). A proteína Vpx é codificada pela linhagem do HIV-2 e é responsável pela inibição de SAMHD1, um fator de restrição responsável pela redução do estoque celular de desoxirribonucleotídeos (Santa-Marta et al. 2013). Vpu é encontrado somente no HIV-1 e SIVcpz, sendo responsável pela degradação do receptor viral CD4 e promove a liberação da progênese viral das células infectadas pelo HIV-1 através da antagonização do fator de restrição humano Teterina (Santa-Marta et al. 2013) (Figura 3).

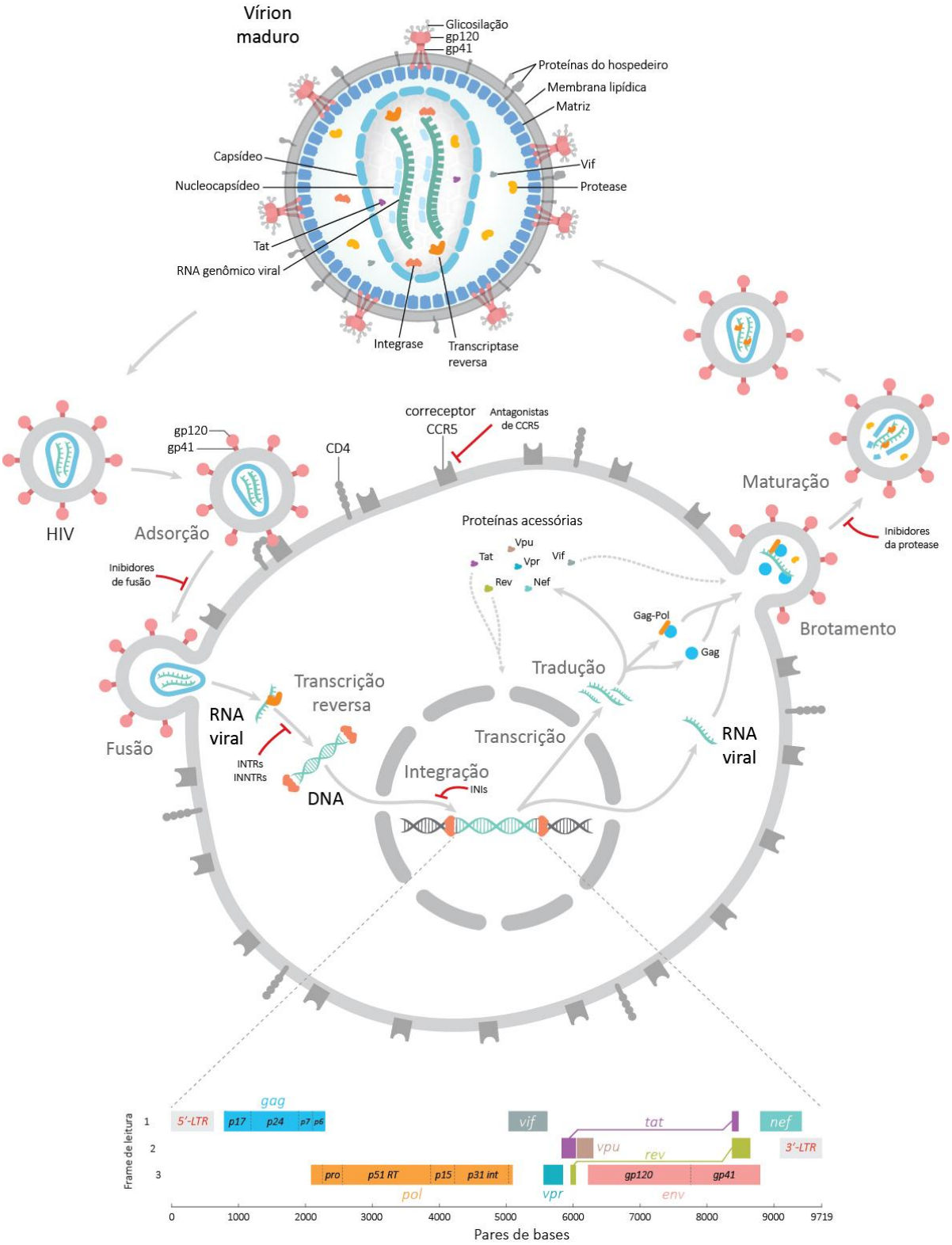


Figura 3 – Ciclo replicativo do HIV. A infecção da célula começa quando a proteína gp120 se liga ao receptor CD4 e subsequentemente ao correceptor CCR5. O resultado é a fusão das membranas e posterior liberação do capsídeo viral no citoplasma. Após a entrada, o RNA viral é retrotranscrito e em seguida direcionado ao núcleo para ser integrado ao genoma celular. A transcrição do provírus gera RNA mensageiros de diferentes comprimentos que serão traduzidos nas diferentes proteínas virais. O brotamento e liberação das partículas virais são acompanhadas ou são precedidas brevemente pela maturação mediada pela protease, originando a partícula viral infecciosa. Cada passo do ciclo replicativo do HIV é um alvo em potencial para a intervenção de antirretrovirais, sendo as principais classes denotadas na figura. INTR - inibidor nucleosídeo da transcriptase reversa, INNTR – inibidor não-nucleosídeo da transcriptase reversa, INI – inibidor da integrase. Modificado de Thomas Splettstoesser (http://www.scistyle.com), com base em Engelman e Cherepanov 2012; Goodsell 2012.

As altas taxas evolutivas descritas para o HIV (Rambaut et al. 2004) são uma consequência dos erros que ocorrem durante a replicação viral, fundamentalmente durante a transcrição reversa do RNA em DNA pela TR, uma vez que esta enzima não possui atividade de correção (Hu e Hughes 2012b). Recentemente, foi proposto que o mecanismo intrínseco de controle antirretroviral promovido pelas citidinas deaminases da família gênica APOBEC3 humana atuando em níveis subótimos, causado pelo bloqueio parcial de sua ação pela proteína viral Vif por exemplo, pode promover mutações esporádicas no genoma do HIV que fornecem uma fonte de variação genética em que a seleção natural pode agir, em adição à variabilidade gerada pela TR (Kim et al. 2014).

A baixa fidelidade do processo replicativo e os erros de pareamento durante a replicação viral geram inserções, deleções, duplicações (Abram et al. 2010). A taxa de mutação para a replicação do HIV-1, é de aproximadamente 2×10^{-5} por nucleotídeo por ciclo replicativo (Hu e Hughes 2012b). Devido à alta frequência de erros, essa dinâmica garante que cada novo vírus tenha praticamente um genoma único e diferente de seu parental. A baixa fidelidade da TR, o surgimento de aproximadamente 10^{10} - 10^{12} diferentes vírions por dia dentro de um mesmo hospedeiro (Perelson et al. 1996), o grande número de indivíduos infectados e a persistência da infecção por vários anos compõem um conjunto de fatores ideais que favorecem a geração e expansão da diversidade viral atual (Aldrich e Hemelaar 2012; Hemelaar et al. 2011).

Além das mutações geradas durante o processo de replicação, a recombinação homóloga também faz parte dos mecanismos de geração de diversidade HIV. Múltiplos passos são necessários para a geração de um novo recombinante, e antes de tudo, a célula produtora de vírus precisa estar infectada por mais de um vírus. A dupla infecção pode ocorrer em um intervalo de tempo próximo (co-infecção) ou uma nova infecção pode ocorrer em um indivíduo com uma infecção pelo HIV já estabelecida (superinfecção). Os RNAs genômicos provenientes destes dois provírus precisam ser co-empacotados num mesmo vírion, de forma que durante a transcrição reversa dos genomas ocorra uma mudança de fita molde, gerando uma cópia de DNA quimérica, que então deve se integrar no genoma da célula alvo. Finalmente, este provírus recombinante tem de ser capaz de gerar vírus por sua vez aptos de se replicar para que o evento de recombinação possa ser observado (Nájera et al. 2002). É importante notar que o processo de recombinação pode ocorrer mesmo que as duas sequências presentes no vírion

sejam do mesmo genótipo, porém, somente a recombinação entre duas moléculas de RNA diferentes pode gerar um genótipo distinto dos parentais (Burke 1997).

1.4 Origem do HIV

Logo após a descoberta do HIV-1 e do HIV-2, outros vírus, coletivamente denominados vírus da imunodeficiência símia (SIV) foram detectados em diferentes primatas da África Subsaariana, incluindo o macaco verde africano (*Chlorocebus sabaeus*), macacos mangabeys (*Cercocebus atys*), mandris (*Mandrillus sphinx*), chimpanzés (*Pan troglodytes*) e gorilas (*Gorilla gorila*) (Beer et al. 1999; Peeters et al. 2001; Peeters e Courgnaud 2002). Estes vírus mostraram-se não patogênicos aos seus hospedeiros naturais, embora fossem muito relacionados ao HIV (Hirsch et al. 1995).

Pesquisas recentes demonstram que os SIVs estão presentes na África infectando seus hospedeiros símios há mais de 30.000 anos (Worobey et al. 2010). Desta forma, as diversas espécies símias representam um antigo e extenso reservatório de vírus com o potencial para infectar e disseminar uma nova epidemia em outras espécies (Hahn 2000). Estes achados forneceram evidências para a formulação da hipótese de surgimento do HIV como uma provável transmissão zoonótica de vírus que infectavam símios para o ser humano. Esta hipótese foi testada através de análises filogenéticas comparando sequências do genoma dos diferentes linhagens de SIVs e HIVs (Gao et al. 1999; Keele et al. 2006; Locatelli e Peeters 2012; Sharp et al. 1995).

Diferentes estudos caracterizaram o surgimento das duas linhagens do HIV (HIV-1 e HIV-2) como resultado de, pelo menos, doze transmissões zoonóticas independentes do SIV para humanos (Sharp e Hahn 2011). As duas linhagens humanas não apresentam relação genética entre si, porém estão relacionadas com diferentes linhagens de SIV que circulam na África Centro-Ocidental (D'arc et al. 2015; Keele et al. 2006; Sharp e Hahn 2011).

Análises filogenéticas do HIV-1 demonstraram que este vírus está relacionado aos SIVs de grandes primatas, no caso, chimpanzé (*Pan troglodytes*, SIVcpz) (Keele et al. 2006; Paraskevis et al. 2003) e gorila (*Gorilla gorilla*, SIVgor) (D'arc et al. 2015; Van Heuverswyn et al. 2006; Takehisa et al. 2009). 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

que permitiram a ocorrência de transmissão entre humanos (Gürtler et al. 1994; De Leys et al. 1990; Plantier et al. 2009; Simon et al. 1998). Estas quatro introduções independentes deram origem a diferentes linhagens denominadas grupos M (*major*), N (*non-M, non-O*), O (*outlier*) e P (Plantier et al. 2009; Robertson et al. 2000).

Todos os quatro grupos do HIV-1, assim como o SIVgor são relacionados ao SIVcpz, particularmente, com cepas SIVcpz*Ptt* da subespécie *Pan troglodytes troglodytes* nativos da região central da África, sugerindo que esta subespécie atua como o reservatório natural das infecções tanto de humanos quanto de gorilas (Keele et al. 2006; Takehisa et al. 2009) (Figura 4). Os grupos M e N são muito relacionados a isolados do SIVcpz*Ptt* do sul de Camarões, indicando este local como a provável origem destes grupos (Van Heuverswyn et al. 2007; Keele et al. 2006). As prováveis fontes zoonóticas imediatas dos grupos O e P foram recentemente confirmadas como o SIVgor, que foi transmitido de gorilas para humanos. Apesar dos isolados de SIVgor terem sido encontrados no sul dos Camarões, a identificação exata do local onde a transmissão zoonótica ocorreu não foi possível (D'arc et al. 2015; Plantier et al. 2009).

Estudos filogenéticos mais robustos, utilizando estratégias de reconstrução temporal baseadas na teoria do relógio molecular, estimaram a origem das epidemias dos grupos M e O entre 1910 e 1930 (Faria et al. 2014; Korber 2000; Lemey et al. 2004; Worobey et al. 2008). Em contraste, os grupos N e P parecem ter emergido mais recentemente, devido aos baixos níveis de diversidade genética observados (Sharp e Hahn 2011). O grupo O está restrito ao centro-oeste da África, tendo se disseminado por Camarões, Gabão, Nigéria e países vizinhos, e estima-se que tenha infectado cerca de 100.000 indivíduos. Os grupos N e P são extremamente raros, encontrados somente nos Camarões, sendo o grupo N descrito em 20 indivíduos e o grupo P em dois (Peeters et al. 1997; Plantier et al. 2009; Vallari et al. 2011; Vallari et al. 2010). O grupo M foi o primeiro a ser descoberto e representa a forma pandêmica do HIV-1, infectando milhões de pessoas no mundo (UNAIDS 2013).

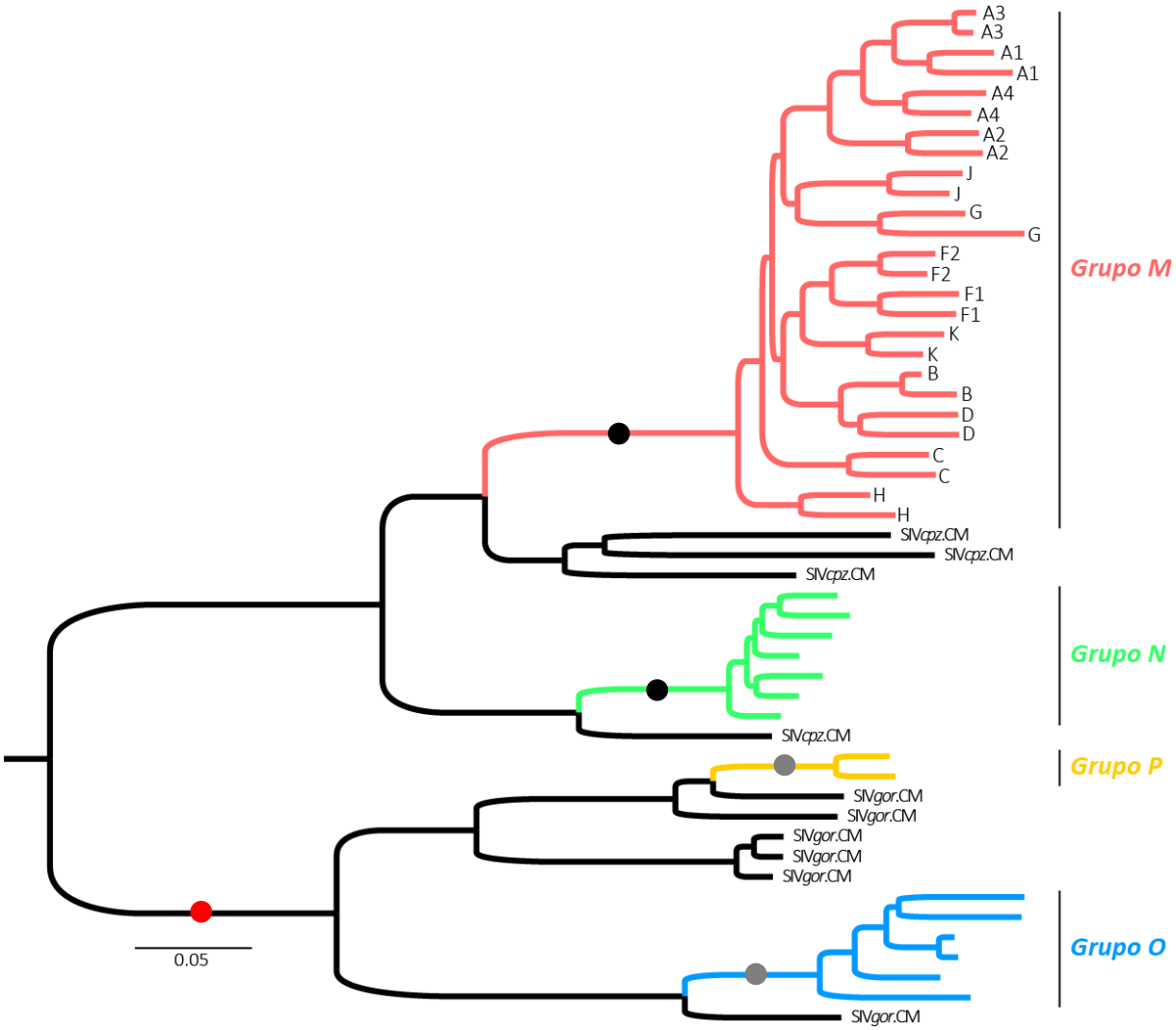


Figura 4 – Origens do HIV-1. Árvore filogenética de máxima-verossimilhança construída a partir da sequência do gene *pol* dos grupos do HIV-1 e dos isolados do SIV mais próximos a cada um destes grupos. Cada grupo do HIV-1 forma um clado monofilético distinto (M, N, O e P, indicado pelas diferentes cores) relacionado aos SIVs isolados de distintos grupos de primatas não humanos do Camarões (CM). Isto evidencia os quatro eventos de transmissão independentes a partir do SIV*gor* (*Gorilla gorilla*) e SIV*cpzPtt* (*Pan troglodytes troglodytes*), cujos ramos estão coloridos em preto. Círculos indicam os diferentes eventos de transmissão zoonótica: preto - SIV*cpzPtt* para humanos; cinza - SIV*gor* para humanos; vermelho – SIV*cpzPtt* para gorilas. Cada subtipo e subsubtipo do grupo M do HIV-1 está representado por duas sequências como indicado. Os comprimentos dos ramos horizontais são proporcionais à barra da parte inferior, indicando substituição de nucleotídeos por sítio.

A origem do HIV-2 é atribuída ao contato entre humanos e macacos mangabey (*Cercocebus atys*) distribuídos na região ocidental África, se estendendo do Senegal à Costa do Marfim (Lemey et al. 2003; Reeves e Doms 2002; Santiago et al. 2005). Desde o isolamento inicial, pelo menos oito linhagens distintas do HIV-2 foram identificadas (Figura 5), cada uma aparentando se tratar de uma transferência zoonótica independente do SIV infectando macacos mangabey (SIVsmm) para o homem (Sharp e Hahn 2011). Cada transferência pode ser categorizada como um

grupo epidêmico, classificados de A até H, embora somente os grupos A e B tenham se disseminado na população humana de forma considerável (Sharp e Hahn 2011).

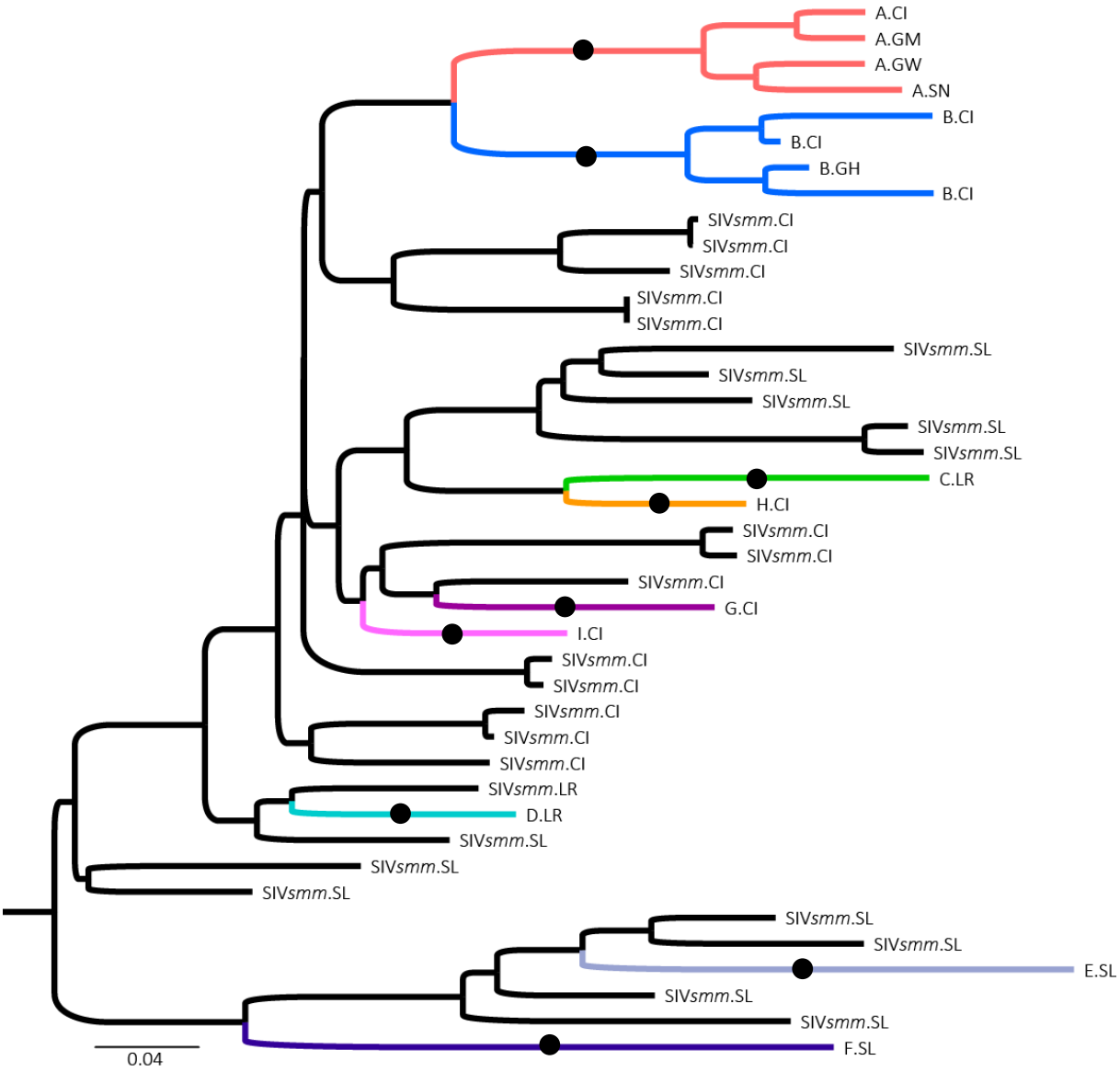


Figura 5 – Origens do HIV-2. Árvore filogenética de máxima-verossimilhança construída a partir da sequência do gene *gag* dos grupos do HIV-2 e os isolados do SIV*smm* mais próximos. Cada um dos oito grupos do HIV-2 (A-H) representa uma introdução zoonótica inpendente nos humanos, e estão representados por diferentes cores. A sequência do suposto grupo I foi incluída (Ayouba et al., 2013). Círculos pretos indicam os diferentes eventos de transmissão zoonótica do SIV*smm* para humanos. Os ramos de SIVsmm estão representadas em preto, com o país de isolamento indicado: CI = Costa do Marfim, SL = Serra Leoa, GH = Gana, LR = Libéria, GM = Gâmbia, GW = Guiné Bissau, SN = Senegal. Os comprimentos dos ramos horizontais são proporcionais à barra da parte inferior, indicando substituição de nucleotídeos por sítio.

Dentre os oito grupos que compõem o HIV-2, somente os grupos A e B infectam significativamente a população humana. Dados de soroprevalência e estimativas evolutivas sugerem que o HIV-2 tenha surgido em Guiné-Bissau e os

eventos de transmissão interespecífica que originaram os grupos A e B do HIV-2 ocorreram na década de 1940, enquanto que a difusão efetiva do vírus para populações humanas só tenha ocorrido após 20 a 30 anos (Lemey et al. 2003). Atualmente o HIV-2 circula moderadamente em países do África Ocidental, principalmente Senegal, Guiné, Gambia e Cabo Verde (Campbell-Yesufu e Gandhi 2011; de Pina-Araujo et al. 2014).

Mesmo após a identificação da origem espacial e temporal dos principais eventos de transmissão zoonótica dos diferentes SIVs de primatas não humanos para o homem, permanece a questão de como tais eventos ocorreram. Existem várias hipóteses, porém uma das mais aceitas é a “teoria do caçador” (Moore 2004), na qual a transferência do SIV para humanos teria sido o resultado da caça de chimpanzés e outros símios infectados, tendo ocorrido através da alimentação da carne e/ou do contato do sangue contaminado com algum ferimento do caçador. Normalmente, o sistema imunológico do caçador conseguiria combater o SIV, porém, em algumas ocasiões o vírus teria se adaptado ao novo hospedeiro, dando origem ao HIV (Sharp et al. 2001; Sharp e Hahn 2011). Cada passagem para o novo hospedeiro humano que resultou em uma adaptação bem sucedida ao mesmo teria originado a uma nova linhagem (grupos M, N, O e P no HIV-1, e grupos A-H do HIV-2) e existem evidências de que essas transmissões zoonóticas continuam ocorrendo (Ayouba et al. 2013).

1.5 A diversidade genética global do HIV

As sequências genômicas do HIV variam consideravelmente, não somente entre indivíduos, mas também no indivíduo infectado. A habilidade do vírus acumular mutações rapidamente o torna capaz de burlar os mecanismos de resposta imune do hospedeiro e se tornar resistente às drogas antirretrovirais, o que torna o tratamento e o desenvolvimento de uma vacina efetiva cada vez mais desafiadores.

Devido ao seu grau de diversidade genética e pelo grande número de indivíduos infectados, o grupo M do HIV-1 se diversificou em diferentes linhagens, denominadas subtipos. Esta diversificação ocorreu logo após a transmissão zoonótica para o homem, ainda na região do centro-oeste da África. Diferentes linhagens destes subtipos disseminaram-se globalmente, em função do resultado de efeitos fundadores (estabelecimentos independentes de uma linhagem proveniente de uma região que atua como fonte para novas populações) e recombinações entre

e dentro de diferentes linhagens (Rambaut et al. 2004). As chances distintas de disseminação e estabelecimento das diferentes linhagens originadas no epicentro da origem do HIV-1 para outras regiões geográficas moldaram a distribuição global dos subtipos do HIV-1 observada atualmente (Hemelaar 2012; Hemelaar et al. 2011).

A variação genética dentro de um subtipo (intrassubtipo) é da ordem de 8%, enquanto que a variação entre subtipos (intersubtipo) é geralmente próxima a 15% dependendo de quais subtipos e regiões genômicas são comparados (Li et al. 2015). Este elevado grau de diversidade é uma das causas que limitam a reatividade cruzada da resposta imune intra- e intersubtipos. Este fator torna a escolha dos subtipos que comporiam uma provável vacina um critério decisivo em sua eficiência (Hemelaar et al. 2011). Além da resposta vacinal, a diversidade do HIV-1 pode impactar o diagnóstico (Kim et al. 2007), afetar a resposta ao tratamento e o desenvolvimento de resistência (Camacho e Vandamme 2007; Geretti et al. 2009) e a progressão da doença (Baeten et al. 2007; Tarosso et al. 2013), o que torna estudos de epidemiologia molecular fundamentais para avaliar o impacto da progressão da diversidade na epidemia do HIV.

Atualmente, o grupo M do HIV-1 é composto por nove subtipos (A-D, F-H, J, K), dois conjuntos de subsubtipos (A1-A4, F1-F2), além de uma miríade de formas recombinantes circulantes (CRFs, nomeadas de acordo com a ordem de descoberta e subtipos que os compõem) e únicas (URFs) (Taylor e Hammer 2008). Recombinantes entre subtipos são designados como CRFs se o genoma completo de três ou mais vírus de indivíduos não diretamente relacionados epidemiologicamente forem sequenciados e estes apresentarem os mesmos pontos de recombinação; e URFs se este critério não for cumprido. Até setembro de 2015, 74 diferentes CRFs do HIV-1 tinham sido descritos (Los Alamos HIV Database, http://www.hiv.lanl.gov) e a tendência é que este número continue a crescer.

Como resultado do aumento da amostragem e sequenciamento de isolados do HIV provenientes de países da África Central, novas propostas de classificação das linhagens encontradas do grupo M do HIV-1 têm surgido. Dentre as novas propostas estão a identificação do subtipo L (Vallari et al. 2015) e da existência do subsubtipo A5 (Vidal et al. 2009) na República Democrática do Congo. Adicionalmente, também foi proposta a existência dos subsubtipos A5 e A6 em Angola (Bártolo et al. 2009). Contudo, estas propostas são preliminares, não cumprindo todos os requisitos para uma mudança concreta na nomenclatura do HIV-1 (Robertson et al. 2000)

A disseminação global do grupo M do HIV-1 na segunda metade do século 20 levou a uma distribuição heterogênea dos subtipos e recombinantes (Figura 6). Numa escala global, o subtipo C é a linhagem predominante, responsável por quase 50% das infecções, seguido pelo subtipo A (12%), subtipo B (11%), CRF02_AG (7%), CRF01_AE (5%) e subtipo G (5%). Formas recombinantes representam aproximadamente 20% das infecções (Hemelaar et al. 2011).

O predomínio do subtipo C do HIV-1 pode ser explicado graças a região geográfica no qual ele se distribui, composta pelos países da África Meridional, Índia e adicionalmente países da África Oriental (Hemelaar et al. 2011), países com as mais altas taxas de infecção e bastante populosos. Dos dez países com o maior número de pessoas infectadas pelo HIV-1, sete são dominados por este subtipo (UNAIDS 2013).

A maior diversidade do HIV-1 é encontrada na região da África Central, berço de origem do HIV-1, onde todos os subtipos e muitos CRFs e URFs são verificados (Figura 6). Na região do oeste da África, uma grande variedade de subtipos também é encontrada, porém com predomínio do CRF02_AG e do subtipo G. Nestas regiões também circulam muitos CRFs complexos, formados por subtipos raramente encontrados na forma pura, como os subtipos H, J e K. Dentre os CRF complexos mais importantes presentes, destacam-se o CRF06_cpx que atinge prevalências de >10% na África Ocidental, além do CRF09_cpx, CRF11_cpx, CRF13_cpx e CRF45_cpx, que normalmente apresentam prevalências inferiores a 5% (Lihana et al. 2012).

O subtipo C corresponde a mais de 95% das infecções pelo HIV em todos os países da África Meridional. Enquanto que na região leste da África a maioria das infecções é pelo subtipo A, seguido pelos subtipos C e D que exibem prevalência variadas. As exceções são o Burundi, Djibuti e Etiópia, nos quais predomina o subtipo C, com prevalências maiores que 70%. Na região norte da África predominam o subtipo B e vários CRFs (Aldrich e Hemelaar 2012; Hemelaar et al. 2011).

Nas regiões sul e sudeste da Ásia, o CRF01_AE é responsável pela grande maioria das infecções, com exceção da Índia, onde ocorre o predomínio do subtipo C. A epidemia na região norte da Ásia é composta pelos subtipos A e B. Na região leste da Ásia, a epidemia é dominada por CRF07_BC, CRF08_BC, CRF01_AE e subtipo B. Na Oceania, a epidemia é composta pelos subtipos B e C (Hemelaar et al. 2011). Na Europa Central e Ocidental predomina o subtipo B, enquanto que na

Europa Oriental predominam os subtipos A e B. No continente americano predomina o subtipo B, espalhado por toda sua área geográfica (Hemelaar et al. 2011).

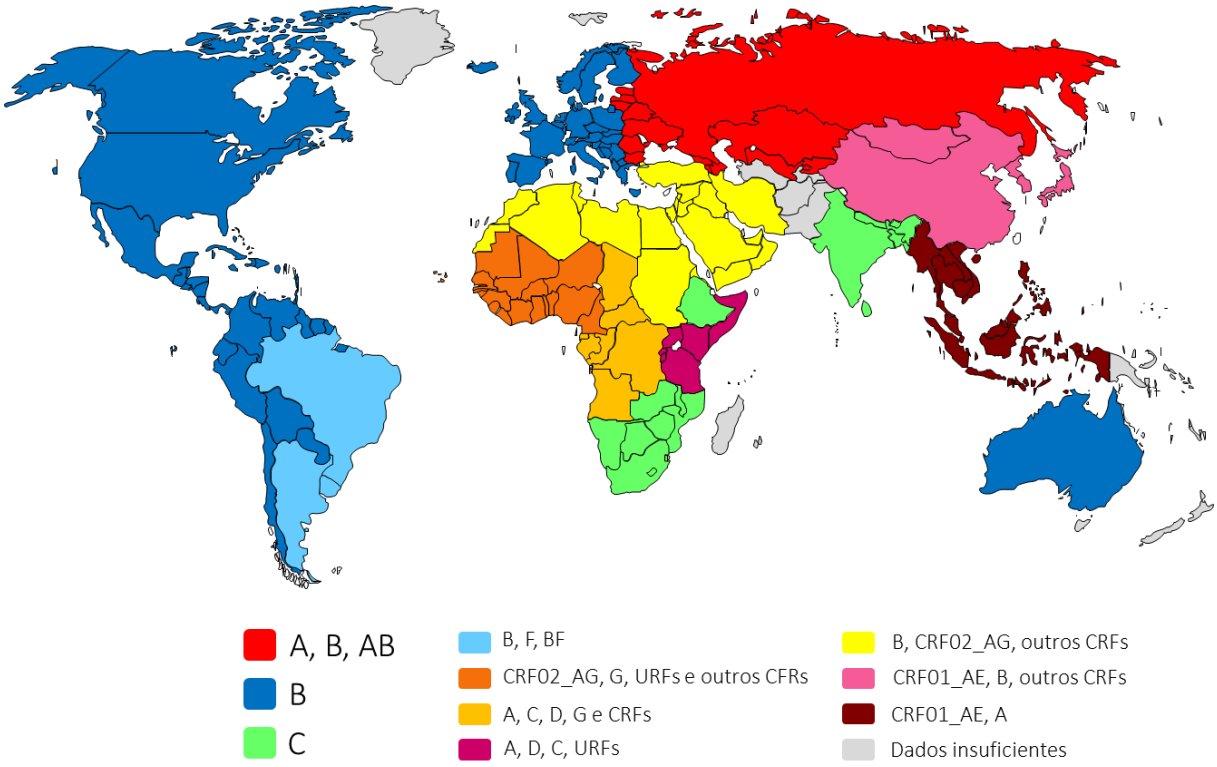


Figura 6 – Distribuição global dos subtipos e formas recombinantes do HIV-1. Formas recombinantes são majoritariamente encontradas na região central de África (com exceção do CRF02_AG na região oeste deste continente), enquanto o subtipo B é a linhagem mais dispersa e predomina em grande parte da Europa, Oceania e América (baseado em Hemelaar et al. 2011; Taylor e Hammer 2008).

Mesmo que a epidemiologia molecular do grupo M nas Américas não seja tão complexa quanto a encontrada no continente Africano (Hemelaar et al. 2011), outros subtipos além do B também têm valor epidemiológico, tais como os subtipos F, C e recombinantes BF e BC que circulam na América do Sul, especialmente no Brasil, na Argentina e no Uruguai (Dilernia et al. 2007; Quarleri et al. 2004; Thomson et al. 2000) (figura 5). Dentre os CRFs caracterizados nesta região estão o CRF12_BF na Argentina, Bolívia, Chile, Paraguai e Uruguai (Aguayo et al. 2008; Carr et al. 2001; Ríos et al. 2007), CRF17 na Argentina e Paraguai (Aguayo et al. 2008; Carr et al. 2001), CRF28_BF, CRF29_BF, CRF39_BF, CRF40_BF, CRF46_BF, CRF70_BF, CRF71_BF e CRF72_BF no Brasil (Guimarães et al. 2008; Pessôa et al. 2014a; Pessôa et al. 2014b; De Sá Filho et al. 2006; Sanabani et al. 2010), CRF38 no Uruguai (Ruchansky et al. 2009), CRF44 no Chile (Delgado et al. 2010) e o CRF31_BC na região sul do Brasil (Santos et al. 2006).

Casos isolados de outros subtipos do grupo M também foram descritos na América do Sul, tais como os subtipos A (Caride et al. 2000), D (Morgado et al. 1998), G (Liegler et al. 2014) H (Yabar et al. 2012), CRF02_AG (Carrion et al. 2003; Delatorre et al. 2012), CRF06_cpx (Castillo et al. 2009; Dilernia et al. 2007; Fernández et al. 2015) e CRF45_cpx (Pessôa et al. 2015), porém muitos destes vírus foram introduzidos provavelmente por imigrantes africanos e não estabeleceram um número significativo de infecções locais.

1.5.1 Diversidade genética do HIV-1 no Brasil

A epidemia brasileira é composta majoritariamente pelos subtipos B, F1 e C, além das formas recombinantes entre eles. O subtipo B apresenta uma prevalência de 70-90%, seguido pelo subsubtipo F1 (5-20%), subtipo C (1-10%). Porém, a distribuição dos subtipos do HIV-1 não é homogênea por todo o país, como por exemplo os estados da região Sul do país, onde o subtipo C e recombinantes BC alcançam altas prevalências (25-66%) (Alcântara et al. 2013; Cavalcanti et al. 2007; Couto-Fernandez et al. 2005; Gaspareto et al. 2012; Gräf et al. 2011; López-Lopes et al. 2013; Machado et al. 2009; Silveira et al. 2012; Soares et al. 2005; Soares et al. 2003).

Com relação à distribuição dos CRFs pelas diferentes regiões do país, nota-se uma grande variedade de formas recombinantes na região sudeste, especificamente o CRF28_BF, CRF29_BF, CRF39_BF, CRF40_BF, CRF46_BF e CRF72_BF (Guimarães et al. 2008; Pessôa et al. 2014a; De Sá Filho et al. 2006; Sanabani et al. 2010; Santos et al. 2006). Na região Sul, foi identificado o CRF31_BC, que atinge altas prevalências (25-35%) e provavelmente surgiu por conta da co-circulação dos subtipos B e C (Almeida et al. 2012; Gräf e Pinto 2013; Santos et al. 2006). E mais recentemente os recombinantes CRF70_BF e CRF71_BF foram identificados na região Nordeste (Pessôa et al. 2014b).

Embora a diversidade molecular do HIV-1 no Brasil tenha aparentemente se mantido estável nos últimos anos, algumas mudanças tem sido notadas recentemente, como o avanço do subtipo C do Sul do país em direção aos estados mais ao norte, principalmente aqueles situados na região Centro-Oeste, onde atingem as maiores prevalências depois dos estados do Sul, como por exemplo nos estados de Mato Grosso (5%) (Ferreira et al. 2011), Mato Grosso do Sul (10%) (da Silveira et al. 2012) e Goiás (11%) (Alcântara et al. 2013).

Além da disseminação do subtipo C em direção ao norte do Brasil, também tem sido notado um aumento na incidência de clados do HIV normalmente encontrados na epidemia africana em diferentes estados, como por exemplo os subtipos A e D e os recombinantes CRF02_AG e CRF45_cpx (Alencar et al. 2013a; Caride et al. 2000; Delatorre et al. 2012; Eyer-Silva e Morgado 2007; Morgado et al. 1998; Pessôa et al. 2015; Pimentel et al. 2013)

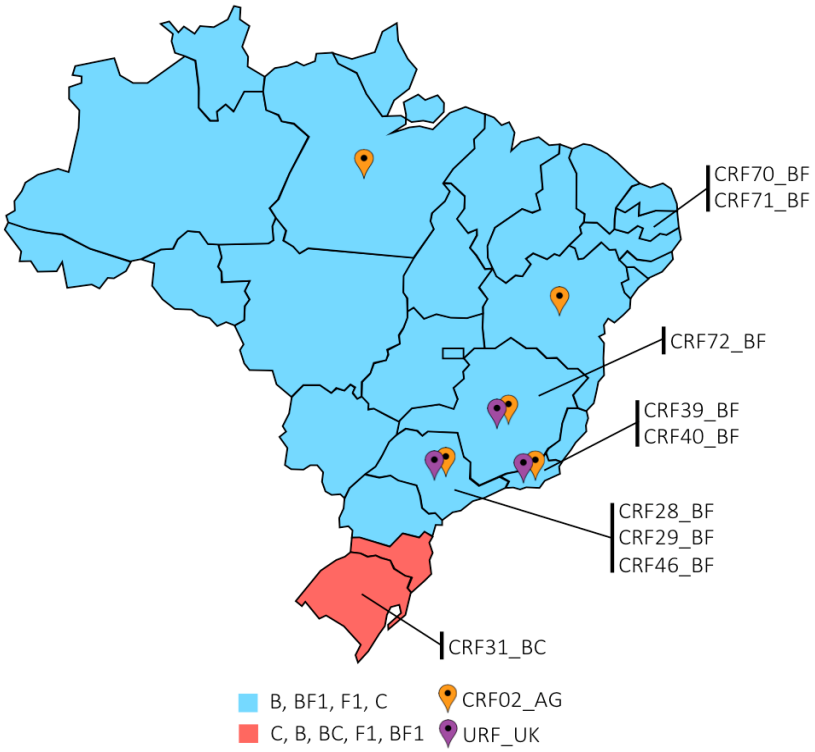


Figura 7 – Mapa da distribuição dos clados do HIV no Brasil. Cada cor representa as diferentes prevalências dos subtipos e formas recombinantes presentes em cada localidade. A ordem dos subtipos na legenda está organizada da maior para a menor prevalência. As CRFs brasileiras estão assinaladas de acordo com o estado onde ocorreu a descrição. Os estados onde o CRF02_AG e/ou URF_UK foram encontrados estão marcados com os pinos de acordo com a legenda. Mapa adaptado a partir dos dados de Bello et al. 2011; Gräf e Pinto 2013; Morgado et al. 2002.

O clado CRF02_AG já foi descrito nos estados do Rio de Janeiro (Couto-Fernandez et al. 2005; Delatorre et al. 2012; Eyer-Silva e Morgado 2007; Pires et al. 2004; Velasco-de-Castro et al. 2014), São Paulo (Ferreira et al. 2013; Sanabani et al. 2011), Pará (Machado et al. 2009) e Bahia (Inocencio et al. 2009), com prevalências na ordem de 0.2-1.9%, indicando a dispersão mesmo que limitada deste CRF no país. Além disso, formas recombinantes únicas carreando o subtipo K, raríssimo em sua forma pura, já foram descritos em baixas prevalências (<2%) nos estados do Rio de Janeiro (Brindeiro et al. 2002; Pilotto et al. 2013; Velasco-de-Castro et al. 2014), São Paulo (Alencar et al. 2013a; Pessôa et al. 2015) e Minas

Gerais (Ferreira et al. 2010). Estes estudos foram realizados com amostras de conveniência, e talvez as prevalências observadas não representem a real circulação destes recombinantes no país (Figura 7).

A maior parte do subtipo B que circula no Brasil pertence ao clado “pandêmico” (>98%) (Cabello et al. 2015). Esta linhagem do subtipo B provavelmente migrou da República Democrática do Congo para o Haiti no fim da década de 1960, e após um pequeno período de expansão local, se moveu para os EUA de onde se disseminou para todo o mundo, principalmente Europa Ocidental e América Latina, inclusive para o Brasil (Gilbert et al. 2007). No Brasil o clado B “pandêmico” foi introduzido diversas vezes, até se tornar o predominante (Bello et al. 2007). Dentro da linhagem do subtipo B circulante no Brasil foram encontradas diferenças genéticas e antigênicas, com a presença de um variante do subtipo B denominada BBr que se caracteriza por apresentar um motivo GWGR no topo da alça V3 da glicoproteína do envelope gp120, em vez do motivo GPGR (Morgado et al. 1994). As frequências desta variante atualmente correspondem a 20-60% do subtipo B encontrado no Brasil, entretanto a monofilia desta variante ainda é alvo de controvérsia (Bello et al. 2007; Covas et al. 1998; Diaz et al. 2008; Leal e Villanova 2010).

Análises filogenéticas estimam a maioria das cepas do subsubtipo F1 e subtipo C que circulam no Brasil são provenientes de eventos fundadores envolvendo linhagens circulantes na África (Aulicino et al. 2007; Bello et al. 2012; Bello et al. 2008; Bello et al. 2007; Fontella et al. 2008; Véras et al. 2011). O subsubtipo F1 brasileiro provavelmente se originou a partir da introdução de uma linhagem proveniente da África Central, possivelmente a República Democrática do Congo, porém a pequena quantidade de sequências de outros países desta região dificulta a obtenção de uma estimativa precisa. A data provável da introdução do subsubtipo F1 no Brasil foi estimada entre o fim da década de 1970 e início da década de 1980 (Aulicino et al. 2007; Bello et al. 2012; Bello et al. 2007). Atualmente, grande parte do subsubtipo F1 que circula no país está presente na forma de fragmentos que compõem diversas formas recombinantes BF1 e poucas cepas com o genoma composto completamente por este subsubtipo foram descritas (Bello et al. 2012; Bello et al. 2011; Guimarães et al. 2009).

Estudos filogeográficos indicam que a epidemia do subtipo C no Brasil se iniciou pela introdução de uma cepa ou cepas geneticamente relacionadas no estado do Paraná, de onde em seguida se disseminou rapidamente para os estados

vizinhos (Bello et al. 2008; Véras et al. 2011). Esta cepa do subtipo C fundadora da epidemia brasileira provavelmente se originou na África Oriental, embora o país e a data precisa deste evento variem entre os estudos. Em trabalhos conduzidos pelo nosso grupo, o Burundi foi apontado como a origem mais provável do clado brasileiro do subtipo C (Bello et al. 2008; Delatorre et al. 2013), enquanto outros autores indicam a Etiópia ou Quênia (Fontella et al. 2008; Véras et al. 2011). As estimativas iniciais da escala temporal do evento fundador deste clado propõem que o mesmo ocorreu por volta do início da década de 1980 (Bello et al. 2008), enquanto outros indicam uma introdução mais antiga entre as décadas de 1960 e 1970 (Véras et al. 2011). Em um trabalho mais recente conduzido por nosso grupo e que também compõe esta tese, utilizando amostras obtidas de diferentes regiões do Brasil, a data de introdução foi estimada em meados da década de 1970, coincidindo a origem de outros subclados do subtipo C em países da África Oriental, como Etiópia, Quênia, Tanzânia e Uganda (Delatorre et al. 2013).

Embora exista um número razoável de trabalhos abordando a reconstrução espaço-temporal da origem dos principais subtipos do HIV-1 que compõem a epidemia brasileira, existe uma carência de estudos sobre a origem dos outros subtipos também presentes de forma minoritária na epidemia. Estas linhagens menos prevalentes representam o resultado do processo dinâmico da distribuição global dos diferentes subtipos do HIV, que envolvem diversos fatores, como efeitos fundadores, crescimento populacional, migrações, entre muitos outros. A coleta de informações precisas e atualizadas sobre a distribuição, origem e disseminação dos diferentes subtipos do HIV-1 podem auxiliar nos esforços de prevenção e tratamento, com ações direcionadas para rotas de transmissão e/ou grupos de risco específicos de cada país ou região geográfica, além de revelar a existência de redes de transmissão desconhecidas.

1.6 Resistência do HIV aos antirretrovirais

O acúmulo de mutações causado pela baixa fidelidade da enzima RT como também a elevada frequência de replicação geram uma população viral altamente heterogênea, composta por um conjunto de variantes virais distintas, porém relacionadas entre si, conhecidas como *quasispecies* (Domingo et al. 2006). A alta diversidade genotípica existente nesta população viral permite que, ao ocorrer uma mudança nas pressões seletivas existentes no hospedeiro, como as

desempenhadas pela resposta imune e/ou TARV, seja possível a emergência de uma variante mais bem adaptada às novas condições, que pode acabar suplantando a população selvagem original.

Algumas das mutações presentes podem alterar o fenótipo do vírus, afetando seu *fitness* (adaptabilidade), ou seja, sua capacidade de transmissão, infecção, replicação, escape do sistema imunológico e resistência aos tratamentos antirretrovirais utilizados. Desta forma, a emergência de variantes virais carreando mutações de resistência durante o tratamento é um fenômeno constante (Martinez-Picado e Martínez 2008). Mesmo sob terapia, a replicação residual do HIV pode ocorrer, permitindo a emergência de variantes resistentes (Sahu 2015). A replicação pode permanecer em compartimentos onde a concentração de inibidores fica abaixo dos valores ideais (Pierson et al. 2000), ou em regiões onde múltiplas infecções derivadas da transmissão célula a célula induzem a perda de sensibilidade ao inibidor, sem o requerimento de mutações de resistência às drogas (Sigal et al. 2011).

Mutações que conferem redução da susceptibilidade viral já foram descritas para todas as drogas antirretrovirais utilizadas na rotina clínica. Tipicamente, a resistência aos inibidores não-nucleosídicos de transcriptase reversa está associada a mutações pontuais, enquanto a resistência aos inibidores de protease na maioria dos casos envolve mutações múltiplas (Wensing et al. 2014). Desta forma, é recomendado realizar testes genotípicos para detectar mutações de resistência antes de se iniciar a TARV e no monitoramento da falha virológica, definida como a não-obtenção ou não-manutenção de carga viral indetectável após seis meses do início ou modificação do tratamento antirretroviral, ou por detecção da carga viral nos indivíduos que a mantinham indetectável na vigência do tratamento (Brasil 2012b).

As mutações relacionadas à resistência aos antirretrovirais são classificadas como mutações principais e acessórias, porém, não existem critérios específicos consolidados para a classificação das mutações nestas duas categorias. Usualmente, as mutações principais são aquelas que sozinhas conseguem reduzir a susceptibilidade do vírus a uma ou mais drogas. Adicionalmente, mutações altamente frequentes em indivíduos em falha virológica, localizadas em sítios estruturalmente importantes ou em regiões conservadas das proteínas também são regularmente consideradas principais (Arts e Hazuda 2012).

Em contraste, mutações acessórias são aquelas com pouca ou nenhuma capacidade para redução da susceptibilidade viral aos ARV *per se*, sendo, porém capazes na presença de mutações principais. Além disso, as mutações acessórias também podem aumentar o *fitness* de vírus contendo mutações principais, contribuindo para a manutenção da população viral resistente (Chang e Torbett 2011; Wensing et al. 2014).

As mutações de resistência aos antirretrovirais podem estar presentes na população viral do indivíduo infectado antes do uso da medicação (resistência transmitida) ou serem selecionadas após o início da terapia (resistência adquirida) (Coffin 1995). Em pacientes virgens de tratamento muitas das mutações acessórias não chegam a conferir resistência, mas facilitam fixação das mutações principais relacionadas, uma vez que ao surgir, o vírus carreando ambas possui um fitness superior ao que teria somente com a mutação principal (Arts e Hazuda 2012).

Até o momento, a terapia antirretroviral encontra-se dividida em seis classes, classificadas de acordo com quais etapas do ciclo replicativo do HIV são bloqueadas (Figura 3): antagonistas do correceptor CCR5 (adsorção e entrada); inibidores de fusão (fusão de membranas); inibidores da transcriptase reversa análogos de nucleosídeos e nucleotídeos, e inibidores não-nucleosídeos da transcriptase reversa (transcrição reversa do RNA genômico viral); inibidores de integrase (integração do cDNA no genoma da célula hospedeira); inibidores da protease (maturação da partícula viral) (Arts e Hazuda 2012; Thompson et al. 2014).

A classe com o maior número de antirretrovirais aprovados são os inibidores da transcrição reversa, que atuam inibindo a ação da transcriptase reversa. Esta enzima é um heterodímero com ação DNA polimerase RNA dependente, ribonuclease H (RNase H) e DNA polimerase DNA dependente. Os medicamentos que atuam nesta enzima podem ser divididos em inibidores nucleosídicos (INTRs) e inibidores não nucleosídicos (INNTRs). Os INNTRs (efavirenz, nevirapina, delavirdina, etravirina e mais recentemente rilpivirina) se ligam a um bolso hidrofóbico situado no sítio alostérico, e com isso causam mudanças conformacionais que impedem a síntese de DNA (Esnouf et al. 1995).

Os INTRs atuam como terminadores de cadeia, impedindo a continuidade da replicação uma vez que não possuem o grupo hidroxila no carbono 3’ da ribose, responsável pela formação da ligação fosfodiéster (Cheng et al. 1987; Richman 2001). Portanto esta classe impossibilita o ataque nucleofílico do fosfato 5' do nucleotídeo que será adicionado à hidroxila 3', e consequentemente a extensão da

cadeia nascente (Esnouf et al. 1995). Dentro da classe dos INTRs, diferenciam-se os análogos de nucleosídeos, que competem com os dNTPs celulares e requerem três passos de fosforilação para serem convertidos em metabólitos ativos (i.e., nucleotídeos) e os análogos de nucleotídeos, que já possuem um grupo fosforilado e requerem apenas dois passos para serem ativados (Menéndez-Arias 2013). Desta forma, todos os INTRs são considerados pró-fármacos, necessitando da ativação através de uma biotransformação, que neste caso corresponde a fosforilação intracelular, para se tornarem drogas farmacologicamente ativas.

Atualmente, sete inibidores análogos de nucleosídeos foram aprovados (zidovudina, didanosina, zalcitabina, estavudina, lamivudina, abacavir e emtricitabina), enquanto há apenas um análogo de nucleotídeo (tenofovir) (Thompson et al. 2014). Existem dois mecanismos principais que conferem resistência aos INTRs. O primeiro corresponde a um mecanismo discriminatório através de inibição alostérica, no qual a TR se torna capaz de evitar a incorporação do análogo de nucleotídeo, enquanto retém a habilidade de incorporar o dNTP natural. Neste caso, mutações pontuais reduzem a afinidade da enzima por certos INTRs. O outro mecanismo corresponde ao aprimoramento da capacidade de remoção hidrolítica do análogo incorporado na terminação 3` da cadeia de cDNA, processo conhecido como pirofosforólise. As mutações associadas aos análogos de timina (TAMs) caracterizam-se por promover este aumento de capacidade fosforolítica. As TAMs estão envolvidas na resistência para todos os INTRs, exceto a lamivudina, entretanto o grau de resistência depende do INRT considerado e a quantidade de TAMs presentes (Marcelin 2006).

Os inibidores da protease exibem sua ação farmacológica durante a maturação dos vírions do HIV, que ocorre na fase tardia no ciclo de replicação (Kohl et al. 1988). Através da ligação com proteases, os IPs induzem o bloqueio das atividades proteolíticas da enzima, o que resulta na inabilidade de formação de vírions maduros infecciosos (Ali et al. 2010). Atualmente, nove IPs estão aprovados para uso: saquinavir, indinavir, nelfinavir, lopinavir, ritonavir, fosamprenavir, tipranavir, atazanavir, darunavir (Wensing et al. 2014). De maneira geral, IPs se caracterizam por possuírem uma alta barreira genética, sendo necessário o acúmulo de múltiplas mutações de resistência para que ocorra uma diminuição sensível da susceptibilidade viral a esta classe (Ali et al. 2010). Entretanto, o tratamento contínuo e a longo prazo com esta classe de ARV acaba sendo frequentemente associado ao aparecimento de efeitos colaterais. Dentre aqueles mais comuns estão

as síndromes metabólicas, como a dislipidemia, a resistência à insulina, e lipodistrofia/lipoatrofia, bem como doenças cardiovasculares e cerebrovasculares (Lv et al. 2015). O aparecimento destes efeitos colaterais afeta gravemente a adesão dos pacientes ao tratamento, um dos maiores obstáculos para o controle da progressão para aids.

Os outros alvos terapêuticos incluem a integração do DNA viral no genoma da célula hospedeira, através da ação de inibidores da integrase. Os inibidores raltegravir, elvitegravir e dolutegravir são capazes de se ligar aos dois íons $Mg2^{+}$ presentes no sítio ativo e interromper a transferência de nucleotídeos entre o complexo de integração e o cromossomo impedindo a formação do provírus (Hare et al. 2010). Os inibidores de adsorção e fusão impedem o processo de entrada do HIV na célula alvo, o que depende da ligação da proteína gp120 com o receptor CD4 e a posterior fusão das membranas viral e celular mediada pela proteína gp41 (Dorr et al. 2005). A inibição da etapa de ligação é realizada pelo fármaco maraviroque, capaz de se ligar à gp120 (Westby e van der Ryst 2010). Já a enfuvirtida é um pequeno peptídeo injetável que impede a mudança conformacional da gp41, inibindo a fusão das membranas (Matthews et al. 2004). O uso destas classes é direcionado para a combinação com outros antirretrovirais compondo o resgate terapêutico de pacientes com vírus multirresistentes (Brasil 2012b). A multirresistência é usualmente definida como a presença de resistência fenotípica ou genotípica a todas as três classes de antirretrovirais originais (IPs, INTRs e INNTRs), uma condição que diminui sensivelmente as opções terapêuticas disponíveis, o que está associado a um maior risco da progressão clínica e aumento na mortalidade (Imaz et al. 2011).

O Ministério da Saúde brasileiro oferece acesso gratuito e universal da terapia antirretroviral, composta por uma combinação de pelo menos três antirretrovirais e envolvendo ao menos duas classes de inibidores, inclusive para a profilaxia da transmissão vertical do HIV. Esta combinação proporciona a supressão máxima da replicação do HIV, auxiliando na redução da morbimortalidade e melhorando a qualidade de vida dos indivíduos, além de permitir a reconstituição do sistema imunológico, reduzindo a ocorrência de doenças oportunistas (Richman 2001).

O aumento da disponibilidade da terapia antirretroviral, é benéfica para os indivíduos vivendo com HIV/aids, mas também pode ter consequências negativas, como o favorecimento da emergência de variantes resistentes às drogas, devido às baixas barreiras genéticas de alguns medicamentos, problemas de adesão aos

regimes terapêuticos, entre outros fatores (Tang e Pillay 2004). Estas variantes resistentes poderiam então serem transmitidas para indivíduos recém-infectados, o que limitaria as opções terapêuticas disponíveis. Isto é ainda mais grave no caso gestantes, nas quais a terapia quimioprofilática tem o objetivo de reduzir rapidamente os níveis de carga viral para a prevenção da transmissão do HIV da mãe para o seu bebê.

Estudos transversais realizados no Brasil apontam para um nível moderado (5-15%) de resistência transmitida do HIV-1 (Brindeiro et al. 2003; Inocencio et al. 2009; Sprinz et al. 2009; Alencar et al. 2013), embora um cenário mais diversificado seja observado nos estudos realizados com mulheres grávidas, que mostram taxas de resistência transmitidas em níveis baixos [0% (Cardoso et al. 2010)], moderados [5.9% (Alcântara et al. 2012), 9.8% (de Lourdes Teixeira et al. 2014), 10.7% (Pilotto et al. 2013)] e altas [25% (da Costa et al. 2013)] de acordo com os critérios estabelecidos pela OMS. Se a prevalência de resistência transmitida aos antirretrovirais em mulheres infectadas pelo HIV entre 15-49 anos alcança níveis superiores à 15%, a OMS recomenda, como uma ação de saúde pública, o início de testes de genotipagem para resistência do HIV para todas as gestantes antes de se iniciar a profilaxia para prevenção da transmissão vertical (World Health Organization 2012), tornando o monitoramento contínuo dos níveis de vírus carreando mutações de resistência aos antirretrovirais imprescindível para que se obtenha o controle da replicação viral, otimizar o sucesso da profilaxia para prevenção da transmissão vertical, prevenir a emergência de mutações de resistência adicionais e prevenir a transmissão de vírus resistentes para outros indivíduos.

1.7 A região metropolitana do Rio de Janeiro como sentinela das mudanças nos perfis de resistência e diversidade do HIV

A região metropolitana do Rio de Janeiro engloba 21 municípios situados nas cercanias da cidade do Rio de Janeiro que se estendem por mais de oito milhões de km^2. Esta região concentra uma população estimada superior a 12 milhões de indivíduos, sendo a segunda maior área metropolitana do Brasil, terceira da América do Sul e 20ª maior do mundo (IBGE 2014). Dentro do estado do Rio de Janeiro, a região metropolitana é responsável por 90% das infecções, sendo os municípios do Rio de Janeiro, Nova Iguaçu, Niterói, Duque de Caxias e São Gonçalo os que apresentam os maiores números de casos de HIV/aids acumulados (Brasil 2012a).

No Brasil, a via heterossexual é a forma dominante de transmissão do HIV entre adolescentes e adultos, correspondendo a >75% de todos os casos de aids relatados nos últimos 10 anos (Brasil 2014). Desta forma, a população de gestantes seria um importante grupo, formado exclusivamente por indivíduos com vida sexual ativa, fornecendo uma boa representação da epidemia do HIV circulante na população em geral.

Estudos prévios conduzidos com esta população no Rio de Janeiro, já apontaram para níveis moderados de mutações transmitidas de resistência às drogas antirretrovirais (MTRD) próximos à 10% (de Lourdes Teixeira et al. 2014; Pilotto et al. 2013). Mesmo tendo encontrado níveis moderados de MTRDs, estes estudos recomendaram a realização da genotipagem do HIV-1 pré-tratamento como uma forma de auxiliar a prescrição da terapia antirretroviral apropriada durante a gravidez, com o intuito de suprimir a carga viral e prevenir a TV.

Com o aumento da oferta de antirretrovirais para os indivíduos infectados pelo HIV como uma nova medida implementada pelo Ministério da Saúde em 2013, o número de pessoas em tratamento tende a aumentar progressivamente. Desta forma, também aumentam a probabilidade de que ocorram mudanças nos níveis de MTRDs para patamares mais elevados, como consequência de falhas na supressão virológica dos regimes terapêuticos adotados, seja por conta da baixa barreira genética de algumas drogas e/ou baixa adesão à terapia (Tang e Pillay 2004). Maiores taxas de MTRDs impactam diretamente o sucesso da TARV, com reflexos negativos no controle da morbimortalidade dos indivíduos infectados como também da transmissão do HIV.

Além disso, já foi sugerido que a região metropolitana do Rio de Janeiro é formada por uma população com um alto nível de infecções importadas do HIV-1 (Velasco-de-Castro et al. 2014). Isto se deve provavelmente ao caráter turístico da região e seu importante papel econômico e cultural, atraindo um grande número de visitantes do mundo todo. Como consequência, se tem um aumento da circulação de clados importados do HIV, alterando o perfil da diversidade de subtipos encontrados na epidemia fluminense. Isto foi evidenciado em dois estudos conduzidos com gestantes desta região, onde foram encontrados pacientes infectados pelo subsubtipo A1 (de Lourdes Teixeira et al. 2014) e recombinantes formados por fragmentos do subtipo K (Pilotto et al. 2013).

Esta alta diversidade e a prevalência moderada de MTRDs do HIV-1 nos motivaram a conduzir um novo estudo para monitorar as mudanças nos perfis de

diversidade e resistência do HIV nesta população gestantes infectadas pelo HIV-1, e comparar com o estudo de Pilotto e colaboradores (2013), realizado com casuística similar no mesmo centro de atendimento (Hospital Geral de Nova Iguaçu), cinco anos antes que o presente estudo. O Hospital Geral de Nova Iguaçu (HGNI) é um hospital geral do Ministério da Saúde com 350 leitos e uma maternidade de alto risco, onde são realizados cerca de 3.000 partos/ano. Sua maternidade é o centro de referência para atendimento de gestantes portadoras de infecção pelo HIV-1 em toda a Baixada Fluminense, atendendo cerca de 120 gestantes nesta situação por ano.

A partir dos clados do HIV-1 raros na epidemia brasileira evidenciados nesta população, procuramos avaliar a história evolutiva dos mesmos em conjunto com outras amostras geneticamente relacionadas encontradas em outros estudos conduzidos no laboratório e/ou disponíveis em bancos de dados públicos de sequências para tentar traçar os padrões de disseminação e a data mais provável de introdução destas linhagens no Brasil.

2 OBJETIVOS

2.1 Objetivo Geral

Avaliação da prevalência da resistência transmitida aos ARVs e da história evolutiva de clados raros do HIV presentes em uma população de gestantes do Rio de Janeiro submetidas a terapia antirretroviral.

2.2 Objetivos Específicos

- Avaliar a diversidade viral e prevalência de mutações de resistência transmitidas às drogas antirretrovirais entre gestantes infectadas pelo HIV virgens de terapia do Rio de Janeiro entre os anos de 2012 e 2015.

- Determinar com mais precisão as origens temporal e geográfica das linhagens do subtipo C do HIV-1 que circulam no Rio de Janeiro e outros estados brasileiros e a sua dinâmica de disseminação no país.

- Determinar a provável origem geográfica e temporal de linhagens do CRF02_AG introduzidas no Rio de Janeiro.

- Caracterizar os genomas e reconstruir a história evolutiva de recombinantes do HIV-1 relacionados ao CRF45_cpx identificados no Rio de Janeiro e outros estados da região Sudeste.

3 RESULTADOS

3.1 Artigo 1

Título: *High HIV-1 diversity and prevalence of transmitted drug resistance mutations among antiretroviral-naïve HIV-infected pregnant women in Rio de Janeiro, Brazil*

Autores: Edson Delatorre, Carlos Silva-de-Jesus, José Carlos Couto-Fernandez, Jose H. Pilotto e Mariza G. Morgado

Resumo:

Mutações de resistência aos antirretrovirais (ARV) durante a infecção pelo HIV-1 podem reduzir a eficácia da terapia profilática para a prevenção da transmissão vertical (PTV) e tratamentos futuros. Este estudo avaliou a diversidade viral e a prevalência de mutações transmitidas de resistência às drogas (MTRD) entre gestantes infectadas pelo HIV-1 virgens de terapia do Rio de Janeiro - Brasil entre os anos de 2012 e 2015. As regiões da protease (PR) e transcriptase reversa (TR) do gene *pol* do HIV-1 de 87 amostras de plasma foram sequenciadas utilizando um método de genotipagem *in house* validado. Os subtipos do HIV-1 e MTRD foram determinadas por análises filogenéticas e pela ferramenta *Calibrated Population Resistance*, respectivamente. O subtipo B do HIV-1 foi identificado em 69% das amostras, seguido pelo sub-subtipo F1 (16,1%) e C (4,6%). As formas recombinantes representaram 10,3% das amostras. A prevalência global para qualquer MTRD foi 17,2%, das quais 5,7% para inibidores nucleosídicos da TR (INTRs), 5,7% para inibidores não-nucleosídicos da TR e 8% para inibidores da protease (IPs). Um aumento na prevalência de subtipos não-B e uma mudança na tendência das MTRD com maiores taxas (de INTRs para IPs) foi observada quando comparada a um estudo prévio conduzido no mesmo local. O nível de MTRD encontrado nesta população pode afetar o desfecho virológico dos regimes de ARV padrão para PTV, reforçando a importância do seu contínuo monitoramento.

High HIV-1 diversity and prevalence of transmitted drug resistance mutations among antiretroviral-naïve HIV-infected pregnant women from Rio de Janeiro, Brazil

Authors: Edson Delatorre[1*], Carlos Silva-de-Jesus[1], José Carlos Couto-Fernandez[1], Jose H. Pilotto[1,2] and Mariza G. Morgado[1]

Running title: TDRM and HIV-1 diversity in pregnant women

Author's affiliation:

[1] Laboratório de AIDS & Imunologia Molecular. Instituto Oswaldo Cruz – FIOCRUZ. Rio de Janeiro, Brazil. [2] Hospital Geral de Nova Iguaçu, Rio de Janeiro, Brazil.

*Corresponding author:
E-mail: edsonod@ioc.fiocruz.br/delatorre.ioc@gmail.com
Address: Lab. de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz - FIOCRUZ. Av. Brasil 4365, 21040-360 Rio de Janeiro, RJ, Brazil.
Phone: +55 21 3865 8154. Fax: +55 21 3865 8173.

Abstract: 211 words

Main text: 3569 words

ABSTRACT

Background

Antiretroviral (ARV) resistance mutations in HIV-1 infection may reduce the efficacy of prophylactic therapy to prevent mother-to-child transmission (PMTCT) and future treatments. This study evaluated the viral diversity and the prevalence of transmitted drug resistance mutations (TDRM) among ARV-naïve HIV-1-infected pregnant women from Rio de Janeiro – Brazil between 2012 and 2015

Methods

The HIV-1 protease and reverse transcriptase (PR/RT) regions of the *pol* gene from 87 plasma samples were sequenced using a validated in house genotyping method. TDRM and HIV-1 subtypes were determined by the Calibrated Population Resistance tool and phylogenetic analyses, respectively.

Results

The HIV-1 subtype B was identified in 67.8% of the samples, followed by subtypes F1 (17.2%) and C (4.6%). The recombinant forms accounted for 10.3% of the samples. The overall prevalence of any TDRM was 17.2%, of which 5.7% were for nucleoside RT inhibitors (NRTIs), 5.7% for nonnucleoside RT inhibitors and 8% for protease inhibitors (PIs). An increase in the prevalence of non-B subtypes and a change in the trend of TDRM with higher rates (from NRTIs to PIs) were observed when compared to a previous study at the same locale.

Conclusions

The TDRM level found in this population may affect the virological outcome of the standard PMTCT ARV-regimens, reinforcing the importance of its continuous monitoring.

KEYWORDS: HIV-1, pregnancy, TDRM, subtypes, PMTCT, Brazil

BACKGROUND

The human immunodeficiency virus type 1 (HIV-1) is the causative agent of acquired immune deficiency syndrome (AIDS), and currently, 35 years after the first description of this disease, it is estimated that approximately 36 (34.3 - 41.4) million people are living with HIV worldwide [1]. According to estimates of the Brazilian Ministry of Health, approximately 734.000 people are HIV-infected in Brazil, which corresponds to a prevalence of 0.4% [2]. However, in populations at higher risk, such as men who have sex with men and sex workers, prevalence rates >5% are observed [3, 4]. Regarding pregnant woman, it is expected that 12.000 cases of HIV infection occur annually in Brazil, however, only 3700 cases were reported to the public health system in 2014, of which 8% corresponds to Rio de Janeiro state [2]. The HIV-1 epidemic of Rio de Janeiro is highly concentrated in its metropolitan region, which is responsible by nearly 90% of the HIV infections reported in the state [2].

Since 1996, universal free access to highly active antiretroviral therapy (HAART) has been available in Brazil for HIV/AIDS treatment, including its use for the prevention of HIV mother-to-child transmission (PMTCT). Over 350.000 HIV-1 infected individuals are currently under treatment through the Public Health System, and since 2013, the Brazilian government recommends the immediate start of antiretroviral (ARV) therapy for all individuals living with HIV/AIDS, regardless the CD4 T-cells counts [5]. The increase in the use of ARV drugs has had a strong impact with reduction of morbidity and mortality of HIV-1 infected individuals [6]. However, it may also have favored the emergence of HIV-1 drug-resistant variants, due to the lower genetic barrier of some drugs, poor adherence to treatment regimens, among other factors [7]. These drug-resistant HIV-1 variants could therefore limit the therapeutic options available for newly infected people, especially for pregnant women, in which the prophylactic therapy intends to achieve a rapid reduction of viral load levels to prevent vertical transmission of HIV. So far, national and regional cross-sectional surveys conducted in Brazil have pointed out a moderate level (5-15%) of HIV-1 transmitted drug

resistance (TDR) [8–11]. However, a more diverse scenario was observed in studies conducted with pregnant women, that showed low (0% [12]), moderate (5.9% [13], 9.8% [14], 10.7% [15]) and high TDR prevalences (25% [16]). Nevertheless, these results should be interpreted with caution, since most studies had small sampling sizes, due to the difficulty in obtaining the number of samples representative of the population of women living with HIV/AIDS at the locations analyzed. If the estimated prevalence of TDR in HIV-infected women between 15 and 49 years reaches >15%, the World Health Organization (WHO) recommends, as a public health action, to start conducting HIV TDR testing of all HIV infected pregnant women before starting PMTCT prophylaxis [17].

Another factor that could possibly influence the drug susceptibility of treatment-naïve subjects are the different HIV-1 subtypes. It has been reported that some non-B subtypes, notably the subtypes C, F, G and CRF02_AG, display different susceptibilities to specific protease inhibitors (PI) [18, 19]. In Brazil, the HIV-1 molecular epidemiology is dominated by subtypes B (70-90%), F1 (5-15%) and C (1-10%), as the recombinants between them [20–25]. The exceptions are the Southern states, where subtype C and CRF31_BC predominate [26, 27]. However, the continuous finding of other HIV clades, such as subtypes A, D and CRF02_AG [Morgado et al. 1998; Caride et al. 2000; Delatorre et al. 2012; Alencar, et al. 2013; Pimentel et al. 2013], as well as the northward spread of subtype C [33], indicate that the HIV-1 molecular diversity in Brazil is dynamic and in constant evolution.

Here, we evaluated the viral diversity and the prevalence of TDR among antiretroviral-naïve HIV-1-infected pregnant women followed at the Hospital Geral de Nova Iguaçu in Rio de Janeiro – Brazil, between 2012 and 2015, and compared the results with a similar previous study conducted by our group at the same locale from 2005 to 2008 [15]. Moreover, we also present in detail the protease and reverse transcriptase in house sequencing protocol developed by our group, which is Virology Quality Assessment Program (VQA)-validated and used for the genotyping of HIV-1 positive individuals in Brazil.

RESULTS

Characteristics of the study population

All patients analyzed in this study lived in the metropolitan region of the Rio de Janeiro state - Brazil at the time of analysis and have a median age of 25 years (15–36 years range). The median pre-HAART CD4+ T-cell count was 413 cells/mm3 (IQR, 316–589), with a homogeneous distribution among the cell count ranges (~33%) when considering the whole study period. The median HIV-1 RNA plasma viral load for our study population was 4.04 $\log_{10}$ copies/ml (IQR, 3.35–4.53), which was mostly distributed (75.4%) between 3 and 5 $\log_{10}$ copies/ml. It is noteworthy that approximately 14% of the samples amplified and sequenced using our protocol had viral loads below 3 $\log_{10}$ copies/ml. Table 1 summarizes the clinical characteristics of the patients.

HIV-1 subtype diversity in the *pol* gene

The HIV-1 subtype B was the most prevalent clade in our sample (67.8%, 59/87), followed by subsubtype F1 (17.2%, 15/87), and subtype C (4.6%, 4/87). The intersubtype recombinant forms (10.3%, 9/87) were divided into circulating recombinant forms (CRF; 5.7%, 5/87) and unique recombinant forms (URF; 4.6%, 4/87) (Fig. 1 A and B). It is worth mentioning that together, the intersubtype recombinant forms represented the third more prevalent HIV-1 group of the pregnant women tested in this study. Most CRFs were composed of subtypes B and F1, as follows: CRF12_BF (2.3%, 2/87), CRF28/29_BF (1.1%, 1/87) and CRF39_BF (1.1%, 1/87). One CRF02_AG (1.1%, 1/87) was also found, as well as two URF_BF (2.3%, 2/87), one URF_BC (1.1%, 1/87) and one URF_CF (1.1%, 1/87). The classification of one sequence as CRF28/29_BF was indicated because these two CRFs share a very similar pattern of recombination in the protease/reverse transcriptase region (PR/RT), forming a unique clade in the maximum likelihood (ML) phylogenetic analysis (Fig. 1B), which makes the genetic distinction between them difficult.

Prevalence and patterns of transmitted drug resistance mutations in PR/RT

Resistance analysis among the 87 samples with both PR and RT sequences showed that at least one drug resistance mutation was observed in 15 samples, leading to an overall prevalence of TDR of 17.2%. TDRM for protease inhibitors (PIs) were detected in 8.0% (7/87) of the patients. Of these, five had the M46L mutation (5.7%, 5/87) and one had the M46I mutation (1.1%, 1/87). The other mutations found were D30N and N88D, both harboured by the same pregnant woman. Additionally, the most frequent minor drug resistance mutations observed were: M36I/L (49.4%, 43/87), L10I/V (18.4%, 16/87) and K20M/R (11.5%, 10/87). TDRM for nucleoside reverse transcriptase inhibitors (NRTIs) were detected in 5.7% (5/87) of the women in the study. The mutation M41L was found in one patient (1.1%, 1/87), while the D67N mutation was found in three (3.4%, 3/87). Furthermore, the mutations T69D and K219Q were detected in one subject each (1.1%, 1/87) that both also exhibited the D67N mutation. TDRM to non-nucleoside reverse transcriptase inhibitors (NNRTIs) were detected in 5.7% (5/87) of the patients analysed. A total of 2.3% (2/87) of women had the K103N mutation. The same frequency was observed for the G190A mutation (2.3%, 2/87). The mutation K101E was found in one patient (1.1%, 1/87) that also had the G190A mutation. The following minor drug resistance mutations associated to NNRTIs were observed: V90I (6.9%, 6/87), E138A/G (3.4%, 3/87), V106I (2.3%, 2/87), V179D (2.3%, 2/87), A98G (1.1%, 1/87). One sequence carrying both NRTI and NNRTI mutations was detected, however, no sequence with triple class resistance was found. Table 2 summarizes PI, NRTI, and NNRTI transmitted drug resistance mutations according to the subtype. Almost all TDRM were found in women infected with HIV-1 subtype B viruses (73%, 11/15). Three patients with TDRM were infected with subsubtype F1 and one with CRF28/29_BF.

DISCUSSION

This TDR and HIV-1 diversity survey was performed in pregnant women from an impoverished and populous area located in the metropolitan region of the Rio de Janeiro state in Brazil. The list of mutations proposed by Bennett et al. [2009] was used as marker of HIV TDRM. The 17.2% overall prevalence of TDR found in this study places the region analyzed in the high level stratum of TDR prevalence (>15%) according to the WHO criteria [17]. Reaching this level, WHO has suggested that it becomes cost effective to perform genotypic resistance testing in individuals with newly diagnosed HIV infection, prior to the beginning of the HAART.

Our findings are affected by the low number of patients sampled as well as some different criteria between the survey method used in this study and the WHO HIV drug resistance threshold survey method, such as limit of inclusion age (< 25 years), and the CD4 T-cell count limit (>500 cells/µl) [17]. However, some criteria were fulfilled, such as the exclusion of all patients who may have been previously exposed to ART, including PMTCT prophylaxis. These WHO criteria have been established to predict the likelihood of recent HIV infection.

In Brazil, national and regional cross-sectional surveys pointed out a moderate level of HIV-1 transmitted drug resistance (5-15%) [8-10]. However, when comparing specific regions of the country and/or population groups, TDR becomes relatively more frequent, such as in the general population of São Paulo (19.4%) [11], men who have sex with men from Santos (28%) [35] and newly HIV-1 diagnosed individuals in Rio de Janeiro (15%) [36]. The prevalence of TDR found in this study (17.2%) was higher than that found in other studies conducted with HIV-1-infected pregnant woman, depicting TDR rates of 0% and 5.9% in Goiás [12, 13], as well as 9.8% and 10.7% in Rio de Janeiro [14, 15]. However, the prevalence of TDR in our study was lower than that described by da Costa et al. (25%) in Goiás [16]. It is important to highlight that the work of Pilotto et al. [15] was conducted for

the same study population five years earlier (2005-2008) than the present study, and at the same location. The overall prevalence of any HIV-1 TDRM found in this study was higher (17.2%) than the previous estimate (10.7%), although the difference was not statistically significant (P = 0.1271), indicating that the TDRM prevalences seem not to have changed between these time periods despite the adoption of new practices regarding ART use. However, the influence of these actions may only be suitable for observation in the future.

Mutations to NRTI were the most prevalent (5.6%) in pregnant women sampled in the 2005-2008 period, which were associated with the use of NRTI monotherapy (zidovudine and stavudine) and NRTI dual therapy before the HAART era in Brazil. In this study, the TDRM for IPs were the most prevalent, with a rate of 8.0%, much higher than the previous observations (3.0%). This may reflect the widespread adoption of HAART in Brazil, in which a regimen composed by two NRTI and one boosted PI is the alternative choice of the main regimen composed by two NRTI and one NNRTI [5]. This high TDRM rate for PIs may be the effect of poor patient adherence to this alternative therapeutic regimen, frequently caused by high pill burden and low tolerance to these ARV regimens [37]. Boosted PIs have a short half-life and for this reason, regularly missed doses may cause more problems for ritonavir-boosted PIs than for NNRTI-based regimens [38]. Therefore, when patients are undergoing partial PI treatment interruption, wild-type virus may be suppressed by the backbone therapy, which prevents the outgrowth of wild-type virus [39]. This allows the replication of the HIV variants carrying PI resistance mutations, which might increase the likelihood of transmission of the resistant virus.

In this study, the PI resistance mutations M46I/L, N88D and D30N were found, which are responsible for low, intermediate, and high levels of resistance to nelfinavir (NFV), respectively [40, 41]. Therapeutic regimens containing NFV were adopted in Brazil since 1998, becoming one of the major PIs used in the country, until its replacement by newly developed PIs, such as lopinavir/r and atazanavir [42]. NFV initially used in Brazil had a high

pill burden (five pills twice a day), a characteristic that is associated with both worse adherence and virological suppression [37]. Among the RT gene resistance mutations, thymidine analogue-associated mutations (TAM)-related mutations D67N and K219Q were found. These mutations are associated to low-level resistance to zidovudine (AZT) and stavudine (d4T) [40, 41]. Two patients exhibited the mutation K103N that causes high-level resistance to efavirenz (EFV) and nevirapine (NVP). The others mutations found (K101E, Y188L and G190A) are responsible of variable levels of resistance to all NNRTIs, but share a high level resistance to NVP [40, 41]. Taken together, these resistance mutations found can abrogate the first line treatment regimen and PMTCT prophylaxis commonly employed in Brazil, which recommends the use of two NRTI (AZT and lamivudine) plus a PI, or alternatively a NNRTI (specifically NVP) [43]. These findings enforce the necessity to promote routine and universal genotype ARV-resistance susceptibility testing to all Brazilian HIV-infected pregnant women for PMTCT purposes.

Regarding HIV-1 diversity, despite the predominance of subtype B, subsubtype F1 was the second most frequent, followed by subtype C and recombinant forms. These findings are in accordance with previous descriptions of the HIV diversity found in the Rio de Janeiro state [29, 36] and with the previous study of pregnant women recruited from 2005 to 2008 at the same clinical center [15]. However, when comparing the differences among the prevalence of subtypes, a statistically significant increase in the prevalence of non-B viruses between the two time periods (P = 0.0330) was observed. This trend was expected, since it has already been observed that the molecular epidemiology of HIV-1 in Rio de Janeiro is changing over the years towards a marked increase in the circulation of non-B strains, mainly intersubtype recombinants [36]. One CRF02_AG sample was among the recombinant forms described in this study. CRF02_AG is the most prevalent recombinant form around the world and is mainly distributed in West Africa [44], however, it has been continuously isolated in Brazil [8, 20, 21, 32, 36]. In a previous study from our group, this CRF02_AG sample was

described as participating in an autochthonous transmission network, which has been disseminated locally in the Rio de Janeiro state over the last 30 years [45]. It is also noteworthy that two CRF12_BF samples were found in this study. This CRF is highly prevalent in Argentina and Uruguay, but rarely detected in Brazil [46]. These findings demonstrate the role of the Rio de Janeiro metropolitan area (a major touristic destination and an important commercial hub) as a hot spot for the introduction and dissemination of new viral clades in the Brazilian epidemic.

CONCLUSIONS

The results presented here indicate an increase in the HIV-1 molecular diversity and a probable change in the TDRM profile in Rio de Janeiro between the periods of 2005-2008 and 2012-2015. The HIV-1 TDRM found in this population of women could affect the virological outcome of the standard ARV regimens employed to prevent HIV vertical transmission. This highlights the importance of continuous monitoring of the HIV-1 genetic diversity and TDRM in Brazil for all HIV-infected pregnant women, intended to optimize the success of PMTCT prophylaxis, prevent the emergence of further resistance mutations, and prevent transmission of resistant viruses to other individuals.

METHODS

Study population

A total of 87 plasma samples were obtained from HIV-1-infected pregnant women followed at the Department of Sexually Transmitted Diseases of the Hospital Geral de Nova Iguaçu (HGNI). This department is the main referral center for HIV-infected patients, located in a populous and impoverished area in the metropolitan region of Rio de Janeiro – Southeast Region of Brazil, providing assistance, free ARVs and services for PMTCT of HIV. Inclusion criteria were confirmed reactive HIV testing and pregnancy of ARV-naïve women, with blood

samples collected before the beginning of treatment. Samples were collected between January/2012 and June/2015. Women with a previous history of HIV-1 PMTCT prophylaxis were excluded from this cohort. Plasma samples were stored at -70°C and later transported to the Laboratory of AIDS and Molecular Imunology (FIOCRUZ) in Rio de Janeiro for HIV subtype and resistance mutations analyses. This study was approved by the Ethics Committee of the Oswaldo Cruz Foundation/Oswaldo Cruz Institute (Number CAAE: 0067.0.011.316-11), and all women signed an informed consent form.

CD4 T cell count and HIV viral load determination

CD4+ T-cell counts were evaluated using flow cytometry (Trucount; FACSCalibur™; BD Biosciences Immunocytometry Systems, USA). Plasma HIV-1 RNA levels were determined by the VERSANT® HIV bDNA assay (Bayer 3.0 Quantitative Assay, USA) and by the Abbott RealTi*m*e HIV-1 assay (Abbott Molecular Inc., USA).

Sample processing and sequencing of HIV-1 *pol* gene

Viral RNA was extracted from plasma samples using the QIAamp Viral RNA Mini Kit (QIAGEN, Germany) according to the manufacturer's protocol, with a final elution volume of 50 µl. The cDNA synthesis was performed using 10µl of purified viral RNA in a 20µl final volume reagent mix containing 1X First-Strand Buffer (Tris-HCl 250mM pH8.3; KCl 375mM; MgCl2 15mM), 10mM of DTT, 40U of RNaseOUT™ (Life Technologies, USA), 0.5mM of dNTP mix, 0.1mM of reverse primer (MMRT6, Invitrogen, USA) and 200U of SuperScript® III reverse transcriptase enzyme (Life Technologies, USA). Samples were then incubated at 37°C for 90 minutes followed by 70°C for 15 minutes for enzyme denaturation.

The PR/RT fragment of the HIV-1 *pol* gene was amplified by nested polymerase chain reaction (nested-PCR). Nested-PCR reactions were carried out using 5 µl input template, 1.5U

of Platinum® Taq DNA Polymerase (Invitrogen, EUA), 1X PCR Buffer (20 mM Tris-HCl (pH 8.4), 50 mM KCl), 2.5mM of $MgCl_2$, 0.3mM of dNTPs mix (Invitrogen, USA), 0.35µM of each specific primer (forward and reverse) and diethylpyrocarbonate-treated water (Invitrogen, USA) to complete 50µl. The G17S and MMRT6 primers (Invitrogen, USA) were used to amplify the outer fragment. The inner amplification was conducted using MMRT10 and MMRT5 primers (Invitrogen, USA), resulting in a fragment of approximately 1,500bp. Cycling conditions were: one cycle of two minutes at 94ºC; 35 cycles of 30 seconds at 94ºC, 30 seconds at 55ºC and two minutes at 72ºC. When the second round of amplification was negative an alternative protocol was used, splitting the second round in two fragments following the same cycling described above and using the primer combinations MMRT10-RT4 (A fragment) and Pol4-MMRT5 (B fragment) (Invitrogen, USA). Table S1 describes the primers used in this study.

The final PCR products were purified using the Illustra GFX PCR DNA purification kit (GE Healthcare, USA) according to the manufacturer's protocol. Purified DNA was sequenced using the ABI BigDye Terminator v.3.1 cycle sequencing ready reaction kit (Applied Biosystems, USA). For each sample, eight primers were used to sequencing the entire PR and the first 335 codons of RT (HXB2: 2253-3554): MMRT10, LR51, DP16, RT4, RT9, Pol4, SEQ-RT and MMRT5 (Invitrogen, USA), ensuring that each nucleotide position was covered by at least two sequencing reactions. After isopropanol precipitation, the sequencing was performed with the automated ABI 3130xL and 3730xL Genetic Analyzers (Applied Biosystems, USA). This protocol is validated by the Virology Quality assessment (VQA) from Rush University, EUA, and is periodically tested for quality control assurance.

HIV-1 subtypes and resistance analyses in the *pol* gene

Chromatograms were assembled and carefully inspected for the presence of ambiguous sites using the SeqMan v7 software, Lasergene package (DNAStar, USA). HIV-1

genetic subtypes were initially determined with the REGA HIV subtyping tool v.3 [47] and later confirmed by ML phylogenetic and bootscaning analyses with HIV-1 reference sequences retrieved from the Los Alamos HIV database (ww.hiv.lanl.gov). ML trees were reconstructed with the PhyML 3.0 program [48], using the SPR branch-swapping algorithm for heuristic tree search and the approximate likelihood-ratio test (aLRT) [49] for estimating the reliability of the obtained tree topology. The ML trees were visualized using the FigTree v1.4.2 program [50]. Bootscaning analyses were performed with by Simplot 3.5.1 software [51], in which bootstrap values supporting branching with reference sequences were determined by the neighbour-joining method using the Kimura-2-parameter distance model and 100 bootstrap replicates of 250 nucleotides sliding windows moving in steps of 10 bases. The Calibrated Population Resistance Tool (CPR version 6.0; http://cpr.stanford.edu/cpr.cgi, available through the Stanford University HIV Drug Resistance Database) was used to analyse the PR and RT sequences [52]. Transmitted drug resistance mutation (TDRM) was defined as the presence of at least one major mutation described in the Surveillance Drug Resistance Mutation (SDRM–2009) list proposed by Bennett et al. [2009]. The minor resistance mutations were analysed based on the list of drug resistance mutations described by the International AIDS Society (IAS)-USA [40], however, these were not considered for the calculation of TDR prevalence.

Statistical analysis

Descriptive analyses (frequency, medians, averages, standard deviations) and univariate analyses using the two-tailed Fisher's exact test were performed in GraphPad Prism v6 (GraphPad Software, USA). Statistical significance was defined as an alpha level of 0.05.

Nucleotide sequence accession numbers

Sequences were deposited in GenBank under accession numbers KX357219 to KX357305.

Availability of supporting data

The data sets supporting the results of this article are included within the article.

ABBREVIATIONS

AIDS	acquired immune deficiency syndrome
aLRT	approximate likelihood-ratio test
ARV	antiretroviral
AZT	zidovudine
D4T	stavudine
EFV	efavirenz
HAART	highly active antiretroviral therapy
HIV-1	human immunodeficiency virus type 1
ML	maximum likelihood
NNRTI	non-nucleoside reverse transcriptase inhibitor
NRTI	nucleoside reverse transcriptase inhibitor
NVP	nevirapine
PI	protease inhibitor
PMTCT	prevent mother-to-child transmission
PR/RT	protease and reverse transcriptase
TDR	transmitted drug resistance
TDRM	transmitted drug resistance mutations
WHO	World Health Organization

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

MGM, ED and JHP designed the study. JHP and JCCF contributed with reagents/samples. ED and CSJ performed the experiments. ED and MGM analyzed the data. ED wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

ACKNOWLEDGMENTS

ED is supported by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). We thank all the patients and the staff of the Department of Sexually Transmitted Diseases (HGNI) who participated in this study. We also thank Dr. Vera Bongertz for the grammar review of the manuscript. The authors are grateful to the Plataforma Genômica – Sequenciamento de DNA-RPT01A-PDTIS/FIOCRUZ.

