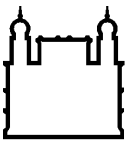
MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
INSTITUTO OSWALDO CRUZ

PROGRAMA DE PÓS-GRADUAÇÃO EM
BIOLOGIA COMPUTACIONAL E SISTEMAS

Comparative Phylodynamics of HIV-1 Subtype B, Subtype C and CRF02_AG Epidemics Across Different Geographical Regions.

Daiana MIR DA SILVA

Rio de Janeiro
March, 2018



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

Author:

Daiana MIR DA SILVA

Comparative Phylodynamics of HIV-1 Subtype B, Subtype C and CRF02_AG Epidemics Across Different Geographical Regions.

Thesis submitted in partial fulfillment of the requirements

for the degree of Doctor in Sciences

Supervisor:

Dr. Gonzalo BELLO BENTANCOR

March, 2018

Mir da Silva, Daiana.

Comparative Phylodynamics of HIV-1 Subtype B, Subtype C and CRF02_AG Epidemics Across Different Geographical Regions. / Daiana Mir da Silva. - Rio de janeiro, 2018.
216 f.; il.

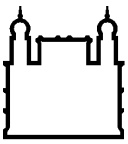
Tese (Doutorado) – Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2018.

Orientador: Gonzalo Bello Bentancor.

Bibliografia: Inclui Bibliografias.

1. Phylodynamics. 2. HIV-1 . I. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da Biblioteca de Manguinhos/ICICT com os dados fornecidos pelo(a) autor(a).



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

Author:

Daiana MIR DA SILVA

Comparative Phylodynamics of HIV-1 Subtype B, Subtype C and CRF02_AG Epidemics Across Different Geographical Regions.

Supervisor:

Dr. Gonzalo BELLO BENTANCOR

Aprovada em: 02/03/2018

EXAMINADORES:

Prof. Dr. Ana Carolina Paulo Vicente - Presidente (IOC/FIOCRUZ)

Prof. Dr. Francisco Inacio Bastos (CICT/FIOCRUZ)

Prof. Dr. André Felipe Andrade dos Santos (UFRJ)

Prof. Dr. Thiago Estevam Parente Martins (IOC/FIOCRUZ)

Prof. Dr. Márcio Galvão Pavan (IOC/FIOCRUZ)

"If you know the question,
you know half."

—Herb Boyer

Abstract

The emergence and reemergence of infectious diseases and the persistent circulation of human pathogens, have stimulated a growing interest and effort to understand the origins and causes of disease outbreaks and dissemination. The field of phylodynamics, has proven to be a proper framework for the integration of epidemiological and evolutionary dynamics, and a key source of information about the spatiotemporal patterns of epidemics and their way of transmission through a susceptible population. This thesis aimed to characterize the demographic dynamics exhibited by some clades of HIV-1 across multiple epidemiological scales in different geographical regions by applying a phylodynamic approach. The specific HIV-1 strains analyzed were: the subtype B, the CRF02_AG and the subtype C. Despite that subtype B is the most widely spread HIV-1 variant in Latin America, the precise number, geographic extension and dissemination dynamics of major subtype B pandemic clades circulating in the region was largely unknown. Among the 90 circulating recombinant forms (CRFs) currently described, CRF02_AG is responsible for the largest number of global infections and the most prevalent HIV variant in West and West-Central Africa. Little was known, however, about the number and demographic patterns of major CRF02_AG lineages in those African regions. In addition to being the most prevalent lineage globally, the HIV-1 subtype C accounts for an important fraction of HIV infections in East Africa and South Brazil, but its evolutionary dynamics in those regions had not been fully analyzed. The main results obtained show that one third of HIV-1 subtype B infections in Latin America originated from the spread of a few B pandemic founder strains probably introduced in the region since the late 1960s onwards. Our analyses also revealed that major West and West-Central African CRF02_AG clades are the result of a few founder strains introduced from Central Africa between late 1960s and mid-1980s. The West African CRF02_AG clade showed a high geographic dissemination and its introduction in Asia and Europe lead to the emergence of local epidemics. The coalescent and the birth-death phylodynamic approaches applied for the analyses of the subtype C lineages circulating in Eastern Africa and Southern Brazil pointed to an initial stage of exponential growth in all them until roughly mid-1990s. Both models are also congruent in a subsequent drop of the epidemic growth of the analyzed C sub-epidemics and their stabilization until the early 2000s. However, very divergent epidemiological patterns were supported by these two phylodynamic approaches for the most recent stages of the C sub-epidemics, underscoring the importance of their joint use in the reconstruction of past population dynamics of HIV epidemics.

Acknowledgements

This thesis is the result of a challenging academic stage. So far I come by a capricious need for answers about the origin and processes of life and paradoxically (although I should have predicted it) the path taken to find absolute answers also led me to find particular answers of great value at a personal level.

Thanks to Dr. Gonzalo Bello, for guiding, encouraging and developing in me a critical and honest scientific look. Their competence, commitment and enthusiasm motivated me to explore and fully exploit my abilities and directed my dedication to the search for the precise answer but above all to the formulation of the correct question. And it was thanks to him, that between good questions and worked answers my PhD flowed and came to fruition giving me the tools to move forward with more strength, more passion and with more conviction than ever that I rightly chose this exciting path to investigate life. I also want to thank Gonza my friend, and creator of unforgettable moments, with whom I had the pleasure of enjoying more than one 'samba no pé ' !

The path for the Doctorate gave me the chance to meet Dr. Edson Delatorre, who honored me daily with his infinitely diverse teachings and in whom I found a noble friend. Thank you, Edson, for your invaluable support throughout the course of my PhD. Together with you and Gonzalo I felt as if I were part of a well tuned symphony orchestra with which I expect many more concerts ahead.

I would also like to thank the entire team at the Laboratório de AIDS e Imunologia Molecular of the Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro - Brazil, with whom I learned about other HIV infection research approaches that enriched my biased Bioinformatics approach.

Thanks also to Dr. Ana Carolina Paulo Vicente from the Laboratório de Genética Molecular de Microorganismos of the Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro - Brazil, who accompanied my entire journey through the postgraduation in Computational Biology and Systems and gave me a space in her own laboratory to do my alchemies.

During the development of this thesis I had the opportunity to work with several collaborators, with whom the different chapters that make it up were born. Thanks to Yuyo (Dr. Héctor Romero) of the Universidad de la República - Uruguay (UdelaR) for his collaboration in Chapter 3. Chapter 4 was possible thanks to Martine Peeters, Matthieu Jung and Nicole Vidal of the Université de Montpellier - France and Chapter 5 saw the light thanks to Tiago Gräf of the Universidade Federal do Rio de Janeiro – Brazil, Sabrina Esteves

de Matos Almeida of the Universidade Federal do Rio Grande do Sul - Brazil and Aguinaldo Roberto Pinto of the Universidade Federal de Santa Catarina - Brazil.

During my stay in Rio de Janeiro, valuable people joined my life ... Thanks to Suwellen and Lili, my very powerful daily colleagues, friends with whom I learned to laugh and cry in portuñol! Mais do que obrigada a Duduca e Zeze pelo seu abrazador carinho e por me revelar deliciosos sabores brasileiros! Agradeço especialmente a Duduca ter compartilhado comigo sua perspicácia e alegria sem igual que me ensinaram o significado de ser uma honorável carioca da gema. Thanks also to Fernanda and Leo, remarkable companions of gatherings, lives and drinks. Eternal thanks to all of them, adding Edson and Gonzalo, for presenting me the seductive, irrational and bohemian Cidade Maravilhosa, its culture and its people. Saudades!

From the Uruguayan crowd thanks to my friends Sabrina, Cecilia, Lau, Pato, Emiliano, Mariana, la Negra, Dario, Alejandra and Maria Laura, who stayed close despite being in different geographic coordinates.

Gracias a mi familia, por su eterno apoyo, paciencia y respaldo a mi carrera y a mi vida. En especial a mis padres, Patricia y Jorge, que me enseñaron a buscar la verdad y el conocimiento siempre y que junto a mi hermana que justamente se llama Victoria, me inspiran una fuerza movilizadora de montañas sin la cual nunca hubiera llegado a ésta meta.

Finally, thanks to Guille, who grabbed the backpack and ventured with me to try life in another point of the map, encouraging me in each challenge, celebrating with me every achievement, reminding me to enjoy the journey and teaching me to be beyond the scientifically verifiable.

Daiana Mir da Silva

March, 2018

Contents

List of Figures

List of Tables

List of Supplementary Figures

List of Supplementary Tables

Abbreviations

AIDS	Acquired immunodeficiency syndrome
aLRT	Approximate Likelihood ratio test
APOBEC	Apolipoprotein B mRNA-editing catalytic polypeptide-like enzyme
ART	Antiretroviral treatment
BDSKY	Birth-death skyline
BSP	Bayesian skyline plot
BSSVS	Bayesian stochastic search variable selection
CRF	Circulating recombinant form
DNA	Deoxyribonucleic acid
DRC	Democratic Republic of the Congo
ESS	Effective sample size
GII	Gender Inequality Index
GTR	General time reversible
HET	Heterosexual
HIV	Human immunodeficiency virus
HPD	Highest Probability Density
HTLV	Human T- cell lymphotropic virus
ICTV	International Committee on Taxonomy of Viruses
IDU	Injection drug user
LAV	Lymphadenopathy associated virus
LTT	Lineages-through-time
MCC	Maximum clade credibility
MCMC	Markov chain Monte Carlo
MEP	Measurable evolving population
ML	Maximum likelihood
MLE	Marginal likelihood estimation
MSM	Men who have sex with men
Ne	Effective population size
PEPFAR	President's Emergency Plan for AIDS Relief
pol	*polymerase*
PS	Path sampling
r	Growth rate
$\mathbf{R_0}$	Basic reproductive number
$\mathbf{R_e}$	Effective reproductive number
RNA	Ribonucleic acid
RT	Reverse transcriptase ou transcriptase reversa
SIR	Susceptible/Infected/Recovered-Removed
SIV	Simian immunodeficiency virus
SS	Stepping-stone sampling
tMRCA	Time of the most recent common ancestor
UNAIDS	Joint United Nations Programme on HIV/AIDS
URF	Unique recombinant form
US	United States
WHO	World health organization

Chapter 1

Introduction

1.1 HIV Discovery and Classification

The Human Immunodeficiency Virus (HIV) belongs to the *Lentivirus* genus of the *Retroviridae* family and it is enough to think about the meaning of the prefixes "Retro" and "Lenti" to extrapolate two of the most important characteristics of this viruses:

- *Retro*: means that they possess a positive sense RNA single-stranded genome, which is back-transcribed in an intermediate form of double-stranded DNA by an enzyme called reverse transcriptase.
- *Lenti*: comes from slow and points to a very important evolutionary strategy of the *Lentivirus* genus, which is the long interval between the initial infection and the appearance of severe symptoms (incubation period).

The genus *Lentivirus* is classified into seven major groups according to their hosts: lagomorphs, small ruminants, cattle, horses, cats, anthropoids and prosimians [1] (Fig. 1.1). The *Lentivirus* found in anthropoids are the HIV and the Simian Immunodeficiency Virus (SIV). The HIV is the cause of one of the largest and most devastating pandemics in human history, being on par with the 1918 Spanish flu as the world's leading infectious cause of

adult death since the bubonic plague of the 14th century [2]. The natural history of the HIV infection is characterized by an initial acute phase with very high circulating levels of virus and a rapid decline in CD4+ T-cells. Despite a strong immune response, the host is not capable of clearing the virus and allows HIV to establish a chronic infection that after some years of asymptomatic course evolved to the Acquired Immunodeficiency Syndrome (AIDS) phase [3].

The first cases of AIDS were described in the United States (US) in 1981 and corresponded to homosexual males affected by infrequent diseases (until that moment) such as *Pneumocystis* pneumonia caused by *Pneumocystis jirovecii* (formerly known as *Pneumocystis carinii*) and Kaposi's sarcoma [4], [5]. Quickly, however, similar cases were reported in other social groups including injecting drug users (IDUs) [6], hemophiliacs [7], recipients of blood transfusions [8], children [9], female sex partners of infected men [10], and prisoners [11], providing evidence that the transmission of this new syndrome occur by exposure to organic fluids of infected individuals via parenteral, sexual or vertical. Two years later, at the Pasteur Institute in Paris, the research group led by Luc Montagnier detected reverse transcriptase activity in culture supernatant of lymph node cells from a patient with AIDS, indicating the presence of a retrovirus, which they called Lymphadenopathy Associated Virus (LAV) [12]. In 1984, four papers of the scientific team led by Robert Gallo of the U.S. National Institute of Health reported the isolation of a new retrovirus found in peripheral mononuclear cells as the probable cause of AIDS and denominate it as Human T-Lymphotropic Virus Type III (HTLV III) [13]–[16]. Subsequently, these two viruses (LAV and HTLV III) were defined as variants of the same retrovirus, etiologic agent of AIDS [17] and the International Committee on Taxonomy of Viruses (ICTV) recommended the name of HIV in 1986 [18]. In the same

year, Montagnier's team isolated a variant of HIV from an AIDS patient from West Africa [19], which led to a modification of the nomenclature becoming as HIV type 1 (HIV-1) the first variant discovered and as HIV type 2 (HIV-2) the subsequently isolated variant.

1.2 The current state of the HIV epidemic

By 2016, after more than three decades of the discovery of the HIV, it was estimated that existed 36.7 million (30.8 - 42.9 million) people infected with that virus globally. In the same year, 1 million (830.000 - 1.2 million) people died from AIDS-related causes worldwide and approximately 5.000 new HIV infections occurred daily around the world [20]. If trends over time are analyzed, the number of newly infected individuals in 2016 was 18% lower than in 2010 and the HIV incidence was brought down substantially in both the overall population (11%) and children (47%) (Fig. 1.2).

The sharp decline in the HIV incidence in children was fully consistent with the increase in the percentage of pregnant women (29%) with access to antiretroviral therapy (ART) in the same period [20]. However, even though new HIV infections have declined, the number of people living with HIV continues to increase, in large part because more people globally are accessing ART and are living longer that at initial phase of the pandemic [21], [22]. As of June 2017, 20.9 million (18.4 - 21.7 million) people living with HIV were accessing ART, which is one of the greatest achievements in the history of global health [21].

International and regional HIV prevalence and incidence estimates, however, should be interpreted with caution, since they reflect average numbers and can mask infection clusters among severe hard-to-reach populations and significant unmet needs [23], [24]. The last Global AIDS data update from

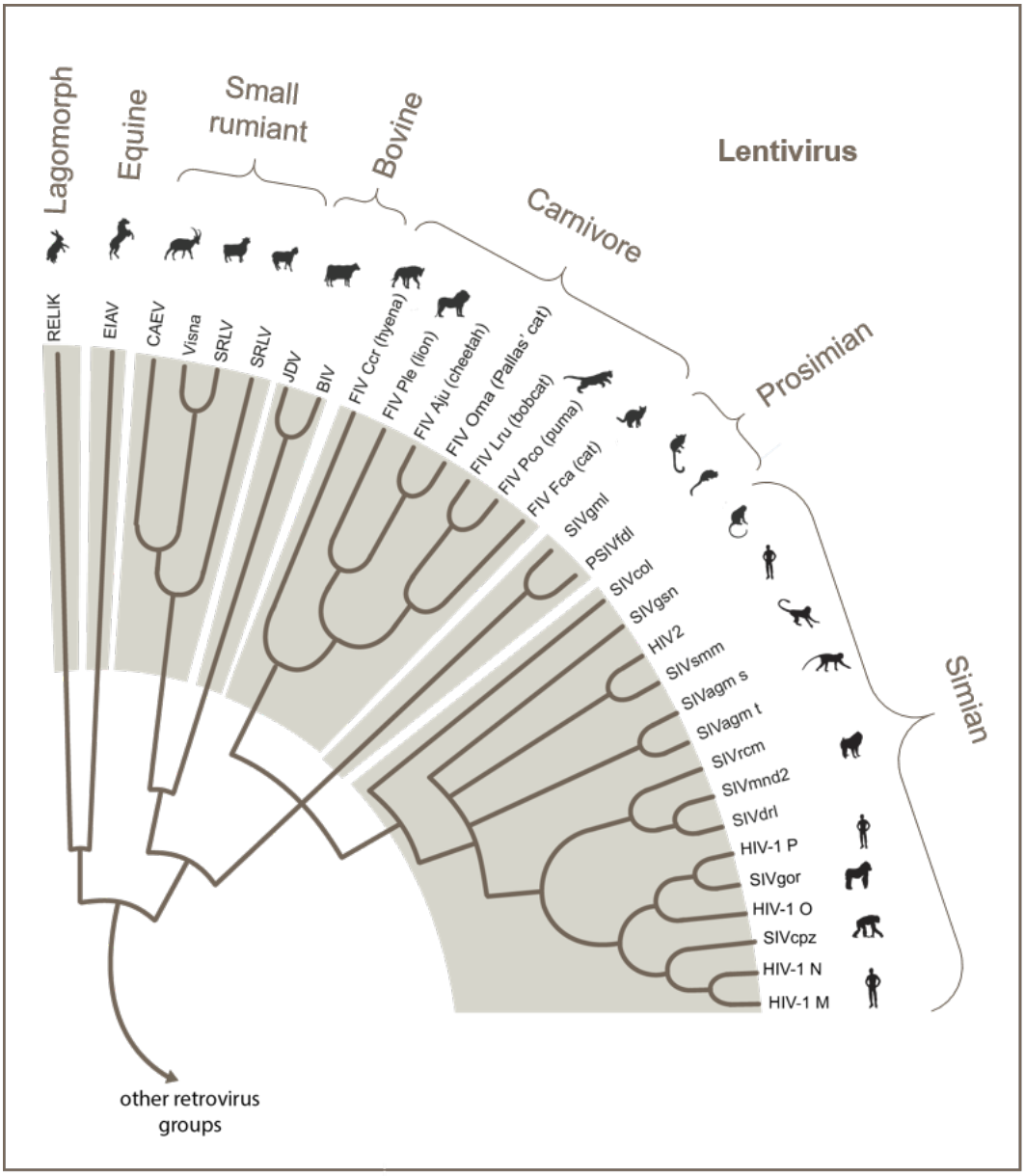


Figure 1.1: Phylogenetic tree summarizing the evolutionary relationships among known *Lentiviruses*. The tree shows branching relationships within and between the seven major *Lentivirus* lineages so far described (lagomorph, equine, small ruminant, bovine, feline, prosimian and simian).Virus names: BIV, bovine immunodeficiency virus; CAEV, caprine arthritis encephalitis virus; EIAV, equine infectious anemia virus; FIV, feline immunodeficiency virus; HIV, human immunodeficiency virus; JDV, jembrana disease virus; PSIV, prosimian immunodeficiency virus; RELIK, rabbit endogenous lentivirus K; SRLV, small ruminant lentivirus. Adapted from: Gifford et al. 2012

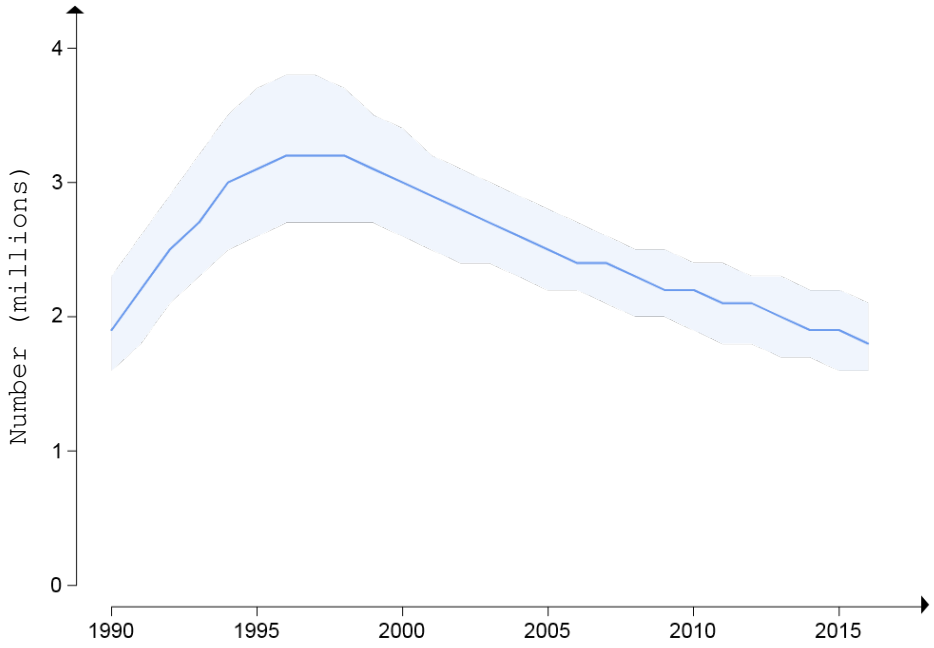


Figure 1.2: Global new HIV infections (all ages). Data Source: UNAIDS Estimates 2017 http://aidsinfo.unaids.org/

the Joint United Nations Programme on HIV/AIDS (UNAIDS) [25], draws attention about the alarming rise in new HIV infections in eastern Europe and central Asia and the discouraging proportion of people knowing their HIV status, initiating ART, and reaching virologic suppression among key populations in India and in the Russian Federation. Such data, coupled with the potential 17% slash in the US President's Emergency Plan for AIDS Relief (PEPFAR) announced in September 2017 [26], are shrinking the optimism to achieve the UNAIDS 90–90–90 HIV treatment targets by 2020 [27] and the ambitious Sustainable Development Goal of ending the AIDS epidemic by 2030 [28].

1.2.1 Zooming into regional HIV epidemics

Sub-Saharan Africa is home of 13% of the world's population and it has been the most heavily affected region throughout the history of the HIV epidemic; currently hosting the 70% of all people living with HIV in the world (∼26

million) (Fig. 1.3). The HIV prevalence varies considerably within the African continent and is disproportionately elevated in the Eastern and Southern regions. Together, these two African regions account for about 53% (∼19 million) of all people living with HIV in the world and harbor a mean HIV prevalence of 7% among adults aged 15–49, reaching extremely high values in Swaziland (27%), Lesotho (25%), Botswana (22%), South Africa (19%), Zimbabwe (13%) and Zambia (12%) [22]. In most Western, Central and Northern African countries, by contrast, the HIV prevalence remains low (<2%) [22]. The observed differences in the HIV epidemic among African regions have been associated to the complex interplay between biological, behavioural and structural (cultural and social) factors [29]. Heterosexual (HET) intercourse is the most common mode of HIV transmission in sub-Saharan Africa (>80% of infections), generally associated with a high Gender Inequality Index (GII) [30]. In sub-Saharan Africa, adolescent girls and young women experience the most elevated HIV risk and vulnerability linked primarily to gender-based violence including sexual abuse, lack of access to education and health services, and infection rates among young women are twice as high as among young men [31].

An estimated 2 million (1.7-2.5 million) people in **Latin America** and the **Caribbean** were living with HIV in 2016 (Fig. 1.3). While the number of new HIV infections among the general population for both sub-regions remained roughly stable between 2010 and 2015, it was observed a rise by 3% among adults (15-49 years age) and a sharp decline by more than 50% among children in the same period. An estimated 32% of adults living with HIV in Latin America and 50% of adults in the Caribbean are women [22].

The **Caribbean** has the second highest HIV prevalence in adults (1.3%) after sub-Saharan Africa. Haiti accounts for 48% of all people living with HIV in the region but the highest prevalence in the Caribbean region is

found in the Bahamas (3.3%) [22]. In this region, the primary mode of HIV transmission is HET contact (79%), followed by men who have sex with other men (MSM, 12%) [32]. There was great progress in the Caribbean related to the elimination of mother-to-child transmission which plummeted from an estimated 2.300 (1.600–3.000) in 2010 to 400 (200–700) in 2015. Among low and middle-income countries, Cuba became the first board-certified country in the world to have eliminated mother-to-child HIV and syphilis transmission by 2015 [33].

Latin America harbors a regional HIV prevalence of 0.5% among the general adult population. The epidemic in the region predominantly affects MSM, transgender women, female sex workers and IDUs [22], [34]. Brazil accounts for 46% of all HIV infections in Latin America, with an estimate of about 830.000 people living with HIV in 2016, followed by Argentina, Venezuela and Colombia that together host 20% of all HIV infections in the region ($\sim$120.000 people living with HIV by 2016) [22]. Latin America has one of the highest HIV treatment coverage in the world, with around 55% of people living with HIV receiving antiretroviral therapy by 2015 [35]. However, the number of new HIV infections has risen above 20% between 2010 and 2015 among the adult populations from some countries including Belize, Nicaragua, and Guatemala. Increases, albeit smaller, were also reported for the same time period in Mexico and Panama (8%), Chile (6%), Colombia (5%) and Brazil (4%) [35].

In 2016, $\sim$160.500 people were newly diagnosed with HIV in the **WHO European Region**, the majority (80%) were from the eastern part, 17% from the western part and 4% from the central part. Those amount of new diagnoses represent the highest number of people ever newly diagnosed in one year since HIV case reporting began in the 1980s [36]. Russia and Ukraine were responsible for 73% of the newly diagnosed infections

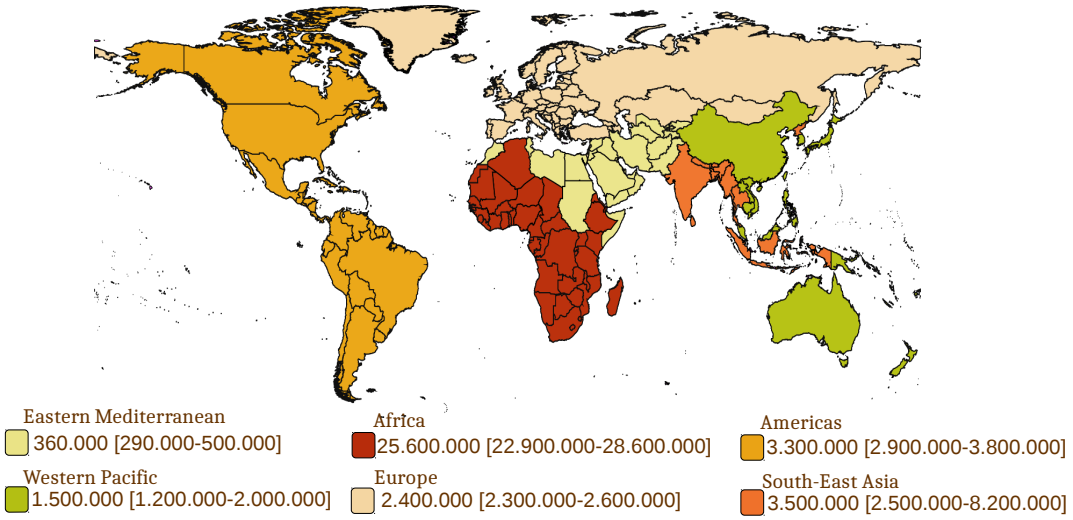


Figure 1.3: Estimated number of people living with HIV, 2016. Data Source : World Health Organization http://gamapserver.who.int/mapLibrary/app/searchResults.aspx

in the whole WHO European Region and of 92% of those in the Eastern European sub-region. Data on transmission mode indicate that among those new diagnoses, HET transmission accounted for 52%, IDUs for 32%, MSM for 15% and mother-to-child transmission for 0.8% [36]. It has been observed an increasing trend between 2006-2015 of new HIV diagnoses among natives (originating from the reporting country) and European migrants (originating from another European country) and a decline among non-European migrants, mainly among people from sub-Saharan Africa [37].

The countries of Central Asia (Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan and Uzbekistan) are currently experiencing a rapidly expanding HIV epidemic largely fueled by IDUs [38]. While HIV surveillance among MSM is subject to great bias due to high levels of stigma and criminalization of homosexuality, studies suggest that there is a growing epidemic in this group [39]. Turkmenistan is a special case, as no data related to HIV infection has been reported to the international community and the government denies that there is any problem [38]. Similar prevalence estimates of HIV infection among the adult populations in Kazakhstan,

Tajikistan and Kyrgyzstan (0.2-0.3%) were reported in 2016 [22]. There is no updated HIV data for Uzbekistan but its known that this country harbor the largest epidemic in Central Asia [40]. Coverage of ART in Central Asia remained significantly lower than the global average, holding in conjunction with Eastern Europe only 21% of people living with HIV on treatment [35].