REFERENCES

1. UNAIDS Joint United Nations Program on HIV/AIDS: *The Gap Report*. Geneva, Switzerland; 2014.

2. Brasil: *Boletim Epidemiológico Aids E DST*. Brasília, DF - Brasil; 2014.

3. Malta M, Magnanini MMF, Mello MB, Pascom ARP, Linhares Y, Bastos FI: **HIV prevalence among female sex workers, drug users and men who have sex with men in Brazil: a systematic review and meta-analysis.** *BMC Public Health* 2010, **10**:317.

4. Kerr LRFS, Mota RS, Kendall C, Pinho AD a., Mello MB, Guimarães MDC, Dourado I, de Brito AM, Benzaken A, McFarland W, Rutherford G: **HIV among MSM in a large middle-income country**. *AIDS* 2013, **27**:427–435.

5. Brasil: *Protocolo Clínico E Diretrizes Terapêuticas Para Manejo Da Infecção Pelo HIV Em Adultos*. Brasília, DF - Brasil; 2013.

6. Marins JRP, Jamal LF, Chen SY, Barros MB, Hudes ES, Barbosa A a, Chequer P, Teixeira PR, Hearst N: **Dramatic improvement in survival among adult Brazilian AIDS patients.** *AIDS* 2003, **17**:1675–1682.

7. Tang JW, Pillay D: **Transmission of HIV-1 drug resistance**. *J Clin Virol* 2004, **30**:1–10.

8. Inocencio L a., Pereira A a., Sucupira MC a, Fernandez JCC, Jorge CP, Souza DFC, Fink HT, Diaz RS, Becker IM, Suffert T a., Arruda MB, MacEdo O, Simão MBG, Tanuri A: **Brazilian Network for HIV Drug Resistance Surveillance: A survey of individuals recently diagnosed with HIV**. *J Int AIDS Soc* 2009, **12**:1–6.

9. Brindeiro RM, Diaz RS, Sabino EC, Morgado MG, Pires IL, Brigido L, Dantas MC, Barreira D, Teixeira PR, Tanuri A: **Brazilian Network for HIV Drug Resistance Surveillance (HIV-BResNet): a survey of chronically infected individuals.** *AIDS* 2003, **17**:1063–1069.

10. Sprinz E, Netto EM, Patelli M, Lima JS, Furtado JJD, da Eira M, Zajdenverg R, Madruga J V, Lewi DS, Machado A a, Pedro RJ, Soares M a: **Primary antiretroviral drug resistance among HIV type 1-infected individuals in Brazil.** *AIDS Res Hum Retroviruses* 2009, **25**:861–867.

11. Alencar CS, Sabino EC, Carvalho SMF, Leao SC, Carneiro-Proietti AB, Capuani L, Oliveira CL, Carrick D, Birch RJ, Gonçalez TT, Keating S, Swanson P a, Hackett J, Busch MP: **HIV Genotypes and Primary Drug Resistance Among HIV-Seropositive Blood Donors in Brazil**. *JAIDS J Acquir Immune Defic Syndr* 2013, **63**:387–392.

12. Cardoso LPV, Pereira GAS, Viegas ÂA, Schmaltz LEPR, Stefani MM de A: **HIV-1 primary and secondary antiretroviral drug resistance and genetic diversity among pregnant women from central Brazil**. *J Med Virol* 2010, **82**:351–357.

13. Alcântara KC, Lins JBA, Albuquerque M, Aires LM, Cardoso LPV, Minuzzi AL, Stefani MMA: **HIV-1 mother-to-child transmission and drug resistance among Brazilian pregnant women with high access to diagnosis and prophylactic measures**. *J Clin Virol* 2012, **54**:15–20.

14. De Lourdes Teixeira M, Nafea S, Yeganeh N, Santos E, Isabel Gouvea M, Joao E, Ceci L, Bressan C, Leticia Cruz M, Claude Sidi L, Nielsen-Saines K: **High rates of baseline antiretroviral resistance among HIV-infected pregnant women in an HIV referral centre in Rio de Janeiro, Brazil**. *Int J STD AIDS* 2014, **0**:1–7.

15. Pilotto JH, Grinsztejn B, Veloso VG, Velasque LS, Friedman RK, Moreira RI, Rodrigues-Pedro A, Oliveira SM, Currier JS, Morgado MG: **Moderate prevalence of transmitted drug resistance mutations among antiretroviral-naive HIV-infected pregnant women in Rio de Janeiro, Brazil.** *AIDS Res Hum Retroviruses* 2013, **29**:681–6.

16. Da Costa ZB, de Lima YAR, Martelli CMT, Stefani MMDA: **Transmitted HIV resistance among pregnant young women infected with HIV-1 in Brazil.** *AIDS Patient Care STDs* 2013, **27**:439–41.

17. World Health Organization: *World Health Organization Global Strategy for the Surveillance and Monitoring of HIV Drug Resistance: An Update.* Geneva, Switzerland; 2012.

18. Abecasis AB, Deforche K, Bacheler LT, McKenna P, Carvalho AP, Gomes P, Vandamme AM, Camacho RJ: **Investigation of baseline susceptibility to protease inhibitors in HIV-1 subtypes C, F, G and CRF02_AG**. *Antivir Ther* 2006, **11**:581–589.

19. Poveda E, de Mendoza C, Parkin N, Choe S, García-Gasco P, Corral A, Soriano V: **Evidence for different susceptibility to tipranavir and darunavir in patients infected with distinct HIV-1 subtypes.** *AIDS* 2008, **22**:611–616.

20. Couto-Fernandez JC, Silva-de-Jesus C, Veloso VG, Rachid M, Gracie RSG, Chequer-Fernandez SL, Oliveira SM, Arakaki-Sanchez D, Chequer PJN, Morgado MG: **Human immunodeficiency virus type 1 (HIV-1) genotyping in Rio de Janeiro, Brazil: assessing subtype and drug-resistance associated mutations in HIV-1 infected individuals failing highly active antiretroviral therapy.** *Mem Inst Oswaldo Cruz* 2005, **100**:73–8.

21. Machado LF a, Ishak MOG, Vallinoto ACR, Lemos JAR, Azevedo VN, Moreira MRC, Souza MIM, Fernandes LM, Souza LLB, Ishak R: **Molecular epidemiology of HIV type 1 in northern Brazil: identification of subtypes C and D and the introduction of CRF02_AG in the Amazon region of Brazil.** *AIDS Res Hum Retroviruses* 2009, **25**:961–6.

22. Gaspareto KV, Mello FMMDA, Dias JRC, Meneguetti VAF, Storti MEG, Ferreira JLDP, Lança AM, Rodrigues R, Brígido LFDM, Teixeira JJV, Bertolini DA: **Genetic diversity and primary resistance among HIV-1-positive patients from Maringá, Paraná, Brazil**. *Rev Inst Med Trop Sao Paulo* 2012, **54**:207–213.

23. López-Lopes GIS, Lança AM, de Paula Ferreira JL, Souza LO, de Macedo Brígido LF: **Discrepancies of HIV-1 Reverse Transcriptase Resistance Interpretation of Insertions and Deletions between Two Genotypic Algorithms**. *Intervirology* 2013, **56**:217–223.

24. Alcântara KC, Reis MNG, Cardoso LPV, Bello G, Stefani MMA: **Increasing heterosexual transmission of HIV-1 subtype C in Inland Central Western Brazil**. *J Med Virol* 2013, **85**:396–404.

25. Cavalcanti a. MS, Lacerda HR, De Brito AM, Pereira S, Medeiros D, Oliveira S: **Antiretroviral resistance in individuals presenting therapeutic failure and subtypes of the human immunodeficiency virus type 1 in the Northeast Region of Brazil**. *Mem Inst Oswaldo Cruz* 2007, **102**:785–792.

26. Almeida SE, de Medeiros RM, Junqueira DM, Gräf T, Passaes CPB, Bello G, Morgado MG, L Guimarães M: **Temporal dynamics of HIV-1 circulating subtypes in distinct exposure categories in southern Brazil.** *Virol J* 2012, **9**:306.

27. Santos AF, Sousa TM, Soares EA, Sanabani S, Martinez AM, Sprinz E, Silveira J, Sabino EC, Tanuri A, Soares M a: **Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil.** *AIDS* 2006, **20**:2011–9.

28. Caride E, Brindeiro R, Hertogs K, Larder B, Dehertogh P, Machado E, de Sá C a, Eyer-Silva W a, Sion FS, Passioni LF, Menezes J a, Calazans a R, Tanuri a: **Drug-resistant reverse transcriptase genotyping and phenotyping of B and non-B subtypes (F and A) of human immunodeficiency virus type I found in Brazilian patients failing HAART.** *Virology* 2000, **275**:107–15.

29. Pimentel VF, Morgado MG, Bello G, Guimarães MDC, Castilho E a, Veloso VG, Guimarães ML: **Temporal trends and molecular epidemiology of HIV type 1 infection in Rio de Janeiro, Brazil.** *AIDS Res Hum Retroviruses* 2013, **29**:1553–61.

30. Morgado MG, Guimarães ML, Gripp CB, Costa CI, Neves I, Veloso VG, Linhares-Carvalho MI, Castello-Branco LR, Bastos FI, Kuiken C, Castilho EA, Galvão-Castro B, Bongertz V: **Molecular epidemiology of HIV-1 in Brazil: high prevalence of HIV-1 subtype B and identification of an HIV-1 subtype D infection in the city of Rio de Janeiro, Brazil. Evandro Chagas Hospital AIDS Clinical Research Group.** *J Acquir Immune Defic Syndr Hum Retrovirol* 1998, **18**:488–94.

31. Alencar CS, Sabino EC, Carvalho SMF, Leao SC, Carneiro-proietti AB, Capuani L, Oliveira CL, Carrick D, Birch RJ, Gonçalez TT, Keating S, Swanson PA, Hackett J, Busch MP, Retrovirus N: **HIV Genotypes and Primary Drug Resistance Among HIV-Seropositive Blood Donors in Brazil : Role of Infected Blood Donors as Sentinel Populations for Molecular Surveillance of HIV**. *JAIDS* 2013, **63**:387–392.

32. Delatorre EO, Bello G, Eyer-Silva WA, Chequer-Fernandez SL, Morgado MG, Couto-Fernandez JC: **Evidence of multiple introductions and autochthonous transmission of the HIV type 1 CRF02_AG clade in Brazil.** *AIDS Res Hum Retroviruses* 2012, **28**:1369–72.

33. Delatorre E, Couto-Fernandez JC, Guimarães ML, Vaz Cardoso LP, de Alcantara KC, Stefani MM de A, Romero H, Freire CCM, Iamarino A, de A Zanotto PM, Morgado MG, Bello G: **Tracing the origin and northward dissemination dynamics of HIV-1 subtype C in Brazil.** *PLoS One* 2013, **8**:e74072.

34. Bennett DE, Camacho RJ, Otelea D, Kuritzkes DR, Fleury H, Kiuchi M, Heneine W, Kantor R, Jordan MR, Schapiro JM, Vandamme AM, Sandstrom P, Boucher C a B, van de Vijver D, Rhee SY, Liu TF, Pillay D, Shafer RW: **Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update**. *PLoS One* 2009, **4**.

35. Bermúdez-Aza EH, Kerr LRFS, Kendall C, Pinho AA, de Mello MB, Mota RS, Guimarães MDC, Alencar CS, de Brito AM, Dourado IC, da Batista SMB, Abreu F, de Oliveira LC, de Souza Moraes A, Benzaken AS, Merchan-Hamann E, de Freitas GMB, McFarland W, Albuquerque E, Rutherford GW, Sabino E: **Antiretroviral Drug Resistance in a Respondent-Driven Sample of HIV-Infected Men Who Have Sex With Men in Brazil**. *JAIDS J Acquir Immune Defic Syndr* 2011, **57**:S186–S192.

36. Velasco-de-Castro CA, Grinsztejn B, Veloso VG, Bastos FI, Pilotto JH, Fernandes N, Morgado MG: **HIV-1 diversity and drug resistance mutations among people seeking HIV diagnosis in voluntary counseling and testing sites in Rio de Janeiro, Brazil.** *PLoS One* 2014, **9**:e87622.

37. Nachega JB, Parienti JJ, Uthman O a., Gross R, Dowdy DW, Sax PE, Gallant JE, Mugavero MJ, Mills EJ, Giordano TP: **Lower pill burden and once-daily antiretroviral treatment regimens for HIV infection: A meta-analysis of randomized controlled trials**. *Clin Infect Dis* 2014, **58**:1297–1307.

38. Parienti JJ, Das-Douglas M, Massari V, Guzman D, Deeks SG, Verdon R, Bangsberg DR: **Not all missed doses are the same: Sustained NNRTI treatment interruptions predict HIV rebound at low-to-moderate adherence levels**. *PLoS One* 2008, **3**:5–9.

39. Nijhuis M, van Maarseveen NM, Boucher C a B: **HIV protease resistance and viral fitness.** *Curr Opin HIV AIDS* 2007, **2**:108–115.

40. Wensing AM, Calvez V, Günthard HF, Johnson V a, Paredes R, Pillay D, Shafer RW, Richman DD: **2014 update of the drug resistance mutations in HIV-1.** *Top Antivir Med* 2014, **22**:642–50.

41. Liu TF, Shafer RW: **Web Resources for HIV Type 1 Genotypic-Resistance Test Interpretation**. *Clin Infect Dis* 2006, **42**:1608–1618.

42. Nunn AS, Fonseca EM, Bastos FI, Gruskin S, Salomon J a.: **Evolution of antiretroviral drug costs in Brazil in the context of free and universal access to AIDS treatment**. *PLoS Med* 2007, **4**:1804–1817.

43. Brasil: *Recomendações Para Profilaxia Da Transmisão Vertical Do HIV E Terapia Antirretroviral Em Gestantes*. Brasília, DF - Brasil; 2010.

44. Hemelaar J, Gouws E, Ghys PD, Osmanov S: **Global trends in molecular epidemiology of HIV-1 during 2000–2007**. *Aids* 2011, **25**:679–689.

45. Delatorre E, Velasco-De-Castro CA, Pilotto JH, Couto-Fernandez JC, Bello G, Morgado MG: **Reassessing the Origin of the HIV-1 CRF02_AG Lineages Circulating in Brazil**. *AIDS Res Hum Retroviruses* 2015, **0**:aid.2015.0183.

46. Bello G, Soares M a, Schrago CG: **The Use of Bioinformatics for Studying HIV Evolutionary and Epidemiological History in South America.** *AIDS Res Treat* 2011, **2011**:154945.

47. Pineda-Peña AC, Faria NR, Imbrechts S, Libin P, Abecasis AB, Deforche K, Gómez-López A, Camacho RJ, De Oliveira T, Vandamme AM: **Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools**. *Infect Genet Evol* 2013, **19**:337–348.

48. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.** *Syst Biol* 2010, **59**:307–21.

49. Anisimova M, Gascuel O: **Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative.** *Syst Biol* 2006, **55**:539–52.

50. **FigTree v1.4** [http://tree.bio.ed.ac.uk/software/figtree/]

51. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC: **Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination.** *J Virol* 1999, **73**:152–60.

52. Gifford RJ, Liu TF, Rhee SY, Kiuchi M, Hue S, Pillay D, Shafer RW: **The calibrated population resistance tool: Standardized genotypic estimation of transmitted HIV-1 drug resistance.** *Bioinformatics* 2009, 25:1197–1198.

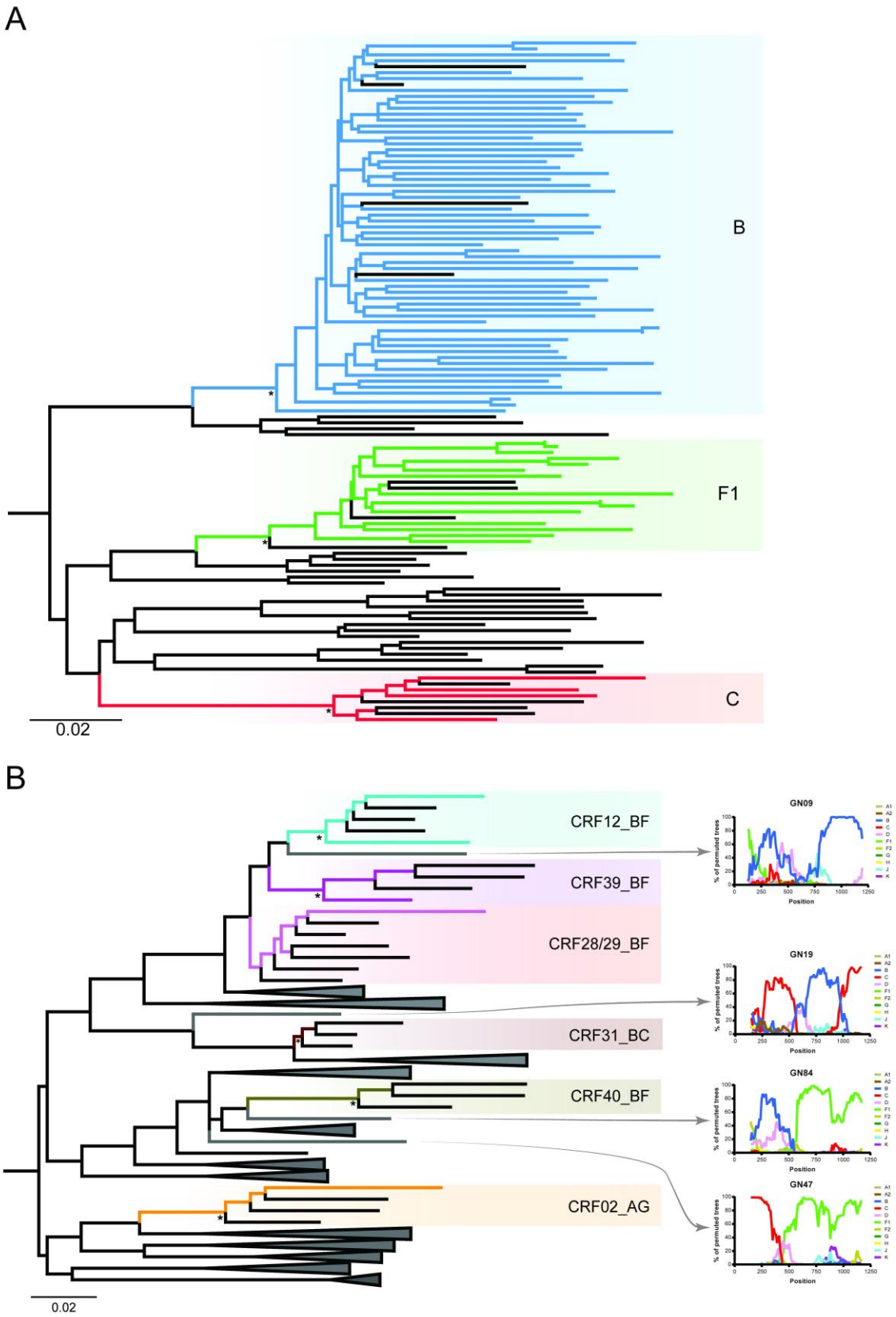

A
B
F1
C
0.02
B
CRF12_BF
CRF39_BF
CRF28/29_BF
CRF31_BC
CRF40_BF
CRF02_AG
0.02
GN09
GN19
GN84
GN47
% of permuted trees
Position

Figure 1

Figure 1. Subtype classification of sequences. A) Maximum likelihood (ML) tree of the PR/RT region. HIV-1 reference sequences of pure subtypes (A-D, F-H, J and K) were included and colored black. The branch support values are indicated as asterisks (aLRT > 0.90) at key nodes. Each HIV-1 clade found in the Brazilian samples is colored as indicated in the figure. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site. B) ML tree of the PR/RT region of the intersubtype recombinant samples. HIV-1 reference sequences of pure subtypes and some CRFs (CRF02_AG, CRF12_BF, CRF28_BF, CRF29_BF, CRF31_BC, CRF39_BF, CRF40_BF, CRF70_BF, CRF71_BF and CRF72_BF) were included and colored black. Each HIV-1 clade found or closely related to the Brazilian samples is colored as indicated in the figure. For visual clarity, some clades are collapsed into triangles and the branch support values are indicated as asterisks (aLRT > 0.90). Horizontal branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site. Unique recombinant forms are colored in gray and their bootscanning plots are depicted on the right.

TABLES

Table 1. Clinical data from the pregnant women included in this study stratified by year

Characteristic[a]	**2012**	**2013**	**2014/15**[b]	**2012-2015**
CD4+ T-cell count (cells/mm^3)	*N*=27 (%)	*N*=35 (%)	*N*=22 (%)	*N*=84 (%)
<350	6 (22.2)	11 (31.4)	10 (45.5)	27 (32.2)
350-500	13 (48.1)	8 (22.9)	7 (31.8)	28 (33.3)
>500	8 (29.6)	16 (45.7)	5 (22.7)	29 (34.5)
Viral load (RNA copies/ml)	*N*=27 (%)	*N*=38 (%)	*N*=20 (%)	*N*=85 (%)
<3 $\log_{10}$	6 (22.2)	5 (13.2)	1 (5.0)	12 (14.1)
3-4 $\log_{10}$	14 (51.9)	10 (26.3)	6 (30.0)	30 (35.3)
4-5 $\log_{10}$	5 (18.5)	20 (52.6)	9 (45.0)	34 (40.1)
5-6 $\log_{10}$	2 (7.4)	3 (7.9)	3 (15.0)	8 (9.4)
>6 $\log_{10}$	0 (0.0)	0 (0.0)	1 (5.0)	1 (1.1)

[a] Numbers vary because of missing information. [b] Samples from years 2014 and 2015 were joined to keep a comparable sample size.

Table 2. Protease inhibitor, nucleoside reverse transcriptase inhibitor, and nonnucleoside reverse transcriptase inhibitor transmitted drug resistance mutations found according to subtypes.

Sequence ID	Subtype	Mutations			Resistance profile		
		PI	NRTI	NNRTI	Low	Intermediate	High
GN02	B	M46I	–	–	NFV	–	–
GN03	F1	–	D67N	–	AZT/ d4T	–	–
GN12	B	M46L	–	–	NFV	–	–
GN13	B	–	D67N, K219Q	–	AZT/ d4T	–	–
GN14	B	M46L	–	–	NFV	–	–
GN16	F1	–	–	K101E, G190A	–	ETR/RPV	EFV/NVP
GN22	B	–	K70R	–	d4T	AZT	–
GN24	B	–	D67N, T69D	Y188L	AZT/ d4T/ETR	ddI	EFV/NVP/RPV
GN26	B	M46L	M41L	–	NFV/AZT/ d4T	–	–
GN31	B	M46L	–	–	NFV	–	–
GN44	F1	–	–	G190A	ETR/RPV	EFV	NVP
GN48	28/29_BF	D30N, N88D	–	–	–	–	NFV
GN68	B	M46L	–	–	NFV	–	–
GN70	B	–	–	K103N	–	–	EFV/NVP
GN72	B	–	–	K103N	–	–	EFV/NVP

PI, protease inhibitor; NRTI, nucleoside reverse transcriptase inhibitors; NNRTI, non-nucleoside reverse transcriptase inhibitors; AZT, zidovudine; ddI, didanosine; d4T, stavudine; EFV, efavirenz; ETR, etravirine ; NFV, nelfinavir; NVP, nevirapine; RPV, rilpivirine.

Table S1. Primers used in this study for RT/PCR and cDNA sequencing.

Name	Reaction[a]	Direction	Sequence	Nucleotide position[b]
G17S	PCR 1R	Forward	5'-AAAAAGGGCTGTTGGAAATGTGGA-3'	2017-2040
MMRT6	RT/PCR 1R	Reverse	5'-TTTTACATCATTAGTGTGGG-3'	3628-3647
MMRT10	PCR 2R/PCR 2RA/SEQ	Forward	5'-CAGGCTAATTTTTTAGGGAA-3'	2077-2096
MMRT5	PCR 2R/PCR 2RB/SEQ	Reverse	5'-TAAATTTGATATGTCCATTG-3'	3555-3574
LR51	SEQ	Reverse	5'-GTATTCCTAATTGAACYTCC-3'	2813-2832
DP16	SEQ	Forward	5'-CCTCARRTCACTCTTTGGCARC-3'	2253-2274
RT4	PCR 2RA/SEQ	Reverse	5'-AGTTCATAMCCCATCCA-3'	3234-3250
RT9	SEQ	Forward	5'-GTACAGTRTTAGTAGGACCTACA-3'	2470-2492
Pol4	PCR 2RB/SEQ	Forward	5'-CARTAYAATGTGCTTCCAC-3'	2982-3000
SEQ-RT	SEQ	Forward	5'-ATGGAAAGGATCACCAGCAA-3'	3005-3024

[a] RT – reverse transcription; PCR 1R – first round; PCR 2R – second round; PCR 2RA and PCR 2RB – alternative second round amplification; SEQ – sequencing.
[b] Positions relative to the HXB2 HIV genome (GenBank accession number K03455).

3.2 Artigo 2

Título: *Tracing the Origin and Northward Dissemination Dynamics of HIV-1 Subtype C in Brazil*

Autores: Edson Delatorre, José C. Couto-Fernandez, Monick L. Guimarães, Ludimila P. V. Cardoso, Keila C. de Alcantara, Mariane M. A. Stefani, Hector Romero, Caio C. M. Freire, Atila Iamarino, Paolo M. de A. Zanotto, Mariza G. Morgado e Gonzalo Bello

Publicado no periódico PLOS ONE

Resumo:

Estudos prévios indicam que a epidemia do subtipo C do HIV-1 na região Sul do Brasil foi iniciada pela introdução de uma única cepa fundadora provavelmente oriunda da África Oriental. Entretanto, o exato país de origem de tal cepa fundadora assim como a origem dos vírus do subtipo C detectados fora da região Sul do Brasil permanecem desconhecidos. Sequências do gene *pol* do HIV-1 isoladas nas regiões Sul, Sudeste e Centro-Oeste do Brasil (*n* = 209) forma comparadas com uma grande quantidade (*n* ~ 2000) de sequências do gene *pol* do subtipo C de origem africana. Análises de máxima-verossimilhança revelaram que a maioria das sequências brasileiras do subtipo C do HIV-1 agruparam-se em um clado único monofilético (C_{BR-I}), aninhado dentro de uma maior linhagem monofilética, característica da África Oriental. Análises bayesianas indicaram que o clado C_{BR-I} se originou provavelmente no Burundi e foi introduzido no estado do Paraná (região Sul) por volta de meados da década de 1970, após o qual rapidamente se disseminou para as regiões vizinhas. Os estados do Paraná e Santa Catarina foram os mais importantes eixos de disseminação do subtipo C, e viagens de rotina e acessibilidade espacial parecem ser as principais forças motrizes deste processo. Cinco introduções adicionais de cepas do subtipo C do HIV-1 provavelmente originadas em países da África Oriental (*n* = 2), Meridional (*n* = 2) e Central (*n* = 1) foram detectadas no estado do Rio de Janeiro (região Sudeste). Estes resultados indicam um influxo contínuo de cepas do subtipo C do HIV-1 de origem africana no Brasil e também revelam a existência de redes de transmissão desconhecidas ligando este país à África Oriental.

Tracing the Origin and Northward Dissemination Dynamics of HIV-1 Subtype C in Brazil

Edson Delatorre[1], José C. Couto-Fernandez[1], Monick Lindenmayer Guimarães[1], Ludimila Paula Vaz Cardoso[2], Keila Correia de Alcantara[2], Mariane Martins de Araújo Stefani[2], Hector Romero[3], Caio C. M. Freire[4], Atila Iamarino[4], Paolo M. de A Zanotto[4], Mariza G. Morgado[1], Gonzalo Bello[1*]

1 Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil, **2** Instituto de Patologia Tropical e Saúde Pública, Universidade Federal de Goiás, Goiânia, Brazil, **3** Laboratorio de Organización y Evolución del Genoma, Sección Biomatemáticas, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay, **4** Laboratório de Evolução Molecular e Bioinformática, Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, Brazil

Abstract

Previous studies indicate that the HIV-1 subtype C epidemic in southern Brazil was initiated by the introduction of a single founder strain probably originating from east Africa. However, the exact country of origin of such a founder strain as well as the origin of the subtype C viruses detected outside the Brazilian southern region remains unknown. HIV-1 subtype C *pol* sequences isolated in the southern, southeastern and central-western Brazilian regions (n = 209) were compared with a large number (n ~ 2,000) of subtype C *pol* sequences of African origin. Maximum-likelihood analyses revealed that most HIV-1 subtype C Brazilian sequences branched in a single monophyletic clade (C_{BR-I}), nested within a larger monophyletic lineage characteristic of east Africa. Bayesian analyses indicate that the C_{BR-I} clade most probably originated in Burundi and was introduced into the Paraná state (southern region) around the middle 1970s, after which it rapidly disseminated to neighboring regions. The states of Paraná and Santa *Catarina* have been the most important hubs of subtype C dissemination, and routine travel and spatial accessibility seems to have been the major driving forces of this process. Five additional introductions of HIV-1 subtype C strains probably originated in eastern (n = 2), southern (n = 2) and central (n = 1) African countries were detected in the Rio de Janeiro state (southeastern region). These results indicate a continuous influx of HIV-1 subtype C strains of African origin into Brazil and also unveil the existence of unrecognized transmission networks linking this country to east Africa.

Citation: Delatorre E, Couto-Fernandez JC, Guimarães ML, Vaz Cardoso LP, de Alcantara KC, et al. (2013) Tracing the Origin and Northward Dissemination Dynamics of HIV-1 Subtype C in Brazil. PLoS ONE 8(9): e74072. doi:10.1371/journal.pone.0074072

Editor: Rongge Yang, Chinese Academy of Sciences, Wuhan Institute of Virology, China

Received April 16, 2013; **Accepted** July 26, 2013; **Published** September 12, 2013

Funding: This work was supported by Public Health Service grants E-26/111.758/2012 from the FAPERJ and 472896/2012-1 from the CNPq. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-Mail: gbellobr@gmail.com

Introduction

The global spread of the Human immunodeficiency virus type 1 (HIV-1) group M that took place in the second half of the twentieth century was associated to the random exportation of some viral strains out of the epicenter in Central Africa into previously unexposed human populations [1]. The subsequent dissemination and diversification of the virus within such populations has resulted in the differential global distribution of HIV-1 group M subtypes and inter-subtype recombinants. The most prevalent HIV-1 group M variant worldwide is subtype C, which accounts for nearly half (48%) of all global infections [2]. This HIV-1 subtype is particularly prevalent in several countries from southern, eastern and central Africa, India and Brazil.

The Brazilian HIV-1 subtype C epidemic has been mostly restricted to the states of the southern region (Rio Grande do Sul [RS], Santa Catarina [SC] and Paraná [PR]) where this subtype accounts for between 20% and 80% of HIV-1 infections [3]. Previous phylogeographic studies indicate that the subtype C epidemic in southern Brazil was probably initiated by the introduction of a single founder strain into PR, followed by a rapid dissemination of the virus to the neighboring southern states [4,5]. The founder Brazilian subtype C strain probably originated in east Africa, although the exact country of origin and the precise time-scale of such an event remain uncertain [6]. One study conducted by our group points to Burundi as the most probable origin of the Brazilian subtype C clade [4], while other studies point to Ethiopia or Kenya [5,7]. Initial estimates based on viral strains

mostly sampled in Rio Grande do Sul propose that the founder event occurred around the early 1980s [4,8], but another study based on samples from several states traced back the origin of the Brazilian subtype C epidemic to 1960-1970 [5].

Recent studies have also documented a significant proportion of HIV-1 subtype C infections among individuals from different states across the southeast, central-west and north Brazilian regions, supporting a northward spread of HIV-1 subtype C in Brazil [3]. Subtype C was observed in 6-8% of patients from the São Paulo (SP) state [9,10], 0.5-1% of patients from the Rio de Janeiro (RJ) state [11,12,13], 3-11% of patients from the Goiás (GO) state [14,15,16], 5% of patients from the Mato Grosso (MT) state [17], 10% of patients from the Mato Grosso do Sul (MS) state [18] and 6% of patients from the Tocantins state [19]. Although those studies support an influx of variants from the southern region, the exact origin and dissemination dynamics of Brazilian subtype C viruses circulating outside the southern states has not been studied in detail up to date.

In the present study, we used a comprehensive data set of Brazilian (n = 209) and African (n > 2,000) HIV-1 subtype C *pol* sequences to reconstruct with more precision the geographic origin and the onset date of the HIV-1 subtype C clade introduced into southern Brazil. Moreover, we traced the dissemination dynamics of the HIV-1 subtype C epidemic in the southeast and central-west Brazilian regions. Spatial and temporal information were combined in a Bayesian framework to reconstruct migration events both within Brazil and between African countries and Brazil.

Materials and Methods

HIV-1 subtype C Brazilian sequences

The Brazilian HIV-1 subtype C dataset was composed of 209 sequences covering the entire protease and partial reverse transcriptase (PR/RT) genes (nt 2253-3272 relative to HXB2 clone) collected in eight states from the south (RS, SC and PR), southeast (SP and RJ), and central-west (GO, MT and MS) regions of Brazil (Figure 1). New PR/RT subtype C sequences were obtained from 32 individuals from RJ selected from a larger cohort of about 3,000 HIV-infected patients followed at outpatient clinics from the Public Health System distributed throughout the state that underwent HIV genotyping tests at the Laboratory of AIDS and Molecular Immunology (FIOCRUZ) between 2002 and 2011, as previously described [20]. The HIV-1 subtype C *pol* sequences from RJ were combined with sequences from SP (n = 18), GO (n = 16), MT (n = 4) and MS (n = 4) available at the Los Álamos HIV Sequence Database (www.hiv.lanl.gov) and described elsewhere [10,14,15,16,17,18,21,22,23,24], and with a dataset of sequences isolated in the south region (RS = 55, SC = 41 and PR = 39) described in detail in a previous study [25]. The study was approved by the Instituto Oswaldo Cruz - Ethics Committee. No informed consent from participants was obtained as the data were analyzed anonymously.

HIV-1 subtype C reference dataset

The HIV-1 subtype C Brazilian sequences were initially aligned with a reference set of 1,961 subtype C *pol* gene sequences of African origin obtained from the Los Álamos HIV Sequence Database. This reference data set, described in more detail in our previous study [26], includes subtype C sequences from therapy-naïve patients representative of the east (Burundi, Ethiopia, Kenya, Tanzania and Uganda), southern (Botswana, Malawi, Mozambique, South Africa, Zambia and Zimbabwe) and central (Angola and Democratic Republic of Congo) African regions sampled over a time period of 25 years (1986-2010). The basic local alignment search tool (BLAST) (www.ncbi.nlm.nih.gov/BLAST) was also used to select 50 subtype C reference sequences with known sampling dates isolated world-wide that displayed a high similarity score (> 95%) to specific Brazilian subtype C strains. The subtype assignment of all sequences here included was confirmed using the REGA HIV subtyping tool v.2 [27].

Sequence alignment and analysis of phylogenetic signal

Sequences were aligned using the CLUSTAL X program [28]. To avoid any bias on the phylogenetic reconstructions, all sites with major antiretroviral drug resistance mutations in PR (50, 82 and 90) and RT (41, 67, 70, 98, 103, 106, 179, 184, 190, 215 and 219) detected in at least two sequences were excluded from those alignments containing Brazilian subtype C sequences retrieved from treated-patients. All alignments are available from the authors upon request. Substitution saturation was evaluated in each alignment by plotting the estimated number of transitions and transversions against genetic distance for each pairwise comparison, using the DAMBE program [29]. The phylogenetic signal in each alignment was also investigated with the likelihood mapping method [30] by analyzing 10,000 random quartets. Likelihood mapping analyses were performed with the TREE-PUZZLE program [31], using the online web platform Phylemon 2.0 [32].

Phylogenetic analysis

Maximum Likelihood (ML) phylogenetic trees were inferred under the GTR+I+Γ_4 nucleotide substitution model, selected using the jModeltest program [33]. The ML tree was reconstructed with the PhyML program [34] using an online web server [35]. Heuristic tree search was performed using the SPR branch-swapping algorithm and the reliability of the obtained topology was estimated with the approximate likelihood-ratio test (aLRT) [36] based on the Shimodaira-Hasegawa-like procedure. The ML trees were visualized using the FigTree v1.3.1 program [37].

Analysis of spatiotemporal dispersion pattern

The evolutionary rate (μ, nucleotide substitutions per site per year, subst./site/year), the age of the most recent common ancestor ($T_{mrca,}$ years), and the spatial dynamics of different HIV-1 subtype C clades were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST v1.7.4 [38,39]. Analyses were

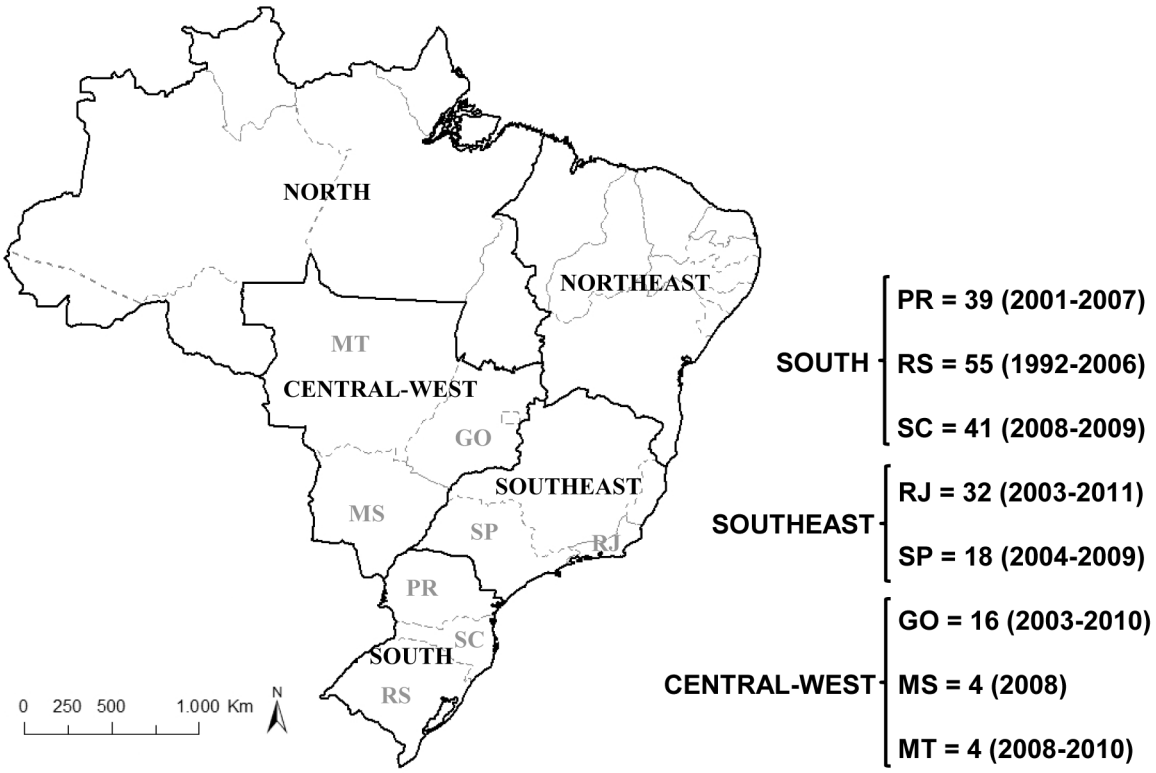


Figure 1. Political map of Brazil showing the country regions and states. Boundaries of regions and states are indicated by black and gray lines, respectively. The position of the eight states analyzed in the present study is indicated with a two letter code: GO (Goiás), MT (Mato Grosso), MS (Mato Grosso do Sul), PR (Paraná), RJ (Rio de Janeiro), RS (Rio Grande do Sul), SC (Santa Catarina) and SP (São Paulo). The number and sampling dates of HIV-1 subtype C pol sequences from each location included in the present study are indicated.

doi: 10.1371/journal.pone.0074072.g001

performed using the GTR+I+Γ_4 nucleotide substitution model, an uncorrelated Lognormal relaxed molecular clock model [40] and a Bayesian Skyline coalescent tree prior [41]. Migration events and the most relevant migration pathways between locations were identified by applying a standard discrete Bayesian phylogeographic model and the Bayesian stochastic search variable selection (BSSVS) approach [42], respectively. Migratory events and significant non-zero rates were summarized using the cross-platform SPREAD application [43] and viewed with Google, Earth (http://earth.google.com). MCMC chains were run for 4-5 × 10^8 generations and adequate chain mixing was checked, after excluding an initial 10%, by calculating the effective sample size (ESS) using the TRACER v1.5 program [44]. Maximum clade credibility (MCC) trees were summarized from the posterior set of trees (PST) with TreeAnnotator and visualized with FigTree v1.3.1.

Viral exchange rates among localities in Brazil were also estimated as transition rates between discrete characters along a PST generated using MrBayes v3.2.1 [45]. The PST was obtained during MCMC convergence from two independent runs with 2 × 10^7 generations and sampled at each 2,000 generations, employing the GTR+I+Γ_4 substitution model. Transition rates (q) were estimated using the APE package v3.06 [46] implemented in the R statistical environment v2.15.2 [47], under three different models: completely asymmetrical (ARD), symmetrical (SYM) and equal rates (ER). The best-fit model to our data was chosen by the comparison of the marginal Likelihoods from each one after 10,000 bootstrap replications, using the method proposed by Suchard et al. [48] implemented in Tracer v1.5. Due to the great uncertainty on the phylogenetic topologies obtained from HIV sequences and given that the estimated *q* values are subject to this issue, outliers from *q*' s distribution were removed using the boxplot function in R.

GenBank accession numbers

Nucleotide sequences obtained during our study have been assigned GenBank accession numbers KF247210, KF255836-KF255866.

Results

Identification of multiple HIV-1 subtype C introductions in Brazil

Four different datasets were used to reconstruct the origin and spatiotemporal dynamics of HIV-1 subtype C in Brazil (Tables S1 to S4). The transition/transversion *vs* divergence graphics and the likelihood-mapping analyses showed that all HIV-1 subtype C *pol* datasets used in this study contain enough evolutionary information for reliable phylogenetic and molecular clock inferences (Figure S1). The first dataset, here called C_{AFR+BR} (Table S1), was used to characterize the relationship between viruses sampled in Brazil (n = 209) with those circulating at 13 African countries (n = 1,961) with an estimated subtype C prevalence >5% [2]. The Brazilian subtype C strains were initially compared with those sequences from South Africa that represent the majority (52%) of subtype C sequences in our dataset. The close relative South African sequences were selected up to a maximum of 100 (Figure S2) and combined with subtype C sequences from the other African countries. The final ML phylogenetic tree revealed that most (98%) subtype C sequences from Brazil branched within a single monophyletic cluster ($C_{BR-I,}$ aLRT = 0.86) that was nested within a highly supported subtype C monophyletic clade ($C_{EA,}$ aLRT = 0.90) (Figure 2). The C_{EA} clade has been previously associated to the east African region [26] and comprise 73% of sequences from east Africa, 9% of sequences from central Africa, and none of sequences from southern Africa included in this analysis. Five (2%) Brazilian subtype C sequences branched outside the C_{BR-I} clade, constituting independent lineages. The lineage C_{BR-II} branched within the C_{EA} clade, while the remaining four lineages (C_{BR-III} to C_{BR-VI}) were dispersed outside that clade (Figure 2). All Brazilian subtype C sequences that branched outside the C_{BR-I} clade were sampled from Brazilian individuals who live in RJ state and were diagnosed with HIV-infection between 2006 and 2011.

Origin of Brazilian HIV-1 subtype C clades

The origin of each Brazilian HIV-1 subtype C clade was reconstructed using a Bayesian statistical framework that allows ancestral reconstruction of the locations at the interior nodes of Bayesian trees while accommodating phylogenetic uncertainty. To trace the origin of the C_{BR-I} clade we used a dataset ($C_{EA+BR-I}$) that combines all sequences from east Africa (n = 236) that branched within the C_{EA} clade and a subset of Brazilian sequences (n = 30) that were representative of the C_{BR-I} lineage (Table S2). The median evolutionary rate of the $C_{EA+BR-I}$ *pol* dataset, estimated under a chronological time-scale employing the dates of the sequences, was 1.8×10^{-3} (95% highest posterior density [HPD]: 1.3×10^{-3} - 2.4×10^{-3}) subst./site/year. The Bayesian MCC tree indicates that the C_{BR-I} clade most probably originated in Burundi (posterior state probability, PSP = 1) at around the middle 1970s, coinciding with the emergence of other major country-specific subclades in several east African countries including: Ethiopia (C_{ET}), Kenya (C_{KE}), Tanzania (C_{TZ}) and Uganda (C_{UG}) (Figure 3).

To determine the most probable geographic origin of the minor Brazilian subtype C clades we used an independent dataset ($C_{AFR+BR-II-VI}$) that combines the sequences C_{BR-II} to $C_{BR-IV,}$ their closest relative African sequences that branched with each minor Brazilian clade until the second ancestral node in the ML phylogenetic tree, and those subtype C sequences isolated world-wide with the highest BLAST search similarity score (> 95%) to each of the minor Brazilian subtype C lineages (Table S3). The Bayesian MCC tree suggests that the C_{BR-II} clade most probably originated in Burundi (PSP = 0.67) or Kenya (PSP = 0.14), the C_{BR-III} clade in Zambia (PSP = 0.65) or South Africa (PSP = 0.34), the C_{BR-IV} clade in Angola (PSP = 0.70) or Zambia (PSP = 0.15), the C_{BR-V} clade in Kenya (PSP = 0.89) and the C_{BR-VI} clade in South Africa (PSP = 0.99) (Figure 4).

Spatiotemporal dispersal pattern of the HIV-1 C_{BR-I} clade

To reconstruct the spatiotemporal dynamics of dissemination of the major Brazilian clade we used a fourth dataset ($C_{BI+BR-I}$) that comprises all Brazilian subtype C sequences that branched within the C_{BR-I} clade and a subset of 10 closely related sequences from Burundi (Table S2). The median evolutionary rate for this subtype C *pol* dataset also estimated under a chronological time-scale employing the dates of the sequences was 2.0×10^{-3} (95% HPD: 1.4×10^{-3} - 2.6×10^{-3}) subst./site/year. The Bayesian analysis placed the most probable root location in the state of PR (PSP = 0.83), followed by SC (PSP = 0.15), and set the maximum and minimum dates for such a founder event to 1972 (median T_{mrca} of the Brazilian and the closest Burundian sequences) and 1976 (median T_{mrca} of the C_{BR-I} clade), respectively (Figure 5). The overall topology of the Bayesian phylogenetic tree showed a great level of phylogenetic intermixing of Brazilian subtype C sequences from different geographic locations, with the exception of RS. A high proportion (78%) of subtype C sequences from RS branched within a single state-specific monophyletic cluster (C_{BR-RS}) (Figure 5). This analysis also identifies seven highly supported (posterior probability, *PP* > 0.85) geographic-specific monophyletic clades of small size (2-3 sequences) outside the southern region (RJ = 4, SP = 1, GO/MT = 1 and GO/MS = 1) (Figure 5). These local clusters comprise 11 (35%) out of 31 subtype C sequences from RJ, 2 (11%) out of 18 sequences from SP and 5 (21%) out of 24 sequences from the central-west region.

Reconstruction of viral migrations across time with the BEAST program revealed a rapid dissemination of the virus across Brazilian regions (Figure 6A). Between 1976 and 1980, the virus moves from PR to SC and from there to RS. During the 1980s, the virus migrates from PR to the southeast and central-west regions and from SC to RJ. In the following years, migration events from PR to RS, from SC to SP and the central-west region, and from RS to the southeast region were also registered. The Bayes factor (BF) tests for significant non-zero rates indicate well-supported rates (BF > 5) between PR/SC, PR/SP and PR/central-west region, and weakly supported rates (BF > 2) between PR/RS, PR/RJ, SC/RS and SC/RJ (Figure 6B). Viral movements among Brazilian localities were also estimated with the APE package. By comparing the

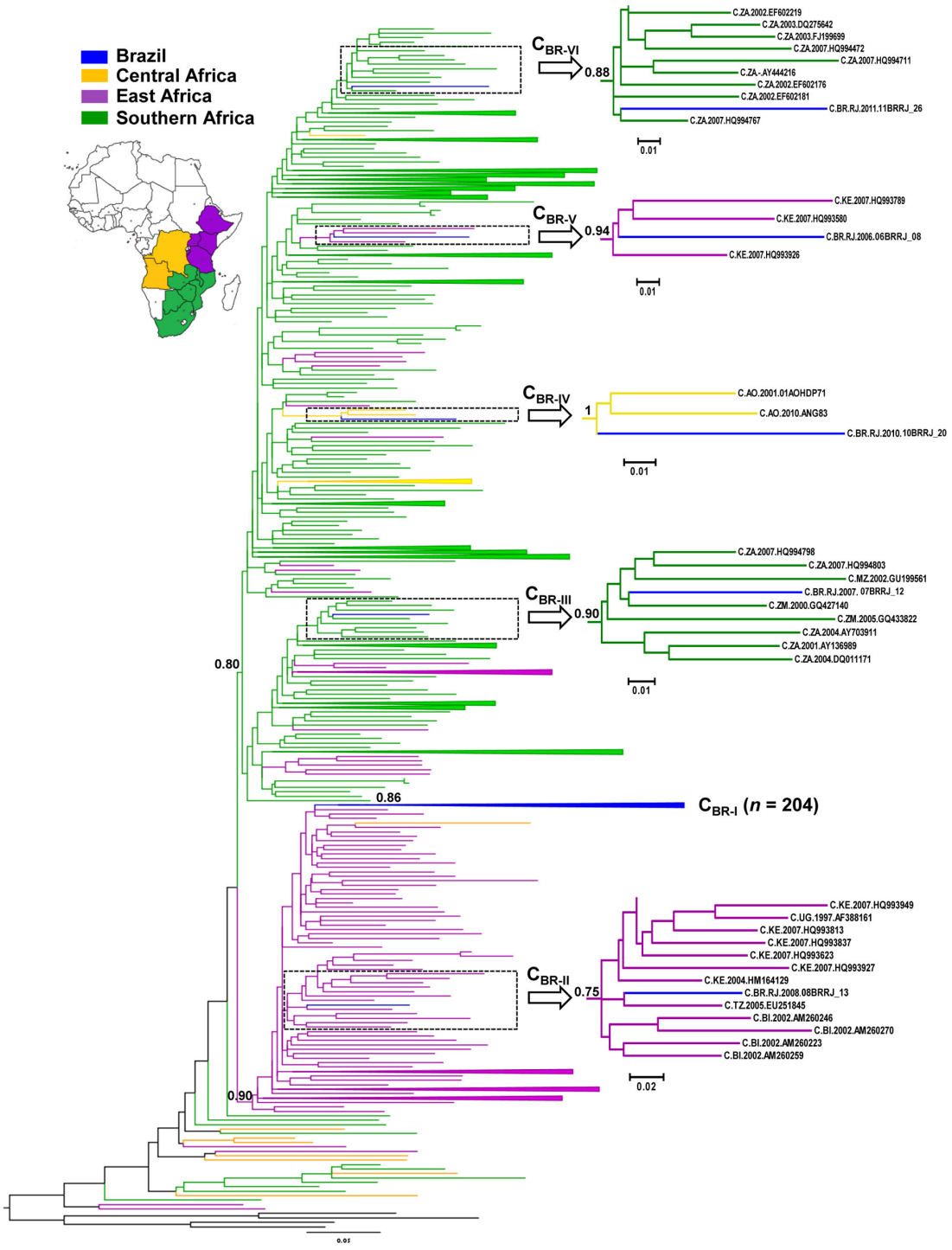


Figure 2. ML tree of HIV-1 subtype C pol (~1,000 bp) sequences from Brazil (*n* = 209) and the central (*n* = 53), eastern (*n* = 332) and southern (*n* = 645) African regions. The color of the branches represents the geographic region from where the subtype C sequences originated, according to the legend and map provide in the figure. The dotted boxes highlight the position of the Brazilian subtype C lineages (C_{BR-II} to C_{BR-IV}) that branched outside the major Brazilian clade (C_{BR-I}). A close view of the minor Brazilian subtype C lineages and the most closely related African sequences is also provided. For visual clarity, the Brazilian clade C_{BR-I} and some clades that comprised mostly sequences from central, eastern or southern Africa were collapsed into triangles. The *aLRT* support values are indicated only at key nodes. The tree was rooted using HIV-1 subtype A1 and D reference sequences (black branches). Horizontal branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.

doi: 10.1371/journal.pone.0074072.g002

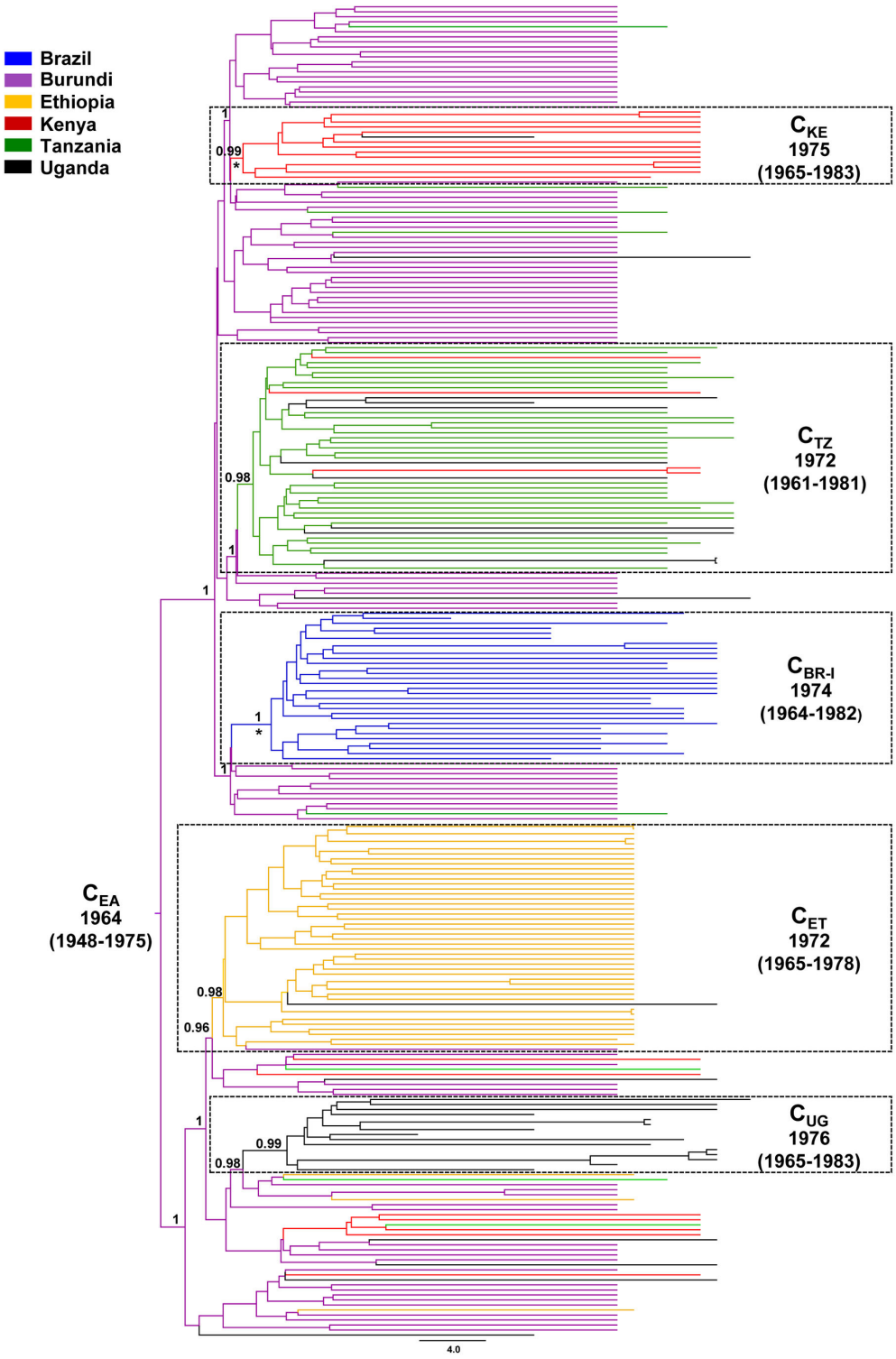


Figure 3. Time-scaled Bayesian MCC tree of the HIV-1 C_{EA} and C_{BR-I} lineages. Branches are colored according to the most probable location state of their descendent nodes. The legend for the colors is shown on the left. The dotted boxes highlight the position of the major country-specific sub-clades detected in our study. The median age (with 95% HPD interval in parentheses) and *PSP* values of key nodes are shown. Asterisks point to key nodes with a high (> 0.85) *PP* support. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating years. The tree was automatically rooted under the assumption of a relaxed molecular clock.

doi: 10.1371/journal.pone.0074072.g003

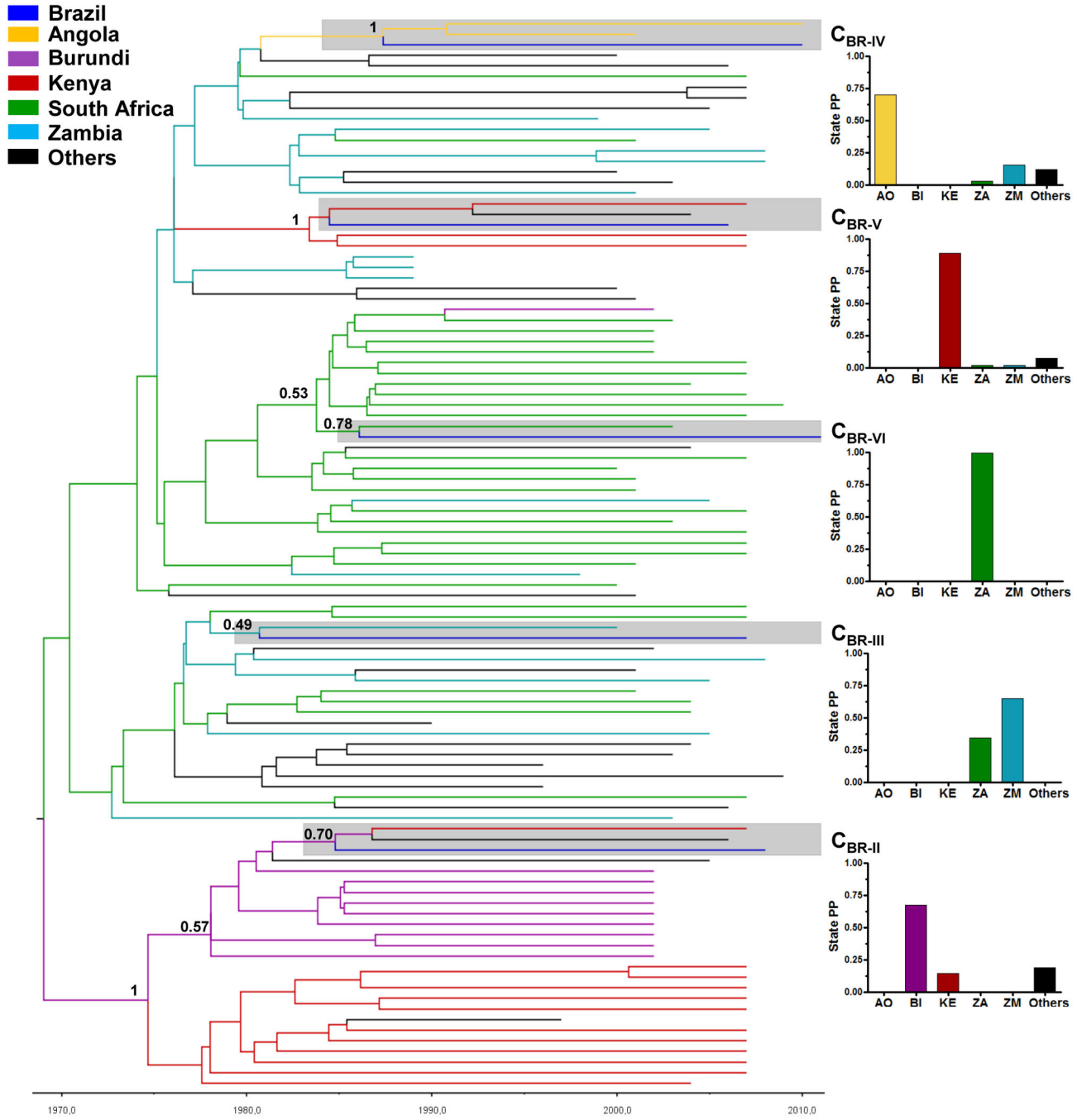


Figure 4. Time-scaled Bayesian MCC tree of the HIV-1 C_{BR-II} to C_{BR-VI} lineages and the most closely related reference sequences. Branches are colored according to the most probable location state of their descendent nodes. The legend for the colors is shown on the left. The boxes highlight the position of the minor Brazilian HIV-1 subtype C clades. The *PP* support is indicated only at key nodes. The scale bar at the bottom indicates years. The tree was automatically rooted under the assumption of a relaxed molecular clock. Graphics on the right depict the *PSP* distributions at the first ancestral nodes of Brazilian subtype C lineages at the Bayesian MCC tree. Countries represented are AO (Angola), BI (Burundi), KE (Kenya), ZA (South Africa), ZM (Zambia) and others (from Asia and Europe).

doi: 10.1371/journal.pone.0074072.g004

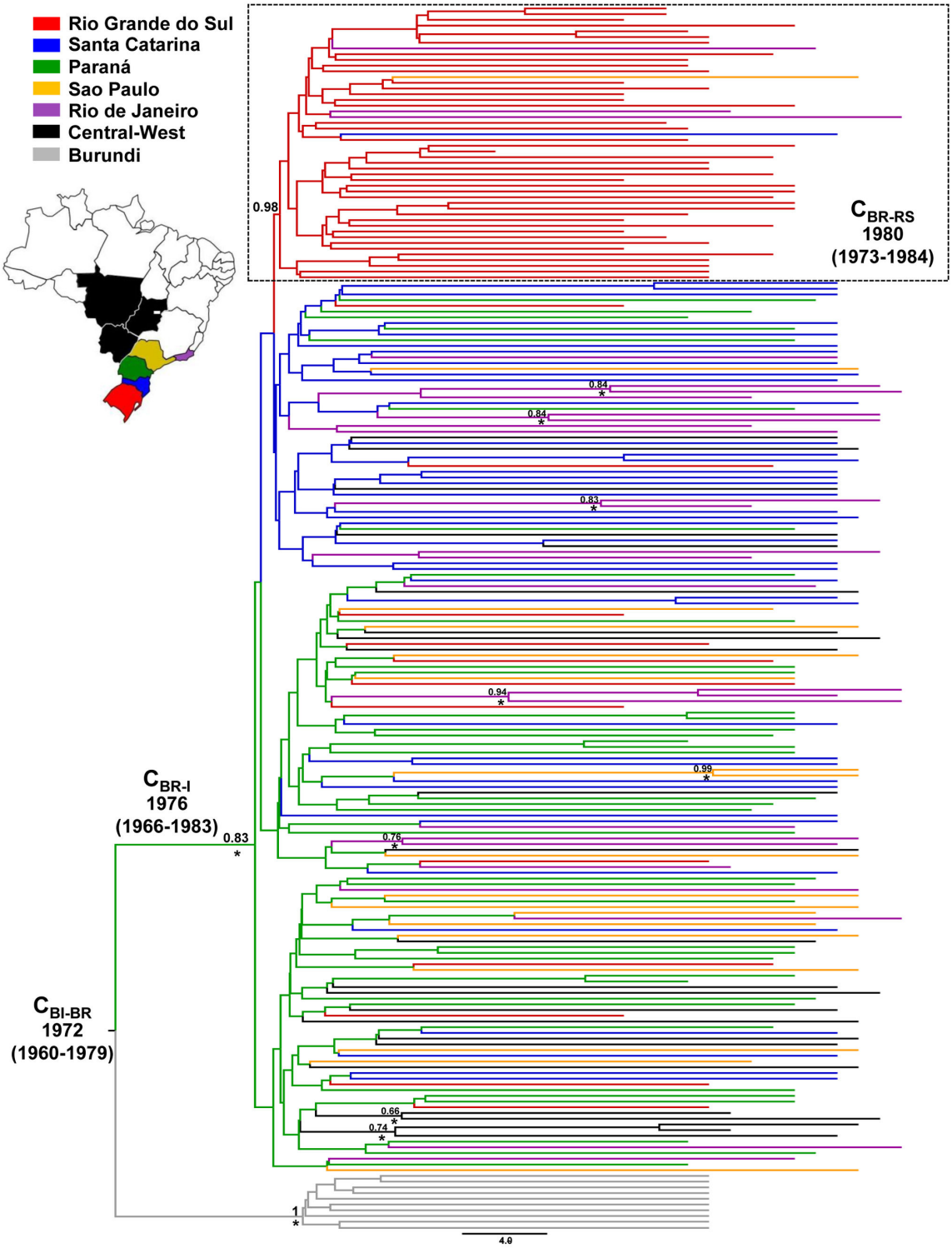


Figure 5. Time-scaled Bayesian MCC tree of the HIV-1 C_{BR-I} lineage. Branches are colored according to the most probable location state of their descendent nodes as indicated in the legend and map shown on the left. The dotted boxes highlight the position of the Brazilian sub-clade characteristic of Rio Grande do Sul (C_{BR-RS}). The median age (with 95% HPD interval in parentheses) and the *PSP* values of some key nodes are shown. Key nodes with a high (> 0.85) *PP* support are marked with an asterisk. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating years. The tree was automatically rooted under the assumption of a relaxed molecular clock.

doi: 10.1371/journal.pone.0074072.g005

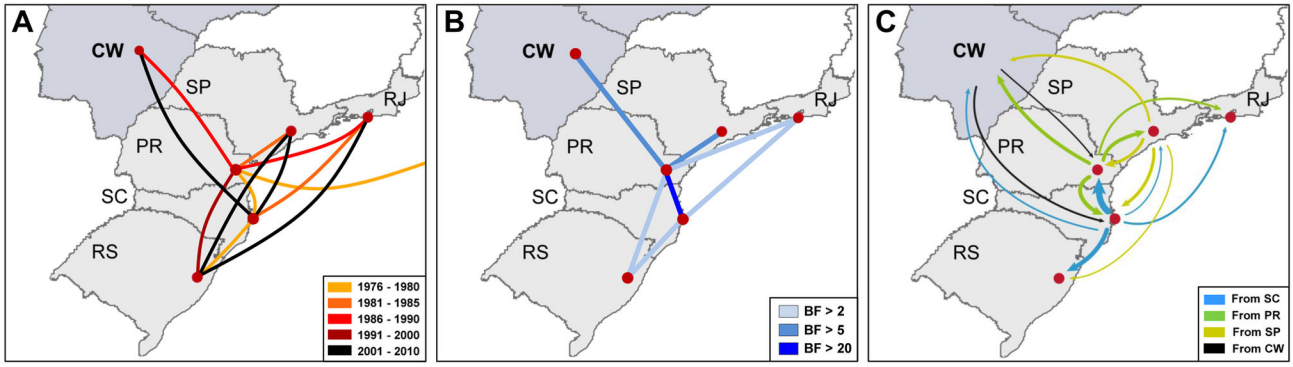


Figure 6. Spatiotemporal dynamics of HIV-1 C_{BR-I} clade dissemination in Brazil. (A) Viral dispersal pattern between 1976 and 2010. Lines between locations represent branches in the Bayesian MCC tree along which location transitions occurs. The yellow-black color gradient of lines informs the date of the earliest viral migrations among each pair of locations. (B) Bayes factor (BF) test for significant non-zero rates. Only rates supported by a BF greater than 2 are indicated. The light-dark color gradient of lines informs the relative strength by which the rates are supported (weak-strong). The maps are based on satellite pictures made available in Google™ Earth (http://earth.google.com). (C) Major estimated viral transitions rates (*q*) as measured by the APES program. The arrows were colored according to the source region and the width is proportional to *q*. All *q* lower than 1.0 were excluded for clarity. RS (Rio Grande do Sul), SC (Santa Catarina), PR (Paraná), RJ (Rio de Janeiro), SP (São Paulo) and CW (Central-west Region).

doi: 10.1371/journal.pone.0074072.g006

marginal likelihoods for each model, we found that the asymmetric one outperformed the other two models (Figure S3 and Table S4). Confirming previous analysis, PR and SC states displayed the most representative estimates of viral exchange and also acted as the most important hubs of spread to the southeast and central-west regions (Figure 6C and S4). This analysis further suggests that SP could be a secondary hub of viral dissemination to the south and central-west regions (Figure 6C and S4); while RS, RJ and the central-west regions came out as receiving ends (i.e., a sink), having few lineages moving to other states (Figure 6C and S4).

Human mobility and spread of the HIV-1 C_{BR-I} clade

To test the relevance of human mobility on the dissemination of HIV-1 subtype C epidemic in Brazil, viral transition rates estimated from the APE package were fitted to the routine travel and road distances between Brazilian localities from the south, southeast and central-west regions (Table S5). We found that viral movement between localities trend to be positively correlated with routine traffic among them (Figure 7A), although such a correlation became statistically significant ($P < 0.05$) only after the routine traffic was adjusted according to the estimated prevalence of subtype C in the state of origin (Figure 7B). We also found a significant negative correlation between viral transition rates and road distance, irrespective of the adjustment to the prevalence of subtype C in the state of origin (Figure 7C and 7D). Despite the statistical significance, the correlation coefficients obtained for all associations were low ($R^2 < 0.4$).

Discussion

The results presented here confirm the hypothesis that the major HIV-1 subtype C lineage circulating in Brazil (C_{BR-I}) originated in east Africa [4,7,8] and further show that this Brazilian lineage belongs to the previously called C_{EA} clade [26]. It has been estimated that the C_{EA} clade comprises 100% of the HIV-1 subtype C sequences from Burundi, 97% from Uganda, 64% from Kenya, 61% from Ethiopia, 49% from Tanzania and 9% from central African countries; while it is absent or extremely rare in southern Africa [26]. Among all African countries where the C_{EA} clade circulates, Burundi is the most probable source of the Brazilian C_{BR-I} lineage.

The median T_{mrca} of the C_{BR-I} clade was previously estimated at 1980-1983 [4], 1977-1980 [8] and 1962-1977 [5]. The two datasets here analyzed ($C_{EA+BR-I}$ and $C_{BI+BR-I}$) consistently traced the origin of the C_{BR-I} clade back to the middle 1970s (1974-1976) and situate the median upper and lower limits for subtype C introduction in Brazil at 1972 and 1976, respectively. This time-frame coincides with the onset date of other major country-specific C_{EA} subclades detected in Ethiopia, Kenya, Tanzania and Uganda [26]. Interestingly, the estimated dissemination of the C_{EA} clade from Burundi to other east African countries and Brazil overlapped with the first major civil conflict that took place in Burundi in 1972 and generated around 300,000 refugees [49]. This large human migration flow exiting Burundi may have played a crucial role in the regional and international dissemination of the C_{EA} clade. The exact route of migration of the C_{EA} clade from Burundi to Brazil, however, remains unclear. It has been suggested that the United Kingdom (UK) may have acted as a staging post in the dissemination of subtype C between east Africa and Brazil [8];

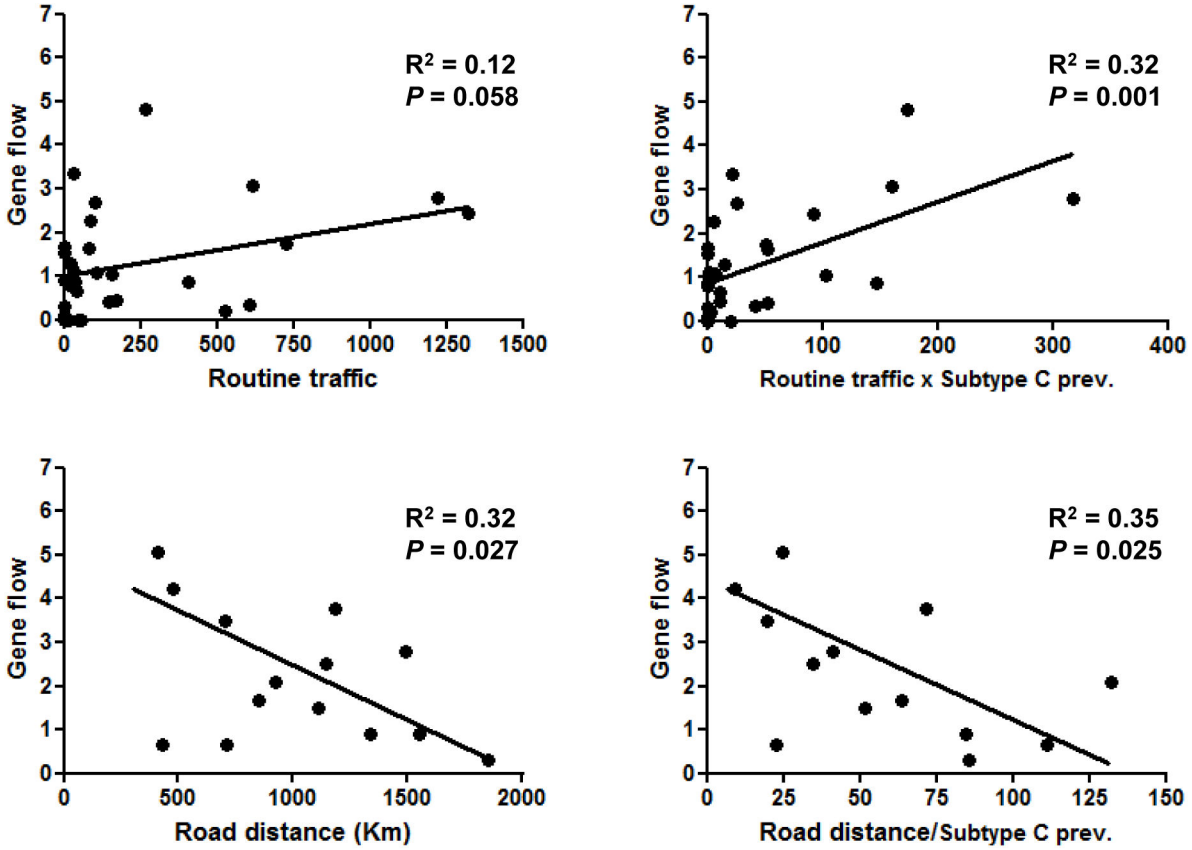


Figure 7. Association between viral migration rates and human mobility data. Viral transition rates (*q*, gene flow) were plotted against: (A) routine traffic ([people × trip]/1,000); (B) routine traffic multiplied by the HIV-1 subtype C prevalence in the site of origin; (C) road distances between state's capitals (the capital of Goiás state was used as reference for the central-west region); and (D) road distances divided by the average subtype C prevalence in the corresponding states. The linear regression line is shown in each graph. The *P*-value and r squared (R^2) from correlations are indicated in each plot.

doi: 10.1371/journal.pone.0074072.g007

but another study found no evidence of viral flow from the UK to Brazil, only from east Africa and Brazil to the UK [5].

Our phylogeographic reconstruction places the root of the C_{BR-I} clade in the state of PR with highest probability (PSP = 0.83), in agreement with previous studies [5,25]. By the early 1980s, the C_{BR-I} clade was already disseminated to the other southern Brazilian states, while between 1983 and 1988 the virus reached the southeast and the central-west regions. Despite a long-standing presence of the C_{BR-I} clade in all Brazilian regions, the final outcome of this HIV-1 clade across localities vary greatly. While subtype C accounts for 20-80% of HIV-1 infections in the southern states, the prevalence of this subtype remains ≤ 10% in the southeast and central-west regions. A recent study also showed an important expansion of HIV-1 subtype C infections amongst pregnant women from interior cities from the GO state, but not among those from the metropolitan area [16]. Thus, factors other than viral genetic characteristic and/or timing of viral introduction have shaped the expansion rate of the C_{BR-I} clade in different Brazilian regions. We propose that difference in the HIV transmission networks operating across localities may have contributed to such a heterogeneous spatial distribution pattern.

The states of PR and SC seems to be the main hubs of dissemination of the C_{BR-I} clade, exporting viruses to the other states. Estimation of viral movements with the APE package suggests that SP could be a secondary hub of viral dissemination sending viruses to the south and central-west regions; although this epidemiological link was not confirmed in the analysis with the BEAST program. Despite the high prevalence of subtype C and the large number of HIV cases in RS, this southernmost state seems to have a marginal role in the dissemination of the C_{BR-I} clade, sending only a few viral lineages to SC, SP and RJ. The results presented here point to a partially isolated subtype C epidemic in RS, consistent with our previous findings [25]. A large proportion (78%) of subtype C infections in RS appeared to be the result of the *in situ* dissemination of a single local sublineage (C_{BR-RS}) that probably emerged at around 1980 and is mostly restricted to that area. Some highly supported geographic-specific monophyletic clades of small size were also identified in RJ, SP and the

central-west region, revealing the existence of local transmission networks operating outside the southern region.

Routine travel and spatial accessibility among Brazilian regions has been pointed as possible driving forces of subtype C dissemination [5,25] and our results are fully consistent with this model. Viral exchanges between Brazilian localities increase as the routine traffic increases and the road distance (accessibility) decreases. The seeding of subtype C in the central-west region mainly from PR is also in line with the recent human migration in that direction, due to soybean plantation and similar agricultural activities. The HIV-1 subtype C prevalence in each locality seems to be another important factor to explain the rates of viral migration. The overall low prevalence of subtype C in SP (<10%) and RJ (<1%), for example, may explain the low viral exchange between both states despite their close geographical proximity and high routine traffic. By contrast, accessibility, human mobility and subtype C prevalence cannot explain the low level of viral migration from RS to SC. Indeed, the low correlation coefficients observed ($R^2 < 0.4$) indicate that additional factors also have influenced the viral dissemination process.

While the C_{BR-I} lineage was the only subtype C clade detected in the southern and central-west Brazilian regions, five additional subtype C introductions were detected in the southeast region, particularly in the state of RJ. Although those five subtype C viruses may have been acquired locally, there is no evidence that they have become widely disseminated in the country as they were represented by only one individual each. Other studies have also identified the circulation of HIV-1 variants of African origin in the states of RJ and SP, including subtype D and the CRF02_AG [10,20,50,51,52]. These states host large international airports, ports and sociocultural and economic events, which create an excellent milieu for the introduction of new HIV-1 strains in the area. Of note, our phylogeographic analyses suggest that two of those five subtype C viruses introduced into RJ were probably imported from Burundi and/or Kenya. The identification of these additional introductions uncovers the existence of unrecognized transmission networks linking Brazil to east Africa.

The unequal number of sequences available from different countries and Brazilian regions can introduce large biases in phylogeographic reconstructions and influence the conclusions. Some of our key findings, however, were robust to the sampling scheme used here. Although most (39%) HIV-1 sequences of the C_{EA} clade were from Burundi, a putative epidemiological link between the Brazilian lineage C_{BR-I} and any other east African country could be easily established because most sequences from Ethiopia, Kenya, Tanzania and Uganda were distributed in well defined country-specific sub-clades. The clade C_{BR-I}, however, was clearly placed among Burundian sequences and outside the major specific sub-clades from the other east African countries, thus supporting the Burundian origin of that major subtype C Brazilian lineage. Our study also indicates that PR was the most probable entrance point and one of the main hubs of dissemination of clade C_{BR-I} in Brazil despite the majority of subtype C Brazilian sequences within the major clade were from RS and SC.

In summary, the results presented here suggest that the HIV-1 subtype C epidemic spreading in most Brazilian states was initiated at around the middle 1970s by the introduction of a single founder strain originated in Burundi. Such a founder subtype C variant was probably introduced into PR and was rapidly disseminated to the other Brazilian states, originating the major C_{BR-I} clade. The states of PR and SC seem to be the most important hubs of the HIV-1 subtype C dissemination in Brazil. The explanation for the dissemination process of C_{BR-I} clade in Brazil is multifactorial and includes human mobility, accessibility, and local founder events among others. This study also identifies a continuous introduction of new HIV-1 subtype C variants of African origin into the RJ state. These results emphasize the importance of the continuous surveillance of HIV-1 subtype C genetic diversity to understand the dissemination dynamics of the C_{BR-I} clade at country level and for earliest detection of the introduction and dissemination of newly emerging subtype C viral clades in the Brazilian population.

Supporting Information

Figure S1. Substitution saturation and likelihood mapping analyses. (A) Transition (blue line) and transversion (green line) versus divergence plot for the different HIV-1 subtype C pol datasets. (B) Percentage of dots plotted in each region of the map after likelihood mapping of 10,000 random quarters selected from the different HIV-1 subtype C pol datasets. Each dot represents the likelihoods of the three possible tree topologies for a set of four sequences (quartets) selected randomly from the dataset. The dots localized on the vertices, in the center and on the laterals represent the tree-like, the star-like and the network-like phylogenetic signals, respectively.
(PDF)

Figure S2. ML tree of HIV-1 subtype C *pol* (~1,000pb) sequences from Brazil (*n* = 209) and South Africa (*n* = 1,031). Branches of Brazilian sequences are represented in red. Those branches of South African sequences that were more closely related to the Brazilian ones and were selected for further phylogenetic analyses are indicated in green. For visual clarity, some Brazilian and South African clades were collapsed. The *aLRT* support values are indicated only at key nodes. The tree was rooted using HIV-1 subtype A1 and D reference sequences (gray branches). Horizontal branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.
(PDF)

Figure S3. Distribution of the likelihood for three distinct models of viral transition rates. ER: model with equal rates among localities (black line). SYM: model with symmetric rates among localities (blue line). ARD: model with asymmetric rates among localities (red line).
(PDF)

Figure S4. Estimated viral transition rates (*q*) to and from each locality. All *q* lower than 0.5 were excluded for clarity. A – RS (Rio Grande do Sul, in red). B-SC (Santa *Catarina*, in blue). C-PR (Paraná, in green). D-SP (São Paulo, in yellow). E-RJ (Rio de Janeiro, in purple). F-CW (Central-west region, in black). The arrows width is proportional to *q* (available in Table S5).
(PDF)

Table S1. HIV-1 C_{AFR+BR} dataset.
(DOC)

Table S2. HIV-1 $C_{EA+BR-I}$ and $C_{BI+BR-I}$ datasets.
(DOC)

Table S3. HIV-1 $C_{AFR+BR-II-VI}$ dataset.
(DOC)

Table S4. Harmonic mean of Likelihoods for distinct models of viral transition rates.
(DOC)

Table S5. Viral transition rates, routine travels and road distances between localities.
(DOC)

Acknowledgements

We wish to thank Dr Vera Bongertz for critical reviewing of the manuscript.

Author Contributions

Conceived and designed the experiments: GB ED MGM PMAZ MMAS. Performed the experiments: ED JCCF MLG LPVC KCA. Analyzed the data: GB ED HR CCMF AI PMAZ. Wrote the manuscript: GB ED MGM PMAZ MMAS HR CCMF AI JCCF MLG LPVC KCA.

References

1. Archer J, Robertson DL (2007) Understanding the diversification of HIV-1 groups M and O. AIDS 21: 1693-1700. doi:10.1097/QAD.0b013e32825eabd0. PubMed: 17690566.
2. Hemelaar J, Gouws E, Ghys PD, Osmanov S (2011) Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS 25: 679-689. doi:10.1097/QAD.0b013e328342ff93. PubMed: 21297424.
3. Gräf T, Pinto AR (2012) The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. Virology, 435: 170–8. PubMed: 22999094.
4. Bello G, Passaes CP, Guimarães ML, Lorete RS, Matos Almeida SE et al. (2008) Origin and evolutionary history of HIV-1 subtype C in Brazil. AIDS 22: 1993-2000. doi:10.1097/QAD.0b013e328315e0aa. PubMed: 18753928.
5. Véras NM, Gray RR, de Macedo Brigido LF, Rodrigues R, Salemi M (2011) High-resolution phylogenetics and phylogeography of human immunodeficiency virus type 1 subtype C epidemic in South America. J Gen Virol 92: 1698-1709. doi:10.1099/vir.0.028951-0. PubMed: 21450946.
6. Bello G, Soares MA, Schrago CG (2011) The Use of Bioinformatics for Studying HIV Evolutionary and Epidemiological History in South America. AIDS Res Treat, 2011: 2011: 154945. PubMed: 22162803
7. Fontella R, Soares MA, Schrago CG (2008) On the origin of HIV-1 subtype C in South America. AIDS 22: 2001-2011. doi:10.1097/QAD.0b013e3283108f69. PubMed: 18784462.
8. de Oliveira T, Pillay D, Gifford RJ (2010) The HIV-1 subtype C epidemic in South America is linked to the United Kingdom. PLOS ONE 5: e9311. doi:10.1371/journal.pone.0009311. PubMed: 20174561.
9. Inocencio LA, Pereira AA, Sucupira MC, Fernandez JC, Jorge CP et al. (2009) Brazilian Network for HIV Drug Resistance Surveillance: a survey of individuals recently diagnosed with HIV. J International AIDS Society 12: 20. PubMed: 19765271
10. Brígido LF, Ferreira JL, Almeida VC, Rocha SQ, Ragazzo TG et al. (2011) Southern Brazil HIV type 1 C expansion into the state of São Paulo, Brazil. AIDS Res Hum Retrovir 27: 339-344. doi:10.1089/aid.2010.0157. PubMed: 20950149.
11. Couto-Fernandez JC, Silva-de-Jesus C, Veloso VG, Rachid M, Gracie RS et al. (2005) Human immunodeficiency virus type 1 (HIV-1) genotyping in Rio de Janeiro, Brazil: assessing subtype and drug-resistance associated mutations in HIV-1 infected individuals failing highly active antiretroviral therapy. Mem Inst Oswaldo Cruz 100: 73-78. doi:10.1590/S0074-02762005000100014. PubMed: 15867968.
12. Pilotto JH, Grinsztejn B, Veloso VG, Velasque LS, Friedman RK et al. (2013) Moderate Prevalence of Transmitted Drug Resistance Mutations Among Antiretroviral-Naive HIV-Infected Pregnant Women in Rio de Janeiro, Brazil. AIDS Res Hum Retrovir, 29: 681–6. PubMed: 23259924.
13. de Castro CA, Grinsztejn B, Veloso VG, Bastos FI, Pilotto JH et al. (2010) Prevalence, estimated HIV-1 incidence and viral diversity among people seeking voluntary counseling and testing services in Rio de Janeiro, Brazil. BMC Infect Dis 10: 224. doi: 10.1186/1471-2334-10-224. PubMed: 20667113.
14. Cardoso LP, Queiroz BB, Stefani MM (2009) HIV-1 pol phylogenetic diversity and antiretroviral resistance mutations in treatment naive patients from Central West Brazil. J Clin Virol 46: 134-139. doi:10.1016/j.jcv.2009.07.009. PubMed: 19682948.
15. Cardoso LP, Pereira GA, Viegas AA, Schmaltz LE, Stefani MM (2010) HIV-1 primary and secondary antiretroviral drug resistance and genetic diversity among pregnant women from central Brazil. J Med Virol 82: 351-357. doi:10.1002/jmv.21722. PubMed: 20087934.
16. Alcântara KC, Reis MN, Cardoso LP, Bello G, Stefani MM (2012) Increasing heterosexual transmission of HIV-1 subtype C in Inland Central western Brazil. J Med Virol, 85: 396–404. PubMed: 23172151.
17. Ferreira AS, Cardoso LP, Stefani MM (2011) Moderate prevalence of transmitted drug resistance and high HIV-1 genetic diversity in patients from Mato Grosso State, Central Western Brazil. J Med Virol 83: 1301-1307. doi:10.1002/jmv.22128. PubMed: 21678433.
18. da Silveira AA, Cardoso LP, Francisco RB, de Araújo Stefani MM (2012) HIV type 1 molecular epidemiology in pol and gp41 genes among naive patients from Mato Grosso do Sul State, central western Brazil. AIDS Res Hum Retrovir 28: 304-307. doi:10.1089/aid.2011.0128. PubMed: 21790471.
19. Carvalho BC, Cardoso LP, Damasceno S, Stefani MM (2011) Moderate prevalence of transmitted drug resistance and interiorization of HIV type 1 subtype C in the inland North State of Tocantins, Brazil. AIDS Res Hum Retrovir 27: 1081-1087. doi:10.1089/aid.2010.0334. PubMed: 21417758.
20. Delatorre EO, Bello G, Eyer-Silva WA, Chequer-Fernandez SL, Morgado MG et al. (2012) Evidence of Multiple Introductions and Autochthonous Transmission of the HIV Type 1 CRF02_AG Clade in Brazil. AIDS Res Hum Retrovir, 28: 1369–72. PubMed: 22333001.
21. Sanabani S, Neto WK, de Sa Filho DJ, Diaz RS, Munerato P et al. (2006) Full-length genome analysis of human immunodeficiency virus type 1 subtype C in Brazil. AIDS Res Hum Retrovir 22: 171-176. doi: 10.1089/aid.2006.22.171. PubMed: 16478399.
22. de Sa-Filho DJ, Soares Mda S, Candido V, Gagliani LH, Cavaliere E et al. (2008) HIV type 1 pol gene diversity and antiretroviral drug resistance mutations in Santos, Brazil. AIDS Res Hum Retrovir 24: 347-353. doi:10.1089/aid.2007.0203. PubMed: 18327988.
23. Cardoso LP, da Silveira AA, Francisco RB, da Guarda Reis MN, Stefani MM (2011) Molecular characteristics of HIV type 1 infection among prisoners from Central Western Brazil. AIDS Res Hum Retrovir 27: 1349-1353. doi:10.1089/aid.2011.0153. PubMed: 21732793.

24. Alcântara KC, Lins JB, Albuquerque M, Aires LM, Cardoso LP et al. (2012) HIV-1 mother-to-child transmission and drug resistance among Brazilian pregnant women with high access to diagnosis and prophylactic measures. J Clin Virol 54: 15-20. doi:10.1016/j.jcv.2012.01.011. PubMed: 22317908.
25. Bello G, Zanotto PM, Iamarino A, Gräf T, Pinto AR et al. (2012) Phylogeographic analysis of HIV-1 subtype C dissemination in Southern Brazil. PLOS ONE 7: e35649. doi:10.1371/journal.pone.0035649. PubMed: 22530062.
26. Delatorre EO, Bello G (2012) Phylodynamics of HIV-1 Subtype C Epidemic in East Africa. PLOS ONE 7: e41904. doi:10.1371/journal.pone.0041904. PubMed: 22848653.
27. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D et al. (2005) An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 21: 3797-3800. doi:10.1093/bioinformatics/bti607. PubMed: 16076886.
28. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876-4882. doi:10.1093/nar/25.24.4876. PubMed: 9396791.
29. Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. J Hered 92: 371-373. doi:10.1093/jhered/92.4.371. PubMed: 11535656.
30. Strimmer K, von Haeseler A (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci U S A 94: 6815-6819. doi:10.1073/pnas.94.13.6815. PubMed: 9192648.
31. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502-504. doi:10.1093/bioinformatics/18.3.502. PubMed: 11934758.
32. Sanchez R, Serra F, Tarraga J, Medina I, Carbonell J et al. (2011) Phylemon. p. 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. Nucleic Acids Res 39: W470-474.
33. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25: 1253-1256. doi:10.1093/molbev/msn083. PubMed: 18397919.
34. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696-704. doi:10.1080/10635150390235520. PubMed: 14530136.
35. Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33: W557-W559. doi:10.1093/nar/gki352. PubMed: 15980534.
36. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol 55: 539-552. doi:10.1080/10635150600755453. PubMed: 16785212.
37. Rambaut A (2009) FigTree v1.3.1: Tree Figure Drawing Tool. Available: http://treebioedacuk/software/figtree/
38. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161: 1307-1320. PubMed: 12136032.
39. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214. doi:10.1186/1471-2148-7-214. PubMed: 17996036.
40. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLOS Biol 4: e88. doi:10.1371/journal.pbio.0040088. PubMed: 16683862.
41. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22: 1185-1192. doi:10.1093/molbev/msi103. PubMed: 15703244.
42. Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. PLOS Comput Biol 5: e1000520. PubMed: 19779555.
43. Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27: 2910-2912. doi:10.1093/bioinformatics/btr481. PubMed: 21911333.
44. Rambaut A, Drummond A (2007) Tracer v1.4. Available: http://beastbioedacuk/Tracer.
45. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A et al. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61: 539-542. doi:10.1093/sysbio/sys029. PubMed: 22357727.
46. Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20: 289-290. doi:10.1093/bioinformatics/btg412. PubMed: 14734327.
47. Team RC (2012) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. Available: http://wwwR-projectorg/
48. Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol 18: 1001-1013. doi:10.1093/oxfordjournals.molbev.a003872. PubMed: 11371589.
49. Fransen S, Ong'ayo A (2010) Migration in Burundi: History, Current Trends, and Future Prospects. Paper Series: Migration and Development Country Profiles Maastricht: Maastricht Graduate School of Governance Available: http://mgsog.merit.unu.edu/ISacademie/docs/CR_burundi.pdf.
50. Couto-Fernandez JC, Eyer-Silva WA, Guimaraes ML, Chequer-Fernandez SL, Grinsztejn B et al. (2006) Phylogenetic analysis of Brazilian HIV type 1 subtype D strains: tracing the origin of this subtype in Brazil. AIDS Res Hum Retrovir 22: 207-211. doi:10.1089/aid.2006.22.207. PubMed: 16478405.
51. Morgado MG, Guimaraes ML, Gripp CB, Costa CI, Neves I Jr. et al. (1998) Molecular epidemiology of HIV-1 in Brazil: high prevalence of HIV-1 subtype B and identification of an HIV-1 subtype D infection in the city of Rio de Janeiro. Brazil: Evandro Chagas Hospital AIDS Clinical Research Group. J Acquir Immune Defic Syndr Hum Retrovirol 18: 488-494
52. Barreto CC, Nishyia A, Araujo LV, Ferreira JE, Busch MP et al. (2006) Trends in antiretroviral drug resistance and clade distributions among HIV-1--infected blood donors in São Paulo, Brazil. J Acquir Immune Defic Syndr 41: 338-341.

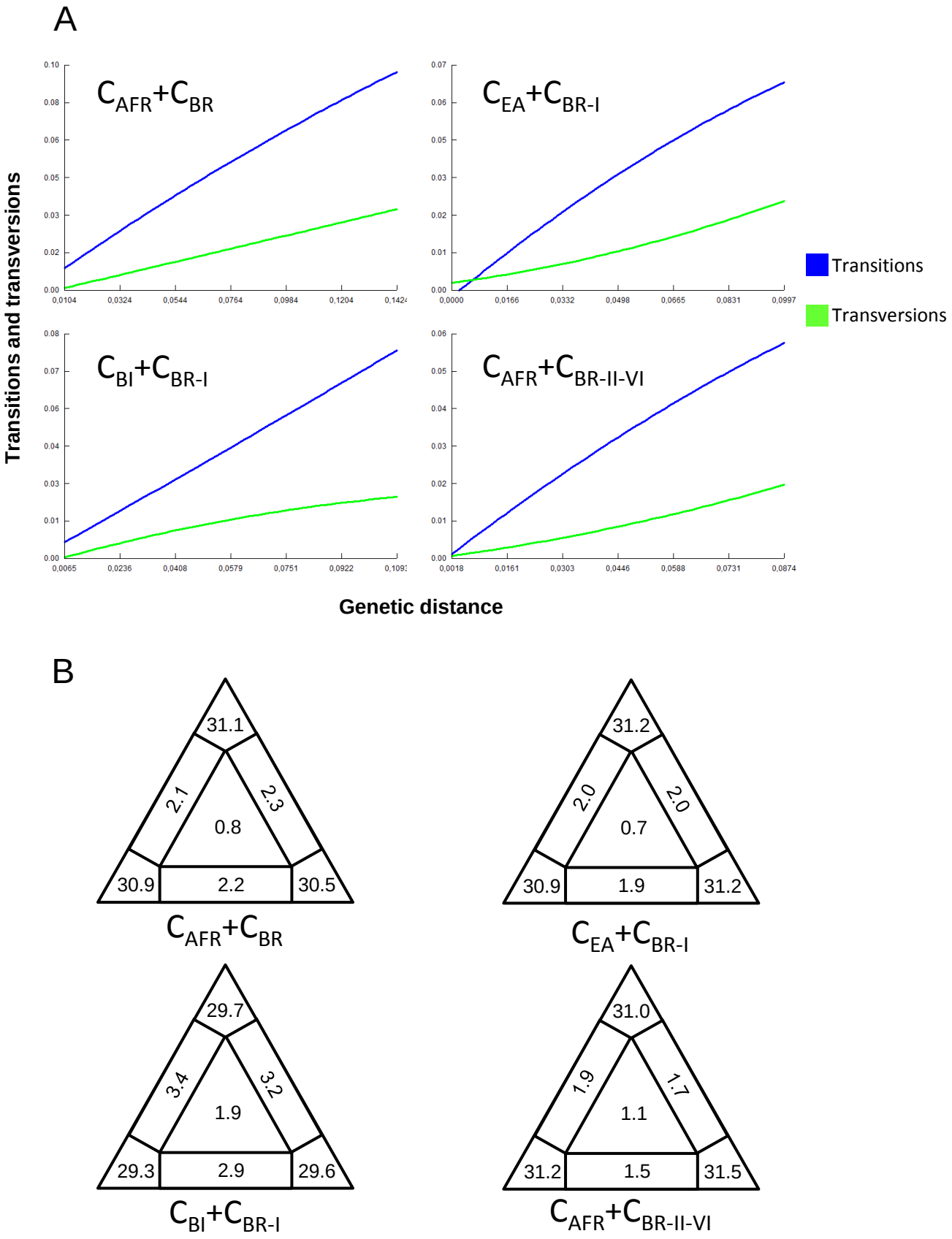

A
$C_{AFR}+C_{BR}$
$C_{EA}+C_{BR-I}$
$C_{BI}+C_{BR-I}$
$C_{AFR}+C_{BR-II-VI}$
Transitions and transversions
Genetic distance
Transitions
Transversions
B
31.1
2.1
2.3
0.8
30.9
2.2
30.5
$C_{AFR}+C_{BR}$
31.2
2.0
2.0
0.7
30.9
1.9
31.2
$C_{EA}+C_{BR-I}$
29.7
3.4
3.2
1.9
29.3
2.9
29.6
$C_{BI}+C_{BR-I}$
31.0
1.9
1.7
1.1
31.2
1.5
31.5
$C_{AFR}+C_{BR-II-VI}$

Figure S1

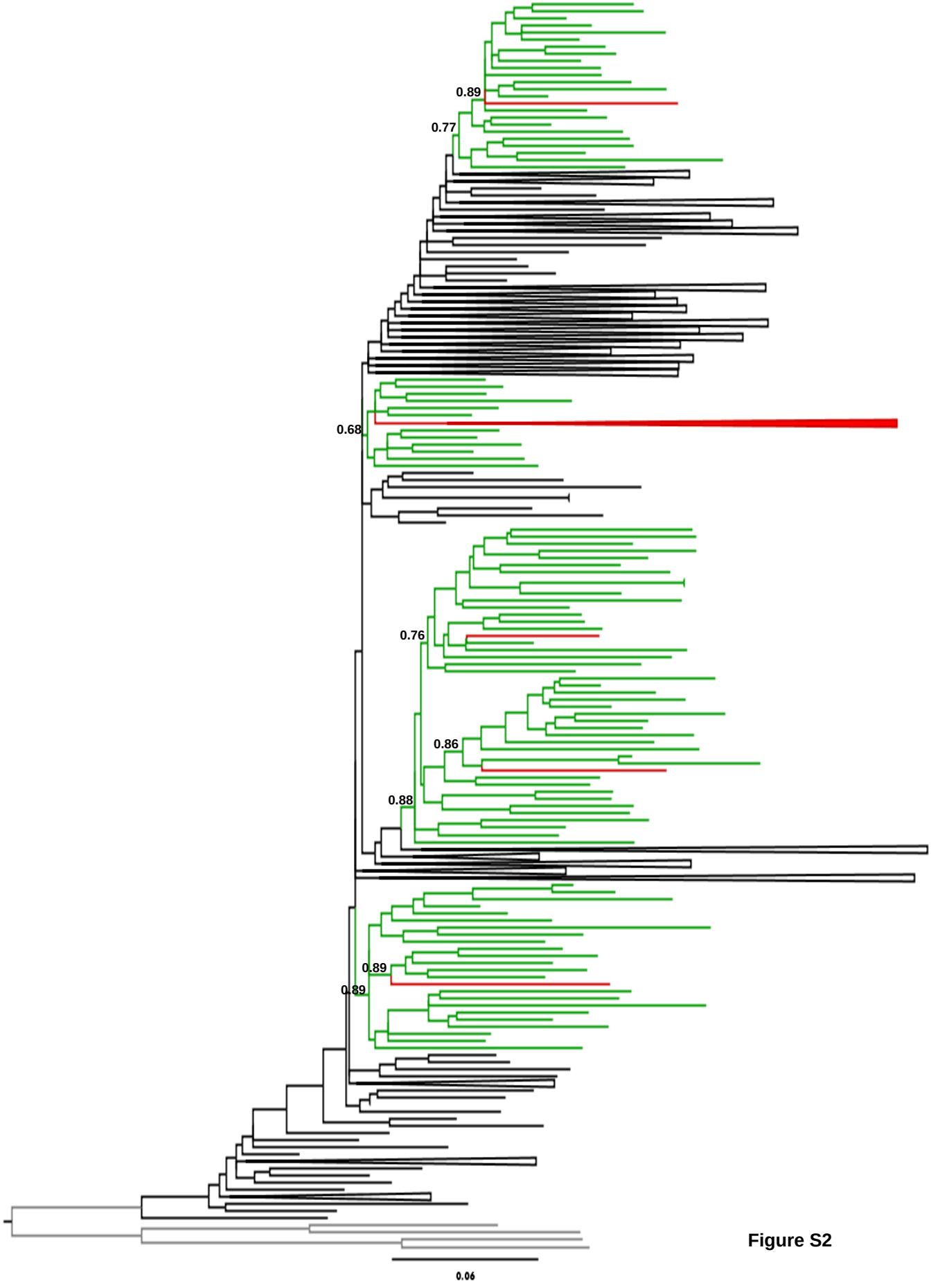


Figure S2

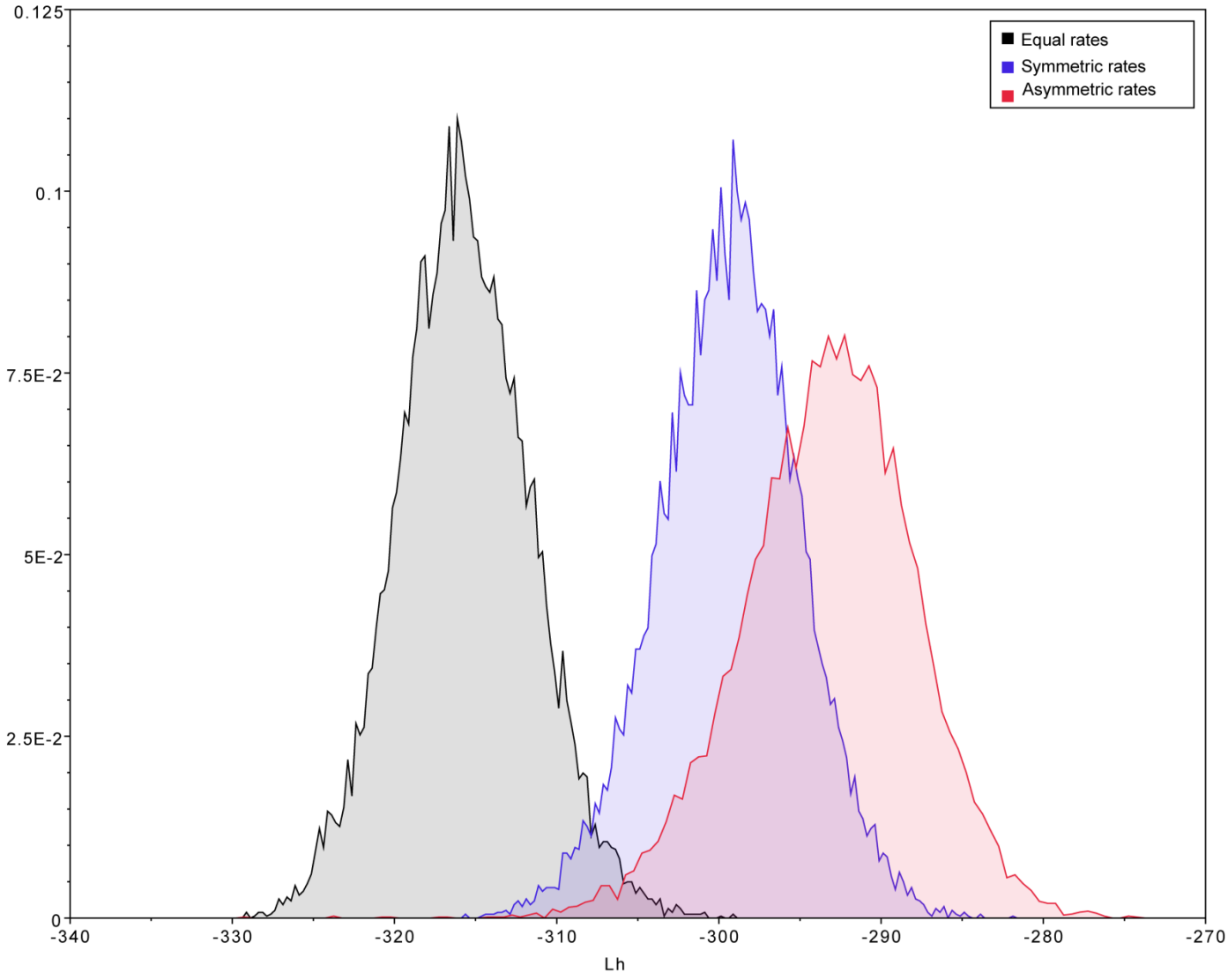

0.125
0.1
7.5E-2
5E-2
2.5E-2
0
-340
-330
-320
-310
-300
-290
-280
-270
Lh
Equal rates
Symmetric rates
Asymmetric rates

Figure S3

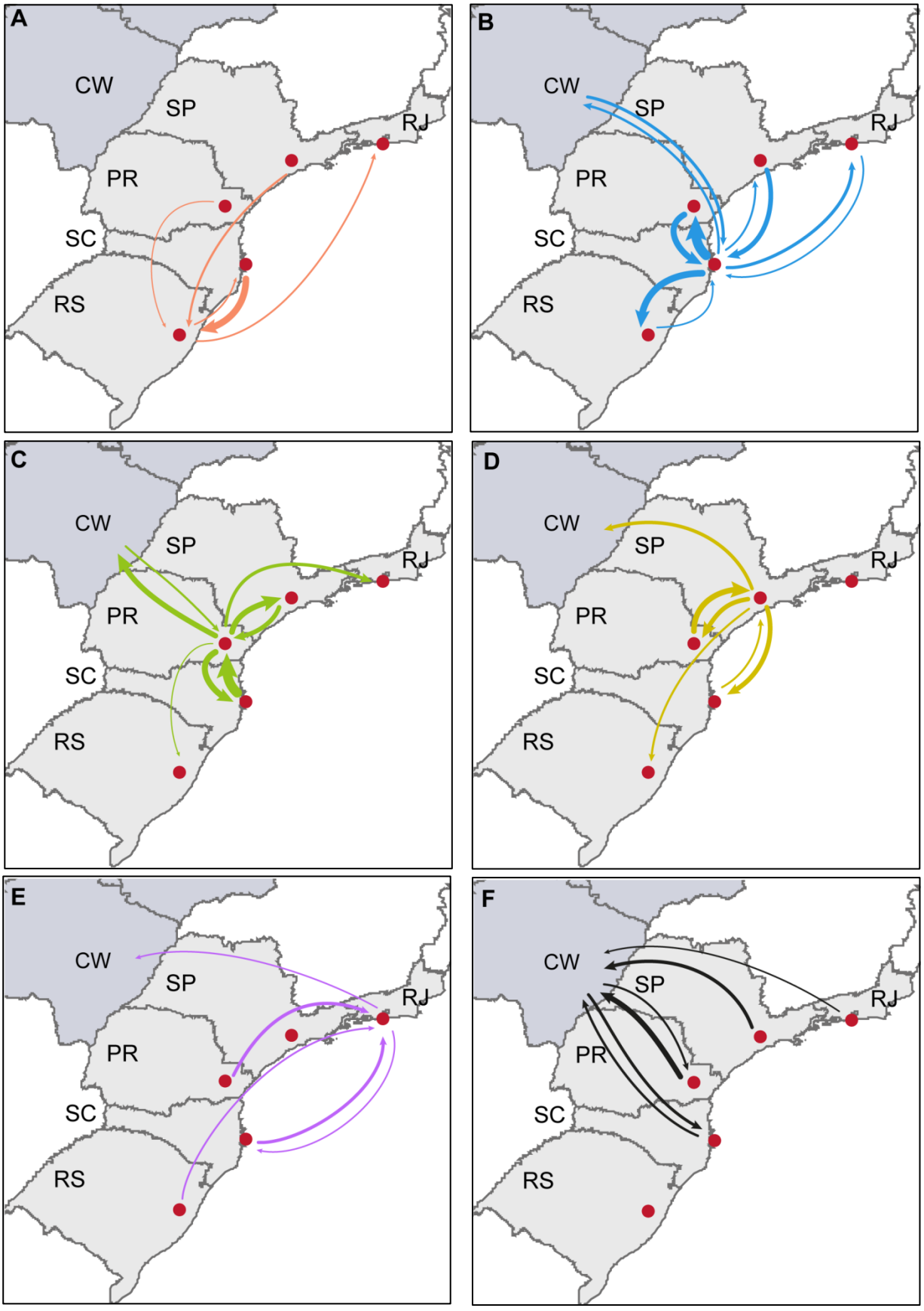

A
CW
SP
RJ
PR
SC
RS
B
CW
SP
RJ
PR
SC
RS
C
CW
SP
RJ
PR
SC
RS
D
CW
SP
RJ
PR
SC
RS
E
CW
SP
RJ
PR
SC
RS
F
CW
SP
RJ
PR
SC
RS

Figure S4

Table S1. HIV-1 C_{AFR+BR} dataset.

Region	Country/State	*N*	Sampling date
Brazil (South)	PR	39	2001-2007
	RS	55	1992-2006
	SC	41	2008-2009
Brazil (Southeast)	RJ	32	2003-2011
	SP	18	2004-2009
Brazil (Central-west)	GO/MT/MS	24	2003-2010
Central Africa	Angola	31	2001-2010
	Democratic Republic of Congo	22	2002-2007
Southern Africa	Botswana	70	2001
	Malawi	46	2002
	Mozambique	101	2002-2004
	South Africa	1,031	1999-2009
	Zambia	150	1998-2008
	Zimbabwe	178	2007
East Africa	Burundi	92	2002
	Ethiopia	82	1986-2003
	Kenya	39	1991-2007
	Tanzania	81	1997-2009
	Uganda	38	1990-2010

Table S2. HIV-1 $C_{EA+BR-I}$ and $C_{BI+BR-I}$ datasets.

Dataset	Region	Country/State	*N*	Sampling date
$C_{EA+BR-I}$	Brazil (South)	PR	10	2001-2006
		RS	10	1992-2006
		SC	10	2008-2009
	East Africa	Burundi	92	2002
		Ethiopia	47	2002-2003
		Kenya	24	2004-2007
		Tanzania	40	2005-2009
		Uganda	33	1990-2010
$C_{BI+BR-I}$	Brazil (South)	PR	39	2001-2007
		RS	55	1992-2006
		SC	41	2008-2009
	Brazil (Southeast)	RJ	27	2003-2011
		SP	18	2004-2009
	Brazil (Central-west)	GO/MT/MS	24	2003-2010
	East Africa	Burundi	10	2002

Table S3. HIV-1 $C_{AFR+BR-II-VI}$ dataset.

Region	Country/State	*N**	Sampling date
Brazil (Southeast)	RJ	5	2006-2011
Central Africa	Angola	2 (2/0)	2001-2010
East Africa	Burundi	9 (3/6)	2002
	Ethiopia	3 (0/3)	1996-2003
	Kenya	15 (14/1)	2004-2007
	Tanzania	2 (2/0)	2005
	Uganda	1 (1/0)	1997
Southern Africa	Botswana	4 (0/4)	2000-2001
	Malawi	4 (0/4)	2000-2009
	Mozambique	1 (1/0)	2002
	South Africa	32 (19/13)	2000-2009
	Zambia	15 (2/13)	1989-2008
	Zimbabwe	2 (2/0)	2007
Western Africa	Senegal	2 (0/2)	1990-2003
Asia	China	1 (0/1)	2004
	India	1 (0/1)	2000
Europe	Austria	1 (0/1)	2004
	Sweden	1 (0/1)	2004

*In parenthesis is indicated the number of non-Brazilian reference sequences selected using ML analysis/BLAST.

Table S4. Differences among harmonic mean of Likelihoods for three distinct transition models.

Model	ln P(model | data)	S.E.	Equal rates	Symmetric rates	Asymmetric rates
Equal rates	-321,307	+/- 0.078	-	-15,145	-18,754
Symmetric rates	-306,162	+/- 0.106	15,145	-	-3,609
Asymmetric rates	-302,553	+/- 0.152	18,754	3,609	-

Table S5. Viral transition rates, routine travels and road distances between localities.

Localities[a]	Transition rate (q)[b]	Routine travels[c] (people × trip)/1,000	Road distances (Km)[d]
RJ to SC	0.88	39	1144
PR to SC	3.06	618	300
SP to SC	2.43	1320	705
RS to SC	0.86	409	476
CW to SC	1.52	0	1493
SC to RJ	1.64	80	1144
PR to RJ	1.67	0	852
SP to RJ	0.46	171	429
RS to RJ	0.90	4	1553
CW to RJ	0.10	1	1338
SC to PR	4.82	267	300
RJ to PR	0.00	19	852
SP to PR	2.27	86	408
RS to PR	0.00	59	711
CW to PR	1.07	107	1186
SC to SP	1.05	158	705
RJ to SP	0.21	526	429
PR to SP	2.79	1223	408
RS to SP	0.41	147	1109
CW to SP	0.34	608	926
SC to RS	3.35	34	476
RJ to RS	0.00	45	1553
PR to RS	0.64	42	711
SP to RS	1.10	33	1109
CW to RS	0.00	0	1847
SC to CW	1.27	24	1493
RJ to CW	0.79	17	1338
PR to CW	2.70	102	1186
SP to CW	1.74	726	926
RS to CW	0.32	0	1847

[a] RS (Rio Grande do Sul), SC (Santa Catarina), PR (Paraná), RJ (Rio de Janeiro), SP (São Paulo) and CW (Central-west Region). [b] Mean transition rates (q) estimated for the posterior set of 18,000 trees using the APE package. [c] Routine traffic amongst states according to estimations of the Public Ministry of Tourism (http://www.dadosefatos.turismo.gov.br/export/sites/default/dadosefatos/demanda_turistica/domestica/downloads_domestica/Relatxrio_Executivo_Tur_Dom_2007.pdf) for the year 2007. [d] Road distances between state's capitals (the capital of Goiás state was used as reference for the central-west region).

3.3 Artigo 3

Título: *Reassessing the Origin of the HIV-1 CRF02_AG Lineages Circulating in Brazil*

Autores: Edson Delatorre, Carlos A. Velasco-De-Castro, Jose H. Pilotto, José C. Couto-Fernandez, Gonzalo Bello e Mariza G. Morgado

Publicado no periódico *AIDS Research and Human Retroviruses*

Resumo:

O CRF02_AG do HIV-1 é responsável por pelo menos 8% das infecções pelo HIV-1 mundialmente e está distribuído principalmente na África Ocidental. O CRF02_AG foi recentemente relatado em países onde não é nativo, incluindo o Brasil. Em um estudo prévio, que incluiu 10 amostras de CRF02_AG brasileiras, nós encontramos pelo menos quatro introduções independentes e duas redes de transmissão autóctones deste clado no Brasil. Como mais amostras de CRF02_AG foram identificadas no Brasil, nós realizamos uma nova análise filogeográfica utilizando um dataset maior que o anterior. Um total de 20 sequências brasileiras (18 do Rio de Janeiro e duas de São Paulo) e 1.485 sequências africanas do gene *pol* de CRF02_AG do HIV-1 forma analisadas utilizando máxima verossimilhança (MV). A árvore de MV demonstrou que as sequências brasileiras se distribuíram em cinco linhagens distintas. A análise filogeográfica bayesiana das sequências brasileiras e das sequências africanas mais intimamente relacionadas (*n* = 212) situaram a origem de todas as linhagens brasileiras na África Ocidental, provavelmente Gana, Senegal e Nigéria. Dois clados monofiléticos foram identificados, formados somente por sequências do Rio de Janeiro, e sua data de origem foi estimada por volta de 1985 (95% da maior densidade *a posteriori*: 1979-1992). Estes resultados suportam a existência de pelo menos cinco eventos independentes de introdução da linhagem CRF02_AG da África Ocidental no Brasil e também indicam que pelo menos duas destas linhagens tem se disseminado localmente no estado do Rio de Janeiro pelos últimos 30 anos.

AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 31, Number 00, 2015

DOI: 10.1089/aid.2015.0183

Reassessing the Origin of the HIV-1 CRF02_AG Lineages Circulating in Brazil

Edson Delatorre,[1] Carlos A. Velasco-De-Castro,[2] José H. Pilotto,[1,3] José Carlos Couto-Fernandez,[1] Gonzalo Bello,[1] and Mariza G. Morgado[1]

Abstract

HIV-1 CRF02_AG is responsible for at least 8% of the HIV-1 infections worldwide and is distributed mainly in West Africa. CRF02_AG has recently been reported in countries where it is not native, including Brazil. In a previous study including 10 CRF02_AG Brazilian samples, we found at least four independent introductions and two autochthonous transmission networks of this clade in Brazil. As more CRF02_AG samples have been identified in Brazil, we performed a new phylogeographic analysis using a larger dataset than before. A total of 20 Brazilian (18 from Rio de Janeiro and two from São Paulo) and 1,485 African HIV-1 CRF02_AG *pol* sequences were analyzed using maximum likelihood (ML). The ML tree showed that the Brazilian sequences were distributed in five different lineages. The Bayesian phylogeographic analysis of the Brazilian and their most closely related African sequences ($n = 212$) placed the origin of all Brazilian lineages in West Africa, probably Ghana, Senegal, and Nigeria. Two monophyletic clades were identified, comprising only sequences from Rio de Janeiro, and their date of origin was estimated at around 1985 (95% highest posterior density: 1979–1992). These results support the existence of at least five independent introductions of the CRF02_AG lineage from West Africa into Brazil and further indicate that at least two of these lineages have been locally disseminated in the Rio de Janeiro state over the past 30 years.

The human immunodeficiency virus type 1 (HIV-1) is classified into four groups (M, N, O, and P), each one representing independent zoonotic transmission events to humans.[1] The M group is responsible for the worldwide pandemic, having evolved and diversified into different clades,[2] characterized by an extensive genetic diversity and further subdivided into pure subtypes (A–D, F–H, J, and K), subsubtypes (A1–A4, F1–F2), 72 circulating recombinant forms (CRFs) and multiple unique recombinant forms (URFs) (Los Alamos HIV Sequence Database—www.hiv.lanl.gov/).

Among the CRFs, the CRF02_AG is the most prevalent worldwide and is responsible for at least 8% of the HIV-1 global infections.[3] This recombinant is distributed mainly in West Africa and, to a lesser extent, in the Middle East and North Africa. In recent years, an increase in the number of CRF02_AG-infected patients has been observed in Europe,[4–6] North America,[7,8] and South America,[9–12] probably due to migrations from African endemic regions. In Brazil, HIV-1 molecular epidemiology studies based in the *pol* gene have identified a high prevalence of HIV-1 subtype B (70–90%), followed by subsubtype F1 (5–15%) and subtype C (1–10%) in almost the entire country, with the exception of the southern region, where subtype C reaches high prevalences (25–66%).[13–20]

Despite the predominance of subtypes B, F1, and C and recombinants between them in the Brazilian HIV-1 epidemic, clade CRF02_AG has already been described in the states of Rio de Janeiro,[10–12,15,21] São Paulo,[22,23] Pará,[16] and Bahia,[24] with prevalences between 0.2% and 1.9%, indicating a limited spread of this CRF in the country. However, these studies were conducted with convenience samples that might not represent the prevalence of this subtype in Brazil.

In a previous attempt to trace the origin of some of those sequences, we have found that at least four introductions of this clade occurred in Brazil and that at least two CRF02_AG Brazilian lineages were successful in getting established and disseminated throughout the Rio de Janeiro state. The precise time and country of origin of those CRF02_AG lineages introduced in Brazil, however, remain unknown.[10] As more CRF02_AG samples have been identified in Brazil,[12,22] we

[1]Laboratório de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz–FIOCRUZ, Rio de Janeiro, Brazil.
[2]Laboratório de Virologia, Departamento de Patologia Clínica, Instituto Nacional de Saúde da Mulher, da Criança e do Adolescente Fernandes Figueira–FIOCRUZ, Rio de Janeiro, Brazil.
[3]Hospital Geral de Nova Iguaçu, Rio de Janeiro, Brazil.

aimed to improve our knowledge about the spread of this CRF infection through the country by using robust Bayesian phylogeographic analysis to estimate the country of origin and date of introduction of Brazilian CRF02_AG lineages.

In this study, all available HIV-1 CRF02_AG *pol* sequences previously reported in Brazil (n = 19, 17 from Rio de Janeiro and two from São Paulo)[10,12,21,25,26] were combined with one new CRF02_AG sequence obtained from a sample collected from one antiretroviral (ARV)-naive pregnant woman recruited at the Hospital Geral de Nova Iguaçu (HGNI), Rio de Janeiro. This sample was amplified and sequenced as described elsewhere.[27] Brazilian sequences were combined with all HIV-1 CRF02_AG *pol* sequences from Sub-Saharan Africa available in the Los Alamos HIV Sequence Database (www.hiv.lanl.gov) that covered the entire protease and partial reverse transcriptase (PR/RT) regions (nt positions 2253–3272 relative to the HIV-1 HXB2 clone) and had a known sampling year. Sub-Saharan Africa has an estimated CRF02_AG prevalence of >7% and as the prevalence of this clade outside this region is very low,[3] sequences from other countries were not included in the study. The subtype assignment of all sequences was confirmed by the REGA HIV subtyping tool v.2,[28] neighbor-joining (NJ) phylogenetic analysis, and bootscanning analysis. The NJ phylogeny with HIV-1 group M subtype reference sequences was constructed with MEGA 5.0 software[29] under the Tamura–Nei nucleotide substitution model.

Bootscanning analyses were performed with the Kimura two-parameter model, within a 250 bp window moving in steps of 10 bases, using SimPlot 3.5.1 software.[30] Sequences with an incorrect CRF02_AG classification and multiple sequences from the same individual were removed, as well as sequences with 100% of identity recognized with the CD-HIT suite online web server.[31] These procedures resulted in a final dataset composed by 1,505 HIV-1 CRF02_AG *pol* sequences (AFR-BR-I) sampled from 17 African countries and Brazil (Table 1).

The phylogenetic signal of all datasets was evaluated by substitution saturation analysis, plotting the observed number of transitions and transversions against genetic distance for each pairwise comparison using DAMBE 5.3 software[32] and with the likelihood mapping method[33] by analyzing 10,000 random quartets using TREE-PUZZLE 5.2 software[34] in the web platform Mobyle@Pasteur (http://mobyle.pasteur.fr/). Maximum likelihood (ML) phylogenetic tree of the AFR-BR-I dataset was inferred under the $GTR + I + \Gamma_4$ nucleotide substitution model, selected using the jModeltest program.[35] The ML tree was reconstructed with PhyML 3.0 software[36] using an online web server[37] and visualized in FigTree 1.4 software.[38] A heuristic tree search was performed using the SPR branch-swapping algorithm and the reliability of the obtained topology was estimated with the approximate likelihood ratio test (aLRT)[39] based on the Shimodaira–Hasegawa-like procedure.

A second alignment (AFR-BR-II) was used for spatio-temporal reconstruction (Table 1). This alignment included all Brazilian sequences and the African sequences more closely related to the Brazilian ones, branching with high support (aLRT > 0.8) up to the second ancestral node in the ML phylogenetic tree. The evolutionary rate (μ, nucleotide substitutions per site per year, subst./site/year), the time of the most recent common ancestor (T_{MRCA}, years), and the geographic transitions of the AFR-BR-II dataset were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST 1.8[40,41] with BEAGLE to increase computational speed.[42] Analyses were performed using the $GTR + I + \Gamma_4$ nucleotide substitution model, the uncorrelated relaxed lognormal molecular clock model,[43] and the nonparametric Bayesian Skyline coalescent tree prior as a coalescent demographic model.[44] Migration

Table 1. HIV-1 CRF02_AG *pol* Sequences Dataset Composition

		AFR-BR-I (ML)		*AFR-BR-II (Bayesian)*		
Region	*Country*	N	*%*	N	*Sampling date*	*%*
South America	Brazil	20	—	19[a]	1998–2012	—
Central Africa	Angola	1	0.1	0	—	0.0
	DRC[b]	8	0.5	0	—	0.0
West-Central Africa	Cameroon	538	36.2	22	1998–2009	10.4
	Equatorial Guinea	21	1.4	1	2008	0.5
	Gabon	57	3.8	2	2008	0.9
West Africa	Benin	93	6.3	10	2004–2009	4.7
	Burkina Faso	79	5.3	7	2003–2006	3.3
	Cape Verde	41	2.8	16	2010–2011	7.5
	Côte d'Ivoire	2	0.1	0	—	0.0
	Gambia	1	0.1	0	—	0.0
	Ghana	184	12.4	43	1993–2007	20.3
	Guinea	1	0.1	1	2007	0.5
	Guinea-Bissau	5	0.3	2	2004–2005	0.9
	Liberia	1	0.1	0	—	0.0
	Nigeria	165	11.1	38	1999–2009	17.9
	Senegal	233	15.7	61	1998–2011	28.8
	Togo	55	3.7	9	2006–2008	4.2

[a]The shortest sequence (KF922174) was excluded to avoid compromising the Bayesian estimates.
[b]The Democratic Republic of Congo.

events throughout the phylogenetic history were identified using a reversible discrete Bayesian phylogeographic model[45] with a continuous-time Markov chain rate reference prior.[46] MCMC chains were run for 2.5×10^8 generations and adequate chain mixing was checked by calculating the effective sample size (ESS) after excluding 10% burn-in using the TRACER 1.6 program.[47] Maximum clade credibility (MCC) trees were summarized from the posterior set of trees with TreeAnnotator 1.8[40] and visualized with FigTree 1.4 software.[38]

Transitions and transversions versus divergence analysis and likelihood mapping indicated that both AFR-BR-I and AFR-BR-II datasets retained enough phylogenetic signals for consistent phylogenetic and molecular clock inferences (Supplementary Fig. S1; Supplementary Data are available online at www.liebertpub.com/aid). The ML phylogenetic analysis of the AFR-BR-I dataset revealed that the 20 Brazilian CRF02_AG sequences formed two monophyletic clusters, consisting of 11 (BR-I) and six (BR-II) sequences. The remaining (three sequences) were sparsely distributed in the ML tree, intermixed among sequences of African origin (Fig. 1). We detected at least three independent introductions of the CRF02_AG clade in Rio de Janeiro (BR-I, BR-II and 06BRRJ34) and two in São Paulo (BR09SP371 and SP3686), but only the introductions into Rio de Janeiro (BR-I and BR-II) seem to have been successfully disseminated locally.

The lineage BR-I formed a highly supported clade (aLRT = 0.93) composed by 11 sequences sampled between 2004 and 2012 from patients living in the state of Rio de Janeiro. This lineage was described in our previous work as being composed of three sequences isolated between 2004 and 2010,[10] thus representing a four-times size increase of this lineage size since our previous work. It is noteworthy that seven of the new sequences were isolated between 2005 and 2007 from patients classified as recent HIV-1 seroconverters,[12] whereas the remaining one was isolated in 2012 from a 29-year-old pregnant woman diagnosed in the same year. The presence of HIV-1 recent seroconverters indicates the occurrence of the local spread of this CRF02_AG lineage in the Rio de Janeiro metropolitan region, at least in the past decade, probably through heterosexual contacts.

The lineage BR-II was composed of six sequences, all isolated from patients from Rio de Janeiro state that clustered together with high support (aLRT = 0.85). This CRF02_AG lineage was also described in our previous work as being composed of five sequences isolated between 2006 and 2011.[10] The new sequence identified within this cluster was isolated in 2002 from a patient attending the Army Health Service in Rio de Janeiro[21] and does not represent a recent transmission event. The remaining CRF02-AG Brazilian sequences (06BRRJ34, BR09SP371, and SP3686) did not form clusters. The sequences 06BRRJ34 and BR09SP371 were already described in our previous work[10] whereas sequence SP3686 corresponds to an CRF02_AG transmission event previously unidentified.

Sequence BR09SP371 was isolated in 2009 from an ARV-naive adult patient newly diagnosed in the state of São Paulo,[26] sequence 06BRRJ34 was isolated in 2006 from a woman from the city of Rio de Janeiro,[10] and sequence SP3686 was isolated in 1998 from a blood donor at the Blood Center of São Paulo.[25]

The Brazilian HIV-1 CRF02_AG sequences were combined with their most closely related African sequences in the ML phylogenetic tree. The lineages BR-I and BR-II and sequence 06BRRJ34 were nested within a highly supported CRF02_AG monophyletic clade (aLRT = 0.92), mainly composed of sequences from Senegal (31.8%), Ghana (23.5%), and Cameroon (11.7%). Sequence BR09SP371 was nested within a cluster with high support (aLRT = 0.87) composed mostly of sequences from Cape Verde (57.1%). Sequence SP3686 was located inside a relatively well-supported cluster (aLRT = 0.88) mostly composed of sequences from Nigeria (82.1%) (Fig. 1).

Comparing the frequencies of each African country in the complete (AFR-BR-I) and Brazilian-related (AFR-BR-II) datasets, an increase in the frequency of some West African countries including Ghana (from 12.4% to 20.3%), Nigeria (from 11.1% to 17.9%), Senegal (from 23.3% to 28.8%), and Cape Verde (from 2.8% to 7.4%) was noted as well as a reduction in the frequency of some West-Central African countries, mainly Cameroon (from 36.2% to 10.4%) and Gabon (from 3.8% to 0.9%) (Table 1).

The geographic origin and time scale of Brazilian HIV-1 CRF02_AG strains were inferred from the AFR-BR-II dataset using a Bayesian phylogeographic framework. The median evolutionary rate of the AFR-BR-II *pol* dataset, estimated under a chronological time scale employing the dates of the sequences, was 1.9×10^{-3} [95% highest posterior density (HPD): $1.4 \times 10^{-3} - 2.4 \times 10^{-3}$] substitutions/site/year, consistent with previous estimations made from this CRF.[48] The estimated coefficient of rate variation in this dataset was 0.34 (HPD: 0.28–0.39), thus supporting a significant variation of substitution rate among branches and the use of a relaxed molecular clock model. The BR-I and BR-II lineages formed highly supported monophyletic clades [posterior probability (PP) = 1] in the reconstructed Bayesian phylogeny and appeared to have originated in Ghana [posterior state probability (PSP) = 1 and PSP = 0.94, respectively] at around 1985 (HPD: 1979–1992). The sequence 06BRRJ34 probably originated in Ghana (PSP = 0.67) or Benin (PSP = 0.27), the sequence BR09SP371 in Senegal (PSP = 0.62) or Guinea-Bissau (PSP = 0.23), and the sequence SP3686in Nigeria (PSP = 1) (Fig. 2).

These results confirm that HIV-1 CRF02_AG strains circulating in Brazil are closely related to those circulating in western African countries from the Bight of Benin (Ghana, Nigeria, and Benin), Senegal, and Guinea-Bissau, where this CRF is highly prevalent.[3] Economic and cultural relationships between these African countries and Brazil are quite restricted, with the exceptions of Nigeria, for which Brazil is an important trading partner (DESA/UNSD, 2013), and Guinea-Bissau, which shares the Portuguese colonization and official language with Brazil.

West African countries have a long history of population migration, internally and externally, traditionally following old colonial linkages.[49] In Ghana and Nigeria, the emigration rate intensified in the 1970s and 1980s, as economic and political conditions worsened,[50] which may have contributed to the spread of HIV-1 CRF02_AG outside these countries, possibly also to Brazil. Thus, the introduction of CRF02_AG strains into Brazil seems to be the result of sporadic events, rather than the consequence of the continuous mobility of people between Brazil and West Africa.

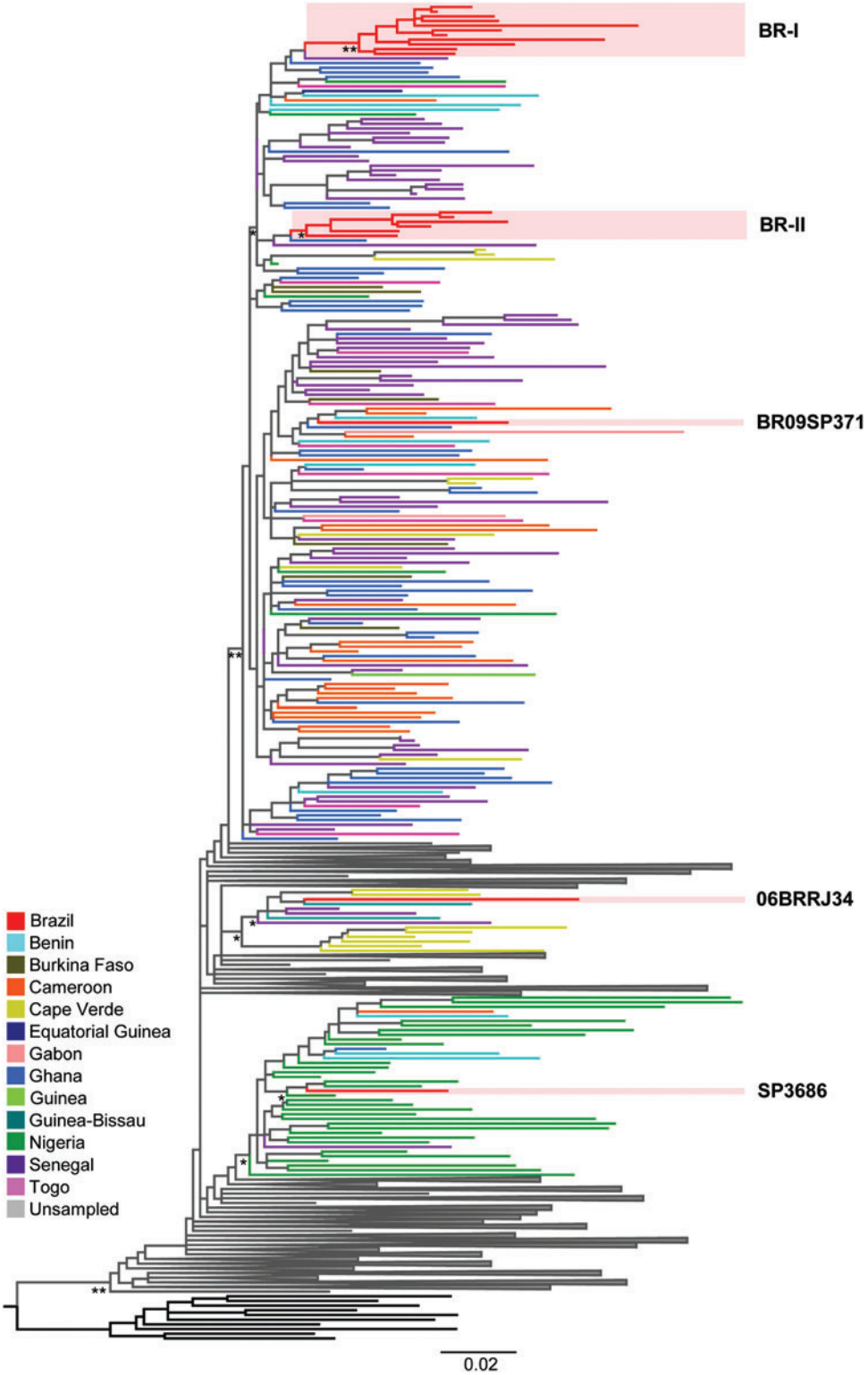


FIG. 1. Maximum likelihood (ML) tree of HIV-1 CRF02_AG *pol* (~1,000 pb) sequences from Brazil ($n = 20$) and African countries ($n = 1,485$). The tree was rooted using HIV-1 subtype G sequences (*black branches*). The branch colors represent the geographic origin of the sequences, according to the legend. The *red boxes* highlight the position of the Brazilian CRF02_AG sequences. For visual clarity, some African clades were collapsed in triangles and only the clades most closely related to the Brazilian lineages and used in the subsequent analysis are shown. The sequences not used in the phylogeographic analysis are in *gray*. *Asterisks* point to key nodes with aLTR support values between 0.80 and 0.9 (*) and >0.9 (**). Horizontal branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site. Color images available online at www.liebertpub.com/aid

However, it is important to note that immigration of Africans to Brazil has grown in recent years, mainly from Central and West African countries.[51] These countries have HIV-1 prevalence rates among adults ranging from 1% to 5%[52] and this migration increase can contribute to the future introduction and establishment of African HIV-1 variants in Brazil.

It is necessary to consider the sampling bias caused by the unequal number of sequences and countries used in the phylogeographic reconstructions. Here, we included only sequences from Sub-Saharan Africa, a region that has a consolidated HIV-1 CRF02_AG epidemic, particularly Western Africa, where the prevalence of this clade is ~50%.[3] Outside this continent, reports of CRF02_AG are anecdotal,[3]

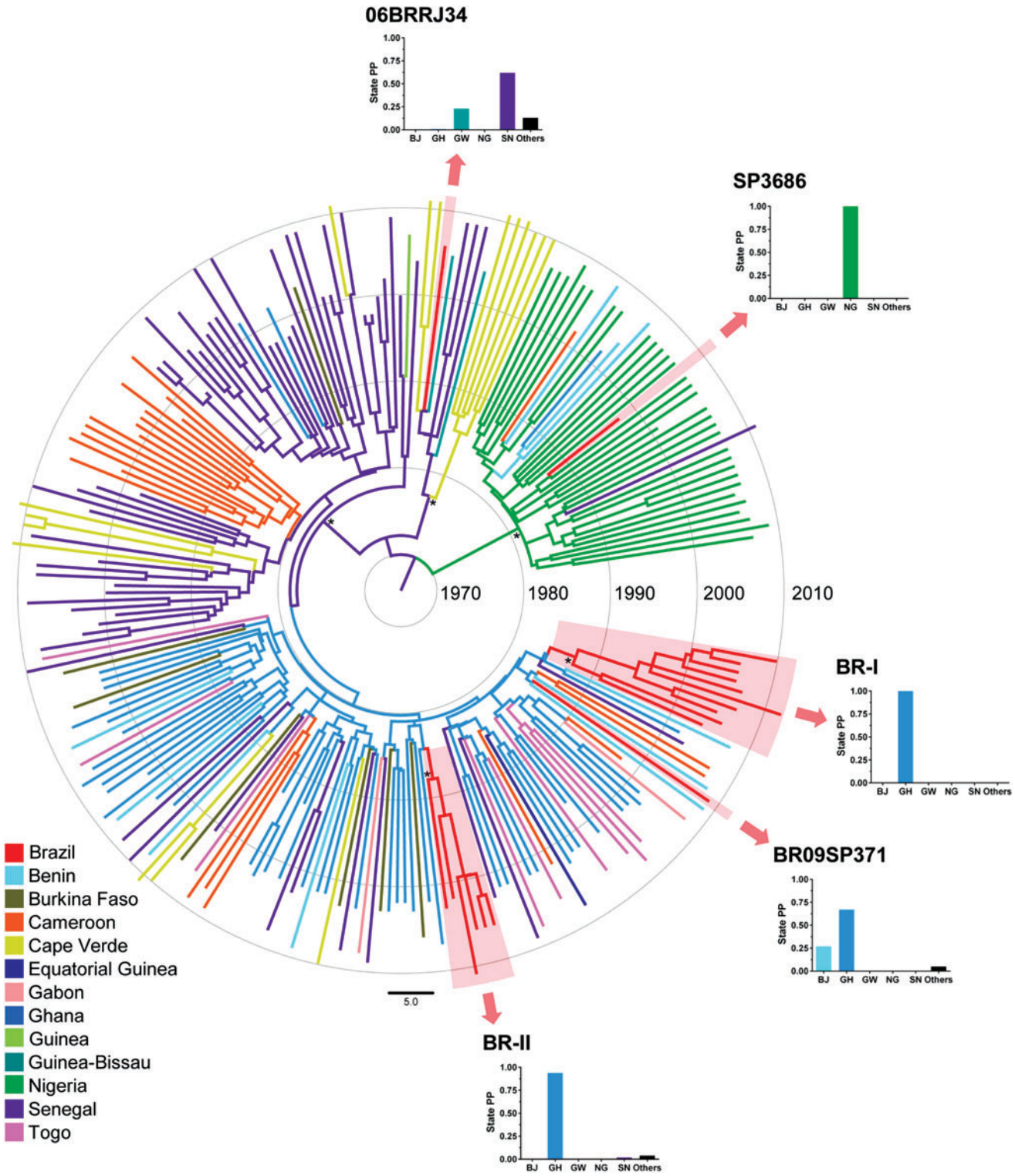


FIG. 2. Time-scaled Bayesian maximum clade credibility (MCC) tree of the HIV-1 CRF02_AG Brazilian strains and the most closely related African sequences ($n = 212$). Branches are colored according to the most probable location state of their descendent nodes as indicated in the legend. The *red boxes* highlight the position of the Brazilian HIV-1 CRF02_AG strains. *Asterisks* point to key nodes with high posterior probability support ($PP > 0.9$). The tree was automatically rooted under the assumption of a relaxed molecular clock and the branch lengths are drawn to scale with dates indicated in the concentric circles. The scale bar at the bottom indicates years. The posterior state probability (*PSP*) distributions at the first ancestral nodes of the Brazilian CRF02_AG strains at the Bayesian MCC tree are indicated in the graphics. Countries represented are BJ (Benin), GH (Ghana), GW (Guinea-Bissau), NG (Nigeria), SN (Senegal), and ''others'' for countries with minor contributions. Color images available online at www.liebertpub.com/aid

except in Europe, where an increasing prevalence of CRF02_AG among newly diagnosed HIV-1 infections was reported.[53] However, the estimated prevalence is <5% and this proportion is mostly due to immigrants originating from Africa.[53] This study clearly demonstrates the dissemination of CRF02_AG lineages of Western African countries to Brazil; however, the exact route of migration of these lineages remains unclear. It is not possible to exclude the hypothesis of Europe acting as a staging post in the dissemination of CRF02_AG to Brazil; however, the low prevalence of this HIV-1 clade in Europe makes this scenario unlikely.

Dated phylogeny made it possible to estimate 1985 (HPD: 1979–1992) as the T_{MRCA} of both BR-I and BR-II lineages, thus suggesting that the HIV-1 CRF02_AG strains started to circulate in Brazil around 10 years later than the other more prevalent HIV-1 subtypes C and F1.[54–56] Although the conditions of introduction of subtypes C, F1, and CRF02_AG in Brazil may have been similar, involving a single or a few related strains of African origin,[54,57–59] the relatively late spread of these CRF02_AG lineages may have limited the dissemination of this viral clade, possibly because it was introduced when the HIV-1 epidemic in Brazil started to stabilize.[60] Variation in the spread of different HIV-1 clades may also be explained by differences in the efficiency of transmission networks that promoted the initial dissemination of these lineages.

HIV-1 subtypes C and F1 may have been introduced and initially disseminated through highly connected networks of injection drug users.[61–63] The CRF02_AG clade, however, seems to have been introduced and mostly disseminated in poorly connected networks primarily involving sexual transmissions, thus resulting in a limited propagation.

The results presented here add a body of evidence to support the existence of at least five independent introductions of the CRF02_AG lineage from West Africa into Brazil and further indicate that at least two of these lineages have been disseminated in the Rio de Janeiro state for about 30 years, reinforcing the important role of this state in the introduction of new HIV-1 strains of African origin. Previous studies have also described the presence of HIV-1 lineages of African origin in the Rio de Janeiro state, such as HIV-1 subtype D variants, probably originating from South Africa,[64] and five HIV-1 subtype C lineages, probably imported from different eastern and southern African countries.[54] In addition, the periodic isolation of typical African HIV-1 subtypes A, D, and CRF02_AG in Rio de Janeiro reinforces the hypothesis that these lineages are being introduced and becoming established in this state, albeit in a minority form.[10,11,65–68] Altogether, the continuous surveillance of HIV-1 genetic diversity is warranted and fundamental to the early detection of the introduction and dissemination of newly emerging viral clades in the Brazilian epidemic, which may expose unknown transmission networks.

Acknowledgments

E.D. is supported by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). We thank Dr. Vera Bongertz for the English review of the manuscript.

Author Disclosure Statement

No competing financial interests exist.

References

1. Sharp PM and Hahn BH: Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med 2011;1:a006841.
2. Hemelaar J: The origin and diversity of the HIV-1 pandemic. Trends Mol Med 2012;18:182–192.
3. Hemelaar J, Gouws E, Ghys PD, and Osmanov S: Global trends in molecular epidemiology of HIV-1 during 2000–2007. AIDS 2011;25:679–689.
4. De Felipe B, Pérez-Romero P, Abad-Fernández M, *et al.:* Prevalence and resistance mutations of non-B HIV-1 subtypes among immigrants in Southern Spain along the decade 2000–2010. Virol J 2011;8:416.
5. Castro E, Khonkarly M, Ciuffreda D, *et al.:* HIV-1 drug resistance transmission networks in southwest Switzerland. AIDS Res Hum Retroviruses 2010;26:1233–1238.
6. Tramuto F, Maida CM, Bonura F, *et al.:* Dynamics and molecular evolution of HIV-1 strains in Sicily among antiretroviral naïve patients. Infect Genet Evol 2013;16:290–297.
7. Brennan C, Yamaguchi J, Devare SG, *et al.:* Expanded evaluation of blood donors in the United States for human immunodeficiency virus type 1 non-B subtypes and antiretroviral drug-resistant strains: 2005 through 2007. Transfusion 2010;50:2707–2712.
8. Chan P, Reitsma MB, Delong A, *et al.:* Phylogenetic and geospatial evaluation of HIV-1 subtype diversity at the largest HIV center in Rhode Island. Infect Genet Evol 2014;28:358–366.
9. Carrion G, Hierholzer J, Montano S, *et al.:* Circulating recombinant form CRF02_AG in South America. AIDS Res Hum Retroviruses 2003;19:329–332.
10. Delatorre EO, Bello G, Eyer-Silva WA, *et al.:* Evidence of multiple introductions and autochthonous transmission of the HIV type 1 CRF02_AG clade in Brazil. AIDS Res Hum Retroviruses 2012;28:1369–1372.
11. Eyer-Silva W and Morgado MG: Autochthonous horizontal transmission of a CRF02_AG strain revealed by a human immunodeficiency virus type 1 diversity survey in a small city in inner state of Rio de Janeiro, Southeast Brazil. Mem Inst Oswaldo Cruz 2007;102:809–815.
12. Velasco-de-Castro CA, Grinsztejn B, Veloso VG, *et al.:* HIV-1 diversity and drug resistance mutations among people seeking HIV diagnosis in voluntary counseling and testing sites in Rio de Janeiro, Brazil. PLoS One 2014;9:e87622.
13. Silveira J, Santos A, and Martínez A: Heterosexual transmission of human immunodeficiency virus type 1 subtype C in southern Brazil. J Clin 2012;54:36–41.
14. Gräf T, Passaes CPB, Ferreira LGE, *et al.:* HIV-1 genetic diversity and drug resistance among treatment naïve patients from Southern Brazil: An association of HIV-1 subtypes with exposure categories. J Clin Virol 2011;51:186–191.
15. Couto-Fernandez JC, Silva-de-Jesus C, Veloso VG, *et al.:* Human immunodeficiency virus type 1 (HIV-1) genotyping in Rio de Janeiro, Brazil: Assessing subtype and drug-resistance associated mutations in HIV-1 infected individuals failing highly active antiretroviral therapy. Mem Inst Oswaldo Cruz 2005;100:73–78.
16. Machado LF, Ishak MOG, Vallinoto ACR, *et al.:* Molecular epidemiology of HIV type 1 in northern Brazil: Identification of subtypes C and D and the introduction of CRF02_AG in the Amazon region of Brazil. AIDS Res Hum Retroviruses 2009;25:961–966.

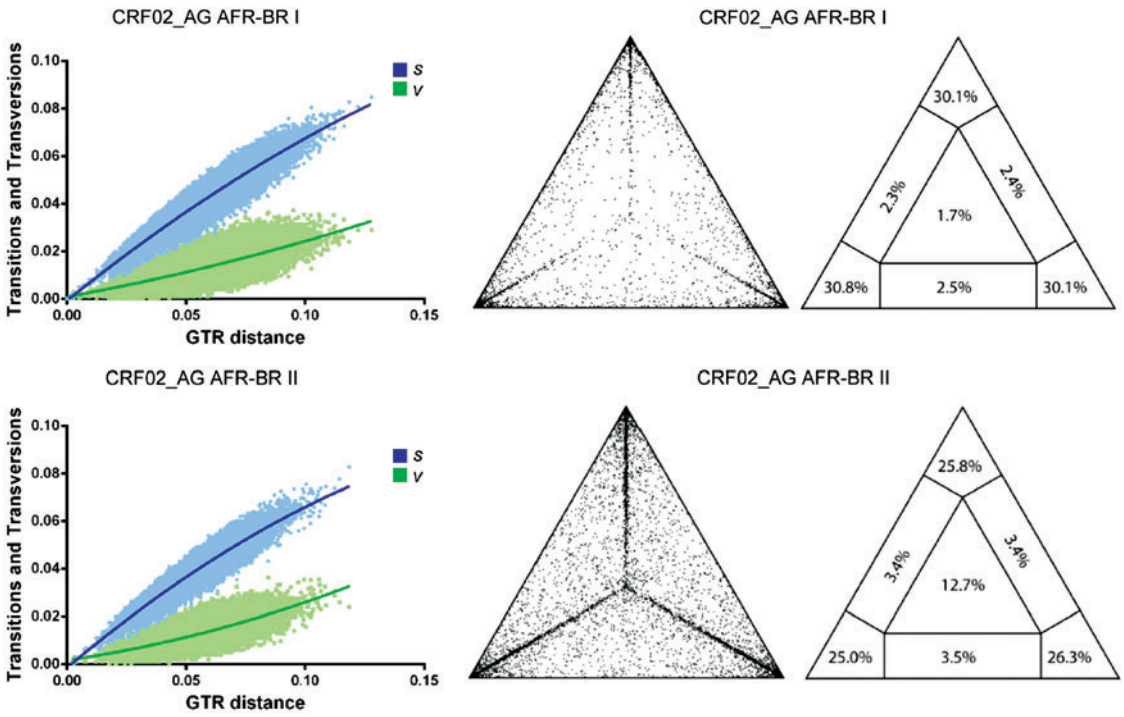
17. Gaspareto KV, Mello FMMDA, Dias JRC *et al.:* Genetic diversity and primary resistance among HIV-1-positive patients from Maringá, Paraná, Brazil. Rev Inst Med Trop Sao Paulo 2012;54:207–213.
18. López-Lopes GIS, Lança AM, de Paula Ferreira JL, *et al.:* Discrepancies of HIV-1 reverse transcriptase resistance interpretation of insertions and deletions between two genotypic algorithms. Intervirology 2013;56:217–223.
19. Alcântara KC, Reis MNG, Cardoso LPV, *et al.:* Increasing heterosexual transmission of HIV-1 subtype C in Inland Central Western Brazil. J Med Virol 2013;85:396–404.
20. Cavalcanti MS, Lacerda HR, De Brito AM, *et al.:* Antiretroviral resistance in individuals presenting therapeutic failure and subtypes of the human immunodeficiency virus type 1 in the Northeast Region of Brazil. Mem Inst Oswaldo Cruz 2007;102:785–792.
21. Pires IL, Soares MA, Speranza FAB, *et al.:* Prevalence of human immunodeficiency virus drug resistance mutations and subtypes in drug-naive, infected individuals in the army health service of Rio de Janeiro. J Clin 2004;42:426–430.
22. Ferreira JLP, Rodrigues R, Lança AM, *et al.:* Transmitted drug resistance among people living with HIV/Aids at major cities of Sao Paulo State, Brazil. Adv Virol 2013; 2013:878237.
23. Sanabani SS, Pastena ÉRDS, da Costa AC *et al.:* Characterization of partial and near full-length genomes of HIV-1 strains sampled from recently infected individuals in São Paulo, Brazil. PLoS One 2011;6:e25869.
24. Inocencio L, Pereira A, Sucupira MC, *et al.:* Brazilian Network for HIV Drug Resistance Surveillance: A survey of individuals recently diagnosed with HIV. J Int AIDS Soc 2009;12:1–6.
25. Barreto CC, Nishyia A, Araújo LV, *et al.:* Trends in antiretroviral drug resistance and clade distributions among HIV-1-infected blood donors in Sao Paulo, Brazil. J Acquir Immune Defic Syndr 2006;41:338–341.
26. Brígido LFM, Ferreira JLP, Almeida VC, *et al.:* Southern Brazil HIV type 1 C expansion into the state of São Paulo, Brazil. AIDS Res Hum Retroviruses 2011;27:339–344.
27. Afonso JM, Bello G, Guimarães ML, *et al.:* HIV-1 genetic diversity and transmitted drug resistance mutations among patients from the North, Central and South regions of Angola. PLoS One 2012;7:e42996.
28. De Oliveira T, Deforche K, Cassol S, *et al.:* An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 2005;21:3797–3800.
29. Tamura K, Peterson D, Peterson N, *et al.:* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011;28:2731–2739.
30. Lole KS, Bollinger RC, Paranjape RS, *et al.:* Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 1999;73:152–160.
31. Huang Y, Niu B, Gao Y, *et al.:* CD-HIT Suite: A web server for clustering and comparing biological sequences. Bioinformatics 2010;26:680–682.
32. Xia X: DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. Mol Biol Evol 2013;30:1720–1728.
33. Strimmer K and von Haeseler A: Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci USA 1997;94:6815–6819.
34. Schmidt H, Strimmer K, Vingron M, and von Haeseler A: TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 2002;18:502–504.
35. Posada D: jModelTest: Phylogenetic model averaging. Mol Biol Evol 2008;25:1253–1256.
36. Guindon S, Dufayard J-F, Lefort V, *et al.:* New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol 2010; 59:307–321.
37. Guindon S, Lethiec F, Duroux P, and Gascuel O: PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 2005;33:W557–559.
38. Rambaut A: FigTree v1.4. 2014.http://tree.bio.ed.ac.uk/software/figtree/.
39. Anisimova M and Gascuel O: Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol 2006;55:539–552.
40. Drummond AJ and Rambaut A: BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 2007;7:214.
41. Drummond AJ, Nicholls GK, Rodrigo AG, and Solomon W: Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 2002;161:1307–1320.
42. Suchard M and Rambaut A: Many-core algorithms for statistical phylogenetics. Bioinformatics 2009;25:1370–1376.
43. Drummond AJ, Ho SYW, Phillips MJ, and Rambaut A: Relaxed phylogenetics and dating with confidence. PLoS Biol 2006;4:e88.
44. Drummond J, Rambaut A, Shapiro B, and Pybus OG: Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 2005;22:1185–1192.
45. Lemey P, Rambaut A, Drummond AJ, and Suchard M: Bayesian phylogeography finds its roots. PLoS Comput Biol 2009;5:e1000520.
46. Ferreira MR and Suchard M: Bayesian analysis of elapsed times in continuous-time Markov chains. Can J Stat 2008; 36:355–368.
47. Rambaut A, Suchard M, and Drummond A: Tracer v1.6. 2013. http://tree.bio.ed.ac.uk/software/tracer/.
48. Faria NR, Suchard M, Abecasis A, *et al.:* Phylodynamics of the HIV-1 CRF02_AG clade in Cameroon. Infect Genet Evol 2012;12:453–460.
49. Gnisci D and Trémolières M: Migration. In: *Regional Atlas on West Africa, West African Studies* (Bossard L, Lebret M-C, Perret C, eds.). OECD Publishing, Paris, France, 2009, pp. 67–85.
50. Black R, Ammassari S, Mouillesseaux S, and Rajkotia R: Migration and pro-poor policy in West Africa. Brighton, 2004, www.migrationdrc.org. Accessed July 13, 2014.
51. UNHCR UNHC for R: UNHCR Statistical Online Population Database. 2015. www.unhcr.org/statistics/populationdatabase. Accessed April 8, 2015.
52. UNAIDS O: GLOBAL REPORT—UNAIDS report on the global AIDS epidemic 2013. Geneva, Switzerland, 2013.
53. Abecasis AB, Wensing AMJ, Paraskevis D, *et al.:* HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. Retrovirology 2013;10:7.
54. Delatorre E, Couto-Fernandez JC, Guimarães ML, *et al.:* Tracing the origin and northward dissemination dynamics of HIV-1 subtype C in Brazil. PLoS One 2013;8:e74072.

55. Bello G, Eyer-Silva W, Couto-Fernandez JC, *et al.:* Demographic history of HIV-1 subtypes B and F in Brazil. Infect Genet Evol 2007;7:263–270.
56. Bello G, Passaes CP, Guimarães ML, *et al.:* Origin and evolutionary history of HIV-1 subtype C in Brazil. AIDS 2008;22:1993–2000.
57. Bello G, Afonso JM, and Morgado MG: Phylodynamics of HIV-1 subtype F1 in Angola, Brazil and Romania. Infect Genet Evol 2012;12:1079–1086.
58. Véras NMC, Gray RR, Brígido LFDM, *et al.:* High-resolution phylogenetics and phylogeography of human immunodeficiency virus type 1 subtype C epidemic in South America. J Gen Virol 2011;92:1698–1709.
59. Fontella R, Soares M, and Schrago CG: On the origin of HIV-1 subtype C in South America. AIDS 2008;22:2001–2011.
60. Ministério da Saúde: Boletim Epidemiológico–Aids e DST, Brasília, DF, Brasil, 2014, www.aids.gov.br.
61. Almeida SE, de Medeiros RM, Junqueira DM, *et al.:* Temporal dynamics of HIV-1 circulating subtypes in distinct exposure categories in southern Brazil. Virol J 2012;9:306.
62. Teixeira SLM, Bastos FI, Telles PR, *et al.:* HIV-1 infection among injection and ex-injection drug users from Rio de Janeiro, Brazil: Prevalence, estimated incidence and genetic diversity. J Clin Virol 2004;31:221–226.
63. Guimarães ML, Bastos FI, Telles PR, *et al.:* Retrovirus infections in a sample of injecting drug users in Rio de Janeiro City, Brazil: Prevalence of HIV-1 subtypes, and co-infection with HTLV-I/II. J Clin Virol 2001;21:143–151.
64. Couto-Fernandez JC, Eyer-Silva W, Guimarães ML, *et al.:* Phylogenetic analysis of Brazilian HIV type 1 subtype D strains: Tracing the origin of this subtype in Brazil. AIDS Res Hum Retroviruses 2006;22:207–211.
65. Caride E, Brindeiro R, Hertogs K, Larder B, *et al.:* Drug-resistant reverse transcriptase genotyping and phenotyping of B and non-B subtypes (F and A) of human immunodeficiency virus type I found in Brazilian patients failing HAART. Virology 2000;275:107–115.
66. Pimentel VF, Morgado MG, Bello G, *et al.:* Temporal trends and molecular epidemiology of HIV type 1 infection in Rio de Janeiro, Brazil. AIDS Res Hum Retroviruses 2013;29:1553–1561.
67. Morgado MG, Guimarães ML, Gripp CB, *et al.:* Molecular epidemiology of HIV-1 in Brazil: High prevalence of HIV-1 subtype B and identification of an HIV-1 subtype D infection in the city of Rio de Janeiro, Brazil. Evandro Chagas Hospital AIDS Clinical Research Group. J Acquir Immune Defic Syndr Hum Retrovirol 1998;18: 488–494.
68. Alencar CS, Sabino EC, Carvalho SMF, *et al.:* HIV genotypes and primary drug resistance among HIV-seropositive blood donors in Brazil: Role of infected blood donors as sentinel populations for molecular surveillance of HIV. JAIDS 2013;63:387–392.

Address correspondence to:
Edson Delatorre
Laboratório de AIDS & Imunologia Molecular
Instituto Oswaldo Cruz–FIOCRUZ
Av. Brasil 4365
21045-900, Rio de Janeiro, RJ
Brazil

E-mail: edsonod@ioc.fiocruz.br/delatorre.ioc@gmail.com

Supplementary Data



SUPPLEMENTARY FIG. S1. Phylogenetic signal measurement of the datasets. Transitions (*s*, *blue line*) and transversions (*v*, *green line*) versus divergence plot for the HIV-1 CRF02_AG *pol* datasets AFR-BR I **(A)** and AFR-BR II **(B)**. Likelihood mapping diagram and percentage of dots plotted in each region of the map after likelihood mapping of 10,000 random quarters selected from the HIV-1 CRF02_AG *pol* datasets AFR-BR I **(C)** and AFR-BR II **(D)**. Each dot represents the probability of the three possible tree topologies for a set of four sequences (quartets) selected randomly from the dataset. The dots localized on the vertices, in the center, and on the laterals represent the tree-like, the star-like, and the network-like phylogenetic signals, respectively.

3.4 Artigo 4

Título: *Near full-length genomes and the date of introduction of a CRF45_cpx HIV-1 lineage identified in Brazil*

Autores: Edson Delatorre, Suwellen S.D. de Azevedo, Adriana Rodrigues-Pedro, Carlos A. Velasco-de-Castro, José C. Couto-Fernandez, Jose H. Pilotto e Mariza G. Morgado.

Resumo:

Pesquisas de epidemiologia molecular recentes identificaram algumas cepas do HIV-1 originadas provavelmente na África circulando no Brasil, incluindo a Forma Recombinante Circulante (CRF)45_cpx, um recombinante A1/K/U complexo que circula na África Central. Amostras recombinantes relacionadas ao CRF45_cpx identificadas em estudos independentes conduzidos com indivíduos HIV+ no Brasil tiveram seus genomas parciais e quase completos caracterizados e sua história evolutiva reconstruída. As sequências virais foram obtidas por amplificações sobrepostas seguidas de sequenciamento direto. Os perfis de recombinação foram determinados por análises filogenéticas e de *bootscanning*. A história evolutiva foi estimada por análises bayesianas utilizando datasets dos genes *gag*, *pol* e *env*. Seis das 10 amostras isoladas no Rio de Janeiro mostraram um padrão similar ao CRF45_cpx ao longo de todo o genoma. As restantes foram classificadas como recombinantes de segunda-geração, exibindo os padrões de mosaico: CRF45_cpx/B/D/F1/U, CRF45_cpx/B/F1/U, CRF45_cpx/B/U e CRF45_cpx/F1. Todas as sequências CRF45_cpx brasileiras, exceto uma, formaram um clado monofilético (CRF45-BR), que parece ser o resultado de um único evento de introdução que se disseminou pelos estados do Rio de Janeiro, São Paulo e Minas Gerais e é relacionado a sequências da Argentina, Itália e Bélgica. As análises bayesianas indicaram datas de origem bastante semelhantes (~1984: 1976-1996) para os três genes. Estes resultados indicam que o clado CRF45-BR está circulando na região Sudeste por cerca de 30 anos, embora sua presença não tenha sido detectada até recentemente devido à sua baixa prevalência. Isso reforça a relevância dos dados de vigilância molecular em larga escala para identificar a emergência de novas variantes do HIV e seu impacto sobre as epidemias locais.

NEAR FULL-LENGTH GENOMES AND THE DATE OF INTRODUCTION OF A CRF45_cpx HIV-1 LINEAGE IDENTIFIED IN BRAZIL

Edson Delatorre[*1], Suwellen S.D. de Azevedo[1], Adriana Rodrigues-Pedro[1], Carlos Augusto Velasco-de-Castro[2], José Carlos Couto-Fernandez[1], Jose H. Pilotto[1,3] and Mariza G. Morgado[1].

Running title: Origin of a Brazilian HIV-1 CRF45_cpx-like lineage

[1] Laboratório de AIDS & Imunologia Molecular. Instituto Oswaldo Cruz – FIOCRUZ. Rio de Janeiro, Brazil. [2] Laboratório de Virologia, Departamento de Patologia Clínica, Instituto Nacional de Saúde da Mulher, da Criança e do Adolescente Fernandes Figueira – FIOCRUZ, Rio de Janeiro, Brazil. [3] Hospital Geral de Nova Iguaçu, Rio de Janeiro, Brazil.

* Corresponding author: edsonod@ioc.fiocruz.br/delatorre.ioc@gmail.com.
Lab. de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz - FIOCRUZ. Av. Brasil 4365, 21045-900, Rio de Janeiro, RJ, Brazil. Phone: +55 21 3865 8154. Fax: +55 21 3865 8173.

WORD COUNT: 3885

ABSTRACT: 280

Keywords: HIV-1; CRF45_cpx; NFLG; Phylogeny; Brazil

ABSTRACT

The HIV-1 epidemiology has changed over the past decade toward a marked increase in the circulation of strains previously restricted to local epidemics. Recent molecular epidemiological surveys identified some HIV-1 strains of probable African origin circulating in Brazil, including the Circulating Recombinant Form (CRF) 45_cpx, a complex A1/K/U recombinant that circulates in Central Africa. Here, we characterize partial and near full-length genomes (NFLG) and reconstruct the evolutionary history of HIV-1 CRF45_cpx-related recombinant samples identified in independent studies carried out with HIV+ individuals in Brazil. The NFLG was obtained by overlapping PCR amplifications followed by direct sequencing. Recombination profiles were determined by phylogenetic and bootscaning analyses. The evolutionary history was estimated by a Bayesian coalescent-based method using datasets representing the *gag*, *pol* and *env* gene fragments. Six of the 10 samples isolated in Rio de Janeiro showed a CRF45_cpx-like pattern throughout the whole genome. The remaining were classified as second-generation recombinants, showing the mosaic patterns: CRF45_cpx/B/D/F1/U, CRF45_cpx/B/F1/U, CRF45_cpx/B/U and CRF45_cpx/F1. All Brazilian CRF45_cpx sequences, except one, formed a monophyletic clade (CRF45-BR), which seems to be the result of a single introduction event that has spread to the Rio de Janeiro, São Paulo and Minas Gerais states and is related to sequences from Argentina, Italy and Belgium. The Bayesian analyses pointed out quite consistent onset dates for CRF45-BR clade (~1984: 1976-1996) in the three gene datasets. These results indicate that the CRF45-BR clade has been circulating in the Southeastern Brazilian region for about 30 years, although its presence was not detected until recently due to its very low prevalence. This reinforces the relevance of large-scale molecular surveillance data to identify the emergence of new HIV variants and their impact on local epidemics.

INTRODUCTION

The human immunodeficiency virus type 1 (HIV-1) originated from multiple zoonotic transmission events of simian immunodeficiency virus from non-human primates to humans in Western-Central Africa in the first decades of the twentieth century [1]. One of these transmission events originated the M group of HIV-1, that shortly thereafter diversified into genetic subtypes (named A-D, F-H and J-K) and disseminated worldwide, being currently the responsible for the HIV pandemic [2]. The distinct chance of spread and establishment of lineages originated at the HIV-1 epicenter to other geographical regions resulted in the current global HIV-1 subtype distribution [3]. Even though this distribution has been broadly stable over the 2000–2007 period, there were dynamic changes in some regions, possibly due to several factors including population growth, increasing migrations and founder effects [4].

In some regions, the epidemiology of HIV-1 has changed over the past decade towards a marked increase in the circulation of non-B strains, such as Western Europe, for example, that although having a characteristic epidemic marked by the predominance of HIV-1 subtype B, increasing in the prevalence of other subtypes has been found, such as the subtype G and CRF02_AG in Spain [5], the subtype D and CRF01_AE in France [6] and subtype F1 in Italy [7]. This pattern is also observed in Brazil, where, despite the HIV-1 epidemic being dominated by subtypes B, C, F1 and recombinants between them, HIV-1 clades A, D and CRF02_AG were sporadically found [8–16]. In recent years, additional isolations of HIV-1 strains containing subtype K and/or unclassified (U) segments in the protease and reverse transcriptase (PR/RT) region of the *pol* gene were described in apparently not linked individuals from the Southeast region of Brazil, including vertically HIV-1-infected children in Rio de Janeiro [17] and Minas Gerais states [18], blood donors in São Paulo state [11,19], HIV-1-positive ARV-naive

pregnant women [20] and individuals seeking HIV diagnosis [21], these last ones both in the Rio de Janeiro state.

This pattern of recombination in PR/RT resembles that found in the HIV-1 complex circulant recombinant forms (CRF) 04_cpx [22], 09_cpx [23], 45_cpx [24] and 49_cpx [25]. CRF04_cpx is composed of fragments of subtypes A/G/H/K/U and has a circulation almost limited to Cyprus and Greece [22]. CRF09_cpx displays a complex A/G/J/K/U mosaic pattern and circulates in Central and West Africa [23]. The CRF45_cpx genome comprises subtypes A/K/U and circulates mainly in West-Central Africa [24]. CRF49_cpx circulates mainly in West Africa and has a genome composed of subtypes A/C/J/K/U [25]. With the exception of CRF04_cpx, already established in Europe, the other complex CRFs are rarely detected outside the African continent, with occasional detections in United Kingdom [26], Spain [27] , France [28], Germany [29], United States of America [30,31] and Brazil [19]. A Brazilian sequence first characterized as an A1/K recombinant, based only on the *pol* gene sequence [11], was reclassified as a CRF45_cpx after amplification of the viral near full-length genome (NFLG) [19], reinforcing the importance of increasing genomic length in order to improve HIV-1 subtype characterization.

As several HIV-1-recombinant strains carrying subtype K in the *pol* gene were identified in HIV-1-positive individuals followed in independent studies carried out in the Rio de Janeiro state in recent years, the objective of this study was to characterize partial and NFLGs sequences of these strains and to reconstruct their phylogenetic relationship and onset date using Maximum Likelihood and Bayesian coalescent-based approaches.

MATERIAL AND METHODS

Ethics statement

The sequences obtained in this study were retrieved from individuals recruited in the context of different research projects involving molecular analyzes. The patients enrolled in the research projects approved by the Evandro Chagas Clinical Research Institute Ethics Committee (CAAE 0040.1.009.000-4; 0032.0.009.000-04 and 0130.0.009.016-05) signed the informed consent or the signature was obtained from their relatives or guardians in the case of minor age individuals (<18 years old). The patients involved in the research project approved by the Oswaldo Cruz Institute Ethics Committee (CAAE 03925112.0.0000.5248) corresponded to HIV-1-infected individuals that underwent standard clinical HIV resistance genotyping tests. These patients do not signed informed consent and their data were analyzed anonymously.

Patient selection

From 2005 to 2014, 10 individuals infected with HIV-1 carrying mosaic genomes composed by subtypes A1, K and/or U in the protease and reverse transcriptase (PR/RT) region of the *pol* gene were recovered from different studies carried out by the Laboratory of AIDS and Molecular Immunology (FIOCRUZ) in the Rio de Janeiro state. The first recombinant sequence found (05BRRJLTS223) came from a HIV-1-positive male, classified as long-term seroconverter based on the BED capture enzyme immunoassay protocol, that had the blood sample collected in 2005 [21]. Two other recombinant sequences came from HIV-1-positive ARV-naive pregnant women, 07BRRJPG144 and 08BRRJPG226, whose samples were collected in 2007 and 2008, respectively [20]. Five patients corresponding to HIV-1 vertically infected children, had samples collected in 2007 (07BRRJPC38) and 2009 (09BRRJCR018, 09BRRJCR019, 09BRRJCR040, 09BRRJPC37). The samples 06BRRJRENK1 and 14BRRJRENK2, collected in 2006 and 2014, respectively, were from HIV-1-infected

individuals that underwent HIV resistance genotyping tests through the Brazilian Network for HIV-1 Genotyping (RENAGENO). Additional patient details are given in Table 1.

Amplification and characterization of partial and near full-length HIV-1 genomes

DNA or RNA samples were extracted from 200 µl of whole blood using QIAamp DNA blood Mini (Qiagen Inc., Valencia, CA) or 140 µl of plasma using QIAamp Viral RNA Mini kits (Qiagen Inc., Valencia, CA), respectively, according to the manufacturer's protocols. Partial and near full-length genomes were amplified by nested polymerase chain reaction (PCR) of overlapped fragments covering the HIV-1 genome and directly sequenced as previously described [32]. The sequences were submitted to GenBank, and have the accession numbers: KX273355 – KX273376.

To determine the subtype of partial and NFLG sequences characterized in this study, bootscaning analyses were performed using the Simplot version 3.5.1 [33]. Bootstrap values supporting branching with reference sequences were determined by the neighbor-joining (NJ) method using the Kimura-2-parameter distance model and 100 bootstrap replicates of 400bp sliding windows with increasing steps of 40bp. Reference sequences available in the Los Alamos HIV Database (http://www.hiv.lanl.gov/) of HIV-1 pure subtypes (A-D, F-H, J, K) and subsubtypes (A1-A4, F1-F2) were included in a first analysis. Next, a second analysis was done, using all pure subtypes (except subtype A, to avoid conflicting tree topologies with CRF45_cpx) and CRF04_cpx, CRF09_cpx, CRF45_cpx and CRF49_cpx sequences. For each query sequence, an alignment composed by each fragment between recombination breakpoints and the HIV-1 subtype reference sequences was used to construct NJ phylogenetic trees under the Tamura-Nei nucleotide substitution model and bootstrapping with 500 replicates using the MEGA 5.0 software [34] to confirm the subtype assignment.

Sequence datasets and alignments

Three datasets were used in the phylogenetic analysis, representing the *gag*, *pol* and *env* genes of HIV-1. The *pol* dataset is composed by whole PR/RT (HXB2: 2253 – 3272) sequences from HIV-1 clades CRF04_cpx, CRF09_cpx, CRF45_cpx and CRF49_cpx, i.e., CRFs with recombination breakpoints involving U and K in this region, available in the Los Alamos HIV database (Figure S1). To complement this dataset, the basic local alignment search tool (BLAST; www.ncbi.nlm.nih.gov/BLAST) was used to select HIV-1 sequences isolated worldwide that displayed a high similarity score (>95%) to the Brazilian strains and shared the genomic mosaic structure. All sequences were inspected by bootscaning analysis and NJ, as described above, to verify the pattern of recombination and those with a non-CRFs_cpx-like recombination pattern were discarded from the dataset. The *gag* (HXB2: 1194 – 2021) and *env* (HXB2: 6850 – 7350) datasets corresponded to regions with absence of recombination and composed by subsubtype A1 in both the Brazilian and some complex CRFs (Figure S1). All sequences that fulfilled these criteria (CRF09_cpx, CRF45_cpx and CRF49_cpx for *gag* and CRF45_cpx for *env*) were retrieved from the Los Alamos HIV Database, as well as, all HIV-1 sequences classified as subtypes A/A1/A2 at each gene region. Those isolates included in the *pol* dataset from the BLAST search that also had sequences available in the *gag* and *env* regions and met the criteria above were also added. The subtype classification of all *pol*, *gag* and *env* sequences were verified by NJ phylogenetic analysis, REGA HIV subtyping tool v.3 [35] and COMET v.2 [36] and those with wrong classification were excluded from the final datasets. These approaches resulted in three datasets containing 108 *pol*, 460 *gag* and 911 *env* sequences each (Table S1). Sequences were aligned using CLUSTAL Omega [37], followed by manual editing of some sites when necessary.

Maximum likelihood phylogenetic analyses

The maximum likelihood (ML) trees were inferred with the PhyML program [38] using an online web server [39] under the GTR+I+Γ_4 nucleotide substitution model selected using the jModeltest v.2 program [40]. The SPR branch-swapping algorithm was used to perform the heuristic tree search and the branch support was estimated with the approximate likelihood-ratio test (aLRT) [41] based on a Shimodaira-Hasegawa-like procedure. Trees were rooted with HIV-1 subtype reference sequences. The ML trees were visualized using the FigTree v1.4.2 program [42].

Temporal analysis of the Brazilian HIV-1 CRF45_cpx clade

A Bayesian Markov Chain Monte Carlo (MCMC) approach was used to jointly estimate the phylogenetic tree, evolutionary rate (μ, nucleotide substitutions per site per year, subst./site/year) and the age of the most recent common ancestor (T_{MRCA}, years) of HIV-1 CRF45_cpx-like dated sequences that composed the clades characterized in the ML analyses and showed the same subtype pattern in each gene fragment. The MCMC chains were implemented in BEAST v1.8.2 [43,44] along with the BEAGLE v2.1 library to improve computation time [45]. The uncorrelated lognormal relaxed molecular clock model [46] and Bayesian Skyline coalescent tree prior [47] were used, along with the nucleotide substitution models GTR+I+Γ_4, TN93+I+Γ_4 and GTR+Γ_4, select by the jModeltest v.2 program [48] for the *pol*, *gag* and *env* datasets, respectively. Informative substitution rate priors for the *pol* (1.5×10^{-3} – 3.0×10^{-3} substitution/site/year) [49,50], *gag* (2×10^{-3} – 4.5×10^{-3} substitution/site/year) [51–53] and *env* (4×10^{-3} – 7×10^{-3} substitution/site/year) [49] were provided for each run. MCMC chains were run for 1-2×10^{7} generations and adequate convergence was checked by calculating the effective sample size (ESS) after excluding an initial 10% burn-in for each run in TRACER v1.6 [54]. All parameter estimates for each run showed ESS

values >200. Maximum clade credibility (MCC) trees were summarized from the posterior distribution of trees with TreeAnnotator and visualized with FigTree v1.4 [42].

RESULTS

Classification of Brazilian HIV-1 A1/K/U recombinant *pol* samples

The BLAST search analysis of the 10 HIV *pol* A1/K/U Brazilian sequences recovered 24 sequences originated from Central (Congo and Democratic Republic of Congo), West-Central (Chad) and West (Senegal) Africa, Europe (Belgium, France, Italy, Norway, Poland, Spain and Switzerland) and South America (Argentina and Brazil) initially classified as unique recombinant forms (URFs), U or even CRF06_cpx. All Brazilian HIV-1 A1/K/U sequences from this study and those sequences recovered by the BLAST analysis were reclassified as CRF45_cpx-like once they clustered with a high support (aLRT = 0.98) inside the CRF45_cpx clade (Fig. 1 A. This clearly expands the list of countries were the CRF45_cpx could be circulating.

Seven sequences recovered by BLAST were from Brazil and jointly with the new samples from this study, except one (07BRRJPC38), formed a monophyletic clade (CRF45-BR, aLRT = 0.70) along with a sequence isolated in Italy (EF488575). It is noteworthy that one Argentinian sample, isolated from a heterosexual woman and previously classified as a CRF06_cpx [55], formed a basal branch to the CRF45-BR clade (aLRT = 0.99). This clade seems to be the result of a single introduction event that has spread to the Southeast region of Brazil, in the Rio de Janeiro, São Paulo and Minas Gerais states (Fig.1 B). The largest number of CRF45_cpx-related sequences in Brazil was found in Rio de Janeiro (n = 14), where this clade has spread to at least five counties of the metropolitan and south state regions (Fig.1 B).