1.3 Origin of HIV

While there are two types of HIV (HIV-1 and HIV-2), many different strains of SIVs are found in more than 40 species of non-human primates in sub-Saharan Africa [41], [42]. Several studies indicate that due to the close phylogenetic relationship, the similarities in organization of the viral genome, and the geographic coincidence and plausible routes of transmission [43]–[45], HIV strains originated as a result of at least 13 independent zoonotic transmissions of SIV to humans (Fig. 1.4) [46]. Four of these zoonotic transmissions originated groups M, N, O and P of the HIV-1. Phylogenetic analyses demonstrated that groups M and N of the HIV-1 are closely related to isolates of SIV infecting chimpanzees of the *Pan troglodytes troglodytes* species (SIVcpz-Ptt) from southeastern and southern-central Cameroon [45], [47]; while groups O and P are more closely related to SIV that infects gorillas of the specie *Gorilla gorilla* (SIVgor) distributed in western and southern-central Cameroon [48], [49]. SIVcpz-Ptt virus in turn, is a recombinant of the SIVs naturally infecting two different monkeys, the red-capped mangabey (*Cercocebus torquatus*) and the greater spot-nosed monkey (*Cercopithecus nictitans*) [50] and SIVgor strains were derived from a single lineage within the SIVcpz radiation [49], [51] (Fig. 1.4). While groups N, O and P remain mainly restricted to Central Africa [52]–[54], the M group was able to spread worldwide being responsible for

>99% of global HIV-1 infections [55]. Using dated molecular phylogenies, it was estimated that the M and O groups have been originated between 1909 and 1930 [56]–[58], while the N and P groups seem to have emerged more recently [46].

The remaining zoonotic events resulted in the nine different groups of HIV-2 (A-I) that probably originated from the transmission of SIV from sooty mangabeys monkeys (SIVsmm) to humans in West Africa [59]–[61]. Of the nine HIV-2 groups, only groups A and B have become widespread in humans and the time of their most recent common ancestor (tMRCA) was dated to 1940-1945 [60]. The lack of sequences of SIVsmm of defined geographic origin makes it difficult to determine the exact origin of HIV-2 strains. However, phylogenetic analyses suggest that HIV-2 groups E and F viruses are closely related to SIVsmm strains from Sierra Leone, group D strain is closely related to the Liberian SIVsmmLIB1 strain [43] and groups A, B, C, G, and H are more closely related to SIVsmm strains from Cote d'Ivoire [61].

1.4 The pandemic spread of the HIV-1 group M

The most accepted hypothesis about how zoonotic transmission events happened, argues that infection of humans by SIV probably occurred in successive exposure events to nonhuman primate blood as a result of hunting, butchering, or other activities such as consumption of uncooked contaminated meat [62]. The opportunities for such cross-species transmissions have endured due to the fact that this activities have been taking place since past times and continue to occur in present days [63], [64]. Indeed, it is likely that other zoonotic transmission events in addition to those already identified occurred and continue happening in isolated rural communities [65]. In this context, it is of utmost importance to analyze the

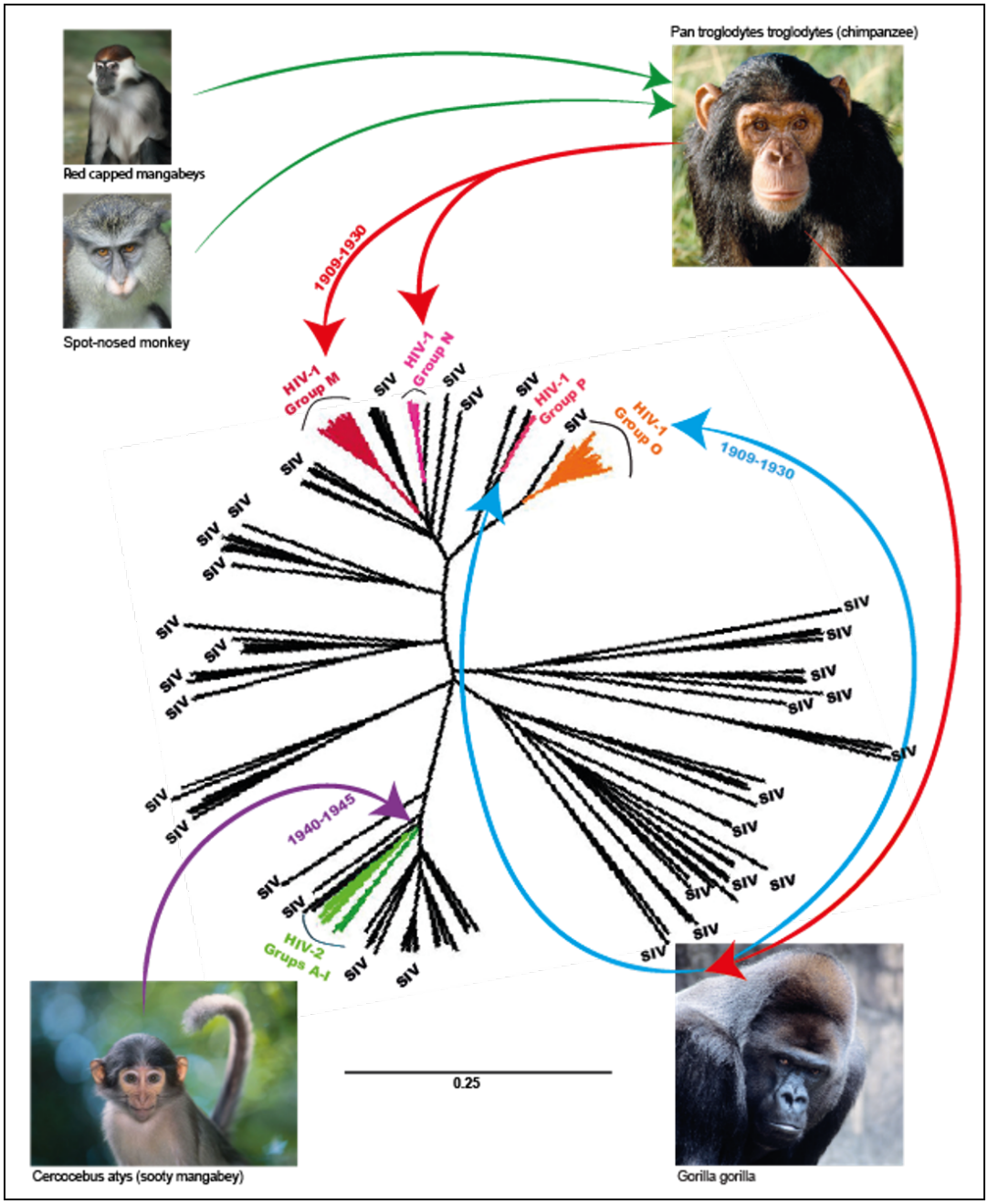


Figure 1.4: Patterns of cross-species transmission of SIV to humans. Clades in color represent viruses identified in humans after their transfer from chimpanzee, gorilla, and sooty mangabey reservoirs. Adapted from: Foley et al. 2016; Tebbit et al. 2010

reasons why among the 13 independent transmissions of SIVs to humans known until now, the HIV-1 group M was the only one who had the ability or opportunity to spread in the human population at a pandemic scale.

To account for the establishment of group M as an epidemic in the early part of the 20th century in Central Africa, a combination of social, economic, and behavioral changes has been proposed. The initial source of HIV-1 group M epidemic dispersion after the transmission of SIVcpz-Ptt from chimpanzees to humans in southeast Cameroon, was recently traced to Kinshasa, capital of the Democratic Republic of the Congo (DRC) [58]. It was there, where human interventions such as colonization, urbanization, prostitution and extensive re-use of needles and syringes in medical practice and campaigns against endemic tropical diseases took place and were the likely cause of the early establishment, dissemination and epidemic growth of the HIV-1 group M [66]. After accumulating a substantial diversity while confined for several decades in Central Africa [57], the chance exportation of different group M strains to other geographical regions since the 1960s onwards, led to the differential distribution of viral variants reported today around the world [56], [67], [68].

1.5 Generation of HIV genetic diversity

Several diversity-generating mechanisms make the RNA viruses in general and the HIV in particular one of the fastest evolving organisms studied so far [69], [70]. Its rapid evolutionary rate provide HIV an effective mechanism to evade immune pressure and drug therapy and make difficult to produce an effective vaccine. The fast evolutionary dynamics exhibited by HIV-1 reflect the low fidelity of its replicative process. The HIV-1 mutation rate, which encompass the combined error rate of the viral reverse transcriptase

(RT) and human RNA polimerase II, is approximately 2 x 10^{-5} per nucleotide per replication cycle (0.2 error per genome per replication cycle) [71]. This high mutation rate coupled with the fast viral turnover, fuel the HIV rapid evolutionary change and enable on average a unique genome in each generated virion [72].

Recombination also has a central role in generating HIV diversity. It happens at all phylogenetic levels (between and within HIV-1 groups) and is boosted by the diploid RNA genome that allow the RT to "jump" from one strand to the other during the replication process. Therefore, cells infected with two different HIV-1 strains might produce heterozygous virions with the ability to infect new cells and generate recombinant proviral DNA that will generate a progeny of recombinant viruses. It has been determined a minimum estimated rate of 2.8 crossovers events/genome/replication-cycle, which is a very high rate given the small size of the genome ($\sim$10 kb) [73]. Finally, the HIV replication process can also be affected by cellular factors, in particular, the cytidine deaminase activity of the host-encoded Apolipoprotein B mRNA-editing catalytic polypeptide-like enzymes (APOBECs) responsible for G-to-A hypermutations [74]–[76] that elevates the HIV-1 mutation rate by 40-fold above the RT/RNA polimerase II error rate [77].

1.6 Global distribution of HIV-1 group M strains

The extensive HIV-1 genetic variation coupled with historical and ecological factors such as founder effects and global travels, eventually resulted in the extreme diversification observed within the HIV-1 M group [65], [66], which is currently phylogenetically divided into ten subtypes (A, B, C, D, F, G, H, J, K and L; denomination that does not follow an alphabetical order because it was observed that subtypes E and I were not real subtypes

but recombinat forms) [78]–[80], two sets of sub-subtypes (A1-A4, F1-F2) and various unique and circulating recombinant forms (URFs and CRFs, respectively) that comprise segments of different subtypes. While the HIV-1 intra-subtype genetic variation was found to be around 8.2% (5.3-10.0%), the inter-subtype divergence was estimated near 14.7% (12.2-15.8%) depending on the gene compared [81]. For the description of a new CRF it is necessary to identify three or more epidemiologically unrelated individuals infected by a virus with the same recombinant points across the entire genome [82]. By December 2017, 90 different HIV-1 CRFs had been described (Los Alamos National Laboratory, http://www.hiv.lanl.gov).

There is a great variance in the distribution of HIV-1 M subtypes and CRFs from one region of the globe to another (Fig. 1.5). On a global scale, the subtype C is the most abundant strain, responsible for approximately half (48%) of all HIV-1 infections worldwide. Subtypes A and B caused 12% and 11% of infections respectively, followed by CRF02_AG (8%), CRF01_AE (5%), subtype G (5%) and D (2%). Subtypes F, H, J and K together caused < 1% of infections worldwide while others CRFs and URFs are each responsible for $\sim$ 4% of global infections [83]. The greatest genetic diversity of HIV-1 group M has been found in Central Africa, which is consistent with the notion that this region is the likely epicenter of the group M pandemic [84].

1.6.1 Origin and dissemination of HIV-1 subtype B

Despite not being the most globally prevalent HIV-1 variant, subtype B (HIV-1B) was the first HIV lineage to infect individuals outside the African continent and is currently the most widely spread, being found at high prevalence in the Americas, Western and Central Europe, Australia, several Asian countries (Hong Kong, Japan, Korea and Taiwan), Northern Africa and the Middle East [83]. The hypothesis with greater scientific consistency to

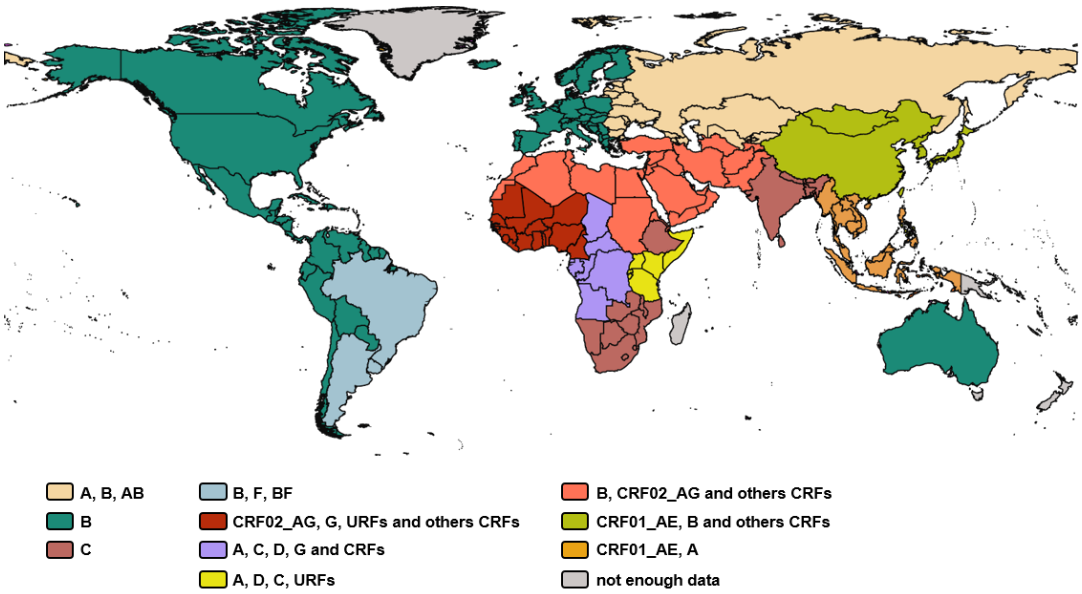


Figure 1.5: Global distribution of HIV-1 M group subtypes and CRFs. Data Source : Hemelaar et al. 2011; Taylor & Hammer 2008

date, suggests that HIV-1B was introduced from the DRC to the American continent through Haiti around the mid-1960s (1962–1970) [85]. Such date suggests that its arrival in Haiti may have occurred with the massive return of Haitian to their country of origin after years of work in the former Belgian Congo (current DRC) [66], [86] (Fig. 1.6). From Haiti, several subtype B strains were disseminated to neighboring Caribbean and Latin American countries from the early 1970s onwards. The widely dissemination in the Caribbean region of HIV-1B strains from the epidemic established in Haiti, allowed the establishment of effective local epidemics and the emergence of the so-called subtype 'B Caribbean' (B_{CAR}) clades [87]. This was not the case for the Latin American countries, where despite that some HIV-1B strains directly introduced from the Caribbean achieved locally spread, overall, these lineages disseminate with low prevalence in this region [88]. Of note, due to the non-pandemic circulation of the B_{CAR} clades as well as of their descendant strains disseminated through Latin America, they have been jointly identified as 'subtype B non-pandemic' ($B_{NON\text{-}PANDEMIC}$) clades [87], [88]. Another subtype B strain directly introduced from Haiti to the US

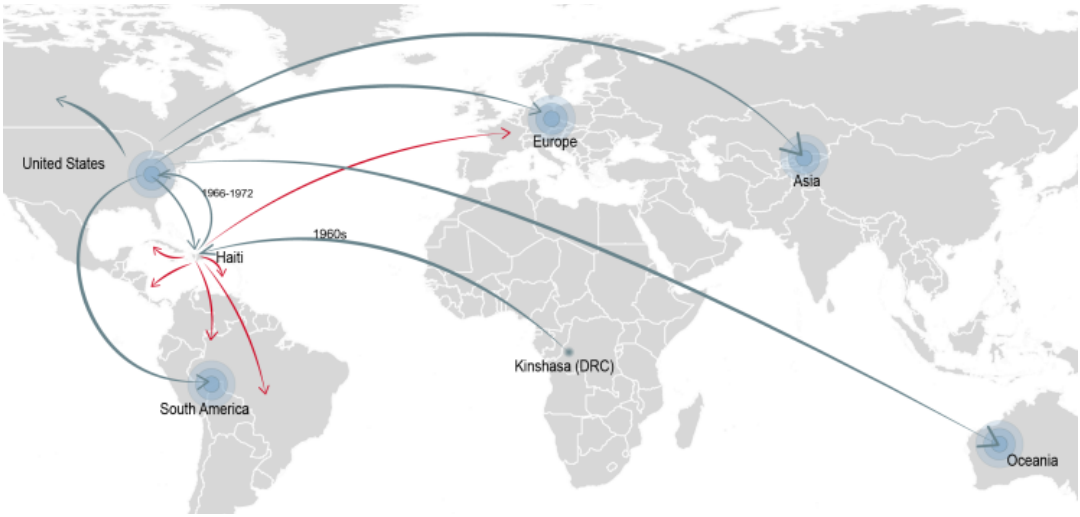


Figure 1.6: Spatial dynamics of HIV-1 Subtype B around the world. Blue arrows represent the main dissemination routes of HIV-1$B_{PANDEMIC}$ form while blue circles indicate its main disease epicenters. Red lines represent the dissemination of the $B_{NON\text{-}PANDEMIC}$ lineage.

around 1969 (1966–1972) was a major epidemiological event in the genesis of the HIV-1 M group pandemic, given its subsequent spread throughout the globe, defining the 'subtype B pandemic' ($B_{PANDEMIC}$) clade [85], [89].

Archived US HIV sequences, have recently allowed to trace New York City (NYC) as the source location of early US HIV-1B diversification and date by 1972 (1970–1974) its MRCA [90]. The inferred time of the US HIV-1B ancestor, uncovers a time-gap of ∼10 years of cryptic circulation of the virus in the US until the initial recognition of the HIV/AIDS in 1981. Among the several cases of sexual transmission reported during the early 1980s in the US, some of them corresponded to HET transmissions, however, the HIV epidemic was by far primarily disseminated through MSM transmission networks. Indeed, it has been suggested that the HIV-1B introduction into US from Haiti, would have occurred as a result of sex tourism practiced on the island, which by that time was a prime destination especially for male homosexuals from NYC [66]. The intense international connection and large-scale migration existing in the US, surely had important implications in the spread of HIV-1B beyond this country limits and fostered the emergence of the $B_{PANDEMIC}$ clade.

Despite the successful dissemination of the B_{CAR} clades in several American countries including: Haiti and the Dominican Republic (75% of HIV-1B infections), Jamaica (∼50% of HIV-1B infections), Trinidad and Tobago (∼95% of HIV-1B infections), other Lesser Antilles (∼40–75% of HIV-1B infections), French Guiana (56% of HIV-1B infections) and Suriname (54% of HIV-1B infections), the HIV-1B epidemic in Caribbean countries such as Cuba and Puerto Rico is primarily driven by the $B_{PANDEMIC}$ clade (∼96-99% of HIV-1B infections) which was back-introduced from the US [87], [88]. In most Latin American countries, the dominance of the $B_{PANDEMIC}$ clade is overwhelming and account for >90% of the current HIV-1B infections [88]. In addition, a single introduction of the $B_{PANDEMIC}$ clade, has been indicated as the reponsible for the majority of the current HIV-1B cases (62%) in Central America [91].

Subtype B dominates the HIV-1 epidemic in Western and Central Europe [83], [92], where it was introduced with few exceptions by multiple founder events, some of which achieved the establishment of local transmission networks [93] mainly at the country level [94]. Notwithstanding, phylogeographic analyses suggest a high level of geographical mixing of the subtype B strains across Europe which highlight the importance of human mobility across international borders in the spread of this clade [99]. HIV-1 molecular epidemiology in Europe is highly stratified according to gender and risk group. It is so that subtype B in this continent was significantly more common in men than in women and in the MSM population than in IDUs or HET individuals [98]. Early epidemiological data on the most probable modes of acquisition of HIV in Europe reported that the majority of the infected MSMs had travelled to the US, where they experienced high risk sexual relations [95], reinforcing thus, the role of the US as crucial hub of early dissemination of the $B_{PANDEMIC}$ clade.

Despite the epidemiological importance of the HIV-1B in Asia (prevalence ∼25%) [83], its spatiotemporal dissemination patterns through the overall population are currently not well known. However, some works on the characterization of the HIV-1B lineages associated with transmission among MSMs in East Asia, revealed that approximately one-third of the Japan subtype B strains belong to a large monophyletic $B_{PANDEMIC}$ cluster that has been widely disseminated into nearby regions of Asia (China and Taiwan) and to the western hemisphere (the US and European countries), most likely through global MSM networks [96], [97].

1.6.2 Origin and dissemination of HIV-1 CRF02_AG

CRFs and URFs contributed substantially to the worldwide epidemic, with prevalence ranging from 10% to 50% depending of the region under analysis [98]. The most prevalent CRF worldwide among the 90 CRFs currently described is the CRF02_AG, responsible for at least 8% of total HIV-1 infections [83]. The epidemic emergence of the CRF02_AG clade has been dated back to 1973 (1972-1975) in the DRC [99]. From DRC, the earliest migratory events of this lineage were traced to Cameroon and Gabon (Fig. 1.7). Given that this migratory trends did not reflect sustained human migratory flows, it has been suggested that the early dissemination of the CRF02_AG clade probably occurred as a consequence of chance exportation [99]. Currently, there is a low prevalence of the CRF02_AG in its native African region (8%) [83], [100], which has been explained as a reflection of the poor accessibility (i.e. large estimated travel time) among the major population centers in Central Africa [101], [102]. A growing trend of the CRF02_AG prevalence is discernible toward West and West-Central Africa, where it stands for about 50% of the HIV-1 infections [83], [103]. Notably, Cameroon has been proposed as the hub

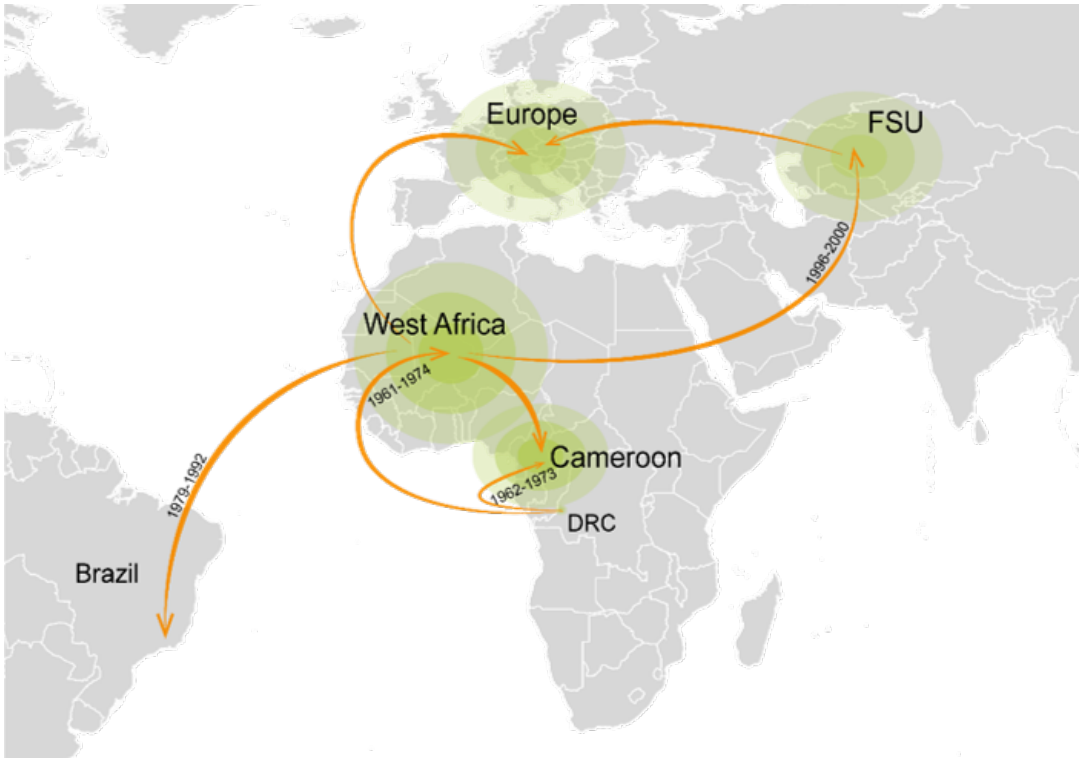


Figure 1.7: Spatial dynamics of HIV-1 CRF02_AG around the world. Yellow arrows represent the main dissemination routes of HIV-1 CRF02_AG lineages while green circles indicate its main disease epicenters.

of the CRF02_AG dispersal and where this lineage reaches its highest proportion of infections [104], [105].

The presence of the CRF02_AG lineage outside Africa is increasingly common, particularly in former Soviet Union (FSU) countries where it is mainly associated with IDUs networks [106]–[111]. Large increasing proportions of CRF02_AG have been also reported among newly diagnosed patients living in Europe [112]. While the presence of HIV-1 non-B subtypes in Europe is commonly attributed to migration of individuals from non-European countries, there is increasing evidence of indigenous transmission networks of CRF02_AG in this continent. Particularly in France, Belgium and Italy, recent analysis evidentiate the existence of some native CRF02_AG transmission clusters particularly among highly connected networks of MSM [113]–[116].

In America, particularly in Brazil, at least five introduction events of the

CRF02_AG lineage have been reported, of which at least two have been disseminated locally in the state of Rio de Janeiro since 1985 (1979–1992) [117]. These CRF02_AG strains circulating in Brazil are closely related to those circulating in western African countries where this CRF is highly prevalent [117], [118] which demonstrates the potential for larger-scale dissemination of this HIV-1 variant.

1.6.3 Origin and dissemination of HIV-1 subtype C

Subtype C predominates widely in Southern Africa [119], Burundi, Djibouti and Ethiopia [120]–[123], and also reaches high prevalence in India and Southern Brazil [83], [124]. Little is known about the potential factors that might explain why the HIV-1 subtype C (HIV-1C) is the most prevalent lineage worldwide. Some recent studies support that the successful dissemination of subtype C could be linked to virological factors that resulted in a lower replicative fitness and virulence, that therefore resulted in a higher transmissibility, in comparison with subtypes A and D [125]–[127].

The origin of HIV-1C was traced to the Katanga region of the DRC in the late 1930s [58] from where it spread independently to Eastern and Southern Africa probably through migrant labor, leading to a phylogeographic subdivision between the HIV-1C strains circulating in those two African regions [128], [129] (Fig. 1.8). The subtype C epidemic in the Southern African region present a strong panmixia pattern, which suggests multiple introductions of C strains into the countries of the region. Molecular clock analyses place the origin of this Southern African epidemic between 1956 and 1964 and the phylodynamic reconstruction revealed a strong early stage of exponential epidemic growth during the 1970s and 1980s that coincide with a period of socio-political changes in that region [130]. In turn, phylogenetic analysis indicate that most HIV-1C strains (70%) from east

Figure 1.8: Spatial dynamics of HIV-1 Subtype C around the world. Pink arrows represent the main dissemination routes of HIV-1C while purple circles indicate its main disease epicenters.

Africa belong to a single regional monophyletic clade, the so-called 'C_{EA}' [131]. The C_{EA} clade was probably originated in Burundi around the early 1960s from where it was subsequently disseminated through the Eastern African region and ignited several local C_{EA} sub-epidemics between the early 1970s and early 1980s [131].

Interestingly, the epidemic of subtype C in Brazil was probably initiated by the introduction into the Paraná state of a single founder C_{EA} strain originated in Burundi [132]–[134]. Estimates of the time scale of this founding event suggest that it took place in the mid-1970s (1974-1976), coinciding with the origin of the African C_{EA} sub-epidemics. The states of Paraná and Santa Catarina have been the most important hubs of HIV-1C dissemination through Brazil [133]. The HIV-1C dissemination occurred rapidly through the southern Brazilian states where this lineage accounts for 20-80% of the HIV-1 infections but its dispersion to other Brazilian regions has been much slower reaching a prevalence $\leq$ 10% in the southeast and central-west

regions [133], [135]. Notably, phylogenetic analysis revealed a clade of HIV-1C sequences from MSMs of United Kingdom (UK) origin that branch within the HIV-1C Brazilian radiation, supporting a scenario of direct viral flow from Brazil to the UK [134], [136]. Recent phylogeographic analysis, found the occurrence of multiple introductions of HIV-1C in Cuba, but the successfull establishment and dissemination of only two of them [131]. While the major of the two characterized HIV-1C Cuban clades probably belongs from the C_{EA} clade, the minor one, by contrast, probably derives from Southern African HIV-1C lineages.

In Asia, the highest number of HIV-1C infections were diagnosed in India. Phylogenetic analysis showed that the majority of Indian HIV-1C sequences belong to a single monophyletic clade with close ancestral linkage to South Africa. The tMRCA of the predominant Indian HIV-1C clade has been dated back to the early 1970s (1965–1982) [137], [138]. There is also evidence supporting the Indian origin of the HIV-1C epidemic in China and the existence of an 'Indian-Chinese' cluster whose emergence is likely the result of transmigration of people between China and India via the Burmese heroin trafficking route across the northeast of India [137], [139].

1.7 Phylodynamics

The area of study that allows the estimation of epidemiological and evolutionary parameters through detailed phylogenetic analysis is known as phylodynamics [140]. The central tenet of the phylodynamic approach is that in rapidly evolving pathogens, population genetic processes occur in a similar timescale that epidemiological and ecological processes [141]. Due to overlapping scales of both types of processes, epidemiological and ecological

drivers that shape genetic patterns can be recovered using phylogenetic techniques coupled with population and epidemiological dynamic models.

1.7.1 Molecular Clock hypothesis and Neutral theory

The concept of a molecular clock is essential to the phylodynamic analyses by enabling the estimates of divergence dates and hence, dating epidemic origins and providing a timescale to the population dynamics and the processes that shape viral genetic diversity. The molecular clock hypothesis proposed in the 1960s [142], [143] postulates that DNA and protein sequences evolve at a constant probability rate over time and between different organisms. Notably, the molecular clock as originally proposed by Pauling and Zuckerkandl was stochastic rather than metronomic and held as corollary that the genetic distance between two species is proportional to the time elapsed since the species diverged from their MRCA, thus allowing the estimation of evolutionary time scales from genetic data (Fig. 1.9).

The hypothesis of a molecular clock was controversial with the Darwinian evolution since there was no concrete evidence underpinning the uniform rate of accumulation of adaptive variations. The neutral theory of molecular evolution proposed by Motoo Kimura in 1968 [144], provided a big support to the molecular clock hypothesis proposing that the vast majority of substitutions in DNA and proteins are neutral and not positively selected. Under the neutral theory, the fate of each neutral mutation (fixation in the population or lost) is determined stochastically by genetic drift. In this way, the neutral theory predicts that the molecular evolution has a roughly constant 'ticking' rate when the proportion of mutations that are neutral is constant over time, thus providing a rational basis for the molecular clock hypothesis.

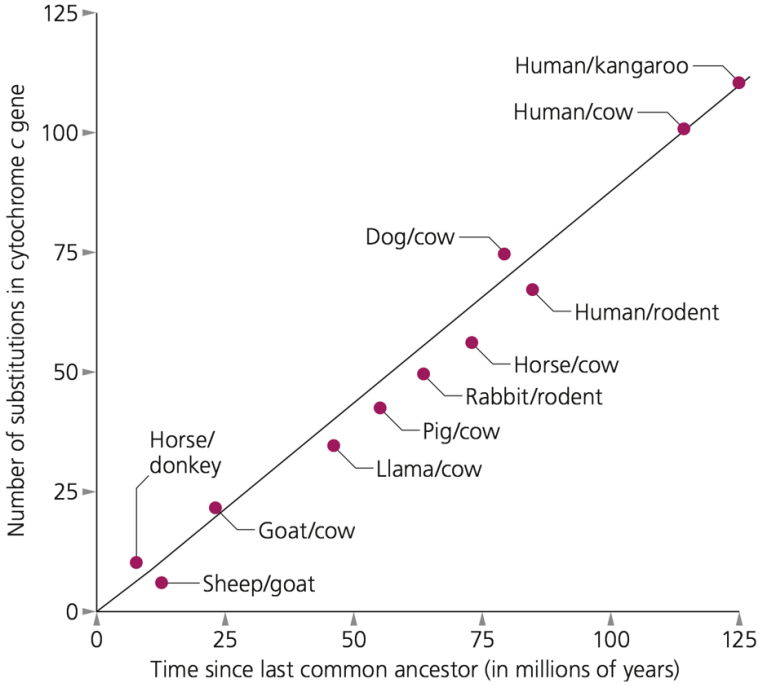


Figure 1.9: Molecular Clock hypothesis: Rates of amino acid changes in cytochrome c. Observations about amino acid changes that occurred during the divergence between species, show that molecular evolution takes place at an approximately constant rate.

A persistent problem for the molecular clock hypothesis is the existence of departures from the clock-like evolutionary behavior and variations in the rate of evolution of the same gene between lineages [145]–[148]. Indeed, there is substantial evidence that the ideal clock-like assumption is rejected at a general level for RNA viruses and particularly for HIV [149]–[151]. To contemplate this non-uninform evolutionary behavior, alternative models that relax the assumption of a 'strict' molecular clock have been described. Among the 'relaxed' approaches are those which estimate a separate molecular rate for groups of branches in the tree (local molecular clock models) [152] and the ones that model the rates per branch in an autocorrelated (as function of the rate of the parent branch) [153], [154] or uncorrelated [155] way from a parametric distribution.

1.7.2 Calibrating the clock

To provide measurements on an absolute timescale, rather than in relative ages, the molecular clock needs to be calibrated. This can be done by incorporating information about the absolute rate or by assigning or constraining the ages of one or more nodes in the phylogeny which can then act as time reference for the whole tree. The age information of the nodes can come from various sources such as fossil record, geological evidence or historical migration events [156]. In measurably evolving populations (MEPs), the use of sampling times (age of external nodes) turns out to be an alternative calibration approach. MEPs populations are defined as those populations for which a statistically significant number of accumulated genetic differences within the sampling temporal range exist. Following this line of reasoning, MEPs are characterized by either a high mutation rate (i.e. RNA virus), or a wide range of sequence sampling times (i.e. ancient DNA from fossil record) [157]. Dated molecular phylogenies of pathogen MEPs in conjunction with population and/or epidemiological models can be used to estimate population dynamics and timing epidemiological events.

1.7.3 Coalescent and Birth-Death models

Two types of models are currently widely used to estimate key population (e.g. growth rate [r]) and epidemiological (e.g. transmission rate [λ] and basic reproductive number [R_0]) parameters from pathogen sequence data within the phylodynamic framework: the coalescent [158]–[160] and the birth-death process [161], [162].

The coalescent theory originally formulated by Kingman [158]–[160] provides a conceptual framework for the analysis of processes that historically influenced populations and that led to the current distribution

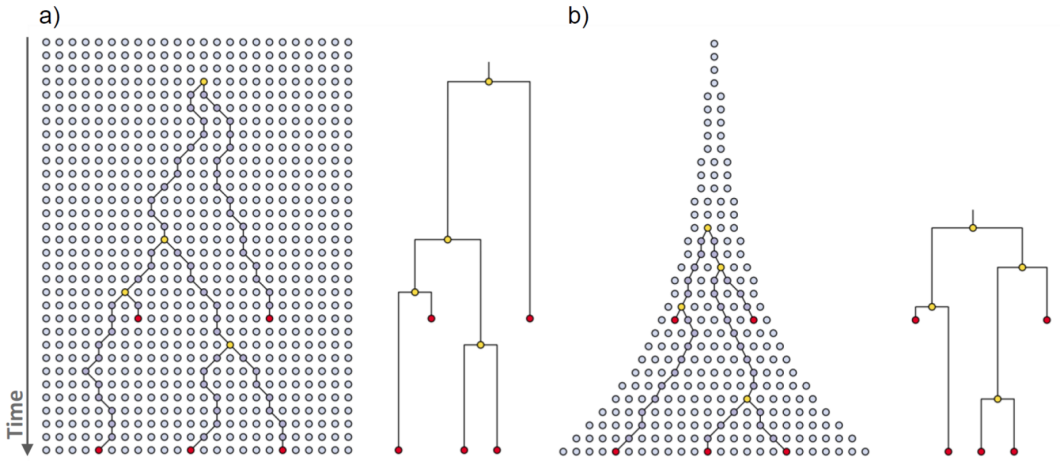


Figure 1.10: Schematic drawing representing the coalescence process in two populations. The red dots represent individuals sampled at two time points (heterochronic), the yellow dots represent the common ancestor between them and the gray dots represent individuals of the non-sampled population. The two schemes represent the phylogenetic relationships between individuals sampled from a population of constant size (a) and exponential growth (b). Moving back in time from the present, we follow the number of lineages in the genealogy in each generation. This value decreases when two lineages share a common ancestor (a coalescence event), and increases when sampled individuals are encountered (a sampling event). The probability that a coalescence event occurs at a particular time is inversely proportional to the population size at that time. This relationship is reflected in the size of the branches of the phylogenetic tree and can be used to estimate the demographic history of the population. Adapted from Drummond et al. 2003.

of its genetic variability. Through this population genetic model, the probable relations among sampled individuals from a population of interest are characterized retrospectively until their MRCA by reconstructing the probabilistic structure of the underlying tree [163]. (Fig. 1.10). The inference of coalescent times is made progressively backward in time, as a function of the effective population size (*Ne*) [163]. Therefore, any population process that affects the relatedness of randomly selected individuals will also affects the distribution of its coalescent times, enabling the inference of the shape and magnitude of the perturbation [164]. Within a viral disease context, the viral population size equals the number of infected individuals and coalescence events occur when the same donor infects two individuals. In essence, the method assumes that the estimated phylogeny from viral sequences accurately reflects the underlying transmission tree [165].

In its seminal paper [159], Kingman presents the statistical description of the genealogies expected from n individuals sampled at random from the present generation of a Fisher-Wright population (i.e. panmictic population not subjected to natural selection or recombination events and with non-overlapping generations) of size N, with $n \ll N$. Although in its most basic formulation the coalescent theory relies on a Fisher-Wright population model [166], [167], it was extended to include changes in the *Ne* [168], [169], compartmentalization [170]–[172], recombination [173]–[175], selection [176]–[178], and sampling at different time-points ('serial-sampling') [179]. Through these extensions, coalescent-based approaches assuming deterministic and non-parametric population size changes over time have been extensively used to shed light on historical population dynamics [180]–[182].

The use of deterministic demographic functions (e.g. constant, exponential, logistic or expansional) within the coalescent approach, allows to obtain the growth rate (r), with which it is possible to estimate the basic reproductive number through the expression: $R_0 = rD + 1$ (where D is the expected infectivity time). R_0 is an important epidemiological parameter, defined as the number of secondary infections produced by an infected individual in a population that is totally susceptible, and it is used as a key indicator of the increase ($R_0 > 1$), decline ($R_0 < 1$) or stabilization ($R_0 = 1$) in the number of new cases when their values are higher, less than or equal to 1 respectively (transmission potential).

Some shortcomings of the coalescent as a model of epidemiological transmission include: 1) it is appropriate only if the number of sampled infected individuals is small compared with the size of the total infected population ($n \ll N$); 2) to estimate R_0 under this approach it is necessary an independent estimate of the average duration of infectiousness [183], [184];

and 3) both r and R_0 inferences are only possible through modeling the population dynamics under a deterministic assumption.

More generally, temporal variation in the transmission potential of infectious diseases is monitored via the effective reproduction number, denoted by R_e, which is defined as equivalent to R_0 when the population is no longer fully susceptible. It is precisely the R_e, what can be estimated using the birth-death process. Formulated under the infectious disease frame, the birth-death process provides a description of the transmission tree of an infectious disease for which n individuals were sampled. However, unlike the coalescent, this model yields a forward in time description of the epidemiological process, from a single infected individual that starts an epidemic at time t, and as a function of the sampling (δ), birth (λ), and the death (μ) rates. The λ rate, is defined as the rate with which one infected individual will infect an uninfected individual (transmission rate), and the μ rate, is the rate with which an infected individual becomes non-infectious. The non-infectious state may be caused by a number of factors such as death, behavior change, or successful treatment. Therefore, if the birth–death process is used as a model for the spread of an infectious disease for which n individuals were sampled with rate δ, the task is to recover the process that led to those n individuals, namely the transmission tree (sampled tree) (Fig. 1.11) in parallel with the λ, μ and δ rates estimation [185]–[188].

As the coalescent, the basic birth-death model have been subject to a number of extensions among which are those taking into account: a fixed time since the MRCA [189], a fixed number of extant individuals [190], an incomplete sampling of extant individuals [191], [192], an heterochronus sampling [193], or the fluctuation of the λ, μ and δ rates over different equidistant time intervals [188].

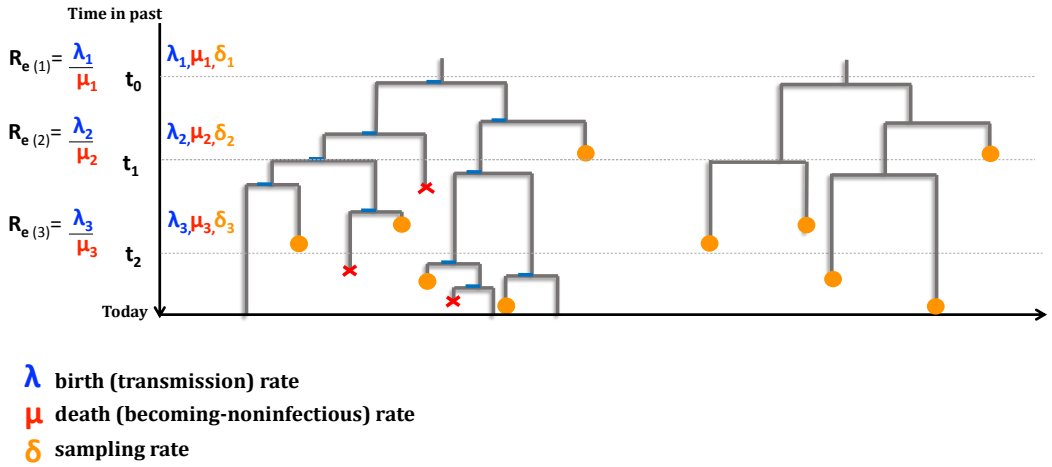


Figure 1.11: Schematic drawing representing the birth–death process (Left) Tree that evolved under the birth–death model with two rate shifts (at time t1 and t2). The sampled individuals are marked with an orange circle, an infection event is displayed as a split of one lineage into two lineages (blue lines) and becoming non-infectious is shown as a lineage that stops growing (red cross). (Right) Corresponding transmission tree, i.e., all lineages with no sampled descendants are pruned. Adapted from : Stadler et al. 2013

It is precisely through the estimation of the λ, μ and δ rates that best fit the topology of the transmission tree in different intervals of the epidemic process, that it is possible to make the inference of the R_e in a piece-wise manner over different time intervals as the ratio of the λ and μ rates: $R_e = \lambda/\mu$.

Some major advantage of the birth-death model over the coalescent include: 1) under the birth–death process the population size can exhibit fluctuations as a result of stochasticity whereas under the parametric coalescent the population size change deterministically; 2) the birth-death process model do not assume a small random sample from a large background population; and 3) the birth-date process has the ability to recover birth and death rates separately, and thus provide an estimate for the R_e directly from sequence data [194].

1.8 References

1. R. J. Gifford, "Viral evolution in deep time: lentiviruses and mammals.," Trends Genet., vol. 28, no. 2, pp. 89–100, Feb. 2012.

2. P. Hunter, "Inevitable or avoidable? Despite the lessons of history, the world is not yet ready to face the next great plague.," EMBO Rep., vol. 8, no. 6, pp. 531–4, Jun. 2007.

3. D. C. Douek, "Disrupting T-cell homeostasis: how HIV-1 infection causes disease.," AIDS Rev., vol. 5, no. 3, pp. 172–7, Jan. 2003.

4. M. S. Gottlieb et al., "Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency.," N. Engl. J. Med., vol. 305, no. 24, pp. 1425–31, Dec. 1981.

5. K. B. Hymes et al., "Kaposi's sarcoma in homosexual men-a report of eight cases.," Lancet (London, England), vol. 2, no. 8247, pp. 598–600, Sep. 1981.

6. V. Quagliarello, "The Acquired Immunodeficiency Syndrome: current status.," Yale J. Biol. Med., vol. 55, no. 5–6, pp. 443–52, Jan. .

7. Centers for Disease Control and Prevention (CDC), "Epidemiologic Notes and Reports Pneumocystis carinii Pneumonia among Persons with Hemophilia A." Online. Available: http://www.cdc.gov/mmwr/preview/mmwrhtml/00001126.htm Accessed: 29-Apr-2016.

8. Centers for Disease Control and Prevention (CDC), "Epidemiologic Notes and Reports Possible Transfusion-Associated Acquired Immune Deficiency Syndrome (AIDS) – California." Online. Available: http://www.cdc.gov/mmwr/preview/mmwrhtml/00001203.htm. Accessed: 29-Apr-2016.

9. Centers for Disease Control and Prevention (CDC), "Unexplained Immunodeficiency and Opportunistic Infections in Infants – New York, New Jersey, California." Online. Available: http://www.cdc.gov/mmwr/preview/mmwrhtml/00001208.htm. Accessed: 29-Apr-2016.

10. Centers for Disease Control and Prevention (CDC), "Epidemiologic Notes and Reports Immunodeficiency among Female Sexual Partners of Males with Acquired Immune Deficiency Syndrome (AIDS) – New York." Online. Available: http://www.cdc.gov/mmwr/preview/mmwrhtml/00001221.htm. Accessed: 29-Apr-2016.

11. Centers for Disease Control and Prevention (CDC), "Acquired Immune Deficiency Syndrome (AIDS) in Prison Inmates – New York, New Jersey," 1983.

12. E. Al Barre-Sinoussi, F., J. C. Chermann, "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).," Science (80-.)., vol. 220(4599), pp. 868–71, 1983.

13. R. C. Gallo et al., "Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS.," Science, vol. 224, no. 4648, pp. 500–3, May 1984.

14. M. Popovic, M. G. Sarngadharan, E. Read, and R. C. Gallo, "Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS.," Science, vol. 224, no. 4648, pp. 497–500, May 1984.

15. M. G. Sarngadharan, M. Popovic, L. Bruch, J. Schüpbach, and R. C. Gallo, "Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS.," Science, vol. 224, no. 4648, pp. 506–8, May 1984.

16. J. Schüpbach, M. Popovic, R. V Gilden, M. A. Gonda, M. G. Sarngadharan, and R. C. Gallo, "Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS.," Science, vol. 224, no. 4648, pp. 503–5, May 1984.

17. L. Ratner, R. C. Gallo, and F. Wong-Staal, "HTLV-III, LAV, ARV are variants of same AIDS virus," Nature, vol. 313, no. 6004, pp. 636–637, Feb. 1985.

18. J. Coffin et al., "Human immunodeficiency viruses.," Science, vol. 232, no. 4751, p. 697, May 1986.

19. F. Clavel et al., "Isolation of a new human retrovirus from West African patients with AIDS," Science (80-.)., vol. 233, no. 4761, pp. 343–346, Jul. 1986.

20. UNAIDS, "Core Epidemiology Slides," 2017.

21. UNAIDS, "UNAIDS fact sheet: Latest statistics on the status of the AIDS epidemic," 2017.

22. United Nations Joint Program on HIV/AIDS (UNAIDS), "AIDSinfo." Online. Available: http://aidsinfo.unaids.org/. Accessed: 01-Dec-2017.

23. P. Piot and P. Aggleton, "The global epidemic," AIDS Care, vol. 10, no. October 2014, p. S201–S201 1p, 1998.

24. UNAIDS, "Local epidemics issues brief," 2014.

25. Unaids, "Ending AIDS - report graphics slides," 2017.

26. J. Ottenhoff et al., "Red Ribbon or White Flag?," 2017.

27. UNAIDS, "90-90-90 An ambitious treatment target to help end the AIDS epidemic," http://www.unaids.org/sites/default/files/media_asset/90-90-90_en.pdf, p. 40, 2014.

28. United Nations, "Transforming our world: the 2030 Agenda for Sustainable Development." Online. Available: https://sustainabledevelopment.un.org/post2015/transformingourworld. Accessed: 07-Apr-2016.

29. A. Buvé, K. Bishikwabo-Nsarhaza, and G. Mutangadura, "The spread and effect of HIV-1 infection in sub-Saharan Africa," Lancet, vol. 359, no. 9322, pp. 2011–2017, 2002.

30. E. T. Richardson et al., "Gender inequality and HIV transmission: a global analysis," J Int AIDS Soc, vol. 17, p. 19035, 2014.

31. UNAIDS, "Prevention Gap Report 2016," 2016.

32. M. E. Sutherland, "The Caribbean's HIV/AIDS Epidemic: Theory, Research and Interventions.," J. HIV/AIDS Infect. Dis., 2014.

33. 'Cuba Wins W.H.O. Certification It Ended Mother-to-Child H.I.V. Transmission,' The New York Times, 2015.

34. UNAIDS, "How AIDS changed everything," 2014.

35. UNAIDS, "Prevention Gap Report 2016," 2016.

36. European Centre for Disease Prevention and Control/WHO Regional Office for Europe, "HIV/AIDS surveillance in Europe 2017 – 2016 data.," Stockholm, 2017.

37. European Centre for Disease Prevention and Control/WHO Regional Office for Europe, "HIV/AIDS Surveillance in Europe 2015," Stockholm, 2016.

38. S. Ancker and B. Rechel, "Policy responses to HIV/AIDS in Central Asia," Glob. Public Health, vol. 10, no. 7, pp. 817–833, 2015.

39. A. Latypov, T. Rhodes, and L. Reynolds, "Prohibition, stigma and violence against men who have sex with men: Effects on HIV in Central Asia," Centr. Asian Surv., vol. 32, no. 1, pp. 52–65, 2013.

40. A. Smolak, N. El-Bassel, A. Malin, A. Terlikbayeva, and S. Samatova, "Sex workers, condoms, and mobility among men in Uzbekistan: implications for HIV transmission," Int. J. STD AIDS, vol. 27, no. 4, pp. 268–272, 2016.

41. M. Peeters, V. Courgnaud, and B. Abela, "Genetic diversity of lentiviruses in non-human primates," AIDS Rev., vol. 3, no. 1, pp. 3–10, 2001.

42. M. Peeters and V. Courgnaud, "Overview of Primate Lentiviruses and Their Evolution in Non-human Primates in Africa," HIV Seq. Compend. 2002, no. 43, pp. 2–23, 2002.

43. Z. Chen et al., "Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey t," J. Virol., vol. 70, no. 6, pp. 3617–27, Jun. 1996.

44. S. Corbet et al., "env sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency

virus group N from the same geographic area.," J. Virol., vol. 74, no. 1, pp. 529–34, Jan. 2000.

45. F. Gao et al., "Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes," Nature, vol. 397, no. 6718, pp. 436–441, Feb. 1999.

46. P. M. Sharp and B. H. Hahn, "Origins of HIV and the AIDS epidemic," Cold Spring Harb. Perspect. Med, pp. 1–22, 2011.

47. B. F. Keele, "Chimpanzee Reservoirs of Pandemic and Nonpandemic HIV-1," Science (80-.)., vol. 313, no. 5786, pp. 523–526, 2006.

48. J.-C. Plantier et al., "A new human immunodeficiency virus derived from gorillas.," Nat. Med., vol. 15, no. 8, pp. 871–872, 2009.

49. M. D'arc et al., "Origin of the HIV-1 group O epidemic in western lowland gorillas," Proc. Natl. Acad. Sci., p. 201502022, 2015.

50. P. M. Sharp, G. M. Shaw, and B. H. Hahn, "Simian immunodeficiency virus infection of chimpanzees.," J. Virol., vol. 79, no. 7, pp. 3891–902, Apr. 2005.

51. J. Takehisa et al., "Origin and biology of simian immunodeficiency virus in wild-living western gorillas.," J. Virol., vol. 83, no. 4, pp. 1635–48, Feb. 2009.

52. M. Peeters et al., "Geographical distribution of HIV-1 group O viruses in Africa.," AIDS, vol. 11, no. 4, pp. 493–8, Mar. 1997.

53. A. Vallari et al., "Four New HIV-1 Group N Isolates from Cameroon: Prevalence Continues to Be Low," AIDS Res. Hum. Retroviruses, vol. 26, no. 1, pp. 109–115, Jan. 2010.

54. A. Vallari et al., "Confirmation of putative HIV-1 group P in Cameroon.," J. Virol., vol. 85, no. 3, pp. 1403–7, Feb. 2011.

55. "Main Search Interface of HIV Sequence Database." .

56. B. Korber et al., "Timing the ancestor of the HIV-1 pandemic strains.," Science, vol. 288, no. 5472, pp. 1789–1796, 2000.

57. M. Worobey et al., "Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960.," Nature, vol. 455, no. 7213, pp. 661–4, Oct. 2008.

58. N. R. Faria et al., "The early spread and epidemic ignition of HIV-1 in human populations.," Science, vol. 346, no. 6205, pp. 56–61, 2014.

59. J. D. Reeves and R. W. Doms, "Human immunodeficiency virus type 2," J. Gen. Virol., vol. 83, no. 2002, pp. 1253–1265, 2002.

60. P. Lemey, O. G. Pybus, B. Wang, N. K. Saksena, M. Salemi, and A.-M. Vandamme, "Tracing the origin and history of the HIV-2 epidemic.," Proc. Natl. Acad. Sci., vol. 100, no. 11, pp. 6588–6592, 2003.

61. M. L. Santiago et al., "Simian Immunodeficiency virus infection in free-ranging Sooty Mangabeys (Cercocebus atys atys) from the Taı Forest, Cote d'Ivoire: implications for the origin of epidemic Human Immunodeficiency virus Type 2," Society, vol. 79, no. 19, pp. 12515–12527, 2005.

62. B. H. Hahn, G. M. Shaw, K. M. De Cock, and P. M. Sharp, "AIDS as a Zoonosis: Scientic and Public Health Implications," Science (80-.)., vol. 287, no. 5453, pp. 607–614, 2000.

63. J. G. Robinson, K. H. Redford, and E. L. Bennett, "Conservation - Wildlife harvest in logged tropical forests," Science (80-.)., vol. 284, no. April, pp. 595–596, 1999.

64. A. Ayouba et al., "Evidence for continuing cross-species transmission of SIVsmm to humans: characterization of a new HIV-2 lineage in rural Cote d'Ivoire.," AIDS, vol. 27, no. 15, pp. 2488–2491, 2013.

65. A. Rambaut, D. Posada, K. a Crandall, and E. C. Holmes, "The causes and consequences of HIV evolution.," Nat. Rev. Genet., vol. 5, no. 1, pp. 52–61, Jan. 2004.

66. J. Pepin, The origins of AIDS. 2011.

67. L. Buonaguro, M. L. Tornesello, and F. M. Buonaguro, 'Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic:

pathogenetic and therapeutic implications.,' J. Virol., vol. 81, no. 19, pp. 10209–10219, 2007.

68. A. Rambaut, D. L. Robertson, O. G. Pybus, M. Peeters, and E. C. Holmes, "Phylogeny and the origin of HIV-1," Nature, vol. 410, no. April, pp. 1047–1048, 2001.

69. S. Duffy, L. A. Shackelton, and E. C. Holmes, "Rates of evolutionary change in viruses: patterns and determinants.," Nat. Rev. Genet., vol. 9, no. 4, pp. 267–76, 2008.

70. A. B. Abecasis, A.-M. Vandamme, and P. Lemey, "Quantifying Differences in the Tempo of Human Immunodeficiency Virus Type 1 Subtype Evolution," J. Virol., vol. 83, no. 24, pp. 12917–12924, 2009.

71. W.-S. Hu and S. H. Hughes, "HIV-1 Reverse Transcription," Cold Spring Harb. Perspect. Med., vol. 2, no. 10, pp. a006882–a006882, Oct. 2012.

72. F. Gao, Y. Chen, D. N. Levy, J. A. Conway, T. B. Kepler, and H. Hui, "Unselected mutations in the human immunodeficiency virus type 1 genome are mostly nonsynonymous and often deleterious," J. Virol., vol. 78, no. 5, pp. 2426–2433, Mar. 2004.

73. J. Zhuang et al., "Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots.," J. Virol., vol. 76, no. 22, pp. 11273–82, 2002.

74. B. Mangeat, P. Turelli, G. Caron, M. Friedli, L. Perrin, and D. Trono, "Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts," Nature, vol. 424, no. 6944, pp. 99–103, Jul. 2003.

75. R. S. Harris et al., "DNA Deamination Mediates Innate Immunity to Retroviral Infection," Cell, vol. 113, no. 6, pp. 803–809, Jun. 2003.

76. K. N. Bishop, R. K. Holmes, A. M. Sheehy, and M. H. Malim, "APOBEC-mediated editing of viral RNA.," Science, vol. 305, no. 5684, p. 645, Jul. 2004.

77. R. Sanjuán and P. Domingo-Calap, "Mechanisms of viral mutation," Cell. Mol. Life Sci., vol. 73, no. 23, pp. 4433–4448, Dec. 2016.

78. I. F. Butler, I. Pandrea, P. A. Marx, and C. Apetrei, "HIV Genetic Diversity: Biological and Public Health Consequences," Curr. HIV Res., vol. 5, no. 985, pp. 23–45, 2007.

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

80. J. L. K. Mokili et al., "Identification of a novel clade of human immunodeficiency virus type 1 in Democratic Republic of Congo.," AIDS Res. Hum. Retroviruses, vol. 18, no. 11, pp. 817–23, Jul. 2002.

81. G. Li et al., "An integrated map of HIV genome-wide variation from a population perspective.," Retrovirology, vol. 12, no. 1, p. 18, 2015.

82. S. Vuilleumier and S. Bonhoeffer, "Contribution of recombination to the evolutionary history of HIV," Curr. Opin. HIV AIDS, vol. 10, no. 2, pp. 84–89, 2015.

83. J. Hemelaar, E. Gouws, P. D. Ghys, and S. Osmanov, "Global trends in molecular epidemiology of HIV-1 during 2000-2007.," AIDS, vol. 25, no. 5, pp. 679–689, 2011.