The bootscaning analysis showed that although most Brazilian CRF45_cpx-like *pol* sequences formed a monophyletic clade, some sequences were actually second generation recombinants (SGRs). Two recombination patterns were found, one composed by CRF45_cpx and subsubtype F1 detected in two samples from Rio de Janeiro recovered from heterosexual women, one of which was as a recent HIV seroconverter (GenBank accession number (AN): KF922200) [21] and the other (08BRRJPG226) an ARV-naïve pregnant woman [20]. The other recombination pattern comprised the CRF45_cpx and subtype B clades and was also found in two heterosexual women, one recent seroconverter (AN: KF922154) [21] and a pregnant-woman (07BRRJPG144) [20]. The sequence 09BRRJCR019 clustered together with the CRF45_cpx/B recombinants, however, it did not show the same recombinant pattern, displaying a distinct U fragment in *pol* (Fig.1 A).

NFLGs of the CRF45_cpx-like and SGRs sequences identified in Rio de Janeiro

Among the 10 A1/K/U *pol* samples isolated in Rio de Janeiro, six were classified as CRF45_cpx-like and the others as SGRs between CRF45_cpx and B, F1 or U clades (Fig.1 A). Partial and NFLG of these isolates were obtained to unveil the mosaic subtype structure along the genome. Nine isolates had at least portions of the three viral structural genes *gag*, *pol* and *env* sequenced (Fig.2). Among the samples analyzed, six showed a CRF45_cpx-like pattern throughout the whole genome. The remaining four consisted of mosaic recombinants between HIV-1 clades CRF45_cpx, B, D, F1 and U. None of the Brazilian SGRs shared a similar mosaic profile and no obvious preferred location for breakpoints was found. The observed mosaic patterns were CRF45_cpx/B/D/F1/U, CRF45_cpx/B/F1/U, CRF45_cpx/B/U and CRF45_cpx/F1, all displaying at least one recombinant fragment with subtypes B and/or F1, characteristic of the Brazilian HIV epidemic (Table 1). Subregion tree analyses indicated that the

subtype F1 fragments found in SGRs were of Brazilian origin (Fig. S2). This analysis was not done with subtype B fragments, since the Brazilian epidemic is characterized by multiple introductions of this subtype [56].

To confirm the shared ancestry of the Brazilian CRF45-BR *pol* clade previously identified, partial genome sequences derived from the *gag* and *env* genes classified as subtype A1 (same subtype as observed in the CRF45_cpx genome) were analyzed. Brazilian CRF45_cpx-like sequences were combined with HIV-1 sequences of subtypes A/A1/A2 and other CRFs_cpx that share the same subtype composition in those regions (Fig. S1). The ML phylogenetic analyses of *gag* and *env* regions confirmed the monophyletic origin of the Brazilian CRF45-BR clade, that formed a highly supported subgroup (aLRT $\geq$ 0.95) nested within the CRF45_cpx radiation (Fig. 3 A and B). Notably, two subtype A1 sequences from Belgium clustered inside the CRF45-BR clade in the *env* dataset. As expected, the sequence 07BRRJPC38 remained inside de CRF45_cpx clade, but did not cluster within the CRF45-BR sub-clade, confirming the independent origin of this CRF45_cpx-like strain.

Time-scale of the HIV-1 CRF45_cpx Brazilian clade

All HIV-1 sequences that clustered inside the CRF45_cpx clade in the ML analyses and were not classified as SGRs were included in the Bayesian analyses. This resulted in three datasets composed by 44 *pol*, 30 *gag* and 14 *env* sequences sampled between 1985 and 2014 (Table S2). Bayesian analyses revealed that the CRF45-BR clade formed a highly (posterior probability, *PP* > 0.95) or relatively highly (*PP* = 0.77) supported monophyletic group in all genomic regions (Fig. 4), consistent with the ML topology. The median evolutionary rates calculated with informative priors under a relaxed molecular clock model for *pol*, *gag* and *env* genes, were 1.58×10^{-3} subst./site/year, 2.15×10^{-3} subst./site/year and 4.77×10^{-3} subst./site/year, respectively (Table 2). The

coefficient of rate variation estimated from all datasets was significantly higher than zero (Table 2), thus supporting the use of a relaxed molecular clock model for all genomic regions. Bayesian analyses pointed to quite similar median onset dates for the CRF45_cpx clade between 1970 and 1975 and for the CRF45-BR sub-clade between 1981 and 1986 across genes (Table 2 and Fig. 4), indicating consistency in the estimates despite the limited amount of sequences available.

DISCUSSION

In this study, we describe partial and near full-length genome sequences of HIV-1 samples previously classified as URFs bearing subtype K found in independent studies carried out with HIV+ individuals from Brazil and reconstruct their evolutionary history. Ten out of 16 Brazilian HIV-1 *pol* sequences bearing subtype K fragments described in this study and recovered by BLAST were classified as CRF45_cpx-like sequences and the remaining six were classified as SGRs between CRF45_cpx and subtypes B, F1 and/or unclassified clades. Most CRF45_cpx-like and SGRs Brazilian sequences were detected in heterosexual patients and vertically infected children from the Rio de Janeiro state, indicating that these strains are mainly disseminating by heterosexual contacts. This is in agreement with the actual main transmission route among teens and adults in Brazil, in which heterosexual sex accounted for >75 percent of all AIDS cases in the last ten years [57].

The BLAST search analysis of the CRF45_cpx-like Brazilian sequences recovered 30 *pol* sequences from Africa, Europe and the Americas previously designated as U, URFs_UK or even CRF06_cpx that were reclassified in this study as CRF45_cpx-like. It was already suggested that the classification of some sequences deposited in the Los Alamos sequence database needs to be revised as a consequence of errors in the original

subtype assignments [58]. The complex recombinants HIV-MAL and HIV-NOGIL were among the sequences retrieved by BLAST, confirming the hypothesis that CRF45_cpx is probably the common ancestor of these lineages [24]. Originally restricted to Central African countries [59–61], the reclassification done in this study expanded the geographic dissemination of the CRF45_cpx clade to Europe (Belgium, France, Italy, Norway, Poland, Spain and Switzerland) and South America (Argentina and Brazil), indicating that at least fragments of this lineage could be circulating worldwide, as previously proposed [24].

Nine samples isolated in Rio de Janeiro that branched inside the CRF45_cpx in the *pol* ML tree had partial and near-full length genomes characterized. Five samples showed a CRF45_cpx-like pattern throughout the whole sequenced genome. The remaining were classified as SGRs, following the classification based in the *pol* gene alone, reinforcing the power of this region to monitor changes in HIV molecular epidemiology [62]. The SGRs displayed at least one recombinant fragment composed by subtypes B and/or F1, highly prevalent in the Brazilian HIV epidemic. The low prevalence of subtype B in the Central Africa [4], and the phylogenetic relationship of the F1 genomic fragments of SGRs with Brazilian subtype F1 clade, suggests that SGRs were generated locally after the introduction of the CRF45_cpx lineage in Brazil. An issue with this hypothesis is the presence of a subtype D fragment in one SGR. Although subtype D is rarely detected in the Brazilian HIV-1 epidemic, it was repeatedly isolated in the Rio de Janeiro state [63], which makes plausible the chance of local recombination between CRF45_cpx and subtype D.

The HIV-1 epidemic seems to be evolving toward a more complex epidemiological landscape. Even with limited circulation in Brazil, the CRF45_cpx lineage described here seems to have already recombined with local HIV-1 strains, forming SGRs. In

addition to constituting a further challenge to accurate virus genotyping and breakpoints identification, the recombination may play an important role in the evolution of HIV-1, providing biological advantage in replicative capacity [64] or favoring drug resistance acquisition during antiretroviral therapy [65]. Brazilian recombinants bearing B and F1 subtypes seem to be associated with an accelerated CD4+ T-cell loss [66], however, the clinical implications of the SGRs here described remain to be clarified.

The ML phylogenetic analyses of *pol*, *gag* and *env* genes revealed that nearly all Brazilian CRF45_cpx-like sequences formed a highly supported monophyletic sub-group, here denominated CRF45-BR, within the CRF45_cpx radiation. This strongly support that most Brazilian CRF45_cpx-like sequences derived from a single introduction event of this clade followed by local dissemination. Most HIV-1 sequences composing the CRF45-BR clade were isolated in Rio de Janeiro, although a few sequences from São Paulo and Minas Gerais were also detected, indicating that this clade was probably introduced and is mostly circulating in the Southeast region of Brazil. Previous studies also identified the circulation and local dissemination of other HIV-1 clades of African origin (CRF02_AG and subtypes A and D) in Rio de Janeiro [8,10,12,13,63,67]. The Brazilian Southeast region hosts the largest international airports, ports, and sociocultural and economic events, which create a favorable environment for the introduction and spread of new HIV-1 strains.

Of note, a few sequences from Italy and Belgium branched inside the CRF45-BR clade in the *pol* and *env* ML analysis. The presence of sequences from Europe inside the CRF45-BR clade may support a more global spread of this clade, however, the lack of epidemiological information of these sequences [68,69] creates uncertainty about their actual origin. Since a substantial portion of non-B clades found in Italy come from HIV-1-infected Brazilian patients followed at Italian clinics [7], the lineage exportation

scenario should be interpreted with caution. It is also interesting to note that an Argentinian CRF45_cpx-like sequence branched with a high support at the base of the CRF45-BR clade, supporting a single introduction event of the CRF45_cpx clade in South America. Clearly, analyses using more samples are necessary to determine the exact route of introduction and dissemination of the CRF45_cpx between Argentina and Brazil.

Bayesian analyses of *gag*, *pol* and *env* genes traced the origin of the CRF45_cpx clade back to the first half of 1970 decade (1971 – 1976). Interestingly, this time period coincides with the upper and lower limits of the onset date of the CRF06_cpx [70] , another complex recombinant carrying subtypes A and K. Despite the recent detection of the CRF45_cpx-like variants in Brazil, our estimations traced the origin of the CRF45-BR clade to the first half of 1980 decade (1981 – 1986), suggesting that this CRF has been circulating in Brazil for about 20-30 years before it has been detected by the public surveillance system. It was proposed that the relatively late introduction and differences of transmission network efficacies might have caused the limited spread of CRF02_AG in Brazil [71] and this could be the same for the CRF45-BR clade. The onset date of the CRF45-BR clade coincides to the time of introduction of the CRF02_AG clade in Rio de Janeiro [71] and is later than that estimated for HIV-1 subtypes C and F1 in Brazil [56,72,73]. Furthermore, the majority of CRF45_cpx samples were collected from or related to mother-to-child and heterosexual transmission routes that are associated to a lower risk of HIV acquisition [74], thus resulting in a more limited propagation.

One Brazilian CRF45_cpx sample branched outside the CRF45-BR clade, constituting an independent introduction event of this clade in the country. This sequence showed a U fragment in the PR region of the *pol* gene and although being isolated from a

vertically-infected child that was born in Brazil, her mother was from Angola, where she probably got HIV infected. Thus, this CRF45_cpx-like sequence probably represents an independent introduction event of a CRF45_cpx-like strain in Brazil from Angola, and not an autochthonous transmission event. This also demonstrates that more CRF45_cpx lineages are being introduced in Brazil.

In summary, the results presented here unveil the existence of CRF45_cpx isolates in Brazil, forming a monophyletic clade that originated around the middle of 1980s by the introduction of a single founder strain in the Brazilian Southeast region, where it probably recombined with local HIV-1 B and F1 lineages originating SGRs. The Brazilian CRF45_cpx-like sequences are also closely related to HIV-1 sequences from Argentina, Belgium and Italy, indicating the possible existence of an international transmission network of this complex CRF. These results highlight the importance of continuous molecular surveillance studies to identify the emergence and establishment of new HIV variants in Brazil and its impact on local AIDS epidemic.

ACKNOWLEDGMENTS

E.D. is supported by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). We thank the researchers whose publicly available data made this work possible, Dr. Gonzalo Bello for helpful and critical comments and Dr. Vera Bongertz for review of the manuscript grammar.

REFERENCES

1. Hemelaar J. The origin and diversity of the HIV-1 pandemic. Trends Mol Med. Elsevier Ltd; 2012;18: 182–92. doi:10.1016/j.molmed.2011.12.001
2. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature. 2008;455: 661–4. doi:10.1038/nature07390
3. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. The early spread and epidemic ignition of HIV-1 in human populations. Science (80-). 2014;346: 56–61. doi:10.1126/science.1256739

4. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global trends in molecular epidemiology of HIV-1 during 2000–2007. Aids. 2011;25: 679–689. doi:10.1097/QAD.0b013e328342ff93
5. Lospitao E, Álvarez a., Soriano V, Holguín A. HIV-1 subtypes in Spain: A retrospective analysis from 1995 to 2003. HIV Med. 2005;6: 313–320. doi:10.1111/j.1468-1293.2005.00313.x
6. Brand D, Cazein F, Lot F, Brunet S, Thierry D, Moreau A, et al. Continuous spread of HIV-1 subtypes D and CRF01-AE in France from 2003 to 2009. J Clin Microbiol. 2012;50: 2484–2488. doi:10.1128/JCM.00319-12
7. Lai a., Riva C, Marconi a., Balestrieri M, Razzolini F, Meini G, et al. Changing patterns in HIV-1 non-B clade prevalence and diversity in Italy over three decades. HIV Med. 2010;11: 593–602. doi:10.1111/j.1468-1293.2010.00832.x
8. Caride E, Brindeiro R, Hertogs K, Larder B, Dehertogh P, Machado E, et al. Drug-resistant reverse transcriptase genotyping and phenotyping of B and non-B subtypes (F and A) of human immunodeficiency virus type I found in Brazilian patients failing HAART. Virology. 2000;275: 107–15. doi:10.1006/viro.2000.0487
9. Pimentel VF, Morgado MG, Bello G, Guimarães MDC, Castilho E a, Veloso VG, et al. Temporal trends and molecular epidemiology of HIV type 1 infection in Rio de Janeiro, Brazil. AIDS Res Hum Retroviruses. 2013;29: 1553–61. doi:10.1089/AID.2013.0050
10. Morgado MG, Guimarães ML, Gripp CB, Costa CI, Neves I, Veloso VG, et al. Molecular epidemiology of HIV-1 in Brazil: high prevalence of HIV-1 subtype B and identification of an HIV-1 subtype D infection in the city of Rio de Janeiro, Brazil. Evandro Chagas Hospital AIDS Clinical Research Group. J Acquir Immune Defic Syndr Hum Retrovirol. 1998;18: 488–94.
11. Alencar CS, Sabino EC, Carvalho SMF, Leao SC, Carneiro-proietti AB, Capuani L, et al. HIV Genotypes and Primary Drug Resistance Among HIV-Seropositive Blood Donors in Brazil : Role of Infected Blood Donors as Sentinel Populations for Molecular Surveillance of HIV. JAIDS. 2013;63: 387–392. doi:10.1097/QAI.0b013e31828ff979
12. Eyer-Silva W a, Morgado MG. Autochthonous horizontal transmission of a CRF02_AG strain revealed by a human immunodeficiency virus type 1 diversity survey in a small city in inner state of Rio de Janeiro, Southeast Brazil. Mem Inst Oswaldo Cruz. 2007;102: 809–15.
13. Delatorre EO, Bello G, Eyer-Silva WA, Chequer-Fernandez SL, Morgado MG, Couto-Fernandez JC. Evidence of multiple introductions and autochthonous transmission of the HIV type 1 CRF02_AG clade in Brazil. AIDS Res Hum Retroviruses. 2012;28: 1369–72. doi:10.1089/AID.2011.0381
14. Ferreira JLP, Rodrigues R, Lança AM, de Almeida VC, Rocha SQ, Ragazzo TG, et al. Transmitted Drug Resistance among People Living with HIV/Aids at Major Cities of Sao Paulo State, Brazil. Adv Virol. 2013;2013: 878237. doi:10.1155/2013/878237
15. Machado LF a, Ishak MOG, Vallinoto ACR, Lemos JAR, Azevedo VN, Moreira MRC, et al. Molecular epidemiology of HIV type 1 in northern Brazil: identification of subtypes C and D and the introduction of CRF02_AG in the Amazon region of Brazil. AIDS Res Hum Retroviruses. 2009;25: 961–6. doi:10.1089/aid.2009.0027
16. Inocencio L a., Pereira A a., Sucupira MC a, Fernandez JCC, Jorge CP, Souza DFC, et al. Brazilian Network for HIV Drug Resistance Surveillance: A survey of individuals recently diagnosed with HIV. J Int AIDS Soc. 2009;12: 1–6. doi:10.1186/1758-2652-12-20
17. Brindeiro P a., Brindeiro RM, Mortensen C, Hertogs K, De Vroey V, Rubini NPM, et al. Testing genotypic and phenotypic resistance in human immunodeficiency virus type 1 isolates of clade B and other clades from children failing antiretroviral therapy. J Clin Microbiol. 2002;40: 4512–4519. doi:10.1128/JCM.40.12.4512-4519.2002
18. Ferreira FGF, Pinto J a, Kakehasi FM, Cleto S, Tupinambás U, Aleixo a W, et al. Prevalence of primary drug resistance-associated mutations among HIV type 1 vertically Infected children in Belo Horizonte, Brazil. AIDS Res Hum Retroviruses. 2010;26: 229–232. doi:10.1089/aid.2009.0146
19. Pessôa R, Watanabe JT, Calabria P, Alencar CS, Loureiro P, Lopes ME, et al. Enhanced detection of viral diversity using partial and near full-length genomes of human immunodeficiency virus Type 1 provirus deep sequencing data from recently infected donors at four blood centers in Brazil. Transfusion. 2015;55: 980–90. doi:10.1111/trf.12936
20. Pilotto JH, Grinsztejn B, Veloso VG, Velasque LS, Friedman RK, Moreira RI, et al. Moderate prevalence of transmitted drug resistance mutations among antiretroviral-naive HIV-infected pregnant women in Rio de Janeiro, Brazil. AIDS Res Hum Retroviruses. 2013;29: 681–6. doi:10.1089/AID.2011.0333
21. Velasco-de-Castro CA, Grinsztejn B, Veloso VG, Bastos FI, Pilotto JH, Fernandes N, et al. HIV-1 diversity and drug resistance mutations among people seeking HIV diagnosis in voluntary

counseling and testing sites in Rio de Janeiro, Brazil. PLoS One. 2014;9: e87622. doi:10.1371/journal.pone.0087622

22. Paraskevis D, Magiorkinis M, Vandamme a. M, Kostrikis LG, Hatzakis A. Re-analysis of human immunodeficiency virus type 1 isolates from Cyprus and Greece, initially designated "subtype I", reveals a unique complex A/G/H/K/? mosaic pattern. J Gen Virol. 2001;82: 575–580.

23. McCutchan FE, Sankale J-L, M'Boup S, Kim B, Tovanabutra S, Hamel DJ, et al. HIV type 1 circulating recombinant form CRF09_cpx from west Africa combines subtypes A, F, G, and may share ancestors with CRF02_AG and Z321. AIDS Res Hum Retroviruses. 2004;20: 819–826. doi:10.1089/0889222041725163

24. Niama FR, Vidal N, Bazepeo SE, Mpoudi E, Toure-Kane C, Parra HJ, et al. CRF45_AKU, a circulating recombinant from Central Africa, is probably the common ancestor of HIV type 1 MAL and HIV type 1 NOGIL. AIDS Res Hum Retroviruses. 2009;25: 1345–1353. doi:10.1089/aid.2009.0169

25. De Silva TI, Turner R, Hué S, Trikha R, van Tienen C, Onyango C, et al. HIV-1 subtype distribution in the Gambia and the significant presence of CRF49_cpx, a novel circulating recombinant form. Retrovirology. BioMed Central Ltd; 2010;7: 82. doi:10.1186/1742-4690-7-82

26. Barber TJ, Harrison L, Asboe D, Williams I, Kirk S, Gilson R, et al. Frequency and patterns of protease gene resistance mutations in HIV-infected patients treated with lopinavir/ritonavir as their first protease inhibitor. J Antimicrob Chemother. 2012;67: 995–1000. doi:10.1093/jac/dkr569

27. De Mendoza C, Garrido C, Poveda E, Corral A, Zahonero N, Treviño A, et al. Changes in drug resistance patterns following the introduction of HIV type 1 non-B subtypes in Spain. AIDS Res Hum Retroviruses. 2009;25: 967–972. doi:10.1089/aid.2008.0166

28. Galimand J, Frange P, Rouzioux C, Deveau C, Avettand-Fenoël V, Ghosn J, et al. Short communication: evidence of HIV type 1 complex and second generation recombinant strains among patients infected in 1997-2007 in France: ANRS CO06 PRIMO Cohort. AIDS Res Hum Retroviruses. 2010;26: 645–651. doi:10.1089/aid.2009.0201

29. Frentz D, Wensing AMJ, Albert J, Paraskevis D, Abecasis AB, Hamouda O, et al. Limited cross-border infections in patients newly diagnosed with HIV in Europe. Retrovirology. 2013;10: 36. doi:10.1186/1742-4690-10-36

30. Pyne MT, Hackett J, Holzmayer V, Hillyard DR. Large-scale analysis of the prevalence and geographic distribution of HIV-1 non-B variants in the United States. J Clin Microbiol. 2013;51: 2662–2669. doi:10.1128/JCM.00880-13

31. Tovanabutra S, Brodine SK, Mascola JR, Sankale J-L, Sanders-Buell E, Kim B, et al. Characterization of complete HIV type 1 genomes from non-B subtype infections in U.S. military personnel. AIDS Res Hum Retroviruses. 2005;21: 424–429. doi:10.1089/aid.2005.21.424

32. Passaes CPB, Guimarães ML, Bello G, Morgado MG. Near full-length genome characterization of HIV type 1 unique BC recombinant forms from Southern Brazil. AIDS Res Hum Retroviruses. 2009;25: 1339–1344. doi:10.1089/aid.2009.0167

33. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol. 1999;73: 152–60. Available: http://jvi.asm.org/content/73/1/152.short

34. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28: 2731–9. doi:10.1093/molbev/msr121

35. Pineda-Peña AC, Faria NR, Imbrechts S, Libin P, Abecasis AB, Deforche K, et al. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. Infect Genet Evol. Elsevier B.V.; 2013;19: 337–348. doi:10.1016/j.meegid.2013.04.032

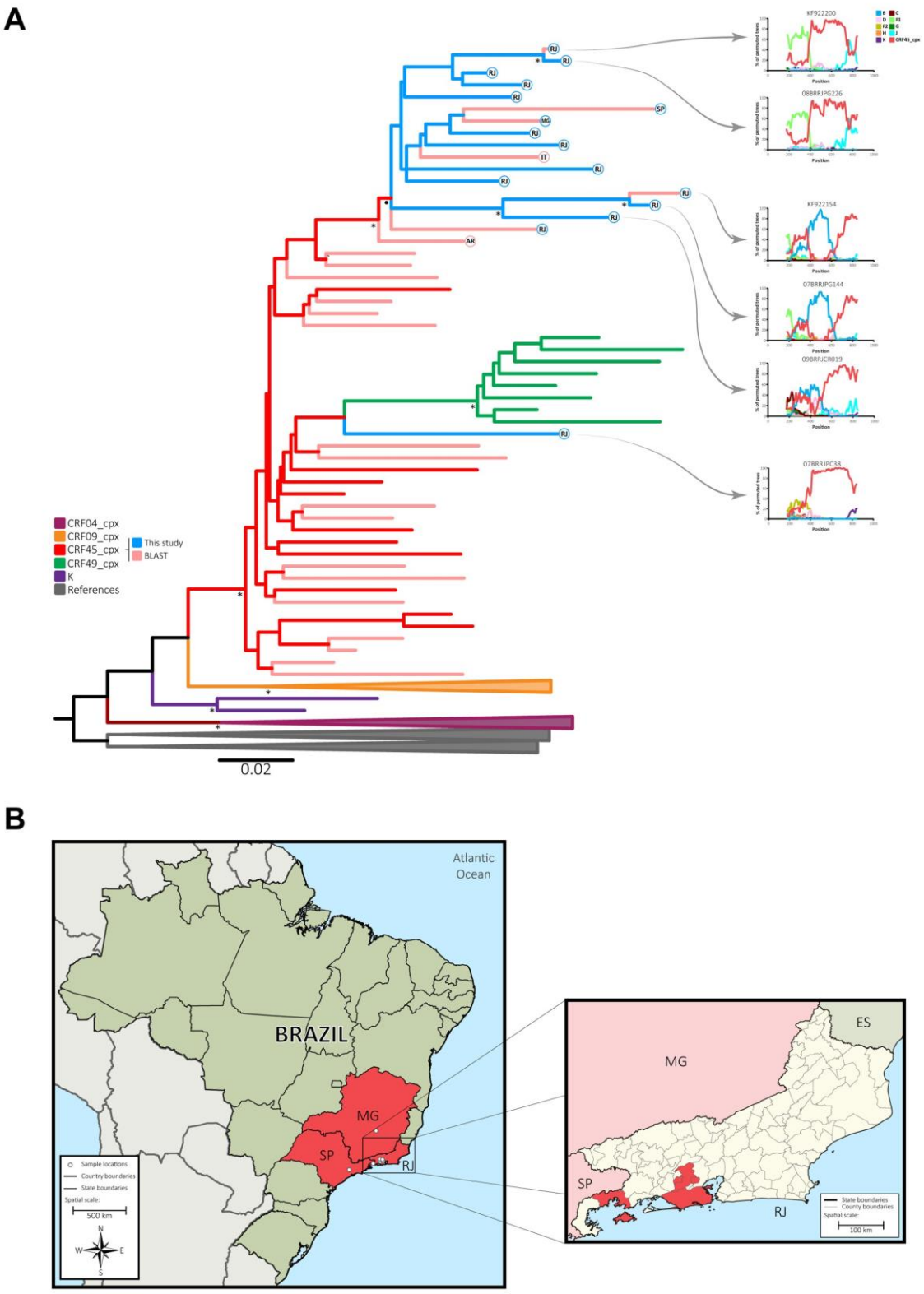
36. Struck D, Lawyer G, Ternes A-M, Schmit J-C, Perez Bercoff D. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic Acids Res. 2014; 1–11. doi:10.1093/nar/gku739

37. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7. doi:10.1038/msb.2011.75

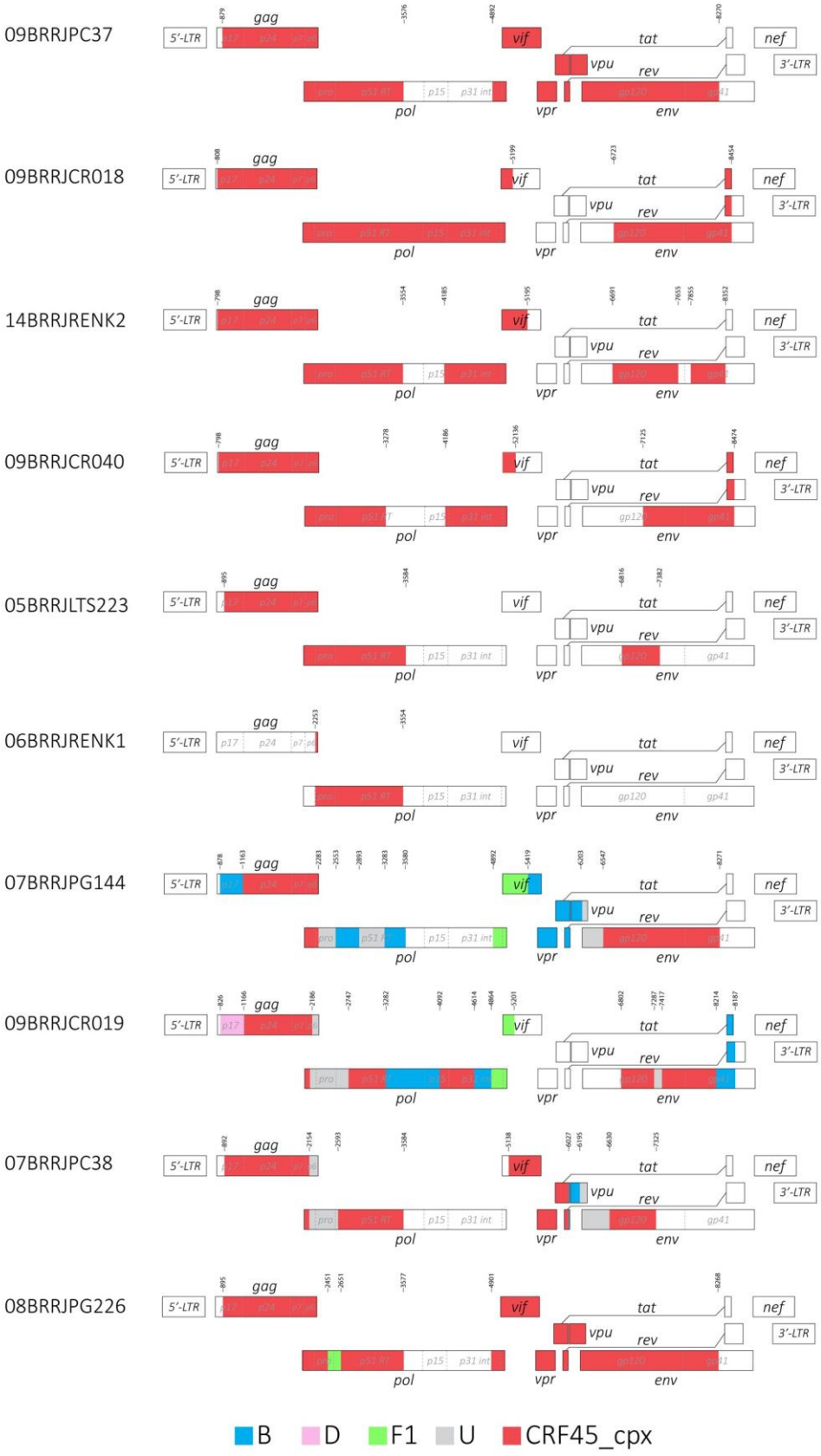
38. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59: 307–21. doi:10.1093/sysbio/syq010

39. Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 2005;33: W557–9. doi:10.1093/nar/gki352
40. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. Nature Publishing Group; 2012;9: 772–772. doi:10.1038/nmeth.2109
41. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol. 2006;55: 539–52. doi:10.1080/10635150600755453
42. Rambaut A. FigTree v1.4 [Internet]. 2014. Available: http://tree.bio.ed.ac.uk/software/figtree/
43. Drummond AJ, Suchard M a., Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29: 1969–1973. doi:10.1093/molbev/mss075
44. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics. 2002;161: 1307–20.
45. Suchard M a, Rambaut A. Many-core algorithms for statistical phylogenetics. Bioinformatics. 2009;25: 1370–6. doi:10.1093/bioinformatics/btp244
46. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006;4: e88. doi:10.1371/journal.pbio.0040088
47. Drummond a J, Rambaut a, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005;22: 1185–92. doi:10.1093/molbev/msi103
48. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. Nature Publishing Group; 2012;9: 772–772. doi:10.1038/nmeth.2109
49. Abecasis AB, Vandamme A-M, Lemey P. Quantifying differences in the tempo of human immunodeficiency virus type 1 subtype evolution. J Virol. 2009;83: 12917–12924. doi:10.1128/JVI.01022-09
50. Paraskevis D, Magiorkinis E, Magiorkinis G, Sypsa V, Paparizos V, Lazanas M, et al. Increasing prevalence of HIV-1 subtype A in Greece: estimating epidemic history and origin. J Infect Dis. 2007;196: 1167–1176. doi:10.1086/521677
51. Tee KK, Pybus OG, Parker J, Ng KP, Kamarulzaman A, Takebe Y. Estimating the date of origin of an HIV-1 circulating recombinant form. Virology. Elsevier Inc.; 2009;387: 229–234. doi:10.1016/j.virol.2009.02.020
52. Tee KK, Pybus OG, Li X-J, Han X, Shang H, Kamarulzaman A, et al. Temporal and spatial dynamics of human immunodeficiency virus type 1 circulating recombinant forms 08_BC and 07_BC in Asia. J Virol. 2008;82: 9206–9215. doi:10.1128/JVI.00399-08
53. Paraskevis D, Magiorkinis E, Magiorkinis G, Kiosses VG, Lemey P, Vandamme AM, et al. Phylogenetic reconstruction of a known HIV-1 CRF04_cpx transmission network using maximum likelihood and Bayesian methods. J Mol Evol. 2004;59: 709–717. doi:10.1007/s00239-004-2651-6
54. Rambaut A, Suchard M, Drummond A. Tracer v1.6 [Internet]. 2013. Available: http://tree.bio.ed.ac.uk/software/tracer/
55. Dilernia D a, Gomez AM, Lourtau L, Marone R, Losso MH, Salomón H, et al. HIV type 1 genetic diversity surveillance among newly diagnosed individuals from 2003 to 2005 in Buenos Aires, Argentina. AIDS Res Hum Retroviruses. 2007;23: 1201–1207. doi:10.1089/aid.2007.0068
56. Bello G, Eyer-Silva W a, Couto-Fernandez JC, Guimarães ML, Chequer-Fernandez SL, Teixeira SLM, et al. Demographic history of HIV-1 subtypes B and F in Brazil. Infect Genet Evol. 2007;7: 263–70. doi:10.1016/j.meegid.2006.11.002
57. Ministério da Saúde. Boletim Epidemiológico - Aids e DST [Internet]. Brasília, DF - Brasil; 2014. Available: www.aids.gov.br
58. Zhang M, Foley B, Schultz A-K, Macke JP, Bulla I, Stanke M, et al. The role of recombination in the emergence of a complex and dynamic HIV epidemic. Retrovirology. 2010;7: 25. doi:10.1186/1742-4690-7-25
59. Afonso JM, Bello G, Guimarães ML, Sojka M, Morgado MG. HIV-1 genetic diversity and transmitted drug resistance mutations among patients from the North, Central and South regions of Angola. PLoS One. 2012;7: e42996. doi:10.1371/journal.pone.0042996
60. Djoko CF, Rimoin AW, Vidal N, Tamoufe U, Wolfe ND, Butel C, et al. High HIV type 1 group M pol diversity and low rate of antiretroviral resistance mutations among the uniformed services in Kinshasa, Democratic Republic of the Congo. AIDS Res Hum Retroviruses. 2011;27: 323–329. doi:10.1089/aid.2010.0201

61. Bruzzone B, Saladini F, Sticchi L, Mayinda Mboungou F a., Barresi R, Caligiuri P, et al. Prevalence of HIV-1 Subtypes and Drug Resistance-Associated Mutations in HIV-1-Positive Treatment-Naive Pregnant Women in Pointe Noire, Republic of the Congo (Kento-Mwana Project). AIDS Res Hum Retroviruses. 2015;31: 150602103920001. doi:10.1089/aid.2015.0105
62. Gifford RJ, de Oliveira T, Rambaut A, Pybus OG, Dunn D, Vandamme A-M, et al. Phylogenetic surveillance of viral genetic diversity and the evolving molecular epidemiology of human immunodeficiency virus type 1. J Virol. 2007;81: 13050–13056. doi:10.1128/JVI.00889-07
63. Couto-Fernandez JC, Eyer-Silva W a, Guimarães ML, Chequer-Fernandez SL, Grinsztejn B, Delaporte E, et al. Phylogenetic analysis of Brazilian HIV type 1 subtype D strains: tracing the origin of this subtype in Brazil. AIDS Res Hum Retroviruses. 2006;22: 207–11. doi:10.1089/aid.2006.22.207
64. Konings FAJ, Burda ST, Urbanski MM, Zhong P, Nadas A, Nyambi PN. Human immunodeficiency virus type 1 (HIV-1) circulating recombinant form 02_AG (CRF02_AG) has a higher in vitro replicative capacity than its parental subtypes A and G. J Med Virol. 2006;78: 523–534. doi:10.1002/jmv.20572
65. Carvajal-Rodríguez A, Crandall K a., Posada D. Recombination favors the evolution of drug resistance in HIV-1 during antiretroviral therapy. Infect Genet Evol. 2007;7: 476–483. doi:10.1016/j.meegid.2007.02.001
66. Tarosso LF, Sanabani SS, Ribeiro SP, Sauer MM, Tomiyama HI, Sucupira MC, et al. HIV Type 1 Subtype BF Leads to Faster CD4(+) T Cell Loss Compared to Subtype B. AIDS Res Hum Retroviruses. 2013;29: 1–5. doi:10.1089/AID.2012.0243
67. Pimentel VF, Morgado MG, Bello G, Guimarães MDC, Castilho E a, Veloso VG, et al. Temporal trends and molecular epidemiology of HIV type 1 infection in Rio de Janeiro, Brazil. AIDS Res Hum Retroviruses. 2013;29: 1553–61. doi:10.1089/AID.2013.0050
68. Galli A, Lai A, Corvasce S, Saladini F, Riva C, Dehò L, et al. Recombination analysis and structure prediction show correlation between breakpoint clusters and RNA hairpins in the pol gene of human immunodeficiency virus type 1 unique recombinant forms. J Gen Virol. 2008;89: 3119–3125. doi:10.1099/vir.0.2008/003418-0
69. Chalmet K, Dauwe K, Foquet L, Baatz F, Seguin-Devaux C, Van Der Gucht B, et al. Presence of CXCR4-Using HIV-1 in patients with recently diagnosed infection: Correlates and evidence for transmission. J Infect Dis. 2012;205: 174–184. doi:10.1093/infdis/jir714
70. Delatorre E, Bello G. Spatiotemporal dynamics of the HIV-1 CRF06_cpx epidemic in Western Africa. AIDS. 2013;27: 1313–20. doi:10.1097/QAD.0b013e32835f1df4
71. Delatorre E, Velasco-de-Castro CA, Pilotto JH, Couto-Fernandez JC, BELLO G, MORGADO MG. Reassessing the origin of the HIV-1 CRF02_AG lineages circulating in Brazil. AIDS Res Hum Retroviruses. 2015;0: 1–17.
72. Delatorre E, Couto-Fernandez JC, Guimarães ML, Vaz Cardoso LP, de Alcantara KC, Stefani MM de A, et al. Tracing the origin and northward dissemination dynamics of HIV-1 subtype C in Brazil. PLoS One. 2013;8: e74072. doi:10.1371/journal.pone.0074072
73. Bello G, Passaes CP, Guimarães ML, Lorete RS, Matos Almeida SE, Medeiros RM, et al. Origin and evolutionary history of HIV-1 subtype C in Brazil. AIDS. 2008;22: 1993–2000. doi:10.1097/QAD.0b013e328315e0aa
74. Patel P, Borkowf CB, Brooks JT, Lasry A, Lansky A, Mermin J. Estimating per-act HIV transmission risk. Aids. 2014; 1. doi:10.1097/QAD.0000000000000298



635
636 **FIGURE 1**



637 **FIGURE 2**
638

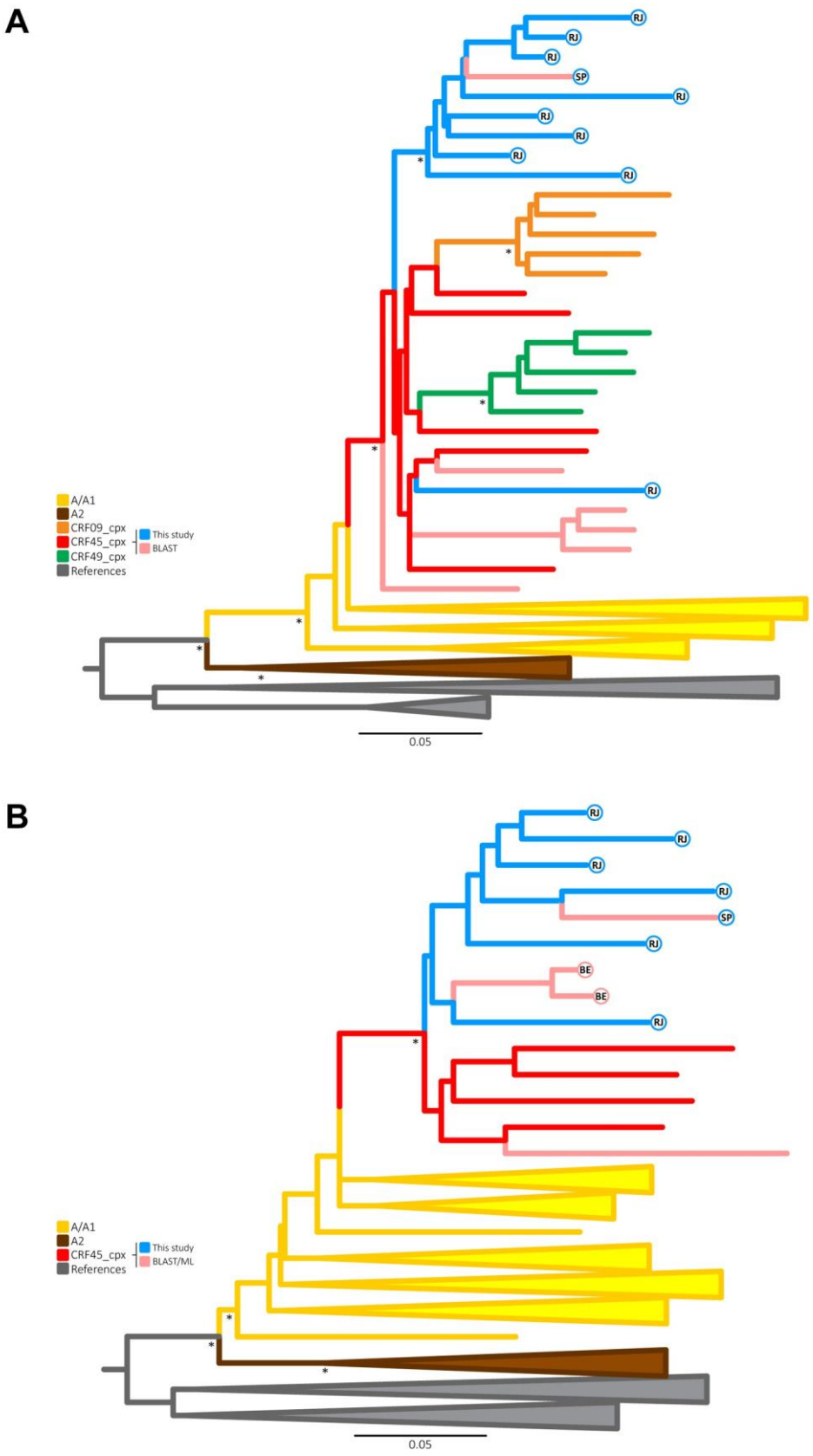
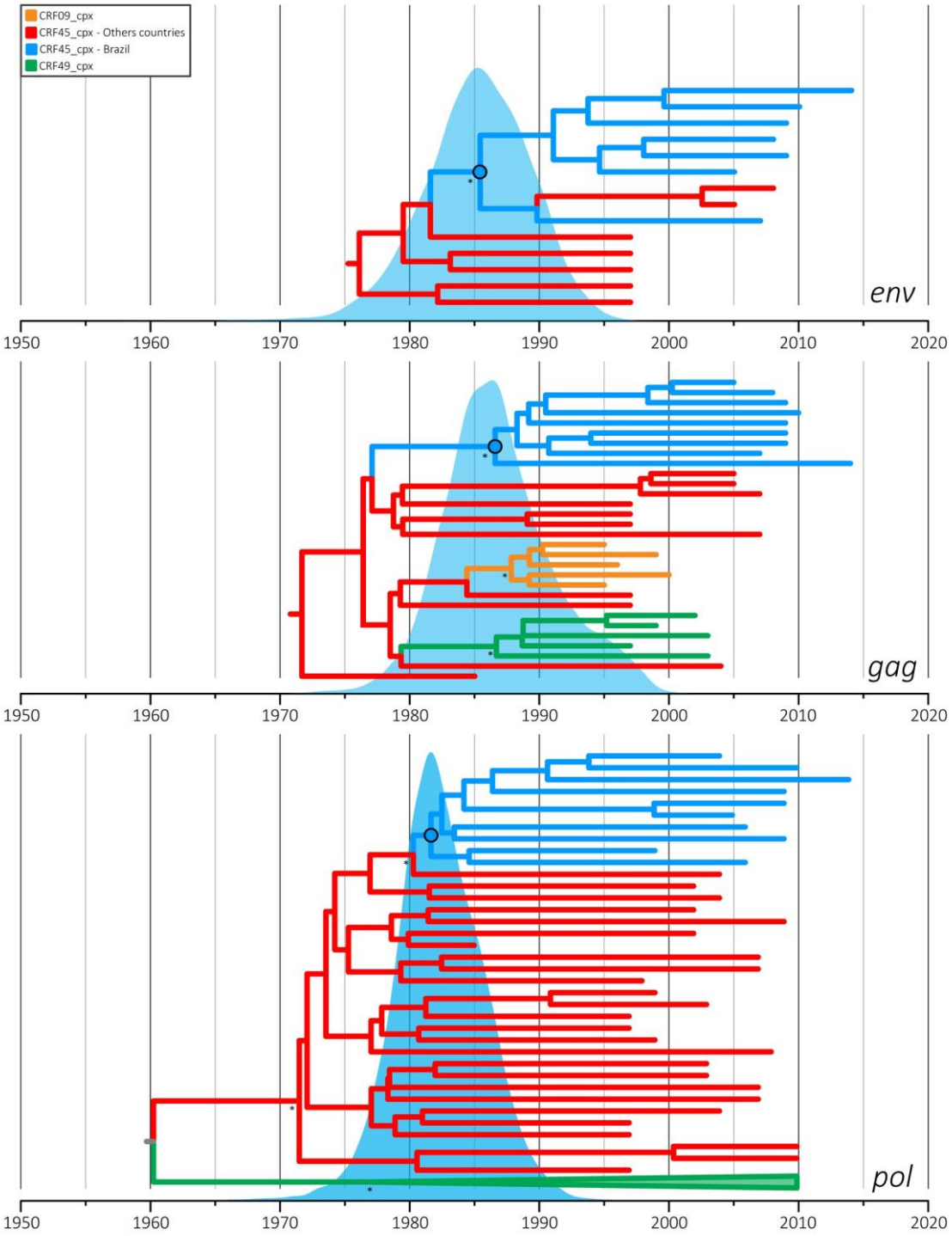

A
RJ
SP
BE
A/A1
A2
CRF09_cpx
CRF45_cpx
CRF49_cpx
References
This study
BLAST
BLAST/ML
0.05
B

FIGURE 3



644

645 **FIGURE 4**

FIGURE LEGENDS

Figure 1. Classification and distribution of the HIV-1 CRF45_cpx-related *pol* sequences isolated in Brazil. A - Maximum likelihood (ML) phylogenetic tree of the HIV-1 *pol* fragment from CRFs04/09/45/49_cpx, Brazilian sequences and the results of the BLAST search. Branches are colored according to the legend at the bottom left. Sequences from Brazil and those recovered by BLAST closely related to the Brazilian clade are identified using a two letter code on the tip of the branch (RJ – Rio de Janeiro, SP – São Paulo, MG – Minas Gerais, AR – Argentina and IT – Italy). For visual clarity, some clades are collapsed into triangles. The branch support values are indicated as black dots (aLRT > 0.80) or asterisks (aLRT > 0.90) at key nodes. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site and the tree is rooted using HIV-1 subtype B-K reference sequences. Representative bootscanning plots of recombinant sequences are depicted on the right. Query sequences were compared to reference sequences of HIV-as indicated in the legend at top right. B – Map showing the locations where at least one HIV-1 Brazilian CRF45_cpx-related sequence (described in this study or retrieved by BLAST) were isolated. A close view in the distribution of the samples isolated in Rio de Janeiro at county level are depicted in the right. States or counties with sample isolation are colored red.

Figure 2. Schematic representation of the partial and near full-length genomic structure and breakpoints profile of ten CRF45_cpx-related isolates identified in this study. Each genomic fragment is colored in accordance with the legend at bottom and the positions of breakpoints are numbered according to the HIV-1 HXB2 sequence and indicated in the top.

Figure 3. ML tree of the HIV-1 *gag* (A) and *env* (B) gene fragments from the HIV-1 Brazilian, CRF45_cpx and related sequences. Branches are colored according to the legend at the left. Sequences from Brazil, and those recovered by BLAST and/or ML that are closely related to the Brazilian clade are identified using a two letter code on the tip of the branch (RJ – Rio de Janeiro, SP – São Paulo and BE – Belgium). For visual clarity, some clades are collapsed into triangles. The branch support values are indicated as an asterisk (aLRT > 0.90) at key nodes. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site and the tree was rooted using HIV-1 subtype B-K reference sequences.

Figure 4. Time-calibrated maximum clade credibility tree of HIV-1 CRF45_cpx clades identified in the *env, gag* and *pol* ML analyses. Branches in the phylogeny are colored according to HIV-1 clade as indicated at the legend at top-left. The posterior probability is indicated at key nodes as an asterisk (>0.90). In each tree, the node indicating the most recent common ancestor of the Brazilian CRF45_cpx sequences is marked as a circle and the posterior probability density of the estimated age of theses nodes are superimposed in blue.

TABLES

Table 1 – Characteristics of the patient samples included in this study.

Sample	Year of collection	Age (years)	Mode of transmission	City of residence	Subtype identification
05BRRJLTS223	2005	15-24	Heterosexual	Rio de Janeiro*	CRF45_cpx
06BRRJRENK1	2006	35-49	N/A	Rio de Janeiro	CRF45_cpx
07BRRJPG144	2007	15-24	Heterosexual	Rio de Janeiro	CRF45_cpx/B/F1/U
07BRRJPC38	2007	<15	MTC	Nova Iguaçu**	CRF45_cpx/B/U
08BRRJPG226	2008	15-24	Heterosexual	Nova Iguaçu	CRF45_cpx/F1
09BRRJCR018	2009	<15	MTC	Rio de Janeiro	CRF45_cpx
09BRRJCR019	2009	<15	MTC	Rio de Janeiro	CRF45_cpx/B/D/F1/U
09BRRJCR040	2009	<15	MTC	Angra dos Reis	CRF45_cpx
09BRRJPC37	2009	<15	MTC	Queimados	CRF45_cpx
14BRRJRENK2	2014	25-34	Bisexual	Belford Roxo	CRF45_cpx

* City of sample collection. ** Parents from Angola. MTC – Mother-to-child, N/A - not available.

Table 2 - Bayesian estimates of evolutionary parameters of the global and Brazilian HIV-1 CRF45_cpx epidemic.

Gene	μ (substitutions site^{-1} year^{-1})	Coefficient of variation	CRF45_cpx T_{MRCA}	CRF45$_{BR}$ T_{MRCA}
gag	2.15×10^3 ($2.0 \times 10^3 – 3.14 \times 10^3$)	0.33 (0.15 – 0.52)	1974 (1967 – 1981)	1986 (1980 – 1996)
pol	1.58×10^3 ($1.50 \times 10^3 – 1.85 \times 10^3$)	0.49 (0.35 – 0.67)	1971 (1963 – 1978)	1981 (1976 – 1987)
env	4.77×10^3 ($4.0 \times 10^3 – 5.85 \times 10^3$)	0.33 (0.13 – 0.58)	1976 (1967 – 1984)	1985 (1977 – 1993)

The 95% HPD is displayed in parentheses.

SUPPORTING INFORMATION

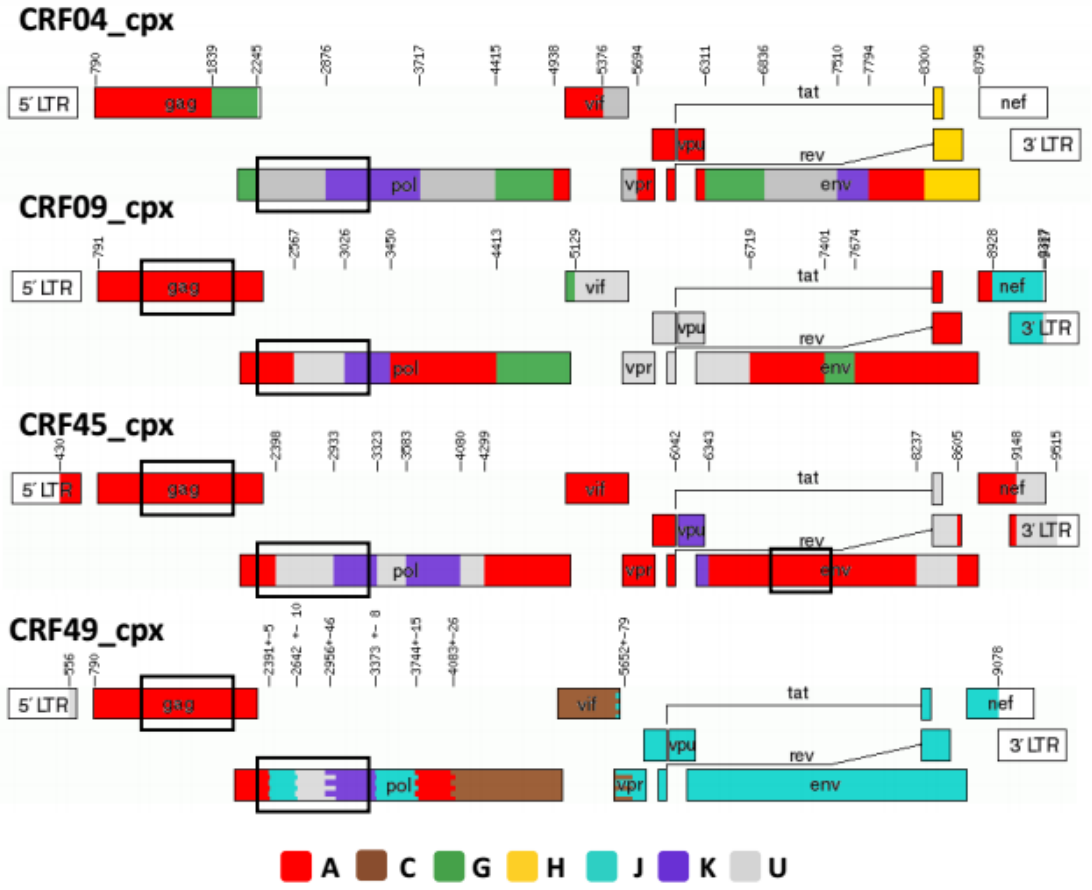
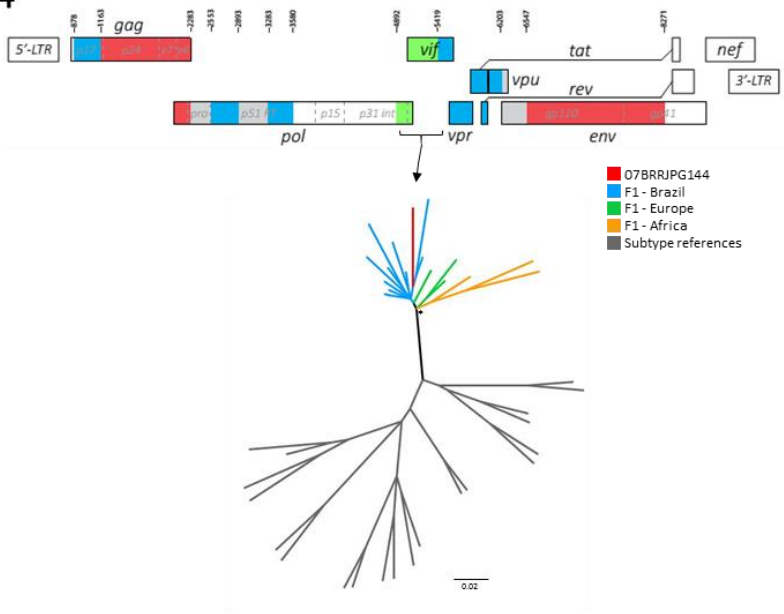
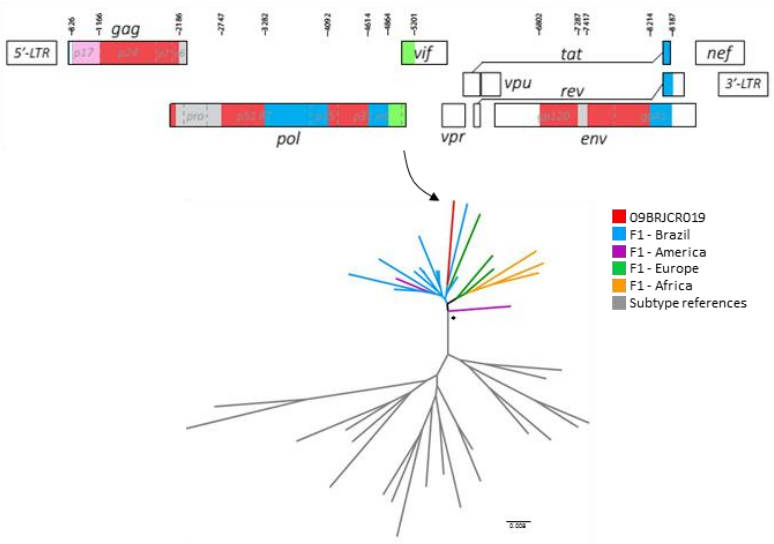


FIGURE S1

07BRRJPG144



09BRJCR019



08BRRJPG226

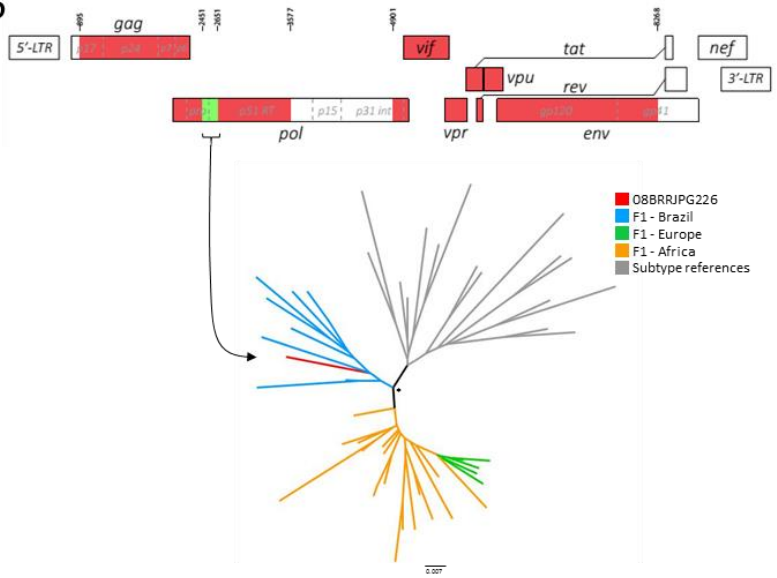


FIGURE S2

Figure S1. Recombination pattern of the circulant recombinant forms analyzed in this study. Boxes representing *gag* (HXB2: 1194 – 2021), *pol* (HXB2: 2253 – 3272) and *env* (HXB2: 6850 – 7350) gene fragments used in this study are superimposed on the graphical illustrations of the CRFs04\09\45\49_cpx genomes based on breakpoint data available in Los Alamos HIV database and colored according to the legend at bottom.

Figure S2. Exploratory NJ trees of the HIV-1 subsubtype F1 fragments identified in the second generation recombinants (SGR). The schematic representations of the NFLG are the same presented in the Figure 2. The HIV-1 subsubtype F1 region analyzed of each SGR is indicated by brackets. Bootstrap values ≥ 70% are indicated by asterisks in key nodes. The scale bar at the bottom of each tree represents nucleotide substitutions per site. The branches are colored by the location of origin of the F1 sequences, as depicted by the legend.

Table S1. HIV-1 *pol*, *gag* and *env* datasets used in this study.

Gene	Classification	*N*
pol	CRF45_cpx-like (This study)	10
	CRF45_cpx-like (BLAST)	25
	CRF04_cpx	13
	CRF09_cpx	42
	CRF45_cpx	10
	CRF49_cpx	8
gag	CRF45_cpx-like (This study)	9
	CRF45_cpx-like (BLAST/ML)	6
	CRF09_cpx	5
	CRF45_cpx	5
	CRF49_cpx	6
	A/A1	417
	A2	12
env	CRF45_cpx-like (This study)	6
	CRF45_cpx-like (BLAST/ML)	4
	CRF45_cpx	4
	A/A1	881
	A2	16

Table S2. HIV-1 *pol*, *gag* and *env* datasets used in the Bayesian analyses.

Gene	HIV-1 clade	Country	*N*	Sampling interval
pol	CRF45_cpx	Angola	2	2010
		Argentina	1	2004
		Belgium	1	2007
		Brazil	10	2005-2014
		Cameroon	1	1997
		Chad	1	2007
		Congo	3	2003
		DRC	8	1985-2007
		France	1	2002
		Gabon	1	1997
		Norway	1	1997
		Poland	1	2004
		Senegal	2	1998-2007
		Spain	2	1999-2008
		Switzerland	1	1999
	CRF49_cpx	Botswana	1	2001
		Gambia	3	1997-2003
		Germany	1	2007
		Nigeria	1	2007
		Senegal	1	2010
		United Kingdom	1	2003
gag	CRF09_cpx	Cote d'Ivoire	1	2000
		Ghana	1	1996
		Senegal	2	1995
		USA	1	1999
	CRF45_cpx	Brazil	10	2005-2014
		Cameroon	1	1997
		Cyprus	3	2005-2007
		DRC	4	1985-2004
		Gabon	1	1997
		Norway	1	1997
	CRF49_cpx	Gambia	5	1997-2003
env	CRF45_cpx	Brazil	7	2005-2014
		Belgium	2	2005-2008
		Cameroon	1	1997
		DRC	2	1997
		Gabon	2	1997

DRC – Democratic Republic of Congo; USA – United States of America

4 DISCUSSÃO

A epidemia global do HIV está evoluindo em direção à um aumento nas taxas de mutações de resistência e diversidade molecular. O aumento da incidência de mutações de resistência as drogas antirretrovirais é influenciado grandemente pela crescente ampliação da disponibilidade de drogas antirretrovirais, que vem ocorrendo tanto em países desenvolvidos quanto em desenvolvimento (UNAIDS 2015a). Em uma revisão recente sobre as tendências geográficas e temporais das mutações transmitidas de resistência às drogas (MTRD), foi estimado que na América Latina houve um aumento anual de 1,1 vez na probabilidade global de MTRDs e de 1,2 vez na probabilidade de MTRDs relacionadas a INNTRs (Rhee et al. 2015). Com relação ao aumento da diversidade, mesmo que o subtipo C continue predominando mundialmente e o subtipo B seja o dominante na Europa Ocidental e Américas, mudanças dinâmicas no perfil de subtipos do HIV circulantes em algumas regiões são impulsionadas por diversos fatores, que incluem crescimento populacional, aumento da migração, recombinação entre diferentes linhagens e efeitos fundadores (Hemelaar 2012; Hemelaar et al. 2011).

O Brasil, desde 1996 oferece gratuitamente medicamentos antirretrovirais e acesso à assistência médica e laboratorial relacionada à infecção pelo HIV. Desde 2013, o Governo Brasileiro recomenda o início imediato da TARV para todos os indivíduos vivendo com HIV/aids, independentemente da contagem de células T $CD4^+$. Como reflexo deste aumento da oferta de antirretrovirais, entre 2004-2014 ocorreu um aumento de 2,5 vezes no número de pessoas vivendo com HIV/aids em tratamento no Brasil, onde atualmente 398.000 pessoas recebem TARV (Brasil 2014). O que se espera com a adoção destas politicas é a redução da morbidade e mortalidade dos indivíduos recebendo a terapia, além da redução das taxas de transmissão do HIV, especialmente a vertical, na qual a quimioprofilaxia exerce um papel fundamental. Entretanto, o sucesso da TARV depende de diversos fatores, tais como o acesso ao tratamento e serviços de saúde, características virais e do sistema imune do hospedeiro, além de características comportamentais do indivíduo, como a aderência ao tratamento (Malta et al. 2005).

O resultado da convergência entre o aumento da disponibilidade de medicamentos antirretrovirais e problemas na aderência de pacientes cria condições favoráveis para que surja um conjunto amplo de vírus resistentes disponíveis para estabelecer novas infecções. Estas cepas do HIV-1 carreando mutações de

resistência podem reduzir a eficiência de futuros regimes terapêuticos dos indivíduos os quais infectaram e/ou terapias profiláticas para a prevenção da transmissão vertical do HIV, desfechos que caracterizam a falha virológica ao tratamento.

Mesmo com uma política bem estabelecida de tratamento da infecção pelo HIV/aids no país, perfis distintos de taxas de mutações transmitidas de resistência são encontrados nas diferentes regiões brasileiras. Estas diferenças podem estar relacionadas com a acessibilidade dos indivíduos aos medicamentos, baixa barreira genética de alguns dos medicamentos prescritos e a forma como os médicos mantem um relacionamento/diálogo com pacientes, experiências que podem impor barreiras ou facilitar a aderência dos indivíduos em tratamento (Malta et al. 2005).

Na avaliação da prevalência de MTRDs realizada em gestantes infectadas pelo HIV-1 acompanhadas num centro de referência para prevenção da transmissão vertical do HIV localizado na Baixada Fluminense, Rio de Janeiro apresentado no ARTIGO 1, encontramos uma prevalência global de MTRDs de 17,2%, o que coloca esta região no mais alto estrato de taxas de resistência transmitida de acordo com a OMS (>15%). Ao atingir estes níveis, a OMS recomenda a realização da testagem de resistência genotípica em indivíduos recém diagnosticados com infecção pelo HIV-1, antes do início do primeiro esquema antirretroviral (World Health Organization 2012), porém cabe ressaltar, que alguns dos critérios recomendados pela OMS para vigilância e monitoramento do nível de resistência do HIV às drogas não foram cumpridos, como a idade máxima de inclusão de 25 anos e contagem de células T $CD4^+$ acima de 500 células/mm^3.

Outros estudos de avaliação da prevalência de MTRDs realizados em escala local, regional e nacional apontaram taxas moderadas, situadas entre 5-15%, mesmo quando o grupo de gestantes foi analisado (Alcântara et al. 2012; Alencar et al. 2013b; Brindeiro et al. 2003; Cardoso et al. 2010; Inocencio et al. 2009; Sprinz et al. 2009). Entretanto, quando se compara os resultados obtidos no ARTIGO 1 com outro trabalho realizado com uma casuística similar no mesmo centro de atendimento cinco anos antes (2005-2008), se nota um aumento na prevalência de MTR (10,7% para 17,2%) porém sem significância estatística. Isto indica que as estimativas obtidas em nosso estudo podem ter sido influenciadas pelo pequeno número de amostras (n = 87) e que na realidade a prevalência de MTRDs se manteve estável entre os dois períodos analisados. Porém, é possível que a influência das novas práticas de tratamento antirretroviral adotadas no Brasil possa ser observada somente no futuro.

Entre as MTRDs encontradas neste trabalho, aquelas relacionadas à classe dos IPs foram as mais prevalentes, atingindo 8,0% das gestantes analisadas. Esta classe compõe o regime terapêutico secundário de tratamento para adultos vivendo com HIV e primário para a quimioprofilaxia da transmissão vertical do HIV (2 ITRN + 1 IP) (Brasil 2012b; Brasil 2010). Considerando as MTRDs encontradas para os IPs, INTRs e INNTRs, fica claro que a eficiência dos esquemas terapêuticos para controle da infecção do HIV amplamente adotados no Brasil pode ser prejudicada, diminuindo o sucesso da profilaxia para prevenção da transmissão vertical, além de permitir que a replicação viral ocorra, o que consequentemente possibilita a emergência de variantes carreando mutações de resistência adicionais que podem ser transmitidas para outros indivíduos.

A alta prevalência de MTRDs encontrada em nossa casuística permite reforçar a necessidade que, ao se considerar as recomendações da OMS, seja indicada a realização de testes de resistência genotípica como rotina antes do início da terapia para que os níveis iniciais de supressão virológica obtidos durante o tratamento sejam os otimizados, garantindo uma queda robusta nos níveis virais plasmáticos. Esta prática se torna ainda mais importante neste momento no qual se busca o início cada vez mais precoce da terapia, principalmente durante a fase de infecção aguda pelo HIV, com o objetivo de reduzir o tamanho dos reservatórios virais precocemente e consequentemente diminuir o prejuízo do sistema imune, para se aproximar da cura funcional da infecção pelo HIV.

Uma das maiores limitações do nosso estudo foi o pequeno número de amostras analisadas, o que pode significar que as frequências de MTRD detectadas neste estudo podem ter sido superestimadas, o que exige cautela na interpretação dos resultados obtidos. Outro fator limitante foi a utilização do sequenciamento pelo método de Sanger para a caracterização das mutações de resistência. Testes convencionais de genotipagem que utilizam este método de sequenciamento raramente detectam variantes que compõem menos de 20% da população de vírus circulantes no plasma. Testes mais sensíveis, capazes de detectar e quantificar variantes minoritários na população do HIV presente no indivíduo infectado surgiram nos últimos anos (Hedskog et al. 2010; Li 2011; Margulies et al. 2005; Mild et al. 2011). As informações obtidas por estas técnicas de sequenciamento de nova geração podem contribuir para a caracterização das populações minoritárias do HIV-1 carreando MTRDs, o que pode melhorar a definição das prevalências de MTR

nesta população que podem comprometer futuros regimes terapêuticos, auxiliando a melhor escolha do esquema terapêutico (Li 2011).

Além do aumento na prevalência de MRTDs, também foi observado um aumento estatisticamente significativo na prevalência de subtipos não-B do HIV-1 quando os resultados da diversidade genética foram comparados com o estudo anterior de Pilotto e colaboradores (2013) realizado no mesmo local. A maior parte dos clados não-B encontrados corresponderam às formas recombinantes, tanto circulantes quanto únicas. Além dos recombinantes BF comumente encontrados compondo a epidemia brasileira, convém salientar as presenças do CRF02_AG, descrito inicialmente na Nigéria (Howard e Rasheed 1996) e que circula em altas prevalências (~50%) na África Ocidental (Hemelaar et al. 2011), e do CRF12_BF, descrito inicialmente no Uruguai e Argentina (Carr et al. 2001) e que circula principalmente na porção meridional da América do Sul, sendo porém raramente encontrado no Brasil (Bello et al. 2011).

Existe uma tendência global no aumento da diversidade do HIV-1, que se reflete no padrão de distribuição dos diferentes subtipos. Embora a distribuição tenha se mantido relativamente sem mudanças entre 2000 e 2007, ocorreram mudanças importantes em algumas regiões e países, como o aumento da prevalência dos clados G e CRF02_AG na Espanha (Lospitao et al. 2005), dos clados D e CRF01_AE na França (Brand et al. 2012) e do subsubtipo F1 na Itália (Lai et al. 2010).

No Brasil, embora a epidemia do HIV-1 seja dominada pelos subtipos B, F1 e na região sul pelo subtipo C, além dos recombinantes entre eles, outros clados do HIV-1 são esporadicamente encontrados, tais como o A, D, CRF02_AG, CRF45_cpx e recombinantes carreando fragmentos do subtipo K (Alencar et al. 2013a; Brindeiro et al. 2002; Caride et al. 2000; Delatorre et al. 2012; Eyer-Silva e Morgado 2007; Ferreira et al. 2010; Ferreira et al. 2013; Inocencio et al. 2009; Machado et al. 2009; Morgado et al. 1998; Pessôa et al. 2015; Pilotto et al. 2013; Pimentel et al. 2013; Velasco-de-Castro et al. 2014).

Englobando os dois trabalhos realizados em nosso laboratório que analisaram a casuística de gestantes acompanhadas na região metropolitana fluminense (Pilotto et al., 2013 e ARTIGO 1), foram encontrados clados raros na epidemia do HIV-1 no Rio de Janeiro (subtipo C) e no Brasil (CRF02_AG e URF_UK). No trabalho de Pilotto (2013), foi observada uma prevalência do subtipo C na ordem de 1%, enquanto no ARTIGO 1 uma prevalência de 4,6% foi encontrada. Embora esta

diferença na prevalência num intervalo de 5 anos não tenha sido estatisticamente significativa, uma tendência de aumento na prevalência deste subtipo nesta casuística e de forma mais abrangente no Rio de Janeiro se torna aparente.

A epidemia do subtipo C do HIV-1 no Brasil está concentrada principalmente nos estados da Região Sul, onde atinge altas prevalências (Gräf e Pinto 2013). Recentemente, o avanço da disseminação deste subtipo em direção ao norte, alcançando diferentes estados das regiões Sudeste, Centro-Oeste e Norte se tornou evidente. Ao agrupar novas sequências do subtipo C obtidas no Rio de Janeiro com sequências provenientes de outros estados brasileiros, o ARTIGO 2 propôs a determinação da dinâmica de disseminação deste subtipo no interior do Brasil.

Os resultados obtidos no ARTIGO 2 confirmam a hipótese de que a maior parte do subtipo C circulante no Brasil faz parte de um clado monofilético provavelmente originado na África Oriental (Bello et al. 2008; Fontella et al. 2008) e que o evento de introdução desta linhagem no Brasil ocorreu mais provavelmente a partir do Burundi em meados da década de 1970 (1974-1976). Trabalhos anteriores, estimaram a data mais provável do ancestral comum mais recente do clado do subtipo C prevalente no Brasil na década de 1960 (Véras et al. 2011) e entre o fim da década de 1970 e início da década de 1980 (Bello et al. 2008; De Oliveira et al. 2010), ou seja, períodos anteriores e posteriores ao estimado em nosso trabalho.

Entretanto a escala temporal traçada no ARTIGO 2 coincide com as datas de origem de outros subclados do subtipo C circulante na África Oriental, encontrados na Etiópia, Quênia, Tanzânia e Uganda (Delatorre e Bello 2012). Neste período ocorreu um grande conflito civil em Burundi, que gerou cerca de 300.000 refugiados. Este pode ter sido um dos fatores que contribuiu para a dispersão do subtipo C nos diferentes países da África Oriental e também para o Brasil. Entretanto, os passos da migração desta linhagem do subtipo C para o Brasil, seja diretamente do Burundi ou com a atuação de outros países como intermediários dos eventos de transmissão ainda necessitam ser melhor esclarecidos.

De acordo reconstrução filogeográfica realizada neste trabalho, o clado do subtipo C predominante no Brasil provavelmente entrou no país pelo estado do Paraná, de onde rapidamente se disseminou para outros estados da região Sul do país e em 10 anos após sua entrada já tinha se disseminado para as regiões Sudeste e Centro-Oeste. Embora tenha sofrido um rápido processo de expansão geográfica, o mesmo não ocorreu com a prevalência deste clado. Outros fatores, além das características genéticas virais e do tempo de introdução parecem ter

influenciado a taxa de expansão deste clado nas diferentes regiões brasileiras, como por exemplo diferenças nas redes de transmissão do HIV-1 ativas, além do tráfego rotineiro e acessibilidade espacial entre as diferentes regiões.

Em adição à introdução do subtipo C proveniente do Burundi que originou o clado majoritário no Brasil, cinco outras introduções deste subtipo foram encontradas no estado do Rio de Janeiro, uma delas proveniente de uma gestante acompanhada no HGNI. Estas introduções se originaram provavelmente de países das regiões Central, Sul e Oriental da África e fornecem indícios da existência de redes de transmissão desconhecidas ligando a epidemia do HIV de diferentes países africanos ao estado do Rio de Janeiro, principalmente sua região metropolitana, com casos detectados em Belford Roxo e São Gonçalo, além da capital.

A presença de clados raros do HIV na região metropolitana do Rio de Janeiro tem sido documentada deste o fim da década de 1990, com a descrição da circulação do subtipo D (Couto-Fernandez et al. 2006; Morgado et al. 1998) e A (Caride et al. 2000) e do primeiro caso CRF02_AG no Brasil (Couto-Fernandez et al. 2005; Pires et al. 2004). Além da circulação na região metropolitana do Rio de Janeiro, o CRF02_AG também já foi encontrado nos municípios mais interioranos de Saquarema (região da Baixada Litorânea) situado ao norte do estado e Angra dos Reis na região Sul fluminense (Delatorre et al. 2012; Eyer-Silva e Morgado 2007).

Em um trabalho anterior do nosso grupo, verificou-se a existência de três linhagens do CRF02_AG circulando no Rio de Janeiro, concentradas na região metropolitana, porém também distribuídas em diferentes regiões do estado (Delatorre et al. 2012). Naquele trabalho foi possível estabelecer uma relação das linhagens circulantes no Rio de Janeiro, como também em outros estados brasileiros com a epidemia do CRF02_AG em países da região Ocidental da África, uma região onde a prevalência deste recombinante pode atingir patamares superiores a 50%. No ARTIGO 3, através da utilização de ferramentas para reconstrução filogenética e filogeográfica mais robustas do que as utilizadas anteriormente, e um conjunto de dados maior, com adição de novas sequências descritas (incluindo um novo isolado encontrado no ARTIGO 1), foi possível confirmar a hipótese da circulação de diferentes linhagens do CRF02_AG no Rio de Janeiro, provavelmente originadas do Golfo do Benin (Gana, Nigéria e Benin), Senegal e Guiné-Bissau.

A provável data de introdução de duas linhagens do CRF02_AG circulantes no Rio de Janeiro foram estimadas em 1985, cerca de 10 anos após a introdução dos clados majoritários dos subtipos C e F1 no Brasil (Bello et al. 2012; Bello et al.

2008; Bello et al. 2007; Delatorre et al. 2013; Fontella et al. 2008). A introdução mais tardia do CR02_AG no Brasil, pode ter sido um fator preponderante no processo de disseminação deste clado do HIV-1 no Brasil, uma vez que ela ocorreu num período no qual a epidemia estava iniciando sua estabilização e muitas das redes de transmissão que promoveram a disseminação dos subtipos do HIV-1 introduzidos mais precocemente poderiam estar próximas a saturação. Entretanto, mesmo com uma circulação que parece limitada a poucas dezenas de indivíduos, as diferentes linhagens do CRF02_AG presentes no Rio de Janeiro já ganharam acesso as populações interioranas do estado, sugerindo a existência de uma alta conectividade entre as epidemias do HIV-1 que circulam entre as diferentes regiões fluminenses.

Além da circulação do subtipo C e do recombinante CRF02_AG já descritas no Rio de Janeiro, alguns estudos relatam a identificação de formas recombinantes carreando fragmentos do subtipo K e porções não classificadas (Brindeiro et al. 2002; Pilotto et al. 2013; Velasco-de-Castro et al. 2014). Além do Rio de Janeiro, estas sequências também já foram isoladas nos estados de Minas Gerais (Ferreira et al. 2010) e São Paulo (Alencar et al. 2013a; Pessôa et al. 2015), todos pertencentes à região Sudeste brasileira. Entre os anos de 2005 e 2014, 10 indivíduos infectados com HIV-1 que exibiam tais características foram identificados a partir de diferentes estudos conduzidos em nosso laboratório, incluindo duas amostras provenientes de gestantes atendidas no HGNI recrutadas no contexto do trabalho de Pilotto e colaboradores (2013), cinco crianças infectados pelo HIV-1 por transmissão vertical (Azevedo 2015; Marina 2008) e indivíduos procurando diagnóstico do HIV em centros de testagem (Velasco-de-Castro et al. 2014). Todos estes trabalhos avaliaram indivíduos residentes na região Metropolitana do Rio de Janeiro.

Através de buscas pode similaridade em banco de dados públicos, sete novas sequências brasileiras com perfis similares de recombinação foram recuperadas, totalizando 17 sequências recombinantes com fragmentos K. Após análises filogenéticas, foi possível reclassificar todas as amostras recombinantes como CRF45_cpx-like, que com exceção de uma, formaram um clado monofilético indicando um evento único de introdução desta linhagem no Brasil.

Após expandir as regiões genômicas analisadas destes isolados do CRF45_cpx, ficou evidente que alguns correspondiam na verdade a recombinantes de segunda geração, que após a introdução recombinaram com os subtipos B e F1, os mais prevalentes no Brasil. É notável que a data mais provável de origem do

clado brasileiro do CRF45_cpx foi estimada por volta da primeira metade de década de 1980 e que nesta escala temporal, mesmo com uma disseminação aparentemente restrita, o CRF45_cpx circulante no Brasil tenha sido capaz de recombinar-se com outros subtipos locais, exemplificando a importância da recombinação como uma força poderosa para a evolução do genoma do HIV-1.

Ao analisar as datas prováveis de origem das linhagens do CRF02_AG e CRF45_cpx estimadas nos ARTIGOS 3 e 4, nota-se uma sobreposição entre as estimativas durante o início da década de 1980. Embora nenhuma circunstância aparente possa ser relacionada com os eventos de migração destas linhagens para o Brasil, este período de tempo coincide com um aumento do número de migrações internacionais na África, ocorrido entre os anos de 1960 e 1980, impulsionado pelos processos de descolonização que se iniciaram e pela piora das situações econômicas e políticas (Zlotnik 2004). Entretanto, a introdução destas linhagens parece ser o resultado de eventos esporádicos e não uma consequência direta da mobilidade de pessoas entre a África e Brasil.

Em conjunto, os dados apresentados sobre clados raros do HIV-1 presentes no Rio de Janeiro, ressaltam a importância deste estado e da casuística de gestantes como sentinelas para o monitoramento da introdução e estabelecimento de novas variantes do HIV-1 de origem africana disseminando-se principalmente pela via heterossexual. O Rio de Janeiro é um dos principais destinos turísticos internacionais do Brasil, além de sediar uma série de eventos socioculturais e econômicos. Dentre os municípios fluminenses nos quais foi descrita a introdução de linhagens Africanas do HIV-1, além daqueles que compõem a região metropolitana, estão Saquarema e Angra dos Reis. Ambas são cidades turísticas que atraem milhares de visitantes domésticos e estrangeiros todos os anos.

Estas características parecem contribuir para a atuação do estado do Rio de Janeiro como um hot-spot para importação de variantes do HIV-1, que podem modificar a dinâmica epidêmica local, tendo como consequência um aumento na diversidade. Esta característica singular da epidemia do HIV-1 no Rio de Janeiro faz com que o contínuo monitoramento da diversidade molecular seja de grande importância para a identificação da emergência e estabelecimento de novas variantes no Brasil que podem impactar a epidemia local e sinalizar para a existência de redes de transmissão internacionais desconhecidas.

5 CONCLUSÕES

- A epidemia do HIV-1 na região metropolitana Rio de Janeiro se modificou entre os períodos de 2005-2008 e 2012-2015, com um aumento significativo na circulação de subtipos não-B (de 19% para 31%) e uma tendência de aumento na prevalência de mutações transmitidas de resistência às drogas (de 10,7 *vs* 17,2%), principalmente para IPs.
- A população de gestantes da região metropolitana do Rio de Janeiro se mostrou um importante grupo sentinela para o monitoramento da entrada e estabelecimento de linhagens importadas do HIV-1 a partir da África no Brasil, com destaque para o subtipo C, CRF02_AG e CRF45_cpx, alvos deste estudo.
- Além da linhagem majoritária do subtipo C que circula pelo Brasil, introduzida a partir do Burundi no estado do Paraná por volta de meados de década de 1970, mais cinco variantes deste subtipo originadas de diferentes regiões africanas foram encontradas no Rio de Janeiro.
- Dentre as cinco linhagens do CRF02_AG originadas a partir da África Ocidental encontradas no Brasil, duas são compostas exclusivamente por sequências do Rio de Janeiro e parecem estar se disseminando no estado por cerca de 30 anos.
- A linhagem do CRF45_cpx existente no Brasil parece ser fruto de um único evento de introdução ocorrido em meados da década de 1980 na região Sudeste, formando um clado monofilético composto majoritariamente por isolados do Rio de Janeiro.
- A circulação de diferentes linhagens do HIV-1 no Rio de Janeiro oriundas do continente africano indica a existência de redes de transmissão com cerca de 30 anos, que estão atuando não só na região metropolitana, mas também adentrando o interior do estado, principalmente através da via heterossexual.
- O contínuo monitoramento da diversidade genética do HIV é justificada e fundamental para a detecção precoce da introdução e disseminação de clados emergentes na epidemia brasileira.

6 PERSPECTIVAS

- Realizar a análise de prevalência de mutações transmitidas de resistência às drogas utilizando sequenciamento de nova geração, capaz de detectar populações virais minoritárias carreando mutações de resistência, permitindo estimativas de prevalência mais acuradas.
- Reconstruir a filodinâmica dos principais clados da linhagem Leste-Africana do subtipo C do HIV-1 circulando globalmente.
- Verificar se houve algum passo intermediário na introdução do CRF02_AG do HIV-1 a partir da África para o Brasil.
- Identificar a origem geográfica da linhagem do CRF45_cpx do HIV-1 introduzida no Brasil.

7 REFERÊNCIAS BIBLIOGRÁFICAS

Abela I a., Berlinger L, Schanz M, Reynell L, Günthard HF, Rusert P, et al. Cell-cell transmission enables HIV-1 to evade inhibition by potent CD4bs directed antibodies. PLoS Pathog. 2012;8(4).

Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. J. Virol. outubro de 2010;84(19):9864–78.

Aguayo N, Laguna-Torres VA, Villafane M, Barboza A, Sosa L, Chauca G, et al. Epidemiological and molecular characteristics of HIV-1 infection among female commercial sex workers, men who have sex with men and people living with AIDS in Paraguay. Rev. Soc. Bras. Med. Trop. junho de 2008;41(3):225–31.

Alcântara KC, Lins JBA, Albuquerque M, Aires LM, Cardoso LPV, Minuzzi AL, et al. HIV-1 mother-to-child transmission and drug resistance among Brazilian pregnant women with high access to diagnosis and prophylactic measures. J. Clin. Virol. Elsevier B.V.; 2012;54(1):15–20.

Alcântara KC, Reis MNG, Cardoso LPV, Bello G, Stefani MMA. Increasing heterosexual transmission of HIV-1 subtype C in Inland Central Western Brazil. J. Med. Virol. março de 2013;85(3):396–404.

Aldrich C, Hemelaar J. Global HIV-1 diversity surveillance. Trends Mol. Med. Elsevier Ltd; dezembro de 2012;18(12):691–4.

Alencar CS, Sabino EC, Carvalho SMF, Leao SC, Carneiro-proietti AB, Capuani L, et al. HIV Genotypes and Primary Drug Resistance Among HIV-Seropositive Blood Donors in Brazil : Role of Infected Blood Donors as Sentinel Populations for Molecular Surveillance of HIV. JAIDS. 2013a;63(3):387–92.

Alencar CS, Sabino EC, Carvalho SMF, Leao SC, Carneiro-Proietti AB, Capuani L, et al. HIV Genotypes and Primary Drug Resistance Among HIV-Seropositive Blood Donors in Brazil. JAIDS J. Acquir. Immune Defic. Syndr. julho de 2013b;63(3):387–92.

Ali A, Bandaranayake RM, Cai Y, King NM, Kolli M, Mittal S, et al. Molecular basis for drug resistance in HIV-1 protease. Viruses. 2010;2(11):2509–35.

Almeida SE, de Medeiros RM, Junqueira DM, Gräf T, Passaes CPB, Bello G, et al. Temporal dynamics of HIV-1 circulating subtypes in distinct exposure categories in southern Brazil. Virol. J. janeiro de 2012;9:306.

Arts EJ, Hazuda DJ. HIV-1 Antiretroviral Drug Therapy. Cold Spring Harb. Perspect. Med. abril de 2012;2(4):a007161.

Aulicino PC, Bello G, Rocco C, Romero H, Mangano A, Morgado MG, et al. Description of the first full-length HIV type 1 subtype F1 strain in Argentina: implications for the origin and dispersion of this subtype in South America. AIDS Res. Hum. Retroviruses. outubro de 2007;23(10):1176–82.

Ayouba A, Akoua-Koffi C, Calvignac-Spencer S, Esteban A, Locatelli S, Li H, et al. Evidence for continuing cross-species transmission of SIVsmm to humans: characterization of a new HIV-2 lineage in rural Côte d'Ivoire. AIDS. 24 de setembro de 2013;27(15):2488–91.

Azevedo SSD. Análise da diversidade genética do HIV-1 e da prevalência de mutações de resistência aos antirretrovirais em uma população de crianças e

adolescentes do estado do Rio de Janeiro. Instituto Oswaldo Cruz - Fiocruz; 2015.

Baeten JM, Chohan B, Lavreys L, Chohan V, McClelland RS, Certain L, et al. HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. J. Infect. Dis. 15 de abril de 2007;195(8):1177–80.

Barin F, Denis F, Allan JS, M'Boup S, Kanki P, Lee TH, et al. SEROLOGICAL EVIDENCE FOR VIRUS RELATED TO SIMIAN T-LYMPHOTROPIC RETROVIRUS III IN RESIDENTS OF WEST AFRICA. Lancet. dezembro de 1985;326(8469–8470):1387–9.

Barre-Sinoussi F, Chermann J, Rey F, Nugeyre M, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science (80-.). 20 de maio de 1983;220(4599):868–71.

Bártolo I, Rocha C, Bartolomeu J, Gama A, Marcelino R, Fonseca M, et al. Highly divergent subtypes and new recombinant forms prevail in the HIV/AIDS epidemic in Angola: New insights into the origins of the AIDS pandemic. Infect. Genet. Evol. 2009;9(4):672–82.

Beer B, Bailes E, Sharp PM, Hirsch VM. Diversity and evolution of primate lentiviruses. In: Kuiken C, Foley B, Hahn B, Korber B, McCutchan F, Marx P, et al., organizadores. Hum. Retroviruses AIDS 1999. Los Alamos, NM: heoretical Biology and Biophysics Group, Los Alamos National Laboratory; 1999. p. 460–74.

Bello G, Afonso JM, Morgado MG. Phylodynamics of HIV-1 subtype F1 in Angola, Brazil and Romania. Infect. Genet. Evol. Elsevier B.V.; julho de 2012;12(5):1079–86.

Bello G, Eyer-Silva W a, Couto-Fernandez JC, Guimarães ML, Chequer-Fernandez SL, Teixeira SLM, et al. Demographic history of HIV-1 subtypes B and F in Brazil. Infect. Genet. Evol. março de 2007;7(2):263–70.

Bello G, Passaes CP, Guimarães ML, Lorete RS, Matos Almeida SE, Medeiros RM, et al. Origin and evolutionary history of HIV-1 subtype C in Brazil. AIDS. 1 de outubro de 2008;22(15):1993–2000.

Bello G, Soares M a, Schrago CG. The Use of Bioinformatics for Studying HIV Evolutionary and Epidemiological History in South America. AIDS Res. Treat. janeiro de 2011;2011:154945.

Beyrer C, Baral SD, Van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet. Elsevier Ltd; 2012;380(9839):367–77.