84. N. Vidal et al., "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., vol. 74, no. 22, pp. 10498–10507, 2000.

85. M. T. P. Gilbert, A. Rambaut, G. Wlasiuk, T. J. Spira, A. E. Pitchenik, and M. Worobey, "The emergence of HIV/AIDS in the Americas and beyond.," Proc. Natl. Acad. Sci. U. S. A., vol. 104, no. 47, pp. 18566–18570, 2007.

86. P. Piot et al., "Acquired Immunodeficiency Syndrome in a Heterosexual Population in Zaire," Lancet, vol. 324, no. 8394, pp. 65–69, 1984.

87. M. Cabello, Y. Mendoza, and G. Bello, “Spatiotemporal Dynamics of Dissemination of Non-Pandemic HIV-1 Subtype B Clades in the Caribbean Region.,” PLoS One, vol. 9, no. 8, p. e106045, 2014.

88. M. Cabello, D. M. Junqueira, and G. Bello, “Dissemination of nonpandemic Caribbean HIV-1 subtype B clades in Latin America.,” AIDS, vol. 29, no. 4, pp. 483–92, 2015.

89. K. E. Robbins et al., “U.S. Human Immunodeficiency Virus Type 1 Epidemic: Date of Origin, Population History, and Characterization of Early Strains,” J. Virol., vol. 77, no. 11, pp. 6359–6366, 2003.

90. M. Worobey et al., “1970s and ‘Patient 0’ HIV-1 genomes illuminate early HIV/AIDS history in North America,” Nature, 2016.

91. W. Murillo et al., “A Single Early Introduction of HIV-1 Subtype B into Central America Accounts for Most Current Cases.”

92. A. B. Abecasis et al., “HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics.,” Retrovirology, vol. 10, p. 7, Jan. 2013.

93. D. Paraskevis et al., “Tracing the HIV-1 subtype B mobility in Europe: A phylogeographic approach,” Retrovirology, vol. 6, no. 1, p. 49, May 2009.

94. D. Frentz et al., “Limited cross-border infections in patients newly diagnosed with HIV in Europe,” Retrovirology, vol. 10, no. 1, p. 36, Apr. 2013.

95. M. P. Glauser, M. Infeetieuses, and C. Hospitatier, “Clinical and Epidemiological Survey of Acquired Immune Deficiency Syndrome in Europe,” Eur J Clin Microbiol, vol. 3, no. 1, pp. 55–58, 1984.

96. Y. Takebe et al., “Intercontinental Dispersal of HIV-1 Subtype B Associated with Transmission among Men Who Have Sex with Men in Japan on behalf of the UK Collaborative Group on HIV Drug Resistance,” J. Virol., vol. 88, no. 17, 2014.

97. M. Kondo et al., "Emergence in Japan of an HIV-1 variant associated with MSM transmission in China: First indication for the international dissemination of the Chinese MSM lineage.," J. Virol, vol. 87, no. 10, pp. 5351–5361, 2013.

98. F. E. Mccutchan, "Global Epidemiology of HIV," J. Med. Virol., 2006.

99. N. R. Faria et al., "Phylodynamics of the HIV-1 CRF02_AG clade in Cameroon," Infect. Genet. Evol., vol. 12, no. 2, pp. 453–460, 2012.

100. M. Peeters, C. Toure-Kane, and J. N. Nkengasong, "Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials.," AIDS, vol. 17, no. 18, pp. 2547–2560, 2003.

101. J. P. Messina et al., "Spatial and socio-behavioral patterns of HIV prevalence in the Democratic Republic of Congo.," Soc. Sci. Med., vol. 71, no. 8, pp. 1428–35, Oct. 2010.

102. R. R. Gray et al., "Spatial phylodynamics of HIV-1 epidemic emergence in east Africa," Wolters Kluwer Heal., vol. 23, pp. 9–17, 2009.

103. R. Lihana, D. Ssemwanga, A. Abimiku, and N. Ndembi, "Update on HIV-1 Diversity in Africa: A Decade in Review," AIDS Rev, vol. 14, pp. 83–100, 2012.

104. C. a Brennan et al., "The prevalence of diverse HIV-1 strains was stable in Cameroonian blood donors from 1996 to 2004.," J. Acquir. Immune Defic. Syndr., vol. 49, no. 4, pp. 432–439, 2008.

105. J. K. Carr et al., "HIV-1 recombinants with multiple parental strains in low-prevalence, remote regions of Cameroon: evolutionary relics?," Retrovirology, vol. 7, p. 39, 2010.

106. P. B. Baryshev, V. V. Bogachev, and N. M. Gashnikova, "Genetic characterization of an isolate of HIV type 1 AG recombinant form circulating in Siberia, Russia," Arch. Virol., vol. 157, no. 12, pp. 2335–2341, 2012.

107. J. K. Carr et al., "Outbreak of a West African Recombinant of HIV-1 in Tashkent, Uzbekistan," J Acquir Immune Defic Syndr, vol. 39, pp. 570–575, 2005.

108. L. M. Eyzaguirre et al., “Genetic Characterization of HIV-1 Strains Circulating in Kazakhstan,” J. Acquir. Immune Defic. Syndr., vol. 46, no. 1, pp. 19–23, 2007.

109. E. Kazennova et al., “HIV-1 Genetic Variants in the Russian Far East,” AIDS Res. Hum. Retroviruses, vol. 30, no. 8, 2014.

110. V. Laga et al., “The Genetic Variability of HIV-1 in Kyrgyzstan: The Spread of CRF02_AG and Subtype A1 Recombinants.,” J. HIV AIDS, vol. 1.2, 2015.

111. I. Lapovok et al., “Short communication: molecular epidemiology of HIV type 1 infection in Kazakhstan: CRF02_AG prevalence is increasing in the southeastern provinces.,” AIDS Res. Hum. Retroviruses, vol. 30, no. 8, pp. 769–774, 2014.

112. L. M. Hofstra et al., “Transmission of HIV Drug Resistance and the Predicted Effect on Current First-line Regimens in Europe,” Clin. Infect. Dis., vol. 62, no. 5, pp. 655–663, Mar. 2016.

113. A. Chaillon et al., “Spatiotemporal dynamics of HIV-1 transmission in France (1999-2014) and impact of targeted prevention strategies,” Retrovirology, vol. 14, no. 1, p. 15, Dec. 2017.

114. C. Tamalet et al., “Emergence of Clusters of CRF02_AG and B Human Immunodeficiency Viral Strains Among Men Having Sex With Men Exhibiting HIV Primary Infection in Southeastern France,” J. Med. Virol. J. Med. Virol, vol. 87, no. 87, 2015.

115. K. Dauwe et al., “Characteristics and spread to the native population of HIV-1 non-B subtypes in two European countries with high migration rate,” BMC Infect. Dis., 2015.

116. M. Giuliani et al., “Circulation of HIV-1 CRF02_AG among MSM population in central Italy: a molecular epidemiology-based study.,” Biomed Res. Int., vol. 2013, p. 810617, Jan. 2013.

117. E. Delatorre, C. A. Velasco-De-Castro, J. H. Pilotto, J. C. Couto-Fernandez, G. Bello, and M. G. Morgado, “Short Communication: Reassessing the Origin of the

HIV-1 CRF02_AG Lineages Circulating in Brazil," AIDS Res. Hum. Retroviruses, vol. 31, no. 12, pp. 1230–1237, 2015.

118. E. O. Delatorre, G. Bello, W. A. Eyer-Silva, S. L. Chequer-Fernandez, M. G. Morgado, and J. C. Couto-Fernandez, "Evidence of Multiple Introductions and Autochthonous Transmission of the HIV Type 1 CRF02_AG Clade in Brazil," AIDS Res. Hum. Retroviruses, vol. 28, no. 10, pp. 1369–1372, 2012.

119. V. A. Novitsky et al., "Molecular cloning and phylogenetic analysis of human immunodeficiency virus type 1 subtype C: a set of 23 full-length clones from Botswana.," J. Virol., vol. 73, no. 5, pp. 4427–32, May 1999.

120. N. Koch, J.-B. Ndihokubwayo, N. Yahi, C. Tourres, J. Fantini, and C. Tamalet, "Genetic Analysis of HIV Type 1 Strains in Bujumbura (Burundi): Predominance of Subtype C Variant," AIDS Res. Hum. Retroviruses, vol. 17, no. 3, pp. 269–273, 2001.

121. N. Vidal et al., "HIV Type 1 Diversity and Antiretroviral Drug Resistance Mutations in Burundi," AIDS Res. Hum. Retroviruses, vol. 23, no. 1, pp. 175–180, 2007.

122. J. Me Maslin et al., "Epidemiology and Genetic Characterization of HIV-1 Isolates in the General Population of Djibouti (Horn of Africa)," J. Acquir. Immune Defic. Syndr., pp. 129–132, 2005.

123. A. Ababa et al., "HIV Type 1 Subtype C in Addis Ababa, Ethiopia.," AIDS Res. Hum. Retroviruses, vol. 13, no. 12, 1997.

124. L. F. M. Brígido et al., "HIV Type 1 Subtype C and CB Pol Recombinants Prevail at the Cities with the Highest AIDS Prevalence Rate in Brazil," AIDS Res. Hum. Retroviruses, vol. 23, no. 12, pp. 1579–1586, Dec. 2007.

125. K. K. Ariën, G. Vanham, and E. J. Arts, "Is HIV-1 evolving to a less virulent form in humans?," Nat. Rev. Microbiol., vol. 5, no. 2, pp. 141–151, 2007.

126. A. Abraha et al., "CCR5- and CXCR4-tropic subtype C human immunodeficiency virus type 1 isolates have a lower level of pathogenic fitness than

other dominant group M subtypes: implications for the epidemic.," J. Virol., vol. 83, no. 11, pp. 5592–605, 2009.

127. C. M. Venner et al., "Infecting HIV-1 Subtype Predicts Disease Progression in Women of Sub-Saharan Africa," EBioMedicine, vol. 13, pp. 305–314, 2016.

128. M. M. Thomson and A. Fernández-garcía, "Phylogenetic structure in African HIV-1 subtype C revealed by selective sequential pruning," Virology, vol. 415, no. 1, pp. 30–38, 2011.

129. E. O. Delatorre and G. Bello, "Phylodynamics of HIV-1 subtype C epidemic in East Africa," PLoS One, vol. 7, no. 7, pp. 1–10, 2012.

130. E. Wilkinson, S. Engelbrecht, and T. de Oliveira, "History and origin of the HIV-1 subtype C epidemic in South Africa and the greater southern African region.," Sci. Rep., vol. 5, no. November, p. 16897, 2015.

131. E. Delatorre and G. Bello, "Phylodynamics of the HIV-1 Epidemic in Cuba.," PLoS One, vol. 8, p. e72448, 2013.

132. G. Bello et al., "Origin and evolutionary history of HIV-1 subtype C in Brazil.," AIDS, vol. 22, no. 15, pp. 1993–2000, 2008.

133. E. Delatorre et al., "Tracing the Origin and Northward Dissemination Dynamics of HIV-1 Subtype C in Brazil," PLoS One, vol. 8, no. 9, p. e74072, Sep. 2013.

134. N. M. C. Véras, R. R. Gray, L. F. D. M. Brígido, R. Rodrigues, and M. Salemi, "High-resolution phylogenetics and phylogeography of human immunodeficiency virus type 1 subtype C epidemic in South America.," J. Gen. Virol., vol. 92, no. Pt 7, pp. 1698–1709, 2011.

135. T. Gräf and A. R. Pinto, "The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent.," Virology, vol. 435, no. 1, pp. 170–8, Jan. 2013.

136. T. de Oliveira, D. Pillay, R. J. Gifford, and UK Collaborative Group on HIV Drug Resistance, "The HIV-1 subtype C epidemic in South America is linked to the United Kingdom.," PLoS One, vol. 5, no. 2, p. e9311, Feb. 2010.

137. C. Shen, J. Craigo, M. Ding, Y. Chen, and P. Gupta, "Origin and dynamics of HIV-1 subtype c infection in India," PLoS One, 2011.

138. U. Neogi et al., "Molecular epidemiology of HIV-1 subtypes in India: Origin and evolutionary history of the predominant subtype C," PLoS One, 2012.

139. C. Beyrer, M. H. Razak, K. Lisam, J. Chen, W. Lui, and X. F. Yu, "Overland heroin trafficking routes and HIV-1 spread in south and south-east Asia," Aids, vol. 14, no. 1, pp. 75–83, 2000.

140. B. T. Grenfell et al., "Unifying the epidemiological and evolutionary dynamics of pathogens.," Science, vol. 303, no. 5656, pp. 327–332, 2004.

141. E. C. Holmes and E. C. Holmes, "Evolutionary history and phylogeography of human viruses.," Annu. Rev. Microbiol., vol. 62, pp. 307–28, 2008.

142. E. Zuckerkandl and L. Pauling, "Evolutionary Divergence and Convergence in Proteins," in Evolving Genes and Proteins, vol. 42, no. 2, Elsevier, 1965, pp. 97–166.

143. Zuckerkandl and L. Pauling, "Molecular disease, evolution, and genetic heterogeneity," Horizons Biochem., pp. 189–225, 1962.

144. M. Kimura, "Evolutionary Rate at the Molecular Level," Nature, vol. 217, no. 5129, pp. 624–626, Feb. 1968.

145. F. J. Ayala, "Vagaries of the molecular clock," Proc. Natl. Acad. Sci. U. S. A., vol. 94, no. 15, pp. 7776–7783, 1997.

146. M. Hasegawa and H. Kishino, "Heterogeneity of tempo and mode of mitochondrial DNA evolution among mammalian orders," Idengaku zasshi, vol. 64, no. 4. pp. 243–258, Aug-1989.

147. F. Tajima, "Simple methods for testing the molecular evolutionary clock hypothesis.," Genetics, vol. 135, no. 2, pp. 599–607, Oct. 1993.

148. C. I. Wu and W. H. Li, “Evidence for higher rates of nucleotide substitution in rodents than in man.,” Proc. Natl. Acad. Sci. U. S. A., vol. 82, no. 6, pp. 1741–5, Mar. 1985.

149. D. Posada and K. A. Crandall, “Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1).,” Mol. Biol. Evol., vol. 18, no. 6, pp. 897–906, Jun. 2001.

150. G. M. Jenkins, A. Rambaut, O. G. Pybus, and E. C. Holmes, “Rates of molecular evolution in RNA viruses: A quantitative phylogenetic analysis,” J. Mol. Evol., vol. 54, no. 2, pp. 156–165, Feb. 2002.

151. B. Korber, “Limitations of a Molecular Clock Applied to Considerations of the Origin of HIV-1,” Science (80-.)., vol. 280, no. 5371, pp. 1868–1871, 1998.

152. M. Hasegawa, H. Kishino, and T. Yano, “Estimation of branching dates among primates by molecular clocks of nuclear DNA which slowed down in Hominoidea,” J. Hum. Evol., vol. 18, no. 5, pp. 461–476, Aug. 1989.

153. J. L. Thorne, H. Kishino, and I. S. Painter, “Estimating the rate of evolution of the rate of molecular evolution,” Mol. Biol. Evol., vol. 15, no. 12, pp. 1647–1657, Dec. 1998.

154. S. Aris-Brosou and Z. Yang, “Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny,” Syst. Biol., vol. 51, no. 5, pp. 703–714, 2002.

155. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, and A. Rambaut, “Relaxed phylogenetics and dating with confidence.,” PLoS Biol., vol. 4, no. 5, p. e88, May 2006.

156. S. Y. W. Ho and S. Duchêne, “Molecular-clock methods for estimating evolutionary rates and timescales,” Mol. Ecol., vol. 23, no. 24, pp. 5947–5965, Dec. 2014.

157. A. J. Drummond, O. G. Pybus, A. Rambaut, R. Forsberg, and A. G. Rodrigo, "Measurably evolving populations," Trends Ecol. Evol., vol. 18, no. 9, pp. 481–488, 2003.

158. J. F. . Kingman, "Exchangeability and the Evolution of Large Populations," Exch. Probab. Stat., pp. 97–112, 1982.

159. J. F. C. Kingman, "The coalescent," Stoch. Process. their Appl., vol. 13, no. 3, pp. 235–248, 1982.

160. J. F. C. Kingman, "On the Genealogy of Large Populations," J. Appl. Probab., vol. 19, no. 1982, p. 27, 1982.

161. D. G. Kendall, "On Some Modes of Population Growth Leading to R. A. Fisher's Logarithmic Series Distribution.," Biometrika, vol. 35, no. 1/2, pp. 6–15, 1948.

162. D. G. Kendall, "On the Generalized 'Birth-and-Death' Process.," Ann. Math. Stat., vol. 19, no. 1, pp. 1–15, 1948.

163. P. Donnelly and S. Tavaré, "Coalescents and genealogical structure under neutrality.," Annu. Rev. Genet., vol. 29, no. 29, pp. 401–21, 1995.

164. A. G. Rodrigo et al., "Coalescent estimates of HIV-1 generation time in vivo.," Proc. Natl. Acad. Sci. U. S. A., vol. 96, no. 5, pp. 2187–91, 1999.

165. T. Leitner and W. Fitch, "The Phylogenetics of Known Transmission Histories.," in The Evolution of HIV, Crandall K, Ed. Baltimore: Johns Hopkins University Press, 1999, pp. 315–45.

166. R. a. Fisher, "The genetical theory of natural selection.," The Genetical Theory of Natural Selection. The Clarendon Press, Oxford, 1 edition, 1930.

167. S. Wright, "Evolution in Mendelian Populations.," Genetics, vol. 16, no. 2, pp. 97–159, Mar. 1931.

168. R. C. Griffiths and S. Tavaré, "Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to Statistical Science. ® www.jstor.org," Stat. Sci., 1994.

169. M. K. Kuhner, J. Yamato, and J. Felsenstein, "Maximum likelihood estimation of population growth rates based on the coalescent.," Genetics, vol. 149, no. 1, pp. 429–434, 1998.

170. M. Slatkin and W. P. Maddison, "A cladistic measure of gene flow inferred from the phylogenies of alleles," Genetics, vol. 123, no. 3, pp. 603–613, 1989.

171. M. Slatkin and W. P. Maddison, "Detecting isolation by distance using phylogenies of genes," Genetics, vol. 126, no. 1, pp. 249–260, 1990.

172. P. Beerli and J. Felsenstein, "Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach," Genetics, vol. 152, no. 2, pp. 763–773, 1999.

173. R. C. Griffiths and P. Marjoram, "Ancestral inferences from samples of DNA sequences with recombination," J. Comput. Biol., vol. 3, no. 4, pp. 479–502, 1996.

174. R. R. Hudson, "Estimating the recombination parameter of a finite population model without selection.," Genet. Res., vol. 50, no. 3, pp. 245–50, Dec. 1987.

175. R. R. Hudson and N. L. Kaplan, "Statistical properties of the number of recombination events in the history of a sample of DNA sequences," Genetics, vol. 111, no. 1, pp. 147–164, 1985.

176. N. L. Kaplan, T. Darden, and R. R. Hudson, "The coalescent process in models with Selection and Recombination," Genetics, vol. 120, no. 3, pp. 819–829, 1988.

177. Krone and Neuhauser, "Ancestral Processes with Selection," Theor. Popul. Biol., vol. 51, no. 3, pp. 210–237, 1997.

178. C. Neuhauser and S. M. Krone, "The genealogy of samples in models with selection," Genetics, vol. 145, no. 2, pp. 519–534, 1997.

179. J. Felsenstein and A. G. Rodrigo, "Coalescent approaches to HIV population genetics.," in Molecular evolution of HIV, K. Crandall, Ed. Baltimore: Johns Hopkins University Press, 1999, pp. 233–272.

180. O. G. Pybus, A. Rambaut, and P. H. Harvey, "An integrated framework for the inference of viral population history from reconstructed genealogies," Genetics, vol. 155, no. 3, pp. 1429–1437, 2000.

181. A. J. Drummond, A. Rambaut, B. Shapiro, and O. G. Pybus, "Bayesian coalescent inference of past population dynamics from molecular sequences," Mol. Biol. Evol., vol. 22, no. 5, pp. 1185–1192, 2005.

182. K. Strimmer and O. G. Pybus, "Exploring the Demographic History of DNA Sequences Using the Generalized Skyline Plot," Mol. Biol. Evol, vol. 18, no. 12, pp. 2298–2305, 2001.

183. O. G. Pybus, M. A. Charleston, S. Gupta, A. Rambaut, E. C. Holmes, and P. H. Harvey, "The epidemic behavior of the hepatitis C virus.," Science, vol. 292, no. 5525, pp. 2323–5, Jun. 2001.

184. J. Wallinga and M. Lipsitch, "How generation intervals shape the relationship between growth rates and reproductive numbers.," Proc. Biol. Sci., vol. 274, no. 1609, pp. 599–604, 2007.

185. V. Boskova, Estimation of epidemiological parameters based on simulated phylogenetic trees using birth-death process and coalescent models, no. August. 2013.

186. T. Stadler et al., "Estimating the basic reproductive number from viral sequence data," Mol. Biol. Evol., vol. 29, no. 1, pp. 347–357, 2012.

187. L. Du Plessis and T. Stadler, "Getting to the root of epidemic spread with phylodynamic analysis of genomic data," Trends Microbiol., vol. 23, no. 7, pp. 383–386, 2015.

188. T. Stadler, D. Kuhnert, S. Bonhoeffer, and a. J. Drummond, "Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV)," Proc. Natl. Acad. Sci., vol. 110, no. 1, pp. 228–233, 2013.

189. E. A. Thompson, Human Evolutionary Trees. Cambridge University Press, 1975.

190. T. Gernhard, “The conditioned reconstructed process,” J. Theor. Biol., vol. 253, no. 4, pp. 769–778, Aug. 2008.

191. T. Stadler, “On incomplete sampling under birth-death models and connections to the sampling-based coalescent,” J. Theor. Biol., vol. 261, no. 1, pp. 58–66, 2009.

192. Z. Yang and B. Rannala, “Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method.,” Mol. Biol. Evol., vol. 14, no. 7, pp. 717–24, Jul. 1997.

193. T. Stadler and Z. Yang, “Dating phylogenies with sequentially sampled tips,” Syst. Biol., vol. 62, no. 5, pp. 674–688, 2013.

194. V. Boskova, S. Bonhoeffer, and T. Stadler, “Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models.,” PLoS Comput. Biol., vol. 10, no. 11, p. e1003913, Nov. 2014.

Chapter 2

Objectives

2.1 General objective

To characterize the spatiotemporal dynamics of different strains of HIV-1 in different geographical regions, in order to identify their origin, population dynamics and potential factors underlying their propagation.

2.2 Specific objectives

- To identify and characterize the evolutionary and demographic history of major HIV-1 subtype B pandemic clades circulating in Latin America, using coalescent- based phylodynamic inference.
- To identify and characterize the evolutionary and demographic history of major HIV-1 CRF02_AG clades circulating in West and Central Africa and other regions around the world, under coalescent-based phylodynamic inference.
- To reconstruct the evolutionary and demographic history of major HIV-1 subtype C clades circulating in East Africa and Brazil using both coalescent-based and birth-death phylodynamic frameworks.

Chapter 3

Phylodynamics of major HIV-1 subtype B pandemic clades circulating in Latin America.

Article published in:

Daiana Mir, Marina Cabello, Hector Romero and Gonzalo Bello.

AIDS (London, England) 2015; 29(14):1863-1869.

3.1 Abstract

The HIV-1 epidemic in Latin America is dominated by subtype B, which accounts for nearly 70% of infections in the region. The aim of this work was to identify the major HIV-1 subtype $B_{PANDEMIC}$ clades circulating in Latin America and to reconstruct their evolutionary and demographic history. To this end a total of 6789 HIV-1 subtype B pol sequences collected from seven different Latin American countries were combined with $B_{PANDEMIC}$ reference sequences (n = 500) from the United States and France. Major $B_{PANDEMIC}$ clades were identified by Maximum Likelihood phylogenetic analysis. Time-scale and demographic reconstructions were performed using a Bayesian coalescent-based method. We identified 10 major $B_{PANDEMIC}$ monophyletic lineages mainly composed by Latin American sequences. Two clades were classified as regional-specific lineages as comprises sequences from at least two neighboring countries; whereas the other eight clades were country-specific. The median age of major Latin American $B_{PANDEMIC}$ clades encompass a period of two decades (1968-1988); although most of them probably arose before the early 1980s. All major clades seem to have experienced an initial period of exponential growth, with median epidemic growth rates that range from 0.50 $year^{-1}$ to 0.94 $year^{-1}$, followed by a recent decline in growth rate. In conclusion, about one third of HIV-1 subtype B infections in Latin America originated from the spread of a few $B_{PANDEMIC}$ founder strains probably introduced in the region since the late 1960s. Despite the successful dissemination, all major $B_{PANDEMIC}$ clades showed signs of subsequent epidemic stabilization.

3.2 Introduction

It has been estimated that 1.4 million (1.1–1.7 million) people were living with the HIV type 1 (HIV-1) in Latin America at the end of 2011 [1]. The HIV adult (15–49 years) prevalence (0.4%) and the number of people newly infected each year (80.000–90.000) in Latin America have remained relatively stable since the early 2000s [1]. Despite this stable epidemiologic scenario, very high HIV prevalence rates are still observed in some vulnerable populations including transsexual people (30–40%), male sex workers (10–25%), MSM (5–20%), intravenous drug users (IDUs) (5–20%), and female sex workers (0.5–5%) [2,3].

The HIV-1 epidemic in Latin America is dominated by subtype B, which accounts for nearly 70% of infections in the region [4]. The subtype B epidemic in the America is characterized by the expansion of both pandemic viral strains that are widely disseminated in the Caribbean region [5] and the subtype $B_{PANDEMIC}$ clade that seems to be the dominant subtype B variant outside the Caribbean [6,7]. A recent study conducted by our group estimates that most (>90%) of HIV-1 subtype B infections in Latin America resulted from the expansion of the globally disseminated $B_{PANDEMIC}$ clade [8], thus displaying a molecular epidemiologic profile similar to North America and Europe.

The $B_{PANDEMIC}$ clade probably arose in the United States in the late 1960s [6] and most studies performed up-to-date support a very rapid dissemination of this viral clade to Latin America. Some studies estimated the origin of major HIV-1 subtype B clades in Argentina and Brazil between the middle 1960s and the early 1970s [9,10]. One study suggests that a single introduction of the $B_{PANDEMIC}$ clade around the middle 1960s accounts for most current cases in Honduras and El Salvador [11], whereas other study

showed that Panamanian HIV-1 epidemic is mainly driven by the expansion of several country-specific $B_{PANDEMIC}$ clades probably introduced between the early 1970s and the early 1980s [12]. The precise number, geographic extension, and dissemination dynamics of major HIV-1 $B_{PANDEMIC}$ clades circulating in Latin America, however, remain largely unknown.

The objective of this study was to characterize and reconstruct the evolutionary and demographic history of major HIV-1 $B_{PANDEMIC}$ clades circulating in Latin America. For this, we used a comprehensive dataset of HIV-1 subtype B *pol* sequences (n = 6789) isolated from seven Latin American countries between 1990 and 2011. These Latin American sequences were combined with subtype B reference sequences, representative of the $B_{PANDEMIC}$ clade circulating in the United States (n = 300) and France (n = 200), and subjected to maximum likelihood (ML) and Bayesian coalescent analyses.

3.3 Materials and Methods

3.3.1 HIV-1 subtype B *pol* sequences

A total of 6789 HIV-1 subtype B *pol* sequences with known sampling dates isolated from Argentina, Brazil, El Salvador, Honduras, Mexico, Peru, and Venezuela were selected from a previous study [8] (Table S3.1). Other Latin American countries were not included because of the very low number of sequences available (n < 50). Sequences covered the entire protease and partial reverse transcriptase regions (nucleotides 2253–3260 relative to HXB2 genome) and were initially retrieved from the Los Alamos HIV Database http://www.hiv.lanl.gov. Latin American subtype B *pol* sequences were aligned with $B_{PANDEMIC}$ reference sequences from the United States (n =

300) and France (n = 200) [5] using the Clustal W program [13]. All sites associated with major antiretroviral drug resistance in protease and reverse transcriptase were excluded as described previously [8]. Alignments are available from the authors upon request.

3.3.2 Detection of country-specific HIV-1 subtype B clades

HIV-1 subtype B sequences from each Latin American country were combined with 500 $B_{PANDEMIC}$ *pol* reference sequences from the United States and France and were subjected to ML phylogenetic analysis with sequential pruning of nonclustered and ambiguously positioned taxa [14]. ML trees were reconstructed with the PhyML program [15] using an online web server [16] and major Latin American $B_{PANDEMIC}$ transmission clusters were defined as those monophyletic clades (approximate likelihood-ratio test ≥ 0.90) with at least 40 sequences, 95% or more of which were of Latin American origin (see Supplementary materials and methods for details). These criteria allow us to select $B_{PANDEMIC}$ clusters with both regional epidemic importance (>1% of subtype B infections of one country) and adequate sample sizes to give reliable coalescent estimates.