Beyrer C, Trapence G, Motimedi F, Umar E, Iipinge S, Dausab F, et al. Bisexual concurrency, bisexual partnerships, and HIV among Southern African men who have sex with men. Sex. Transm. Infect. 2010;86(4):323–7.

Brand D, Cazein F, Lot F, Brunet S, Thierry D, Moreau A, et al. Continuous spread of HIV-1 subtypes D and CRF01-AE in France from 2003 to 2009. J. Clin. Microbiol. 2012;50(7):2484–8.

Brasil. Plano nacional de enfrentamento da epidemia de aids e das DST entre gays, HSH e travestis. Brasília, DF - Brasil; 2008.

Brasil. Recomendações para Profilaxia da Transmisão Vertical do HIV e Terapia Antirretroviral em Gestantes. Brasília, DF - Brasil; 2010.

Brasil. Sistema nacional de vigilância em saúde : relatório de situação : Rio de Janeiro. 5. ed. BRASIL M da SS de V em SD de DA e HV, organizador. Brasília, DF -

Brasil; 2012a.

Brasil. Recomendações de terapia antirretroviral para adultos vivendo com HIV / aids no Brasil- 2012 versão preliminar. Ministério da Saúde Brasília, Bras. 2012b;.

Brasil. Boletim Epidemiológico Aids e DST. Brasília, DF - Brasil; 2014.

Brasil. Protocolo Clínico e Diretrizes Terapêuticas para Manejo da Infecção pelo HIV em Adultos. Brasília, DF - Brasil; 2013.

Brindeiro P a., Brindeiro RM, Mortensen C, Hertogs K, De Vroey V, Rubini NPM, et al. Testing genotypic and phenotypic resistance in human immunodeficiency virus type 1 isolates of clade B and other clades from children failing antiretroviral therapy. J. Clin. Microbiol. 2002;40(12):4512–9.

Brindeiro RM, Diaz RS, Sabino EC, Morgado MG, Pires IL, Brigido L, et al. Brazilian Network for HIV Drug Resistance Surveillance (HIV-BResNet): a survey of chronically infected individuals. AIDS. 2003;17(7):1063–9.

Bryceson A, Tomkins A, Ridley D, Warhurst D, Goldstone A, Bayliss G, et al. HIV-2-associated AIDS in the 1970s. Lancet (London, England). 23 de julho de 1988;2(8604):221.

Burke DS. Recombination in HIV: An Important Viral Evolutionary Strategy. Emerg. Infect. Dis. 1997;3(3):253–9.

Cabello M, Junqueira DM, Bello G. Dissemination of nonpandemic Caribbean HIV-1 subtype B clades in Latin America. Aids. 2015;29(4):483–92.

Camacho RJ, Vandamme A-M. Antiretroviral resistance in different HIV-1 subtypes: impact on therapy outcomes and resistance testing interpretation. Curr. Opin. HIV AIDS. março de 2007;2(2):123–9.

Campbell-Yesufu OT, Gandhi RT. Update on human immunodeficiency virus (HIV)-2 infection. Clin. Infect. Dis. 15 de março de 2011;52(6):780–7.

Cardoso LPV, Pereira GAS, Viegas ÂA, Schmaltz LEPR, Stefani MM de A. HIV-1 primary and secondary antiretroviral drug resistance and genetic diversity among pregnant women from central Brazil. J. Med. Virol. março de 2010;82(3):351–7.

Caride E, Brindeiro R, Hertogs K, Larder B, Dehertogh P, Machado E, et al. Drug-resistant reverse transcriptase genotyping and phenotyping of B and non-B subtypes (F and A) of human immunodeficiency virus type I found in Brazilian patients failing HAART. Virology. 15 de setembro de 2000;275(1):107–15.

Carr JK, Avila M, Gomez Carrillo M, Salomon H, Hierholzer J, Watanaveeradej V, et al. Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America. AIDS. 19 de outubro de 2001;15(15):F41-7.

Carrion G, Hierholzer J, Montano S, Alava a, Perez J, Guevara a, et al. Circulating recombinant form CRF02_AG in South America. AIDS Res. Hum. Retroviruses. abril de 2003;19(4):329–32.

Castillo J, Comegna M, Quijada W, Jauvin V, Pinson P, Masquelier B, et al. Surveillance of HIV type 1 drug resistance among naive patients from Venezuela. AIDS Res. Hum. Retroviruses. 2009;25(12):1329–33.

Cavalcanti a. MS, Lacerda HR, De Brito AM, Pereira S, Medeiros D, Oliveira S. Antiretroviral resistance in individuals presenting therapeutic failure and subtypes of the human immunodeficiency virus type 1 in the Northeast Region of Brazil. Mem.

Inst. Oswaldo Cruz. 2007;102(7):785–92.

CDC Centers for Disease Control and Prevention. Pneumocystis pneumonia--Los Angeles. 1981. Morb. Mortal. Wkly. Rep. junho de 1981;30(21):1–3.

CDC Centers for Disease Control and Prevention. Epidemiologic Notes and Reports Pneumocystis carinii Pneumonia among Persons with Hemophilia A. Morb. Mortal. Wkly. Rep. 1982a;31(27):365–7.

CDC Centers for Disease Control and Prevention. Epidemiologic Notes and Reports Possible Transfusion-Associated Acquired Immune Deficiency Syndrome. Morb. Mortal. Wkly. Rep. 1982b;31(48):652–4.

CDC Centers for Disease Control and Prevention. Unexplained immunodeficiency and opportunistic infections in infants--New York, New Jersey, California. Morb. Mortal. Wkly. Rep. 17 de dezembro de 1982c;31(49):665–7.

CDC Centers for Disease Control and Prevention. Opportunistic Infections and Kaposi ' s Sarcoma among Haitians in the United States. Morb. Mortal. Wkly. Rep. 1982d;31(26):353–4.

CDC Centers for Disease Control and Prevention. Current Trends Update on Acquired Immune Deficiency Syndrome (AIDS) -- United States. Morb. Mortal. Wkly. Rep. 1982e;31(37):513–4.

CDC Centers for Disease Control and Prevention. Epidemiologic Notes and Reports Immunodeficiency among Female Sexual Partners of Males with Acquired Immune Deficiency Syndrome (AIDS) -- New York. Morb. Mortal. Wkly. Rep. 1983a;31(52):697–8.

CDC Centers for Disease Control and Prevention. Acquired immune deficiency syndrome (AIDS) in prison inmates--New York, New Jersey. Morb. Mortal. Wkly. Rep. 7 de janeiro de 1983b;31(52):700–1.

CDC Centers for Disease Control and Prevention. HIV and AIDS--United States, 1981-2000. MMWR. Morb. Mortal. Wkly. Rep. 1 de junho de 2001;50(21):430–4.

Chang MW, Torbett BE. Accessory mutations maintain stability in drug-resistant HIV-1 protease. J. Mol. Biol. Elsevier Ltd; 2011;410(4):756–60.

Cheng YC, Dutschman GE, Bastow KF, Sarngadharan MG, Ting RY. Human immunodeficiency virus reverse transcriptase. General properties and its interactions with nucleoside triphosphate analogs. J. Biol. Chem. 1987;262(5):2187–9.

Clavel F, Guetard D, Brun-Vezinet F, Chamaret S, Rey M, Santos-Ferreira M, et al. Isolation of a new human retrovirus from West African patients with AIDS. Science (80-.). 18 de julho de 1986;233(4761):343–6.

Clumeck N, Mascart-Lemone F, De Maubeuge J, Brenez D, Marcelis L. ACQUIRED IMMUNE DEFICIENCY SYNDROME IN BLACK AFRICANS. Lancet. março de 1983;321(8325):642.

De Cock KM, Jaffe HW, Curran JW. The evolving epidemiology of HIV/AIDS. AIDS. junho de 2012;26(10):1205–13.

Coffin J, Haase A, Levy J, Montagnier L, Oroszlan S, Teich N, et al. Human immunodeficiency viruses. Science (80-.). 9 de maio de 1986;232(4751):697–697.

Coffin JM. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science. janeiro de 1995;267(5197):483–9.

Connor EM, Sperling RS, Gelber R, Kiselev P, Scott G, O'Sullivan MJ, et al.

Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. N. Engl. J. Med. 3 de novembro de 1994;331(18):1173–80.

da Costa ZB, de Lima YAR, Martelli CMT, Stefani MMDA. Transmitted HIV resistance among pregnant young women infected with HIV-1 in Brazil. AIDS Patient Care STDs. 2013;27(8):439–41.

Couto-Fernandez JC, Eyer-Silva W a, Guimarães ML, Chequer-Fernandez SL, Grinsztejn B, Delaporte E, et al. Phylogenetic analysis of Brazilian HIV type 1 subtype D strains: tracing the origin of this subtype in Brazil. AIDS Res. Hum. Retroviruses. fevereiro de 2006;22(2):207–11.

Couto-Fernandez JC, Silva-de-Jesus C, Veloso VG, Rachid M, Gracie RSG, Chequer-Fernandez SL, et al. Human immunodeficiency virus type 1 (HIV-1) genotyping in Rio de Janeiro, Brazil: assessing subtype and drug-resistance associated mutations in HIV-1 infected individuals failing highly active antiretroviral therapy. Mem. Inst. Oswaldo Cruz. 2005;100(1):73–8.

Covas DT, Bíscaro TA, Kashima S, Duarte G, Machado AA. High Frequency of the GWG (Pro Trp) Envelope Variant of HIV-1 in Southeast Brazil. J. Acquir. Immune Defic. Syndr. Hum. Retrovirology. setembro de 1998;19(1):74–9.

Craigie R, Bushman FD. HIV DNA integration. Cold Spring Harb. Perspect. Med. 2012;2(7).

D'arc M, Ayouba A, Esteban A, Learn GH, Boué V, Liegeois F, et al. Origin of the HIV-1 group O epidemic in western lowland gorillas. Proc. Natl. Acad. Sci. 17 de março de 2015;112(11):E1343–52.

Das SR, Jameel S. Biology of the HIV Nef protein. Indian J. Med. Res. 2005;121(4):315–32.

Deeks SG, Walker BD. Human Immunodeficiency Virus Controllers: Mechanisms of Durable Virus Control in the Absence of Antiretroviral Therapy. Immunity. 2007;27(3):406–16.

Delatorre E, Couto-Fernandez JC, Guimarães ML, Vaz Cardoso LP, de Alcantara KC, Stefani MM de A, et al. Tracing the origin and northward dissemination dynamics of HIV-1 subtype C in Brazil. PLoS One. janeiro de 2013;8(9):e74072.

Delatorre EO, Bello G. Phylodynamics of HIV-1 subtype C epidemic in East Africa. PLoS One. 2012;7(7):1–10.

Delatorre EO, Bello G, Eyer-Silva WA, Chequer-Fernandez SL, Morgado MG, Couto-Fernandez JC. Evidence of multiple introductions and autochthonous transmission of the HIV type 1 CRF02_AG clade in Brazil. AIDS Res. Hum. Retroviruses. outubro de 2012;28(10):1369–72.

Delgado E, Ríos M, Fernández J, Pérez-Alvarez L, Nájera R, Thomson MM. Identification of a new HIV type 1 BF intersubtype circulating recombinant form (CRF44_BF) in Chile. AIDS Res. Hum. Retroviruses. julho de 2010;26(7):821–6.

Diaz RS, Leal É, Sanabani S, Sucupira MC a, Tanuri A, Sabino EC, et al. Selective regimes and evolutionary rates of HIV-1 subtype B V3 variants in the Brazilian epidemic. Virology. Elsevier Inc.; 2008;381(2):184–93.

Dilernia D a, Gomez AM, Lourtau L, Marone R, Losso MH, Salomón H, et al. HIV type 1 genetic diversity surveillance among newly diagnosed individuals from 2003 to 2005 in Buenos Aires, Argentina. AIDS Res. Hum. Retroviruses. outubro de 2007;23(10):1201–7.

Domingo E, Martín V, Perales C, Grande-Pérez a., García-Arriaza J, Arias a. Viruses as quasispecies: Biological implications. Curr. Top. Microbiol. Immunol. 2006;299:51–82.

Dorr P, Westby M, Dobbs S, Griffin P, Irvine B, Macartney M, et al. Maraviroc (UK-427 , 857), a Potent, Orally Bioavailable, and Selective Small-Molecule Inhibitor of Chemokine Receptor CCR5 with Broad-Spectrum Anti-Human Immunodeficiency Virus Type 1 Activity. Antimicrob. Agents Chemother. 2005;49(11):4721–32.

Douek DC. Disrupting T-cell homeostasis: how HIV-1 infection causes disease. AIDS Rev. 2003;5(3):172–7.

Engelman A, Cherepanov P. The structural biology of HIV-1: mechanistic and therapeutic insights. Nat. Rev. Microbiol. 2012;10(4):279–90.

Esnouf R, Ren J, Ross C, Jones Y, Stammers D, Stuart D. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. Nat. Struct. Biol. 1995;2(4):303–8.

Eyer-Silva W a, Morgado MG. Autochthonous horizontal transmission of a CRF02_AG strain revealed by a human immunodeficiency virus type 1 diversity survey in a small city in inner state of Rio de Janeiro, Southeast Brazil. Mem. Inst. Oswaldo Cruz. novembro de 2007;102(7):809–15.

Fanales-Belasio E, Raimondo M, Suligoi B, Buttò S. HIV virology and pathogenetic mechanisms of infection: a brief overview. Ann. Ist. Super. Sanita. SciELO Public Health; janeiro de 2010;46(1):5–14.

Faria NR, Hodges-Mameletzis I, Silva JC, Rodés B, Erasmus S, Paolucci S, et al. Phylogeographical footprint of colonial history in the global dispersal of human immunodeficiency virus type 2 group A. J. Gen. Virol. 2012;93(4):889–99.

Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. The early spread and epidemic ignition of HIV-1 in human populations. Science (80-.). 3 de outubro de 2014;346(6205):56–61.

Fauci AS, Pantaleo G, Stanley S, Weissman D. Immunopathogenic mechanisms of HIV infection. Ann. Intern. Med. 1996;124(7):654–63.

Fernández MF, Golemba MD, Terrones C, Paradiso P, Nassif JC, Bologna R, et al. Short Communication: Emergence of Novel A/G Recombinant HIV-1 Strains in Argentina. AIDS Res. Hum. Retroviruses. 2015;31(3):293–7.

Ferreira AS, Cardoso LPV, Stefani MM de A. Moderate prevalence of transmitted drug resistance and high HIV-1 genetic diversity in patients from Mato Grosso State, Central Western Brazil. J. Med. Virol. agosto de 2011;83(8):1301–7.

Ferreira FGF, Pinto J a, Kakehasi FM, Cleto S, Tupinambás U, Aleixo a W, et al. Prevalence of primary drug resistance-associated mutations among HIV type 1 vertically Infected children in Belo Horizonte, Brazil. AIDS Res. Hum. Retroviruses. 2010;26(2):229–32.

Ferreira JLP, Rodrigues R, Lança AM, de Almeida VC, Rocha SQ, Ragazzo TG, et al. Transmitted Drug Resistance among People Living with HIV/Aids at Major Cities of Sao Paulo State, Brazil. Adv. Virol. janeiro de 2013;2013:878237.

Fontella R, Soares M a, Schrago CG. On the origin of HIV-1 subtype C in South America. AIDS. 2008;22(15):2001–11.

Frøland SS, Jenum P, Lindboe CF, Wefring KW, Linnestad PJ, Böhmer T. HIV-1 infection in Norwegian family before 1970. Lancet (London, England). 11 de junho de 1988;1(8598):1344–5.

Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science. 4 de maio de 1984;224(4648):500–3.

Galvão-Castro B, Ivo-dos-Santos J, Couto-Fernandez JC, Bongertz V, Chequer-Bou-Habib D, Sion FS, et al. Isolation and antigenic characterization of human immunodeficiency virus (HIV) in Brazil. Mem. Inst. Oswaldo Cruz. dezembro de 1987;82(4):453–6.

Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, et al. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature. 4 de fevereiro de 1999;397(6718):436–41.

Gaspareto KV, Mello FMMDA, Dias JRC, Meneguetti VAF, Storti MEG, Ferreira JLDP, et al. Genetic diversity and primary resistance among HIV-1-positive patients from Maringá, Paraná, Brazil. Rev. Inst. Med. Trop. Sao Paulo. 2012;54(4):207–13.

Geretti AM, Harrison L, Green H, Sabin C, Hill T, Fearnhill E, et al. Effect of HIV-1 subtype on virologic and immunologic response to starting highly active antiretroviral therapy. Clin. Infect. Dis. 1 de maio de 2009;48(9):1296–305.

Giesecke J, Scalia-Tomba G, Håkansson C, Karlsson A, Lidman K. Incubation time of AIDS: progression of disease in a cohort of HIV-infected homo- and bisexual men with known dates of infection. Scand. J. Infect. Dis. 1990;22(4):407–11.

Gilbert MTP, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. The emergence of HIV/AIDS in the Americas and beyond. Proc. Natl. Acad. Sci. U. S. A. 2007;104(47):18566–70.

Gonçalves VLMA, Troiani C, Ribeiro AA, Spir PRN, Gushiken EKK, Vieira RB, et al. Vertical transmission of HIV-1 in the western region of the State of São Paulo. Rev. Soc. Bras. Med. Trop. fevereiro de 2011;44(1):4–7.

Goodsell DS. Illustrating the machinery of life: Viruses. Biochem. Mol. Biol. Educ. 2012;40(5):291–6.

Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf R a, et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N. Engl. J. Med. 10 de dezembro de 1981;305(24):1425–31.

Gräf T, Passaes CPB, Ferreira LGE, Grisard EC, Morgado MG, Bello G, et al. HIV-1 genetic diversity and drug resistance among treatment naïve patients from Southern Brazil: an association of HIV-1 subtypes with exposure categories. J. Clin. Virol. julho de 2011;51(3):186–91.

Gräf T, Pinto AR. The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. Virology. 2013;435(1):170–8.

Guimarães ML, Eyer-Silva WA, Couto-Fernandez JC, Morgado MG. Identification of two new CRF_BF in Rio de Janeiro State, Brazil. AIDS. 30 de janeiro de 2008;22(3):433–5.

Guimarães ML, Vicente ACP, Otsuki K, da Silva RFFC, Francisco M, da Silva FG, et al. Close phylogenetic relationship between Angolan and Romanian HIV-1 subtype F1 isolates. Retrovirology. janeiro de 2009;6:39.

Gürtler LG, Hauser PH, Eberle J, von Brunn A, Knapp S, Zekeng L, et al. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. J. Virol. março de 1994;68(3):1581–5.

Hahn BH. AIDS as a Zoonosis: Scientific and Public Health Implications. Science (80-.). 28 de janeiro de 2000;287(5453):607–14.

Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P. Retroviral intasome assembly and inhibition of DNA strand transfer. Nature. Nature Publishing Group; 2010;464(7286):232–6.

Harm Reduction International. The Global State of Harm Reduction 2014. Stone K, organizador. London, United Kingdom: Harm Reduction International; 2014.

Hedskog C, Mild M, Jernberg J, Sherwood E, Bratt G, Leitner T, et al. Dynamics of HIV-1 quasispecies during antiviral treatment dissected using ultra-deep pyrosequencing. PLoS One. 2010;5(7).

Hemelaar J. The origin and diversity of the HIV-1 pandemic. Trends Mol. Med. Elsevier Ltd; março de 2012;18(3):182–92.

Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global trends in molecular epidemiology of HIV-1 during 2000–2007. Aids. março de 2011;25(5):679–89.

Herschhorn A, Hizi A. Retroviral reverse transcriptases. Cell. Mol. Life Sci. agosto de 2010;67(16):2717–47.

Van Heuverswyn F, Li Y, Bailes E, Neel C, Lafay B, Keele BF, et al. Genetic diversity and phylogeographic clustering of SIVcpzPtt in wild chimpanzees in Cameroon. Virology. 10 de novembro de 2007;368(1):155–71.

Van Heuverswyn F, Li Y, Neel C, Bailes E, Keele BF, Liu W, et al. Human immunodeficiency viruses: SIV infection in wild gorillas. Nature. 9 de novembro de 2006;444(7116):164.

Hirsch VM, Dapolito G, Goeken R, Campbell BJ. Phylogeny and natural history of the primate lentiviruses, SIV and HIV. Curr. Opin. Genet. Dev. dezembro de 1995;5(6):798–806.

Howard TM, Rasheed S. Genomic structure and nucleotide sequence analysis of a new HIV type 1 subtype A strain from Nigeria. AIDS Res. Hum. Retroviruses. 10 de outubro de 1996;12(15):1413–25.

Hu W-S, Hughes SH. HIV-1 Reverse Transcription. Cold Spring Harb. Perspect. Med. 1 de outubro de 2012a;2(10):a006882–a006882.

Hu W-S, Hughes SH. HIV-1 reverse transcription. Cold Spring Harb. Perspect. Med. outubro de 2012b;2(10).

Hudson CP. AIDS in rural Africa: a paradigm for HIV-1 prevention. Int. J. STD AIDS. 1996;7(4):236–43.

Hymes KB, Greene JB, William DC, Cheung T, Prose NS, Ballard H, et al. Kaposi's sarcome in homosexual men - A report of eight cases. Lancet. 1981;318(8247):598–600.

IBGE. Estimativas Da População Dos Municípios Brasileiros. Brasília, DF - Brasil; 2014.

Imaz A, Falcó V, Ribera E. Antiretroviral salvage therapy for multiclass drug-resistant HIV-1-infected patients: from clinical trials to daily clinical practice. AIDS Rev. 2011;13(3):180–93.

Inocencio L a., Pereira A a., Sucupira MC a, Fernandez JCC, Jorge CP, Souza DFC, et al. Brazilian Network for HIV Drug Resistance Surveillance: A survey of individuals recently diagnosed with HIV. J. Int. AIDS Soc. 2009;12(1):1–6.

Karn J, Stoltzfus CM. Transcriptional and Posttranscriptional Regulation of

HIV-1 Gene Expression. Cold Spring Harb. Perspect. Med. 1 de fevereiro de 2012;2(2):a006916–a006916.

Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science. julho de 2006;313(5786):523–6.

Kerr LRFS, Mota RS, Kendall C, Pinho AD a., Mello MB, Guimarães MDC, et al. HIV among MSM in a large middle-income country. AIDS. janeiro de 2013;27(3):427–35.

Kim E-Y, Lorenzo-Redondo R, Little SJ, Chung Y-S, Phalora PK, Maljkovic Berry I, et al. Human APOBEC3 Induced Mutation of Human Immunodeficiency Virus Type-1 Contributes to Adaptation and Evolution in Natural Infection. PLoS Pathog. 2014;10(7):e1004281.

Kim JE, Beckthold B, Chen Z, Mihowich J, Malloch L, Gill MJ. Short communication: identification of a novel HIV type 1 subtype H/J recombinant in Canada with discordant HIV viral load (RNA) values in three different commercial assays. AIDS Res. Hum. Retroviruses. novembro de 2007;23(11):1309–13.

Kohl NE, Emini E a, Schleif W a, Davis LJ, Heimbach JC, Dixon R a, et al. Active human immunodeficiency virus protease is required for viral infectivity. Proc. Natl. Acad. Sci. U. S. A. 1988;85(13):4686–90.

Korber B. Timing the Ancestor of the HIV-1 Pandemic Strains. Science (80-.). 9 de junho de 2000;288(5472):1789–96.

Lai a., Riva C, Marconi a., Balestrieri M, Razzolini F, Meini G, et al. Changing patterns in HIV-1 non-B clade prevalence and diversity in Italy over three decades. HIV Med. 2010;11(9):593–602.

Leal É, Villanova FE. Diversity of HIV-1 Subtype B: Implications to the Origin of BF Recombinants. Martin DP, organizador. PLoS One. 28 de julho de 2010;5(7):e11833.

Lemey P, Pybus OG, Rambaut A, Drummond AJ, Robertson DL, Roques P, et al. The molecular population genetics of HIV-1 group O. Genetics. julho de 2004;167(3):1059–68.

Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M, Vandamme A-M. Tracing the origin and history of the HIV-2 epidemic. Proc. Natl. Acad. Sci. U. S. A. maio de 2003;100(11):6588–92.

De Leys R, Vanderborght B, Vanden Haesevelde M, Heyndrickx L, van Geel A, Wauters C, et al. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. J. Virol. março de 1990;64(3):1207–16.

Li G, Piampongsant S, Faria NR, Voet A, Pineda-Peña A-C, Khouri R, et al. An integrated map of HIV genome-wide variation from a population perspective. Retrovirology. ???; dezembro de 2015;12(1):18.

Li JZ. Low-Frequency HIV-1 Drug Resistance Mutations and Risk of NNRTI-Based Antiretroviral Treatment Failure. JAMA. 6 de abril de 2011;305(13):1327.

Liegler T, Abdel-Mohsen M, Bentley LG, Atchison R, Schmidt T, Javier J, et al. HIV-1 Drug Resistance in the iPrEx Pre-Exposure Prophylaxis Trial. J. Infect. Dis. 2014;210:1217–27.

Lihana RW, Ssemwanga D, Abimiku A, Ndembi N. Update on HIV-1 diversity in Africa: A decade in review. AIDS Rev. 2012;14(2):83–100.

Locatelli S, Peeters M. Non-Human Primates , Retroviruses , and Zoonotic Infection Risks in the Human Population. Nat. Educ. Knowl. 2012;3(10):62.

López-Lopes GIS, Lança AM, de Paula Ferreira JL, Souza LO, de Macedo Brígido LF. Discrepancies of HIV-1 Reverse Transcriptase Resistance Interpretation of Insertions and Deletions between Two Genotypic Algorithms. Intervirology. 2013;56(4):217–23.

Lospitao E, Álvarez a., Soriano V, Holguín A. HIV-1 subtypes in Spain: A retrospective analysis from 1995 to 2003. HIV Med. 2005;6(5):313–20.

de Lourdes Teixeira M, Nafea S, Yeganeh N, Santos E, Isabel Gouvea M, Joao E, et al. High rates of baseline antiretroviral resistance among HIV-infected pregnant women in an HIV referral centre in Rio de Janeiro, Brazil. Int. J. STD AIDS. 9 de dezembro de 2014;0(0):1–7.

Lv Z, Chu Y, Wang Y. HIV protease inhibitors: a review of molecular selectivity and toxicity. HIV. AIDS. (Auckl). 2015;7:95–104.

Machado LF a, Ishak MOG, Vallinoto ACR, Lemos JAR, Azevedo VN, Moreira MRC, et al. Molecular epidemiology of HIV type 1 in northern Brazil: identification of subtypes C and D and the introduction of CRF02_AG in the Amazon region of Brazil. AIDS Res. Hum. Retroviruses. outubro de 2009;25(10):961–6.

Malim MH. APOBEC proteins and intrinsic resistance to HIV-1 infection. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 2009;364(1517):675–87.

Malta M, Magnanini MMF, Mello MB, Pascom ARP, Linhares Y, Bastos FI. HIV prevalence among female sex workers, drug users and men who have sex with men in Brazil: a systematic review and meta-analysis. BMC Public Health. 2010;10:317.

Malta M, Petersen ML, Clair S, Freitas F, Bastos FI. Adherence to antiretroviral therapy: a qualitative study with physicians from Rio de Janeiro, Brazil. Cad. Saude Publica. outubro de 2005;21(5):1424–32.

Marcelin A. Resistance to nucleoside reverse transcriptase inhibitors. In: AM G, organizador. Antiretrovir. Resist. Clin. Pract. Chapter 1. London, United Kingdom: Mediscript; 2006. p. 1–11.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057):376–80.

Marina GC. Perfil de Resistência Genotípica do HIV-1 aos Anti-Retrovirais em uma População de Crianças Infectadas pelo HIV-1 por Transmissão Vertical. Instituto Oswaldo Cruz - Fiocruz; 2008.

Marlink R, Kanki P, Thior I, Travers K, Eisen G, Siby T, et al. Reduced rate of disease development after HIV-2 infection as compared to HIV-1. Science. 1994;265(5178):1587–90.

Martinez-Picado J, Martínez MA. HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: A view from the clinic and ex vivo. Virus Res. 2008;134(1–2):104–23.

Masur H. Opportunistic Infection in Previously Healthy Women. Ann. Intern. Med. 1 de outubro de 1982;97(4):533–9.

Matida LH, Santos NJS, Ramos AN, Gianna MC, da Silva MH, Domingues CSB, et al. Eliminating Vertical Transmission of HIV in São Paulo, Brazil: Progress and Challenges. JAIDS J. Acquir. Immune Defic. Syndr. 2011. p. S164–70.

Matthews T, Salgo M, Greenberg M, Chung J, DeMasi R, Bolognesi D. Enfuvirtide: the first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes. Nat. Rev. Drug Discov. 2004;3(3):215–25.

Menéndez-Arias L. Molecular basis of human immunodeficiency virus type 1 drug resistance: Overview and recent developments. Antiviral Res. 2013;98(1):93–120.

Menezes Succi RC De. Mother-to-child transmission of HIV in Brazil during the years 2000 and 2001: results of a multi-centric study. Cad. saude publica / Minist. da Saude, Fund. Oswaldo Cruz, Esc. Nac. Saude Publica. 2007;23 Suppl 3:S379–89.

Mild M, Hedskog C, Jernberg J, Albert J. Performance of Ultra-Deep Pyrosequencing in Analysis of HIV-1 pol Gene Variation. PLoS One. janeiro de 2011;6(7):e22741.

Moore J. The Puzzling Origins of AIDS. Am. Sci. 2004;92(6):540.

Morgado MG, Guimarães ML, Galvão-Castro B. HIV-1 polymorphism: A challenge for vaccine development - A review. Mem. Inst. Oswaldo Cruz. 2002;97(2):143–50.

Morgado MG, Guimarães ML, Gripp CB, Costa CI, Neves I, Veloso VG, et al. Molecular epidemiology of HIV-1 in Brazil: high prevalence of HIV-1 subtype B and identification of an HIV-1 subtype D infection in the city of Rio de Janeiro, Brazil. Evandro Chagas Hospital AIDS Clinical Research Group. J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 15 de agosto de 1998;18(5):488–94.

Morgado MG, Sabino EC, Shpaer EG, Bongertz V, Brigido L, Guimaraes MD, et al. V3 region polymorphisms in HIV-1 from Brazil: prevalence of subtype B strains divergent from North American/European prototype and detection of subtype F. AIDS Res. Hum. Retroviruses. maio de 1994;10(5):569–76.

Nájera R, Delgado E, Pérez-Alvarez L, Thomson MM. Genetic recombination and its role in the development of the HIV-1 pandemic. AIDS. janeiro de 2002;16 Suppl 4:S3-16.

De Oliveira T, Pillay D, Gifford RJ. The HIV-1 subtype C epidemic in South America is linked to the United Kingdom. PLoS One. 2010;5(2).

Paraskevis D, Lemey P, Salemi M, Suchard M, Van De Peer Y, Vandamme a-M. Analysis of the evolutionary relationships of HIV-1 and SIVcpz sequences using bayesian inference: implications for the origin of HIV-1. Mol. Biol. Evol. dezembro de 2003;20(12):1986–96.

Parker R, Bastos C, Galvão J, Pedrosa JS. A AIDS No Brasil, 1982-1992. Rio de Janeiro: ABIA: IMS/UERJ; 1994.

Patricio FRL, Rutherford GW, Barreto JHS, Rodamilans C, Badaró R. Effectiveness of the prevention of mother-to-child HIV transmission in Bahia, Brazil. Brazilian J. Infect. Dis. Elsevier Editora Ltda; agosto de 2015;(x x):2–6.

Peeters M, Courgnaud V. Overview of Primate Lentiviruses and Their Evolution in Non-human Primates in Africa. In: Kuiken C, Foley B, Freed E, Hahn B, Korber B, Marx P, et al., organizadores. HIV Seq. Compend. 2002. Los Alamos, New Mexico: heoretical Biology and Biophysics Group, Los Alamos National Laboratory; 2002. p. 2–23.

Peeters M, Courgnaud V, Abela B. Genetic diversity of lentiviruses in non-human primates. AIDS Rev. 2001;3:3–10.

Peeters M, Gueye a, Mboup S, Bibollet-Ruche F, Ekaza E, Mulanga C, et al.

Geographical distribution of HIV-1 group O viruses in Africa. AIDS. 15 de março de 1997;11(4):493–8.

Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time. Science (80-.). 15 de março de 1996;271(5255):1582–6.

Van de Perre P, Rouvroy D, Lepage P, Bogaerts J, Kestelyn P, Kayihigi J, et al. Acquired immunodeficiency syndrome in Rwanda. Lancet (London, England). 14 de julho de 1984;2(8394):62–5.

Pessôa R, Carneiro Proietti ABDF, Busch MP, Sanabani SS. Identification of a Novel HIV-1 Circulating Recombinant Form (CRF72_BF1) in Deep Sequencing Data from Blood Donors in Southeastern Brazil. Genome Announc. 2014a;2(3):e00386-14.

Pessôa R, Watanabe JT, Calabria P, Alencar CS, Loureiro P, Lopes ME, et al. Enhanced detection of viral diversity using partial and near full-length genomes of human immunodeficiency virus Type 1 provirus deep sequencing data from recently infected donors at four blood centers in Brazil. Transfusion. 2015;55(5):980–90.

Pessôa R, Watanabe JT, Calabria P, Felix AC, Loureiro P, Sabino EC, et al. Deep Sequencing of HIV-1 near Full-Length Proviral Genomes Identifies High Rates of BF1 Recombinants Including Two Novel Circulating Recombinant Forms (CRF) 70_BF1 and a Disseminating 71_BF1 among Blood Donors in Pernambuco, Brazil. PLoS One. 2014b;9(11):e112674.

Pieniazek D, Peralta JM, Ferreira JA, Krebs JW, Owen SM, Sion FS, et al. Identification of mixed HIV-1/HIV-2 infections in Brazil by polymerase chain reaction. AIDS. novembro de 1991;5(11):1293–9.

Pierson T, McArthur J, Siliciano RF. Reservoirs for HIV-1: mechanisms for viral persistence in the presence of antiviral immune responses and antiretroviral therapy. Annu. Rev. Immunol. 2000;18:665–708.

Pilotto JH, Grinsztejn B, Veloso VG, Velasque LS, Friedman RK, Moreira RI, et al. Moderate prevalence of transmitted drug resistance mutations among antiretroviral-naive HIV-infected pregnant women in Rio de Janeiro, Brazil. AIDS Res. Hum. Retroviruses. 2013;29(4):681–6.

Pimentel VF, Morgado MG, Bello G, Guimarães MDC, Castilho E a, Veloso VG, et al. Temporal trends and molecular epidemiology of HIV type 1 infection in Rio de Janeiro, Brazil. AIDS Res. Hum. Retroviruses. 2013;29(12):1553–61.

de Pina-Araujo IIM, Guimarães ML, Bello G, Vicente ACP, Morgado MG. Profile of the HIV Epidemic in Cape Verde: Molecular Epidemiology and Drug Resistance Mutations among HIV-1 and HIV-2 Infected Patients from Distinct Islands of the Archipelago. PLoS One. 2014;9(4):e96201.

Piot P, Taelman H, Quinn TC, Feinsod FM, Minlangu KB, Wobin O, et al. Acquired immunodeficiency syndrome in a heterosexual population in Zaire. Lancet. 1984;2(8394):65–9.

Pires IL, Soares MA, Speranza FAB, Ishii SK, Vieira MCG, Gouve MIFS, et al. Prevalence of human immunodeficiency virus drug resistance mutations and subtypes in drug-naive, infected individuals in the army health service of Rio de Janeiro,. J. Clin. 8 de janeiro de 2004;42(1):426–30.

Plantier J, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemée V, et al. A new human immunodeficiency virus derived from gorillas. Nat. Med. agosto de 2009;15(8):871–2.

Poorolajal J, Molaeipoor L, Mohraz M, Mahjub H, Ardekani MT, Mirzapour P, et al. Predictors of progression to AIDS and mortality post-HIV infection: a long-term retrospective cohort study. AIDS Care. 20 de julho de 2015;(August):1–8.

Popovic M, Sarngadharan MG, Read E, Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science. 4 de maio de 1984;224(4648):497–500.

Prüss-Ustün A, Wolf J, Driscoll T, Degenhardt L, Neira M, Calleja JMG. HIV Due to Female Sex Work: Regional and Global Estimates. PLoS One. 2013;8(5):1–7.

Quagliarello V. The Acquired Immunodeficiency Syndrome: current status. Yale J. Biol. Med. 1982;55(5–6):443–52.

Quarleri JF, Rubio A, Carobene M, Turk G, Vignoles M, Harrigan RP, et al. HIV type 1 BF recombinant strains exhibit different pol gene mosaic patterns: descriptive analysis from 284 patients under treatment failure. AIDS Res. Hum. Retroviruses. outubro de 2004;20(10):1100–7.

Rambaut A, Posada D, Crandall K a, Holmes EC. The causes and consequences of HIV evolution. Nat. Rev. Genet. janeiro de 2004;5(1):52–61.

Ratner L, Gallo RC, Wong-Staal F. HTLV-III, LAV, ARV are variants of same AIDS virus. Nature. 21 de fevereiro de 1985;313(6004):636–7.

Reeves JD, Doms RW. Human immunodeficiency virus type 2. J. Gen. Virol. junho de 2002;83(Pt 6):1253–65.

Rhee S-Y, Blanco JL, Jordan MR, Taylor J, Lemey P, Varghese V, et al. Geographic and Temporal Trends in the Molecular Epidemiology and Genetic Mechanisms of Transmitted HIV-1 Drug Resistance: An Individual-Patient- and Sequence-Level Meta-Analysis. PLOS Med. 2015;12(4):e1001810.

Richman DD. HIV chemotherapy. Nature. 2001;410(6831):995–1001.

Ríos M, Delgado E, Pérez-Alvarez L, Fernández J, Gálvez P, de Parga EV, et al. Antiretroviral drug resistance and phylogenetic diversity of HIV-1 in Chile. J. Med. Virol. junho de 2007;79(6):647–56.

Robertson DL, Anderson JP, Bradac J a, Carr JK, Foley B, Funkhouser RK, et al. HIV-1 nomenclature proposal. Science (80-.). 7 de abril de 2000;288(5463):55–6.

Ruchansky D, Casado C, Russi JC, Arbiza JR, Lopez-Galindez C. Identification of a new HIV Type 1 circulating recombinant form (CRF38_BF1) in Uruguay. AIDS Res. Hum. Retroviruses. março de 2009;25(3):351–6.

De Sá Filho DJ, Sucupira MC a, Caseiro MM, Casiero MM, Sabino EC, Diaz RS, et al. Identification of two HIV type 1 circulating recombinant forms in Brazil. AIDS Res. Hum. Retroviruses. janeiro de 2006;22(1):1–13.

Sahu GK. Potential Implication of Residual Viremia in Patients on Effective Antiretroviral Therapy. AIDS Res. Hum. Retroviruses. 2015;31(1):25–35.

Sanabani SS, Pastena ÉRDS, da Costa AC, Martinez VP, Kleine-Neto W, de Oliveira ACS, et al. Characterization of partial and near full-length genomes of HIV-1 strains sampled from recently infected individuals in São Paulo, Brazil. PLoS One. janeiro de 2011;6(10):e25869.

Sanabani SS, Pastena ERDS, Neto WK, Martinez VP, Sabino EC. Characterization and frequency of a newly identified HIV-1 BF1 intersubtype circulating recombinant form in São Paulo, Brazil. Virol. J. janeiro de 2010;7:74.

Santa-Marta M, de Brito PM, Godinho-Santos A, Goncalves J. Host Factors

and HIV-1 Replication: Clinical Evidence and Potential Therapeutic Approaches. Front. Immunol. 2013;4(October):343.

Santiago ML, Range F, Keele BF, Li Y, Bailes E, Bibollet-Ruche F, et al. Simian immunodeficiency virus infection in free-ranging sooty mangabeys (Cercocebus atys atys) from the Taï Forest, Côte d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. J. Virol. outubro de 2005;79(19):12515–27.

Santos AF, Sousa TM, Soares EA, Sanabani S, Martinez AM, Sprinz E, et al. Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. AIDS. 24 de outubro de 2006;20(16):2011–9.

Sharp PM, Bailes E, Chaudhuri RR, Rodenburg CM, Santiago MO, Hahn BH. The origins of acquired immune deficiency syndrome viruses: where and when? Philos. Trans. R. Soc. Lond. B. Biol. Sci. 29 de junho de 2001;356(1410):867–76.

Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. Cold Spring Harb. Perspect. Med. setembro de 2011;1(1):a006841.

Sharp PM, Robertson DL, Hahn BH. Cross-species transmission and recombination of "AIDS" viruses. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 29 de julho de 1995;349(1327):41–7.

Sigal A, Kim JT, Balazs AB, Dekel E, Mayo A, Milo R, et al. Cell-to-cell spread of HIV permits ongoing replication despite antiretroviral therapy. Nature. Nature Publishing Group; 2011;477(7362):95–8.

da Silveira AA, Cardoso LPV, Francisco RBL, de Araújo Stefani MM. HIV Type 1 Molecular Epidemiology in pol and gp41 Genes Among Naive Patients from Mato Grosso do Sul State, Central Western Brazil. AIDS Res. Hum. Retroviruses. março de 2012;28(3):304–7.

Silveira J, Santos A, Martínez A. Heterosexual transmission of human immunodeficiency virus type 1 subtype C in southern Brazil. J. Clin. …. Elsevier B.V.; maio de 2012;54(1):36–41.

Simon F, Mauclère P, Roques P, Loussert-Ajaka I, Müller-Trutwin MC, Saragosti S, et al. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat. Med. setembro de 1998;4(9):1032–7.

Simon V, Ho DD, Abdool Karim Q. HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. Lancet. 5 de agosto de 2006;368(9534):489–504.

Soares EAJM, Martínez AMB, Souza TM, Santos AFA, Da Hora V, Silveira J, et al. HIV-1 subtype C dissemination in southern Brazil. AIDS. outubro de 2005;19 Suppl 4:S81-6.

Soares MA, De Oliveira T, Brindeiro RM, Diaz RS, Sabino EC, Brigido L, et al. A specific subtype C of human immunodeficiency virus type 1 circulates in Brazil. AIDS. 2003;17(1):11–21.

Soeiro CMDO, Miranda AE, Saraceni V, Lucena NO De, Talhari S, Ferreira LCDL. Mother-to-child transmission of HIV infection in Manaus, State of Amazonas, Brazil. Rev. Soc. Bras. Med. Trop. outubro de 2011;44(5):537–41.

Sprinz E, Netto EM, Patelli M, Lima JS, Furtado JJD, da Eira M, et al. Primary antiretroviral drug resistance among HIV type 1-infected individuals in Brazil. AIDS Res. Hum. Retroviruses. 2009;25(9):861–7.

Sundquist WI, Kräusslich H-G. HIV-1 Assembly, Budding, and Maturation. Cold Spring Harb. Perspect. Med. julho de 2012;2(7):a006924.

Takehisa J, Kraus MH, Ayouba A, Bailes E, Van Heuverswyn F, Decker JM, et al. Origin and biology of simian immunodeficiency virus in wild-living western gorillas. J. Virol. fevereiro de 2009;83(4):1635–48.

Tang JW, Pillay D. Transmission of HIV-1 drug resistance. J. Clin. Virol. 2004;30(1):1–10.

Tarosso LF, Sanabani SS, Ribeiro SP, Sauer MM, Tomiyama HI, Sucupira MC, et al. HIV Type 1 Subtype BF Leads to Faster CD4(+) T Cell Loss Compared to Subtype B. AIDS Res. Hum. Retroviruses. 6 de setembro de 2013;29(0):1–5.

Taylor BS, Hammer SM. The challenge of HIV-1 subtype diversity. N. Engl. J. Med. 30 de outubro de 2008;359(18):1965–6.

Thompson M a, Aberg J a, Hoy JF, Benson C, Gu HF, Hammer SM, et al. Antiretroviral Treatment of Adult HIV Infection. Jama. 2014;308(4):387–402.

Thomson MM, Villahermosa ML, Vázquez-de-Parga E, Cuevas MT, Delgado E, Manjón N, et al. Widespread circulation of a B/F intersubtype recombinant form among HIV-1-infected individuals in Buenos Aires, Argentina. AIDS. 5 de maio de 2000;14(7):897–9.

Torres-Anjel MJ. Macroepidemiology of the HIVs-AIDS (HAIDS) pandemic. Insufficiently considered zoological and geopolitical aspects. Ann. N. Y. Acad. Sci. 16 de junho de 1992a;653(1 Tropical Vete):257–73.

Torres-Anjel MJ. Macroepidemiology of the HIVs-AIDS (HAIDS) pandemic. Insufficiently considered zoological and geopolitical aspects. Ann. N. Y. Acad. Sci. junho de 1992b;653(1 Tropical Vete):257–73.

UNAIDS. The Gap Report. Geneva, Switzerland; 2014.

UNAIDS. How AIDS changed everything. Geneva, Switzerland; 2015a.

UNAIDS. 90-90-90: Uma meta ambiciosa de tratamento para contribuir para o fim da epidemia de AIDS. Genebra, Suíça; 2015b.

UNAIDS. GLOBAL REPORT - UNAIDS report on the global AIDS epidemic 2013. Geneva, Switzerland; 2013.

Vallari A, Bodelle P, Ngansop C, Makamche F, Ndembi N, Mbanya D, et al. Four New HIV-1 Group N Isolates from Cameroon : Prevalence Continues to Be Low. 2010;26(1).

Vallari A, Holzmayer V, Harris B, Yamaguchi J, Ngansop C, Makamche F, et al. Confirmation of putative HIV-1 group P in Cameroon. J. Virol. fevereiro de 2011;85(3):1403–7.

Vallari AS, Carole M, Larry S, Brennan C. Searching for Rare HIV Strains in Rural Democratic Republic of Congo (2001-2003). Conf. Retroviruses Opportunistic Infect. 2015. p. #234.

Velasco-de-Castro CA, Grinsztejn B, Veloso VG, Bastos FI, Pilotto JH, Fernandes N, et al. HIV-1 diversity and drug resistance mutations among people seeking HIV diagnosis in voluntary counseling and testing sites in Rio de Janeiro, Brazil. PLoS One. janeiro de 2014;9(1):e87622.

Véras NMC, Gray RR, Brígido LFDM, Rodrigues R, Salemi M. High-resolution phylogenetics and phylogeography of human immunodeficiency virus type 1 subtype C epidemic in South America. J. Gen. Virol. 2011;92(Pt 7):1698–709.

Vidal N, Bazepeo SE, Mulanga C, Delaporte E, Peeters M. Genetic characterization of eight full-length HIV type 1 genomes from the Democratic

Republic of Congo (DRC) reveal a new subsubtype, A5, in the A radiation that predominates in the recombinant structure of CRF26_A5U. AIDS Res. Hum. Retroviruses. 2009;25(8):823–32.

Vieira ACBC, Miranda AE, Vargas PRM De, Maciel ELN. HIV prevalence in pregnant women and vertical transmission in according to socioeconomic status, Southeastern Brazil. Rev. Saude Publica. 2011;45(4):644–51.

Wensing AM, Calvez V, Günthard HF, Johnson V a, Paredes R, Pillay D, et al. 2014 update of the drug resistance mutations in HIV-1. Top. Antivir. Med. 2014;22(3):642–50.

Westby M, van der Ryst E. CCR5 antagonists: host-targeted antiviral agents for the treatment of HIV infection, 4 years on. Antivir. Chem. Chemother. 2010;20(5):179–92.

Wilen CB, Tilton JC, Doms RW. HIV: Cell binding and entry. Cold Spring Harb. Perspect. Med. 2012;2(8):1–14.

World Health Organization. Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach. Geneva, Switzerland; 2010.

World Health Organization. World Health Organization global strategy for the surveillance and monitoring of HIV drug resistance: an update. Geneva, Switzerland; 2012.

Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature. 2 de outubro de 2008;455(7213):661–4.

Worobey M, Telfer P, Souquière S, Hunter M, Coleman C a, Metzger MJ, et al. Island biogeography reveals the deep history of SIV. Science. 17 de setembro de 2010;329(5998):1487.

Yabar CA, Acuña M, Gazzo C, Salinas G, Cárdenas F, Valverde A, et al. New Subtypes and Genetic Recombination in HIV Type 1-Infecting Patients with Highly Active Antiretroviral Therapy in Peru (2008–2010). AIDS Res. Hum. Retroviruses. dezembro de 2012;28(12):1712–22.

Yoshida N, Deichmann U. Measurement of Accessibility and Its Applications. J. Infrastruct. Dev. 2009;1(1):1–16.

Zhu T, Korber BT, Nahmias a J, Hooper E, Sharp PM, Ho DD. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature. 5 de fevereiro de 1998;391(6667):594–7.

Zlotnik H. International Migration in Africa: An Analysis Based on Estimates of the Migrant Stock [Internet]. Online J. Migr. Policy Inst. 2004 [citado 20 de setembro de 2015]. Recuperado de: http://www.migrationpolicy.org

8 APÊNDICES

Trabalhos científicos realizados em período concomitante e diretamente relacionados ao tema central desta tese de Doutorado

8.1 APÊNDICE 1

Evidence of Multiple Introductions and Autochthonous Transmission of the HIV Type 1 CRF02_AG Clade in Brazil

Autores: **Edson Oliveira Delatorre**, Gonzalo Bello, Walter A. Eyer-Silva, Saada L. Chequer-Fernandez, Mariza G. Morgado, e José Carlos Couto-Fernandez

Jornal: Aids Research and Human Retroviruses, 2012

AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 28, Number 10, 2012

DOI: 10.1089/aid.2011.0381

Evidence of Multiple Introductions and Autochthonous Transmission of the HIV Type 1 CRF02_AG Clade in Brazil

Edson Oliveira Delatorre,[1,*] Gonzalo Bello,[1,*] Walter A. Eyer-Silva,[2] Saada L. Chequer-Fernandez,[1] Mariza G. Morgado,[1] and José Carlos Couto-Fernandez[1]

Abstract

HIV-1 CRF02_AG is the most prevalent intersubtype recombinant form worldwide. Six HIV-1 samples from patients living in Rio de Janeiro, Brazil, were subtyped as CRF02_AG at the *pol* gene between 2004 and 2011. To trace the origin of these viruses, they were compared with 793 CRF02_AG *pol* sequences of African origin and another four Brazilian CRF02_AG *pol* sequences previously described. Phylogenetic analysis reveals that there have been at least four introductions of the CRF02_AG clade in Brazil, as signified by the presence of four phylogenetically distinct lineages, probably originated from western African countries (Benin, Ghana, and Guinea-Bissau). At least two CRF02_AG Brazilian lineages were successful in getting established and disseminated throughout the Rio de Janeiro state, with evidence of both horizontal and vertical transmission. Continuous epidemiological surveillance of HIV-1 strains circulating in Brazil is of paramount importance to the early detection of newly emerging viral lineages.

Human immunodeficiency virus type 1 (HIV-1) exhibits a high degree of genetic diversity due to the occurrence of mutations and the possibility of recombination events across its genome. The pandemic clade of HIV-1, group M, has been classified into nine genetic subtypes designated A–D, F–H, J, and K, and a large variety of intersubtype recombinant genomes designated unique recombinant forms (URFs) and circulating recombinant forms (CRFs). The intersubtype recombinant viruses are responsible for over 20% of HIV infections worldwide.[1] The most prevalent HIV-1 recombinant clade is the CRF02_AG. This recombinant is the fourth largest variant globally, accounting for nearly 8% of all global infections, and is mainly concentrated in West/Central/North Africa and the Middle East.[1]

The molecular epidemiologic scenario of the HIV-1 epidemic in Brazil is dominated by subtypes B, F1, C, and a large collection of intersubtype recombinant forms among those strains.[2,3] Isolated cases of other HIV-1 group M subtypes have been also described in the country, such as subtype A1 and subtype D.[4–7] A few cases of the CRF02_AG strain have been also identified over the past years, including four individuals in the state of Rio de Janeiro,[8–10] two in the state of São Paulo,[11,12] and one in the state of Para (Amazon region).[13] Of note, two CRF02_AG sequences detected in Rio de Janeiro were identified from a married heterosexual couple who never traveled abroad, thus providing the first molecular evidence of autochthonous horizontal transmission of this lineage in Brazil.[10]

In this study, we report the identification of six new CRF02_AG *pol* sequences in the state of Rio de Janeiro, Brazil. These new Brazilian CRF02_AG strains were compared with CRF02_AG sequences of African origin deposited on public databases and with Brazilian CRF02_AG sequences previously described.

The Brazilian CRF02_AG strains identified in the present study were recovered from six chronically HIV-1-infected patients followed at outpatient clinics from the Public Health System distributed throughout the state of Rio de Janeiro, Brazil. Whole-blood samples from patients were sent for genotyping analysis at the Laboratory of AIDS and Molecular Immunology of the Oswaldo Cruz Institute–Fiocruz between 2004 and 2011.

A fragment of around 1250 pb encompassing the whole protease (PR) and part of the reverse transcriptase (RT) regions of the *pol* gene (positions 2297–3539 of the HXB2 genome) was amplified by using two HIV-1 genotyping systems: ViroSeq HIV-1 Genotyping System (Celera Diagnostic, Abbott Laboratories, USA) and TruGene (Siemens Diagnostics, NY), under conditions recommended by the manufacturers. The new Brazilian CRF02_AG *pol* sequences and four Brazilian CRF02_AG sequences (three from Rio de Janeiro and one from São Paulo) described previously[9,10,12]

[1]Laboratorio de AIDS and Imunologia Molecular, Instituto Oswaldo Cruz–Fiocruz, Rio de Janeiro, Brazil.
[2]Hospital Universitário Gafreé Guinle, Rio de Janeiro, Brazil.
*These authors contributed equally to this work.

were aligned against a set of 793 CRF02_AG reference sequences of African origin gathered from the Los Alamos HIV Database (http://hiv-web.lanl.gov/). The reference data set includes all available CRF02_AG *pol* sequences from those African regions (west, central, and north) with an estimated prevalence of this variant higher than 7%, according to Hemelaar *et al.*[1] The final alignment is available from the authors upon request.

The phylogenetic tree was inferred by the maximum likelihood (ML) method under the GTR + I + Γ_4 nucleotide substitution model, selected using the jModeltest program.[14] The ML tree was reconstructed with program PhyML[15] using an online web server.[16] Heuristic tree search was performed using the SPR branch-swapping algorithm and the reliability of the obtained topology was estimated with the approximate likelihood-ratio test (*aLRT*)[17] based on a Shimodaira–Hasegawa-like procedure.

Phylogenetic analysis of the data revealed that the 10 Brazilian CRF02_AG sequences were distributed in four independent lineages (I to IV) that were intermixed among sequences of African origin (Fig. 1). The Brazilian lineages I and II are composed by only one sequence. Lineage I contain the sequence 06BRRJ34 that was collected in 2006 from a 35-year-old women from the city of Rio de Janeiro. This Brazilian lineage branched (*aLRT* = 0.88) with two sequences from Guinea Bissau and one sequence from Senegal (West Africa) (Fig. 1). Of note, Guinea Bissau is a former Portuguese colony that maintains strong cultural and political relationship with Brazil. Lineage II comprises the sequence HM534195 that was collected in 2009 from an antiretroviral treatment (ART)-naive adult patient newly diagnosed in the state of São Paulo.[12] This sequence clustered (aLTR = 0.76) with strains recovered from Burkina Faso, Côte D'Ivoire, and Mali (West Africa) (Fig. 1).

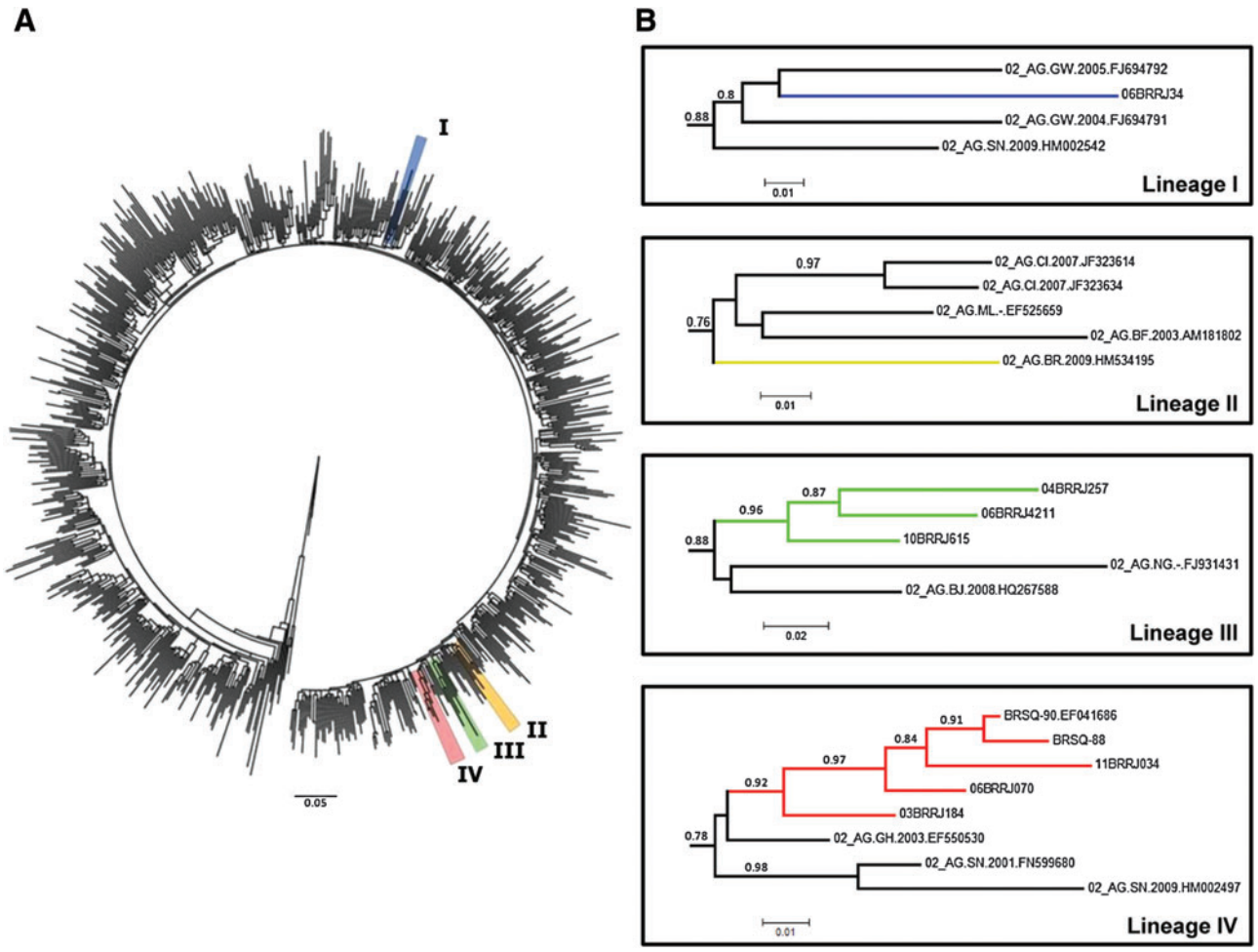


FIG. 1. **(A)** Maximum likelihood (ML) tree of 803 HIV-1 CRF02_AG *pol* (~1250 pb) sequences from African countries and from Brazil (nine from Rio de Janeiro state and one from São Paulo state). The colored boxes highlight the position of the Brazilian CRF02_AG lineages. The tree was rooted using HIV-1 subtype B and C reference sequences. **(B)** A close view of the four Brazilian CRF02_AG lineages (colored branches) and the most closely related sequences of African origin (black branches). The names of the CRF02_AG isolates include reference to country of origin, year of isolation, and GenBank accession number. Countries represented are Benin (BJ), Burkina Faso (BF), Côte D'Ivoire (CI), Ghana (GH), Guinea Bissau (GW), Mali (ML), Nigeria (NG), and Senegal (SN). The approximate likelihood-ratio test (*aLRT*) support values are indicated only at key nodes. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site. Color images available online at www.liebertonline.com/aid

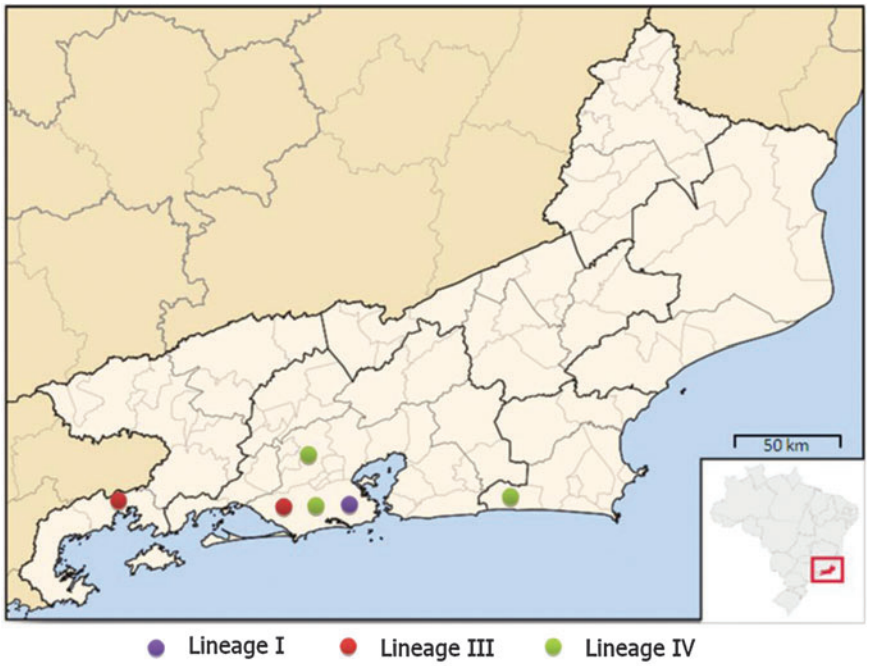


FIG. 2. Political map of Rio de Janeiro state, showing the localization of the three HIV-1 CRF02_AG Brazilian lineages circulating in this state. The maximum intercounty distance found in related sequences was about 200 km. Color images available online at www.liebertonline.com/aid

The Brazilian lineage III is composed by three sequences from Rio de Janeiro (04BRRJ257, 06BRRJ4211, and 10BRRJ615) that clustered together with a high support (*aLRT* = 0.96) (Fig. 1). The sequences 06BRRJ4211 and 04BRRJ257 sampled at 2006 and 2004, respectively, segregate in a subcluster within lineage III and correspond to a mother–child pair, providing the first phylogenetic evidence of autochthonous vertical transmission of the CRF02_AG lineage in Brazil. The third sequence, 10BRRJ615, was collected in 2010 from a 24-year-old pregnant woman. This Brazilian lineage branched (*aLRT* = 0.88) with sequences from Benin and Nigeria (Fig. 1).

The other five sequences from Rio de Janeiro (BRSQ-88, BRSQ-90, 03BRRJ184, 06BRRJ070, and 11BRRJ034) branched in a well-supported monophyletic clade (aLTR = 0.92), here called lineage IV (Fig. 1). The sequences BRSQ-88 and BRSQ-90 were collected in 2006 from a married heterosexual couple from Saquarema, a small city in the Rio de Janeiro north coast.[10] Patient BRSQ-88 was a 28-year-old male who attributed the acquisition of HIV-1 infection to unprotected sexual contacts in the city of Rio de Janeiro, while his 29-year-old wife (patient BRSQ-90) reported having always lived in Saquarema. The other sequences were collected in 2003 (03BRRJ184) from a 37-year-old man, 2006 (06BRRJ070) from a 34-year-old women, and 2011 (11BRRJ034) from a 22-year-old woman. This Brazilian lineage branched (*aLRT* = 0.78) with sequences from Ghana and Senegal (West Africa).

Although patients from Brazilian CRF02_AG clades III and IV had no known direct epidemiologic relationship (with the exception of patients 04BRRJ257/06BRRJ4211 and BRRQ-88/BRSQ-90), the high support of both clusters (aLTR > 0.9) indicates that sequences from each clade were probably recovered from patients who took part in the same chain of viral spread.

Both transmission chains have spread outside the metropolitan region of Rio de Janeiro, reaching small cities located almost 200 km away (Fig. 2). The date of HIV diagnosis (1998–2002) of some individuals from Brazilian clades III and IV reveals that both lineages have been circulating in the state of Rio de Janeiro for over 10 years.

This study demonstrates the existence of at least four independent introductions of the HIV-1 CRF02_AG clade into Brazil, probably from western African countries (Benin, Ghana, and Guinea-Bissau) where this recombinant form is highly prevalent. None of individuals from Rio de Janeiro included in the present study reported a history of travel to African countries, thus indicating that those CRF02_AG infections were acquired in Brazil.

Indeed, we find evidence of the occurrence of at least two autochthonous transmission networks of CRF02_AG, spreading by both horizontal and vertical forms. So far, the dissemination of CRF02_AG seems to be limited to a few individuals in the state of Rio de Janeiro. However, the identification of autochthonous transmission networks involving individuals from different counties with no known direct epidemiologic link demonstrates the potential for larger-scale dissemination of the CRF02_AG variant. These results emphasize the importance of continuous surveillance of HIV-1 genetic diversity to the early detection of newly emerging viral clades in the Brazilian population.

Sequence Data

The GenBank accession numbers of the CRF02_AG HIV-1 Brazilian sequences generated in this study are JQ514094–JQ514099.

Acknowledgments

We thank Carlos Silva de Jesus and Suwellen Sardinha Dias de Azevedo for technical support.

Author Disclosure Statement

No competing financial interests exist.

References

1. Hemelaar J, Gouws E, Ghys PD, and Osmanov S: Global trends in molecular epidemiology of HIV-1 during 2000–2007. AIDS 2011;25(5):679–689.
2. Bongertz V, Bou-Habib DC, Brigido LF, *et al.:* HIV-1 diversity in Brazil: Genetic, biologic, and immunologic characterization of HIV-1 strains in three potential HIV vaccine evaluation sites. Brazilian Network for HIV Isolation and Characterization. J Acquir Immune Defic Syndr 2000;23(2): 184–193.
3. Brindeiro RM, Diaz RS, Sabino EC, *et al.:* Brazilian Network for HIV Drug Resistance Surveillance (HIV-BResNet): A survey of chronically infected individuals. AIDS 2003;17(7): 1063–1069.
4. Morgado MG, Guimaraes ML, Gripp CB, *et al.:* Molecular epidemiology of HIV-1 in Brazil: High prevalence of HIV-1 subtype B and identification of an HIV-1 subtype D infection in the city of Rio de Janeiro, Brazil. Evandro Chagas Hospital AIDS Clinical Research Group. J Acquir Immune Defic Syndr Hum Retrovirol 1998;18(5):488–494.
5. Guimaraes ML, dos Santos Moreira A, Loureiro R, Galvao-Castro B, and Morgado MG: High frequency of recombinant genomes in HIV type 1 samples from Brazilian southeastern and southern regions. AIDS Res Hum Retroviruses 2002; 18(17):1261–1269.
6. Couto-Fernandez JC, Eyer-Silva WA, Guimaraes ML, *et al.:* Phylogenetic analysis of Brazilian HIV type 1 subtype D strains: Tracing the origin of this subtype in Brazil. AIDS Res Hum Retroviruses 2006;22(2):20–211.
7. Caride E, Brindeiro R, Hertogs K, *et al.:* Drug-resistant reverse transcriptase genotyping and phenotyping of B and non-B subtypes (F and A) of human immunodeficiency virus type I found in Brazilian patients failing HAART. Virology 2000;275(1):107–115.
8. Pires IL, Soares MA, Speranza FA, *et al.:* Prevalence of human immunodeficiency virus drug resistance mutations and subtypes in drug-naive, infected individuals in the army health service of Rio de Janeiro, Brazil. J Clin Microbiol 2004; 42(1):426–430.
9. Couto-Fernandez JC, Silva-de-Jesus C, Veloso VG, *et al.:* Human immunodeficiency virus type 1 (HIV-1) genotyping in Rio de Janeiro, Brazil: Assessing subtype and drug-resistance associated mutations in HIV-1 infected individuals failing highly active antiretroviral therapy. Mem Inst Oswaldo Cruz 2005;100(1):73–78.
10. Eyer-Silva WA and Morgado MG: Autochthonous horizontal transmission of a CRF02_AG strain revealed by a human immunodeficiency virus type 1 diversity survey in a small city in inner state of Rio de Janeiro, Southeast Brazil. Mem Inst Oswaldo Cruz 2007;102(7):809–815.
11. Barreto CC, Nishyia A, Araujo LV, Ferreira JE, Busch MP, and Sabino EC: Trends in antiretroviral drug resistance and clade distributions among HIV-1-infected blood donors in Sao Paulo, Brazil. J Acquir Immune Defic Syndr 2006;41(3): 338–341.
12. Brigido LF, Ferreira JL, Almeida VC, *et al.:* Southern Brazil HIV type 1 C expansion into the state of Sao Paulo, Brazil. AIDS Res Hum Retroviruses 2011;27(3):339–344.
13. Machado LF, Ishak MO, Vallinoto AC, *et al.:* Molecular epidemiology of HIV type 1 in northern Brazil: Identification of subtypes C and D and the introduction of CRF02_AG in the Amazon region of Brazil. AIDS Res Hum Retroviruses 2009; 25(10):961–966.
14. Posada D: jModelTest: Phylogenetic model averaging. Mol Biol Evol 2008;25(7):1253–1256.
15. Guindon S and Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 2003;52(5):696–704.
16. Guindon S, Lethiec F, Duroux P, and Gascuel O: PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 2005;33(Web Server issue):W557–559.
17. Anisimova M and Gascuel O: Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol 2006;55(4):539–552.

Address correspondence to:
Edson Oliveira Delatorre and Gonzalo Bello
Laboratorio de AIDS and Imunologia Molecular
Instituto Oswaldo Cruz–Fiocruz
Av. Brasil 4365
21045-900 Rio de Janeiro, RJ
Brazil

E-mail: edsonod@ioc.fiocruz.br; gbello@ioc.fiocruz.br

8.2 APÊNDICE 2

Phylodynamics of HIV-1 Subtype C Epidemic in East Africa

Autores: **Edson Oliveira Delatorre** e Gonzalo Bello

Jornal: PLOS ONE, 2012

OPEN ACCESS Freely available online

PLoS one

Phylodynamics of HIV-1 Subtype C Epidemic in East Africa

Edson Oliveira Delatorre, Gonzalo Bello*

Laboratório de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz, Rio de Janeiro, Brazil

Abstract

The HIV-1 subtype C accounts for an important fraction of HIV infections in east Africa, but little is known about the genetic characteristics and evolutionary history of this epidemic. Here we reconstruct the origin and spatiotemporal dynamics of the major HIV-1 subtype C clades circulating in east Africa. A large number ($n = 1,981$) of subtype C *pol* sequences were retrieved from public databases to explore relationships between strains from the east, southern and central African regions. Maximum-likelihood phylogenetic analysis of those sequences revealed that most (>70%) strains from east Africa segregated in a single regional-specific monophyletic group, here called C_{EA}. A second major Ethiopian subtype C lineage and a large collection of minor Kenyan and Tanzanian subtype C clades of southern African origin were also detected. A Bayesian coalescent-based method was then used to reconstruct evolutionary parameters and migration pathways of the C_{EA} African lineage. This analysis indicates that the C_{EA} clade most probably originated in Burundi around the early 1960s, and later spread to Ethiopia, Kenya, Tanzania and Uganda, giving rise to major country-specific monophyletic sub-clusters between the early 1970s and early 1980s. The results presented here demonstrate that a substantial proportion of subtype C infections in east Africa resulted from dissemination of a single HIV local variant, probably originated in Burundi during the 1960s. Burundi was the most important hub of dissemination of that subtype C clade in east Africa, fueling the origin of new local epidemics in Ethiopia, Kenya, Tanzania and Uganda. Subtype C lineages of southern African origin have also been introduced in east Africa, but seem to have had a much more restricted spread.

Citation: Delatorre EO, Bello G (2012) Phylodynamics of HIV-1 Subtype C Epidemic in East Africa. PLoS ONE 7(7): e41904. doi:10.1371/journal.pone.0041904

Editor: Marco Salemi, University of Florida, United States of America

Received April 23, 2012; **Accepted** June 27, 2012; **Published** July 27, 2012

Funding: This study was supported by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro and Conselho Nacional de Desenvolvimento Científico e Tecnológico grants. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gbellobr@gmail.com

Introduction

Human immunodeficiency virus type 1 (HIV-1) sequences belonging to the pandemic group M are classified into nine subtypes (A–D, F–H, J, and K), six sub-subtypes (A1–A4, and F1–F2), and a variety of inter-subtype recombinant forms (Los Alamos HIV sequence database: http://hiv-web.lanl.gov/). Subtype C is the most prevalent variant, accounting for nearly half (48%) of all global infections [1]. This high prevalence is due to the predominance of subtype C in southern Africa, east Africa and India, with further infections in central Africa and Brazil.

Subtype C accounts for >95% of HIV infections in all southern African countries [1]. Several studies showed that subtype C sequences from neighboring southern African nations display a great degree of phylogenetic intermixing with no evidence of significant geographical clustering [2,3,4,5,6,7], indicating a largely unrestricted viral movement across the entire subcontinent. A more recent phylogenetic study revealed that after sequential pruning of ambiguously positioned taxa 10 strongly supported subtype C clusters becomes apparent in southern Africa, showing that the geographic subdivision of subtype C viruses circulating in this region is higher than expected by chance [8]. Most subtype C clusters identified, however, circulate in more than one southern African country and all four countries analyzed (Botswana, Malawi, South Africa and Zambia) comprise strains from multiple clusters. Thus, HIV epidemics in southern African countries are probably the result of the introduction and circulation of multiple subtype C strains with a variable level of local and regional dissemination.

In contrast to the southern African region, the prevalence of HIV-1 subtype C clade displays a great variation among eastern African countries. Subtype C reaches high prevalence in Burundi (>80%) [9,10], Djibouti (>70%) [11] and Ethiopia (>95%) [12,13,14,15], medium prevalence in Tanzania (20–40%) [16,17,18,19,20], and relatively low prevalence in Rwanda (14%) [21] and Uganda (<5%) [22,23,24,25,26,27,28]. Subtype C also accounts for a minor fraction (<15%) of HIV infections in western [29,30,31], coastal [28,32,33,34] and central [28,35,36,37] regions of Kenya; but displays a much higher frequency (25–50%) in some cities of the northern region that borders Ethiopia [38,39].

Little is known about the genetic characteristics of HIV-1 subtype C strains circulating in east Africa. Previous studies showed that two genetically different subtype C strains designated C and C′, have been co-circulating in roughly similar prevalence and among the same risk groups and geographical areas in Ethiopia [13,15,40]. A recent study of Thomson and Fernández-García [8] revealed that the Ethiopian-C clade corresponds to one subtype C cluster also found in other east African countries including Burundi, Djibouti, Kenya, and Uganda; while the Ethiopian-C′ clade was assigned to an independent cluster

associated to southern Africa. Other studies performed in Kenya showed that subtype C samples from this country are not concentrated in a single cluster, but distributed in several independent lineages associated to sequences from both east and southern Africa [34,39]. Despite these previous studies, we still have an incomplete understanding of the number, onset date, and migration pattern of the distinct HIV-1 subtype C lineages circulating in the eastern African region.

To obtain a more comprehensive picture of the spatiotemporal dynamics of the HIV-1 subtype C epidemic in east Africa, we analyzed a large number of subtype C *pol* sequences sampled from the east (Burundi, Ethiopia, Kenya, Tanzania and Uganda), southern (Botswana, Malawi, Mozambique, South Africa, Zambia and Zimbabwe) and central (Angola and Democratic Republic of Congo) African regions over a time period of 25 years (1986–2010).

Materials and Methods

Sequence dataset

HIV-1 subtype C *pol* sequences from east, southern, and central African countries, that matched the selected genomic region (nt 2253–3272 relative to HXB2 clone) were retrieved from the Los Alamos HIV Database (http://hiv.lanl.gov). Countries were grouped in geographical regions according to the classification proposed by Hemelaar *et al* [1]. In order to improve the accuracy of phylogenetic inference only sequences from antiretroviral therapy naïve individuals were selected. The subtype assignment of all sequences was confirmed by the REGA HIV subtyping tool v.2 [41] and by maximum likelihood (ML) phylogenetic analysis (see below) with HIV-1 subtype reference sequences. Those sequences with incorrect subtype C classification, sequences containing frame-shift mutations or deletions, multiple sequences from the same individual and those sequences from countries poorly represented (<5 sequences) were removed. This resulted in a final dataset of 1,981 HIV-1 subtype C *pol* sequences sampled from 12 different African countries (Table 1). Sequences were aligned using the CLUSTAL X program [42] and alignment is available from the authors upon request.

Table 1. HIV-1 subtype C sequences.

African region	Country	*N*	Sampling date
Central	Angola	31	2001–2010
	Democratic Republic of Congo	22	2002–2007
Southern	Botswana	70	2001
	Malawi	46	2002
	Mozambique	101	2002–2004
	South Africa	1,031	1999–2009
	Zambia	150	1998–2008
	Zimbabwe	178	2007
East	Burundi	92	2002
	Ethiopia	102	1986–2003
	Kenya	39	1991–2007
	Tanzania	81	1997–2009
	Uganda	38	1990–2010

doi:10.1371/journal.pone.0041904.t001

Substitution saturation and likelihood mapping analyses

Substitution saturation was evaluated by plotting the estimated number of transitions and transversions against genetic distance for each pairwise comparison in our alignment of 1,981 HIV-1 subtype C *pol* sequences using DAMBE program [43]. The phylogenetic signal in the *pol* dataset was investigated with the likelihood mapping method [44] by analyzing 10,000 random quartets. Likelihood mapping was performed with TREE-PUZZLE program [45] using the online web platform Phylemon 2.0 [46].

Phylogenetic analysis

ML phylogenetic trees were inferred under the GTR+I+Γ_4 nucleotide substitution model, selected using the jModeltest program [47]. ML tree was reconstructed with PhyML program [48] using an online web server [49]. Heuristic tree search was performed using the SPR branch-swapping algorithm and the reliability of the obtained topology was estimated with the approximate likelihood-ratio test (*aLRT*) [50] based on the Shimodaira-Hasegawa-like procedure. The ML trees were visualized using the FigTree v1.3.1 program [51].

Characterization of intrasubtype C/C′ recombinant sequences

Putative intrasubtype C/C′ recombinant sequences in Ethiopia were identified by Bootscanning using Simplot version 3.5.1 [52], following the same procedure described by Pollakis *et al* [40]. Bootstrap values supporting branching with reference sequences were determined in Neighbor-Joining (NJ) trees constructed using the K2-P nucleotide substitution model, based on 100 re-samplings, with a 300 bp sliding window moving in steps of 10 bases.

Analysis of spatiotemporal dispersion pattern

The evolutionary rate (μ, units are nucleotide substitutions per site per year, subst./site/year), the age of the most recent common ancestor (T_{mrca}, years), and the spatial dynamics of major subtype C clades from east Africa were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach implemented in the BEAST software package v1.6.2 [53,54]. Analyses were performed using the GTR+I+Γ_4 nucleotide substitution model, an uncorrelated Lognormal relaxed molecular clock model [55], a Bayesian Skyline coalescent tree prior [56], and a discrete phylogeographic model in which all possible reversible exchange rates between locations were equally likely [57]. Two separate MCMC chains were run for 4×10^8 generations and adequate chain mixing was checked by calculating the effective sample size (ESS) after excluding an initial 10% for each run using program TRACER v1.4 [58]. MCMC runs converged to almost identical values and combined estimates showed ESS values >200. Maximum clade credibility (MCC) trees were summarized from the posterior distribution of trees with TreeAnnotator and visualized with FigTree v1.3.1. Migratory events were summarized using the cross-platform SPREAD application [59].

Results

Phylogenetic analysis

A large dataset of HIV-1 subtype C *pol* sequences ($n = 1,981$) downloaded from the Los Alamos HIV Database (http://hiv.lanl.gov) was used to characterize the relationship between subtype C sequences sampled from east, central and southern African countries. The transition/transversion *vs* divergence graphics

showed that both type of nucleotide substitutions increase linearly with the genetic distance, with transitions being higher than transversions (Figure S1a), thus indicating no substitution saturation in our alignment. While, the likelihood-mapping analysis showed that most (90%) of the randomly chosen quartets from the HIV-1 subtype C alignment were equally distributed in the three corners of the likelihood map (Figure S1b), indicating a strong tree-like phylogenetic signal in the data. Both analyses indicate that the HIV-1 subtype C *pol* dataset used in this study contains enough evolutionary information for reliable phylogenetic and molecular clock inferences.

The ML phylogenetic analysis revealed that most (73%) subtype C sequences from east Africa branched within a highly supported ($aLRT = 0.93$) monophyletic cluster, here called C_{EA}, that contains sequences from all five east African countries analyzed (Figure 1). Notably, the C_{EA} clade comprises a minor proportion (9%) of the 54 sequences from central Africa, but none of the 1,576 sequences from southern Africa here included. A minor fraction (11%) of subtype C sequences from east Africa branched in a second well supported ($aLRT = 0.94$) monophyletic cluster that comprises sequences from Ethiopia, and corresponds to the so called Ethiopian-C′ (C'_{ET}) clade (Figure 1). The remaining subtype C east African sequences (16%) were distributed in several independent lineages of small size ($n \leq 5$ sequences) that were intermixed among strains from southern African countries (Figure 1).

The analysis of sequence distribution among clades by country of origin revealed three different patterns within east Africa represented by Burundi/Uganda, Ethiopia and Kenya/Tanzania (Figure 2a). All or most subtype C strains circulating in Burundi and Uganda belong to the major clade C_{EA}. Subtype C strains from Ethiopia, by contrast, were mainly distributed into clades C_{EA} (61%) and C'_{ET} (37%). Finally, about 64% of subtype C sequences from Kenya and 49% from Tanzania branched within the major clade C_{EA}, while the remaining sequences were distributed in the multiple minor clades of southern African origin. Such geographical variation in the prevalence of different subtype C clades could be also observed at a more local scale in Tanzania (Fig. 2b). In the Kagera and Mwanza regions (north), most (>70%) subtype C strains belong to the C_{EA} clade. In the Kilimanjaro region (northeast), sequences from both the C_{EA} and "southern African" clades reach a roughly similar prevalence. In the Mbeya region (southwest), only "southern African" clades were detected.

Migration pattern of HIV-1 C_{EA} clade

A closer inspection of the HIV-1 C_{EA} clade showed that sequences from Burundi occupies the most basal position in the clade (Figure S2), thus suggesting that Burundi was the most probable epicenter of dissemination of this subtype C lineage. The migration pattern of the C_{EA} lineage was reconstructed using a Bayesian statistical framework that allows ancestral reconstruction of the locations at the interior nodes of Bayesian tree while accommodating phylogenetic uncertainty. Sequences with no information about sampling date ($n = 2$), sequences with unexpectedly long branches in the phylogenetic analysis ($n = 10$), and Ethiopian sequences with evidence of intra-subtype recombination ($n = 8$, see below) were excluded from this analysis. This resulted in a final dataset of 236 sequences (Burundi = 92, Ethiopia = 47, Kenya = 24, Tanzania = 40, and Uganda = 33) sampled between 1990 and 2010.

The Bayesian MCC tree supports the hypothesis that the C_{EA} clade originated in Burundi ($PP = 1$) and was later exported to the other east African countries where it further spread, establishing new local epidemics (Figures 3 and 4). Estimation of viral movement among countries, obtained by counting the state changes along the tree nodes, points to the role of Burundi as the most important hub of dissemination of this subtype C lineage in east Africa, followed by Tanzania (Table 2). Several migration events of the lineage C_{EA} from Burundi to Ethiopia ($n = 4$), Kenya ($n = 5$), Tanzania ($n = 8$) and Uganda ($n = 8$) were detected, as well as from Tanzania to both Kenya ($n = 3$) and Uganda ($n = 7$). Importation of the C_{EA} lineage into Burundi from other east African countries, and viral exchanges between Ethiopia, Kenya and Uganda were seldom detected in our dataset.

The Bayesian analysis also supports an important phylogeographic subdivision within the C_{EA} lineage. Consistent with the ML topology (Figure S2), most subtype C sequences from Ethiopia, Kenya, Tanzania and Uganda branched in country-specific monophyletic sub-clusters that most probably ($PP \geq 0.93$) had a Burundian origin (Fig. 3). The C_{ET1} and C_{ET2} lineages, that correspond to the so called Ethiopian-C clade, comprise 44% of all Ethiopian sequences here included and were almost exclusively composed by sequences from this country. The C_{KE} and C_{UG} lineages comprise 33% and 37% of all sequences from Kenya and Uganda, respectively, and their circulation seems to be mainly restricted to those countries. Finally, the C_{TZ} lineage comprises 39% of all Tanzanian sequences analyzed and has also been disseminated to Kenya and Uganda. Both ML and Bayesian analyses further suggest that the C_{EA} clade branched in two major sub-clades: one composed by sequences from Burundi and lineages C_{ET1}, C_{ET2} and C_{UG}; the other one composed by sequences from Burundi and lineages C_{KE} and C_{TZ}. The statistical support of such major sub-clades in Bayesian analysis, however, was not significant ($PP < 0.50$) and this observation should be interpreted with caution.

Time-scale of the HIV-1 C_{EA} clade

The median estimated evolutionary rate for the *pol* region of the C_{EA} clade was 1.8×10^{-3} (95% highest posterior density [HPD]: 1.1×10^{-3}–2.4×10^{-3}) subst/site/year, similar to that previously estimated for HIV-1 subtype C lineages circulating in South America [60] and southern Africa [6]. Importantly, the coefficient of rate variation was higher than zero (0.26 [95% HPD: 0.21–0.31]), thus demonstrating a significant variation of substitution rate among branches in the C_{EA} clade and supporting the use of a relaxed molecular clock model to reconstruct the time-scale of this lineage. According to this analysis the C_{EA} clade started to spread in Burundi at 1962 (95% HPD: 1942–1975), while major sub-clades C_{ET1}/C_{ET2}, C_{KE}, C_{TZ} and C_{UG} began to expand in Ethiopia, Kenya, Tanzania and Uganda, respectively, by the early 1970s (Figure 3).

Time-scale of the HIV-1 subtype C Ethiopian clades

The time-scale of the two major Ethiopian clades (C_{ET} and C'_{ET}) was also estimated by combining all sequences from this country in a single dataset and incorporating the posterior distribution of the substitution rate previously estimated for the C_{EA} lineage as an informative prior. This analysis resulted in a Bayesian MCC tree in which clades C_{ET} and C'_{ET} were poorly supported ($PP < 0.5$) and several strains branched outside those major clades (Figure S3). A careful exploration of Ethiopian sequences, revealed that some strains initially classified within clades C_{ET} ($n = 8$) or C'_{ET} ($n = 10$) and those strains that branched outside major Ethiopian clades ($n = 2$) are putative C/C′ intrasubtype recombinant viruses (Figure S3). When those viruses were excluded, the clades C_{ET} and C'_{ET} segregate in two highly supported ($PP > 0.9$) reciprocally monophyletic groups (Figure 5). According to this new Bayesian MCC tree, the median T_{mrca} was

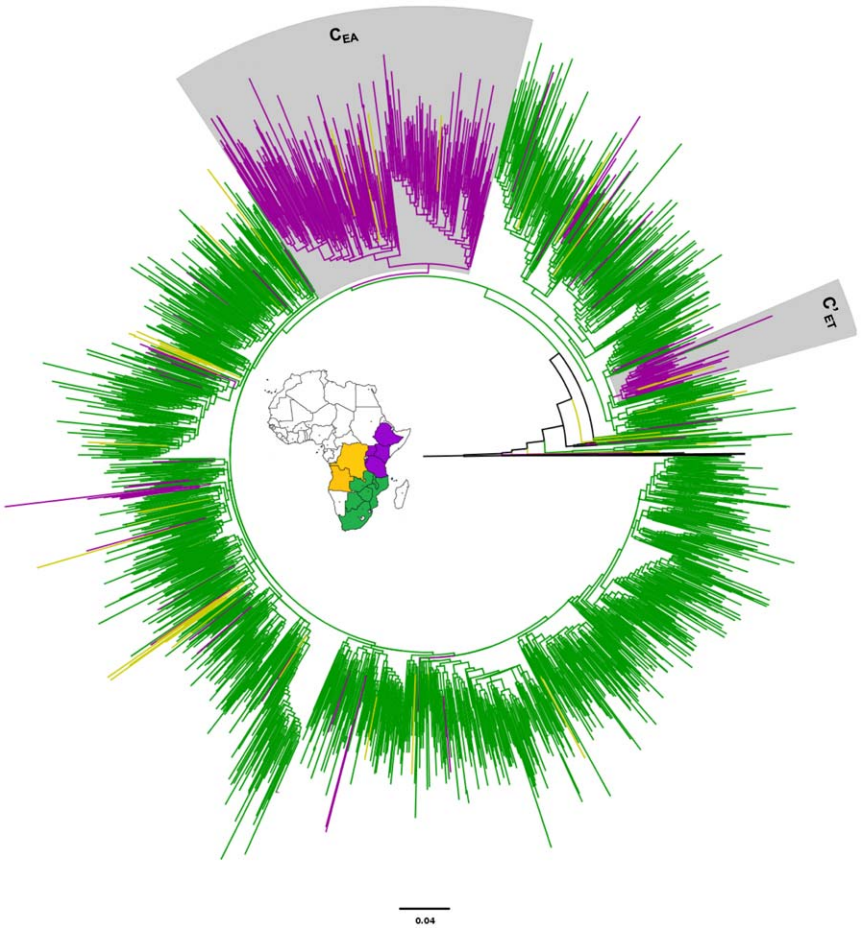


Figure 1. Maximum likelihood phylogenetic tree based on 1,981 HIV-1 subtype C *pol* (~1,000 pb) sequences. Sequences were sampled at different countries from the east ($n = 352$), central ($n = 53$) and southern ($n = 1,576$) African regions shown in Table 1. The color of branches represents the geographic region from where the subtype C sequences originated, according to the map given in the figure. The boxes highlight the position of the major east African subtype C lineages. The tree was rooted using HIV-1 subtype A1 and D reference sequences (black branches). Horizontal branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.
doi:10.1371/journal.pone.0041904.g001

estimated at 1978 for clade C_{ET}, 1981 for sub-clade C_{ET1}, 1984 for sub-clade C_{ET2}, and 1981 for clade C'_{ET} (Figure 5).