3.3.3 Evolutionary and demographic reconstructions

The evolutionary rate, the time of the tMRCA, and the mode and rate (r, years $^{-1}$) of population growth of major HIV-1 $B_{PANDEMIC}$ clades circulating in Latin America were jointly estimated using the Bayesian Markov Chain Monte Carlo approach as implemented in BEAST v1.8 [17,18] with BEAGLE to improve run-time [19] (see Supplementary materials and methods for details). Changes in *Ne* through time were initially estimated using a Bayesian Skyline coalescent tree prior [20] 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 the datasets. Comparison between demographic models was performed using the log marginal likelihood estimation based on path sampling and stepping-stone sampling methods [21].

3.4 Results

3.4.1 Identification of major HIV-1 $B_{PANDEMIC}$ clades in Latin America

A total of 6789 HIV-1 subtype *pol* sequences isolated from seven Latin American countries between 1990 and 2011 were analyzed in this study. ML phylogenetic analysis, combined with sequential pruning of nonclustered sequences and clusters of small size, resulted in the identification of 12 country-specific $B_{PANDEMIC}$ clades of large size (n >40) in Latin America that together comprise 36% (n= 2423) of all subtype B sequences from the region here included (Figs. S3.1-S3.3). Major $B_{PANDEMIC}$ clades comprise a substantial proportion (31–91%) of HIV-1 subtype B sequences from most Latin American countries analyzed, except for Venezuela (Table S3.2). Excluding the large clade previously described in the isolated Warao Amerindian population [22,23], the remaining Venezuelan sequences were distributed in a large collection of small (n < 30) lineages that were intermixed among non-Latin American reference strains (Fig. S3.4).

All major $B_{PANDEMIC}$ clades from South America and Central America/Mexico were then combined into two regional subsets and subjected to ML analysis. Among the 12 $B_{PANDEMIC}$ Latin American clades initially identified, eight contain sequence from one single country, whereas

Table 3.1: Major HIV-1 subtype $B_{PANDEMIC}$ clades identified in Latin America after combined analysis of sequences from Argentina/Brazil/Peru and El Salvador/Honduras/Mexico.

Dataset	Country	Clade	N(%)[a]	Sampling dates
A	Argentina	B_{SAM}[b]	199(15)	2000–2008
		$B_{SAM\text{-}AR}$	87(6)	2001–2007
		$B_{AR\text{-}II}$	142(11)	2001–2007
		$B_{AR\text{-}III}$	65(5)	2001–2007
	Brazil	$B_{SAM} = B_{SAM\text{-}BR}$	445(18)	1997–2010
		$B_{BR\text{-}II}$	174(7)	1997–2010
		$B_{BR\text{-}III}$	96(4)	1997–2009
		$B_{BR\text{-}IV}$	43(2)	1997–2009
	Peru	B_{PE}	126(51)	2006–2010
B	El Salvador	B_{CAM}[c]	69(41)	2008–2010
		$B_{CAM\text{-}SV}$	47(28)	2008–2010
	Honduras	$B_{CAM} = B_{CAM\text{-}HN}$	462(91)	2001–2007
	Mexico	$B_{MX\text{-}I}$	406(25)	2004–2010
		$B_{MX\text{-}II}$	196(12)	2005–2010

[a] Number of sequences within each major $B_{PANDEMIC}$ clade and corresponding percentage of the total number of sequences from each country included in the study: Argentina = 1351, Brazil = 2480, El Salvador = 170, Honduras = 507, Mexico = 1647, and Peru = 249.
[b] Includes all sequence from Argentina that branched within the B_{SAM} clade.
[c] Includes all sequence from El Salvador that branched within the B_{CAM} clade.

the remaining four are branched into two larger regional clades (B_{SAM} and B_{CAM}) (Fig. 3.1 and Table 3.1). The regional clade B_{SAM} comprises sequences from clades $B_{AR\text{-}I}$ and $B_{BR\text{-}I}$ that circulate in Argentina and Brazil, respectively; whereas the clade B_{CAM} comprises sequences from clades B_{HN} and B_{SV} that circulate in Honduras and El Salvador, respectively. A closer inspection of these regional clades revealed that most B_{SAM} Argentinean sequences branched in a monophyletic subclade ($B_{SAM\text{-}AR}$) nested among the basal Brazilian strains, whereas most (68%) B_{CAM} sequences from El Salvador branched in a monophyletic subclade ($B_{CAM\text{-}SV}$) nested among the basal Honduran strains (Fig. 3.1).

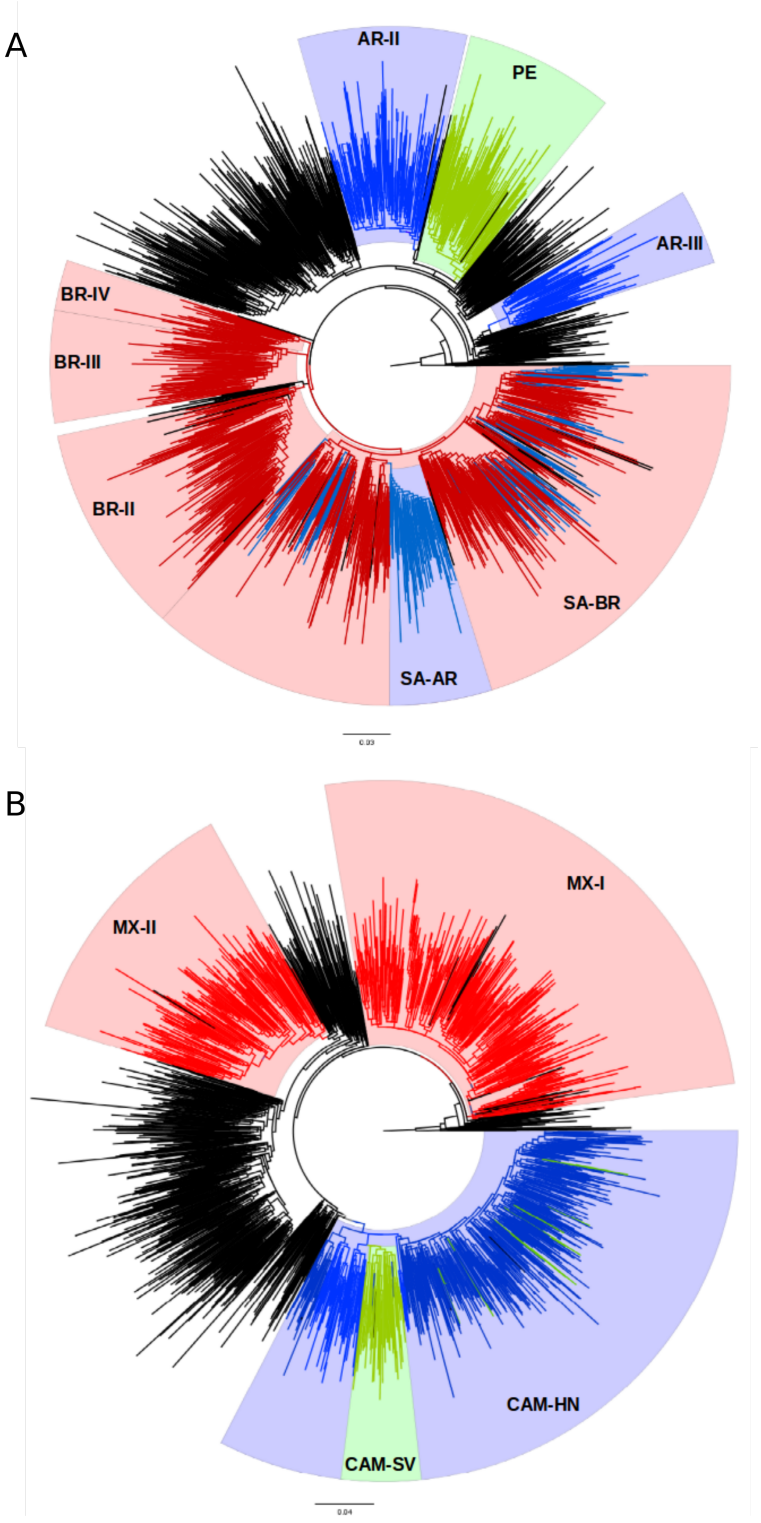


Figure 3.1: Maximum likelihood phylogenetic tree of HIV-1 subtype B *pol* protease and partial reverse transcriptase sequences (∼1000 nt) from major $B_{PANDEMIC}$ lineages circulating in Latin America. Major $B_{PANDEMIC}$ lineages from South America (a) and Central America/Mexico (b) were combined with reference sequences representative of the $B_{PANDEMIC}$ lineages circulating in the United States (n = 300), and France (n = 200). Positions of major Latin American $B_{PANDEMIC}$ clades are indicated by colored shaded boxes. Trees were rooted using HIV-1 subtype D reference sequences. The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site. $B_{PANDEMIC}$, subtype B pandemic.

3.4.2 Time scale of major HIV-1 $B_{PANDEMIC}$ clades in Latin America

Broad similar median substitution rates were observed for the *pol* region of the different HIV-1 clades, ranging from 1.8 x 10^{-3} substitution/site per year to 2.2 x 10^{-3} substitution/site/year (95% highest probability density overlap: 1.7–2.5 x 10^{-3} substitution/site per year) (Table 3.2). With these substitution rates, the median tMRCA of the Latin American $B_{PANDEMIC}$ clades embraces a period of two decades between the late 1960s and the late 1980s. Interestingly, of the four $B_{PANDEMIC}$ clades with the oldest median tMRCA estimates, three were from Brazil ($B_{SAM\text{-}BR}$ = 1968, $B_{BR\text{-}II}$ = 1970, and $B_{BR\text{-}III}$ = 1976) and the fourth one was from Honduras ($B_{CAM\text{-}HN}$ = 1975). Most of the remaining country-specific $B_{PANDEMIC}$ clades in Latin America probably arose between the late 1970s and the early 1980s, except for the clades $B_{CAM\text{-}SV}$ and $B_{MX\text{-}I}$ that probably arose between the middle and late 1980s.

3.4.3 Demographic history of major HIV-1 $B_{PANDEMIC}$ clades in Latin America

The Bayesian skyline plot (BSP) analyses suggest that all $B_{PANDEMIC}$ clades displayed a similar population growth pattern characterized by an initial phase of exponential growth followed by a decline in growth rate (Fig. 3.2). Some differences, however, were observed in the *Ne*, time of slow down (from late 1980s to early 2000s) and duration of the exponential growth period (10–20 years) for each clade (Fig. S3.5). These parameters were significantly correlated with the median tMRCA ($P < 0.005$) of each clade (Fig. S3.6), indicating that older clades will tend to stabilize earlier but maintain longer periods of exponential growth and reach higher Ne values. To determine the epidemic growth rate parameter of each $B_{PANDEMIC}$ lineage,

we selected the logistic growth model that outperforms the other growth models for all clades (Table S??). According to the logistic function, the median growth rates for the $B_{PANDEMIC}$ clades range from 0.50 $year^{-1}$ to 0.94 $year^{-1}$, with a great overlap of the highest probability density intervals for most clades (Figs. 3.2 and S3.7).

Table 3.2: Bayesian time-scale estimates of major HIV-1 subtype $B_{PANDEMIC}$ Latin American clades.

Country	Clade	Substitution rate	Coefficient of variation	tMRCA
Argentina	B_{SAM-AR}	1.9×10^{-3} (1.7×10^{-3} –2.3×10^{-3})	0.35 (0.26–0.45)	1982 (1976–1988)
	B_{AR-II}	1.9×10^{-3} (1.7×10^{-3} –2.3×10^{-3})	0.36 (0.29–0.44)	1979 (1974–1984)
	B_{AR-III}	1.9×10^{-3} (1.7×10^{-3} –2.4×10^{-3})	0.35 (0.26–0.45)	1981 (1974–1987)
Brazil	B_{SAM-BR}	1.8×10^{-3} (1.7×10^{-3} –1.9×10^{-3})	0.27 (0.24–0.31)	1968 (1963–1973)
	B_{BR-II}	1.9×10^{-3} (1.7×10^{-3} –2.3×10^{-3})	0.23 (0.18–0.28)	1970 (1964–1978)
	B_{BR-III}	1.8×10^{-3} (1.7×10^{-3} –2.2×10^{-3})	0.19 (0.11–0.27)	1976 (1971–1982)
	B_{BR-IV}	1.9×10^{-3} (1.7×10^{-3} –2.4×10^{-3})	0.27 (0.17–0.38)	1983 (1978–1988)
El Salvador	B_{CAM-SV}	2.2×10^{-3} (1.8×10^{-3} –2.5×10^{-3})	0.05 (3.3×10^{-5}–0.15)	1988 (1982–1993)
Honduras	B_{CAM-HN}	2.1×10^{-3} (1.7×10^{-3} –2.4×10^{-3})	0.26 (0.22–0.30)	1975 (1967–1981)
Mexico	B_{MX-I}	2.2×10^{-3} (1.8×10^{-3} –2.5×10^{-3})	0.29 (0.24–0.34)	1985 (1980–1989)
	B_{MX-II}	2.2×10^{-3} (1.8×10^{-3} –2.5×10^{-3})	0.12 (0.01–0.20)	1981 (1974–1986)
Peru	B_{PE}	1.9×10^{-3} (1.7×10^{-3} –2.3×10^{-3})	0.24 (0.17–0.31)	1978 (1971–1985)

Bayesian skyline estimates of substitution rate (substitution/site per year), coefficient of rate variation, and time to the most recent common ancestor (tMRCA) for major country-specific $B_{PANDEMIC}$ Latin American clades. The coefficient of variation was above zero for all clades, supporting the use of a relaxed molecular clock model to the time scale reconstructions.

3.5 Discussion

The HIV-1 subtype $B_{PANDEMIC}$ was ignited with the arrival of the virus into the United States around the late 1960s from where a subsequent dissemination to other countries around the world took place, thereby establishing the so–called $B_{PANDEMIC}$ clade [6]. Based upon the analysis of an extensive dataset of HIV-1 subtype B *pol* sequences, we observed that most (64%) Latin American sequences were distributed as sporadic lineages or within country-specific clades of small sizes, suggesting multiple independent viral introductions into the region with a restricted local dissemination. We identified, however, 12 major $B_{PANDEMIC}$ clades with particular successfully epidemic outcomes that together comprise 36% of all HIV-1 subtype B Latin American sequences here included.

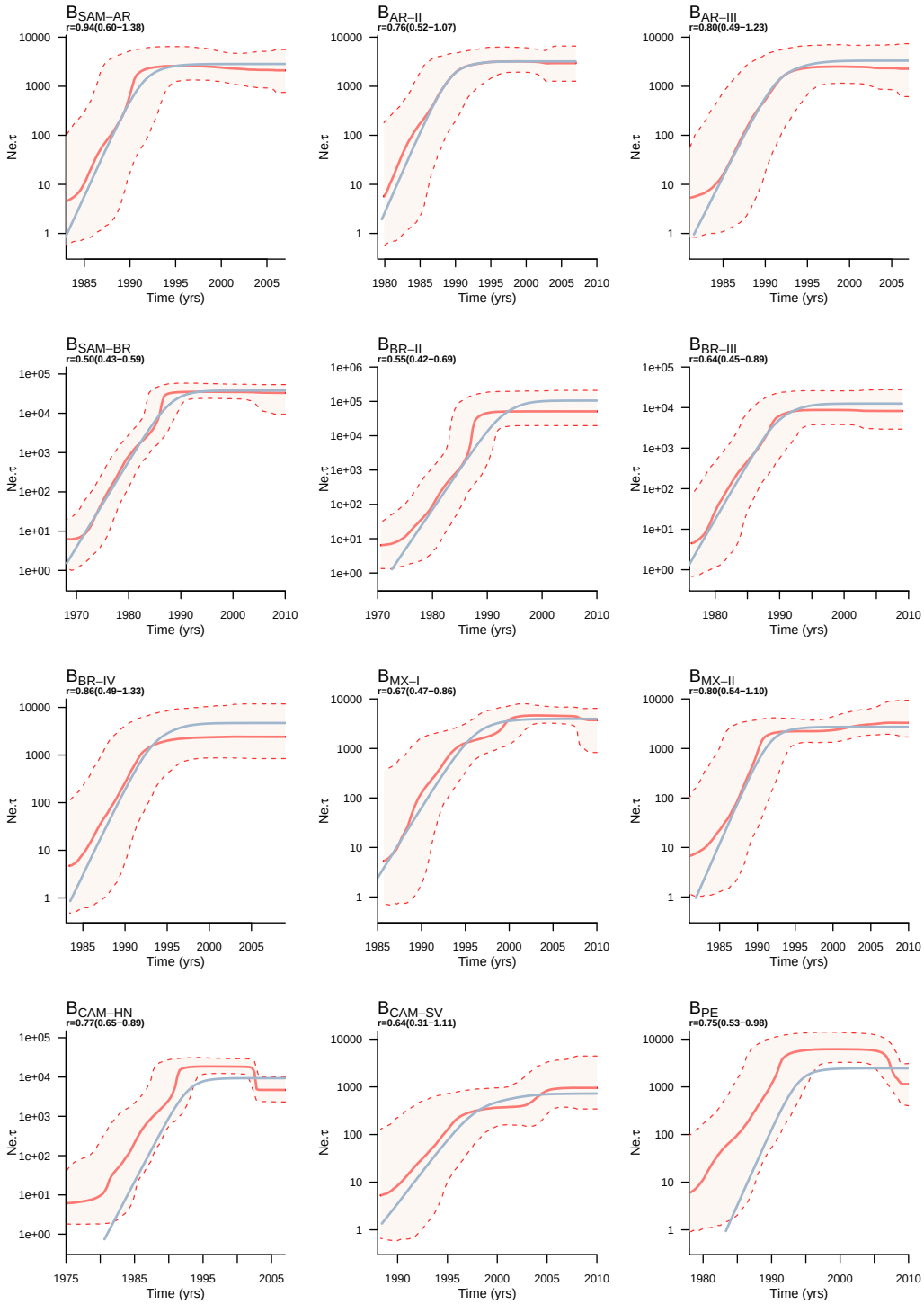


Figure 3.2: Demographic history of major HIV-1 $B_{PANDEMIC}$ Latin American clades. Median estimates of the effective number of infections using Bayesian skyline (solid red line) and logistic growth model (solid blue line) are shown in each graphic together with 95% highest probability density intervals of the Bayesian skyline estimates (dashed red area). The vertical axes represent the estimated effective number of infections on a logarithmic scale. Time scale is in calendar years. The median growth rate (with the corresponding 95% highest probability density interval in parenthesis) estimated for each clade under the logistic growth model is indicated in the upper left corner. $B_{PANDEMIC}$, subtype B pandemic.

Eight major $B_{PANDEMIC}$ Latin American clades here identified correspond to country-specific lineages that comprise between 2 and 51% of subtype B sequences from a given Latin American country. The remaining four $B_{PANDEMIC}$ Latin American clades identified were part of two major regional lineages: clade B_{CAM} that comprises 91 and 41% of subtype B sequences from Honduras and El Salvador, respectively, and clade B_{SAM} that comprises 18 and 15% of subtype B sequences from Brazil and Argentina, respectively. Phylogenetic tree topologies indicate that clades B_{CAM} and B_{SAM} probably arose in Honduras and Brazil, respectively, and were later spread at multiple times to the corresponding neighboring countries. According to Murillo et al. [11], the clade B_{CAM} also circulates at lower prevalence in Belize (24%), Nicaragua (6%), and Costa Rica (5%). These findings support that dissemination of major $B_{PANDEMIC}$ viruses in Latin America is mainly driven by within countries and among neighboring countries transmissions, but not by transmissions among countries without a common border.

The proportion of HIV-1 subtype B infections that resulted from the expansion of major local lineages in Argentina (31%), Brazil (31%), Mexico (37%), El Salvador (41%), Peru (51%), and Honduras (91%) was comparable with that previously estimated for subtype B epidemics in Jamaica (53%), Cuba (70%), Panama (77%), and Trinidad and Tobago (94%) [5,12,24]; but higher than that previously observed for subtype B epidemic in the United Kingdom (14%) [25]. This supports a much greater geographic compartmentalization of the HIV-1 subtype B epidemic in Latin American and the Caribbean regions than in Europe, which could be associated to intraregional differences in human mobility and spatial accessibility. The only exception to this trend was Venezuela where no large $B_{PANDEMIC}$ clusters were detected in the general population.

According to our estimations, the oldest local $B_{PANDEMIC}$ lineages from

South and Central America arose between the late 1960s and the middle 1970s and by when the AIDS epidemic was first recognized in the early 1980s, the $B_{PANDEMIC}$ clade had already reached and established in most Latin American countries analyzed. This timeframe agrees with previous studies that pointed the arrival of the $B_{PANDEMIC}$ clade in South [9,10] and Central [11,12] America between the middle 1960s and the early 1970s, and also coincides with the estimated age of the $USB_{PANDEMIC}$ clade (1966–1972) [6,26], thus supporting a very fast dissemination of this viral clade in the continent. Notably, some nonpandemic subtype B clades of Caribbean origin were also probably introduced in Latin America between the early 1970s and the early 1980s, but displayed a much more restricted local dissemination [8,12]. This indicates that the extraordinary epidemic outcomes of some $B_{PANDEMIC}$ lineages in Latin America is not associated to their earlier introduction into the region, but probably to their introduction into highly connected transmission networks.

The phylodynamic patterns herein inferred for all major Latin American $B_{PANDEMIC}$ clades support an initial exponential growth and a recent phase of epidemic stabilization. It was demonstrated that the flattening of the BSP often coincides with the time of the most recent coalescent event in the maximum clade credibility tree, and may not reflect a real epidemic stabilization [27]. The leveling of the BSP in our analyses, however, occurred before the flattening of the cumulative number of lineages-through-time for all Latin American $B_{PANDEMIC}$ clades (Fig. S3.8), supporting that the observed slowing of epidemic spread in the recent past was probably not an artifact of the inference method. These data are also consistent with the Joint United Nations Programme on HIV/AIDS estimations that show that HIV adult prevalence and the number of people newly infected each year in Latin America have remained stable since the early 2000s [1].

Despite the similarity in the overall epidemic curves, the precise dynamics of epidemic stabilization greatly vary across clades. Older clades seem to stabilize earlier but maintained longer periods of exponential growth and reached higher sizes than the more recent ones. The median estimated growth rate also varies across different Latin American $B_{PANDEMIC}$ clades, as was previously observed for both pandemic and nonpandemic subtype B clades circulating in Panama and the Caribbean [5,12,24] (Fig. S3.7). Because differences could be observed for different clades circulating in the same country and for the same clade (B_{SAM} , for example) across different locations, the most probable source of the observed variations are not country-specific prevention programmes or viral genetics properties, but the characteristics of the contact networks. Older subtype B clades may have gained access to larger contact networks with slower saturation dynamics [28], and clades predominantly transmitted within MSM and IDUs networks (with high rates of partner exchange) may have faster growth rates than those mainly transmitted within heterosexual chains (Fig. S3.7).

In conclusion, this study reveals that about one-third of HIV-1 subtype B infections in Latin America resulted from the expansion of a few $B_{PANDEMIC}$ founder strains probably introduced at a very early stage of the HIV epidemic in the Americas. All major Latin American $B_{PANDEMIC}$ clades seem to have experienced roughly similar populations dynamics with evidence for recent stabilization, although some variation was detected in the specific stabilization time and growth rate across different lineages. Investigation of the epidemiological characteristics of the transmission networks that drives the dissemination of major Latin American $B_{PANDEMIC}$ clades will be of paramount importance to design more effective HIV prevention strategies in the region.

3.6 References

1. UNAIDS. Report on the global AIDS epidemic. 2013. http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf [Accessed April 2015].

2. UNAIDS. Consolidated Regional Analysis of the UNGASS Reports Presented by 17 Latin American Countries in 2010.

3. De Boni R, Veloso VG, Grinsztejn B. Epidemiology of HIV in Latin America and the Caribbean. Curr Opin HIV AIDS 2014; 9:192–198.

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.

5. Cabello M, Mendoza Y, Bello G. Spatiotemporal dynamics of dissemination of nonpandemic HIV-1 subtype B clades in the Caribbean region. PLoS One 2014; 9:e106045.

6. Gilbert MT, 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; 104:18566–18570.

7. Junqueira DM, de Medeiros RM, Matte MC, Araujo LA, Chies JA, Ashton-Prolla P, et al. Reviewing the history of HIV-1: spread of subtype B in the Americas. PLoS One 2011; 6:e27489.

8. Cabello M, Junqueira DM, Bello G. Dissemination of nonpandemic Caribbean HIV-1 subtype B clades in Latin America. AIDS 2015; 29:483–492.

9. Bello G, Eyer-Silva WA, Couto-Fernandez JC, Guimaraes ML, Chequer-Fernandez SL, Teixeira SL, et al. Demographic history of HIV-1 subtypes B and F in Brazil. Infect Genet Evol 2007; 7:263–270.

10. Dilernia DA, Jones LR, Pando MA, Rabinovich RD, Damilano GD, Turk G, et al. Analysis of HIV type 1 BF recombinant sequences from South America dates

the origin of CRF12_BF to a recombination event in the 1970s. AIDS Res Hum Retroviruses 2011; 27:569–578.

11. Murillo W, Veras N, Prosperi M, de Rivera IL, Paz-Bailey G, Morales-Miranda S, et al. A single early introduction of HIV-1 subtype B into Central America accounts for most current cases. J Virol 2013; 87:7463–7470.

12. Mendoza Y, Martinez AA, Castillo Mewa J, Gonzalez C, Garcia- Morales C, Avila-Rios S, et al. Human immunodeficiency virus type 1 (HIV-1) subtype B epidemic in Panama is mainly driven by dissemination of country-specific clades. PLoS One 2014; 9:e95360.

13. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997; 25:4876–4882.

14. Thomson MM, Fernandez-Garcia A. Phylogenetic structure in African HIV-1 subtype C revealed by selective sequential pruning. Virology 2011; 415:30–38.

15. 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.

16. 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–W559.

17. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 2002; 161:1307–1320.

18. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 2007; 7:214.

19. Suchard MA, Rambaut A. Many-core algorithms for statistical phylogenetics. Bioinformatics 2009; 25:1370–1376.

20. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 2005; 22:1185–1192.

21. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol Biol Evol 2012; 29:2157–2167.

22. Rangel HR, Maes M, Villalba J, Sulbaran Y, de Waard JH, Bello G, et al. Evidence of at least two introductions of HIV-1 in the Amerindian Warao population from Venezuela. PLoS One 2012; 7:e40626.

23. Villalba JA, Bello G, Maes M, Sulbaran YF, Garzaro D, Loureiro CL, et al. HIV-1 epidemic in Warao Amerindians from Venezuela: spatial phylodynamics and epidemiological patterns. AIDS 2013; 27:1783–1791.

24. Delatorre E, Bello G. Phylodynamics of the HIV-1 epidemic in Cuba. PLoS One 2013; 8:e72448.

25. 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.

26. Robbins KE, Lemey P, Pybus OG, Jaffe HW, Youngpairoj AS, Brown TM, et al. U.S. Human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. J Virol 2003; 77:6359–6366.

27. de Silva E, Ferguson NM, Fraser C. Inferring pandemic growth rates from sequence data. J R Soc Interface 2012; 9:1797–1808.

28. Blower S. Behaviour change and stabilization of seroprevalence levels in communities of injecting drug users: correlation or causation? J Acquir Immune Defic Syndr 1991; 4:920–923.

3.7 Supplementary Information

Maximum Likelihood (ML) phylogenetic analysis. ML phylogenetic trees were inferred under the GTR+I+Γ nucleotide substitution model as selected by the jModeltest program [1]. 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) [2] based on the Shimodaira-Hasegawa-like procedure. Because ambiguously positioned taxa may reduce the phylogeographic structure of the HIV-1 phylogenies [3], we use a relatively permissive statistical support (aLRT $\geq$ 0.80) for initial definition of countryspecific subtype B clusters. That criterion becomes more stringent as sequences branching outside of the so defined clusters were progressively removed, and only those clades with aLRT support $\geq$ 0.90 were finally selected. Trees were rooted using subtype D sequences (the closets HIV-1 group M lineage relative to subtype B) taken from the Los Alamos HIV Database and visualized using the FigTree v1.4.0 program [4].

Bayesian MCMC analyses. All Bayesian Markov Chain Monte Carlo (MCMC) analyses were performed using the GTR+I+Γ4 nucleotide substitution model, a relaxed uncorrelated lognormal molecular clock model [5], and an informative prior for substitution rate (1.7-2.5 x 10^{-3} subst./site/year) recovered from previous estimations for the subtype B *pol* gene [6-9]. MCMC chains were run for 10-100 x 10^7 generations. Convergence and uncertainty of parameter estimates were assessed by calculating the Effective Sample Size (ESS) and the 95% Highest Probability Density (HPD) values, respectively, after excluding the initial 10% of each run with Tracer v1.6 [10]. Graphical representations of the effective number of infection through time and the number of lineages-through-time (LTT) were generated with R statistical software package.

3.7.1 Supplementary References

1. Posada D. jModelTest: phylogenetic model averaging. Mol Biol Evol 2008,25:1253-1256.

2. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol 2006,55:539-552.

3. Thomson MM, Fernandez-Garcia A. Phylogenetic structure in African HIV-1 subtype C revealed by selective sequential pruning. Virology 2011,415:30-38.

4. Rambaut A. FigTree v1.4: Tree Figure Drawing Tool. Available from http://tree.bio.ed.ac.uk/software/figtree/ 2009.

5. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol 2006,4:e88.

6. 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.

7. 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.

8. Chen JH, Wong KH, Chan KC, To SW, Chen Z, Yam WC. Phylodynamics of HIV-1 subtype B among the men-having-sex-with-men (MSM) population in Hong Kong. PLoS ONE 2011,6:e25286.

9. Mendoza Y, Martinez AA, Castillo Mewa J, Gonzalez C, Garcia-Morales C, Avila-Rios S, et al. Human Immunodeficiency Virus Type 1 (HIV-1) Subtype B Epidemic in Panama Is Mainly Driven by Dissemination of Country-Specific Clades. PLoS ONE 2014,9:e95360.

10. Rambaut A, Drummond A. Tracer v1.6. Available from http://tree.bio.ed.ac.uk/software/tracer/ 2007.

Table S 3.1: HIV-1 subtype B *pol* sequences from Latin America, the Caribbean, US and France used for ML phylogenetic analyses.

Region	Country	*N*	Sampling time
Latin America	Argentina	1,351	1998-2009
	Brazil	2,480	1990-2010
	El Salvador	170	2008-2010
	Honduras	507	2001-2009
	Peru	249	2003-2010
	Mexico	1,647	2004-2010
	Venezuela	385	2004-2011
North America	US	300	1997-2009
Europe	France	200	1985-2008

Table S 3.2: Major HIV-1 $B_{PANDEMIC}$ clades identified in Latin America after analysis of each country separately.

Country	Clade	N (%)[a]	Proportion[b]	Sampling dates
Argentina	$B_{AR\text{-}I}$	199 (15%)	99,00%	2000-2008
	$B_{AR\text{-}II}$	142 (11%)	98,00%	2001-2007
	$B_{AR\text{-}III}$	65 (5%)	100,00%	2001-2007
Brazil	$B_{BR\text{-}I}$	445 (18%)	98,00%	1997-2010
	$B_{BR\text{-}II}$	174 (7%)	98,00%	1997-2010
	$B_{BR\text{-}III}$	96 (4%)	100,00%	1997-2009
	$B_{BR\text{-}IV}$	43 (2%)	100,00%	1997-2009
El Salvador	B_{SV}	69 (41%)	95,00%	2008-2010
Honduras	B_{HN}	462 (91%)	99,00%	2001-2007
Mexico	$B_{MX\text{-}I}$	406 (25%)	97,00%	2004-2010
	$B_{MX\text{-}II}$	196 (12%)	99,00%	2005-2010
Peru	B_{PE}	126 (51%)	99,00%	2006-2010

a Number of Latin American sequences within each major $B_{PANDEMIC}$ clade and percentage of the total number of sequences from each country included in the study: Argentina = 1,351, Brazil = 2,480, El Salvador = 170, Honduras = 507, Mexico = 1,647, and Peru = 249. b Proportion of the total number of sequences within each major $B_{PANDEMIC}$ clade that correspond to a given Latin American country.

Table S 3.3: Best fit demographic model for major HIV-1 $B_{PANDEMIC}$ Latin American clades.

Clade	Model	PSLog ML	Models compared	Log BF	SSLog ML	Models compared	Log BF
B_{SAM-AR}	Log	-9384.16	-	-	-9384.79	-	-
	Expo	-9445.61	Log/Expo	61.45	-9446.01	Log/Expo	61.22
	Expa	-9451.39	Log/Expa	67.23	-9451.61	Log/Expa	66.82
B_{AR-II}	Log	-14433.16	-	-	-14436.00	-	-
	Expo	-14490.42	Log/Expo	57.26	-14492.08	Log/Expo	56.08
	Expa	-14491.81	Log/Expa	58.65	-14493.05	Log/Expa	57.05
B_{AR-III}	Log	-7933.03	-	-	-7933.05	-	-
	Expo	-7954.8	Log/ Expo	21.77	-7955.19	Log/ Expo	22.14
	Expa	-7963.61	Log/Expa	30.58	-7963.92	Log/Expa	30.87
B_{SAM-BR}	Log	-47286.45	-	-	-47310.77	-	-
	Expo	-47527.70	Log/Expo	241.25	-47548.31	Log/Expo	237.54
	Expa	-47557.28	Log/Expa	270.83	-47573.62	Log/Expa	262.85
B_{BR-II}	Log	-20646.60	-	-	-20651.19	-	-
	Expo	-20692.17	Log/Expo	45.57	-20694.97	Log/Expo	43.78
	Expa	-20704.08	Log/Expa	57.48	-20705.78	Log/Expa	54.59
B_{BR-III}	Log	-11978.26	-	-	-11979.36	-	-
	Expo	-12022.05	Log/Expo	43.79	-12022.30	Log/Expo	42.94
	Expa	-12032.57	Log/Expa	54.31	-12033.17	Log/Expa	53.81
B_{BR-IV}	Log	-6411.89	-	-	-6412.27	-	-
	Expo	-6427.87	Log/Expo	15.98	-6428.09	Log/Expo	15.82
	Expa	-6436.06	Log/Expa	24.17	-6436.36	Log/Expa	24.09
B_{CAM-SV}	Log	-6165.89	-	-	-6166.20	-	-
	Expo	-6169.32	Log/Expo	3.43	-6169.64	Log/Expo	3.44
	Expa	-6175.59	Log/Expa	9.7	-6175.66	Log/Expa	9.46
B_{CAM-HN}	Log	-41727.83	-	-	-41753.80	-	-
	Expo	-42070.40	Log/Expo	342.57	-42090.05	Log/Expo	336.25
	Expa	-42062.25	Log/Expa	334.42	-42083.75	Log/Expa	329.95
B_{MX-I}	Log	-33110.35	-	-	-33133.66	-	-
	Expo	-33259.09	Log/Expo	148.74	-33279.31	Log/Expo	145.65
	Expa	-33264.18	Log/Expa	153.83	-33284.78	Log/Expa	151.12
B_{MX-II}	Log	-18972.62	-	-	-18978.04	-	-
	Expo	-19043.66	Log/Expo	71.04	-19047.36	Log/Expo	69.36
	Expa	-19052.33	Log/Expa	79.71	-19055.97	Log/Expa	77.93
B_{PE}	Log	-14256.71	-	-	-14258.72	-	-
	Expo	-14330.68	Log/Expo	73.97	-14332.17	Log/Expo	73.45
	Expa	-14336.29	Log/Expa	79.58	-14337.61	Log/Expa	78.89

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 > 1 indicates that model H1 is more strongly supported by the data than model H0.

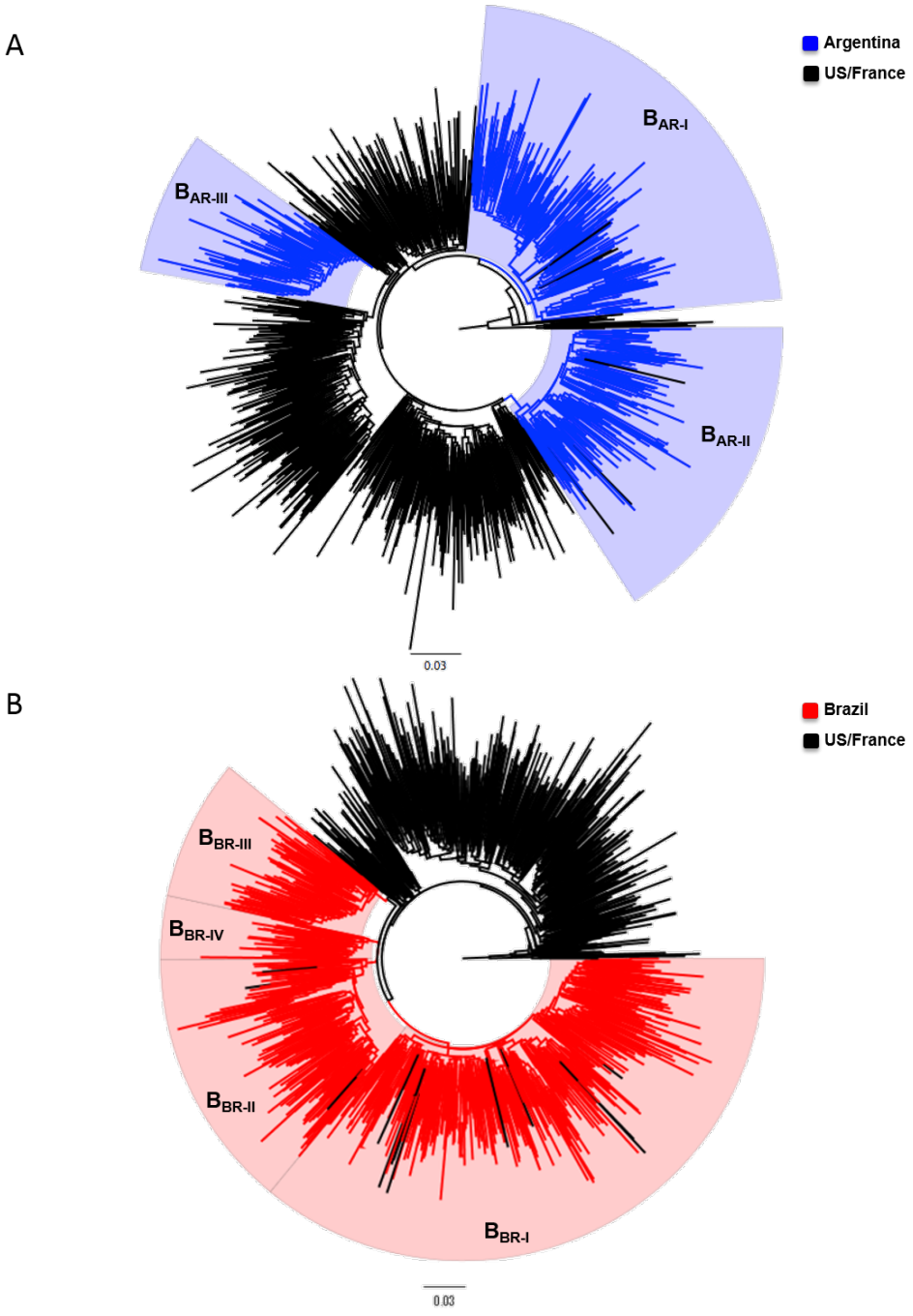


Figure S 3.1: ML phylogenetic tree of HIV-1 subtype B *pol* PR/RT sequences (∼1,000 nt) from major $B_{PANDEMIC}$ lineages circulating in Argentina (A) and Brazil (B) combined with reference sequences representative of the $B_{PANDEMIC}$ lineages circulating in the US (n = 300), and France (n = 200). Branches are colored according to the geographic origin of each sequence as indicated at the legend (bottom right). Positions of major Latin American $B_{PANDEMIC}$ clades are indicated by colored shaded boxes. Trees were rooted using HIV-1 subtype D reference sequences. The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.

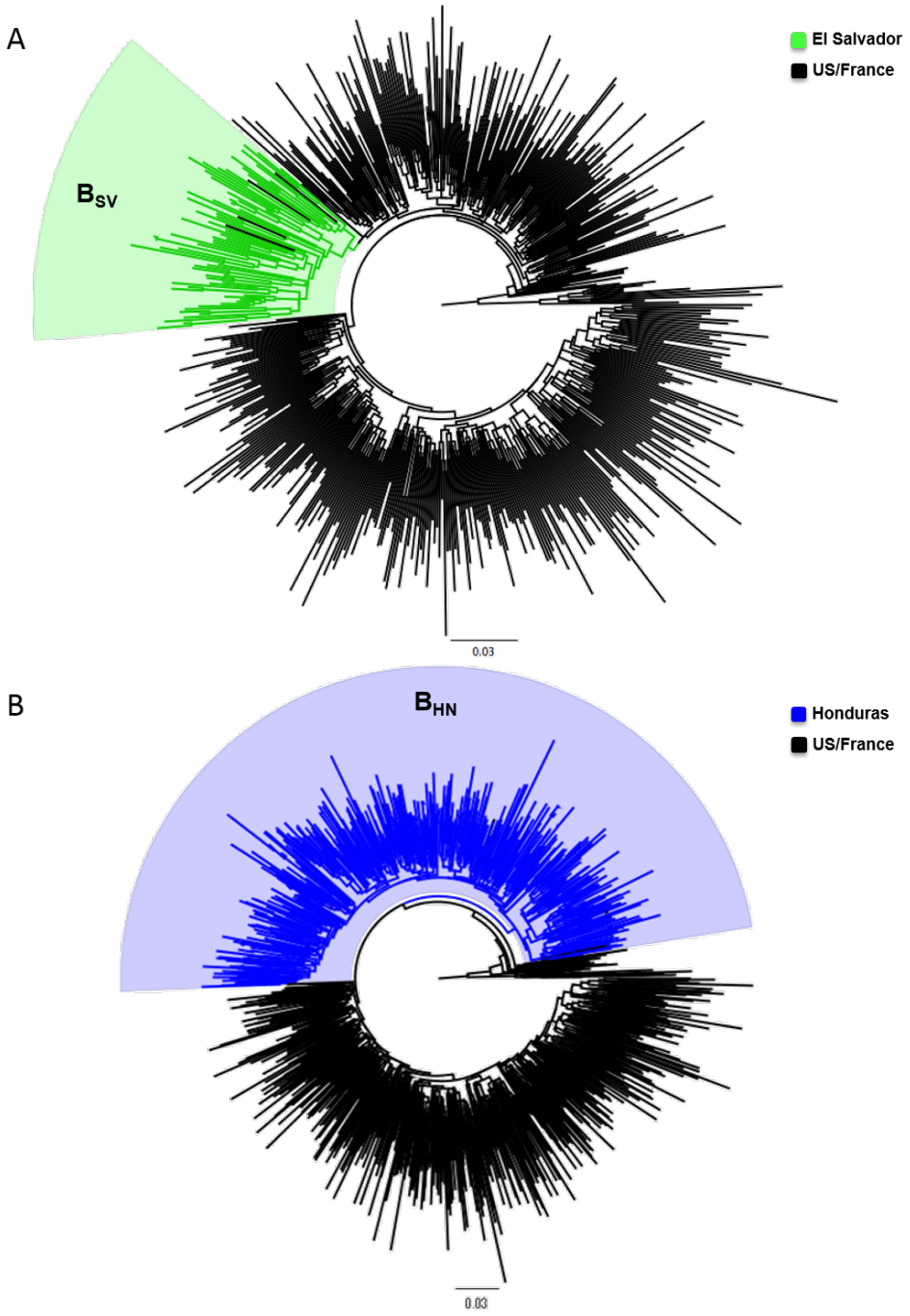


Figure S 3.2: ML phylogenetic tree of HIV1 subtype B *pol* PR/RT sequences (∼1,000 nt) from major $B_{PANDEMIC}$ lineages circulating in El Salvador (A), Honduras (B) combined with reference sequences representative of the $B_{PANDEMIC}$ lineages circulating in the US (n = 300), and France (n = 200). Branches are colored according to the geographic origin of each sequence as indicated at the legend (bottom right). Positions of major Latin American $B_{PANDEMIC}$ clades are indicated by colored shaded boxes. Trees were rooted using HIV-1 subtype D reference sequences. The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.

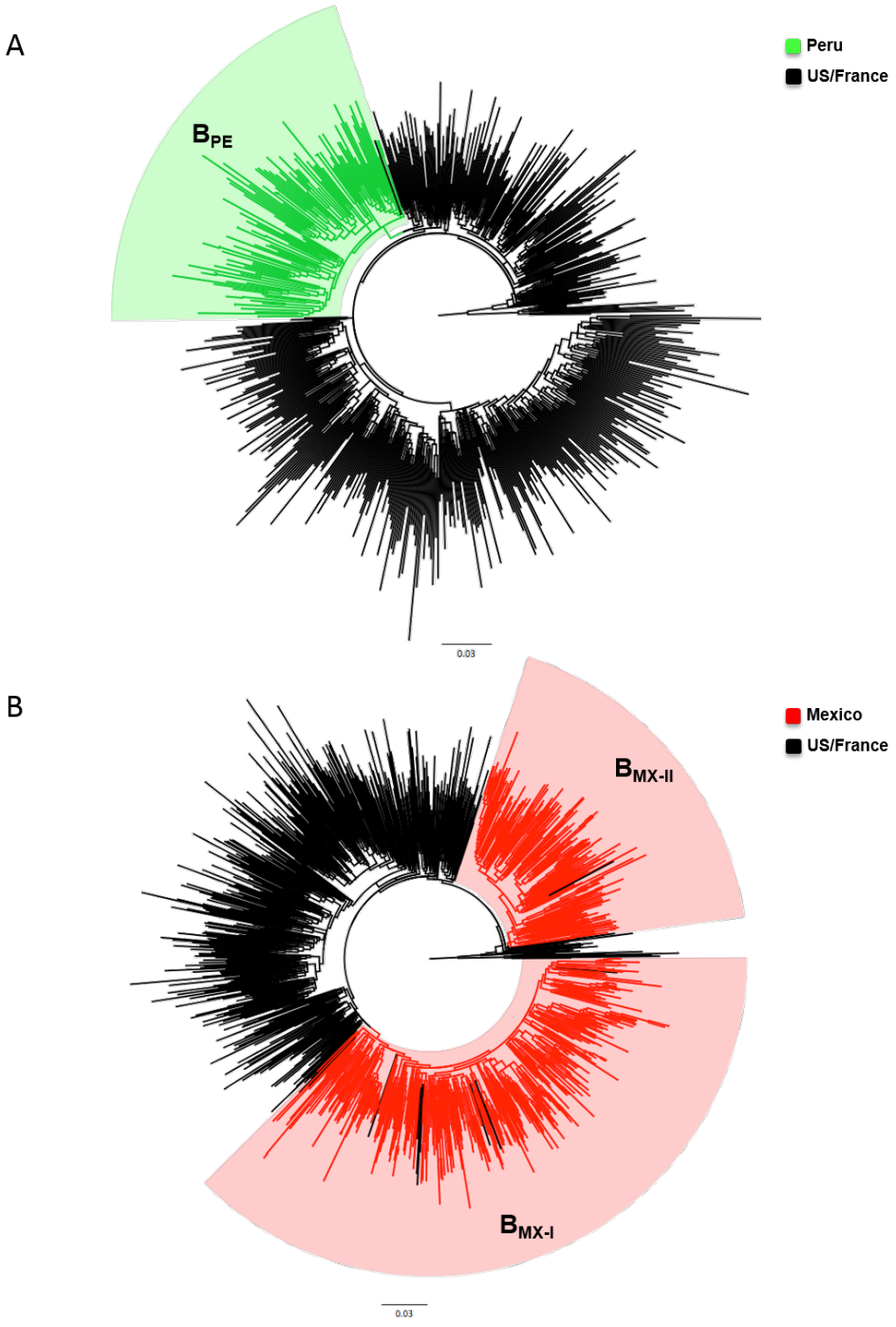


Figure S 3.3: ML phylogenetic tree of HIV-1 subtype B *pol* PR/RT sequences (∼1,000 nt) from major $B_{PANDEMIC}$ lineages circulating in Peru (A) and Mexico (B) combined with reference sequences representative of the $B_{PANDEMIC}$ lineages circulating in the US (n = 300), and France (n = 200). Branches are colored according to the geographic origin of each sequence as indicated at the legend (bottom right). Positions of major Latin American $B_{PANDEMIC}$ clades are indicated by colored shaded boxes. Trees were rooted using HIV-1 subtype D reference sequences. The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.

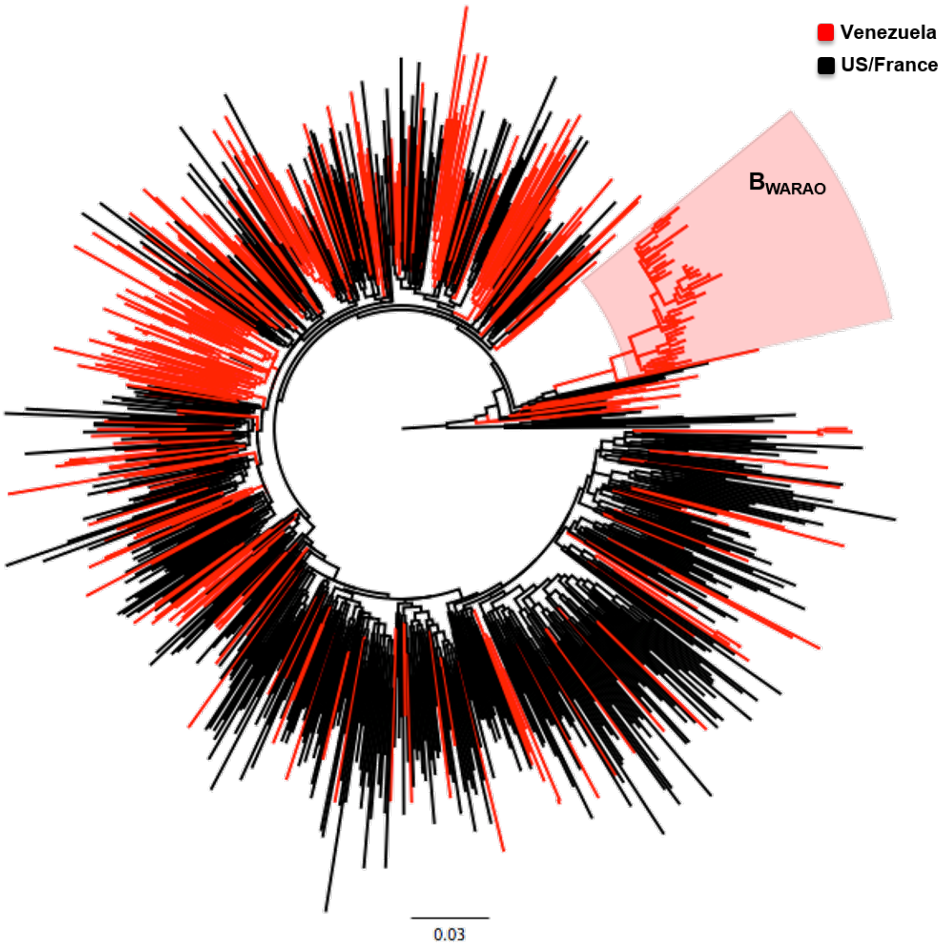


Figure S 3.4: ML phylogenetic tree of HIV-1 subtype B *pol* PR/RT sequences (∼1,000 nt) circulating in Venezuela (n = 909) and representative sequences of the $B_{PANDEMIC}$ (US = 165, France = 135) clade. Branches are colored according to the geographic origin of each sequence as indicated at the legend (bottom right). The position of the only major (n > 40) strongly supported (aLRT > 0.9) Venezuelan $B_{PANDEMIC}$ clade identified is indicated by a colored shaded box and was associated to the Warao Ameridian community. Tree was rooted using HIV-1 subtype D reference sequences. The branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.

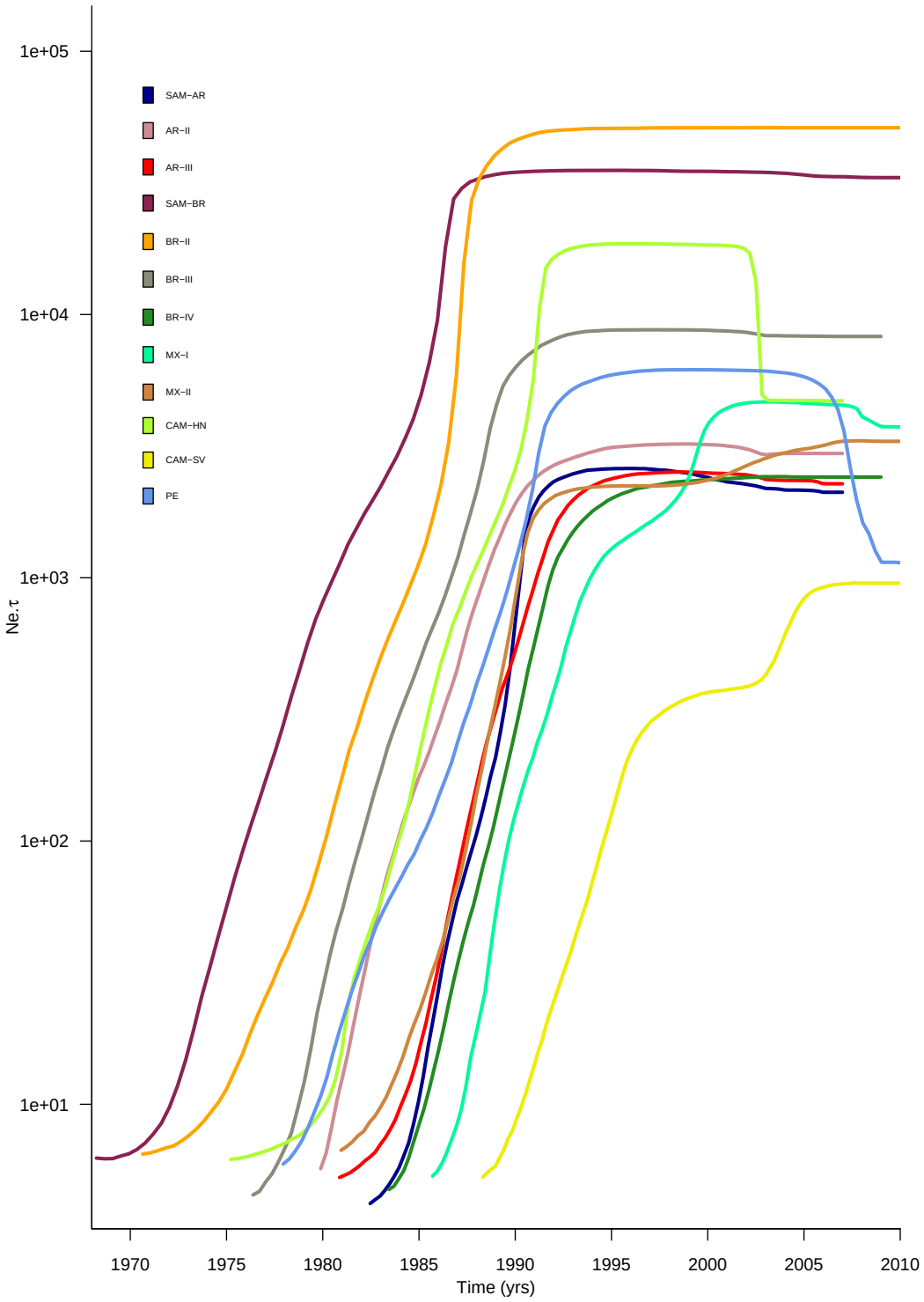


Figure S 3.5: Bayesian skyline estimates of the median *Ne* (y axis, logarithmic scale) over time (x axis, calendar years) obtained for the major HIV-1 $B_{PANDEMIC}$ Latin American clades. Each clade is represented by a color as indicated in the legend at the upper left corner.

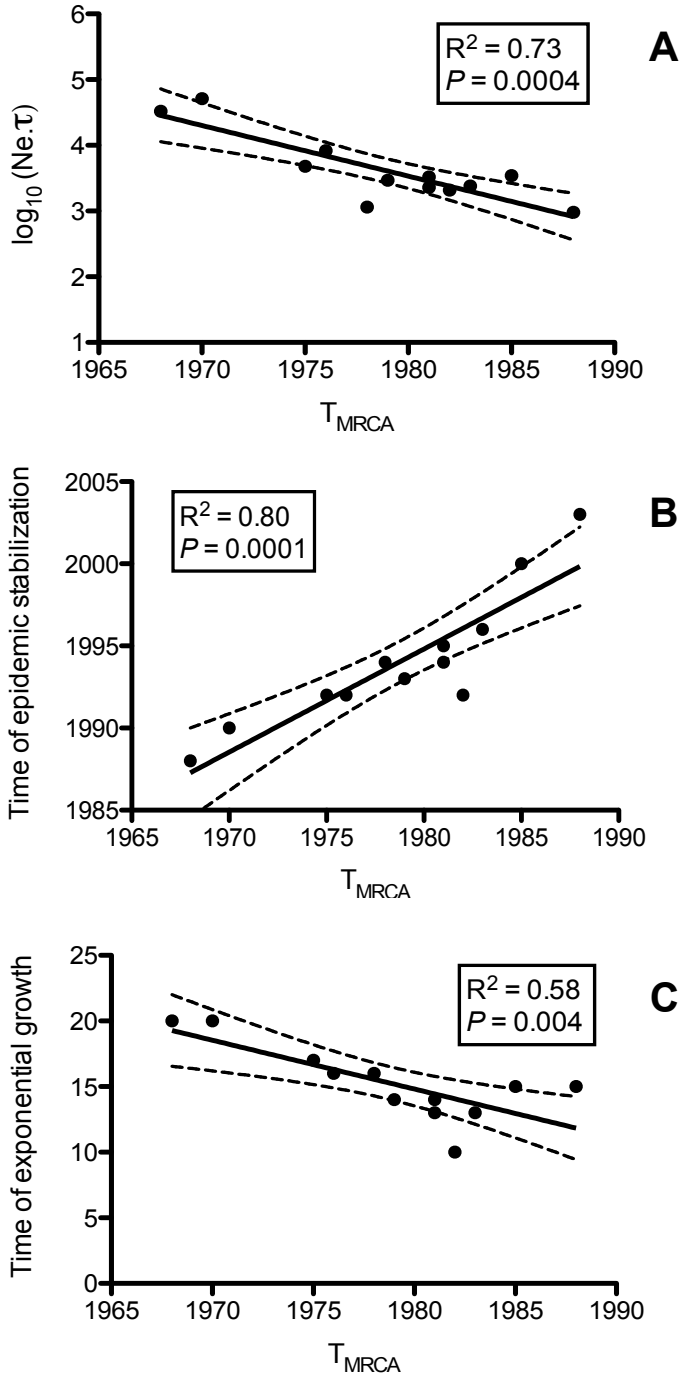


Figure S 3.6: Correlation between the median estimated tMRCA and the corresponding median effective population size (*Ne*) at the most recent time (A), transition year of epidemic stabilization (B) and total length time (years) of exponential growth of each major HIV-1 $B_{PANDEMIC}$ Latin American clade (C). Correlations were adjusted to a linear regression curve and the model fit to the data (R^2 value) as well the probability for slope deviation from zero is indicated in each graph.