Discussion

This study demonstrates a significant phylogeographic subdivision of HIV-1 subtype C strains circulating in the east respect to those circulating in the central and southern African regions, consistent with a recent study [8]. Most (73%) subtype C sequences from east Africa analyzed in this study branched within a highly supported monophyletic clade, here called C_{EA}, that comprise 100% of subtype C sequences from Burundi, 97% from Uganda, 64% from Kenya, 61% from Ethiopia, and 49% from Tanzania. This major east African clade also comprises a minor proportion (<10%) of sequences from central Africa, but no sequence from southern Africa, thus indicating that its circulation is mainly restricted to the east African region. Of note, the genealogies previously inferred for HIV-1 subtypes A and D also support a model of limited introduction of each subtype into east Africa, followed by a subsequent local expansion [61].

Our phylogeographic study suggests that the C_{EA} clade most probably originated in Burundi and after a period of local expansion, this viral lineage was disseminated at multiple times to Ethiopia, Kenya, Tanzania and Uganda, where it generated new local epidemics. Several introductions of the C_{EA} lineage from Tanzania into both Kenya and Uganda were also detected, while viral exchanges between Ethiopia, Kenya and Uganda were less frequent. Five major country-specific monophyletic sub-clusters were detected within the C_{EA} clade that comprise 44%, 33%, 37%, and 39% of all sequences from Ethiopia, Kenya, Uganda and Tanzania here included, respectively. Thus, despite frequent viral movement among east African countries, a significant proportion of subtype C infections in Ethiopia, Kenya, Tanzania and Uganda most likely resulted from the expansion of a few ancestral C_{EA} strains.

It has been suggested that interconnectivity between population centers was a critical factor in the spread of HIV-1 subtypes A and D across Africa [61]. The restricted circulation of the C_{EA} lineage in southern African countries is consistent with this model, considering the relative inaccessibility between the principal population centers of eastern and southern African regions. This model, however, is not consistent with the proposed role of Burundi as the main hub of dissemination of the C_{EA} clade in the region, as this small country is poorly interconnected to other east African countries. Previous studies have also shown a strongly supported phylogenetic relationship between subtype C sequences

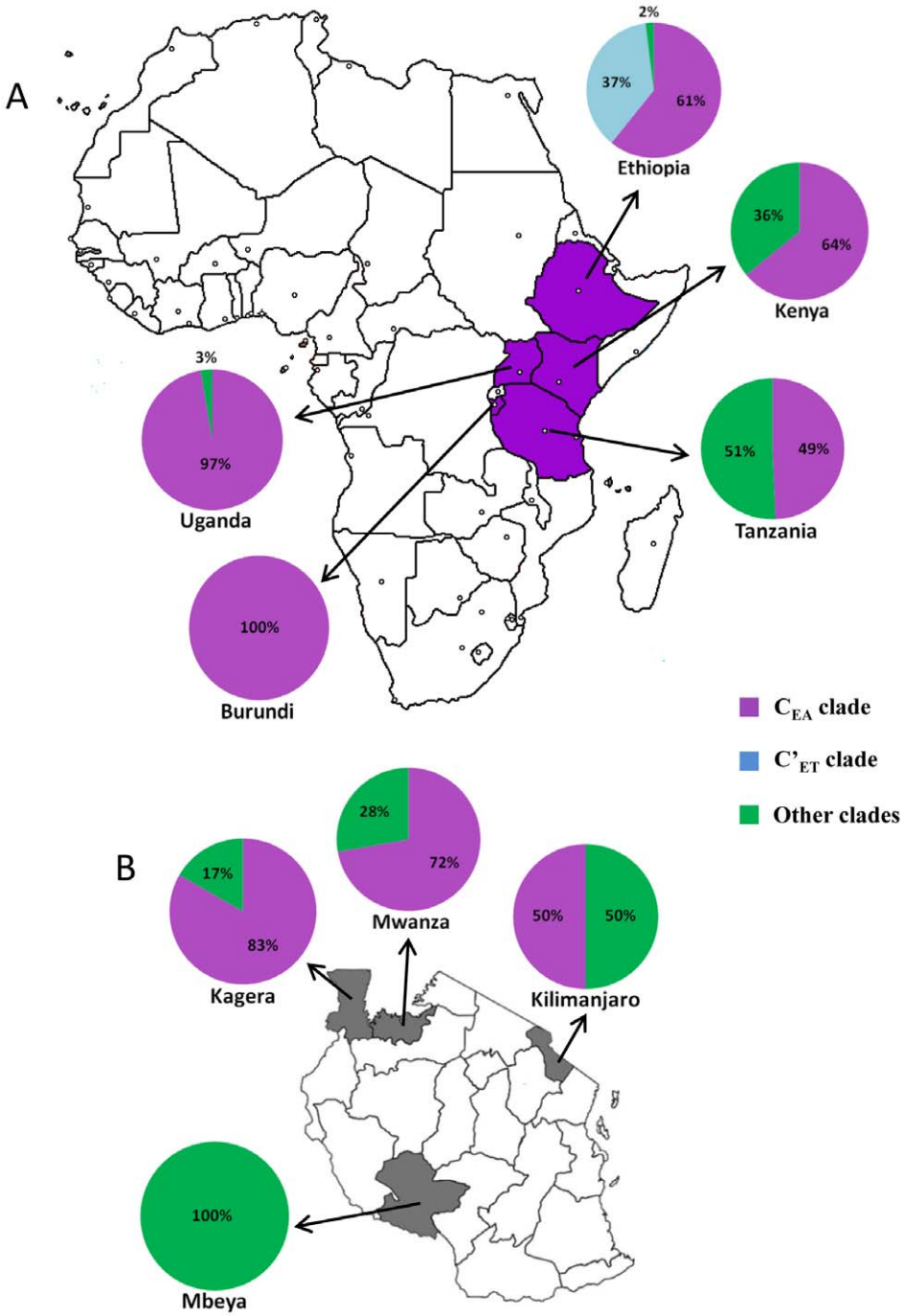


Figure 2. Geographic distribution of HIV-1 subtype C clades in east Africa. a) Map of Africa showing the frequency of distinct HIV-1 subtype C clades across the five countries from the east region here studied (Burundi, Ethiopia, Kenya, Uganda and Tanzania). b) Map of Tanzania showing the frequency of distinct HIV-1 subtype C clades across different country regions where patients included in the present study resided (Kagera, Mwanza, Kilimanjaro and Mbeya). The legend for the colors on graphics is shown on the right.
doi:10.1371/journal.pone.0041904.g002

from Brazil, the UK, Burundi and Kenya; thus indicating that the C_{EA} clade has also been disseminated to South America and Europe [60,62,63]. These evidences suggest that factors other than accessibility may have shaped the dissemination of the C_{EA} clade at both local and global scale.

Burundi has known many violent ethnic conflicts mainly since the 1960s that resulted in large migration flows. Two major civil conflicts that took place in Burundi in 1972 and 1993 generated especially large human movements with the former producing around 300,000 refugees and the latter producing about 687,000 [64]. Most refugees initially crossed the border of their country in the east, fleeing to neighboring Tanzania, followed by movement into other neighboring African countries and later to Europe and North America. It has been estimated that there are about 200,000 Burundians currently living in Tanzania, 18,000 in the Democratic Republic of the Congo, 4,000 in Uganda, 10,000 in the

Figure 3. Time-scaled Bayesian MCC tree of the HIV-1 C_{EA} lineage. Branches are colored according to the most probable location state of their descendent nodes. The legend for the colors is shown on the left. The state posterior probability is indicated only at key nodes. The boxes highlight the position of the major country-specific sub-clades detected in our study. The median age (with 95% HPD interval in parentheses) of those country-specific sub-clades is shown. Horizontal branch lengths are drawn to scale with the bar at the bottom indicating years. The tree was automatically rooted under the assumption of a relaxed molecular clock.
doi:10.1371/journal.pone.0041904.g003

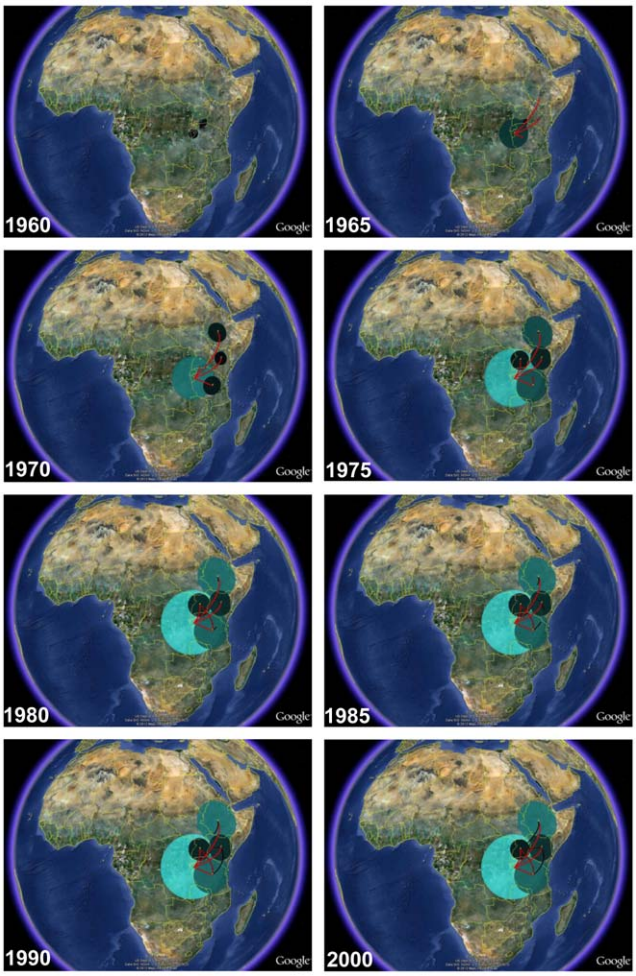


Figure 4. Spatiotemporal dynamic of HIV-1 C_{EA} clade dissemination in east Africa. We provide snapshots of the dispersal pattern for the years 1960, 1965, 1970, 1975, 1980, 1985, 1990 and 2000. Lines between locations represent branches in the Bayesian MCC tree along which location transition occurs. Location circle diameters are proportional to square root of the number of Bayesian MCC branches maintaining a particular location state at each time-point. The white-green color gradient informs the relative age of the transitions (older-recent). The maps are based on satellite pictures made available in Google™ Earth (http://earth.google.com).
doi:10.1371/journal.pone.0041904.g004

European Union, and about 3,000 in the USA and Canada [64]. The molecular clock analysis clade traced the origin of the C_{EA} lineage in Burundi to the early 1960s, while the onset date of the major sub-clades circulating in Ethiopia, Kenya, Tanzania and Uganda was estimated at around the early 1970s, coinciding with the first large Burundian migration flow. These analyses support the notion that the Burundian migration flow occurring in 1972 may have played a fundamental role in the regional and international dissemination of the C_{EA} clade.

While subtype C epidemic in Burundi and Uganda is largely dominated by the C_{EA} clade, a second major subtype C lineage is also circulating in Ethiopia. Our results showed that the two Ethiopian lineages previously designated C and C′ [13], resulted from independent founder strains originated in the eastern and southern African regions, respectively, and further confirmed the circulation of intra-subtype C/C′ recombinants in Ethiopia [40]. The prevalence of C/C′ recombinant viruses estimated in our dataset (20%) was equal to the percentage found in the general Ethiopian population [40]. The onset date of Ethiopian clades C and C′ was dated to between the early 1970s and the early 1980s; consistent with previous estimations [65,66,67].

Table 2. Estimated number of migration events of HIV-1 C_{EA} clade among east African countries.

From/To	Burundi	Ethiopia	Kenya	Tanzania	Uganda
Burundi	-	4	5	8	8
Ethiopia	0	-	0	0	1
Kenya	0	0	-	1	1
Tanzania	0	0	3	-	7
Uganda	0	0	0	0	-

doi:10.1371/journal.pone.0041904.t002

A large collection of minor subtype C lineages of southern African origin were detected in Kenya and Tanzania, which together represent 36% and 51% of sequences from those countries here analyzed, respectively. These lineages seem to have a more restricted expansion than the C_{EA} clade, although they were particularly prevalent (100%) in southwest Tanzania (Mbeya region), close to Zambia and Malawi. The co-circulation of subtype C sequences from both east and southern African origin in Tanzania is consistent with its intermediate geographical position between eastern and southern countries. It is unclear whether subtype C clades of southern African origin detected in Kenya were introduced from Tanzania and/or directly from southern Africa.

It is also unclear the relevance of these findings for HIV-1 vaccine design. Possible correlations of distinct HIV-1 subtype C clades with differential susceptibility to neutralizing antibody and/or cellular immune responses should be explored to justify the selection of vaccines incorporating one or multiple immunogens derived from major African subtype C clades [8]. It is also uncertain whether distinct subtype C lineages may possess different biological properties that affect disease progression and viral transmission. A recent study conducted in Ethiopia showed that infection with clade C_{ET} is associated with initially lower HIV-1 RNA plasma loads but more rapid onset of disease than infections with clade C'_{ET} [68]. The authors proposed that the clade C_{ET} may be less efficiently transmitted than clade C'_{ET}, which is consistent with epidemiological evidence that show that the strain C'_{ET} has gained ground and surpassed the clade C_{ET} over time [40,68]. New studies are necessary to determine if subtype C lineages of east African origin are less transmissible than those originated in southern Africa.

In conclusion, the results presented here point to the existence of a HIV-1 subtype C lineage characteristic of east Africa, which accounts for >70% of subtype C infections in this African region. This lineage probably emerged in Burundi in the 1960s and about 10 years later spread to Ethiopia, Kenya, Uganda and Tanzania, where it disseminated establishing new local epidemics. The subtype C epidemics in Ethiopia, Kenya and Tanzania also resulted from the introduction and dissemination of additional lineages of southern African origin. The explanation for the pattern of spread of the HIV-1 subtype C epidemic in east Africa is probably multifactorial and includes founder effects, massive migration between countries as a consequence of ethnic conflicts and geographical proximity.

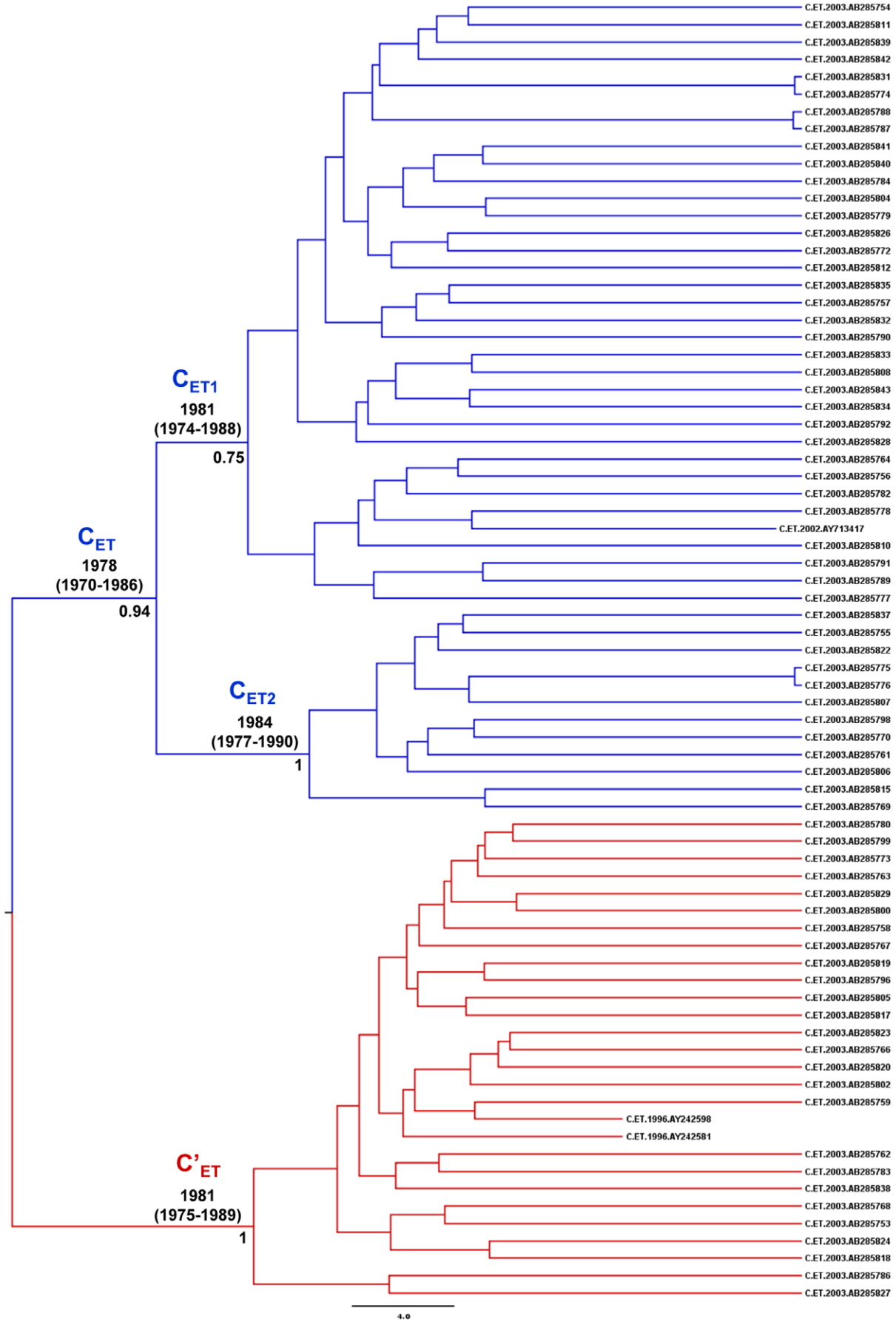

C.ET.2003.AB285754
C.ET.2003.AB285811
C.ET.2003.AB285839
C.ET.2003.AB285842
C.ET.2003.AB285831
C.ET.2003.AB285774
C.ET.2003.AB285788
C.ET.2003.AB285787
C.ET.2003.AB285841
C.ET.2003.AB285840
C.ET.2003.AB285784
C.ET.2003.AB285804
C.ET.2003.AB285779
C.ET.2003.AB285826
C.ET.2003.AB285772
C.ET.2003.AB285812
C.ET.2003.AB285835
C.ET.2003.AB285757
C.ET.2003.AB285832
C.ET.2003.AB285790
C.ET.2003.AB285833
C.ET.2003.AB285808
C.ET.2003.AB285843
C.ET.2003.AB285834
C.ET.2003.AB285792
C.ET.2003.AB285828
C.ET.2003.AB285764
C.ET.2003.AB285756
C.ET.2003.AB285782
C.ET.2003.AB285778
C.ET.2002.AY713417
C.ET.2003.AB285810
C.ET.2003.AB285791
C.ET.2003.AB285789
C.ET.2003.AB285777
C.ET.2003.AB285837
C.ET.2003.AB285755
C.ET.2003.AB285822
C.ET.2003.AB285775
C.ET.2003.AB285776
C.ET.2003.AB285807
C.ET.2003.AB285798
C.ET.2003.AB285770
C.ET.2003.AB285761
C.ET.2003.AB285806
C.ET.2003.AB285815
C.ET.2003.AB285769
C.ET.2003.AB285780
C.ET.2003.AB285799
C.ET.2003.AB285773
C.ET.2003.AB285763
C.ET.2003.AB285829
C.ET.2003.AB285800
C.ET.2003.AB285758
C.ET.2003.AB285767
C.ET.2003.AB285819
C.ET.2003.AB285796
C.ET.2003.AB285805
C.ET.2003.AB285817
C.ET.2003.AB285823
C.ET.2003.AB285766
C.ET.2003.AB285820
C.ET.2003.AB285802
C.ET.2003.AB285759
C.ET.1996.AY242598
C.ET.1996.AY242581
C.ET.2003.AB285762
C.ET.2003.AB285783
C.ET.2003.AB285838
C.ET.2003.AB285768
C.ET.2003.AB285753
C.ET.2003.AB285824
C.ET.2003.AB285818
C.ET.2003.AB285786
C.ET.2003.AB285827
4.0
C'ET
1981
(1975-1989)
1
CET1
1981
(1974-1988)
0.75
CET
1978
(1970-1986)
0.94
CET2
1984
(1977-1990)
1

Figure 5. Time-scaled Bayesian MCC tree of major Ethiopian HIV-1 subtype C lineages. MCC tree was obtained after exclusion of putative C/C′ intrasubtype recombinant sequences. Branches are colored according to the initial clade assignment of each sequence based on ML analysis: C_{ET} (blue) and C'_{ET} (red). The *PP* support and the median age (with 95% HPD interval in parentheses) are indicated only at key nodes. Horizontal branch lengths are drawn to scale with the scale at the bottom indicating years. The tree was automatically rooted under the assumption of a relaxed molecular clock.
doi:10.1371/journal.pone.0041904.g005

Supporting Information

Figure S1 Substitution saturation and likelihood mapping analyses. a) Transition (blue line) and transversion (green line) versus divergence plot for the HIV-1 subtype C *pol* dataset. b) Likelihood mapping of 10,000 random quarters selected from the HIV-1 subtype C *pol* dataset. Distribution (left triangle) and percentage (right triangle) of dots plotted in each region of the map. Each dot represents the likelihoods of the three possible tree topologies for a set of four sequences (quartets) selected randomly from the dataset. The dots localized on the vertices, in the centre and on the laterals represent the tree-like, the star-like and the network-like phylogenetic signals, respectively.
(PPT)

Figure S2 Close view of the HIV-1 C_{EA} lineage despited in Figure 1. The color of branches represents the country from where the sequence originated, according to the legend shown on the left. The boxes highlight the position of the major country-specific sub-clades detected in our study. The aLRT support values are indicated only at key nodes.
(PPTX)

Figure S3 Time-scaled Bayesian MCC tree of HIV-1 subtype C *pol* sequences from Ethiopia. Branches are colored according to the initial clade assignment of each sequence based on ML analysis: C_{ET} (blue), C'_{ET} (red), other clades (green). The *PP* support is indicated only at key nodes. Positions of the putative interclade C/C′ recombinant sequences are marked with asterisks. Horizontal branch lengths are drawn to scale with the scale at the bottom indicating years. The tree was automatically rooted under the assumption of a relaxed molecular clock. Representative bootscanning plots of some putative C/C′ intrasubtype recombinant sequences are depicted on the right. Query sequences were compared to reference sequences of HIV-1 clades A1 (AB253429), D (AY371157), C_{ET} (AY242589), and C'_{ET} (AY242581).
(PPTX)

Acknowledgments

We wish to thank Dr. Vera Bongertz for critical review of the manuscript and anonymous reviewers for comments.

Author Contributions

Conceived and designed the experiments: GB. Performed the experiments: GB EOD. Analyzed the data: GB EOD. Contributed reagents/materials/analysis tools: GB EOD. Wrote the paper: GB EOD.

References

1. Hemelaar J, Gouws E, Ghys PD, Osmanov S (2011) Global trends in molecular epidemiology of HIV-1 during 2000–2007. Aids 25: 679–689.
2. Parreira R, Piedade J, Domingues A, Lobao D, Santos M, et al. (2006) Genetic characterization of human immunodeficiency virus type 1 from Beira, Mozambique. Microbes Infect 8: 2442–2451.
3. Bredell H, Martin DP, Van Harmelen J, Varsani A, Sheppard HW, et al. (2007) HIV type 1 subtype C gag and nef diversity in Southern Africa. AIDS Res Hum Retroviruses 23: 477–481.
4. Deho L, Walwema R, Cappelletti A, Sukati H, Sibandze D, et al. (2008) Subtype assignment and phylogenetic analysis of HIV type 1 strains in patients from Swaziland. AIDS Res Hum Retroviruses 24: 323–325.
5. Lahuerta M, Aparicio E, Bardaji A, Marco S, Sacarlal J, et al. (2008) Rapid spread and genetic diversification of HIV type 1 subtype C in a rural area of southern Mozambique. AIDS Res Hum Retroviruses 24: 327–335.
6. Dalai SC, de Oliveira T, Harkins GW, Kassaye SG, Lint J, et al. (2009) Evolution and molecular epidemiology of subtype C HIV-1 in Zimbabwe. Aids 23: 2523–2532.
7. Novitsky V, Wang R, Lagakos S, Essex M (2010) HIV-1 Subtype C Phylodynamics in the Global Epidemic. Viruses 2: 33–54.
8. Thomson MM, Fernandez-Garcia A (2011) Phylogenetic structure in African HIV-1 subtype C revealed by selective sequential pruning. Virology 415: 30–38.
9. Koch N, Ndihokubwayo JB, Yahi N, Tourres C, Fantini J, et al. (2001) Genetic analysis of hiv type 1 strains in Bujumbura (burundi): predominance of subtype c variant. AIDS Res Hum Retroviruses 17: 269–273.
10. Vidal N, Niyongabo T, Nduwimana J, Butel C, Ndayiragije A, et al. (2007) HIV type 1 diversity and antiretroviral drug resistance mutations in Burundi. AIDS Res Hum Retroviruses 23: 175–180.
11. Maslin J, Rogier C, Berger F, Khamil MA, Mattera D, et al. (2005) Epidemiology and genetic characterization of HIV-1 isolates in the general population of Djibouti (Horn of Africa). J Acquir Immune Defic Syndr 39: 129–132.
12. Abebe A, Kuiken CL, Goudsmit J, Valk M, Messele T, et al. (1997) HIV type 1 subtype C in Addis Ababa, Ethiopia. AIDS Res Hum Retroviruses 13: 1071–1075.
13. Abebe A, Pollakis G, Fontanet AL, Fisseha B, Tegbaru B, et al. (2000) Identification of a genetic subcluster of HIV type 1 subtype C (C′) widespread in Ethiopia. AIDS Res Hum Retroviruses 16: 1909–1914.
14. Hussein M, Abebe A, Pollakis G, Brouwer M, Petros B, et al. (2000) HIV-1 subtype C in commerical sex workers in Addis Ababa, Ethiopia. J Acquir Immune Defic Syndr 23: 120–127.
15. Kassu A, Fujino M, Matsuda M, Nishizawa M, Ota F, et al. (2007) Molecular epidemiology of HIV type 1 in treatment-naive patients in north Ethiopia. AIDS Res Hum Retroviruses 23: 564–568.
16. Renjifo B, Chaplin B, Mwakagile D, Shah P, Vannberg F, et al. (1998) Epidemic expansion of HIV type 1 subtype C and recombinant genotypes in Tanzania. AIDS Res Hum Retroviruses 14: 635–638.
17. Kiwelu IE, Renjifo B, Chaplin B, Sam N, Nkya WM, et al. (2003) HIV type 1 subtypes among bar and hotel workers in Moshi, Tanzania. AIDS Res Hum Retroviruses 19: 57–64.
18. Herbinger KH, Gerhardt M, Piyasirisilp S, Mloka D, Arroyo MA, et al. (2006) Frequency of HIV type 1 dual infection and HIV diversity: analysis of low- and high-risk populations in Mbeya Region, Tanzania. AIDS Res Hum Retroviruses 22: 599–606.
19. Nyombi BM, Kristiansen KI, Bjune G, Muller F, Holm-Hansen C (2008) Diversity of human immunodeficiency virus type 1 subtypes in Kagera and Kilimanjaro regions, Tanzania. AIDS Res Hum Retroviruses 24: 761–769.
20. Mosha F, Urassa W, Aboud S, Lyamuya E, Sandstrom E, et al. (2011) Prevalence of genotypic resistance to antiretroviral drugs in treatment-naive youths infected with diverse HIV type 1 subtypes and recombinant forms in Dar es Salaam, Tanzania. AIDS Res Hum Retroviruses 27: 377–382.
21. Servais J, Lambert C, Karita E, Vanhove D, Fischer A, et al. (2004) HIV type 1 pol gene diversity and archived nevirapine resistance mutation in pregnant women in Rwanda. AIDS Res Hum Retroviruses 20: 279–283.
22. Brennan CA, Lund JK, Golden A, Yamaguchi J, Vallari AS, et al. (1997) Serologic and phylogenetic characterization of HIV-1 subtypes in Uganda. Aids 11: 1823–1832.
23. Rayfield MA, Downing RG, Baggs J, Hu DJ, Pieniazek D, et al. (1998) A molecular epidemiologic survey of HIV in Uganda. HIV Variant Working Group. Aids 12: 521–527.
24. Hu DJ, Baggs J, Downing RG, Pieniazek D, Dorn J, et al. (2000) Predominance of HIV-1 subtype A and D infections in Uganda. Emerg Infect Dis 6: 609–615.
25. Gale CV, Yirrell DL, Campbell E, Van der Paal L, Grosskurth H, et al. (2006) Genotypic variation in the pol gene of HIV type 1 in an antiretroviral treatment-naive population in rural southwestern Uganda. AIDS Res Hum Retroviruses 22: 985–992.
26. Herbeck JT, Lyagoba F, Moore SW, Shindo N, Biryahwaho B, et al. (2007) Prevalence and genetic diversity of HIV type 1 subtypes A and D in women attending antenatal clinics in Uganda. AIDS Res Hum Retroviruses 23: 755–760.
27. Ssemwanga D, Ndembi N, Lyagoba F, Bukenya J, Seeley J, et al. (2011) HIV Type 1 Subtype Distribution, Multiple Infections, Sexual Networks, and

Partnership Histories in Female Sex Workers in Kampala, Uganda. AIDS Res Hum Retroviruses.
28. Hamers RL, Wallis CL, Kityo C, Siwale M, Mandaliya K, et al. (2011) HIV-1 drug resistance in antiretroviral-naive individuals in sub-Saharan Africa after rollout of antiretroviral therapy: a multicentre observational study. Lancet Infect Dis 11: 750–759.
29. Yang C, Li M, Shi YP, Winter J, van Eijk AM, et al. (2004) Genetic diversity and high proportion of intersubtype recombinants among HIV type 1-infected pregnant women in Kisumu, western Kenya. AIDS Res Hum Retroviruses 20: 565–574.
30. Oyaro M, Mbithi J, Oyugi F, Laten A, Anzala O, et al. (2011) Molecular characterization of HIV type 1 among HIV-infected respondents in a cohort being prepared for HIV Phase III vaccine clinical trials, Western Kenya. AIDS Res Hum Retroviruses 27: 257–264.
31. Nyagaka B, Kiptoo MK, Lihana RW, Khamadi SA, Makokha EP, et al. (2011) HIV Type 1 gag Genetic Diversity Among Antenatal Clinic Attendees in North Rift Valley, Kenya. AIDS Res Hum Retroviruses.
32. Rainwater S, DeVange S, Sagar M, Ndinya-Achola J, Mandaliya K, et al. (2005) No evidence for rapid subtype C spread within an epidemic in which multiple subtypes and intersubtype recombinants circulate. AIDS Res Hum Retroviruses 21: 1060–1065.
33. Khamadi SA, Lihana RW, Osman S, Mwangi J, Muriuki J, et al. (2009) Genetic diversity of HIV type 1 along the coastal strip of Kenya. AIDS Res Hum Retroviruses 25: 919–923.
34. Hue S, Hassan AS, Nabwera H, Sanders EJ, Pillay D, et al. (2012) HIV Type 1 in a Rural Coastal Town in Kenya Shows Multiple Introductions with Many Subtypes and Much Recombination. AIDS Res Hum Retroviruses 28: 220–224.
35. Neilson JR, John GC, Carr JK, Lewis P, Kreiss JK, et al. (1999) Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. J Virol 73: 4393–4403.
36. Lihana RW, Khamadi SA, Kiptoo MK, Kinyua JG, Lagat N, et al. (2006) HIV type 1 subtypes among STI patients in Nairobi: a genotypic study based on partial pol gene sequencing. AIDS Res Hum Retroviruses 22: 1172–1177.
37. Kageha S, Lihana RW, Okoth V, Mwau M, Okoth FA, et al. (2012) HIV Type 1 Subtype Surveillance in Central Kenya. AIDS Res Hum Retroviruses 28: 228–231.
38. Khamadi SA, Ochieng W, Lihana RW, Kinyua J, Muriuki J, et al. (2005) HIV type 1 subtypes in circulation in northern Kenya. AIDS Res Hum Retroviruses 21: 810–814.
39. Khamadi SA, Lihana RW, Mwaniki DL, Kinyua J, Lagat N, et al. (2008) HIV type 1 genetic diversity in Moyale, Mandera, and Turkana based on env-C2-V3 sequences. AIDS Res Hum Retroviruses 24: 1561–1564.
40. Pollakis G, Abebe A, Kliphuis A, De Wit TF, Fisseha B, et al. (2003) Recombination of HIV type 1C (C′/C″) in Ethiopia: possible link of EthHIV-1C′ to subtype C sequences from the high-prevalence epidemics in India and Southern Africa. AIDS Res Hum Retroviruses 19: 999–1008.
41. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, et al. (2005) An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 21: 3797–3800.
42. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876–4882.
43. Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. J Hered 92: 371–373.
44. Strimmer K, von Haeseler A (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci U S A 94: 6815–6819.
45. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502–504.
46. Sanchez R, Serra F, Tarraga J, Medina I, Carbonell J, et al. (2011) Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. Nucleic Acids Res 39: W470–474.
47. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25: 1253–1256.
48. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704.
49. Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online–a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33: W557–559.
50. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol 55: 539–552.
51. Rambaut A (2009) FigTree v1.3.1: Tree Figure Drawing Tool. Available from http://treebioedacuk/software/figtree/. Accessed 2012 April 20.
52. Ray S Simplot v2.5.0. Available from: http://sray.med.som.jhmi.edu/SCRoftware/simplot/. Accessed 2012 April 20.
53. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161: 1307–1320.
54. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.
55. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLoS Biol 4: e88.
56. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22: 1185–1192.
57. Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. PLoS Comput Biol 5: e1000520.
58. Rambaut A, Drummond A (2007) Tracer v1.4. Available from http://beastbioedacuk/Tracer. Accessed 2012 April 20.
59. Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27: 2910–2912.
60. Bello G, Passaes CP, Guimaraes ML, Lorete RS, Matos Almeida SE, et al. (2008) Origin and evolutionary history of HIV-1 subtype C in Brazil. AIDS 22: 1993–2000.
61. Gray RR, Tatem AJ, Lamers S, Hou W, Laeyendecker O, et al. (2009) Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. Aids 23: F9–F17.
62. de Oliveira T, Pillay D, Gifford RJ (2010) The HIV-1 subtype C epidemic in South America is linked to the United Kingdom. PLoS ONE 5: e9311.
63. Fontella R, Soares MA, Schrago CG (2008) On the origin of HIV-1 subtype C in South America. AIDS 22: 2001–2011.
64. Fransen S, Ong'ayo A (2010) Migration in Burundi: History, Current Trends, and Future Prospects. Paper Series: Migration and Development Country Profiles Maastricht: Maastricht Graduate School of Governance Available from http://mgsog.merit.unu.edu/ISacademie/docs/CR_burundi.pdf. Accessed 2012 April 20.
65. Abebe A, Lukashov VV, Pollakis G, Kliphuis A, Fontanet AL, et al. (2001) Timing of the HIV-1 subtype C epidemic in Ethiopia based on early virus strains and subsequent virus diversification. AIDS 15: 1555–1561.
66. Abebe A, Lukashov VV, Rinke De Wit TF, Fisseha B, Tegbaru B, et al. (2001) Timing of the introduction into Ethiopia of subcluster C′ of HIV type 1 subtype C. AIDS Res Hum Retroviruses 17: 657–661.
67. Tully DC, Wood C (2010) Chronology and evolution of the HIV-1 subtype C epidemic in Ethiopia. Aids 24: 1577–1582.
68. Ayele W, Mekonnen Y, Messele T, Mengistu Y, Tsegaye A, et al. (2010) Differences in HIV type 1 RNA plasma load profile of closely related cocirculating Ethiopian subtype C strains: C and C′. AIDS Res Hum Retroviruses 26: 805–813.

8.3 APÊNDICE 3

Spatiotemporal dynamics of the HIV-1 CRF06_cpx epidemic in Western Africa

Autores: **Edson Delatorre** e Gonzalo Bello

Jornal: AIDS, 2013

Spatiotemporal dynamics of the HIV-1 CRF06_cpx epidemic in western Africa

Edson Delatorre and Gonzalo Bello

Objective: To investigate the origin and spatiotemporal dynamics of dissemination of the HIV-1 CRF06_cpx clade in western Africa.

Design: A total of 180 HIV-1 CRF06_cpx-like *pol* sequences isolated from 12 different countries from west and west-central Africa over a period of 16 years (1995–2010) were analyzed.

Methods: Evolutionary, phylogeographic and demographic parameters were jointly estimated from sequence data using a Bayesian coalescent-based method and combined with molecular epidemiology and spatial accessibility data.

Results: The CRF06_cpx most probably emerged in Burkina Faso in 1979 (1970–1985). From Burkina Faso, the virus was first disseminated to Mali and Nigeria during the 1980s and later to other countries from west and west-central Africa. Demographic reconstruction indicates that the CRF06_cpx epidemic grew exponentially during the 1980s, with a median growth rate of 0.82 year^{-1} (0.60–1.09 year^{-1}), and after stabilize. We found a negative correlation between CRF06_cpx prevalence and the geographical distance to Burkina Faso's capital. Regional accessibility information agrees with the overall geographical range of the CRF06_cpx, but not fully explains the highly heterogeneous distribution pattern of this CRF at regional level.

Conclusion: The CRF06_cpx epidemic in western Africa probably emerged at the late 1970s and grew during the 1980s at a rate comparable to the HIV-1 epidemics in the United States and Europe. Burkina Faso seems to be the most important epicenter of dissemination of the HIV-1 CRF06_cpx strain at regional level. The explanation for the current geographical distribution of CRF06_cpx is probably multifactorial.

AIDS 2013, **27**:1313–1320

Keywords: CRF06_cpx, HIV-1, phylodynamics, western Africa

Introduction

The dispersion of some human immunodeficiency virus type 1 (HIV-1) group M strains out of the epicenter in Central Africa has given rise to a diverse collection of viral lineages with a complex global distribution, that we know today as subtypes and inter-subtype recombinant forms. The HIV-1 circulating recombinant forms (CRFs) and unique recombinants forms have an epidemiologically relevant contribution to the HIV-1 epidemic being responsible for over 20% of all global infections [1]. Importantly, the global proportion of all CRFs combined increased by 4.5% between 2000–2003 and 2004–2007 [1].

The CRF06_cpx is a complex recombinant that includes genomic segments of subtypes A, G, J, and K [2]. This CRF mainly circulates in western Africa, although its occurrence greatly varies across countries. The CRF06_cpx is the predominant clade in Burkina Faso where it accounts for 40–50% of HIV-1 infections [3–5], whereas its prevalence is reduced to 10–15% in Mali [6,7] and Niger [8,9], 3–8% in Benin [10], Ghana [11,12], Côte d'Ivoire [13,14], Nigeria [15], Senegal [16,17], and Togo [18,19], and 1% or less in Guinea Bissau [20] and Guinea Conakry [9]. This CRF has also been occasionally detected in several west-central African countries including: Cameroon, Central African Republic (CAR), Chad, Equatorial Guinea and Gabon [21].

Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil.
Correspondence to Gonzalo Bello, Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ. Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brazil.
Tel: +55 21 3865 8154; fax: +55 21 3865 8173; e-mail: gbellobr@gmail.com/gbello@ioc.fiocruz.br
Received: 14 November 2012; revised: 20 December 2012; accepted: 17 January 2013.

DOI:10.1097/QAD.0b013e32835f1df4

Bayesian phylogeographic models has been successfully used to reconstruct the spatial and temporal dispersion pattern of different HIV-1 clades at a regional level in the African continent, including the subtypes A, C and D in East Africa [22,23] and the CRF02_AG in the Congo River basin [24]. Despite extensive data about molecular epidemiology, the spatiotemporal dynamics of dissemination of most prevalent HIV-1 lineages circulating in western Africa remains largely unexplored. Previous studies have reconstructed the migration routes and population dynamics of some HIV-1 clades in west Africa at a country scale, such as the CRF02_AG clade in Guinea Bissau [20] and the subtype C clade in Senegal [25]; but none has explored the dissemination dynamics of HIV-1 at a regional level.

In the present study, we used a comprehensive data set of 180 HIV-1 CRF06_cpx-like *pol* sequences isolated from 12 different countries from west and west-central Africa over a period of 16 years (1995–2010). Spatial and temporal information of sequences was combined with Bayesian analyses to reconstruct simultaneously the onset date, the migration routes and the demographic history of the HIV-1 CRF06_cpx epidemic at a regional scale.

Methods

Sequence dataset

A total of 207 HIV-1 CRF06_cpx *pol* sequences from different patients of African origin covering the entire protease and partial reverse transcriptase (PR/RT) regions (nt 2253-3272 relative to HXB2 clone) were downloaded from the Los Alamos HIV Sequence Database (www.hiv.lanl.gov) by July 2012. Five sequences with a mosaic profile different to the CRF06_cpx reference sequences and 22 sequences with no information about sampling date were removed. This resulted in a final data set of 180 CRF06_cpx-like *pol* sequences from west ($n = 174$) and west-central ($n = 6$) Africa sampled over a period of 16 years (Table S1, http://links.lww.com/QAD/A311). Sequences were aligned and all sites with major antiretroviral drug resistance mutations in RT (41, 65, 67, 69, 70, 74, 101, 103, 106, 138, 151, 181, 184, 188, 190, 210, 215 and 219) detected in at least two sequences were excluded. Alignment is available from the authors upon request.

Genetic classification

The CRF06_cpx-like classification of all *pol* sequences here included was confirmed by: maximum likelihood phylogenetic analysis, REGA HIV subtyping tool v.2 [26], and bootscanning analysis using Simplot software v.3.5.1 [27]. The maximum likelihood phylogenetic tree was constructed with the PhyML 3.0 program [28] using an online web server (http://www.atgc-montpellier.fr/phyml/). The maximum likelihood tree was inferred under the GTR+I+G nucleotide substitution model selected using the jModeltest program [29], and the heuristic tree search was performed using the SPR branch-swapping algorithm. The approximate likelihood-ratio test based on a Shimodaira-Hasegawa-like procedure was used as a statistical test to calculate branch support. In bootscanning analyses, supporting branching of query sequences with reference sequences from all HIV-1 group M subtypes was determined in Neighbor-Joining trees constructed using the Kimura two-parameter model, within a 250 bp window moving in steps of 10 bases.

Analysis of the spatiotemporal dispersion pattern

The evolutionary rate (μ, nucleotide substitutions per site per year, subst./site per year), the age of the most recent common ancestor (T_{mrca}, years), the demographic history, and the spatial dynamics of CRF06_cpx circulating in west and west-central Africa were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST v1.6.2 [30]. Analyses were performed using the GTR+I+Γ_4 nucleotide substitution model and an uncorrelated Lognormal relaxed molecular clock model [31] under different coalescent models. Migration events throughout the phylogenetic histories and the most relevant migration pathways between locations were identified by applying a standard discrete Bayesian phylogeographic model and the Bayesian stochastic search variable selection (BSSVS) approach [32], respectively. MCMC chains were run for 5×10^8 generations and adequate chain mixing was checked, after excluding an initial 10%, by calculating the effective sample size using the TRACER v1.4 program (http://beast.bio.ed.ac.uk/Tracer). Maximum clade credibility (MCC) trees were summarized from the posterior distribution of trees with TreeAnnotator and visualized with FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree). Migratory events and significant non-zero rates obtained by the BSSVS approach were summarized using the cross-platform SPREAD application [33] and viewed with Google Earth (http://earth.google.com).

Statistical analysis

The correlation between CRF06_cpx prevalence in each country and the corresponding geographical distance from the country's capital to the Burkina Faso's capital was examined using different regression models. The model with the better fit to the data (higher R^2 value) was selected. Statistical calculations were done using the GraphPad Prism version 2.01 program (GraphPad Software, San Diego, California, USA).

Results

Origin of the HIV-1 CRF06_cpx clade

In the present study we used a dataset of 180 CRF06_cpx-like *pol* sequences isolated from 12 different

countries from west and west-Central Africa between 1995 and 2010 (Table S1, http://links.lww.com/QAD/A311). Most (76%) CRF06_cpx-like *pol* sequences were retrieved from untreated patients (Table S1, http://links.lww.com/QAD/A311). All *pol* sequences here included branched in a highly supported monophyletic clade and displayed the same G/K mosaic structure that the CRF06_cpx reference sequences (Fig. S1, http://links.lww.com/QAD/A311), thus, confirming their original classification. Importantly, our dataset includes sequences from all African countries with description of CRF06_cpx infections at a prevalence more than 1%, with exception of Niger for which no sequence data for the selected *pol* gene segment was available in public databases.

The median evolutionary rate of the HIV-1 CRF06_cpx lineage at *pol* gene directly calculated from the sampling dates of the sequences was estimated at 2.4×10^{-3} (95% HPD: $1.8 \times 10^{-3} - 3.0 \times 10^{-3}$) subst./site per year, consistent with the order of magnitude of 10^{-3} expected for HIV-1. The estimated coefficient of rate variation in our dataset was 0.28 (95% HPD: 0.22–0.33). This demonstrates a significant variation of substitution rate among branches and validates the use of a relaxed molecular clock model to reconstruct the time-scale of the CRF06_cpx clade. According to the Bayesian MCMC analysis, the most probable root location of the CRF06_cpx clade was placed in Burkina Faso (posterior state probability, $PSP = 0.94$), and the onset date of this clade was estimated to be 1979 (95% HPD: 1970–1985) (Fig. 1).

Spatiotemporal dispersal pattern of the HIV-1 CRF06_cpx clade

The spatiotemporal dynamics of CRF06_cpx was reconstructed using a Bayesian phylogeographic diffusion model that takes into account the uncertainty both at the phylogenetic and the viral migration level. The Bayesian MCC tree points to a great level of phylogenetic intermixing of CRF06_cpx sequences from different geographic locations and supports that Burkina Faso, where CRF06_cpx is dominant, has been the epicenter from where this viral strain has spread to neighboring countries (Fig. 1). This analysis identified some highly supported (posterior probability, $PP > 0.90$) country-specific monophyletic clades of small size ($n \leq 5$) in Benin, Chad, Côte d'Ivoire, Mali, Nigeria and Senegal (Fig. 1). A large ($n = 9$) country-specific monophyletic clade was also detected in Mali, but received low support ($PP < 0.50$).

Reconstruction of viral migrations across time revealed a rapid dissemination of CRF06_cpx across West Africa (Fig. 2). After its emergence in Burkina Faso around the late 1970s, the virus was first disseminated to Mali between 1980 and 1985 and later to Nigeria between 1985 and 1990. During the 1990s the virus migrated from Burkina Faso to southern neighboring countries including Benin, Ghana and Côte d'Ivoire, and also to more distant countries like Chad and Senegal. In more recent times, migrations of the CRF06_cpx clade were detected from Burkina Faso to Togo and west-equatorial African countries (Cameroon, Gabon and Equatorial Guinea). The Bayes factor tests for significant nonzero rates, only supports epidemiological linkage between Burkina Faso and Mali (Bayes factor = 6426), Burkina Faso and Nigeria (Bayes factor = 180), and Burkina Faso and Ghana (Bayes factor = 11) (Fig. 2).

Spatial accessibility and the spread of the HIV-1 CRF06_cpx clade

To better characterize the spatial spread of the HIV-1 CRF06_cpx clade at a regional level, the geographical distribution of CRF06_cpx was superimposed to accessibility data reflecting the travel time to major cities (>50 000 people) (http://bioval.jrc.ec.europa.eu/products/gam/index.htm). Inspection of the African accessibility map revealed that western region is well connected and accessible, and that such a corridor of connectivity also extends to some countries from the central region (Cameroon, CAR, Chad, Equatorial Guinea and Gabon) where the CRF06_cpx has been detected (Fig. S2, http://links.lww.com/QAD/A311). Despite this high interconnectivity of west and west-central African populations, molecular epidemiology data reveals a great variation of the CRF06_cpx prevalence at regional level, ranging from less than 1 to more than 40% (Fig. 3). We found that such prevalence was negatively correlated with the geographical distance to the capital of Burkina Faso (Ouagadougou). As we move away from Ouagadougou the prevalence of CRF06_cpx rapidly decreases following an exponential decay curve (Fig. 3).

Demographic history of the HIV-1 CRF06_cpx clade

To reconstruct the population dynamic pattern of the HIV-1 CRF06_cpx clade, estimate of effective population size (*Ne*) over time was obtained using a Bayesian skyline plot (BSP) coalescent model. The BSP analysis suggests that the CRF06_cpx clade experienced an initial phase of fast exponential growth during the 1980s, followed by a more recent decline in growth rate since the early 1990s (Fig. 4). This demographic trend was consistent with a model of logistic growth that was then used to estimate the initial growth rate of the CRF06_cpx epidemic. The logistic growth coalescent model indicates that the CRF06_cpx expanded during the 1980s with a median growth rate of 0.82 $year^{-1}$ (95% HPD: 0.60–1.09 $year^{-1}$). The demographic pattern of the CRF06_cpx clade was then compared with changes in the estimated number of people living with HIV in those west African countries with a CRF06_cpx prevalence more than 2%. The reconstructed demographic pattern fully agrees with the epidemiological profile of Burkina Faso, where the

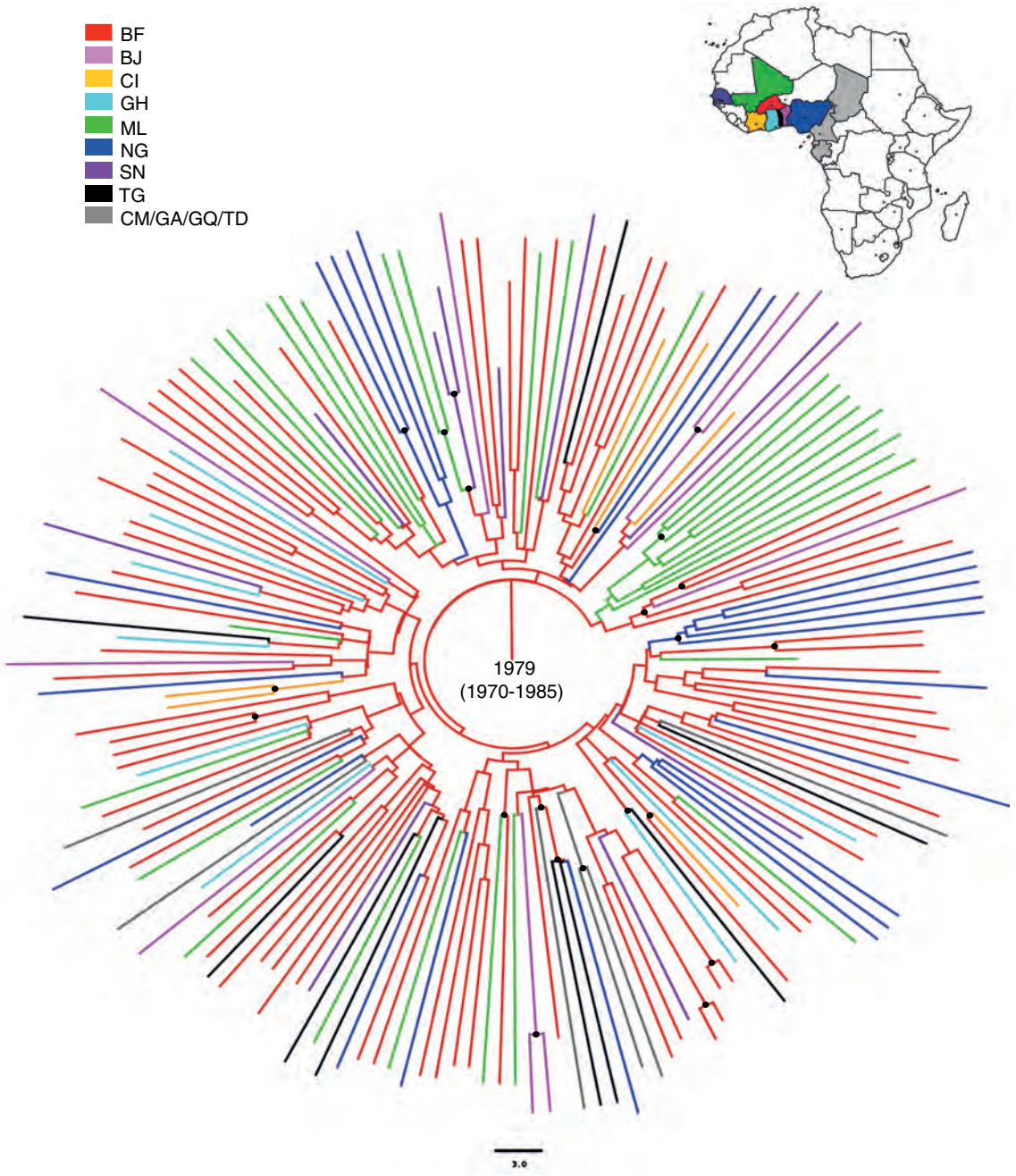


Fig. 1. Time-scaled Bayesian MCC tree of the HIV-1 CRF06_cpx clade. Branches are colored according to the most probable location state of their descendent nodes. Color code is indicated in the legend and in the map shown in the top. The median age (with 95% HPD interval in parentheses) of the most recent common ancestor of CRF06_cpx is shown. Black dots point to key nodes with a high (>0.85) *PP* support. Branch lengths are drawn to scale with the bar at the bottom indicating years. The tree was automatically rooted under the assumption of a relaxed molecular clock. Countries represented are: BF (Burkina Faso), BJ (Benin), CI (Côte d'Ivoire), GH (Ghana), ML (Mali), NG (Nigeria), SN (Senegal), TG (Togo), CM (Cameroon), GA (Gabon), GQ (Equatorial Guinea) and TD (Chad).

number of people living with HIV remained relatively stable between 1990 and 2010; but differs from the epidemiological scenario in other west African countries where the HIV epidemic only began to stabilize after the middle 1990s (Fig. 4).

Discussion

The present study characterized the spatiotemporal dynamics of dispersal of the HIV-1 CRF06_cpx throughout western Africa. The evolutionary and

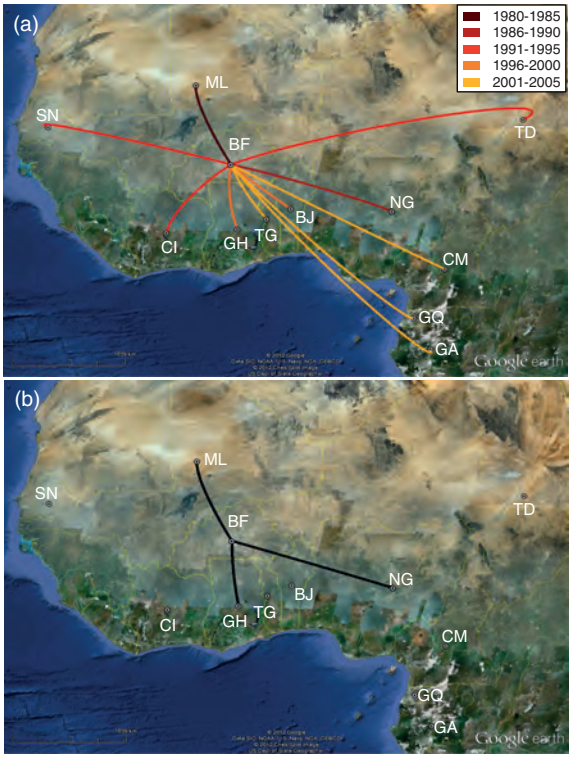


Fig. 2. Spatiotemporal dynamics of HIV-1 CRF06_cpx clade dissemination in west and west-central African regions. (a) Viral migration events between 1980 and 2005. Lines between locations represent branches in the Bayesian MCC tree along which location transitions occurs. The line's color informs the timing of location transitions according to the legend at right corner. (b) Bayes factor test for significant nonzero rates. Only rates supported by a Bayes factor greater than 10 are indicated. BF, Burkina Faso; BJ, Benin; CI, Côte d'Ivoire; CM, Cameroon; GA, Gabon; GH, Ghana; GQ, Equatorial Guinea; ML, Mali; NG, Nigeria; SN, Senegal; TD, Chad; TG, Togo. The maps are based on satellite pictures made available in Google Earth (http://earth.google.com).

phylogeographic analysis performed here suggests that the CRF06_cpx originated in Burkina Faso in the late 1970s. Burkina Faso, a landlocked country that occupies a central geographical position in western Africa, seems to have been the most important epicenter of the CRF06_cpx dissemination at regional level, continuously exporting the virus to other countries from west and west-Central Africa. Whether the recombination event that originates the CRF06_cpx ancestor took place in Burkina Faso or another African country, however, is not clear. One intriguing point is that the mosaic genome of CRF06_cpx comprises subtypes J and K, two HIV-1 variants detected at low prevalence in central-Africa, but not in Burkina Faso [3–5].

Intense migratory flows were registered from Burkina Faso to neighboring countries during the 1970s and 1980s, particularly to Côte d'Ivoire and Ghana (http://www. oecd.org/migration/38409521.pdf). Our phylogeographic analysis of CRF06_cpx detected significant epidemiological links between Burkina Faso and Mali, Burkina Faso and Nigeria, and Burkina Faso and Ghana; but not between Burkina Faso and Côte d'Ivoire. This may be partially due to sampling bias since the majority (73%) of the CRF06_cpx sequences in our data set was from Burkina Faso, Mali and Nigeria; whereas other countries were represented by a small number of sequences. A more comprehensive sampling of CRF06_cpx viruses may certainly result in the identification of new migration events and epidemiological links not detected in this study. The root position of the CRF06_cpx clade in Burkina Faso may be also sensitive to the sampling bias because most sequences (43%) were from that country. Such root location, however, is fully consistent with epidemiological data that show that prevalence of CRF06_cpx reaches a maximum in Burkina Faso and decreases exponentially as we move away from that country.

Tatem *et al.* [34] suggest that accessibility between locations have played a major role in the spatial spread of HIV-1 in sub-Saharan Africa. The concentration of most CRF06_cpx infections in western Africa is fully consistent with the strong connectivity of this region and its relative isolation from other African regions, with exception of some west-central African countries where CRF06_cpx has been also detected. Human mobility may have also played an important role in the spatial spread of the CRF06_cpx clade as western Africa appears as an area of intense intermixing of populations at regional level. Quantitative estimates of intra-regional migration indicate that west African countries currently host about 7.5 million migrants from other countries of the region, representing 3% of the regional population, a rate that is above the African average (2%) and that largely exceeds that of the European Union (0.5%) (http://www.oecd.org/migration/38409521.pdf).

Once an HIV variant entered in western Africa, the strong accessibility and the high human mobility between population centers may promote the rapid and homogenous dissemination of the viral strain throughout the region. This may explain the quite homogenous prevalence of the CRF02_AG clade that usually ranges between 40 and 70% across western African countries [1]. The CRF06_cpx prevalence at regional level, however, is much more heterogeneous, ranging from less than 1 to more than 40% [3–20], and seems to be negatively correlated with the geographical distance to Burkina Faso's capital. A small increase in the distance to Ouagadougou is associated with a drastic reduction in the prevalence of CRF06_cpx. This suggests that accessibility and human mobility are not enough to fully explain the complexity of spatial distribution of distinct HIV-1 clades in western Africa, and additional factors may also have influenced such dissemination process.

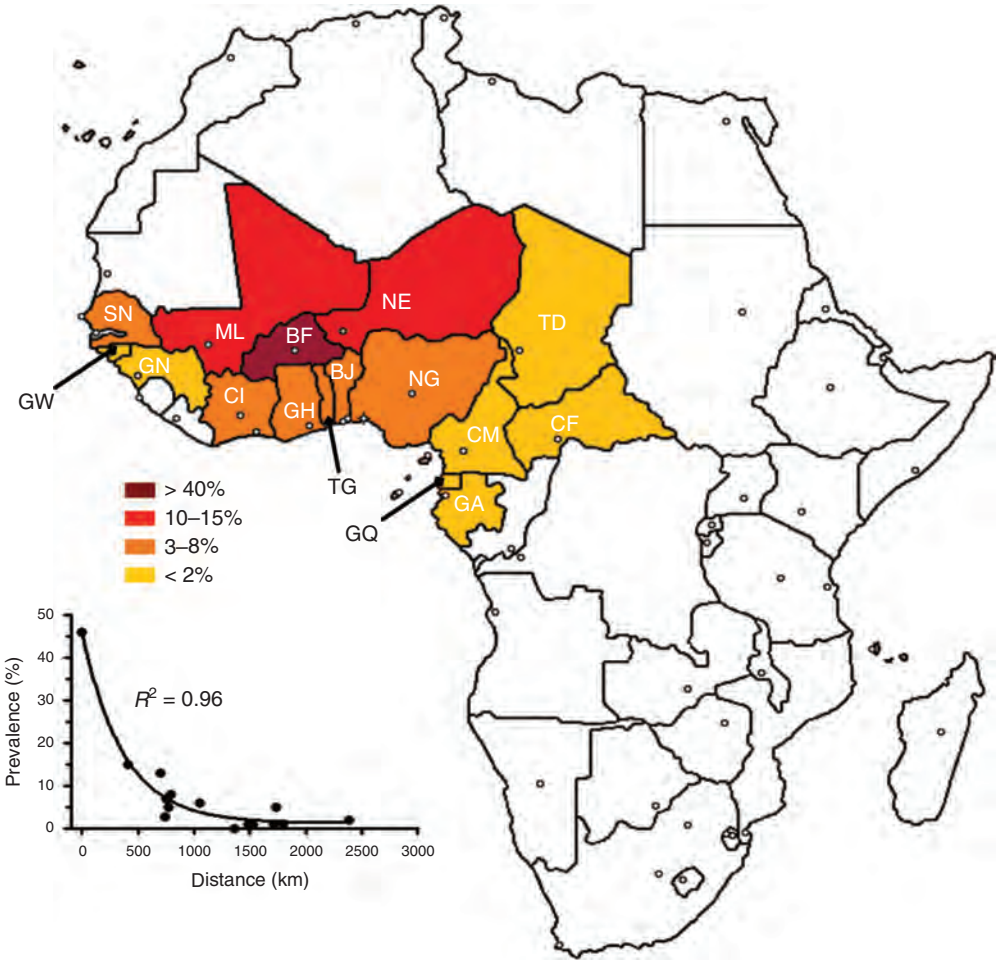


Fig. 3. African map showing the HIV-1 CRF06_cpx prevalence in west and west-central Africa. Countries were colored according to the prevalence of CRF06_cpx as shown in the legend. CRF06_cpx prevalence was estimated from diverse sources [3–20]. The graph on the bottom shows the correlation between the CRF06_cpx prevalence at each country and the geographical distance between the corresponding country's capital and Burkina Faso's capital. Correlation was adjusted to an exponential decay curve and the model fit to the data (R^2 value) is indicated. BF, Burkina Faso; BJ, Benin; CF, Central African Republic; CI, Côte d'Ivoire; CM, Cameroon; GA, Gabon; GH, Ghana; GN, Guinea; GQ, Equatorial Guinea; GW, Guinea Bissau; ML, Mali; NE, Niger; NG, Nigeria; SN, Senegal; TD, Chad; TG, Togo.

Our demographic reconstruction suggests that the CRF06_cpx epidemic in western Africa experienced an initial phase of exponential growth during the 1980s followed by a more recent stabilization since the early 1990s. This pattern resemble that previously described for the CRF02_AG epidemic in the same region [20], with one important difference. The CRF02_AG probably emerged 10 years earlier than the CRF06_cpx and was rapidly disseminated throughout the 1970s and 1980s. At the same time, estimations of UNAIDS indicate that the number of people living with HIV in most western African countries started to stabilize during the 1990s (Fig. 4). We suggest that the later emergence of the CRF06_cpx clade combined with temporal changes in the epidemic growth pattern at regional level during 1990s may have resulted in a less efficient and more heterogeneous dissemination of CRF06_cpx through west-Africa when compared with CRF02_AG.

Coalescent estimations of the HIV-1 epidemic growth rate in southern and eastern African countries (~0.2–0.4 year^{-1}) [35–37] have been lower than those obtained in countries from western Europe, USA and South America (~0.5–1.5 year^{-1}) [37–40]. This may reflects the impact of different transmission routes operating at each region. Although HIV transmission in Africa mainly occurred through heterosexual contacts; in Europe and the United States the virus is also transmitted through networks with high rates of partner exchanges like IDU and MSM. The median growth rate here estimated for the CRF06_cpx epidemic in western Africa (0.82 year^{-1}), however, was similar to that projected for HIV-1 epidemics in Europe and the United States. This may suggests that: HIV-1 transmission by heterosexual contacts was more efficient in western Africa than in other African regions; other HIV-1 transmission routes, like the iatrogenic one, also

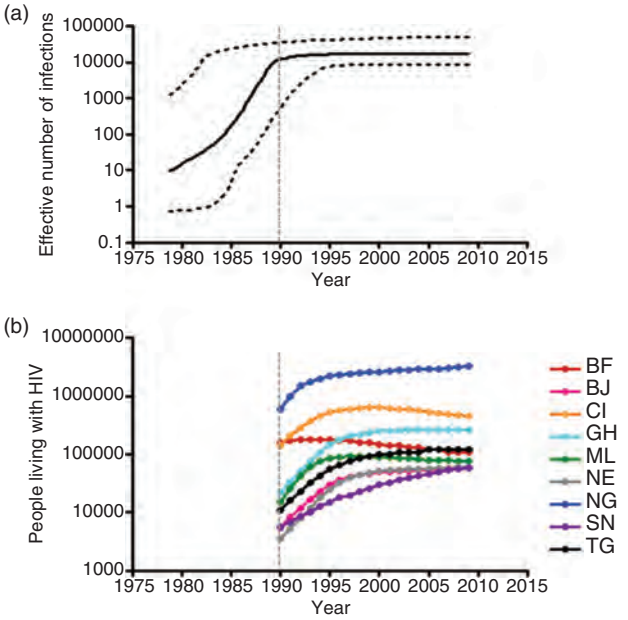


Fig. 4. Demographic analysis of the HIV-1 CRF06_cpx epidemic. (a) Bayesian skyline plot representing nonparametric estimates of effective number of infections through time for the HIV-1 CRF06_cpx epidemic in west Africa. Median estimate of the effective number of infections (solid line) and 95% confidence limits of the estimate (dashed lines) are shown. (b) Plot representing the UNAIDS estimated number of people living with HIV (http://www.unaids.org/globalreport/documents/20101123_GlobalReport_full_en.pdf) from 1990 to 2009 in those western African countries with a CRF06_cpx prevalence more than 2%. Vertical axes were represented on a logarithmic scale. BF, Burkina Faso; BJ, Benin; CI, Côte d'Ivoire; GH, Ghana; ML, Mali; NE, Niger; NG, Nigeria; SN, Senegal; TG, Togo.

operate in western Africa; or HIV-1 epidemic growth rates in Africa were previously underestimated.

In summary, this study suggests that the CRF06_cpx clade started to circulate in Burkina Faso around the late 1970s and was later spread to other countries from west and west-central African regions. The explanation for the current distribution of CRF06_cpx clade in Africa is probably multifactorial and includes human mobility, accessibility, distance from the epicenter, timing of viral dissemination, and temporal changes in regional epidemic growth pattern among others. Our data also highlight that the initial growth rate of the CRF06_cpx epidemic in western Africa was comparable to that estimated for HIV-1 epidemics in Europe and the United States. These findings offer important insights toward an understanding of the current characteristics and dynamics of the HIV-1 epidemic in western Africa. It will be important to incorporate in future studies new sequences from countries not represented or poorly represented here to test the major conclusions of our model.

Acknowledgements

We wish to thank Dr Vera Bongertz for critical review of the manuscript.

The study was conceived and designed by G.B. Data acquisition and analysis was performed by E.D. and G.B. G.B. wrote the first draft and E.D. contributed to the final version of the study.

Conflicts of interest

The authors have no conflict of interest.

References

1. Hemelaar J, Gouws E, Ghys PD, Osmanov S. **Global trends in molecular epidemiology of HIV-1 during 2000–2007.** *AIDS* 2011; **25**:679–689.
2. Montavon C, Toure-Kane C, Nkengasong JN, Vergne L, Hertogs K, Mboup S, *et al*. **CRF06-cpx: a new circulating recombinant form of HIV-1 in West Africa involving subtypes A, G, K, and J.** *J Acquir Immune Defic Syndr* 2002; **29**:522–530.
3. Ouedraogo-Traore R, Montavon C, Sanou T, Vidal N, Sangare L, Sanou I, *et al*. **CRF06-cpx is the predominant HIV-1 variant in AIDS patients from Ouagadougou, the capital city of Burkina Faso.** *AIDS* 2003; **17**:441–442.
4. Vergne L, Diagbouga S, Kouanfack C, Aghokeng A, Butel C, Laurent C, *et al*. **HIV-1 drug-resistance mutations among newly diagnosed patients before scaling-up programmes in Burkina Faso and Cameroon.** *Antivir Ther* 2006; **11**:575–579.
5. Tebit DM, Sangare L, Tiba F, Saydou Y, Makamtse A, Somlare H, *et al*. **Analysis of the diversity of the HIV-1 pol gene and drug resistance associated changes among drug-naive patients in Burkina Faso.** *J Med Virol* 2009; **81**:1691–1701.
6. Derache A, Maiga AI, Traore O, Akonde A, Cisse M, Jarrousse B, *et al*. **Evolution of genetic diversity and drug resistance mutations in HIV-1 among untreated patients from Mali between 2005 and 2006.** *J Antimicrob Chemother* 2008; **62**:456–463.
7. Maiga AI, Fofana DB, Maiga AC, Diallo F, Arkoub ZA, Daou F, *et al*. **Transmitted Antiretroviral Drug Resistance in Newly HIV-Infected and Untreated Patients in Segou and Bamako, Mali.** *AIDS Res Hum Retroviruses* 2012; **29**:182–186.
8. Mamadou S, Montavon C, Ben A, Djibo A, Rabiou S, Mboup S, *et al*. **Predominance of CRF02-AG and CRF06-cpx in Niger, West Africa.** *AIDS Res Hum Retroviruses* 2002; **18**:723–726.
9. Charpentier C, Bellecave P, Cisse M, Mamadou S, Diakite M, Peytavin G, *et al*. **High prevalence of antiretroviral drug resistance among HIV-1-untreated patients in Guinea-Conakry and in Niger.** *Antivir Ther* 2011; **16**:429–433.
10. Chamberland A, Diabate S, Sylla M, Anagounou S, Geraldo N, Zannou DM, *et al*. **Transmission of HIV-1 drug resistance in Benin could jeopardise future treatment options.** *Sex Transm Infect* 2012; **88**:179–183.
11. Fischetti L, Opare-Sem O, Candotti D, Sarkodie F, Lee H, Allain JP. **Molecular epidemiology of HIV in Ghana: dominance of CRF02_AG.** *J Med Virol* 2004; **73**:158–166.
12. Delgado E, Ampofo WK, Sierra M, Torpey K, Perez-Alvarez L, Bonney EY, *et al*. **High prevalence of unique recombinant forms of HIV-1 in Ghana: molecular epidemiology from an antiretroviral resistance study.** *J Acquir Immune Defic Syndr* 2008; **48**:599–606.
13. Ayouba A, Lien TT, Nouhin J, Vergne L, Aghokeng AF, Ngo-Giang-Huong N, *et al*. **Low prevalence of HIV type 1 drug resistance mutations in untreated, recently infected patients from Burkina Faso, Cote d'Ivoire, Senegal, Thailand, and Vietnam: the ANRS 12134 study.** *AIDS Res Hum Retroviruses* 2009; **25**:1193–1196.
14. Toni T, Masquelier B, Minga A, Anglaret X, Danel C, Coulibaly A, *et al*. **HIV-1 antiretroviral drug resistance in recently infected patients in Abidjan, Cote d'Ivoire: A 4-year survey, 2002–2006.** *AIDS Res Hum Retroviruses* 2007; **23**:1155–1160.

15. Chaplin B, Eisen G, Idoko J, Onwujekwe D, Idigbe E, Adewole I, *et al*. **Impact of HIV type 1 subtype on drug resistance mutations in Nigerian patients failing first-line therapy.** *AIDS Res Hum Retroviruses* 2011; **27**:71–80.
16. Hamel DJ, Sankale JL, Eisen G, Meloni ST, Mullins C, Gueye-Ndiaye A, *et al*. **Twenty years of prospective molecular epidemiology in Senegal: changes in HIV diversity.** *AIDS Res Hum Retroviruses* 2007; **23**:1189–1196.
17. Diop-Ndiaye H, Toure-Kane C, Leye N, Ngom-Gueye NF, Montavon C, Peeters M, *et al*. **Antiretroviral drug resistance mutations in antiretroviral-naive patients from Senegal.** *AIDS Res Hum Retroviruses* 2010; **26**:1133–1138.
18. Yaotse DA, Nicole V, Roch NF, Mireille PD, Eric D, Martine P. **Genetic characterization of HIV-1 strains in Togo reveals a high genetic complexity and genotypic drug-resistance mutations in ARV naive patients.** *Infect Genet Evol* 2009; **9**:646–652.
19. Dagnra AY, Vidal N, Mensah A, Patassi A, Aho K, Salou M, *et al*. **High prevalence of HIV-1 drug resistance among patients on first-line antiretroviral treatment in Lome, Togo.** *J Int AIDS Soc* 2011; **14**:30.
20. Esbjornsson J, Mild M, Mansson F, Norrgren H, Medstrand P. **HIV-1 molecular epidemiology in Guinea-Bissau, West Africa: origin, demography and migrations.** *PLoS One* 2011; **6**:e17025.
21. Lihana RW, Ssemwanga D, Abimiku A, Ndembi N. **Update on HIV-1 diversity in Africa: a decade in review.** *AIDS Rev* 2012; **14**:83–100.
22. Gray RR, Tatem AJ, Lamers S, Hou W, Laeyendecker O, Serwadda D, *et al*. **Spatial phylodynamics of HIV-1 epidemic emergence in east Africa.** *AIDS* 2009; **23**:F9–F17.
23. Delatorre EO, Bello G. **Phylodynamics of HIV-1 Subtype C Epidemic in East Africa.** *PLoS ONE* 2012; **7**:e41904.
24. Faria NR, Suchard MA, Abecasis A, Sousa JD, Ndembi N, Bonfim I, *et al*. **Phylodynamics of the HIV-1 CRF02_AG clade in Cameroon.** *Infect Genet Evol* 2012; **12**:453–460.
25. Jung M, Leye N, Vidal N, Fargette D, Diop H, Toure Kane C, *et al*. **The origin and evolutionary history of HIV-1 subtype C in Senegal.** *PLoS ONE* 2012; **7**:e33579.
26. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, *et al*. **An automated genotyping system for analysis of HIV-1 and other microbial sequences.** *Bioinformatics* 2005; **21**:3797–3800.
27. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, *et al*. **Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination.** *J Virol* 1999; **73**:152–160.
28. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. **New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.** *Syst Biol* 2010; **59**:307–321.
29. Posada D. **jModelTest: phylogenetic model averaging.** *Mol Biol Evol* 2008; **25**:1253–1256.
30. Drummond AJ, Rambaut A. **BEAST: Bayesian evolutionary analysis by sampling trees.** *BMC Evol Biol* 2007; **7**:214.
31. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. **Relaxed phylogenetics and dating with confidence.** *PLoS Biol* 2006; **4**:e88.
32. Lemey P, Rambaut A, Drummond AJ, Suchard MA. **Bayesian phylogeography finds its roots.** *PLoS Comput Biol* 2009; **5**:e1000520.
33. Bielejec F, Rambaut A, Suchard MA, Lemey P. **SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics.** *Bioinformatics* 2011; **27**:2910–2912.
34. Tatem AJ, Hemelaar J, Gray RR, Salemi M. **Spatial accessibility and the spread of HIV-1 subtypes and recombinants in sub-Saharan Africa.** *AIDS* 2012; **26**:2351–2360.
35. Salemi M, de Oliveira T, Soares MA, Pybus O, Dumans AT, Vandamme AM, *et al*. **Different epidemic potentials of the HIV-1B and C subtypes.** *J Mol Evol* 2005; **60**:598–605.
36. Tully DC, Wood C. **Chronology and evolution of the HIV-1 subtype C epidemic in Ethiopia.** *AIDS* 2010; **24**:1577–1582.
37. Walker PR, Pybus OG, Rambaut A, Holmes EC. **Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic growth.** *Infect Genet Evol* 2005; **5**:199–208.
38. Hue S, Pillay D, Clewley JP, Pybus OG. **Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups.** *Proc Natl Acad Sci U S A* 2005; **102**:4425–4429.
39. Zehender G, Ebranati E, Lai A, Santoro MM, Alteri C, Giuliani M, *et al*. **Population dynamics of HIV-1 subtype B in a cohort of men-having-sex-with-men in Rome, Italy.** *J Acquir Immune Defic Syndr* 2010; **55**:156–160.
40. Bello G, Soares MA, Schrago CG. **The use of bioinformatics for studying HIV evolutionary and epidemiological history in South America.** *AIDS Res Treat* 2011; **2011**:154945.

8.4 APÊNDICE 4

Phylodynamics of the HIV-1 Epidemic in Cuba

Autores: **Edson Delatorre** e Gonzalo Bello

Jornal: PLOS ONE, 2013

OPEN ACCESS Freely available online

PLOS ONE

Phylodynamics of the HIV-1 Epidemic in Cuba

Edson Delatorre, Gonzalo Bello*

Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil

Abstract

Previous studies have shown that the HIV-1 epidemic in Cuba displayed a complex molecular epidemiologic profile with circulation of several subtypes and circulating recombinant forms (CRF); but the evolutionary and population history of those viral variants remains unknown. HIV-1 *pol* sequences of the most prevalent Cuban lineages (subtypes B, C and G, CRF18_cpx, CRF19_cpx, and CRFs20/23/24_BG) isolated between 1999 and 2011 were analyzed. Maximum-likelihood analyses revealed multiple introductions of subtype B ($n \geq 66$), subtype C ($n \geq 10$), subtype G ($n \geq 8$) and CRF18_cpx ($n \geq 2$) viruses in Cuba. The bulk of HIV-1 infections in this country, however, was caused by dissemination of a few founder strains probably introduced from North America/Europe (clades $B_{CU\text{-}I}$ and $B_{CU\text{-}II}$), east Africa (clade $C_{CU\text{-}I}$) and central Africa (clades G_{CU}, $CRF18_{CU}$ and $CRF19_{CU}$), or locally generated (clades CRFs20/23/24_BG). Bayesian-coalescent analyses show that the major HIV-1 founder strains were introduced into Cuba during 1985–1995; whereas the CRFs_BG strains emerged in the second half of the 1990s. Most HIV-1 Cuban clades appear to have experienced an initial period of fast exponential spread during the 1990s and early 2000s, followed by a more recent decline in growth rate. The median initial growth rate of HIV-1 Cuban clades ranged from 0.4 $year^{-1}$ to 1.6 $year^{-1}$. Thus, the HIV-1 epidemic in Cuba has been a result of the successful introduction of a few viral strains that began to circulate at a rather late time of the AIDS pandemic, but then were rapidly disseminated through local transmission networks.

Citation: Delatorre E, Bello G (2013) Phylodynamics of the HIV-1 Epidemic in Cuba. PLoS ONE 8(9): e72448. doi:10.1371/journal.pone.0072448

Editor: Chiyu Zhang, Institut Pasteur of Shanghai, Chinese Academy of Sciences, China

Received June 14, 2013; **Accepted** July 17, 2013; **Published** September 9, 2013

Funding: This work was supported by Public Health Service grants E-26/111.758/2012 from the FAPERJ and 472896/2012-1 from the CNPq. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gbellobr@gmail.com

Introduction

The global dissemination of the Human immunodeficiency virus type 1 (HIV-1) group M clade during the second half of the twentieth century has resulted in the generation of a diverse collection of genetic variants classified into subtypes, sub-subtypes, circulating recombinant forms (CRFs) and unique recombinants forms (URFs). The HIV-1 epidemic in the Americas is typically dominated by subtype B clade, although substantial proportions (≥20%) of non-B subtype genetic forms are observed in Argentina, Brazil, Cuba and Uruguay [1].

Cuba displayed a unique HIV-1 molecular epidemiologic profile in the Americas characterized by the co-circulation of several subtypes (A1, B, C, F1, G, H and J), CRFs and URFs. Subtype B is the most prevalent variant (~33–40%), followed by CRF19_cpx (~20–28%), CRFs20/23/24_BG (~12–20%) CRF18_cpx (~7–10%), subtype C (~3–10%), and subtype G (~2–7%) [2,3,4,5,6,7]. It has been proposed that the presence of numerous Cuban military and civilian personnel in several sub-Saharan African countries, and particularly those stationed in Angola and neighboring countries between 1975 and 1991, have contributed to the introduction of multiple non-B HIV-1 subtypes into Cuba [2]. Some HIV-1 recombinants including CRF18_cpx and CRF19_cpx were probably also imported into Cuba directly from central Africa, since the parental viruses of these complex genetic forms were only detected in that African region [8,9]. Indeed, a few cases of CRF18_cpx and CRF19_cpx like viruses have been confirmed in Angola [10,11], Democratic Republic of Congo (DRC) [12,13], Republic of Congo [14,15], Central African Republic [16], and Cameroon [17,18,19]. Other HIV-1 recombinants including all three CRFs_BG, however, were probably generated locally by recombination between subtypes B and G already circulating in Cuba [20].

According to this model, most non-B subtype HIV-1 variants circulating in Cuba were probably introduced or locally generated after 1975. Despite the extensive knowledge about the molecular epidemiology of HIV-1 variants, the time-scale and epidemic history of most prevalent HIV-1 clades circulating in Cuba remains to be elucidated. In this study, we used a Bayesian coalescent-based method and a comprehensive data set of HIV-1 subtype B ($n = 322$), and non-B subtypes ($n = 420$) *pol* sequences of Cuban origin isolated between 1999 and 2011, to date the origin and reconstruct the demographic history of major HIV-1 variants circulating in Cuba.

Materials and Methods

HIV-1 Cuban sequence datasets

We downloaded all HIV-1 Cuban sequences covering the entire protease and partial reverse transcriptase (PR/RT) regions of the *pol* gene (nt 2253–3272 relative to HXB2 clone) classified as subtypes B ($n = 322$), C ($n = 49$), G ($n = 35$), CRF18_cpx ($n = 71$), CRF19_cpx ($n = 167$), and CRFs20/23/24_BG ($n = 118$) that were available at the Los Alamos HIV Sequence Database (www.hiv.lanl.gov) by March 2013. HIV-1 *pol* sequences were retrieved from both antiretroviral therapy naïve and HAART treated patients from different Cuban regions between 1999 and 2011, as described in previous studies [2,3,4,5,6]. Sequences were aligned using the CLUSTAL X program [21]. To avoid any bias on the

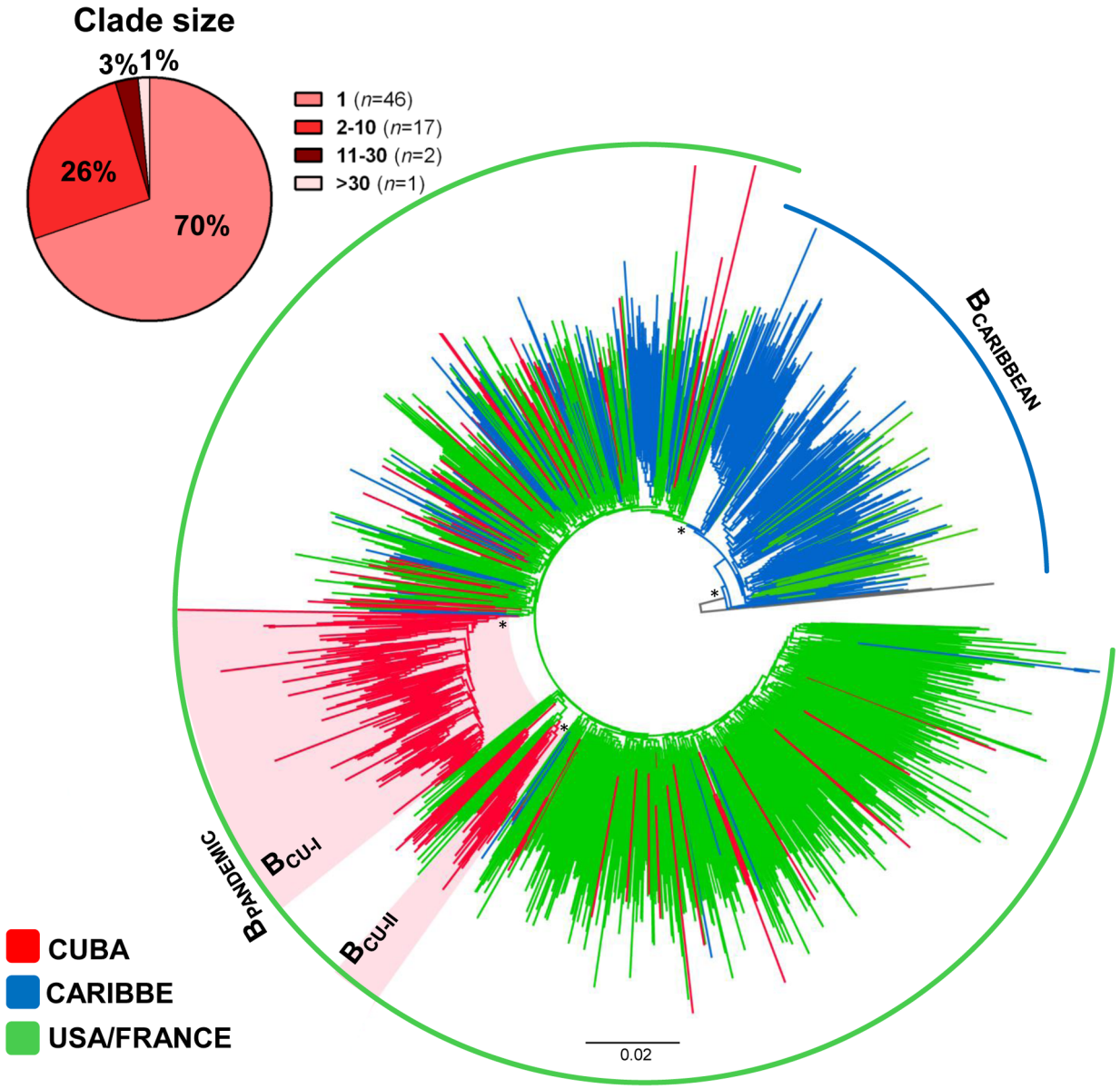


Figure 1. ML phylogenetic tree of HIV-1 subtype B *pol* (~1000 pb) sequences circulating in Cuba ($n=322$), US ($n=525$), France ($n=348$), and other Caribbean countries ($n=418$). The branches are colored according to the origin of each sequence, as indicated at the legend (bottom left). The circular brackets highlight the position of the pandemic ($B_{PANDEMIC}$, green line) and non-pandemic ($B_{CARIBBEAN}$, blue line) HIV-1 subtype B clades. Shaded boxes highlight the position of the two major HIV-1 subtype B Cuban clades (B_{CU-I} and B_{CU-II}). The number of Cuban sequences distributed accordingly to the clade size is shown (top left). Key nodes with *a*LRT support values >0.80 (*) and ≥0.90 (**) are indicated. The tree was rooted using HIV-1 subtype D reference sequences (gray branches). The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.
doi:10.1371/journal.pone.0072448.g001

phylogenetic reconstructions, all sites with major antiretroviral drug resistance mutations in PR (30, 32, 46, 47, 48, 50, 54, 76, 82, 84, 88 and 90) or RT (41, 65, 67, 69, 70, 74, 100, 101, 103, 106, 115, 138, 151, 181, 184, 188, 190, 210, 215, 219 and 230) detected in at least two sequences were excluded from each alignment. All alignments are available from the authors upon request.