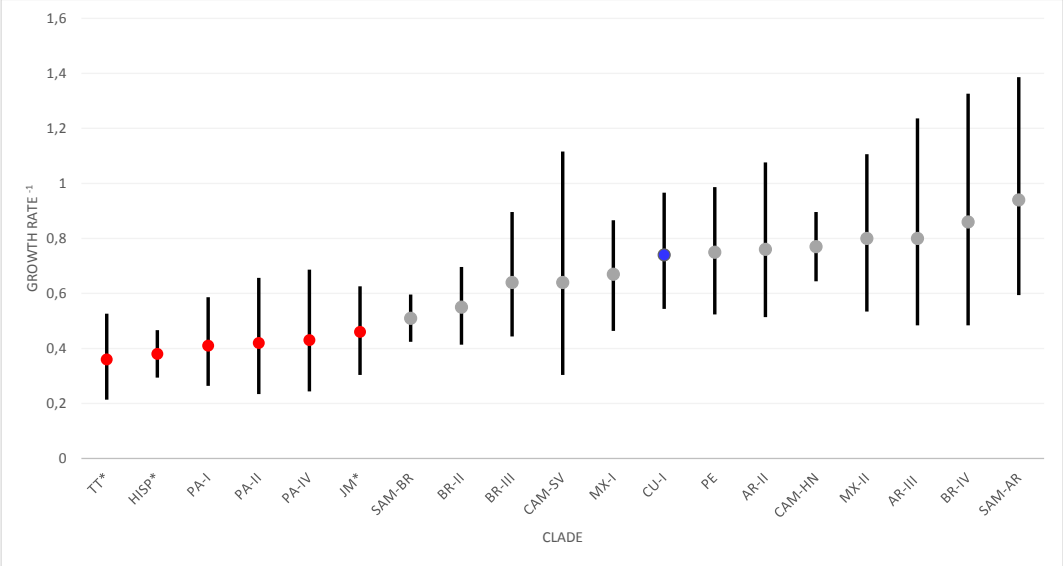


Figure S 3.7: Coalescent estimates of epidemic growth rate of major HIV-1 $B_{PANDEMIC}$ and $B_{NONPANDEMIC}$ (indicated with an asterisk) clades from Latin America and the Caribbean. The circle and the vertical lines represent the median growth rates (years^{-1}) and the corresponding 95% HPD intervals of the posterior distributions estimated under the logistic growth coalescent model for each clade. Subtype B clades mostly associated to heterosexual (red circles), MSM (blue circles) or unknown (green circles) transmission networks are indicated. Epidemic growth rates of subtype B clades circulating in Cuba (CU), Hispaniola (HISP), Jamaica (JM), Trinidad and Tobago (TT) and Panama (PA) were estimated previously in Delatorre et al. 2013, Cabello et al. 2014 and Mendoza et al. 2014.

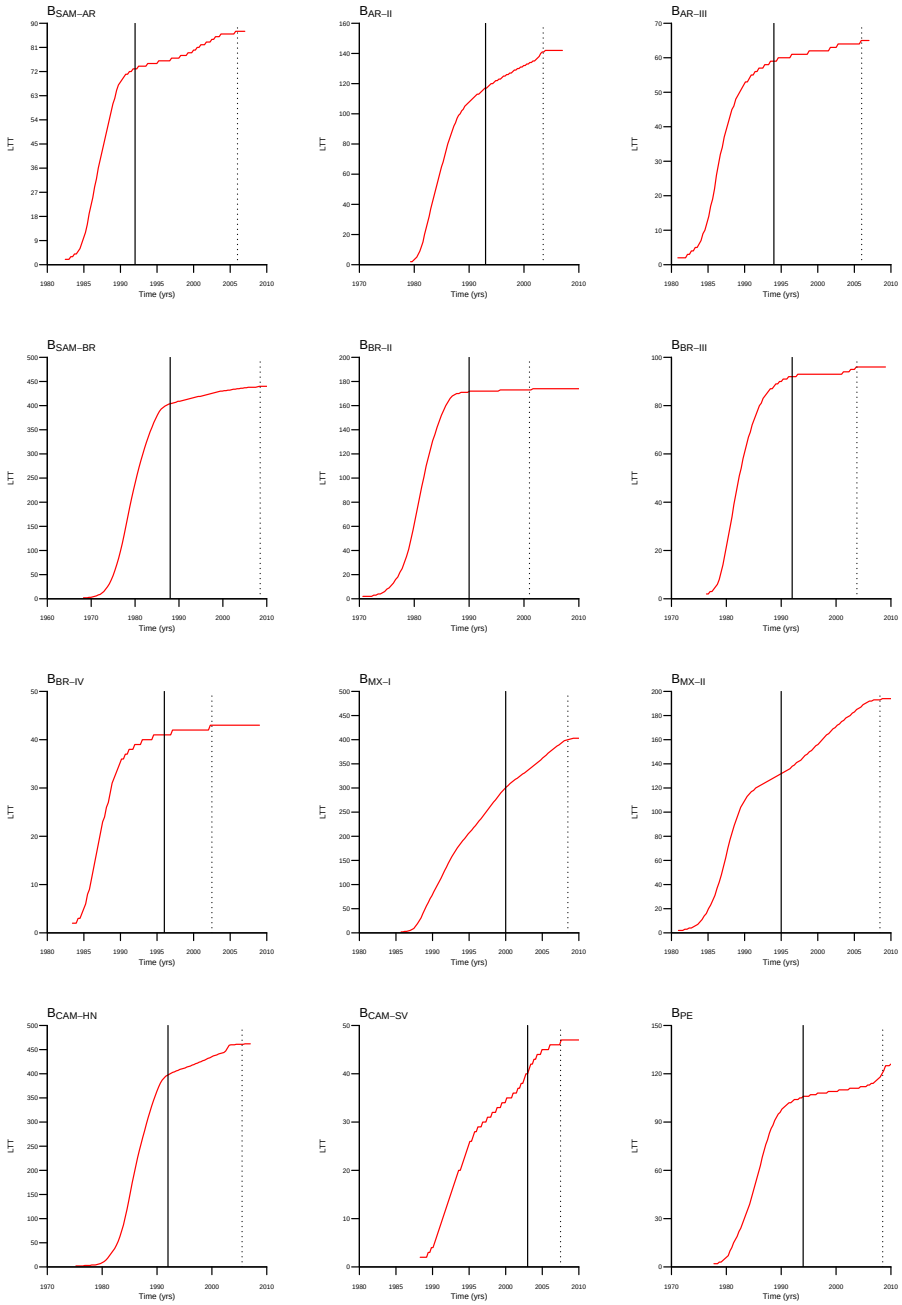


Figure S 3.8: Cumulative number of lineages (y axis) through time (x axis) (LTT) for the major HIV-1 $B_{PANDEMIC}$ Latin American clades. The vertical solid line represents the time of BSP stabilization for each clade. The vertical dashed line represents the time of the most recent coalescent event for each clade.

Chapter 4

Phylodynamics of the major HIV-1 CRF02_AG African lineages and its global dissemination.

Article published in:

Daiana Mir, Matthieu Jung, Edson Delatorre, Nicole Vidal, Martine Peeters and Gonzalo Bello.

Infection, Genetics and Evolution 2016; 46:190-199.

4.1 Abstract

The HIV-1 CRF02_AG clade is the most prevalent HIV variant in West and West-Central Africa and its detection outside Africa is increasingly common. Little is known, however, about the number and phylodynamics of major CRF02_AG lineages circulating worldwide. To this end, a total of 3170 HIV-1 CRF02_AG-like *pol* sequences isolated around the world, over a period of 25 years (1989 to 2013), were analyzed using Maximum Likelihood and Bayesian coalescent-based methods. Our results suggest that most of the current CRF02_AG diversity comes from the dissemination of a few founder strains out of Central Africa into West Africa and Cameroon between the late 1960s and the middle 1980s. The CRF02_AG strain introduced into West Africa established a large regional epidemic with low phylogeographic structure. This strain was also successfully disseminated out of the West African region and originated at least three large secondary outbreaks in Cameroon at around the late 1970s, in the former Soviet Union (FSU) countries at around the late 1990s, and in Bulgaria/Germany at around the early 2000s. The CRF02_AG African lineages introduced into Cameroon remained mostly restricted to this country and its neighbors. Demographic reconstructions indicate that major CRF02_AG clades circulating in Africa exhibited a decline in growth rate since the middle 1980s/1990s, whereas CRF02_AG clades in Europe and the FSU countries continue to grow exponentially until the middle to late 2000s. Substantial differences in the median estimated growth rate of the same CRF02_AG clade circulating in different regions (0.63–2.00 $year^{-1}$), and of different CRF02_AG clades circulating in the same country (0.41–0.75 $year^{-1}$) were observed. Thus, the cause of the epidemic outcome of the different HIV-1 CRF02_AG lineages is probably multifactorial.

4.2 Introduction

The epidemic dispersion of HIV-1 group M from its location root in Kinshasa, capital of the DRC, since the first part of the 20th century [1] has resulted in the extensive diversity of subtypes, sub-subtypes, circulating recombinant forms (CRFs) and unique recombinant forms (URFs) reported across the world. Among the 79 CRFs currently described, CRF02_AG is responsible for the largest number of infections worldwide and is the fourth most prevalent HIV-1 variant accounting for 8% of the global infections [2].

The CRF02_AG variant predominates in West and West-Central African countries where it stand for about 50% of the HIV-1 infections [2,3]; but there is a notable decrease in its frequencies toward Central Africa where it display a prevalence of around 8% of the total infections [2,4] and is rarely detected in other African regions. In recent years, there has been an increase in reported sporadic CRF02_AG cases in Europe and North America, mostly caused by migrant flows from endemic regions and global travel [5–8].

A few indigenous transmission networks of CRF02_AG have been also detected in different regions out of Africa. Autochthonous transmission networks of CRF02_AG have arisen in former Soviet Union (FSU) countries mainly distributed among IDUs, but with increasing prevalence into heterosexual populations [9–14]. Local dissemination of CRF02_AG has been also detected in Brazil, where it has shown the existence of at least two transmission networks with dissemination by both horizontal and vertical pathways [15–17]. Moreover, recent analysis evidence the existence of some native CRF02_AG transmission clusters particularly among HIV-infected men having sex with men (MSM) from France and Belgium and heterosexual population from Switzerland [18–21].

The field of viral phylodynamics coupled with coalescent-based models

has become a powerful tool allowing the recognition of the spatiotemporal dynamics of a variety of viruses and its implementation helped elucidate the origin and dispersion pattern of the CRF02_AG lineage in the Congo River basin [22], as well as the demographic dynamics and/or migration routes of CRF02_AG circulating in Guinea Bissau, Cameroon and Brazil [16,22–24]. The spatiotemporal pattern of dissemination of the CRF02_AG at a global scale, however, remains largely unknown.

The objectives of the present study were to identify and characterize the major HIV-1 CRF02_AG clades circulating in West, West-Central and Central Africa and their dispersion at both regional and global scales. Spatial and temporal information of 3170 CRF02_AG-like *pol* sequences sampled worldwide over a period of 25 years were used in maximum-likelihood and coalescent-based phylodynamic approaches to determine the prevalence of the major HIV-1 CRF02_AG clades and to reconstruct simultaneously their evolutionary and demographic histories.

4.3 Materials and Methods

4.3.1 HIV-1 CRF02_AG-like *pol* sequence datasets

A total of 2246 HIV-1 CRF02_AG-like *pol* sequences, covering the entire protease and partial reverse transcriptase (PR/RT) regions (nucleotides 2253-3260 relative to HXB2 genome), isolated from 20 countries from Central, West-Central and West Africa over a period of 24 years (1990 to 2013) were used in this study (Table S4.1). Sequences were retrieved from the Los Alamos HIV Database (n = 2113) (http://www.hiv.lanl.gov) and from a local database at the Institut de Recherche pour le Développement, Université Montpellier (n = 133). These African sequences were combined

with 924 CRF02_AG-like *pol* sequences isolated from 43 countries from the Americas, Europe, Asia and Oceania and with 38 CRF63_02A1-like *pol* sequences isolated in Russia, covering the same genomic region described above and that were available at the Los Alamos HIV Database (Table S4.1). The subtype assignment of all sequences was confirmed using COMET [25] and REGAv3.0 [26]. Sequences with discordant results were further submitted to Maximum Likelihood (ML) phylogenetic analysis (see bellow) and Bootscan analysis [27] with reference samples. All sites with major antiretroviral drug resistance mutations were excluded, leaving 921 nucleotides in the final alignment that is available from authors upon request.

4.3.2 Phylogenetic analyses

ML trees were inferred with the PhyML v3.0 program [28], under the GTR + I + Γ4 model of nucleotide substitution recommended by the jModeltest program [29]. The Subtree Pruning and Regrafting (SPR) option was selected as the heuristic tree search method and branch support was estimated with the approximate likelihood-ratio (aLRT) SH-like test [30]. Reference sequences of HIV-1 subtypes B, C, D, F, H, J and K from the Los Alamos HIV Database were used as outgroup. The phylogenetic trees were visualized with FigTree v1.4.2 (`http://tree.bio.ed.ac.uk/software/figtree/`). Major CRF02_AG monophyletic clusters were identified by visual inspection and only those including more than 30 sequences and an aLRT score over 0.85 were selected for further analysis.

4.3.3 Evolutionary and demographic reconstructions

For each CRF02_AG clade identified, the evolutionary rate (μ, units are nucleotide substitutions per site per year, subst./site/year), age of most

recent common ancestor (tMRCA, years), and mode and rate (r, $years^{-1}$) of population growth were coestimated by a Bayesian MCMC coalescent-based phylodynamic analyses as implemented in BEAST v1.8 [31] with BEAGLE [32] to improve run performance. Since this methodology is computationally prohibitive on large datasets, those identified clades made up by more than 500 sequences were subjected to a sub-sampling strategy (see Supplementary Information for full details of the procedure). Analyses were carried out under the GTR + I + Γ4 model of nucleotide substitution and a relaxed uncorrelated lognormal molecular clock model [33]. A uniform prior was applied on the clock rate (1.5-3.0 x 10^{-3} subs./site/year) on the basis of estimations reported from previous studies [34]. The dynamics of the *Ne* over time were initially estimated by the non-parametric Bayesian skyline plot model (BSP) [35] as coalescent tree prior. Parametric estimates of the growth rates were obtained under three demographic models (exponential, logistic and expansion growth) whose adjustment to the data were assessed using the log marginal likelihood estimation (MLE) based on path sampling (PS) and stepping-stone sampling (SS) approaches [36]. MCMC were run for 5-50 x 10^7 generations to ensure Effective Sample Size (ESS) values above 200. The ESS and the 95% Highest Probability Density (HPD) values were inspected using Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/). Maximum clade credibility (MCC) trees were summarized using TreeAnnotator v1.8 and visualized with FigTree v1.4.2.

4.4 Results

4.4.1 Characterization of major HIV-1 CRF02_AG clades circulating in Africa

The ML phylogenetic analysis performed with 2246 HIV-1 CRF02_AG-like *pol* sequences from 20 different African countries revealed the existence of five major strongly supported (aLRT > 0.85) clades within the radiation of the CRF02_AG (Fig. 4.1). The clades $CRF02_{CM\text{-}I}$, $CRF02_{CM\text{-}II}$, $CRF02_{CM\text{-}III}$ and $CRF02_{CM\text{-}IV}$ mostly circulate in Cameroon and together comprise 81% of the CRF02_AG sequences from that country, whereas the clade $CRF02_{WA}$ is the most prevalent one circulating in West Africa and comprises 98% of the CRF02 sequences from that region. Clades $CRF02_{WA}$, $CRF02_{CM\text{-}I}$, $CRF02_{CM\text{-}III}$ and $CRF02_{CM\text{-}IV}$ were nested among basal sequences from Central Africa and probably represent independent introductions of CRF02_AG strains from Central Africa into West Africa and Cameroon. The clade $CRF02_{CM\text{-}II}$, by contrast, was nested within the $CRF02_{WA}$ clade and probably resulted from a secondary dissemination event of a $CRF02_{WA}$ strain from West Africa into Cameroon.

Analysis of the relative prevalence of major CRF02_AG clades across African countries revealed the existence of four different epidemiologic scenarios (Fig. 4.2 and Table S4.2). The CRF02_AG epidemic in Central Africa was mainly composed by basal CRF02_AG lineages (60%), followed by the $CRF02_{WA}$ clade (24%). The CRF02_AG epidemic in West Africa was clearly dominated by the $CRF02_{WA}$ clade (98%). In Gabon and Equatorial Guinea, the $CRF02_{WA}$ clade also predominated ($\geq$ 62%), but a significant fraction of sequences ($\geq$ 14%) branched inside the Cameroonian clades $CRF02_{CM\text{-}I}$ and $CRF02_{CM\text{-}II}$. In Cameroon, the CRF02_AG epidemic was dominated by

clades CRF02$_{\text{CM-I}}$ and CRF02$_{\text{CM-II}}$ (71%), and also by a substantial proportion of strains from CRF02$_{\text{WA}}$ clade (19%). Despite its variable prevalence (ranging from 19% to 100%), the CRF02$_{\text{WA}}$ clade was detected in all African countries analyzed.

4.4.2 Worldwide dissemination of the major HIV-1 CRF02_AG African clades

To investigate the role played by each major African CRF02_AG clade in the global dissemination of CRF02_AG, a worldwide set of CRF02_AG-like *pol* gene sequences was subjected to ML phylogenetic analyses, alongside a reference alignment consisting of randomly selected African sequences representative of the major clades. The phylogenetic reconstruction essentially recovered the five major monophyletic CRF02_AG groups (aLRT > 0.85) previously identified, with global samples branching within them (Fig. 4.3). The level of global geographic dispersion observed among the different African CRF02_AG clades varied widely (Fig. 4.4 and Table S4.2). The majority ($\geq$ 84%) of the CRF02_AG sequences detected in the Americas, Asia and Europe branched inside the CRF02$_{\text{WA}}$ clade. The clades CRF02$_{\text{CM-I}}$ and CRF02$_{\text{CM-II}}$ together comprised between 4% and 17% of the CRF02_AG infections out of Africa, whereas the clades CRF02$_{\text{CM-III}}$ and CRF02$_{\text{CM-IV}}$ were only detected in Europe and at very low prevalence (< 1%).

The ML phylogenetic reconstruction using the global HIV-1 CRF02_AG-like *pol* sequences also revealed two major non-African strongly supported (aLRT $\geq$ 0.99) monophyletic sub-clades within the CRF02$_{\text{WA}}$ radiation (Fig. 4.3). The sub-clade CRF02$_{\text{BG-DE}}$ included sequences sampled from Bulgaria and Germany between 2006 and 2012, whereas the sub-clade CRF02$_{\text{FSU}}$ was composed by sequences from the former Soviet Union (FSU) countries

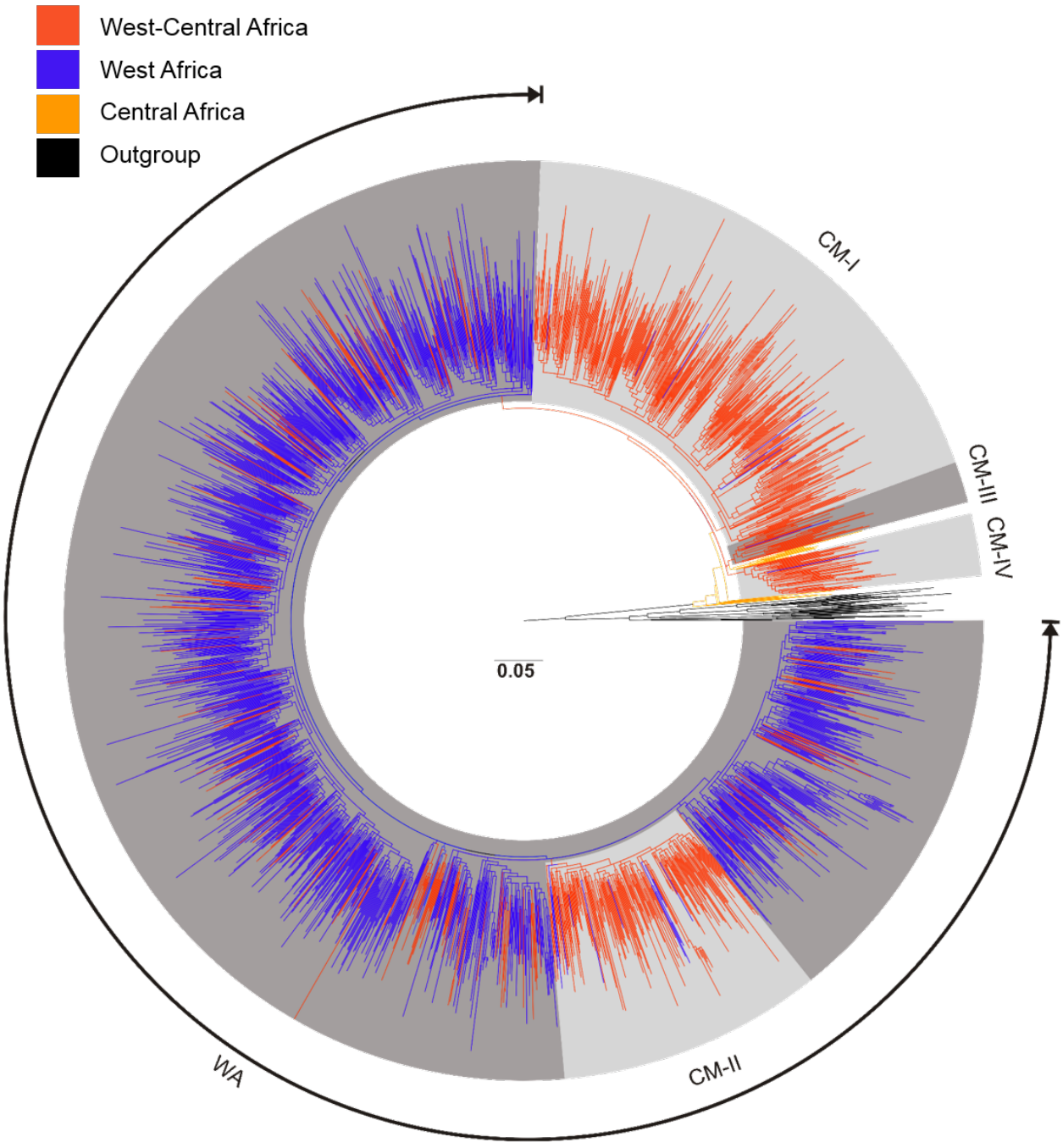


Figure 4.1: ML phylogenetic tree of HIV-1 CRF02_AG-like *pol* PR/RT sequences (∼1,000 nt) from Central, West-Central and West Africa. Branches are colored according to the geographic origin of each sequence as indicated at the legend (top left). Positions of major CRF02_AG African clades are indicated by shaded boxes. The aLRT support value of each identified clade was > 0.85. The tree was rooted using HIV-1 subtypes B, C, D, F, H, J and K reference sequences. The branch lengths are drawn to scale with the bar in the center indicating nucleotide substitutions per site.

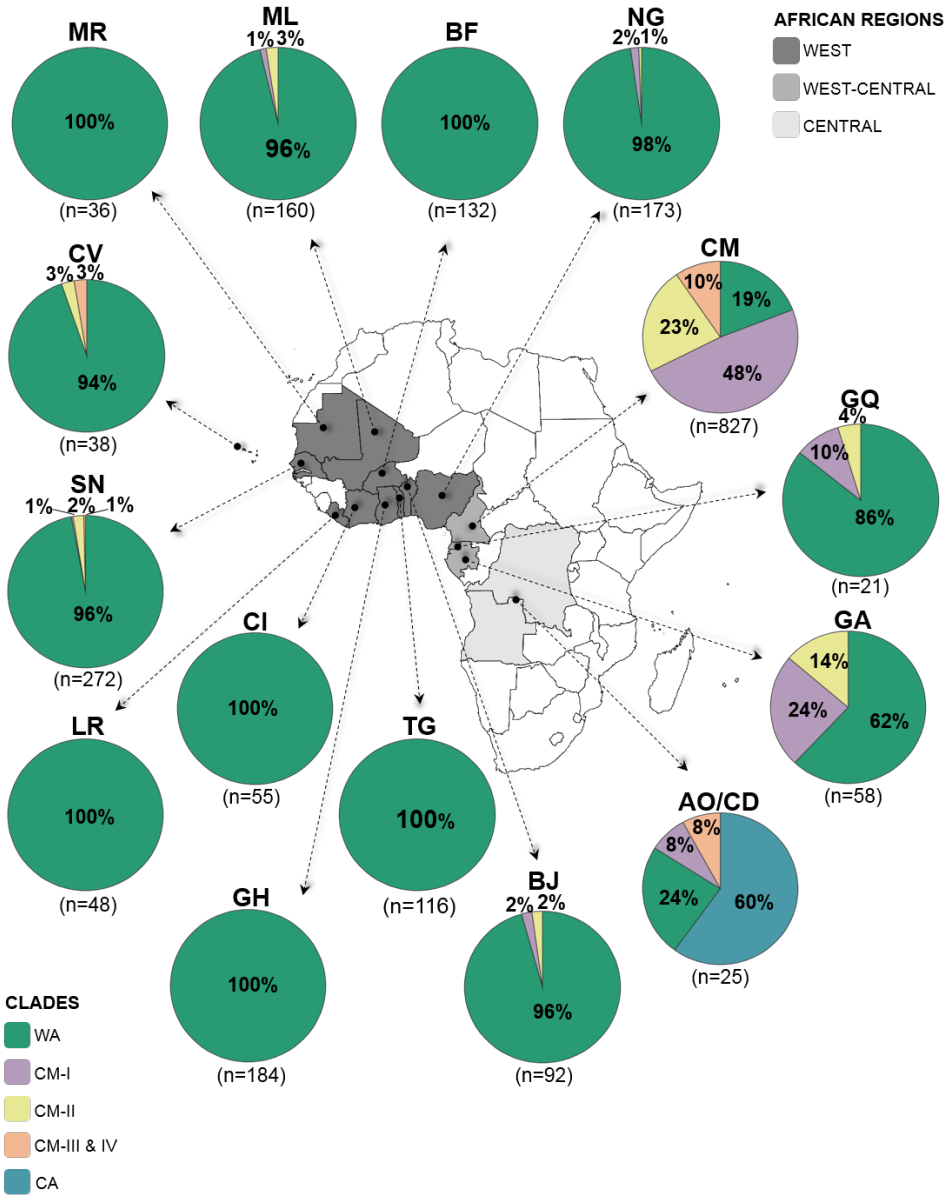


Figure 4.2: Prevalence of CRF02$_{CM-II}$, CRF02$_{CM-II}$, CRF02$_{CM-III}$, CRF02$_{CM-IV}$, CRF02$_{WA}$ clades and basal sequences (CRF02$_{CA}$) among CRF02_AG infected individuals from different African countries, estimated from the phylogenetic analyses presented in Fig. 4.1. The total number of CRF02_AG sequences analyzed in each country is indicated. The CRF02_AG clades and the countries from each African region are represented by a color code as indicated at the legends at bottom. The two-letter country codes are described in Table S4.1

(Russia, Armenia, Kazakhstan, Uzbekistan and Ukraine) isolated between 2002 and 2013. Smaller (n < 20 sequences) country-specific monophyletic sub-clades within the $CRF02_{WA}$ radiation were also detected in others countries (data not shown). These results corroborate the existence of autochthonous transmission networks of CRF02_AG out of Africa that resulted from the introduction and local dissemination of the $CRF02_{WA}$ clade.