HIV-1 reference datasets

HIV-1 Cuban sequences were combined with reference sequences of diverse origin that matched the selected genomic region and were available at the Los Alamos HIV Sequence Database. Subtype B Cuban sequences were aligned with reference sequences representative of the viral diversity in US ($n=525$), France ($n=348$) and the Caribbean ($n=417$) (Table S1). Subtype C Cuban sequences were aligned with representative sequences from central ($n=53$), eastern ($n=330$) and southern ($n=545$) African regions (Table S2). The HIV-1 subtype G Cuban sequences were combined with all available subtype G sequences of African origin ($n=437$) (Table S3). The CRF19_cpx Cuban sequences were aligned with all available CRF19_cpx sequences from other countries ($n=3$) and subtype D sequences of African origin ($n=1,112$) (Table S4). Finally, the HIV-1 CRF18_cpx and CRFs_BG Cuban sequences were combined with all available

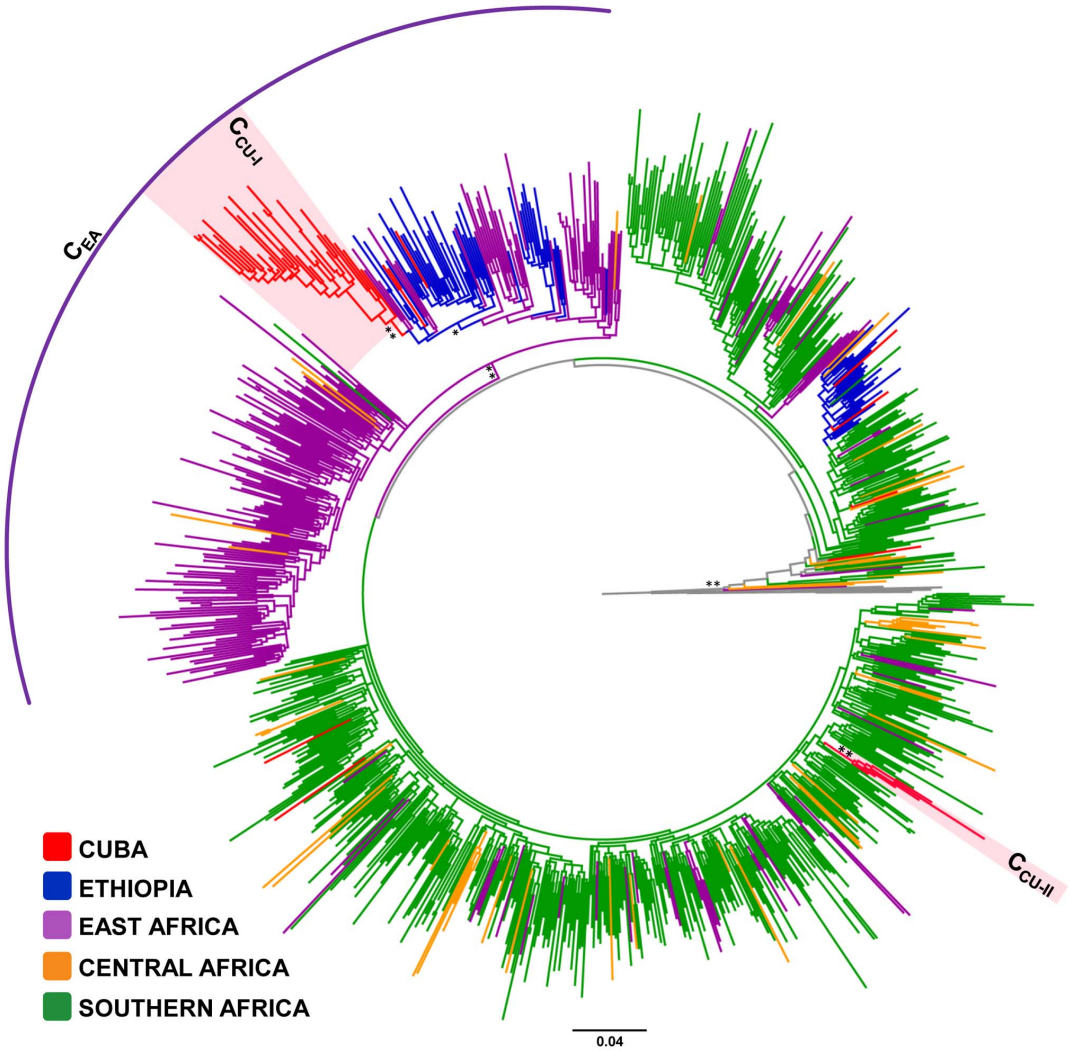


Figure 2. ML phylogenetic tree of HIV-1 subtype C *pol* (~1000 pb) sequences circulating in Cuba ($n=49$), and in central ($n=53$), eastern ($n=330$) and southern ($n=545$) African countries. Branches are colored according to the origin of each sequence, as indicated at the legend (bottom left). The circular bracket highlights the position of the subtype C east African clade (C_{EA}). Shaded boxes highlight the position of the two major HIV-1 subtype C Cuban clades (C_{CU-I} and C_{CU-II}). Key nodes with *a*LRT support values >0.80 (*) and ≥0.90 (**) are indicated. The tree was rooted using HIV-1 subtype A1 and D reference sequences (gray branches). The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.
doi:10.1371/journal.pone.0072448.g002

CRF18_cpx ($n=15$) and CRFs20/23/24_BG ($n=7$) sequences from other countries (Tables S4 and S5).

Subtype assignment

The subtype assignment and recombinant structure of all sequences here included was confirmed by: REGA HIV subtyping tool v.2 [22]; Bootscanning with Simplot software v3.5.1 [23] and Maximum Likelihood (ML) phylogenetic analysis. In bootscan analyses, supporting branching with reference sequences from all HIV-1 group M subtypes were determined in Neighbor-Joining trees based on 100 re-samplings, within a 250 bp window moving in steps of 10 bases. ML phylogenetic trees were inferred under the best nucleotide substitution model selected using the jModeltest program [24] (Table S6). The ML tree was reconstructed with the PhyML program [25] using an online web server [26]. Heuristic tree search was performed using the SPR branch-swapping algorithm and the reliability of the obtained topology was estimated with the approximate likelihood-ratio test (*aLRT*) [27] based on the Shimodaira-Hasegawa-like procedure. The ML trees were visualized using the FigTree v1.4.0 program [28]. All HIV-1 sequences displaying incorrect clade assignment and/or frameshift mutations were excluded from the study, with the exception of four CRF23_BG sequences that were reclassified as CRF20_BG (GenBank accession numbers FJ481689 and FJ585687) and CRF24_BG (GenBank accession numbers JQ585465 and FJ481688).

Reconstruction of evolutionary and demographic history

The evolutionary rate (μ, nucleotide substitutions per site per year, subst./site/year), the age of the most recent common ancestor ($T_{mrca,}$ years), and the mode and rate (r, years^{-1}) of population growth of different Cuban HIV-1 clades were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST v1.7.5 [29,30]. Analyses were performed using the best nucleotide substitution model (Table S6) and an uncorrelated Lognormal relaxed

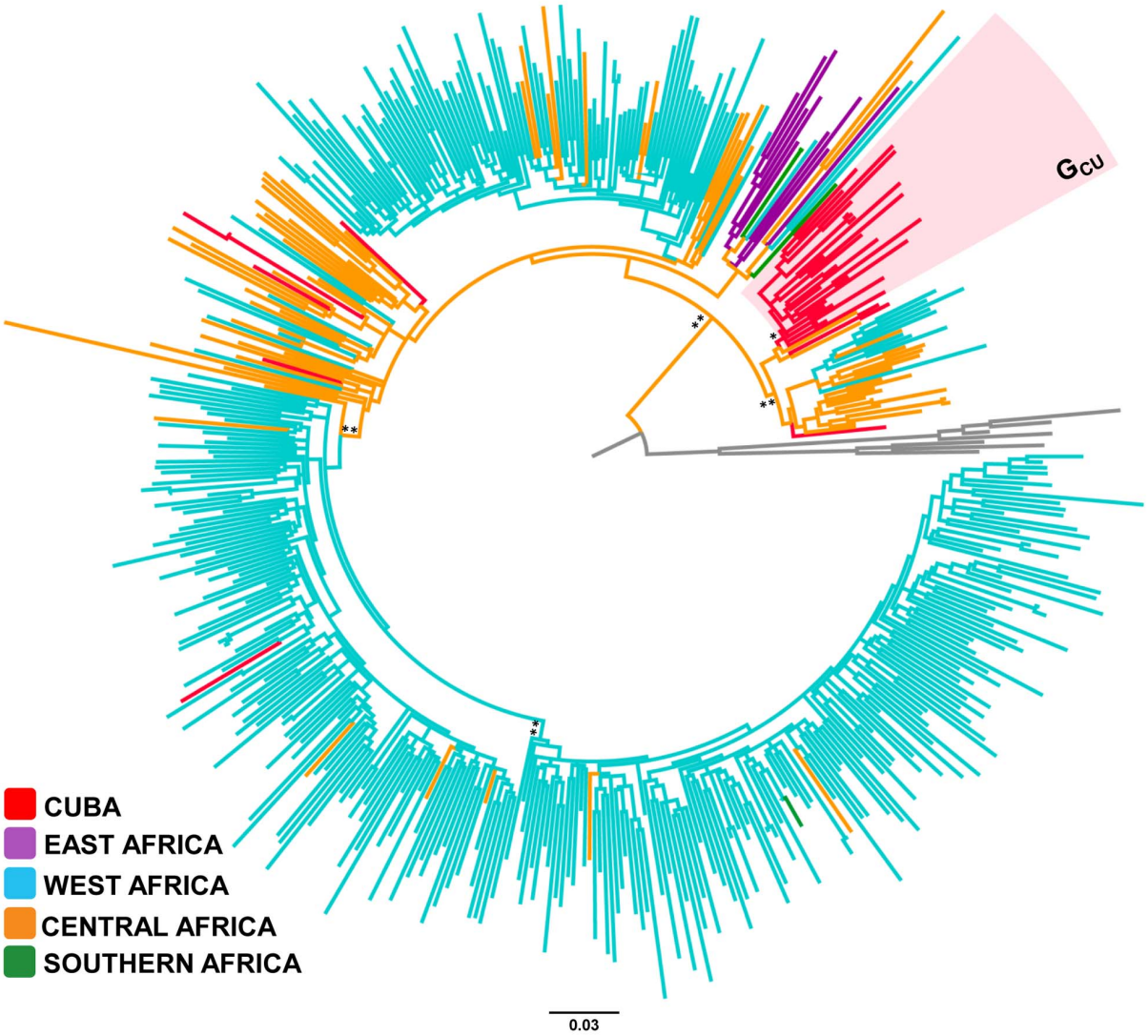


Figure 3. ML phylogenetic tree of HIV-1 subtype G *pol* (~1000 pb) sequences circulating in Cuba ($n=35$), and in central ($n=71$), western ($n=366$), eastern ($n=10$) and southern ($n=3$) African countries. Branches are colored according to the origin of each sequence, as indicated at the legend (bottom left). Shaded boxes highlight the position of the major HIV-1 subtype G Cuban clade (G_{CU}). Key nodes with *a*LRT support values >0.80 (*) and ≥0.90 (**) are indicated. The tree was rooted using HIV-1 subtype A1 and B reference sequences (gray branches). The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.
doi:10.1371/journal.pone.0072448.g003

molecular clock model [31]. A Bayesian Skyline coalescent tree prior [32] was first used to estimate μ, the T_{mrca}, and the change in effective population size through time. Estimates of the population growth rate were subsequently obtained using a logistic growth coalescent tree prior that was the model pointed out by the Bayesian Skyline plot and that also provided the best fit to the demographic signal contained in most datasets. Comparison between demographic models was performed using the log marginal likelihood (ML) estimation based on path sampling (PS) and stepping-stone sampling (SS) methods [33]. MCMC chains were run for $10–50\times10^6$ generations. Adequate chain mixing and uncertainty in parameter estimates were assessed by calculating the effective sample size (ESS) and the 95% Highest Probability Density (HPD) values respectively, after excluding an initial 10% using the TRACER v1.5 program [34].

Results

Identification of major HIV-1 Cuban clades

The ML analysis of HIV-1 subtypes B, C and G sequences from Cuba and other countries from the Americas, Europe and Africa revealed that most Cuban strains branched in well-supported country-specific sub-clades. Of the 322 HIV-1 subtype B Cuban sequences analyzed, 180 (56%) formed a large country-specific monophyletic sub-clade ($B_{CU\text{-}I}$, $aLRT=0.81$), 44 (14%) branched in two clusters of medium size ($15<n<30$), 52 (16%) formed clusters of small size ($n\leq10$), and the remaining 46 (14%) represented non-clustered sequences (Fig. 1). Of note, all subtype B Cuban sequences branched in a large $B_{PANDEMIC}$ monophyletic cluster ($aLRT=0.80$) together with most subtype B sequences from US (92%) and all sequences from France (100%); whereas most non-Cuban Caribbean sequences (60%) occupy the deepest branches within B clade (Fig. 1).

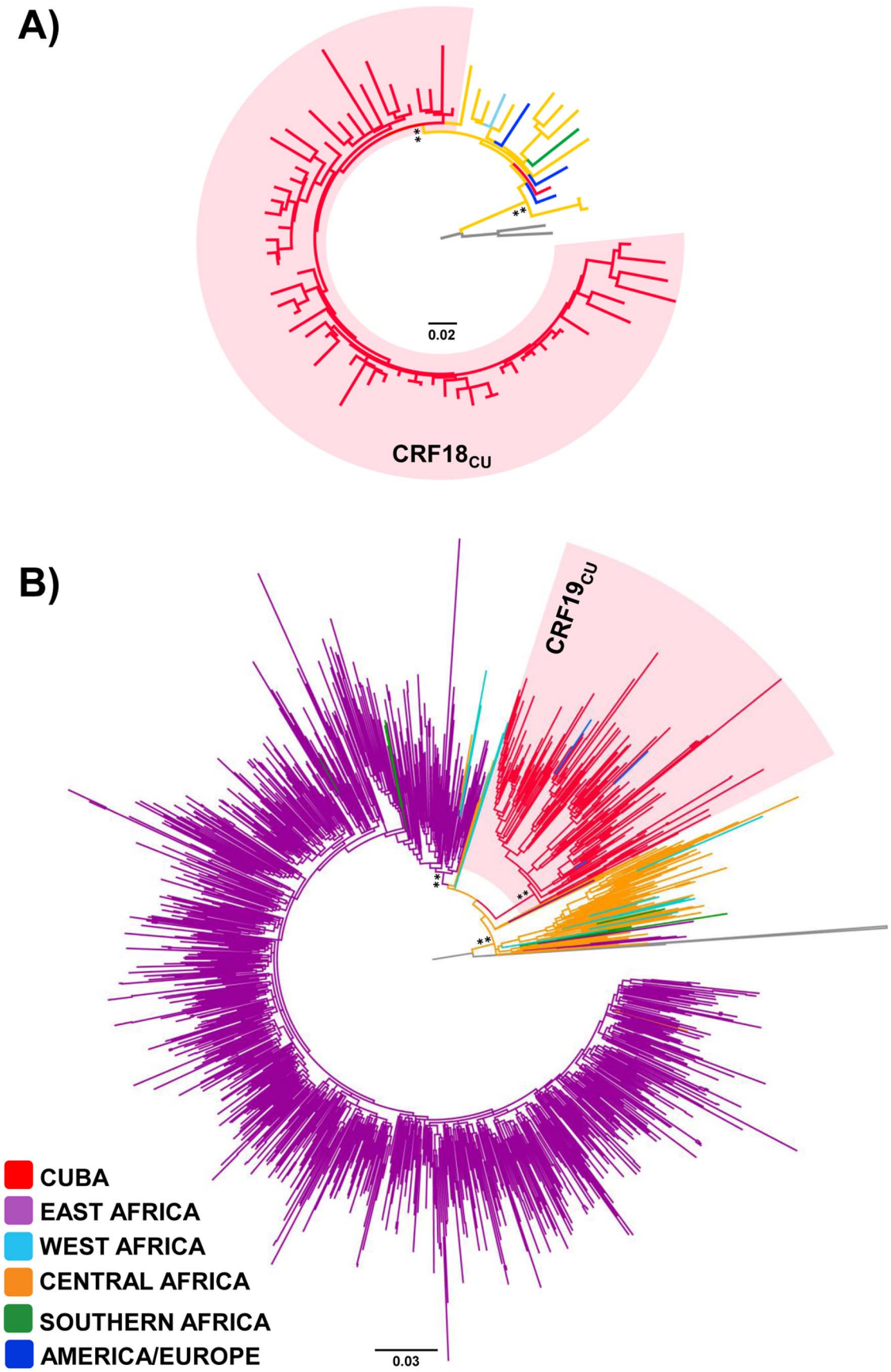

A)
0.02
CRF18CU
B)
CRF19CU
CUBA
EAST AFRICA
WEST AFRICA
CENTRAL AFRICA
SOUTHERN AFRICA
AMERICA/EUROPE
0.03

Figure 4. ML phylogenetic trees of HIV-1 CRFs_cpx *pol* (~1000 pb) sequences. A) HIV-1 CRF18_cpx from Cuba ($n = 62$), were combined with those isolated in African ($n = 12$), American ($n = 1$) and European ($n = 2$) countries. The tree was rooted using HIV-1 subtype G reference sequences (black branches). B) HIV-1 CRF19_cpx sequences from Cuba ($n = 160$) and European countries ($n = 3$) were combined with subtype D sequences of African origin ($n = 1,112$). Branches are colored according to the origin of each sequence, as indicated at the legend (bottom left). Shaded boxes highlight the position of the major HIV-1 CRF18_cpx ($CRF18_{CU}$) and CRF19_cpx ($CRF19_{CU}$) Key nodes with *a*LRT support values >0.80 (*) and ≥0.90 (**) are indicated. The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.
doi:10.1371/journal.pone.0072448.g004

Of the 49 HIV-1 subtype C Cuban sequences analyzed, 34 (69%) branched in a single monophyletic sub-cluster ($C_{CU\text{-}I}$, $aLRT = 0.94$), six (12%) branched in a second well supported minor clade ($C_{CU\text{-}II}$, $aLRT = 0.98$), and the remaining nine (18%) represented non-clustered sporadic lineages (Fig. 2). The major clade $C_{CU\text{-}I}$ was nested within Ethiopian sequences that belongs to the previously called C_{EA} clade [35], a viral lineage characteristic of the east African region (Fig. 2). The minor clade $C_{CU\text{-}II}$, by contrast, was nested within subtype C sequences from southern Africa (Fig. 2). Non-clustered Cuban sequences were scattered among strains from Ethiopia and southern African countries.

Of the 35 HIV-1 subtype G Cuba sequences analyzed, 26 (74%) branched in a single monophyletic sub-cluster (G_{CU}, $aLRT = 0.87$) and the remaining nine (26%) represented sporadic lineages of one or two sequences. Although most subtype G African strains included in our analysis were from the western region ($n = 366$, 84%), the clade G_{CU} and most sporadic subtype G Cuban lineages were nested among basal sequences from the central African region (Angola, DRC and Cameroon) (Fig. 3). There was only one

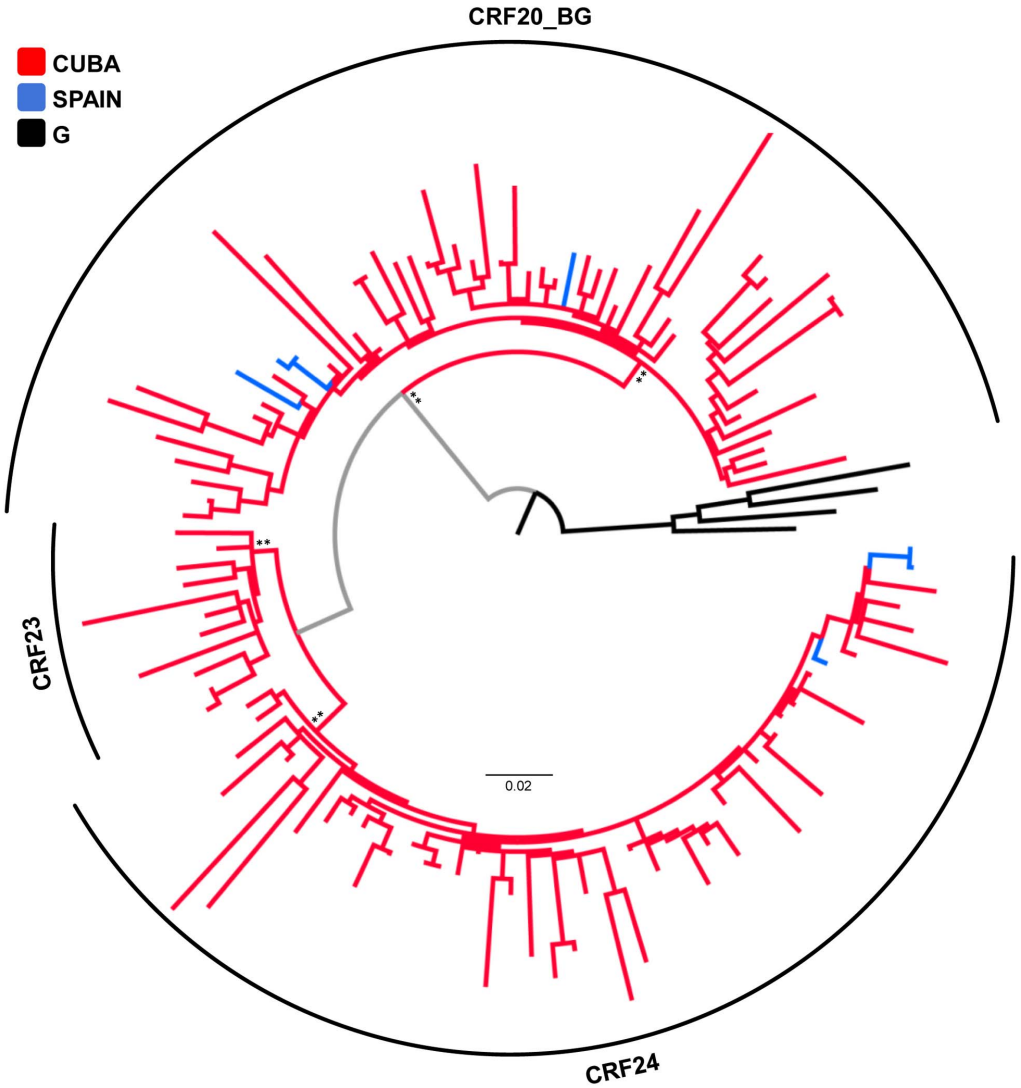


Figure 5. Ml phylogenetic tree of HIV-1 CRFs20/23/24_BG *pol* (~1000 pb) sequences circulating in Cuba ($n = 118$) and Spain ($n = 7$). Branches are colored according to the origin of each sequence, as indicated at the legend (top left). The circular brackets highlight the distribution of the three CRFs_BG clades. The tree was rooted using HIV-1 subtype G reference sequences (black branches). Key nodes with *a*LRT support values >0.80 (*) and ≥0.90 (**) are indicated. The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.
doi:10.1371/journal.pone.0072448.g005

Table 1. Evolutionary rate and time-scale of major HIV-1 Cuban clades.

HIV-1 clade	*N*	Sampling interval	μ (subst./site/year)	Coefficient of variation	T_{MRCA}
B_{CU-I}	176	2003–2011	3.0×10^{-3}	0.30	1992
			(2.4×10^{-3}–3.6×10^{-3})	(0.21–0.39)	(1988–1994)
B_{CU-II}	27	1999–2011	2.4×10^{-3}	0.25	1991
			(1.6×10^{-3}–3.2×10^{-3})	(0.01–0.45)	(1986–1994)
C_{CU}	34	2003–2011	2.8×10^{-3}	0.41	1994
			(2.0×10^{-3}–3.8×10^{-3})	(0.19–0.65)	(1990–1998)
G_{CU}	26	1999–2011	2.0×10^{-3}	0.56	1988
			(1.0×10^{-3}–3.3×10^{-3})	(0.36–0.81)	(1976–1995)
$CRF18_{CU}$	61	1999–2011	2.6×10^{-3}	0.40	1992
			(1.9×10^{-3}–3.5×10^{-3})	(0.25–0.59)	(1987–1996)
$CRF19_{CU}$	158	1999–2011	3.4×10^{-3}	0.38	1987
			(2.9×10^{-3}–4.0×10^{-3})	(0.30–0.47)	(1983–1991)
CRF20/23/24_BG	117	1999–2011	2.6×10^{-3}	0.35	1991
			(2.1×10^{-3}–3.1×10^{-3})*	(0.25–0.45)*	(1986–1994)*
CRF20_BG	56	1999–2011	2.6×10^{-3}	0.35	1996
			(2.1×10^{-3}–3.1×10^{-3})*	(0.25–0.45)*	(1994–1998)*
			2.4×10^{-3}	0.27	1996
			(1.8×10^{-3}–3.0×10^{-3})	(0.10–0.44)	(1994–1998)
CRF23_BG	11	2003–2011	2.6×10^{-3}	0.35	1998
			(2.1×10^{-3}–3.1×10^{-3})*	(0.25–0.45)*	(1996–2000)*
CRF24_BG	50	2003–2011	2.6×10^{-3}	0.35	1997
			(2.1×10^{-3}–3.1×10^{-3})*	(0.25–0.45)*	(1996–1999)*
			2.2×10^{-3}	0.36	1998
			(1.6×10^{-3}–2.8×10^{-3})	(0.19–0.54)	(1996–2000)

*Estimates obtained from the combined CRF20/23/24_BG data set.

doi:10.1371/journal.pone.0072448.t001

Cuban sequence that branched within a major African subtype G sub-clade mainly composed by sequences from Nigeria.

Test the monophyletic origin of the HIV-1 CRFs_cpx Cuban sequences was very much complicated because the scarcity of CRF18_cpx ($n = 12$) and the absence of CRF19_cpx *pol* sequences of African origin available in public databases. Because CRF19_cpx is subtype D in the *pol* fragment analyzed, we decided to include all available subtype D *pol* sequences of African origin in our analysis. ML analysis revealed that all (except one) CRF18_cpx and all CRF19_cpx sequences from Cuba branched in highly supported ($aLRT \geq 0.90$) monophyletic sub-clusters ($CRF18_{CU}$ and $CRF19_{CU}$) that were nested within CRF18_cpx and subtype D *pol* sequences of central African origin, respectively (Fig. 4). The few CRF18_cpx isolated in Europe ($n = 2$) and South America ($n = 1$) were intermixed among basal African strains; whereas all CRF19_cpx detected in Europe ($n = 3$) branched within the clade $CRF19_{CU}$ (Fig. 4).

As expected, the CRFs20/23/24_BG Cuban sequences formed three well-supported ($aLRT \geq 0.90$) monophyletic lineages (Fig. 5). The few CRF20_BG ($n = 4$) and CRF24_BG ($n = 3$) sequences isolated outside Cuba (Spain and Greece) were intermixed among Cuban strains (Fig. 5); thus supporting a Cuban origin for all those European sequences.

Time scale of major HIV-1 Cuban clades

Bayesian MCMC analyses under a relaxed molecular clock model were used to estimate the substitution rate and T_{MRCA} of all HIV-1 Cuban clades with a minimum size of 25 sequences. A few subtype B ($n = 4$) and CRF19_cpx ($n = 2$) sequences with anomalously long branches in the phylogenetic tree, were excluded. The final number of HIV-1 Cuban sequences included in the evolutionary analyses is shown in Table 1. The median estimated evolutionary rates for the *pol* region of the different HIV-1 clades were roughly similar, ranging from 2.0×10^{-3} subst./site/year (G_{CU} clade) to 3.4×10^{-3} subst./site/year ($CRF19_{CU}$ clade), with a considerable overlap of the 95% HPD intervals (Table 1). The coefficient of rate variation was higher than zero for all HIV-1 datasets analyzed (Table 1), thus supporting the use of a relaxed molecular clock model to reconstruct the time-scale of major HIV-1 Cuban lineages.

The median T_{MRCA} of those HIV-1 clades imported into Cuba range between 1987 ($CRF19_{CU}$) and 1994 (C_{CU-I}); whereas the median T_{MRCA} of those CRF_BG variants locally generated varied between 1996 and 1998 (Table 1). The T_{MRCA} of CRF20_BG and CRF24_BG clades estimated from the single CRF datasets were almost identical to those estimated from the combined CRFs20/23/24_BG data set (Table 1), indicating that all Cuban CRFs_BG evolved at quite similar rates. A previous study [20], proposed that Cuban CRF_BG viruses derive from a common recombinant ancestor generated by recombination between clade G_{CU} and the second most prevalent subtype B clade (B_{CU-II}) (Fig. 1). The analysis of the combined CRFs20/23/24_BG data set allows us to estimate the median T_{MRCA} of that putative BG recombinant ancestor at 1991, roughly coinciding

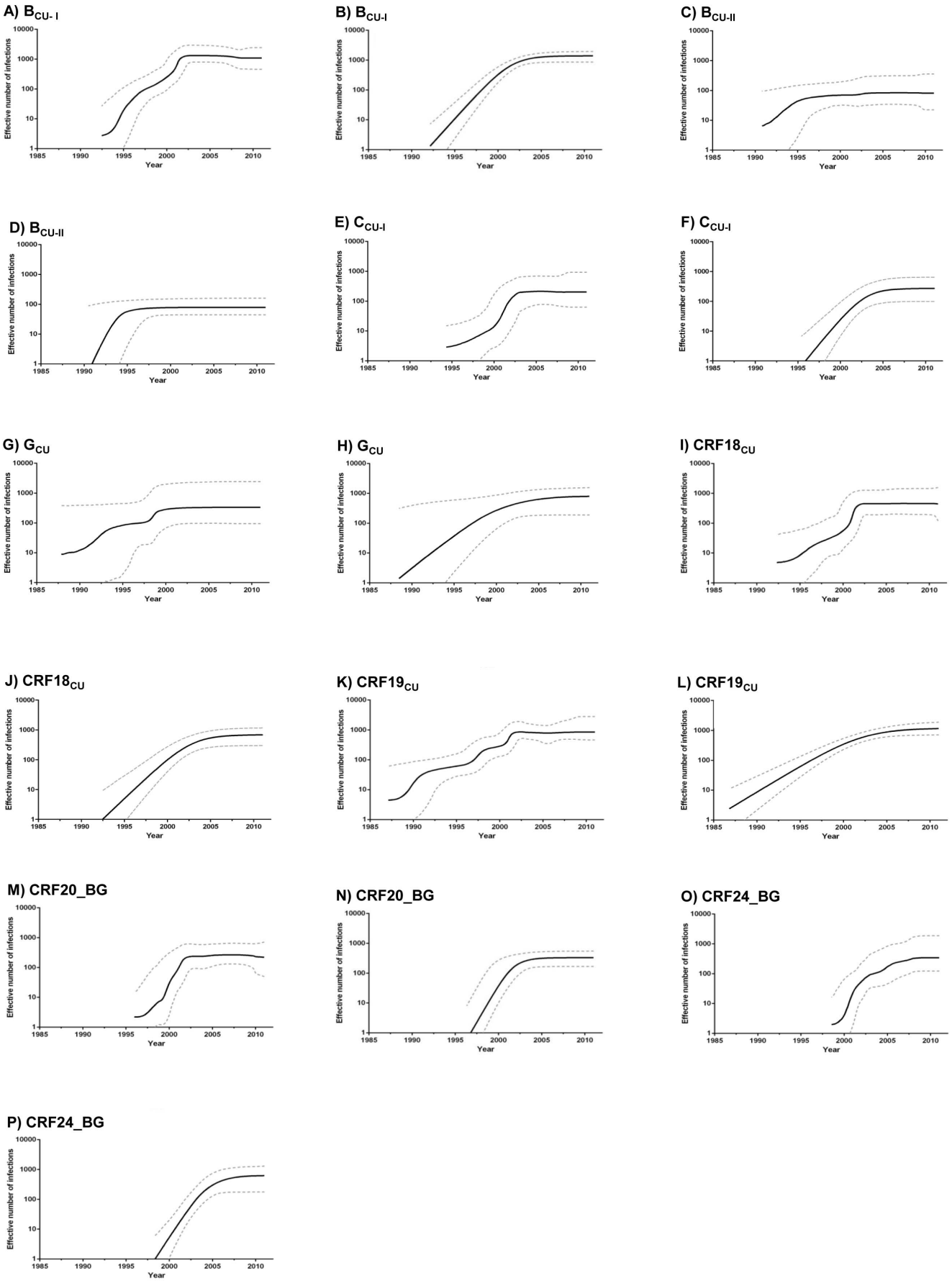

A) B_{CU-I}
B) B_{CU-I}
C) B_{CU-II}
D) B_{CU-II}
E) C_{CU-I}
F) C_{CU-I}
G) G_{CU}
H) G_{CU}
I) $CRF18_{CU}$
J) $CRF18_{CU}$
K) $CRF19_{CU}$
L) $CRF19_{CU}$
M) CRF20_BG
N) CRF20_BG
O) CRF24_BG
P) CRF24_BG
Effective number of infections
Year
10000
1000
100
10
1
1985
1990
1995
2000
2005
2010

Figure 6. Demographic history of the major HIV-1 Cuban clades. Effective number of infections through time estimated using both Bayesian skyline (A, C, E, G, I, K, M and O) and logistic growth (B, D, F, H, J, L, N and P) coalescent models are shown for each of the eight HIV-1 Cuban clades analyzed. Median estimates of the effective number of infections (solid line) and 95% HPD intervals of the estimates (dashed lines) are shown in each graphic. The vertical axes represent the estimated effective number of infections on a logarithmic scale. Time scale is in calendar years.
doi:10.1371/journal.pone.0072448.g006

with the estimated T_{MRCA} of the parental clades G_{CU} and $B_{CU\text{-}II}$ (Table 1).

Demographic history of major HIV-1 Cuban clades

The Bayesian skyline plot (BSP) analyses suggest that all HIV-1 Cuban clades experienced an initial phase of fast exponential growth followed by a more recent decline in growth rate (Fig. 6). The growth rate of most HIV-1 Cuban clades seems to start to decrease around the early 2000s; except for clades $B_{CU\text{-}II}$ and CRF24_BG that seem to stabilize at earlier (before 2000) and later (after 2005) time points, respectively. The BSP analyses also suggests that the coalescent model of logistic population growth fits the demographic information contained in all HIV-1 Cuban data sets better than the exponential one.

To test this, log ML for the logistic and exponential growth models were calculated using both PS and SS methods. The model of logistic population growth was strongly supported over the exponential one for most HIV-1 Cuban clades (log BF$>$3), with exception of G_{CU} and CRF24_BG for which only a weak support was obtained (log BF = 0.9–1.0) (Table 2). Such a low BF support to the logistic growth model could be explained by the low number of sequences in clade G_{CU} ($n = 26$) and the very recent stabilization of clade CRF24_BG (after 2005). Moreover, the overall time-scale and demographic pattern obtained from both BSP and logistic growth coalescent tree priors were very similar for all HIV-1 Cuban clades (Fig. 6). According to the logistic model, the median initial growth rates of HIV-1 Cuban clades range between 0.40 year^{-1} ($CRF19_{CU}$) to 1.57 year^{-1} ($B_{CU\text{-}II}$) with some overlap of the 95% HPD intervals for most lineages, except between $CRF19_{CU}$ and clades $B_{CU\text{-}I}$, $B_{CU\text{-}II}$, $C_{CU\text{-}I}$, CRF20_BG and CRF24_BG (Fig. 7).

Discussion

The Cuban HIV epidemic is unique in the Americas because of the exceptionally low HIV prevalence, estimated at 0.20% in adults in 2011 [36], and the unusually high HIV-1 genetic diversity with circulation of subtype B and several non-B subtypes [2,3,4,5,6,7]. Our study indicates that most HIV-1 infections in Cuba derived from the dissemination of a few founder viruses that were either introduced from the Americas/Europe (subtype B) and Africa (subtype C, subtype G, CRF18_cpx and CRF19_cpx) or were locally generated (CRFs20/23/24_BG).

The most accepted model of worldwide HIV-1 subtype B dissemination suggests that the virus moved from Haiti to other Caribbean islands and to the United States (US), and then from the US to the rest of the world establishing a "$B_{PANDEMIC}$" clade [37]. The phylogenetic analysis here performed revealed multiple ($n \geq 66$) introductions of HIV-1 $B_{PANDEMIC}$ strains in Cuba, although the bulk of the subtype B epidemic in this country resulted from the dissemination of only a few clades. The two most prevalent clades $B_{CU\text{-}I}$ and $B_{CU\text{-}II}$ comprises about 55% and 8% of all subtype B sequences from Cuba here included, respectively. We estimate that these clades most probably emerged in Cuba in the early 1990s, much later than the estimated origin of subtype B epidemics in Haiti and the US (1960–1970) [37,38,39]. The estimated T_{MRCA} of clades $B_{CU\text{-}I}$ and $B_{CU\text{-}II}$ coincides with a crisis in the Cuban economy caused by the collapse of the Soviet Union in 1991 that precipitated important investments in the tourist industry and a sharp increase in the number of tourist mostly from

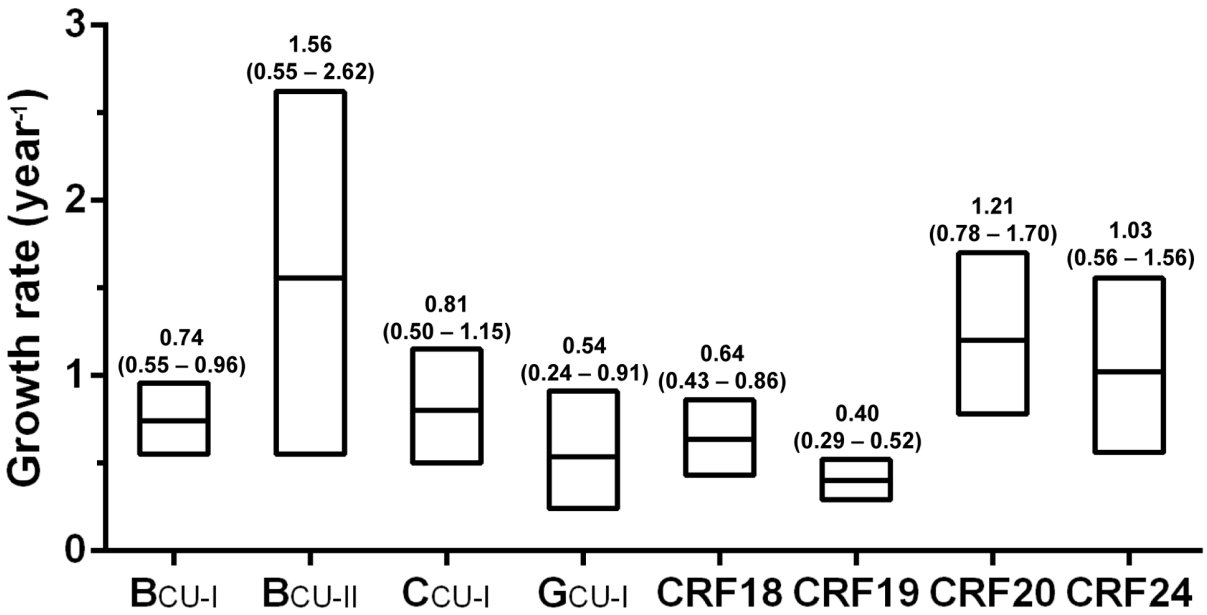


Figure 7. Coalescent estimates of epidemic growth rate of the major HIV-1 Cuban clades. The box plots and the numbers above represent the median growth rates (years^{-1}) and the 95% HPD intervals of the posterior distributions estimated under the logistic growth coalescent model for each of the eight HIV-1 Cuban clades analyzed.
doi:10.1371/journal.pone.0072448.g007

Table 2. Best fit demographic model for major HIV-1 Cuban clades.

Dataset	PS			SS		
	Log ML LG	Log ML EG	Log BF (LG vs EG)	Log ML LG	Log ML EG	Log BF (LG vs EG)
$B_{CUBA\ I}$	−12891.71	−12935.39	43.68	−12894.56	−12939.47	43.91
$B_{CUBA\ II}$	−3563.92	−3568.51	4.59	−3563.99	−3568.71	4.72
C_{CUBA}	−4114.63	−4118.50	3.87	−4114.74	−4118.72	3.98
G_{CUBA}	−4196.64	−4197.66	1.02	−4196.87	−4197.73	0.86
CRF18	−6109.29	−6118.91	9.62	−6109.54	6119.01	9.47
CRF19	−14552.57	−14564.21	11.64	−14553.91	−14565.09	11.18
CRF20	−5083.74	−5100.80	17.06	−5083.91	−5101.05	17.14
CRF24	−4159.40	−4160.43	1.03	−4159.81	−4160.70	0.89

Log marginal likelihood (ML) estimates for logistic growth (LG) and exponential growth (EG) demographic models obtained using the path sampling (PS) and stepping stone sampling (SS) methods. The Log Bayes factor (BF) is the difference of the Log ML between of alternative (H1 = LG) and null (H0 = EG) models. Log BFs>1 indicates that model H1 is more strongly supported by the data than model H0.

doi:10.1371/journal.pone.0072448.t002

North America and Europe [40], regions with a widespread circulation of the subtype $B_{PANDEMIC}$ clade. This may explain the massive influx of subtype $B_{PANDEMIC}$ strains and the apparent absence of "non-pandemic" subtype B Caribbean lineages in Cuba.

Similarly to subtype B, there were multiple introductions of subtype C ($n \geq 10$), subtype G ($n \geq 8$) and CRF18_cpx ($n \geq 2$) viruses in Cuba, but only a few of them were able to get established and disseminate. The clades $C_{CU\text{-}I}$, G_{CU} and $CRF18_{CU}$ comprise 69%, 74% and 98% of all subtype C, subtype G and CRF18_cpx sequences from Cuba included in this study, respectively. The monophyletic clustering of CRF19_cpx-like *pol* Cuban sequences within subtype D radiation, the paucity of this genetic variant in Africa, and the recent T_{MRCA} of Cuban sequences strongly suggests that the $CRF19_{CU}$ clade also derives from a single founder event. HIV-1 clades G_{CU}, $CRF18_{CU}$ and $CRF19_{CU}$ probably originate in central Africa, whereas clade $C_{CU\text{-}I}$ probably derives from east Africa. Our study suggests that clades $CRF19_{CU}$ and G_{CU} began to circulate in Cuba around the late 1980s, followed shortly thereafter by clades $CRF18_{CU}$ and $C_{CU\text{-}I}$. Thus, although Cuban personnel were stationed in several African countries since the 1970s, HIV-1 African strains were successfully disseminated within Cuba only from the late 1980s onwards.

Our data suggest that HIV-1 CRFs_BG (20_BG, 23_BG and 24_BG) started to spread in Cuba in the second half of the 1990s. Such a recent expansion of BG recombinants in Cuba is fully consistent with epidemiological data showing that in samples collected in 2003, none of the individuals harboring BG recombinants were diagnosed with HIV-1 infection earlier than 1996, and all but three were diagnosed since 2000 [20]. Similarly, the proportion of BG infections among MSM in Havana City increased from 0% in those diagnosed in 1998 to 31% in those diagnosed in 2003 [3]. It was proposed that all Cuban CRFs_BG evolved from a common BG recombinant ancestor locally generated by recombination between parental clades $B_{CU\text{-}II}$ and G_{CU} [20]. According to our estimations, that common BG recombinant ancestor was generated in the early 1990s, thus around or immediately after the estimated onset date of parental clades $B_{CU\text{-}II}$ and G_{CU} and some years earlier than the emergence of the CRFs_BG.

The reconstruction of the demographic history indicates that most HIV-1 Cuban clades followed a very similar growth pattern characterized by rapid dissemination until the early 2000s after which the epidemic growth rate of those epidemics started to slow-down. The expansion of the $B_{CU\text{-}II}$ clade, by contrast, seems to decrease during 1990s; whereas the growth rate of the CRF24_BG clade probably only stabilized in the second half of the 2000s. The initial expansion of the major HIV-1 Cuban clades coincides with a sustained increase in the number of infected HIV-positive individuals in Cuba from 1991 to 2000 [41]. UNAIDS estimations indicate that the total number of people living with HIV in Cuba continued to growth in the last decade, rising from 3,100 (2,600–4,300) in 2000 to 14,000 (12,000–16,000) in 2011 [36]. Our demographic analysis, however, suggests a trend toward stability in the effective number of infections of all major HIV-1 Cuban clades over time consistent with recent epidemiological data that shows a decrease of HIV incidence in Cuba, mainly among men, in the biennium 2010–2011 [42].

Our coalescent-based analyses suggest that CRF20_BG, CRF24_BG and $B_{CU\text{-}II}$ have displayed a more explosive initial growth (1.0 $year^{-1}$–1.6 year-^{1}) than clades G_{CU}, $CRF18_{CU}$ and $CRF19_{CU}$ (~0.4–0.6 $year^{-1}$); whereas clades $B_{CU\text{-}I}$ and $C_{CU\text{-}I}$ displayed intermediate initial growth rates (~0.8 $year^{-1}$). Notably, all those HIV-1 Cuban clades with the fastest initial expansion rates ($B_{CU\text{-}I}$, $B_{CU\text{-}II}$, $B_{CU\text{-}I}$ and CRFs_BG) were much more prevalent among MSM than among heterosexual (HET) persons [3]. Thus, some HIV-1 Cuban clades may have spread faster than others because they encountered, by chance, local transmission chains with higher rates of partner exchange. Dissemination within a transmission network of small size and high rate of partner exchange may also explain the fast, but self limited, dissemination phase of clade $B_{CU\text{-}II}$. Additional influence of virological factors cannot be excluded. These results must also be interpreted with caution as most growth rate estimates here obtained displayed quite large overlapping 95% HPD intervals.

It has been proposed that Cuba's low rate of HIV infection is due to several factors that served to prevent sexual transmission of the virus, including: wide-scale HIV screening and subsequent contact tracing of HIV-positive individuals, mandatory quarantine of the first HIV-infected individuals at sanatoria, free access to a well structured public health system, comprehensive HIV education campaigns, coordinated work of Cuban government agencies

and community, and restricted tourism between Cuba and western countries up to the early 1990s [43,44,45,46]. The estimated initial growth rates of the major HIV-1 Cuban clades ($\sim$0.4–1.6 year^{-1}), however, were comparable to those obtained for different HIV-1 epidemics in the Americas ($\sim$0.5–1.3 year^{-1}) [38,47,48,49,50,51,52], Europe ($\sim$0.4–1.5 year^{-1}) [52,53,54,55], Africa ($\sim$0.2–0.8 year^{-1}) [47,52,56,57,58] and Asia ($\sim$0.8 year^{-1}) [59]. This suggests that several factors may have contributed to delay the introduction and/or dissemination of HIV-1 in Cuba for many years; but once some HIV-1 strains got established in vulnerable HET and MSM transmission groups they spread quickly.

In summary, this study indicates that only a few subtype B and non-B subtype founder viral strains were successfully disseminated in Cuba. Some of those HIV-1 viral strains were probably introduced from North America/Europe, central Africa and east Africa between the middle 1980s and the middle 1990s; whereas other were locally generated around the late 1990s. Changes in the social and economic landscapes of Cuba occurring at the beginning of the 1990s may have fueled the introduction and/or initial dissemination of major HIV-1 Cuban clades. Although the main HIV-1 Cuban lineages began to circulate at a rather late time of the AIDS pandemic, further dissemination within vulnerable groups was rapid. These results reinforce the importance of maintaining, reviewing and updating permanently the public health measures aimed at controlling the spread of those HIV-1 variants already established in the Cuban population.

Supporting Information

Table S1 HIV-1 subtype B dataset.
(PDF)

Table S2 HIV-1 subtype C dataset.
(PDF)

Table S3 HIV-1 subtype G dataset.
(PDF)

Table S4 HIV-1 CRF18_cpx and CRF19_cpx/subtype D datasets.
(PDF)

Table S5 HIV-1 CRF20/23/24_cpx datasets.
(PDF)

Table S6 Nucleotide substitution models selected using jModeltest program.
(PDF)

Acknowledgments

We wish to thank Dr Vera Bongertz for critical reviewing of the manuscript.

Author Contributions

Conceived and designed the experiments: GB ED. Performed the experiments: ED GB. Analyzed the data: ED GB. Wrote the paper: GB ED.

References

1. Thomson MM, Najera R (2005) Molecular epidemiology of HIV-1 variants in the global AIDS pandemic: an update. AIDS Rev 7: 210–224.
2. Cuevas MT, Ruibal I, Villahermosa ML, Diaz H, Delgado E, et al. (2002) High HIV-1 genetic diversity in Cuba. Aids 16: 1643–1653.
3. Perez L, Thomson MM, Bleda MJ, Aragones C, Gonzalez Z, et al. (2006) HIV Type 1 molecular epidemiology in cuba: high genetic diversity, frequent mosaicism, and recent expansion of BG intersubtype recombinant forms. AIDS Res Hum Retroviruses 22: 724–733.
4. Kouri V, Aleman Y, Perez L, Perez J, Fonseca C, et al. (2012) High frequency of antiviral drug resistance and non-B subtypes in HIV-1 patients failing antiviral therapy in Cuba. J Clin Virol 55: 348–355.
5. Machado LY, Blanco M, Dubed M, Diaz HM, Ruiz NM, et al. (2012) HIV type 1 genetic diversity in newly diagnosed Cuban patients. AIDS Res Hum Retroviruses 28: 956–960.
6. Machado LY, Dubed M, Diaz H, Ruiz N, Romay D, et al. (2013) Transmitted HIV type 1 drug resistance in newly diagnosed Cuban patients. AIDS Res Hum Retroviruses 29: 411–414.
7. Perez L, Kouri V, Aleman Y, Abrahantes Y, Correa C, et al. (2013) Antiretroviral drug resistance in HIV-1 therapy-naive patients in Cuba. Infect Genet Evol 16C: 144–150.
8. Thomson MM, Casado G, Posada D, Sierra M, Najera R (2005) Identification of a novel HIV-1 complex circulating recombinant form (CRF18_cpx) of Central African origin in Cuba. Aids 19: 1155–1163.
9. Casado G, Thomson MM, Sierra M, Najera R (2005) Identification of a novel HIV-1 circulating ADG intersubtype recombinant form (CRF19_cpx) in Cuba. J Acquir Immune Defic Syndr 40: 532–537.
10. Bartolo I, Rocha C, Bartolomeu J, Gama A, Marcelino R, et al. (2009) Highly divergent subtypes and new recombinant forms prevail in the HIV/AIDS epidemic in Angola: New insights into the origins of the AIDS pandemic. Infect Genet Evol 9: 672–682.
11. Afonso JM, Bello G, Guimaraes ML, Sojka M, Morgado MG (2012) HIV-1 genetic diversity and transmitted drug resistance mutations among patients from the North, Central and South regions of Angola. PLoS ONE 7: e42996.
12. Kalish ML, Robbins KE, Pieniazek D, Schaefer A, Nzilambi N, et al. (2004) Recombinant viruses and early global HIV-1 epidemic. Emerg Infect Dis 10: 1227–1234.
13. Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, 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. J Virol 74: 10498–10507.
14. Niama FR, Toure-Kane C, Vidal N, Obengui P, Bikandou B, et al. (2006) HIV-1 subtypes and recombinants in the Republic of Congo. Infect Genet Evol.
15. Pircher M, Diafouka M, Papuchon J, Recordon-Pinson P, Mahambou DN, et al. (2012) Molecular characterization of HIV type 1 in Brazzaville, Republic of Congo, and first data on resistance to antiretroviral drugs. AIDS Res Hum Retroviruses 28: 1798–1802.
16. Marechal V, Jauvin V, Selekon B, Leal J, Pelembi P, et al. (2006) Increasing HIV type 1 polymorphic diversity but no resistance to antiretroviral drugs in untreated patients from Central African Republic: a 2005 study. AIDS Res Hum Retroviruses 22: 1036–1044.
17. Carr JK, Torimiro JN, Wolfe ND, Eitel MN, Kim B, et al. (2001) The AG recombinant IbNG and novel strains of group M HIV-1 are common in Cameroon. Virology 286: 168–181.
18. Vergne L, Bourgeois A, Mpoudi-Ngole E, Mougnutou R, Mbuagbaw J, et al. (2003) Biological and genetic characteristics of HIV infections in Cameroon reveals dual group M and O infections and a correlation between SI-inducing phenotype of the predominant CRF02_AG variant and disease stage. Virology 310: 254–266.
19. Powell R, Barengolts D, Mayr L, Nyambi P (2010) The Evolution of HIV-1 Diversity in Rural Cameroon and its Implications in Vaccine Design and Trials. Viruses 2: 639–654.
20. Sierra M, Thomson MM, Posada D, Perez L, Aragones C, et al. (2007) Identification of 3 phylogenetically related HIV-1 BG intersubtype circulating recombinant forms in Cuba. J Acquir Immune Defic Syndr 45: 151–160.
21. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876–4882.
22. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, et al. (2005) An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 21: 3797–3800.
23. Ray SC (1999) Simplot v2.5.0. Available: http://sray.med.som.jhmi.edu/SCRoftware/simplot/. Accessed 2013 July 16.
24. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25: 1253–1256.
25. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704.
26. Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online–a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33: W557–559.
27. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol 55: 539–552.
28. Rambaut A (2009) FigTree v1.4.0: Tree Figure Drawing Tool. Available: http://tree.bio.ed.ac.uk/software/figtree/. Accessed 2013 July 16.
29. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161: 1307–1320.
30. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.

31. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLoS Biol 4: e88.
32. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22: 1185–1192.
33. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, et al. (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol Biol Evol 29: 2157–2167.
34. Rambaut A, Drummond A (2007) Tracer v1.4. Available: http://tree.bio.ed.ac.uk/software/tracer/. Accessed 2013 July 16.
35. Delatorre EO, Bello G (2012) Phylodynamics of HIV-1 Subtype C Epidemic in East Africa. PLoS ONE 7: e41904.
36. UNAIDS (2013) AIDS INFO DATABASE. Available: http://www.unaids.org/en/dataanalysis/datatools/aidsinfo/. Accessed 2013 July 16.
37. Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, et al. (2007) The emergence of HIV/AIDS in the Americas and beyond. Proc Natl Acad Sci U S A 104: 18566–18570.
38. Robbins KE, Lemey P, Pybus OG, Jaffe HW, Youngpairoj AS, et al. (2003) U.S. Human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. J Virol 77: 6359–6366.
39. Junqueira DM, de Medeiros RM, Matte MC, Araujo LA, Chies JA, et al. (2011) Reviewing the history of HIV-1: spread of subtype B in the Americas. PLoS ONE 6: e27489.
40. Figueras MA (2001) International Tourism and the Formation of Productive Clusters in the Cuban Economy. Latin American Studies Association (LASA) XXII Congress, Washington, DC Available: http://www3.dogus.edu.tr/memek/DersNotlari/TOUR%20229/TOUR%20229-2004%20fall/cuba_contribution.pdf. Accessed 2013 July 16.
41. Hsieh YH, Chen CW, Lee SM, de Arazoza H (2001) On the recent sharp increase in HIV detections in Cuba. AIDS 15: 426–428.
42. UNAIDS (2012) Country progress reports. Available: http://www.unaids.org/en/dataanalysis/knowyourresponse/countryprogressreports/2012countries/ce_CU_Narrative_Report[1]pdf. Accessed 2013 July 16.
43. Perez-Stable EJ (1991) Cuba's response to the HIV epidemic. Am J Public Health 81: 563–567.
44. Hansen H, Groce NE (2001) From quarantine to condoms: shifting policies and problems of HIV control in Cuba. Med Anthropol 19: 259–292.
45. de Arazoza H, Joanes J, Lounes R, Legeai C, Clemencon S, et al. (2007) The HIV/AIDS epidemic in Cuba: description and tentative explanation of its low HIV prevalence. BMC Infect Dis 7: 130.
46. Anderson T (2009) HIV/AIDS in Cuba: lessons and challenges. Rev Panam Salud Publica 26: 78–86.
47. Salemi M, de Oliveira T, Soares MA, Pybus O, Dumans AT, et al. (2005) Different epidemic potentials of the HIV-1B and C subtypes. J Mol Evol 60: 598–605.
48. Bello G, Eyer-Silva WA, Couto-Fernandez JC, Guimaraes ML, Chequer-Fernandez SL, et al. (2007) Demographic history of HIV-1 subtypes B and F in Brazil. Infect Genet Evol 7: 263–270.
49. Bello G, Guimaraes ML, Passaes CP, Matos Almeida SE, Veloso VG, et al. (2009) Short communication: Evidences of recent decline in the expansion rate of the HIV type 1 subtype C and CRF31_BC epidemics in southern Brazil. AIDS Res Hum Retroviruses 25: 1065–1069.
50. Bello G, Aulicino PC, Ruchansky D, Guimaraes ML, Lopez-Galindez C, et al. (2010) Phylodynamics of HIV-1 circulating recombinant forms 12_BF and 38_BF in Argentina and Uruguay. Retrovirology 7: 22.
51. Ristic N, Zukurov J, Alkmim W, Diaz RS, Janini LM, et al. (2011) Analysis of the origin and evolutionary history of HIV-1 CRF28_BF and CRF29_BF reveals a decreasing prevalence in the AIDS epidemic of Brazil. PLoS ONE 6: e17485.
52. Bello G, Afonso JM, Morgado MG (2012) Phylodynamics of HIV-1 subtype F1 in Angola, Brazil and Romania. Infect Genet Evol 12: 1079–1086.
53. Hue S, Pillay D, Clewley JP, Pybus OG (2005) Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. Proc Natl Acad Sci U S A 102: 4425–4429.
54. Salemi M, de Oliveira T, Ciccozzi M, Rezza G, Goodenow MM (2008) High-Resolution Molecular Epidemiology and Evolutionary History of HIV-1 Subtypes in Albania. PLoS ONE 3: e1390.
55. Zehender G, Ebranati E, Lai A, Santoro MM, Alteri C, et al. (2010) Population dynamics of HIV-1 subtype B in a cohort of men-having-sex-with-men in Rome, Italy. J Acquir Immune Defic Syndr 55: 156–160.
56. Walker PR, Pybus OG, Rambaut A, Holmes EC (2005) Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic growth. Infect Genet Evol 5: 199–208.
57. Tully DC, Wood C (2010) Chronology and evolution of the HIV-1 subtype C epidemic in Ethiopia. Aids 24: 1577–1582.
58. Delatorre E, Bello G (2013) Spatiotemporal dynamics of the HIV-1 CRF06_cpx epidemic in western Africa. Aids.
59. Chen JH, Wong KH, Chan KC, To SW, Chen Z, et al. (2011) Phylodynamics of HIV-1 subtype B among the men-having-sex-with-men (MSM) population in Hong Kong. PLoS ONE 6: e25286.

8.5 APÊNDICE 5

Spatiotemporal Dynamics of the HIV-1 Subtype G epidemic in West and Central Africa

Autores: **Edson Delatorre** e Gonzalo Bello

Jornal: PLOS ONE, 2014

OPEN ACCESS Freely available online

PLOS ONE

Spatiotemporal Dynamics of the HIV-1 Subtype G Epidemic in West and Central Africa

Edson Delatorre, Daiana Mir, Gonzalo Bello*

Laboratório de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil

Abstract

The human immunodeficiency virus type 1 (HIV-1) subtype G is the second most prevalent HIV-1 clade in West Africa, accounting for nearly 30% of infections in the region. There is no information about the spatiotemporal dynamics of dissemination of this HIV-1 clade in Africa. To this end, we analyzed a total of 305 HIV-1 subtype G *pol* sequences isolated from 11 different countries from West and Central Africa over a period of 20 years (1992 to 2011). Evolutionary, phylogeographic and demographic parameters were jointly estimated from sequence data using a Bayesian coalescent-based method. Our analyses indicate that subtype G most probably emerged in Central Africa in 1968 (1956–1976). From Central Africa, the virus was disseminated to West and West Central Africa at multiple times from the middle 1970s onwards. Two subtype G strains probably introduced into Nigeria and Togo between the middle and the late 1970s were disseminated locally and to neighboring countries, leading to the origin of two major western African clades (G_{WA-I} and G_{WA-II}). Subtype G clades circulating in western and central African regions displayed an initial phase of exponential growth followed by a decline in growth rate since the early/middle 1990s; but the mean epidemic growth rate of G_{WA-I} (0.75 $year^{-1}$) and G_{WA-II} (0.95 $year^{-1}$) clades was about two times higher than that estimated for central African lineages (0.47 $year^{-1}$). Notably, the overall evolutionary and demographic history of G_{WA-I} and G_{WA-II} clades was very similar to that estimated for the CRF06_cpx clade circulating in the same region. These results support the notion that the spatiotemporal dissemination dynamics of major HIV-1 clades circulating in western Africa have probably been shaped by the same ecological factors.

Citation: Delatorre E, Mir D, Bello G (2011) Spatiotemporal Dynamics of the HIV-1 Subtype G Epidemic in West and Central Africa. PLoS ONE 9(6): e98908. doi:10.1371/journal.pone.0098908

Editor: Dimitrios Paraskevis, University of Athens, Medical School, Greece

Received March 26, 2014; **Accepted** May 3, 2014; **Published** June 11, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All sequences used in this study were retrieved from Los Alamos HIV Database (http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html). Final alignments and a full list of GenBank accession numbers are available in the Supporting Information files.

Funding: This work was supported by Public Health Service grants E-26/111.758/2012 from the FAPERJ and 472896/2012-1 from the CNPq. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gbellobr@gmail.com

Introduction

The current distribution of human immunodeficiency virus type 1 (HIV-1) group M subtypes and circulating recombinant forms (CRFs) around the world resulted from the chance exportation of different viral strains out of Central Africa into new geographic regions were these initiated secondary epidemics [1]. A recent study suggests that spatial accessibility (human migrations and movements through transportation link availability and quality) has played a significant role in HIV-1 spread across sub-Saharan Africa and may explain the heterogeneous distribution of HIV-1 subtypes and CRFs in the different African regions [2].

West Africa is one of the most strongly connected regions in the continent [2] and also appears as an area of intense intra-regional migration [3]. This coincides with an overall dominance of the CRF02_AG variant, that accounts for about 50% of all HIV-1 infections in West Africa [4]. A closer inspection of the HIV-1 molecular epidemiological profile in this African region, however, reveals an important intra-regional heterogeneity in the distribution of other viral clades, including subtype G and CRF06_cpx. Subtype G is the second most prevalent HIV-1 clade in West Africa accounting for nearly 30% of infections in the region [4]. Its prevalence greatly varies within and between countries, comprising 30–50% of HIV-1 infections across different regions from Nigeria [5,6,7,8,9,10,11,12], 5–15% in Benin, Niger and Togo [13,14,15,16,17], and ≤4% in other western African countries [14,18,19,20,21,22,23,24,25,26,27,28,29,30]. Similarly, the occurrence of the CRF06_cpx clade ranges from 40–50% of HIV-1 infections in Burkina Faso [18,19,20], to 5–15% in Benin, Ghana, Mali, Niger, Nigeria, Senegal and Togo [5,6,7,8,9,10,11,12,13,14,15,16,17,21,22,23,24,28,29], and <3% in other western African countries [14,26,27,30].

The highly heterogeneous distribution of subtype G and CRF06_cpx across the well-connected western African countries, suggests that spatial accessibility is not enough to fully explain the spatial distribution of those HIV-1 clades in this African region. A recent study conducted by our group suggests that Burkina Faso was the most important epicenter of dissemination of the HIV-1 CRF06_cpx strain at regional level and that CRF06_cpx prevalence decreases exponentially as we move away from the epicenter [31]. Our study also estimated that the CRF06_cpx clade started to spread in West Africa around the late 1970s [31], almost 10 years later than the estimated origin of the CRF02_AG clade in West Central Africa [32]. We postulated that the relatively late introduction of the CRF06_cpx clade into western Africa combined with the stabilization of the HIV epidemic in several

countries from the region since the early/middle 1990s may have resulted in a more limited dissemination away from the epicenter and a more heterogeneous regional distribution of CRF06_cpx when compared with CRF02_AG.

It is unclear whether this hypothesis could also explain the complex distribution of subtype G in West Africa. The objective of this study was to reconstruct the onset date, dissemination routes and demographic history of the HIV-1 subtype G clade in the African continent. To this end, we used a Bayesian coalescent-based framework to analyze 305 HIV-1 subtype G *pol* sequences isolated from 11 different countries from West (Benin, Ghana, Nigeria, Senegal and Togo), West Central (Cameroon, Equatorial Guinea and Gabon), and Central Africa (Angola, Democratic Republic of Congo and Republic of Congo) over a period of 20 years (1992 to 2011).

Materials and Methods

Sequence Dataset

All HIV-1 subtype G *pol* sequences from West and Central African countries that covered the entire protease and partial reverse transcriptase (PR/RT) regions (nt 2253–3272 relative to HXB2 clone) and for which the sampling year was known, were downloaded from the Los Alamos HIV Sequence Database (www.hiv.lanl.gov) by August 2013. The subtype assignment of all sequences was confirmed by: REGA HIV subtyping tool v.2 [33], Maximum Likelihood (ML) phylogenetic analysis, and bootscanning analysis. A ML phylogeny with HIV-1 group M subtype reference sequences was constructed with the PhyML 3.0 program [34] using an online web server [35]. The ML tree was inferred under the GTR+I+G nucleotide substitution model recommended by the jModeltest program [36]. The heuristic tree search was performed using the SPR branch-swapping algorithm and branch support was calculated with the approximate likelihood-ratio (aLRT) SH-like test [37]. In bootscanning analyses, supporting branching of query sequences with HIV-1 group M subtypes reference sequences was determined in Neighbor-Joining trees constructed with the Kimura two-parameter model, within a 250 bp window moving in steps of 10 bases, using Simplot software v.3.5.1 [38]. We detected that 4.7% of the subtype G *pol* sequences available in database had incorrect subtype classification, consistent with previous estimations [39]. Sequences with incorrect classification, multiple sequences from the same individual and sequences from countries poorly represented ($n<4$ sequences) were removed, resulting in a final data set of 305 HIV-1 subtype G *pol* African sequences (Table 1). All codon positions known to be associated with major antiretroviral drug resistance were maintained in the final alignment because ML trees constructed on alignments with or without such positions resulted in the same overall topology (data not shown). Final sequence alignment is available from the authors upon request.

Analysis of Spatiotemporal Dispersion Pattern and Demographic History

The evolutionary rate (μ, nucleotide substitutions per site per year, subst./site/year), the age of the most recent common ancestor (T_{MRCA}, years), the ancestral geographic movements, and the mode and rate (r, years-1) of population growth of HIV-1 subtype G clades circulating in Africa were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST v1.8 [40,41] with BEAGLE to improve run-time [42]. Analyses were performed under a GTR+I+G nucleotide substitution model. The temporal scale of evolutionary process was estimated from the sampling dates of the sequences using a relaxed uncorrelated lognormal molecular clock model and a uniform prior on clock rate ($1.0–4.0\times10^{-3}$ subst/site/year) [43]. Migration events throughout the phylogenetic history were inferred using a reversible discrete Bayesian phylogeographic model [44], in which all possible reversible exchange rates between locations were equally likely, and a CTMC rate reference prior [45]. To quantify the dissemination process, we estimated the number of viral migrations among locations using 'Markov Jump' counts [46] of location-state transitions along the posterior tree distribution as previously described [47,48]. Changes in effective population size through time were initially estimated using a flexible Bayesian Skyline coalescent model [49] that does not require strong prior assumptions of demographic history. Estimates of the population growth rate were subsequently obtained using the parametric model (logistic, exponential or expansion) that provided the best fit to the demographic signal contained in datasets. Comparison between demographic models was performed using the log marginal likelihood (ML) estimation based on path sampling (PS) and stepping-stone sampling (SS) methods [50]. MCMC chains were run for $50–500\times10^{6}$ generations. Adequate chain mixing and uncertainty in parameter estimates were assessed by calculating the Effective Sample Size (ESS) and the 95% Highest Probability Density (HPD) values, respectively, using the TRACER v1.6 program [51]. Maximum clade credibility (MCC) trees were summarized from the posterior distribution of trees with TreeAnnotator and visualized with FigTree v1.4.0 [52]. Migratory events across time were summarized using the cross-platform SPREAD application [53].

Results

Origin of the HIV-1 Subtype G and Identification of Major African Clades

We analyzed 305 HIV-1 subtype G *pol* sequences isolated from 11 African countries between 1992 and 2011 that were sampled across seven different location states (Table 1). Neighboring countries from West (Togo/Ghana), West Central (Gabon/Equatorial Guinea) and Central (Angola/Democratic Republic of Congo/Republic of Congo) Africa comprising few samples ($n<15$) were grouped into the same location (Table 1). According to the Bayesian MCMC analysis, the median evolutionary rate of the HIV-1 subtype G lineage at *pol* gene was estimated at 2.3×10^{-3} (95% HPD: $1.8\times10^{-3}-2.8\times10^{-3}$) subst./site/year. The estimated coefficient of rate variation in our dataset was 0.28 (95% HPD: 0.24–0.32), thus supporting a significant variation of substitution rate among branches and the use of a relaxed molecular clock model. The most probable root location of the subtype G clade was placed in Central Africa (posterior state probability, $PSP=0.88$), and the onset date of this clade was estimated to be 1968 (95% HPD: 1956–1976) (Fig. 1).

The Bayesian MCC (Fig. 1) and ML (Fig. S1) trees point to a clear phylogeographic subdivision of subtype G strains from West and Central Africa. Sequences from western Africa branched mostly in two large monophyletic clades (G_{WA-I} and G_{WA-II}) that were nested among the most basal clades from Central and West Central Africa (G_{CA}). Distribution of HIV-1 subtype G clades greatly varies across countries within each region (Fig. 2). The G_{WA-I} clade was the predominant subtype G lineage detected in Nigeria (80%) and the G_{WA-II} clade predominates in Togo/Ghana (86%). The subtype G epidemic in Benin is dominated by both G_{WA-I} (47%) and G_{WA-II} (40%) clades, whereas G_{WA-I} (50%) and G_{CA} (42%) clades prevail among subtype G infections in Senegal. Basal G_{CA} clades predominate in countries from both central (100%) and west central (50–71%) regions.

Table 1. HIV-1 subtype G *pol* dataset.

Region	Country	Location[a]	*N*	Sampling interval
West Africa	Benin	BJ	15	2004–2009
	Ghana	TG/GH	8	2002–2009
	Nigeria	NG	183	1992–2010
	Senegal	SN	12	1998–2010
	Togo	TG/GH	13	2006–2008
West Central Africa	Cameroon	CM	31	1997–2011
	Gabon	GA/GQ	6	2000–2008
	Equatorial Guinea	GA/GQ	4	2005–2009
Central Africa	Angola	AO/CD/CG	13	1997–2010
	DRC	AO/CD/CG	12	1993–2007
	Republic of Congo	AO/CD/CG	8	2003

[a]Location assigned in the Bayesian phylogeographic analysis. DRC: Democratic Republic of Congo.
doi:10.1371/journal.pone.0098908.t001

Spatiotemporal Dispersal Pattern of the HIV-1 Subtype G African's Epidemic

Reconstruction of viral migrations across time revealed the occurrence of multiple introductions of HIV-1 subtype G strains from Central into West Africa since the middle 1970s (Fig. 3). The earliest viral migrations led to the origin of the G_{WA-I} and G_{WA-II} lineages. The G_{WA-I} clade most probably emerged in Nigeria (*PSP* = 1) around 1974 (95% HPD: 1966–1981) and from this country was later disseminated to Benin, Cameroon, Equatorial Guinea, Ghana, and Senegal. The G_{WA-II} clade most probably emerged in Togo/Ghana (*PSP* = 0.68) around 1979 (95% HPD: 1973–1984) and was disseminated to Nigeria in 1981 (95% HPD: 1976–1986), where it further spread locally. In the following years, the G_{WA-II} clade was disseminated from both Togo/Ghana and Nigeria to Benin, Cameroon, Gabon, and Senegal. Our phylogeographic analysis also detected several independent introductions of subtype G variants from Central Africa into Cameroon (Figs 1 and 3). The earliest introductions occurred between the late 1970s and the middle 1980s and gave rise to at least three local Cameroonian clades; one of which was further disseminated to Gabon, Equatorial Guinea, Senegal and Angola.

We next quantified the viral flux between locations using Markov jump counts (Fig. 4 and Table S1). Nigeria (16.4), central African countries (14.8), and Togo/Ghana (8.3) displayed positive net viral migration rates (efflux minus influx), whereas Benin (−14.2), Cameroon (−10.1), Gabon/Equatorial Guinea (−8.4), and Senegal (−6.8) displayed negative net viral migration fluxes. The highest numbers of viral transitions were from Nigeria to Benin (8.1), Togo/Ghana (5.5) and Cameroon (4.4), from Central Africa to Cameroon (6.1) and Senegal (4.2), and from Togo/Ghana to Benin (5.7), Nigeria (4.1) and Cameroon (3.9). The estimated viral flux to Gabon/Equatorial Guinea from Cameroon (2.6), Central Africa (2.3), Nigeria (2.3) and Togo/Ghana (2.3) was very similar.

Demographic History of HIV-1 Subtype G African's Epidemic

Estimations of effective population size (*Ne*) changes over time were initially obtained using a Bayesian skyline plot (BSP) coalescent model. The BSP analysis of the complete dataset suggests that the subtype G African epidemic experienced a fast exponential growth during the 1970s and 1980s, followed by a more recent stabilization since the early 1990s (Fig. 5A). This overall growth pattern, however, represents the combined population dynamics of the different African subtype G clades that are being disseminated within different countries and regions. In order to better understand the regional differences in the demographic histories of HIV-1 subtype G African epidemics, the G_{CA}, G_{WA-I} and G_{WA-II} clades were analyzed separately (Table S2).

The BSP analyses suggest that all African subtype G clades displayed a similar population growth pattern characterized by an initial phase of exponential growth followed by a decline in growth rate since the early/middle 1990s (Figs. 5B, D and F). To estimate the mean epidemic growth rate of the major subtype G African clades, log ML for the logistic, exponential and expansion growth models were calculated using both PS and SS methods. The best-fit demographic model for all subtype G clades was the logistic one (log BF>5) (Table S3) that was then used to estimate the initial epidemic growth rate. The overall time-scale and demographic pattern obtained from both BSP (Figs. 5B, D and F) and logistic growth coalescent tree priors (Figs. 5C, E and G) were very similar and important differences in the epidemic growth rate were detected across subtype G clades from West and Central Africa. According to the logistic growth coalescent model, the mean growth rate of clades G_{WA-I} (0.75 $year^{-1}$) and G_{WA-II} (0.95 $year^{-1}$) was about two times higher than that estimated for the clade G_{CA} (0.47 $year^{-1}$) (Fig. 5).

Discussion

This study indicates that the HIV-1 subtype G likely originated in Central Africa around the late 1960s. The root position of the subtype G clade is fully consistent with the most accepted model that traces the origin of all HIV-1 group M subtypes to the DRC [54,55,56,57] and is also resistant to the problem of sampling bias because sequences from Central Africa represent a minor fraction (9.2%) of the total subtype G sequences included in our study. The T_{MRCA} of subtype G clade here estimated (1968: 1956–1976) is also fully consistent to that previously estimated for this subtype (1970: 1960–1978) [58]. This onset date is comparable to that estimated for subtype F (1967: 1956–1976) [59]; but more recent than that of subtypes A1 (1954: 1940–1968), C (1955: 1934–1972), and D (1947: 1938–1955) [58].

After emerging in Central Africa around the late 1960s, the HIV-1 subtype G was disseminated to West and West Central

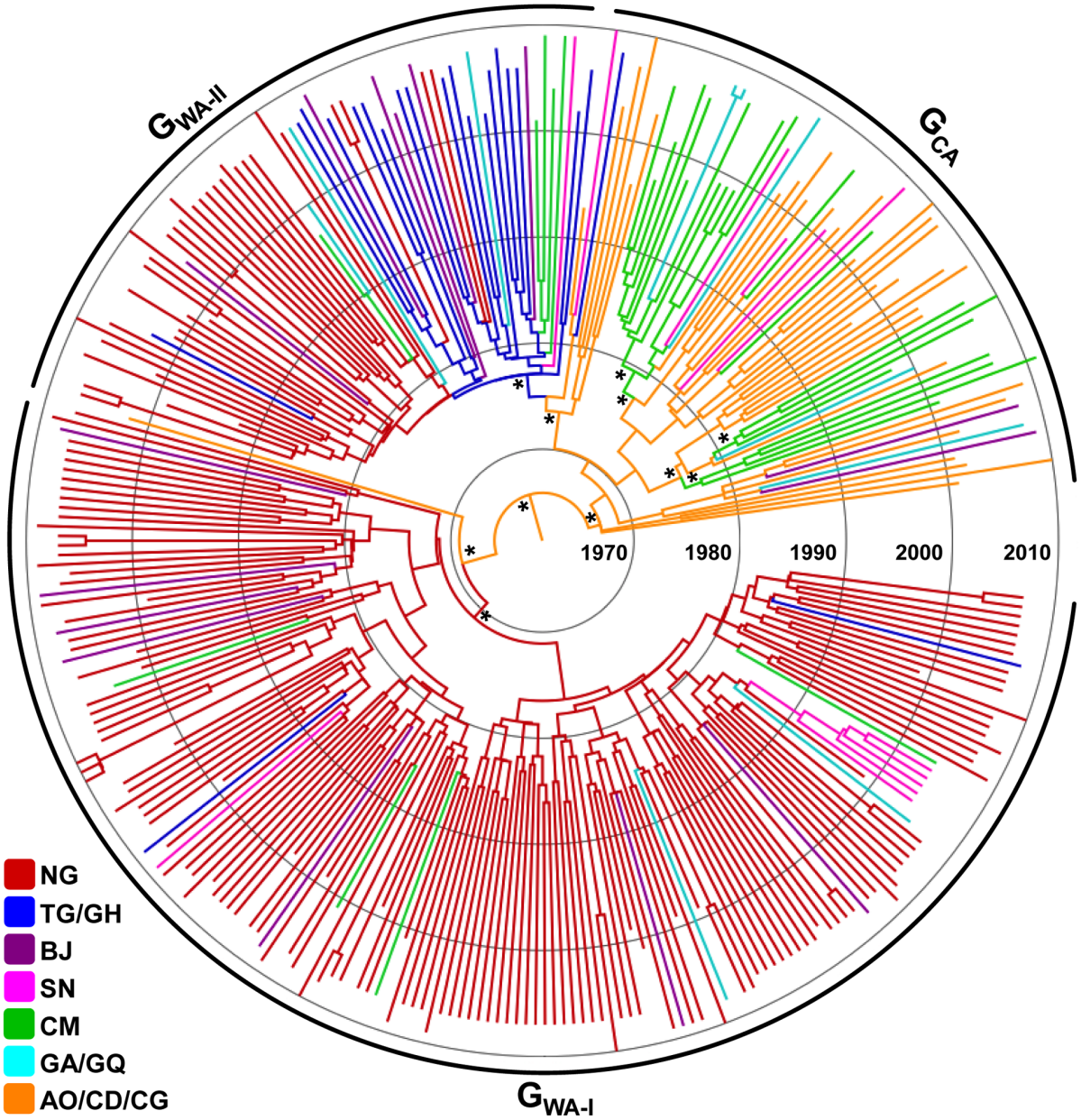


Figure 1. Time-scaled Bayesian MCC tree of the HIV-1 subtype G *pol* PR/RT sequences (~1,000 nt) circulating in West and Central Africa. Branches are colored according to the most probable location state of their descendent nodes as indicated at the legend (bottom left). Arcs indicate the positions of major subtype G clades characteristic of western (G_{WA-I} and G_{WA-II}) and central (G_{CA}) African regions. Asterisks point to key nodes with high posterior state probability support (*PSP*>0.85). Branch lengths are drawn to scale of years. The tree was automatically rooted under the assumption of a relaxed molecular clock.
doi:10.1371/journal.pone.0098908.g001

Africa a few years later (1975–1980). Our phylogeographic analysis supports the occurrence of multiple introductions of HIV-1 subtype G strains from central into the western and west central African regions. Some of the viral strains disseminated during the 1970s fueled secondary outbreaks that led to the origin of specific subtype G clades. The major subtype G clades detected in our study were the G_{WA-I} that most probably emerged in Nigeria around the middle 1970s, and the G_{WA-II} that most probably emerged in Togo or Ghana around the late 1970s. Although we grouped sequences from Togo and Ghana into one single location, the much higher prevalence of subtype G in Togo (9%) [16,17] compared with Ghana (<1%) [23,24] suggests that the G_{WA-II} clade probably arose in Togo. We also detected three minor subtype G clades that resulted of independent introductions of viral strains from central Africa into Cameroon between the late 1970s and the middle 1980s.

Nigeria and Togo/Nigeria were inferred as the most important epicenters of dissemination of the G_{WA-I} and G_{WA-II} clades at regional level, respectively. The G_{WA-I} clade, which corresponds to the clade previously designated G' [5,8], was the predominant subtype G lineage in Nigeria (80%), Senegal (50%), and Benin (47%), and also comprises a significant fraction of subtype G infections in Gabon/Equatorial Guinea (20%), Cameroon (13%) and Togo/Ghana (9%). The G_{WA-II} clade predominates in Togo/

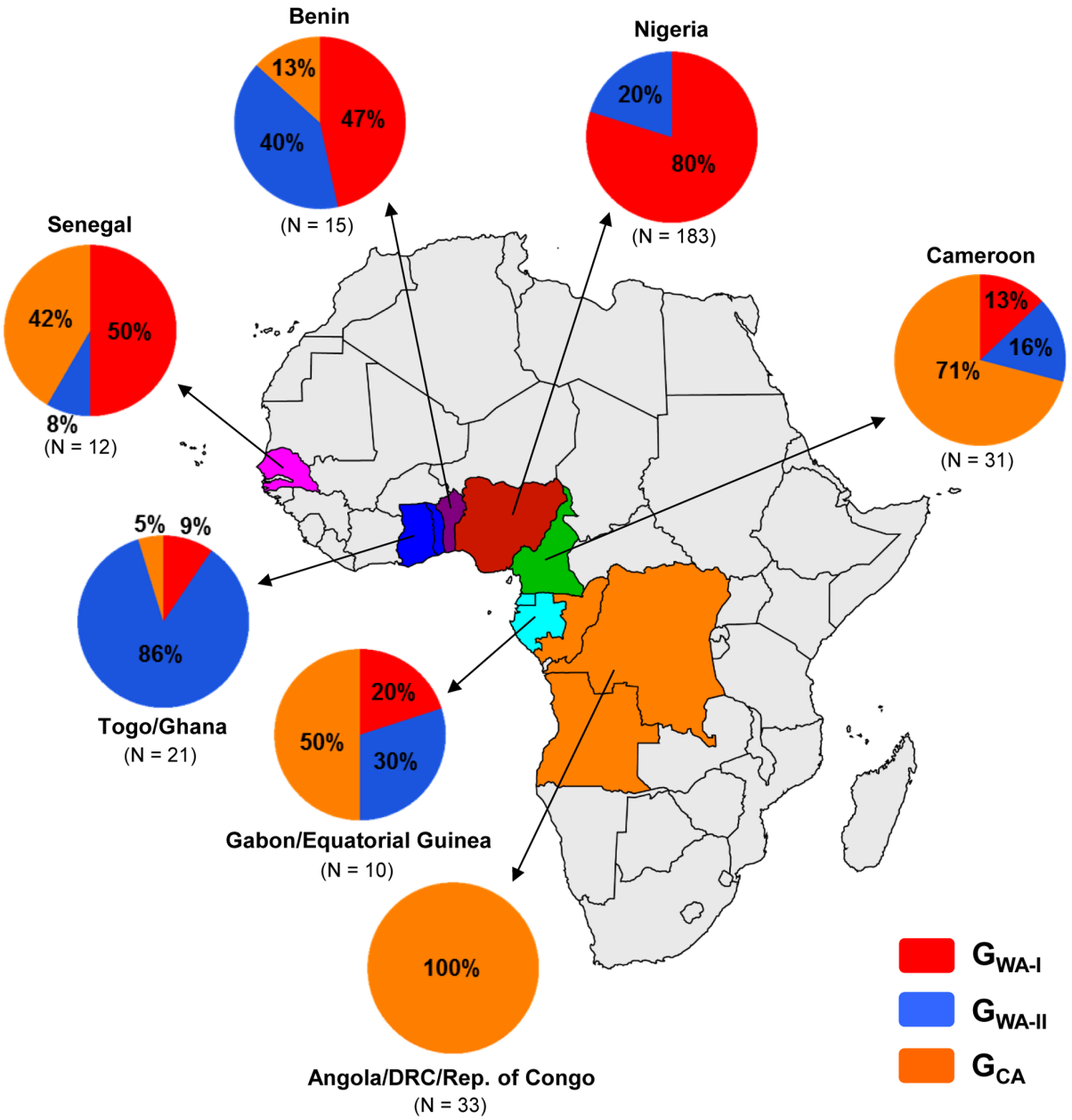


Figure 2. Prevalence of G_{WA-I}, G_{WA-II} and G_{CA} clades among subtype G infected individuals from different African countries, estimated from phylogenetic analyses presented in Figs. 1 and S1. The total number of subtype G sequences analyzed in each locality is indicated. Each clade is represented by a color as indicated at the legend.
doi:10.1371/journal.pone.0098908.g002

Ghana (86%) and is responsible for a significant fraction of subtype G infections in Benin (40%), Gabon/Equatorial Guinea (30%), Nigeria (20%), Cameroon (16%) and Senegal (8%). The subtype G clades introduced into Cameroon were mainly disseminated to the neighboring countries in the central west region (Gabon and Equatorial Guinea), although a few disseminations to Senegal were also detected. These results indicate that founder subtype G strains introduced into Nigeria and Togo have been much more efficiently disseminated at regional level than those introduced into Cameroon.

Our demographic reconstructions also revealed another important difference between African subtype G clades mainly disseminated in the western region (G_{WA-I} and G_{WA-II}) and those mainly disseminated in the west central and central regions (G_{CA}). Although all African subtype G clades displayed a similar population growth pattern characterized by an initial phase of exponential growth followed by a decline in growth rate since the early/middle 1990s; the mean epidemic growth rate of G_{WA-I} (0.75 $year^{-1}$) and G_{WA-II} (0.95 $year^{-1}$) clades was about two times higher than that estimated for G_{CA} (0.47 $year^{-1}$) clades. This suggests that subtype G clades introduced into Nigeria and Togo during the 1970s probably encountered more favorable conditions for local and regional expansion than those disseminated within central and west-central African countries around the same time. The median growth rates of the G_{WA-I} and G_{WA-II} clades were comparable to that estimated for the CRF06_cpx in western

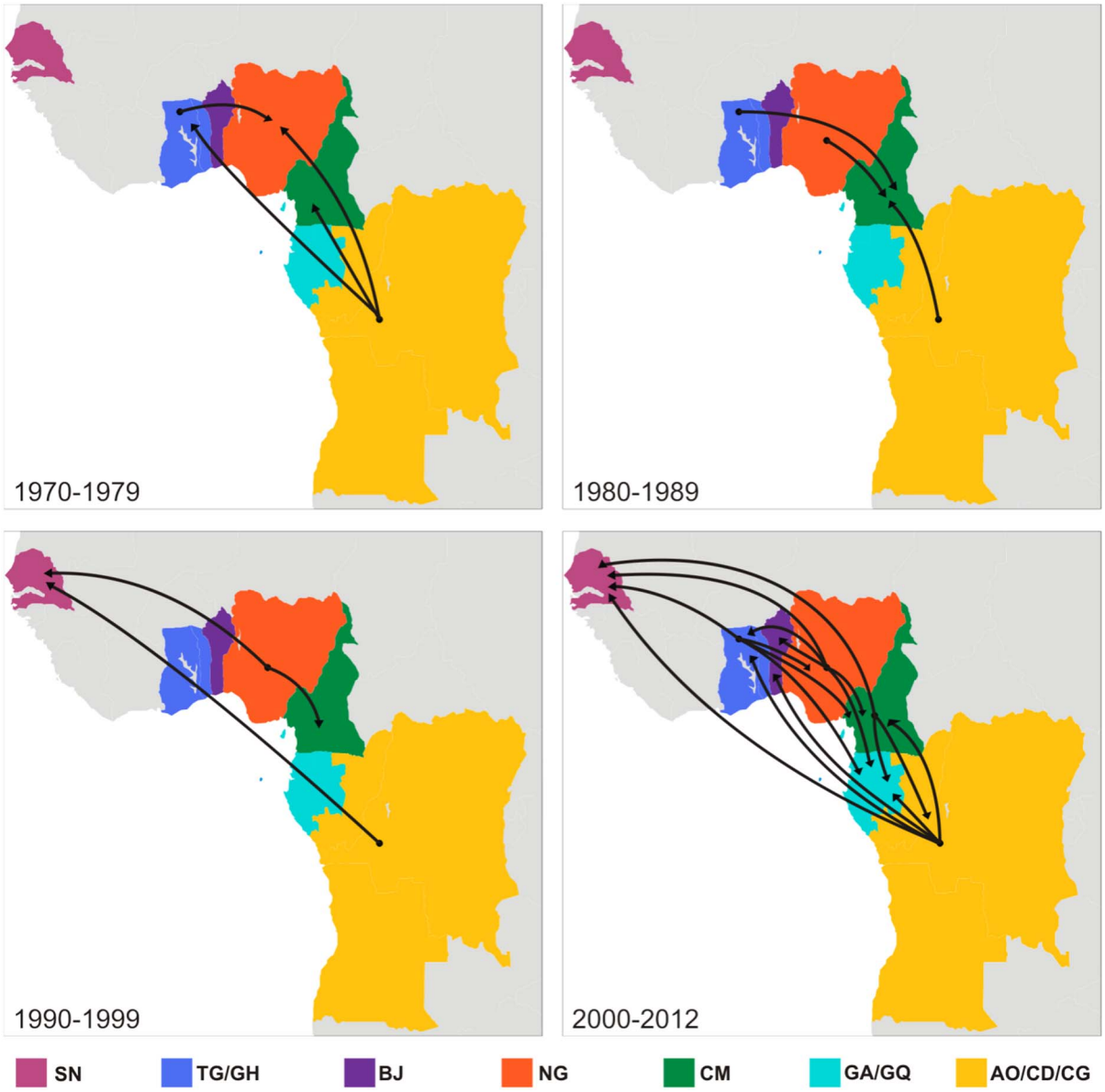


Figure 3. Spatiotemporal dynamics of HIV-1 subtype G clade dissemination in West and Central Africa. Snapshots of viral migration events occurring at different time intervals between 1970 and 2012 are shown. Lines between locations represent branches in the Bayesian MCC tree along which location transitions occur. Each location is represented by a color as indicated at the legend. SN: Senegal; TG/GH: Togo/Ghana; BJ: Benin; NG: Nigeria; CM: Cameroon; GA/GQ: Gabon/Equatorial Guinea; AO/CD/CG: Angola/DRC/Republic of Congo.
doi:10.1371/journal.pone.0098908.g003

Africa (0.82 year^{-1}) [31]; whereas the median growth rate of the G_{CA} clades was roughly similar to that estimated for subtype G in Cuba (0.54 year^{-1}) [60] and higher than that estimated for HIV-1 group M in Democratic Republic of Congo (0.17 year^{-1}) [61].

The faster epidemic growth and the broader geographic dissemination of subtype G strains introduced into West Africa compared with those circulating in the central west and central African regions could be associated to clade-specific or regional-specific differences in viral transmissibility. It has been suggested that accessibility between locations have played a major role in the spatial spread of HIV-1 in sub-Saharan Africa [2]. Notably, West Africa is one of the most strongly connected regions in the continent [2] and also displays an intra-regional migration rate (3%) above the African average (2%) [3]. Others factors including urbanization [56,62], iatrogenic interventions [63,64], and forced migration [62,65] might have also played a role in the emergence and spread of HIV in Africa. Such alternative scenarios can now be tested in a Bayesian framework [66] to find the hypothesis that best explain the variability in the rate of HIV spread across African regions.

Despite the strong regional accessibility, the prevalence of subtype G and CRF06_cpx clades greatly vary across western African countries. The clades CRF06_cpx, G_{WA-I}, and G_{WA-II} seem to have experienced very similar dissemination dynamics; although their origin was traced to different western African countries (Burkina Faso, Nigeria and Togo, respectively) [31]. The three HIV-1 clades probably started to spread in West Africa around the same time (1975–1980), expanded during the 1980s with similar epidemic growth rates (0.75–0.95 year^{-1}), started to stabilize around the early/middle 1990s, and their prevalence is

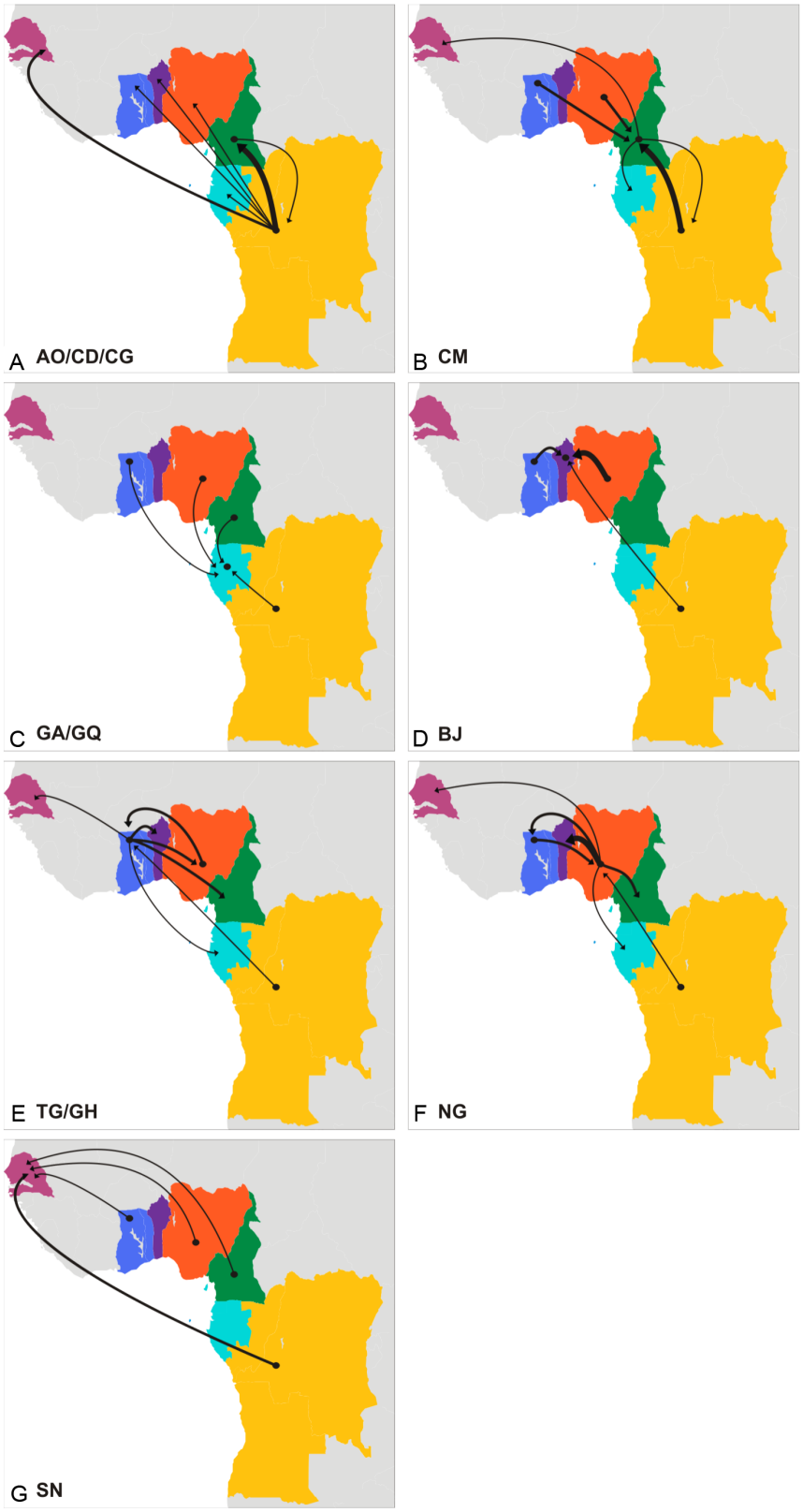


Figure 4. Viral migration rates among locations as measured using 'Markov jump' counts. Each panel represents the estimated viral exchanges from and to Angola/DRC/Republic of Congo (A), Cameroon (B), Gabon/Equatorial Guinea (C), Benin (D), Togo/Ghana (E), Nigeria (F), and Senegal (G). The width of the arrows is proportional to the corresponding mean estimated number of viral transitions between locations according to the following scale: thin arrows = 1.0–2.9 transitions, medium arrows = 3.0–5.9 transitions, thick arrows = 6.0–8.9 transitions. No arrows were displayed when the mean estimated number of transitions was below one.
doi:10.1371/journal.pone.0098908.g004

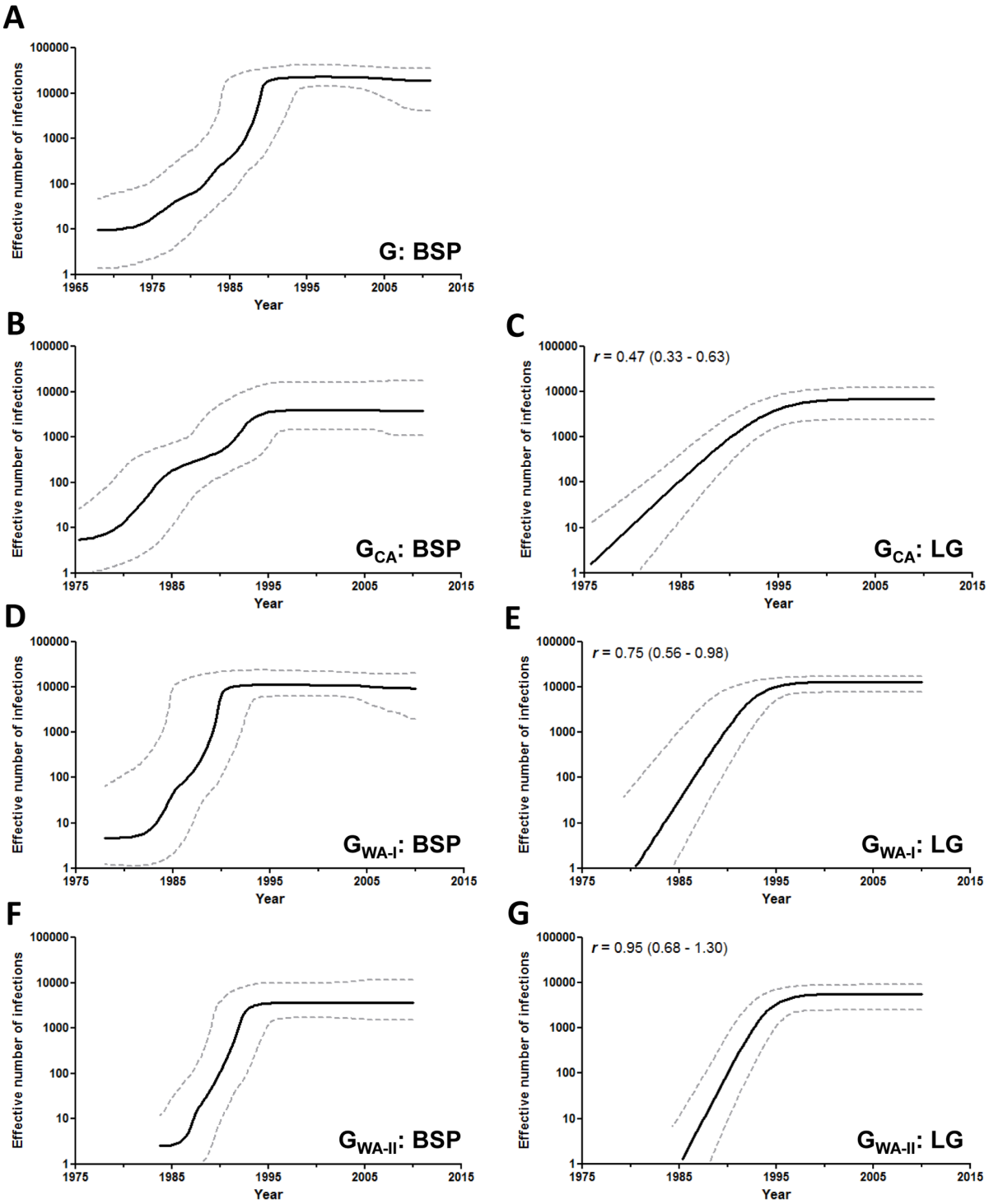


Figure 5. Demographic history of the HIV-1 subtype G and the clades G_{CA}, G_{WA-I} and G_{WA-II} circulating in Central and West Africa. Effective number of infections (y-axis; log10 scale) through time (x-axis; calendar years) estimated using Bayesian skyline (A, B, D, F) and logistic (C, E, G) growth coalescent model. Median estimates of the effective number of infections (solid line) and 95% HPD intervals of the estimates (dashed lines) are shown in each graphic. The median growth rate (with the corresponding 95% credibility interval in parenthesis) of each clade estimated under the logistic growth model is indicated in the upper left corner.
doi:10.1371/journal.pone.0098908.g005

greatly reduced as we moved away from the corresponding epicenters [31]. The relatively late spread of subtype G and CRF06_cpx clades in West Africa combined with: 1) stabilization of the HIV epidemic in several western African countries since the early/middle 1990s, and/or 2) depletion of the susceptible populations most at risk by the firstly introduced CRF02_AG lineage, may have limited the dissemination of these viral clades

far from the epicenter, thus generating a heterogeneous spatial distribution.

The most important limitation of our study was the small sampling size of many African countries. Only Nigeria ($n = 183$) and Cameroon ($n = 31$) were represented by a high or relatively high number of sequences. Other western (Benin, Niger, and Togo) and central (Central African Republic, Chad, Equatorial Guinea, and Gabon) African countries with circulation of subtype G at significant levels (≥5% of all HIV-1 infections) [13,14,15,16,17,67,68,69,70,71,72] were represented by a small number of sequences ($n \leq 15$) that may not fully reflect the country's subtype G diversity, or were not represented at all in our study (Fig. S2). Thus, a more comprehensive and balanced sampling from countries poorly or not represented here would certainly provide more precise estimates of the relative prevalence and migration routes of clades G_{WA-I}, G_{WA-II} and G_{CA} across different African regions, and may also result in the identification of new regional viral clades not detected in this study.

It will be also interesting to trace the origins and global dispersal pathways of those subtype G lineages found in countries outside sub-Saharan Africa, particularly in Cuba [73,74,75], Portugal [76,77,78], and Russia [79] where this subtype has been disseminated among local populations. It has been showed that the spread of HIV-2 outwards Africa mirrors socio historical ties [80] and a previous study conducted by our group showed that most subtype G Cuban lineages are nested among basal sequences from Central Africa [60]. Thus, circulation of subtype G outside sub-Saharan Africa may be linked to the presence of Portuguese, Cuban, and Russian personnel in Angola and neighboring countries during 1960–1990.

In summary, this study suggests that the HIV-1 subtype G clade started to circulate in Central Africa around the late 1960s and was disseminated to West and West Central Africa from the middle 1970s onwards. Nigeria and Togo were pointed out as the major secondary hubs of dissemination of subtype G within western and west central African regions. Our data also highlight that the spatiotemporal dissemination dynamics of western African subtype G clades were very similar to that estimated for the CRF06_cpx epidemic; supporting the notion that current distribution of major HIV-1 clades in West Africa may have been shaped by the same ecological factors. Despite some study limitations, these findings offer important insights toward an understanding of the current characteristics and dynamics of the HIV-1 epidemic in West and West Central Africa.

Supporting Information

Figure S1 ML tree of the of the HIV-1 subtype G *pol* PR/RT sequences (~1,000 nt) circulating in West and Central Africa. Branches are colored according to the geographic origin of each sequence as indicated at the legend (bottom left). Arcs indicate the positions of major subtype G clades characteristic of western (G_{WA-I} and G_{WA-II}) and central (G_{CA}) African regions. Asterisks point to key nodes with high support (aLRT>0.85). The tree was rooted on midpoint. The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.
(PDF)

Figure S2 African map showing the prevalence of subtype G among HIV-1-infected individuals from West and West Central Africa, and the corresponding representativeness of each African country in our subtype G dataset. Countries were colored according to the relative prevalence of subtype G (estimated from references 5–30 and 53–58) as shown in the legend. Asterisks indicate countries represented by very high (***n>100), relatively high (**n>30), and small (*$n \leq 30$) number of sequences. Countries with no asterisks were not represented in our dataset.
(PDF)

Table S1 Number of viral migration between locations estimated using Markov jumps counts.
(PDF)

Table S2 Evolutionary rate and time-scale of HIV-1 subtype G and major regional clades circulating in Africa.
(PDF)

Table S3 Best fit demographic model for HIV-1 subtype G African clades.
(PDF)

Acknowledgments

We wish to thank Dr. Vera Bongertz for critical review of the manuscript and corrections on English language usage. We also thank the article reviewers for their helpful comments.

Author Contributions

Conceived and designed the experiments: GB ED. Performed the experiments: ED DM GB. Analyzed the data: ED GB. Contributed to the writing of the manuscript: ED DM GB.

References

1. Archer J, Robertson DL (2007) Understanding the diversification of HIV-1 groups M and O. Aids 21: 1693–1700.
2. Tatem AJ, Hemelaar J, Gray RR, Salemi M (2012) Spatial accessibility and the spread of HIV-1 subtypes and recombinants in sub-Saharan Africa. AIDS 26: 2351–2360.
3. Gnisci D, Trémolières M, (SWAC/OECD) (2009) West African Studies. Regional Atlas on West Africa. Population and settlement. Charpter 4. Migration. OECD Publishing: 67–85.
4. Hemelaar J, Gouws E, Ghys PD, Osmanov S (2011) Global trends in molecular epidemiology of HIV-1 during 2000–2007. Aids 25: 679–689.
5. Peeters M, Esu-Williams E, Vergne L, Montavon C, Mulanga-Kabeya C, et al. (2000) Predominance of subtype A and G HIV type 1 in Nigeria, with geographical differences in their distribution. AIDS Res Hum Retroviruses 16: 315–325.
6. Agwale SM, Zeh C, Robbins KE, Odama L, Saekhou A, et al. (2002) Molecular surveillance of HIV-1 field strains in Nigeria in preparation for vaccine trials. Vaccine 20: 2131–2139.
7. Ojesina AI, Sankale JL, Odaibo G, Langevin S, Meloni ST, et al. (2006) Subtype-specific patterns in HIV Type 1 reverse transcriptase and protease in Oyo State, Nigeria: implications for drug resistance and host response. AIDS Res Hum Retroviruses 22: 770–779.
8. Sankale JL, Langevin S, Odaibo G, Meloni ST, Ojesina AI, et al. (2007) The complexity of circulating HIV type 1 strains in Oyo state, Nigeria. AIDS Res Hum Retroviruses 23: 1020–1025.
9. Chaplin B, Eisen G, Idoko J, Onwujekwe D, Idigbe E, et al. (2011) Impact of HIV type 1 subtype on drug resistance mutations in Nigerian patients failing first-line therapy. AIDS Res Hum Retroviruses 27: 71–80.
10. Hamers RL, Wallis CL, Kityo C, Siwale M, Mandaliya K, et al. (2011) HIV-1 drug resistance in antiretroviral-naive individuals in sub-Saharan Africa after rollout of antiretroviral therapy: a multicentre observational study. Lancet Infect Dis 11: 750–759.
11. Ajoge HO, Gordon ML, Ibrahim S, Shittu OS, Ndung'u T, et al. (2012) Drug resistance pattern of HIV type 1 isolates sampled in 2007 from therapy-naive pregnant women in North-Central Nigeria. AIDS Res Hum Retroviruses 28: 115–118.
12. Imade GE, Sagay AS, Chaplin B, Chebu P, Musa J, et al. (2014) Short communication: Transmitted HIV drug resistance in antiretroviral-naive pregnant women in north central Nigeria. AIDS Res Hum Retroviruses 30: 127–133.

13. Mamadou S, Montavon C, Ben A, Djibo A, Rabiou S, et al. (2002) Predominance of CRF02-AG and CRF06-cpx in Niger, West Africa. AIDS Res Hum Retroviruses 18: 723–726.
14. Charpentier C, Bellecave P, Cisse M, Mamadou S, Diakite M, et al. (2011) High prevalence of antiretroviral drug resistance among HIV-1-untreated patients in Guinea-Conakry and in Niger. Antivir Ther 16: 429–433.
15. Chamberland A, Diabate S, Sylla M, Anagounou S, Geraldo N, et al. (2012) Transmission of HIV-1 drug resistance in Benin could jeopardise future treatment options. Sex Transm Infect 88: 179–183.
16. Yaotse DA, Nicole V, Roch NF, Mireille PD, Eric D, et al. (2009) Genetic characterization of HIV-1 strains in Togo reveals a high genetic complexity and genotypic drug-resistance mutations in ARV naive patients. Infect Genet Evol 9: 646–652.
17. Dagnra AY, Vidal N, Mensah A, Patassi A, Aho K, et al. (2011) High prevalence of HIV-1 drug resistance among patients on first-line antiretroviral treatment in Lome, Togo. J Int AIDS Soc 14: 30.
18. Ouedraogo-Traore R, Montavon C, Sanou T, Vidal N, Sangare L, et al. (2003) CRF06-cpx is the predominant HIV-1 variant in AIDS patients from Ouagadougou, the capital city of Burkina Faso. AIDS 17: 441–442.
19. Vergne L, Diagbouga S, Kouanfack C, Aghokeng A, Butel C, et al. (2006) HIV-1 drug-resistance mutations among newly diagnosed patients before scaling-up programmes in Burkina Faso and Cameroon. Antivir Ther 11: 575–579.
20. Tebit DM, Sangare L, Tiba F, Saydou Y, Makamtse A, et al. (2009) Analysis of the diversity of the HIV-1 pol gene and drug resistance associated changes among drug-naive patients in Burkina Faso. J Med Virol 81: 1691–1701.
21. Derache A, Maiga AI, Traore O, Akonde A, Cisse M, et al. (2008) Evolution of genetic diversity and drug resistance mutations in HIV-1 among untreated patients from Mali between 2005 and 2006. J Antimicrob Chemother 62: 456–463.
22. Maiga AI, Fofana DB, Maiga AC, Diallo F, Arkoub ZA, et al. (2012) Transmitted Antiretroviral Drug Resistance in Newly HIV-Infected and Untreated Patients in Segou and Bamako, Mali. AIDS Res Hum Retroviruses.
23. Fischetti L, Opare-Sem O, Candotti D, Sarkodie F, Lee H, et al. (2004) Molecular epidemiology of HIV in Ghana: dominance of CRF02_AG. J Med Virol 73: 158–166.
24. Delgado E, Ampofo WK, Sierra M, Torpey K, Perez-Alvarez L, et al. (2008) High prevalence of unique recombinant forms of HIV-1 in Ghana: molecular epidemiology from an antiretroviral resistance study. J Acquir Immune Defic Syndr 48: 599–606.
25. Nii-Trebi NI, Ibe S, Barnor JS, Ishikawa K, Brandful JA, et al. (2013) HIV-1 drug-resistance surveillance among treatment-experienced and -naive patients after the implementation of antiretroviral therapy in Ghana. PLoS One 8: e71972.
26. Ayouba A, Lien TT, Nouhin J, Vergne L, Aghokeng AF, et al. (2009) Low prevalence of HIV type 1 drug resistance mutations in untreated, recently infected patients from Burkina Faso, Cote d'Ivoire, Senegal, Thailand, and Vietnam: the ANRS 12134 study. AIDS Res Hum Retroviruses 25: 1193–1196.
27. Toni T, Masquelier B, Minga A, Anglaret X, Danel C, et al. (2007) HIV-1 antiretroviral drug resistance in recently infected patients in Abidjan, Cote d'Ivoire: A 4-year survey, 2002–2006. AIDS Res Hum Retroviruses 23: 1155–1160.
28. Hamel DJ, Sankale JL, Eisen G, Meloni ST, Mullins C, et al. (2007) Twenty years of prospective molecular epidemiology in Senegal: changes in HIV diversity. AIDS Res Hum Retroviruses 23: 1189–1196.
29. Diop-Ndiaye H, Toure-Kane C, Leye N, Ngom-Gueye NF, Montavon C, et al. (2010) Antiretroviral drug resistance mutations in antiretroviral-naive patients from Senegal. AIDS Res Hum Retroviruses 26: 1133–1138.
30. Esbjornsson J, Mild M, Mansson F, Norrgren H, Medstrand P (2011) HIV-1 molecular epidemiology in Guinea-Bissau, West Africa: origin, demography and migrations. PLoS One 6: e17025.
31. Delatorre E, Bello G (2013) Spatiotemporal dynamics of the HIV-1 CRF06_cpx epidemic in western Africa. AIDS 27: 1313–1320.
32. Faria NR, Suchard MA, Abecasis A, Sousa JD, Ndembi N, et al. (2012) Phylodynamics of the HIV-1 CRF02_AG clade in Cameroon. Infect Genet Evol 12: 453–460.
33. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, et al. (2005) An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 21: 3797–3800.
34. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59: 307–321.
35. Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online–a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33: W557–559.
36. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25: 1253–1256.
37. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol 55: 539–552.
38. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, et al. (1999) Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 73: 152–160.
39. Zhang M, Foley B, Schultz AK, Macke JP, Bulla I, et al. (2010) The role of recombination in the emergence of a complex and dynamic HIV epidemic. Retrovirology 7: 25.
40. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161: 1307–1320.
41. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.
42. Suchard MA, Rambaut A (2009) Many-core algorithms for statistical phylogenetics. Bioinformatics 25: 1370–1376.
43. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLoS Biol 4: e88.
44. Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. PLoS Comput Biol 5: e1000520.
45. Ferreira MAR, Suchard MA (2008) Bayesian analysis of elapsed times in continuous-time Markov chains. Canadian Journal of Statistics 26: 355–368.
46. Minin VN, Suchard MA (2008) Counting labeled transitions in continuous-time Markov models of evolution. J Math Biol 56: 391–412.
47. Talbi C, Lemey P, Suchard MA, Abdelatif E, Elharrak M, et al. (2010) Phylodynamics and human-mediated dispersal of a zoonotic virus. PLoS Pathog 6: e1001166.
48. Nunes MR, Faria NR, Vasconcelos HB, Medeiros DB, Silva de Lima CP, et al. (2012) Phylogeography of dengue virus serotype 4, Brazil, 2010–2011. Emerg Infect Dis 18: 1858–1864.
49. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22: 1185–1192.
50. Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol 18: 1001–1013.
51. Rambaut A, Drummond A (2007) Tracer v1.6. Available from: http://tree.bio.ed.ac.uk/software/tracer. Accessed 2014 Apr 25.
52. Rambaut A (2009) FigTree v1.4: Tree Figure Drawing Tool. Available from: http://tree.bio.ed.ac.uk/software/figtree/. Accessed 2014 April 25.
53. Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27: 2910–2912.
54. Rambaut A, Robertson DL, Pybus OG, Peeters M, Holmes EC (2001) Human immunodeficiency virus. Phylogeny and the origin of HIV-1. Nature 410: 1047–1048.
55. Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, 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. J Virol 74: 10498–10507.
56. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, et al. (2008) Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature 455: 661–664.
57. Kalish ML, Robbins KE, Pieniazek D, Schaefer A, Nzilambi N, et al. (2004) Recombinant viruses and early global HIV-1 epidemic. Emerg Infect Dis 10: 1227–1234.
58. Abecasis AB, Vandamme AM, Lemey P (2009) Quantifying differences in the tempo of human immunodeficiency virus type 1 subtype evolution. J Virol 83: 12917–12924.
59. Mehta SR, Wertheim JO, Delport W, Ene L, Tardei G, et al. (2011) Using phylogeography to characterize the origins of the HIV-1 subtype F epidemic in Romania. Infect Genet Evol 11: 975–979.
60. Delatorre E, Bello G (2013) Phylodynamics of the HIV-1 epidemic in Cuba. PLoS ONE 8: e72448.
61. Yusim K, Peeters M, Pybus OG, Bhattacharya T, Delaporte E, et al. (2001) Using human immunodeficiency virus type 1 sequences to infer historical features of the acquired immune deficiency syndrome epidemic and human immunodeficiency virus evolution. Philos Trans R Soc Lond B Biol Sci 356: 855–866.
62. Buve A, Bishikwabo-Nsarhaza K, Mutangadura G (2002) The spread and effect of HIV-1 infection in sub-Saharan Africa. Lancet 359: 2011–2017.
63. Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M, et al. (2003) Tracing the origin and history of the HIV-2 epidemic. Proc Natl Acad Sci U S A 100: 6588–6592.
64. Gisselquist D (2004) Impact of long-term civil disorders and wars on the trajectory of HIV epidemics in sub-Saharan Africa. SAHARA J 1: 114–127.
65. Salama P, Dondero TJ (2001) HIV surveillance in complex emergencies. Aids 15: S4–S12.
66. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, et al. (2014) Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. PLoS Pathog 10: e1003932.
67. Marechal V, Jauvin V, Selekon B, Leal J, Pelembi P, et al. (2006) Increasing HIV type 1 polymorphic diversity but no resistance to antiretroviral drugs in untreated patients from Central African Republic: a 2005 study. AIDS Res Hum Retroviruses 22: 1036–1044.
68. Aghokeng AF, Vergne L, Mpoudi-Ngole E, Mbangue M, Deoudje N, et al. (2009) Evaluation of transmitted HIV drug resistance among recently-infected antenatal clinic attendees in four Central African countries. Antivir Ther 14: 401–411.
69. Koyalta D, Charpentier C, Beassamda J, Rey E, Si-Mohamed A, et al. (2009) High frequency of antiretroviral drug resistance among HIV-infected adults

receiving first-line highly active antiretroviral therapy in N'Djamena, Chad. Clin Infect Dis 49: 155–159.
70. Djoko CF, Wolfe ND, Vidal N, Tamoufe U, Montavon C, et al. (2010) HIV type 1 pol gene diversity and genotypic antiretroviral drug resistance mutations in Malabo, Equatorial Guinea. AIDS Res Hum Retroviruses 26: 1027–1031.
71. Caron M, Lekana-Douki SE, Makuwa M, Obiang-Ndong GP, Biba O, et al. (2012) Prevalence, genetic diversity and antiretroviral drugs resistance-associated mutations among untreated HIV-1-infected pregnant women in Gabon, central Africa. BMC Infect Dis 12: 64.
72. Pandrea I, Robertson DL, Onanga R, Gao F, Makuwa M, et al. (2002) Analysis of partial pol and env sequences indicates a high prevalence of HIV type 1 recombinant strains circulating in Gabon. AIDS Res Hum Retroviruses 18: 1103–1116.
73. Cuevas MT, Ruibal I, Villahermosa ML, Diaz H, Delgado E, et al. (2002) High HIV-1 genetic diversity in Cuba. Aids 16: 1643–1653.
74. Kouri V, Aleman Y, Perez L, Perez J, Fonseca C, et al. (2012) High frequency of antiviral drug resistance and non-B subtypes in HIV-1 patients failing antiviral therapy in Cuba. J Clin Virol 55: 348–355.
75. Machado LY, Blanco M, Dubed M, Diaz HM, Ruiz NM, et al. (2012) HIV type 1 genetic diversity in newly diagnosed Cuban patients. AIDS Res Hum Retroviruses 28: 956–960.
76. Esteves A, Parreira R, Venenno T, Franco M, Piedade J, et al. (2002) Molecular epidemiology of HIV type 1 infection in Portugal: high prevalence of non-B subtypes. AIDS Res Hum Retroviruses 18: 313–325.
77. Esteves A, Parreira R, Piedade J, Venenno T, Franco M, et al. (2003) Spreading of HIV-1 subtype G and envB/gagG recombinant strains among injecting drug users in Lisbon, Portugal. AIDS Res Hum Retroviruses 19: 511–517.
78. Palma AC, Araujo F, Duque V, Borges F, Paixao MT, et al. (2007) Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. Infect Genet Evol 7: 391–398.
79. Bobkova M (2013) Current status of HIV-1 diversity and drug resistance monitoring in the former USSR. AIDS Rev 15: 204–212.
80. Faria NR, Hodges-Mameletzis I, Silva JC, Rodes B, Erasmus S, et al. (2012) Phylogeographical footprint of colonial history in the global dispersal of human immunodeficiency virus type 2 group A. J Gen Virol 93: 889–899.

8.6 APÊNDICE 6

Origin and Population Dynamics of a Novel HIV-1 Subtype G Clade Circulating in Cape Verde and Portugal

Autores: Isabel Inês M. de Pina-Araujo*, **Edson Delatorre***, Monick L. Guimarães, Mariza G. Morgado, Gonzalo Bello

* Estes autores contribuíram igualmente para este estudo.

Jornal: PLOS ONE, 2015

PLOS | ONE

RESEARCH ARTICLE

Origin and Population Dynamics of a Novel HIV-1 Subtype G Clade Circulating in Cape Verde and Portugal

Isabel Inês M. de Pina-Araujo[1☯], Edson Delatorre[2☯], Monick L. Guimarães[2], Mariza G. Morgado[2], Gonzalo Bello[2]*

1 Departamento de Ciência e Tecnologia, Universidade de Cabo Verde, Praia, Santiago, Cabo Verde,
2 Laboratório de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil

☯ These authors contributed equally to this work.
* gbellobr@gmail.com

Citation: de Pina-Araujo IIM, Delatorre E, Guimarães ML, Morgado MG, Bello G (2015) Origin and Population Dynamics of a Novel HIV-1 Subtype G Clade Circulating in Cape Verde and Portugal. PLoS ONE 10(5): e0127384. doi:10.1371/journal.pone.0127384

Academic Editor: Massimo Ciccozzi, National Institute of Health, ITALY

Received: March 9, 2015

Accepted: April 15, 2015

Published: May 20, 2015

Data Availability Statement: All sequences were retrieved from the Los Alamos HIV Sequence Database (www.hiv.lanl.gov) and corresponding GenBank accession numbers are available in S1 Table.

Funding: This work was sponsored by Public Health Service grants 490468/2008-0 and E-26/111.758/2012 from the CNPq. IIMPA was supported by a fellowship from the CAPES PEC-PG Program. ED is supported by a fellowship from the FAPERJ. The funders had no role in study design, data collection

Abstract

The human immunodeficiency virus type 1 (HIV-1) subtype G is the most prevalent and second most prevalent HIV-1 clade in Cape Verde and Portugal, respectively; but there is no information about the origin and spatiotemporal dispersal pattern of this HIV-1 clade circulating in those countries. To this end, we used Maximum Likelihood and Bayesian coalescent-based methods to analyze a collection of 578 HIV-1 subtype G *pol* sequences sampled throughout Portugal, Cape Verde and 11 other countries from West and Central Africa over a period of 22 years (1992 to 2013). Our analyses indicate that most subtype G sequences from Cape Verde (80%) and Portugal (95%) branched together in a distinct monophyletic cluster (here called $G_{CV\text{-}PT}$). The $G_{CV\text{-}PT}$ clade probably emerged after a single migration of the virus out of Central Africa into Cape Verde between the late 1970s and the middle 1980s, followed by a rapid dissemination to Portugal a couple of years later. Reconstruction of the demographic history of the $G_{CV\text{-}PT}$ clade circulating in Cape Verde and Portugal indicates that this viral clade displayed an initial phase of exponential growth during the 1980s and 1990s, followed by a decline in growth rate since the early 2000s. Our data also indicate that during the exponential growth phase the $G_{CV\text{-}PT}$ clade recombined with a preexisting subtype B viral strain circulating in Portugal, originating the CRF14_BG clade that was later disseminated to Spain and Cape Verde. Historical and recent human population movements between Angola, Cape Verde and Portugal probably played a key role in the origin and dispersal of the $G_{CV\text{-}PT}$ and CRF14_BG clades.

Introduction

The global dissemination of human immunodeficiency virus type 1 (HIV-1) group M, the pandemic clade of HIV, resulted from the random exportation out of Central Africa of a few viral strains designated as subtypes (A–D, F–H, J and K) and inter-subtype circulating recombinant forms (CRFs) [1].

and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Subtype G is the sixth most prevalent HIV-1 clade in the world accounting for nearly 5% of all global infections [2]. This subtype reaches the highest prevalence in some African countries, comprising 30–50% of HIV-1 infections in Cape Verde [3,4] and Nigeria [5–12], and 5–15% of HIV-1 infections in Angola [13–15], Benin [16], Niger [17,18] and Togo [19,20]. A recent study conducted by our group suggests that subtype G most probably emerged in Central Africa around the late 1960s and was rapidly disseminated into the West and West Central African regions [21]. This study showed that basal subtype G lineages (G_{CA}) were mostly restricted to Central and West Central African countries. Two subtype G strains, however, gained access to West Africa between the middle and the late 1970s and fueled secondary local outbreaks, leading to the origin of two major subtype G West African clades (G_{WA-I} and G_{WA-II}).

Some subtype G strains where also disseminated out of the African continent and the most remarkable example is Portugal where subtype G is the second most prevalent HIV-1 clade (>10%), after subtype B (> 40%) [22–25]. The high prevalence of subtypes B and G in Portugal has also promoted the appearance of different types of B/G recombinant strains, including one circulating recombinant form (CRF14_BG) that was initially identified in Galicia, Northern Spain [26]. According to a previous study, the CRF14_BG probably emerged in Portugal in the early 1990s and later spread to Galicia in the late 1990s as a consequence of the mobility of HIV-infected injecting drug users (IDUs) [27].

Notably, about two-thirds of the subtype G viruses previously described in Portugal were found in individuals from Angola and Cape Verde [23]. These countries are two former Portuguese African colonies that have strong historic links and maintain ongoing relationships with Portugal and displayed a relatively high prevalence of subtype G [3,4,13–15]. The high numbers of immigrants from Angola and Cape Verde who enter Portugal and also those Portuguese returning after living in the former Portuguese African colonies from the 1970s onwards [28], supposes a potential risk for introduction of HIV-1 subtype G strains in Portugal. However, the precise evolutionary relationship between subtype G viruses circulating in Angola, Cape Verde and Portugal remains unknown.

The objective of this study was to reconstruct the phylogenetic relationship, onset date and dissemination routes of the HIV-1 subtype G clades circulating in Angola, Cape Verde and Portugal. To this end, we used Maximum Likelihood and Bayesian coalescent-based approaches to analyze 578 HIV-1 subtype G *pol* sequences isolated from Portugal, West Africa and Central Africa over a period of 22 years (1992 to 2013).

Materials and Methods

Sequence dataset

All HIV-1 subtype G *pol* sequences from Portugal and West/Central African countries that covered the entire protease and partial reverse transcriptase (PR/RT) regions (nt 2253–3272 relative to HXB2 clone) and for which the sampling year was known, were downloaded from the Los Alamos HIV Sequence Database (www.hiv.lanl.gov) by December 2014. The same procedure was adopted to obtain the subtype G *pol* fragment of all CRF14_BG sequences from Portugal and Spain (the main countries were this CRF circulates) with full-length genome characterization up to date. The subtype assignment of all sequences was confirmed by REGA HIV subtyping tool v.2 [29] and bootscanning analysis. In bootscanning analyses, supporting branching of query sequences with HIV-1 group M subtype reference sequences was determined in Neighbor-Joining trees constructed with the Kimura two-parameter model, within a 250bp window moving in steps of 10 bases, using Simplot software v.3.5.1 [30]. Sequences with incorrect classification, multiple sequences from the same individual and sequences from countries poorly represented ($n < 4$ sequences) were removed, resulting in a final data set of

Table 1. HIV-1 subtype G *pol* dataset.

Los Alamos classification	Region	Country	*N*	Sampling interval
Subtype G	West Africa	Benin	15	2004–2009
		Ghana	9	2002–2010
		Cape Verde	60	2005–2011
		Nigeria	223	1992–2013
		Senegal	12	1998–2010
		Togo	28	2006–2011
	West Central Africa	Cameroon	62	1997–2012
		Gabon	6	2000–2008
		Equatorial Guinea	4	2005–2009
	Central Africa	Angola	20	1997–2010
		DRC	12	1993–2007
		Republic of Congo	8	2003
	Europe	Portugal	107	1998–2008
CRF14_BG	Europe	Portugal	2[a]	1998–2008
		Spain	10[a]	1999–2005

[a] Subtype G *pol* fragments recovered from full-length CRF14_BG reference sequences.

doi:10.1371/journal.pone.0127384.t001

578 HIV-1 subtype G *pol* sequences (Table 1). Sequence's GenBank accession numbers are available in S1 Table. All codon positions known to be associated with major antiretroviral drug resistance were maintained in the final alignment because phylogenetic trees constructed on alignments with or without such positions resulted in the same overall topology (data not shown). The presence of phylogenetic signal and substitution saturation in our datasets was investigated by: 1) using the likelihood mapping analysis [31] performed with TREE-PUZZLE v5.2 program [32] implemented in the online web platform Mobyle@Pasteur v1.5 [33], and 2) plotting the observed number of transitions and transversions against genetic distance for each pairwise comparison, calculated under the GTR+I+G nucleotide substitution model using DAMBE v5.3 program [34].

Identification of major HIV-1 subtype G clades

Major HIV-1 subtype G clades were identified by Maximum Likelihood (ML) phylogenetic analysis. A ML phylogeny was constructed with the PhyML 3.0 program [35] using an online web server [36]. The ML tree was inferred under the GTR+I+G nucleotide substitution model as recommended by the jModeltest program [37], the heuristic tree search was performed using the SPR branch-swapping algorithm and branch support was calculated with the approximate likelihood-ratio (aLRT) SH-like test [38].

Analysis of spatiotemporal dispersion pattern and demographic history

The evolutionary rate (μ, nucleotide substitutions per site per year, subst./site/year), the age of the most recent common ancestor (T_{MRCA}, years), the ancestral geographic movements, and the mode and rate (r, years-1) of population growth of HIV-1 subtype G clades were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST v1.8 [39,40] with BEAGLE to improve run-time [41]. Analyses were performed under a GTR+I+G nucleotide substitution model. The temporal scale of the evolutionary process was estimated from the sampling dates of the sequences using a relaxed uncorrelated lognormal

molecular clock model and a uniform prior on clock rate (1.5–3.0 x 10^{-3} subst/site/year) [42]. Migration events throughout the phylogenetic history were inferred using a reversible discrete Bayesian phylogeographic model [43], in which all possible reversible exchange rates between locations were equally likely, and a CTMC rate reference prior [44]. Changes in effective population size through time were initially estimated using a flexible Bayesian Skyline coalescent model [45] and estimates of the population growth rate were subsequently obtained using the parametric model (logistic, exponential or expansion) that provided the best fit to the demographic signal contained in datasets. Comparison between demographic models was performed using the log marginal likelihood (ML) estimation based on path sampling (PS) and stepping-stone sampling (SS) methods [46]. MCMC chains were run for 10–100 x 10^6 generations. Adequate chain mixing and uncertainty in parameter estimates were assessed by calculating the Effective Sample Size (ESS) and the 95% Highest Probability Density (HPD) values, respectively, using the TRACER v1.6 program [47]. Maximum clade credibility (MCC) trees were summarized from the posterior distribution of trees with TreeAnnotator and visualized with FigTree v1.4.0 [48].

Results

Identification of major HIV-1 subtype G clades

The likelihood mapping analysis and the transitions/transversions versus divergence plots indicates that all datasets used in our study retained enough phylogenetic signal for consistent phylogenetic inferences and no evidence of substitution saturation (S1 Fig). The ML phylogenetic tree of 578 HIV-1 subtype G *pol* sequences (566 classified as subtype G and 12 classified as CRF14_BG in the Los Alamos HIV Sequence Database) isolated in Portugal, Spain and 12 countries from West and Central Africa between 1992 and 2013 (Table 1) points to a clear phylogeographic subdivision of viral strains (Fig 1). Subtype G sequences from continental western African countries branched mostly in two large monophyletic clades ($G_{WA\text{-}I}$ and $G_{WA\text{-}II}$) that were nested among the most basal clades from Central and West Central Africa (G_{CA}), consistent with our previous findings [21]. Although some subtype G sequences from Cape Verde (n = 10) also branched within the $G_{WA\text{-}I}$ clade; most sequences from this insular West African country (n = 48) branched together with most subtype G sequences from Portugal (n = 102) in a distinct monophyletic clade ($G_{CV\text{-}PT}$) nested among basal G_{CA} lineages. All CRF14_BG sequences and several subtype G *pol* sequences from Portugal (n = 78) and Cape Verde (n = 7) formed a highly supported sub-cluster (CRF14_BG-like) within the $G_{CV\text{-}PT}$ clade. According to the relative prevalence of the distinct subtype G clades, we can describe four basic molecular epidemiologic scenarios (Fig 2 and S2 Table): 1) basal G_{CA} clades are the predominant subtype G lineages circulating in countries from Central (90%) and West-Central (50%) African regions; 2) the $G_{WA\text{-}I}$ clade was the predominant lineage detected in Nigeria (78%), Senegal (50%) and Benin (47%); 3) the $G_{WA\text{-}II}$ clade predominates in Togo/Ghana (84%); and 4) the $G_{CV\text{-}PT}$ clade was the dominant lineage in Cape Verde (80%) and Portugal (95%).

Spatiotemporal dispersal pattern of the HIV-1 $G_{CV\text{-}PT}$ clade

To reconstruct the subtype G migrations between Africa and Portugal, all subtype G sequences belonging to the $G_{CV\text{-}PT}$ clade (excluding the CRF14_BG-like sub-clade) (n = 69) were combined with basal G_{CA} strains of Central African origin (n = 73). Sequences were divided in six geographical locations, as those neighboring countries from Central and West-Central Africa comprising few samples ($n < 20$) were grouped into the same location state (S3 Table), and subjected to Bayesian phylogeographic analysis According to the Bayesian MCMC analysis, the most probable root location of the subtype G clade was placed in Central Africa (posterior state

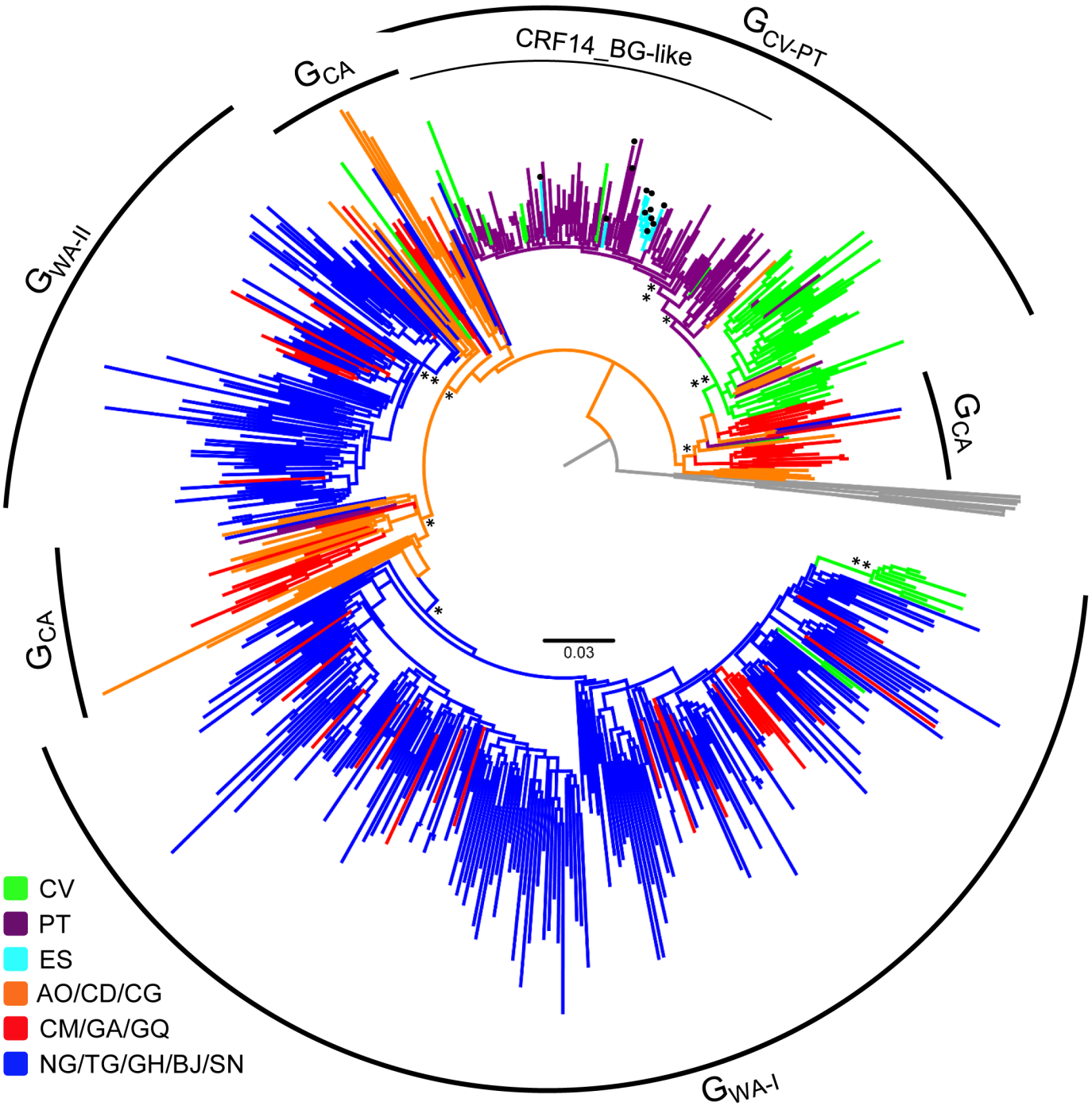


Fig 1. ML tree of the HIV-1 subtype G *pol* PR/RT sequences (~1,000 nt) from Central Africa, West Africa, Portugal and Spain. Branches are colored according to the geographic origin of each sequence as indicated in the legend (bottom left). Arcs indicate the positions of major subtype G clades circulating in Central Africa (G_{CA}), West Africa (G_{WA-I} and G_{WA-II}) and Cape Verde/Portugal (G_{CV-PT}) and the position of the subclade that comprises all CRF14_BG reference sequences (CRF14_BG-like). Black dots indicate the positions of the reference sequences classified as CRF14_BG based on full-length genome analysis. Asterisks point to key nodes with relatively high (*, *a*LRT > 0.80) and high (**, *a*LRT > 0.90) support. The tree was rooted using HIV-1 subtype A-D reference sequences. The branch lengths are drawn to scale with the bar at the center indicating nucleotide substitutions per site. AO/CD/CG: Angola/Democratic Republic of Congo/Republic of Congo; CM/GA/GQ: Cameroon/Gabon/Equatorial Guinea; NG/TG/GH/BJ/SN: Nigeria/Togo/Ghana/Benin/Senegal; CV: Cape Verde; PT: Portugal; ES: Spain.

doi:10.1371/journal.pone.0127384.g001

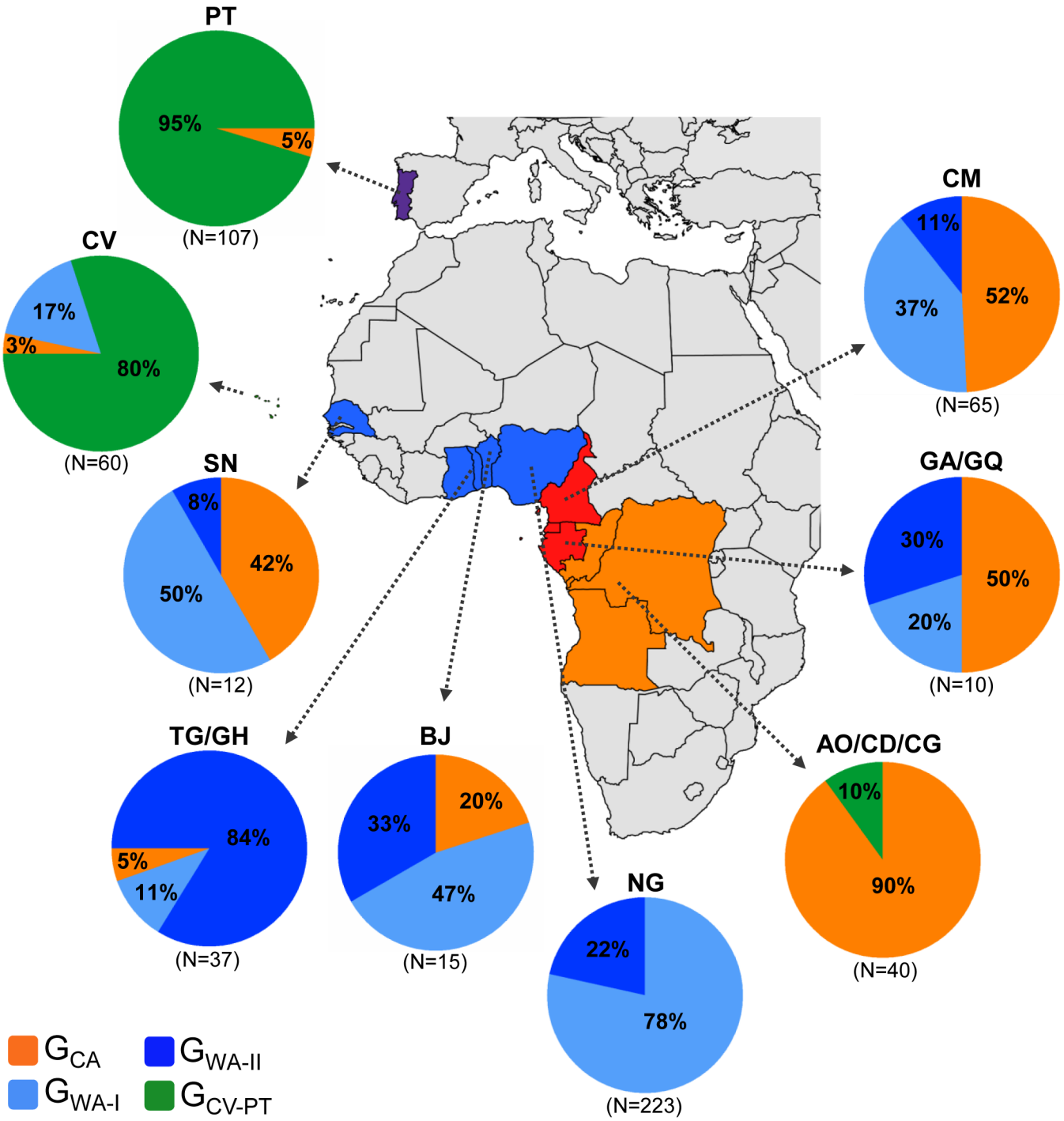


Fig 2. Prevalence of G_{CV-PT}, G_{WA-I}, G_{WA-II} and G_{CA} clades among HIV-1 subtype G infected individuals from different countries, estimated from phylogenetic analyses presented in Fig 1. The total number of subtype G sequences analyzed in each locality is indicated. Each clade is represented by a color as indicated at the legend. AO/CD/CG: Angola/Democratic Republic of Congo/Republic of Congo; BJ: Benin; CM: Cameroon; CV: Cape Verde; GA/GQ: Gabon/Equatorial Guinea; NG: Nigeria; PT: Portugal; TG/GH: Togo/Ghana; SN: Senegal.

doi:10.1371/journal.pone.0127384.g002

probability, *PSP* = 1), and the onset date of this clade was estimated to be 1964 (95% HPD: 1937–1978) (Fig 3). Sequences from Cape Verde branched at the base of the G_{CV-PT} clade,

Fig 3. Time-scaled Bayesian MCC tree of HIV-1 subtype G *pol* PR/RT sequences (~1,000 nt) from the G_{CA} and G_{CV-PT} clades. Branches are colored according to the most probable location state of their descendent nodes as indicated in the legend (upper left). The arc indicates the position of the G_{CV-PT} clade. Key nodes corresponding to the MRCA of the Cape Verdean and Portuguese G_{CV-PT} lineages are indicated with circles and the median T_{MRCA} (with the corresponding 95% HPD interval) of each lineage is indicated at the bottom left. Asterisks point to key nodes with relatively high (*, *PP* > 0.80) and high (**, *PP* > 0.90) posterior probability support. Branch lengths are drawn to a scale of years. The tree was automatically rooted under the assumption of a

relaxed molecular clock. AO/CD/CG: Angola/ Democratic Republic of Congo /Republic of Congo; CM: Cameroon; CV: Cape Verde; GA/GQ: Gabon/ Equatorial Guinea; PT: Portugal.

doi:10.1371/journal.pone.0127384.g003

whereas most sequences from Portugal branched in a monophyletic cluster nested within the Cape Verdean sequences (Fig 3). This analysis suggests that the $G_{CV\text{-}PT}$ clade most probably migrates from Central Africa to Cape Verde (*PSP* = 0.68) at 1977 (95% HPD: 1972–1982) and rapidly moved from Cape Verde to Portugal in 1979 (95% HPD: 1974–1984). A few additional exchanges of the $G_{CV\text{-}PT}$ clade between Cape Verde and Portugal were detected at later times (Fig 3).

Demographic history of the HIV-1 $G_{CV\text{-}PT}$ clade

To reconstruct the demographic history of the $G_{CV\text{-}PT}$ clade, all subtype G sequences from Cape Verde (n = 41) and Portugal (n = 24) that branched within this clade (excluding the CRF14_BG-like sub-clade) were selected. In agreement with our previous analysis, most subtype G sequences from Portugal branched in a sub-cluster nested among basal Cape Verdean sequences (Fig 4A). This new analysis, however, supports a relatively more recent time-scale than previous estimations. According to this new analysis, the $G_{CV\text{-}PT}$ clade probably arose in Cape Verde (*PSP* = 0.76) in 1984 (95% HPD: 1979–1989) and was rapidly disseminated to Portugal in 1987 (95% HPD: 1983–1990). Two additional migrations of the $G_{CV\text{-}PT}$ clade from Cape Verde to Portugal and one migration event from Portugal to Cape Verde were also detected, in agreement with our previous analysis (Fig 4A). The Bayesian skyline plot (BSP) analysis suggests that the $G_{CV\text{-}PT}$ clade experienced a fast exponential growth during the 1980s and 1990s, followed by a more recent stabilization since the early 2000s (Fig 4B). According to the logistic growth coalescent model, selected as the best-fit demographic model for the $G_{CV\text{-}PT}$ clade (log BF $>$ 10) (S4 Table), the mean growth rate of this subtype G clade was 0.52 $year^{-1}$ (95% HPD: 0.32–0.77 $year^{-1}$) (Fig 4C).

Origin of the CRF14_BG clade

To investigate the origin of the parental subtype lineage that gave rise to the CRF14_BG, all *pol* sequences that branched within the CRF14_BG-like subclade (n = 97) were combined with sequences from clades G_{CA} and $G_{CV\text{-}PT}$. The overall topology and temporal structure of the Bayesian MCC trees remains conserved after inclusion of the CRF14_BG-like subclade, but placed most of the posterior root state probability mass of the $G_{CV\text{-}PT}$ clade in Portugal (*PSP* = 0.55–0.81) (Fig 5 and S5 Table). Both Bayesian MCMC analyses showed that all CRF14_BG-like sequences formed a well-supported sub-cluster (*PP* $>$ 0.90) nested among basal subtype G Portuguese sequences within the $G_{CV\text{-}PT}$ radiation (Fig 5). Those analyses support that the CRF14_BG-like clade most probably arose in Portugal (*PSP* = 1) and was later disseminated at multiple times from Portugal to both Spain and Cape Verde (Fig 5). The T_{MRCA} of the CRF14_BG-like clade was traced to 1986 (95% HPD: 1982–1991) when basal G_{CA} strains were included in the analysis (Fig 5A), and to 1991 (95% HPD: 1988–1994) when basal G_{CA} strains were not included (Fig 5B).

Discussion

This and our previous study [21] indicate that the HIV-1 subtype G likely originated in Central Africa around the middle-late 1960s and began to be disseminated to Western and West-Central Africa from the middle 1970s onwards. Some of the subtype G strains disseminated out of Central Africa fueled secondary outbreaks that led to the origin of regional-specific subtype G

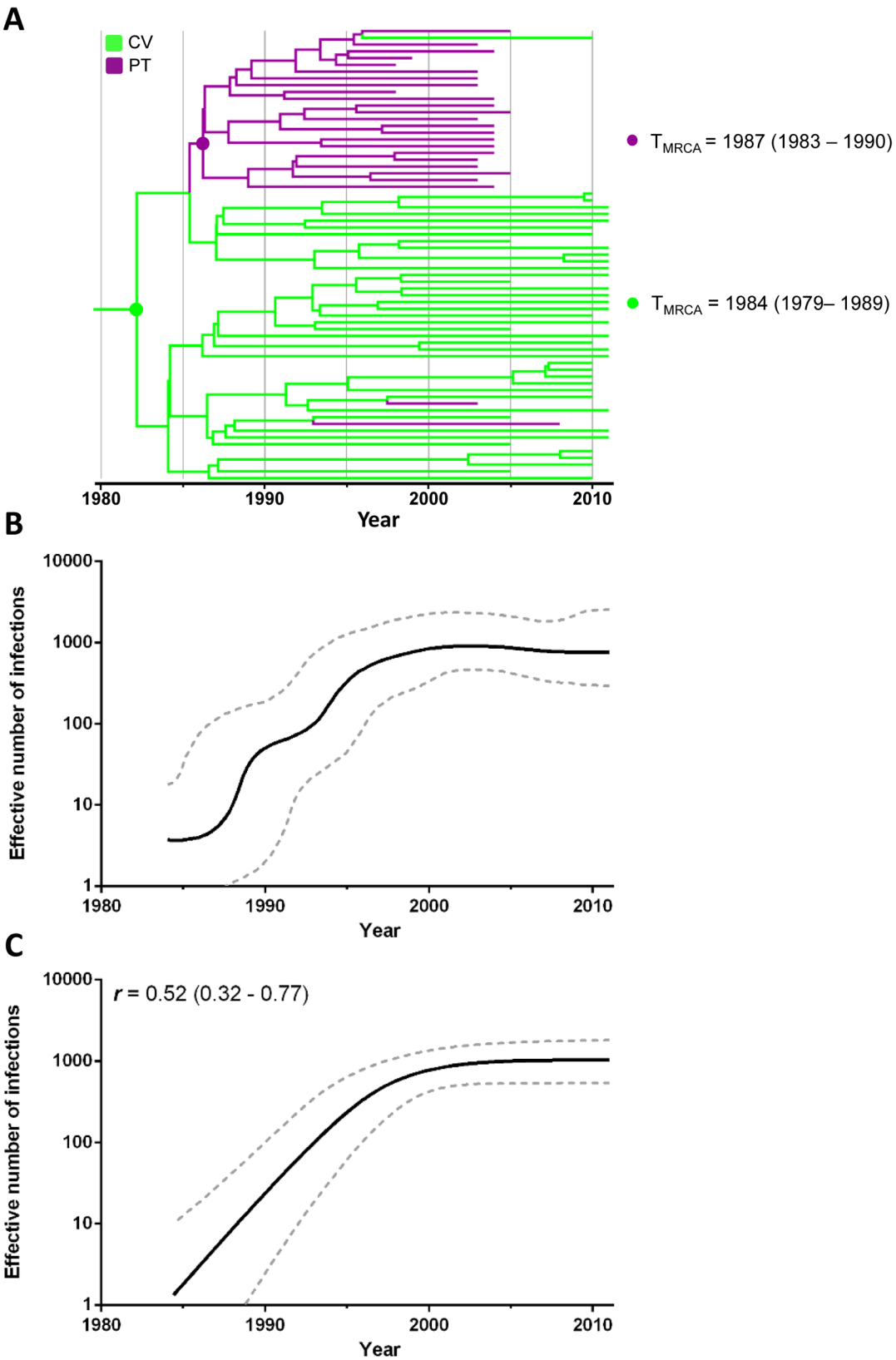


Fig 4. Demographic history of the HIV-1 G_{CV-PT} clade circulating in Cape Verde and Portugal. A) Time-scaled Bayesian MCC tree of the HIV-1 G_{CV-PT} clade. Branches are colored according to the most probable location state of their descendent nodes as indicated in the legend (upper left). Key nodes

corresponding to the MRCA of the Cape Verde and Portuguese $G_{CV\text{-}PT}$ lineages are indicated with circles and the median T_{MRCA} (with the corresponding 95% HPD interval) of each lineage is indicated at right. Branch lengths are drawn to a scale of years. The tree was automatically rooted under the assumption of a relaxed molecular clock. B and C) Effective number of infections (y-axis; log10 scale) through time (x-axis; calendar years) estimated using Bayesian skyline (B) and logistic growth (C) coalescent models. Median (solid line) and 95% HPD intervals (dashed lines) of the effective number of infections estimated through time are shown in each graphic. The median growth rate (with the corresponding 95% HPD interval) of $G_{CV\text{-}PT}$ clade estimated under the logistic growth model is indicated in the upper left corner.

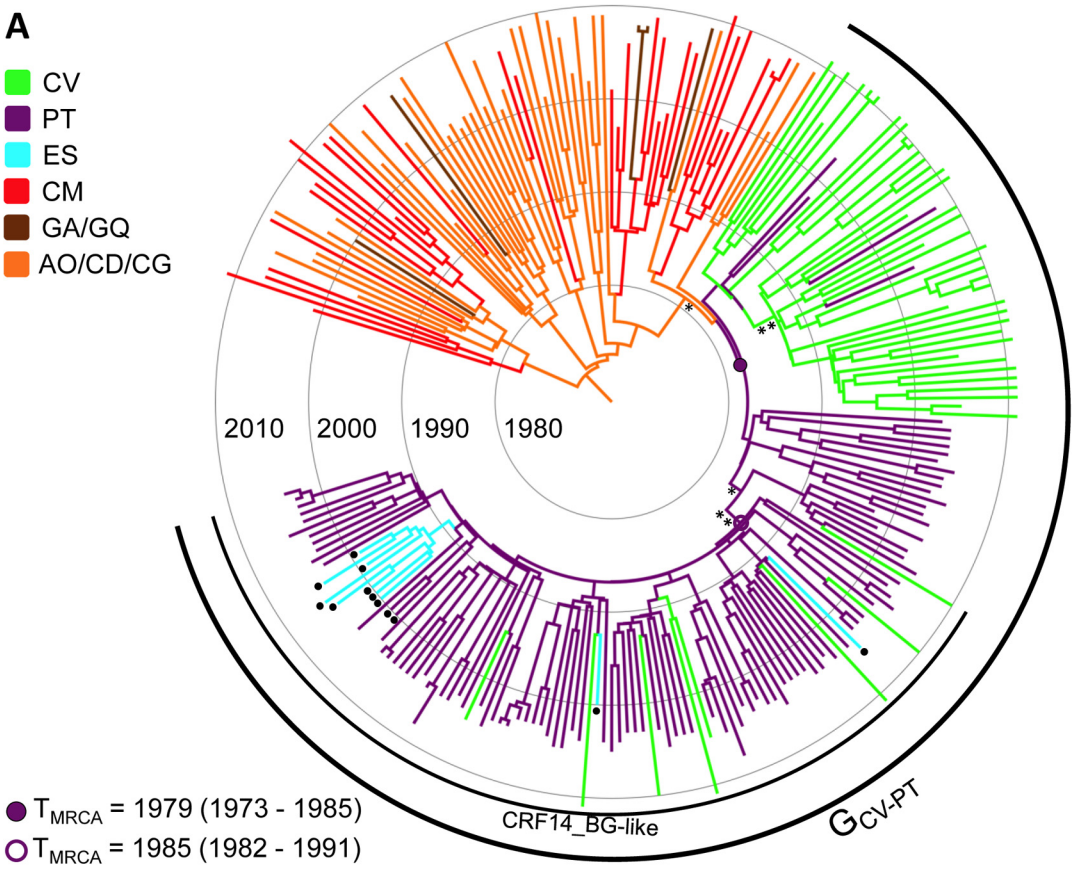
doi:10.1371/journal.pone.0127384.g004

clades. The major subtype G clades detected in our previous study in West Africa were the $G_{WA\text{-}I}$ (that most probably emerged in Nigeria around the middle 1970s) and the $G_{WA\text{-}II}$ (that most probably emerged in Togo or Ghana around the late 1970s) [21]. In the present study we identified a novel major clade ($G_{CV\text{-}PT}$) that probably emerged between the late 1970s and the middle 1980s and circulates in Cape Verde and Portugal.

The $G_{CV\text{-}PT}$ clade comprises 95% and 80% of HIV-1 subtype G *pol* sequences from Portugal and Cape Verde included in our study, respectively. Within the $G_{CV\text{-}PT}$ radiation, most sequences from Portugal (73%) branched in a monophyletic subclade together with the CRF14_BG reference sequences, whereas the remaining Portuguese sequences branched at the base of the CRF14_BG-like subclade. This clearly indicates that the $G_{CV\text{-}PT}$ clade is the parental subtype G lineage of the CRF14_BG variant and that the CRF14_BG clade is probably more prevalent in Portugal than the parental $G_{CV\text{-}PT}$ clade, consistent with previous findings [25]. It is also important to note that a small fraction of $G_{CV\text{-}PT}$ *pol* sequences from Cape Verde (15%) also branched within the CRF14_BG-like clade, indicating that this recombinant lineage not only circulates in Portugal and Spain, but also in Cape Verde. Full-length genome analyses of Cape Verdean HIV-1 subtype G *pol* sequences that branched within the CRF14_BG-like subclade should be performed to confirm this hypothesis.

The phylogeographic analyses that combined subtype G sequences of the $G_{CV\text{-}PT}$ clade (with exception of the CRF14_BG-like lineage) and basal G_{CA} clades consistently pointed to Cape Verde as the most probable root location of the $G_{CV\text{-}PT}$ clade (*PSP* = 0.68–0.76). When CRF14_BG-like sequences are included, the root location of the $G_{CV\text{-}PT}$ clade was most probably placed in Portugal (*PSP* = 0.55–0.81). It has been shown that convenience sampling (particularly sampling heterogeneity) can obfuscate the accurate estimation of ancestral spatial locations based on standard phylogeographic continuous-time Markov chain implementation [49]. When CRF14_BG-like sequences are included, the number of Portuguese sequences (n = 104) far exceeds the number of Cape Verdean sequences (n = 48) within the $G_{CV\text{-}PT}$ clade and such a larger sample from Portugal may results in the higher support for this location as the origin of that clade. Thus, according to the more balanced data sets the founder $G_{CV\text{-}PT}$ ancestor probably moved from Central Africa to Cape Verde and later passed from Cape Verde to Portugal.

Whereas the inclusion of the CRF14_BG-like sequences has a great impact on estimation of the $G_{CV\text{-}PT}$ ancestral root location, ancestral root ages were mainly influenced by the inclusion of basal G_{CA} clades. The median T_{MRCA} of the $G_{CV\text{-}PT}$ clade was traced to the late 1970s when basal G_{CA} clades were included, and to the middle 1980s when those basal sequences were not included (S5 Table). Similarly, the T_{MRCA} of the CRF14_BG clade moved from the middle 1980s to the early 1990s when G_{CA} clades were removed from the analysis (S5 Table). This suggests that inclusion of basal lineages from Central Africa tend to produce slightly older internal node ages, although no significant changes are observed in the mean estimated substitution rates (S5 Table). This observation, however, should be interpreted with caution because those T_{MRCA} estimates displayed a considerable overlap of the confidence interval and thus should not be regarded as statistically different.


A
CV
PT
ES
CM
GA/GQ
AO/CD/CG
2010
2000
1990
1980
CRF14_BG-like
GCV-PT
TMRCA = 1979 (1973 - 1985)
TMRCA = 1985 (1982 - 1991)

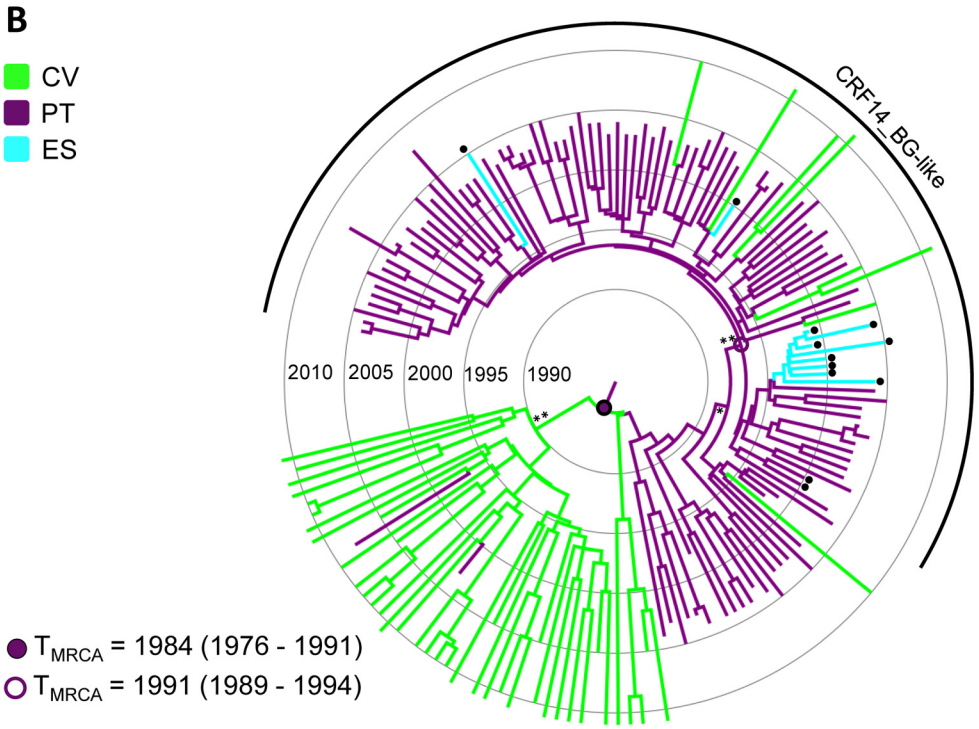

B
CV
PT
ES
CRF14_BG-like
2010
2005
2000
1995
1990
TMRCA = 1984 (1976 - 1991)
TMRCA = 1991 (1989 - 1994)

Fig 5. Time-scaled Bayesian MCC tree of HIV-1 subtype G *pol* PR/RT sequences (~1,000 nt) from the G_{CA}, G_{CV-PT} and CRF14_BG-like clades. Sequences that branched within the CRF14_BG-like subclade were combined with sequences from G_{CA} and G_{CV-PT} clades (A) or only G_{CV-PT} clade (B). Branches are colored according to the most probable location state of their descendent nodes as indicated in the legend (upper left). Arcs indicate the positions of G_{CV-PT} and CRF14_BG-like clades. Nodes corresponding to the MRCA of those clades are indicated with circles and the median T_{MRCA} (with the corresponding 95% HPD interval) of each clade is indicated at the bottom left. Black dots indicate the position of the CRF14_BG reference sequences. Asterisks point to key nodes with high relatively high (*, *PP* > 0.80) and high (**, *PP* > 0.90) posterior probability support. Branch lengths are drawn to a scale of years. The tree was automatically rooted under the assumption of a relaxed molecular clock. AO/CD/CG: Angola/ Democratic Republic of Congo /Republic of Congo; CM: Cameroon; CV: Cape Verde; GA/GQ: Gabon/Equatorial Guinea; PT: Portugal; ES: Spain.

doi:10.1371/journal.pone.0127384.g005

Regardless the precise root age, our phylogeographic analyses support a nearly simultaneous introduction and concurrent dissemination of the G_{CV-PT} clade in Cape Verde and Portugal. Our phylogeographic analyses based on balanced datasets suggest that the G_{CV-PT} clade started to be disseminated in Portugal only a couple of years later than the estimated introduction of the virus into Cape Verde. Of note, the estimated time-frame (1977–1984) for introduction and dissemination of the G_{CV-PT} clade in Cape Verde and Portugal was preceded by a phase of negative migratory outflow in Angola [50], associated to the exodus of thousands of Portuguese citizens of European and African ethnicity from Angola after the country independence in 1974. This may have fueled the chance exportation of the G_{CV-PT} ancestor strain from Angola into Cape Verde and its rapid dissemination to Portugal, thus suggesting that the global route of spread of the G_{CV-PT} clade was probably laid out along the colonial history ties, as has been previously demonstrated for the HIV-2 group A [49].

Despite the continuous and extensive migration of people between Angola, Cape Verde, and Portugal [28,51], subtype G strains sampled in those Portuguese-speaking countries retain a high phylogeographic structure with relative few viral exchanges among them. We have detected a total of: 1) four independent introductions of G_{CA} strains from Central Africa into Portugal, 2) three introductions of G_{CA} strains from Central Africa into Cape Verde, 3) three introductions of G_{CV-PT} strains from Cape Verde into Portugal, and 4) one G_{CV-PT} migration and five CRF14_BG introductions from Portugal into Cape Verde. Although the continuous viral exchanges among these countries may suppose a risk to the emergence of new country-specific subtype G lineages, most viral introductions seem to have failed to sustain new local subtype G epidemics with exception of the G_{CV-PT} founder strain.

According to our analysis, the G_{CV-PT} clade displayed a logistic population growth pattern characterized by an initial phase of exponential growth with a median rate of 0.52 $year^{-1}$ (95% HPD: 0.32–0.77 $year^{-1}$), followed by a decline in growth rate since the early 2000s. The median estimated logistic growth rate of the G_{CV-PT} clade was similar to that estimated for basal G_{CA} clades in Central Africa (0.47 $year^{-1}$) [21] and the G_{CU} clade circulating in Cuba (0.55 $year^{-1}$) [52]; but lower than those previously estimated for the G_{WA-I} (0. 75 $year^{-1}$) and G_{WA-II} (0.95 $year^{-1}$) clades circulating in continental West African countries [21] (S6 Table). The differential growth rates detected among different subtype G clades could be associated to clade-specific or ecological-specific differences in viral transmissibility. Further studies should be performed to understand whether the G_{WA-I} and G_{WA-II} clades introduced into continental West Africa displayed a higher intrinsic transmissibility or encountered more favorable epidemiological conditions for local and regional expansion than those disseminated within Central Africa, Cape Verde, Cuba and Portugal.

A previous study concluded that the CRF14_BG emerged in Portugal in the early 1990s and then spread to the North of Spain in late 1990s following the mobility of HIV-infected IDUs [27]. Our phylogeographic analyses indicate that the CRF14_BG clade probably arose in Portugal between the middle 1980s and the early 1990s, which is fully consistent with the previous estimation and with epidemiological data showing that CRF14_BG was already circulating in Lisbon in 1993 [27]. According to this estimate, the recombinant ancestor of the CRF14_BG

clade was generated about five years after the estimated arrival of the parental $G_{CV\text{-}PT}$ clade into Portugal, thus indicating a very rapid generation of BG recombinants in this country. After a period of local dissemination within Portugal, the CRF14_BG clade was dispersed not only from Portugal to Spain, but also probably to Cape Verde at multiple times.

In summary, this study reveals that most HIV-1 subtype G infections in Cape Verde and Portugal have resulted from the local dissemination of a single clade (here called $G_{CV\text{-}PT}$) that probably emerged after a single migration of the virus out of Central Africa into Cape Verde between the late 1970s and the middle 1980s. Dispersion of the $G_{CV\text{-}PT}$ clade seems to have been shaped by the historical and ongoing human population movements between Angola, Cape Verde and Portugal,. Our data also highlight that once introduced in Portugal, the $G_{CV\text{-}PT}$ was disseminated in the local population and probably recombined with local preexisting subtype B variants, originating the CRF14_BG clade. These findings offer important insights to understanding the origin and current characteristics of the HIV-1 subtype G and CFR14_BG epidemics in Cape Verde and Portugal.

Supporting Information

S1 Fig. Analyses of phylogenetic signal and substitution saturation. (A—E) Likelihood maps of 10,000 random quartets made from every HIV-1 subtype G dataset used in this study as indicated in the figure. The triangles display the distribution (left) and percentage (right) of dots representing the likelihoods of the three possible tree topologies for a group of four sequences (quartets) randomly selected from the dataset. The tree-like, star-like and network-like phylogenetic signals are represented by the dots localized on the vertices, center and on the laterals, respectively. Fully resolved (tree-like) tree topologies ranged from 0.77 (CRF14-like-CV-PT) to 0.93 (G-CA + G-WA + G-CV-PT), thus indicating enough phylogenetic signal for consistent phylogenetic inferences in all datasets. (F—J) Substitution saturation plots of the datasets used in this study as depicted in the figure. The ordinate corresponds to the observed proportion of transitions (*s*, green) and transversions (*v*, blue) while the abscissa refers to the distance calculated using the GTR substitution model. The central lines of each plot correspond to the quadratic nonlinear regressions of the data. CA—Central Africa, WA—West Africa, CV—Cape Verde, PT—Portugal. All analyses indicated an absence of substitution saturation in the data set explored since the plots did not reach an evident plateau nor the transversions outnumbered transitions.
(PDF)

S1 Table. GenBank accession numbers of HIV-1 subtype G *pol* sequences described in Table 1.
(PDF)

S2 Table. Distribution of HIV-1 subtype G *pol* sequences across major regional clades circulating in Central/West-Central Africa (G_{CA}), West Africa ($G_{WA\text{-}I}$ and $G_{WA\text{-}II}$) and Cape Verde/Portugal ($G_{CV\text{-}PT}$). AO/CD/CG: Angola/Democratic Republic of Congo/Republic of Congo. GA/GQ: Gabon/Equatorial Guinea. GH/TG: Ghana/Togo. PT/ES: Portugal/Spain.
(PDF)

S3 Table. HIV-1 subtype G *pol* dataset used for Bayesian phylogeographic analysis. [a]The number of subtype G *pol* fragments recovered from full-length HIV-1 CRF14_BG reference sequences is indicated in parenthesis. DRC: Democratic Republic of Congo.
(PDF)

S4 Table. Best fit demographic model for HIV-1 $G_{CV\text{-}PT}$ clade. Log marginal likelihood (ML) estimates for the logistic (Log), exponential (Expo) and expansion (Expa) growth demographic models obtained using the path sampling (PS) and stepping-stone sampling (SS) methods. The Log Bayes factor (BF) is the difference of the Log ML between of alternative (H1) and null (H0) models (H1/H0). Log BFs > 3 indicates that model H1 is more strongly supported by the data than model H0.
(PDF)

S5 Table. Bayesian estimates of the age and root location of the most recent common ancestor (MRCA) of major HIV-1 subtype G ($G_{CV\text{-}PT}$) and BG (CRF14_BG) clades circulating in Cape Verde and Portugal. [a] substitutions/site/year. CV: Cape Verde. PT: Portugal. TMRCA: time of the most recent common ancestor. PSP: posterior state probability.
(PDF)

S6 Table. Evolutionary and demographic parameters estimated for major HIV-1 subtype G clades circulating in Central/West-Central Africa (G_{CA}), West Africa ($G_{WA\text{-}I}$ and $G_{WA\text{-}II}$), Cuba (G_{CU}) and Cape Verde/Portugal ($G_{CV\text{-}PT}$). [a] Data from Delatorre et al [21]. [b] Data from Delatorre et al [52]. [c] Estimated at this study.
(PDF)

Acknowledgments

We thank Vera Bongertz and the article reviewers for helpful comments.

Author Contributions

Conceived and designed the experiments: GB MGM ED IIMPA. Performed the experiments: IIMPA ED GB. Analyzed the data: GB ED IIMPA MGM MLG. Contributed reagents/materials/analysis tools: IIMPA ED MLG MGM GB. Wrote the paper: GB ED IIMPA MGM MLG.

References

1. Archer J, Robertson DL (2007) Understanding the diversification of HIV-1 groups M and O. Aids 21: 1693–1700. PMID: 17690566
2. Hemelaar J, Gouws E, Ghys PD, Osmanov S (2011) Global trends in molecular epidemiology of HIV-1 during 2000–2007. Aids 25: 679–689. doi: 10.1097/QAD.0b013e328342ff93 PMID: 21297424
3. Oliveira V, Bartolo I, Borrego P, Rocha C, Valadas E, Barreto J, et al. (2012) Genetic diversity and drug resistance profiles in HIV type 1- and HIV type 2-infected patients from Cape Verde Islands. AIDS Res Hum Retroviruses 28: 510–522. doi: 10.1089/AID.2011.0174 PMID: 21902592
4. de Pina-Araujo I, Guimaraes ML, Bello G, Vicente AC, Morgado MG (2014) Profile of the HIV epidemic in Cape Verde: molecular epidemiology and drug resistance mutations among HIV-1 and HIV-2 infected patients from distinct islands of the archipelago. PLOS One 9: e96201. doi: 10.1371/journal.pone.0096201 PMID: 24763617
5. Peeters M, Esu-Williams E, Vergne L, Montavon C, Mulanga-Kabeya C, Harry T, et al. (2000) Predominance of subtype A and G HIV type 1 in Nigeria, with geographical differences in their distribution. AIDS Res Hum Retroviruses 16: 315–325. PMID: 10716369
6. Agwale SM, Zeh C, Robbins KE, Odama L, Saekhou A, Edubio A, et al. (2002) Molecular surveillance of HIV-1 field strains in Nigeria in preparation for vaccine trials. Vaccine 20: 2131–2139. PMID: 11972982
7. Ojesina AI, Sankale JL, Odaibo G, Langevin S, Meloni ST, Sarr AD, et al. (2006) Subtype-specific patterns in HIV Type 1 reverse transcriptase and protease in Oyo State, Nigeria: implications for drug resistance and host response. AIDS Res Hum Retroviruses 22: 770–779. PMID: 16910833
8. Sankale JL, Langevin S, Odaibo G, Meloni ST, Ojesina AI, Olaleye D, et al. (2007) The complexity of circulating HIV type 1 strains in Oyo state, Nigeria. AIDS Res Hum Retroviruses 23: 1020–1025. PMID: 17725419

9. Chaplin B, Eisen G, Idoko J, Onwujekwe D, Idigbe E, Adewole I, et al. (2011) Impact of HIV type 1 subtype on drug resistance mutations in Nigerian patients failing first-line therapy. AIDS Res Hum Retroviruses 27: 71–80. doi: 10.1089/aid.2010.0050 PMID: 20964479
10. Hamers RL, Wallis CL, Kityo C, Siwale M, Mandaliya K, Conradie F, et al. (2011) HIV-1 drug resistance in antiretroviral-naive individuals in sub-Saharan Africa after rollout of antiretroviral therapy: a multicentre observational study. Lancet Infect Dis 11: 750–759. doi: 10.1016/S1473-3099(11)70149-9 PMID: 21802367
11. Ajoge HO, Gordon ML, Ibrahim S, Shittu OS, Ndung'u T, Olonitola SO (2012) Drug resistance pattern of HIV type 1 isolates sampled in 2007 from therapy-naive pregnant women in North-Central Nigeria. AIDS Res Hum Retroviruses 28: 115–118. doi: 10.1089/AID.2011.0115 PMID: 21568761
12. Imade GE, Sagay AS, Chaplin B, Chebu P, Musa J, Okpokwu J, et al. (2014) Short communication: Transmitted HIV drug resistance in antiretroviral-naive pregnant women in north central Nigeria. AIDS Res Hum Retroviruses 30: 127–133. doi: 10.1089/AID.2013.0074 PMID: 24164431
13. Bartolo I, Rocha C, Bartolomeu J, Gama A, Marcelino R, Fonseca M, et al. (2009) Highly divergent subtypes and new recombinant forms prevail in the HIV/AIDS epidemic in Angola: New insights into the origins of the AIDS pandemic. Infect Genet Evol 9: 672–682. doi: 10.1016/j.meegid.2008.05.003 PMID: 18562253
14. Afonso JM, Bello G, Guimaraes ML, Sojka M, Morgado MG (2012) HIV-1 genetic diversity and transmitted drug resistance mutations among patients from the North, Central and South regions of Angola. PLOS ONE 7: e42996. doi: 10.1371/journal.pone.0042996 PMID: 22952625
15. Bartolo I, Zakovic S, Martin F, Palladino C, Carvalho P, Camacho R, et al. (2014) HIV-1 diversity, transmission dynamics and primary drug resistance in Angola. PLOS One 9: e113626. doi: 10.1371/journal.pone.0113626 PMID: 25479241
16. Chamberland A, Diabate S, Sylla M, Anagounou S, Geraldo N, Zannou DM, et al. (2012) Transmission of HIV-1 drug resistance in Benin could jeopardise future treatment options. Sex Transm Infect 88: 179–183. doi: 10.1136/sextrans-2011-050209 PMID: 22158948
17. Mamadou S, Montavon C, Ben A, Djibo A, Rabiou S, Mboup S, et al. (2002) Predominance of CRF02-AG and CRF06-cpx in Niger, West Africa. AIDS Res Hum Retroviruses 18: 723–726. PMID: 12167280
18. Charpentier C, Bellecave P, Cisse M, Mamadou S, Diakite M, Peytavin G, et al. (2011) High prevalence of antiretroviral drug resistance among HIV-1-untreated patients in Guinea-Conakry and in Niger. Antivir Ther 16: 429–433. doi: 10.3851/IMP1754 PMID: 21555827
19. Yaotse DA, Nicole V, Roch NF, Mireille PD, Eric D, Martine P (2009) Genetic characterization of HIV-1 strains in Togo reveals a high genetic complexity and genotypic drug-resistance mutations in ARV naive patients. Infect Genet Evol 9: 646–652. doi: 10.1016/j.meegid.2009.04.002 PMID: 19460333
20. Dagnra AY, Vidal N, Mensah A, Patassi A, Aho K, Salou M, et al. (2011) High prevalence of HIV-1 drug resistance among patients on first-line antiretroviral treatment in Lome, Togo. J Int AIDS Soc 14: 30. doi: 10.1186/1758-2652-14-30 PMID: 21663632
21. Delatorre E, Mir D, Bello G (2014) Spatiotemporal dynamics of the HIV-1 subtype G epidemic in West and Central Africa. PLOS One 9: e98908. doi: 10.1371/journal.pone.0098908 PMID: 24918930
22. Esteves A, Parreira R, Piedade J, Venenno T, Franco M, Germano de Sousa J, et al. (2003) Spreading of HIV-1 subtype G and envB/gagG recombinant strains among injecting drug users in Lisbon, Portugal. AIDS Res Hum Retroviruses 19: 511–517. PMID: 12892060
23. Esteves A, Parreira R, Venenno T, Franco M, Piedade J, Germano de Sousa J, et al. (2002) Molecular epidemiology of HIV type 1 infection in Portugal: high prevalence of non-B subtypes. AIDS Res Hum Retroviruses 18: 313–325. PMID: 11897032
24. Palma AC, Araujo F, Duque V, Borges F, Paixao MT, Camacho R (2007) Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. Infect Genet Evol 7: 391–398. PMID: 17360244
25. Abecasis AB, Martins A, Costa I, Carvalho AP, Diogo I, Gomes P, et al. (2011) Molecular epidemiological analysis of paired pol/env sequences from Portuguese HIV type 1 patients. AIDS Res Hum Retroviruses 27: 803–805. doi: 10.1089/AID.2010.0312 PMID: 21198411
26. Delgado E, Thomson MM, Villahermosa ML, Sierra M, Ocampo A, Miralles C, et al. (2002) Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. J Acquir Immune Defic Syndr 29: 536–543. PMID: 11981372
27. Bartolo I, Abecasis AB, Borrego P, Barroso H, McCutchan F, Gomes P, et al. (2011) Origin and epidemiological history of HIV-1 CRF14_BG. PLOS One 6: e24130. doi: 10.1371/journal.pone.0024130 PMID: 21969855

28. Carling J (2004) Emigration, Return and Development in Cape Verde: The Impact of Closing Borders. Popul Space Place 10: 113–132.
29. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, et al. (2005) An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 21: 3797–3800. PMID: 16076886
30. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. (1999) Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 73: 152–160. PMID: 9847317
31. Strimmer K, von Haeseler A (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci U S A 94: 6815–6819. PMID: 9192648
32. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502–504. PMID: 11934758
33. Neron B, Menager H, Maufrais C, Joly N, Maupetit J, Letort S, et al. (2009) Mobyle: a new full web bioinformatics framework. Bioinformatics 25: 3005–3011. doi: 10.1093/bioinformatics/btp493 PMID: 19689959
34. Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. J Hered 92: 371–373. PMID: 11535656
35. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59: 307–321. doi: 10.1093/sysbio/syq010 PMID: 20525638
36. Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33: W557–559. PMID: 15980534
37. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25: 1253–1256. doi: 10.1093/molbev/msn083 PMID: 18397919
38. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol 55: 539–552. PMID: 16785212
39. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161: 1307–1320. PMID: 12136032
40. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214. PMID: 17996036
41. Suchard MA, Rambaut A (2009) Many-core algorithms for statistical phylogenetics. Bioinformatics 25: 1370–1376. doi: 10.1093/bioinformatics/btp244 PMID: 19369496
42. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLOS Biol 4: e88. PMID: 16683862
43. Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. PLOS Comput Biol 5: e1000520. doi: 10.1371/journal.pcbi.1000520 PMID: 19779555
44. Ferreira MAR, M.A. S (2008) Bayesian analysis of elapsed times in continuous-time Markov chains. Canadian Journal of Statistics 26: 355–368.
45. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22: 1185–1192. PMID: 15703244
46. Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol 18: 1001–1013. PMID: 11371589
47. Rambaut A, Drummond A (2007) Tracer v1.6. Available from http://treebioedacuk/software/tracer/.
48. Rambaut A (2009) FigTree v1.4: Tree Figure Drawing Tool. Available from http://treebioedacuk/software/figtree/.
49. Faria NR, Hodges-Mameletzis I, Silva JC, Rodes B, Erasmus S, Paolucci S, et al. (2012) Phylogeographical footprint of colonial history in the global dispersal of human immunodeficiency virus type 2 group A. J Gen Virol 93: 889–899. doi: 10.1099/vir.0.038638-0 PMID: 22190015
50. Bello G, Afonso JM, Morgado MG (2012) Phylodynamics of HIV-1 subtype F1 in Angola, Brazil and Romania. Infect Genet Evol 12: 1079–1086. doi: 10.1016/j.meegid.2012.03.014 PMID: 22484759
51. Dias P, Machado R, Bento AR, (SEF) SdEeF (2014) Relatório de Imigração, Fronteiras e Asilo—2013 (in Portuguese). Available at: http://sefstatsefpt/Docs/Rifa_2013pdf.
52. Delatorre E, Bello G (2013) Phylodynamics of the HIV-1 epidemic in Cuba. PLOS ONE 8: e72448. doi: 10.1371/journal.pone.0072448 PMID: 24039765