4.4.3 Phylogenetic relationship between $CRF02_{FSU}$ and CRF63_02A1 clades

The CRF63_02A1 is a HIV-1 variant mainly spreading among IDUs and heterosexual populations from the Russian Federation that was generated by recombination between the CRF02_AG and the subtype A1 clades [9,12,37,38] This CRF displays a CRF02_AG-like profile in the *pol* gene fragment here selected, which may complicate the subtyping of HIV-1 CRF02_AG-like sequences from FSU countries. To test this, *pol* sequences classified within the $CRF02_{FSU}$ clade were aligned with CRF63_02A1 sequences and with CRF02_AG African sequences representative of the major clades identified in this study. The ML phylogenetic analysis showed that CRF02_AG-like sequences from Armenia, Kazakhstan, Uzbekistan and Ukraine branched at the base of the $CRF02_{FSU}$ clade (Fig. 4.5). Most Russian sequences, by contrast, are intermixed with CRF63_02A1 viruses in a monophyletic subclade nested within basal $CRF02_{FSU}$ lineages and were thus reclassified as CRF63_02A1-like viruses.

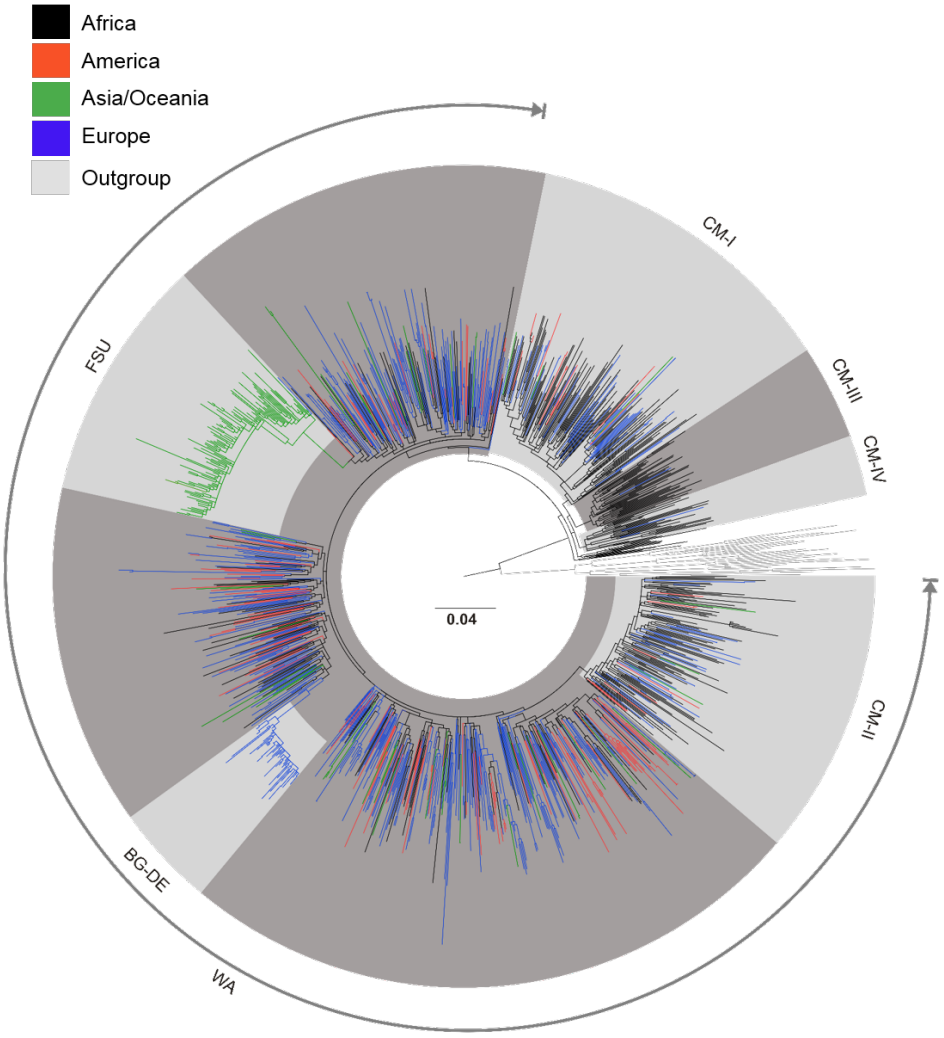


Figure 4.3: ML phylogenetic tree of HIV-1 CRF02_AG-like *pol* PR/RT sequences (∼1,000 nt) isolated around the world alongside reference sequences representative of the major African clades. Branches are colored according to the geographic origin of each sequence as indicated at the legend (top left). Positions of major CRF02_AG clades are indicated by shaded boxes. The aLRT support value of each identified clade was > 0.85. The tree was rooted using HIV-1 subtypes B, C, D, F, H, J and K reference sequences. The branch lengths are drawn to scale with the bar in the center indicating nucleotide substitutions per site.

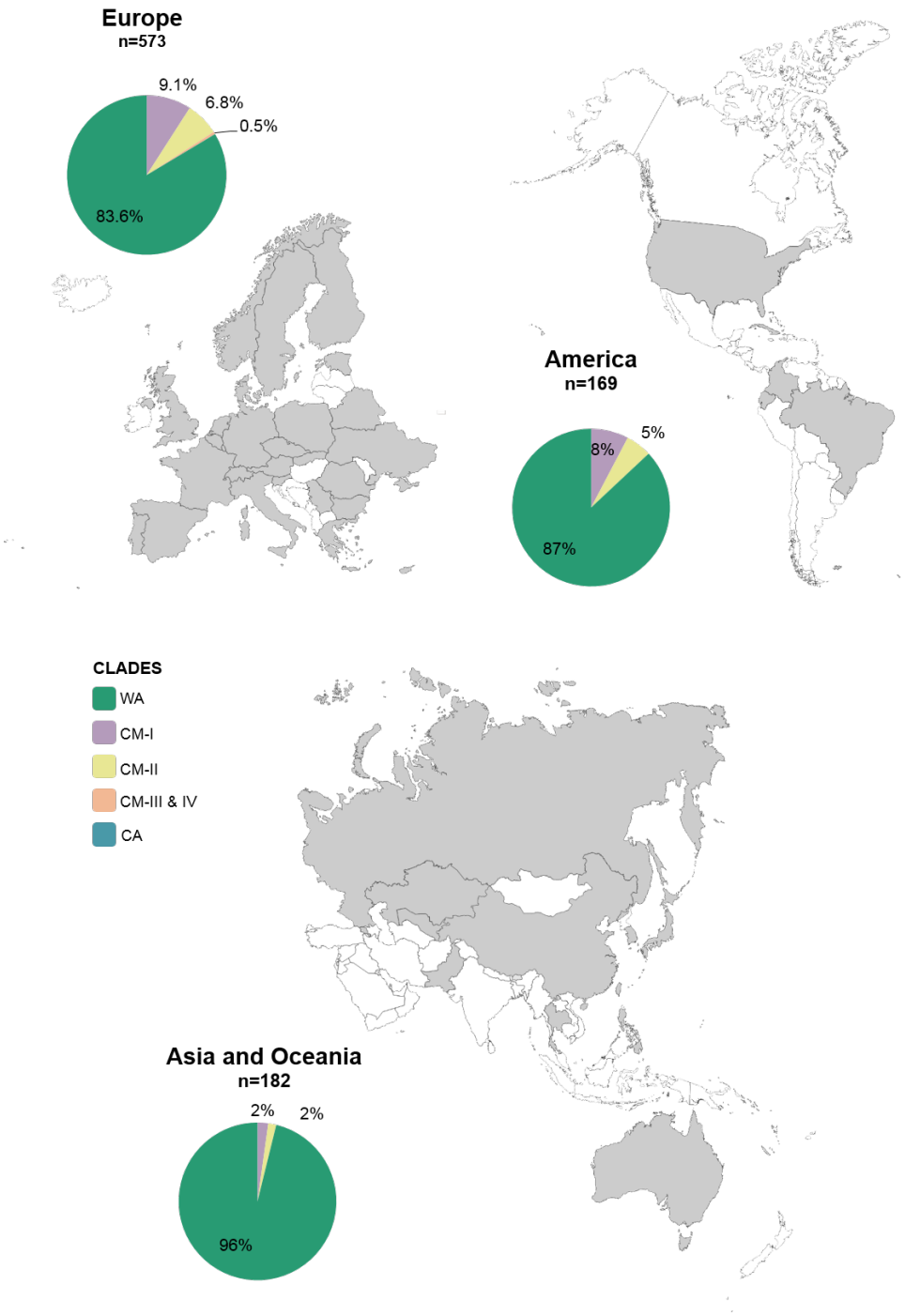


Figure 4.4: Prevalence of $CRF02_{CM\text{-}I}$, $CRF02_{CM\text{-}II}$, $CRF02_{CM\text{-}III}$, $CRF02_{CM\text{-}IV}$ and $CRF02_{WA}$ clades among CRF02_AG infected individuals from different countries out of Africa, estimated from the phylogenetic analyses presented in Fig. 4.3. The total number of CRF02_AG sequences analyzed in each region is indicated. Each CRF02_AG clade is represented by a color as indicated at the legend. The countries of each region included in the analysis are filled in gray.

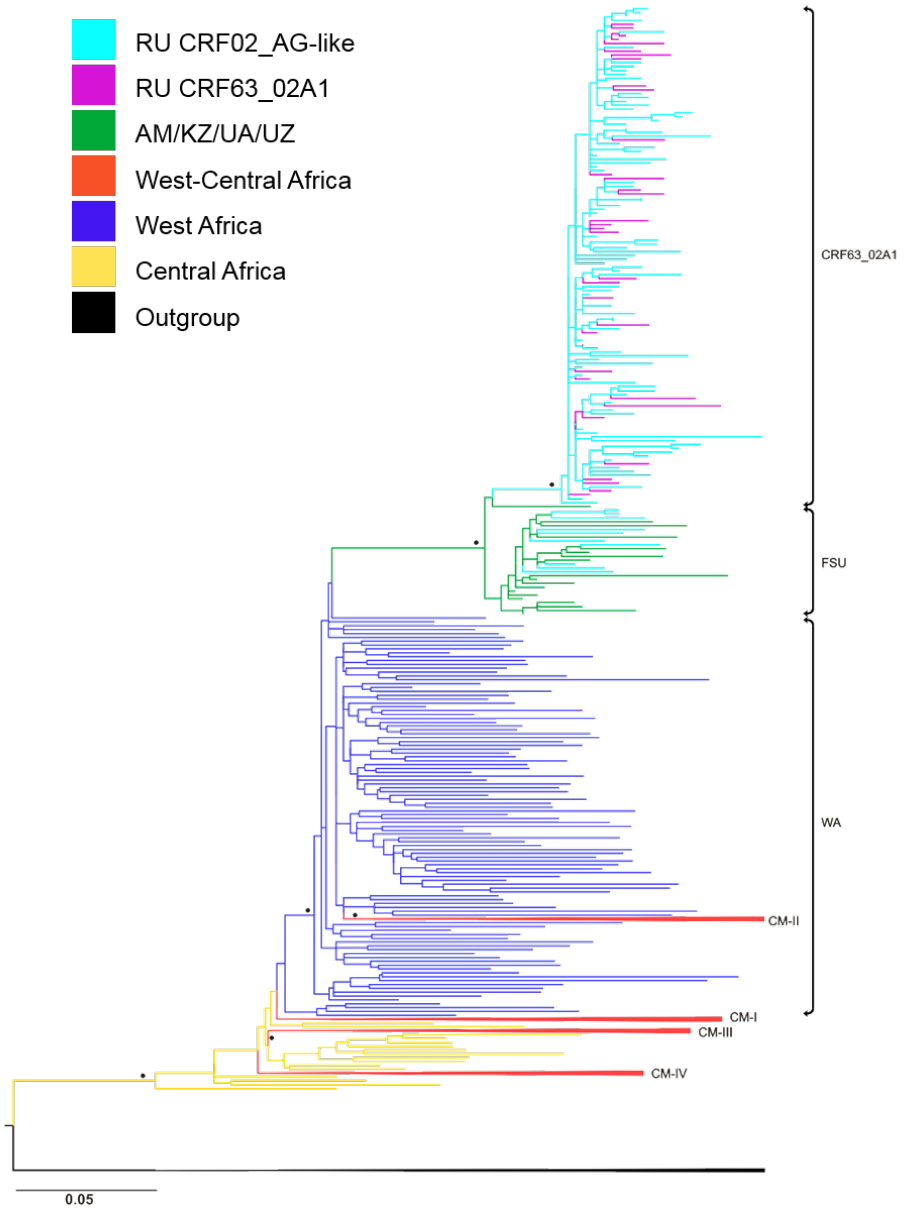


Figure 4.5: ML phylogenetic tree of HIV-1 CRF02_AG-like *pol* PR/RT sequences (∼1,000 nt) of the CRF02$_{FSU}$ and the CRF63_02A1 clades alongside reference sequences representative of the major African clades. Branches are colored according to the geographic origin and clade assignment of each sequence as indicated at the legend (top left). Positions of major CRF02_AG and CRF63_02A1 clades are indicated. The aLRT support value of each identified clade was > 0.85. The Cameroonian CRF02_AG clades as well as outgroup sequences were compressed for visual clarity. The tree was rooted using HIV-1 subtypes B, C, D, F, H, J and K reference sequences. The branch lengths are drawn to scale with the bar indicating nucleotide substitutions per site.

4.4.4 Timescale and demographic history of CRF02_AG and CRF63_02A1 clades

The evolutionary and demographic history of major CRF02_AG African clades and of $CRF02_{BG\text{-}DE}$, $CRF02_{FSU}$ and CRF63_02A1 clades was reconstructed using a Bayesian coalescent-based approach. The $CRF02_{WA}$ clade was subjected to a sub-sampling strategy because of its large size (n = 1507), resulting in six subsets (see Supplementary Information for full details of the procedure) each of which underwent the same analytical pipeline of the other clades. The median estimated evolutionary rate of the different clades were roughly comparable and all displayed a coefficient of rate variation that did not encompass zero (Tables 4.1 and S4.3), thus justifying the use of the relaxed molecular clock model. According to the substitution rates here estimated, the median tMRCA of the different clades dated back to between the late 1960s and the middle 2000s (Tables 4.1 and S4.3).

The demographic history estimated through the nonparametric BSP model, suggested that all CRF02_AG clades as well as the CRF63_02A1 clade underwent an initial period of substantial population expansion, followed by a more recent slowdown in its rates of spread (Fig. 4.6). The growth rate seems to start to decrease between 1985 and 1995 for African CRF02_AG clades, around the middle 2000s for the $CRF02_{FSU}$ clade and around the late 2000s for $CRF02_{BG\text{-}DE}$ and CRF63_02A1 clades. To test the significance of such a recent decline in the epidemic growth rate, different parametric growth models were compared for each clade. The logistic growth model provided the best fit to the demographic signal contained in all African CRF02_AG clades and in the CRF63_02A1 clade, whereas the demographic signal contained in the $CRF02_{FSU}$ and $CRF02_{BG\text{-}DE}$ clades was nearly equally fitted by both logistic and exponential growth models (Table S4.4). The median estimated logistic growth rates of major CRF02_AG African clades

(0.41 year^{-1} to 0.75 year^{-1}) were much lower than those estimated for CRF02_AG and CRF63_02A1 clades circulating in Europe and Asia (1.74 year^{-1} to 2.20 year^{-1}) (Tables 4.1 and S4.3). According to the exponential growth model, however, the median epidemic growth rates of the CRF02$_{FSU}$ (0.40 year^{-1}) and CRF02$_{BG-DE}$ (0.83 year^{-1}) clades were lower than those estimated by the logistic one (Table 4.1).

Table 4.1: Evolutionary and demographic parameters estimated for CRF02_AG and CRF63_02A1 clades.

Clade	N	Sampling interval	Substitution rate (10^{-3})	Coefficient of variation	tMRCA	Growth model	Growth rate
CRF02WA*	1,507	1990-2013	1.8(1.5-2.1)	0.23(0.19-0.26)	1967(1961-1974)	LG	0.63(0.48-0.78)
CRF02CM-I	428	1996-2012	1.6(1.5-1.8)	0.26(0.22-0.30)	1967(1962-1973)	LG	0.41(0.3-0.5)
CRF02CM-II	212	1996-2012	1.7(1.5-2.1)	0.35(0.29-0.41)	1978(1972-1984)	LG	0.75(0.5-1.0)
CRF02CM-III	50	2001-2012	1.5(1.5-1.9)	0.31(0.16-0.47)	1985(1979-1989)	LG	0.58(0.3-1.0)
CRF02CM-IV	33	1999-2012	1.6(1.5-2.0)	0.26(0.01- 0.44)	1983(1976-1989)	LG	0.44(0.2-0.7)
CRF02BG-DE	54	2006-2012	1.7(1.5-2.3)	1.10(0.70- 1.60)	2003(1997-2005)	LG	1.74(0.67-3.59)
						EG	0.84(0.59-1.11)
CRF02FSU	28	2008-2012	1.6(1.5-1.9)	0.33(0.07-0.61)	1998(1996-2000)	LG	2.00(0.48-4.36)
						EG	0.40(0.30-0.49)
CRF63_02A1	130	2009-2013	1.6(1.5-2.0)	0.63(0.51-0.79)	2004(2003-2006)	LG	2.10(1.25-3.25)

* Mean values from all CRF02$_{WA}$ subsets (Table S4.3). LG: logistic growth.
EG: exponential growth

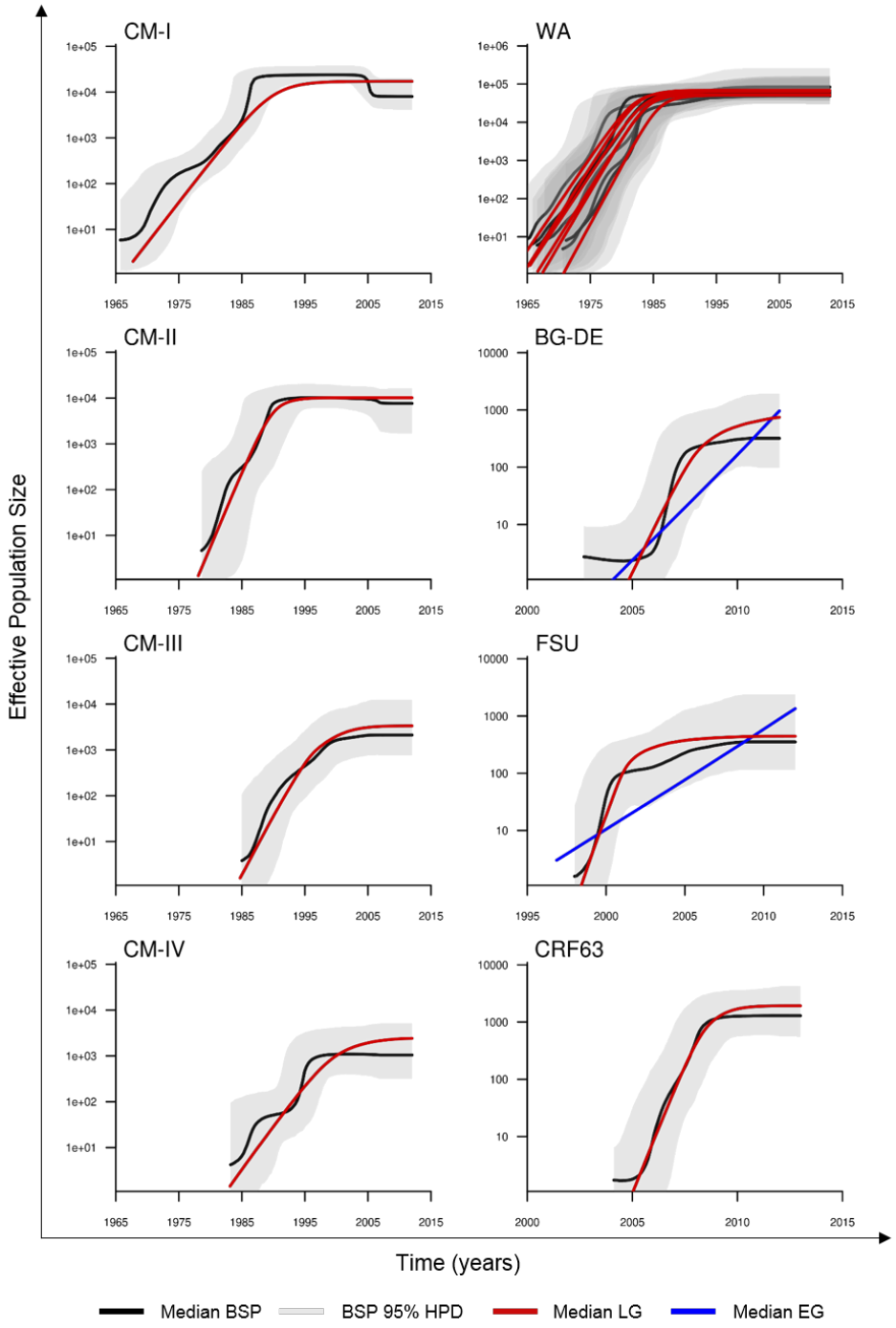


Figure 4.6: Demographic history of major HIV-1 CRF02_AG clades and the CRF63_02A1 clade. Median estimates of the *Ne* using Bayesian skyline (black line) and logistic/exponential growth models (red/blue line) are shown in each graphic together with 95% HPD intervals of the Bayesian skyline estimates (gray area). The vertical axes represent the estimated *Ne* on a logarithmic scale. Time scale is in calendar years.

4.5 Discussion

The present study embodies a major step toward the identification of the main HIV-1 CRF02_AG lineages circulating worldwide and the characterization of its spatiotemporal dynamics of dissemination. The analyses carried out with 3170 CRF02_AG-like *pol* sequences sampled around the world support that the current diversity of this HIV-1 variant mostly resulted from the expansion of a few clades with different epidemic outcomes.

The major CRF02_AG African lineage identified here, called $CRF02_{WA}$, probably arose after the introduction of a single founder strain from Central Africa into West Africa. This founder strain was disseminated throughout West Africa since the late 1960s onwards, establishing a large regional epidemic that comprises about 98% of the CRF02_AG sequences from that region here included and displays a very weak geographical structure characterized by country-specific sub-clades of small size ($n < 10$ sequences). The weak phylogeographic structure of the $CRF02_{WA}$ clade reflects multiple and frequent viral exchanges among West African countries that is fully consistent with the strong spatial accessibility [39] and the frequent human mobility [40,41] between countries from the West African region, and also coincides with the weak phylogeographic structure observed for other HIV-1 lineages (subtype G and CRF06_cpx) circulating in that region [42,43].

Other major CRF02_AG African clades identified here ($CRF02_{CM\text{-}I}$, $CRF02_{CM\text{-}III}$ and $CRF02_{CM\text{-}IV}$), seems to be the result of the expansion of three founder strains probably introduced into Cameroon from Central Africa. The estimated emergence of these CRF02_AG Cameroonian clades encompass a period of almost two decades, ranging from the late 1960s for $CRF02_{CM\text{-}I}$ to the middle 1980s for $CRF02_{CM\text{-}III}$ and $CRF02_{CM\text{-}IV}$. Previous studies

described the existence of two [22] and three [24] major CRF02_AG clades circulating in Cameroon that coincide with clades $CRF02_{CM\text{-}I}$/$CRF02_{CM\text{-}II}$ and $CRF02_{CM\text{-}I}$/$CRF02_{CM\text{-}II}$/$CRF02_{CM\text{-}III}$ detected in this study, respectively. The greater number of CRF02_AG Cameroonian lineages detected here compared to those previously reported probably arose from the larger number of sequences used in this new study.

The $CRF02_{WA}$ clade was not only successfully disseminated within the West African region, but was also introduced multiple times into other regions of Africa and into other countries around the world, originating a number of secondary outbreaks. The largest secondary $CRF02_{WA}$ outbreaks were detected in Cameroon, Bulgaria/Germany and countries from the FSU, leading to the origin of sub-clades called here $CRF02_{CM\text{-}II}$, $CRF02_{BG\text{-}DE}$ and $CRF02_{FSU}$, respectively, which were nested within the $CRF02_{WA}$ radiation. The $CRF02_{WA}$ clade and the descendant sub-clades comprise a significant proportion of sequences from West-Central Africa (> 40%), Central Africa (24%), and other regions around the world (> 84%). Thus, the $CRF02_{WA}$ clade is the most successfully disseminated CRF02_AG lineage at a global scale.

The chance of exportation of the CRF02_AG virus from West-Central and Central African regions seems to be much lower than from West Africa. The Cameroonian CRF02_AG clades ($CRF02_{CM\text{-}I}$, $CRF02_{CM\text{-}II}$, $CRF02_{CM\text{-}III}$ and $CRF02_{CM\text{-}IV}$) reach a high prevalence in Cameroon (81%) and the neighboring Gabon (38%), but comprise only a minor fraction of the CRF02_AG sequences detected in Angola and DRC (16%), Europe (16%), Equatorial Guinea (14%), America (13%), Asia/Oceania (4%), and West Africa (2%). Similarly, basal CRF02_AG lineages that are prevalent in Angola and DRC (60%) were barely detected outside this region. Given that West Africa hosts a much larger number of CRF02_AG-infected people (∼2,500,000) than Cameroon (∼330,000) and Central African countries

(<100,000) [2,3], a more frequent exportation of the $CRF02_{WA}$ clade out of the epicenter like the one supported by our results would be expected.

Demographic reconstructions indicate that major African CRF02_AG clades displayed a similar population growth pattern characterized by an initial phase of exponential growth followed by a decline in growth rate since the middle 1980s-1990s onwards. The median growth rate of the $CRF02_{WA}$ clade (0.63 $year^{-1}$) was somewhat lower than those previously estimated for subtype G clades and the CRF06_cpx lineage circulating in West Africa (0.75-0.95 $year^{-1}$) [42,43] (Fig. S4.1). Similarly, the median growth rate of different CRF02_AG lineages circulating in Cameroon also varied over a large range (0.41-0.75 $year^{-1}$). Although these results should be interpreted with caution because of the overlap of HPD intervals (Table 4.1), they support that spatial accessibility may not the only factor that shaped the rate of expansion of the different HIV-1 clades circulating in those African regions. Differences in the onset date of epidemics, transmission dynamics in distinct risk groups and/or viral transmissibility properties might be also responsible for the growth rate variances observed.

Our analyses revealed the existence of two major CRF02_AG transmission networks outside Africa involving individuals from Bulgaria/Germany ($CRF02_{BG\text{-}DE}$) and FSU countries ($CRF02_{FSU}$) that probably arose at around the early 2000s and the late 1990s, respectively. Although none of these countries host a large number of West African migrants [40], molecular epidemiologic studies showed that CRF02_AG infections detected in Bulgaria [44] as well as in Kazakhstan [11,13], Uzbekistan [10], Kyrgyzstan [14], and the Russian Federation [12] were preferentially associated to IDUs populations. Given the epidemiological link between the CRF02_AG and IDUs transmission networks described in the above countries, the origin of clades $CRF02_{BG\text{-}DE}$ and $CRF02_{FSU}$ could have been shaped by the rise

of international heroin traffic routes linking Afghanistan (the world largest opium producer) to the markets of the Russian Federation and Western Europe [45].

The HIV-1 epidemic in IDUs from FSU countries has been mainly driven by a subtype A1 variant characteristic of that region (A_{FSU}), that probably began to spread among IDUs from Ukraine in the early 1990s and was later disseminated to other FSU countries [46]. According to our estimations, the $CRF02_{FSU}$ variant began to spread in FSU countries around the late 1990s, probably resulting in a high number of co-infections with the AFSU variant already circulating and the subsequent generation of the AFSU/$CRF02_{FSU}$ recombinant called CRF63_02A1 clade. We estimated the origin of the CRF63 _02A1 clade around the middle 2000s, consistent with a previous study [37], supporting a very short time interval (<10 years) between the emergence of the $CRF02_{FSU}$ lineage and the origin of the CRF63_02A1 clade in the IDUs from FSU countries.

Inspection of the BSP of the $CRF02_{BG\text{-}DE}$, $CRF02_{FSU}$ and CRF63_02A1 clades supports a trend toward very recent epidemic stabilization since 2005-2010. The median logistic growth rate estimated for these clades circulating in European and FSU countries were very similar among each other (1.74-2.20 $year^{-1}$) and between three and five times faster than those estimated for African CRF02_AG clades. These extremely fast epidemic growth rates are fully consistent with the preferential dissemination of $CRF02_{BG\text{-}DE}$, $CRF02_{FSU}$ and CRF63_02A1 clades through highly connected IDUs transmission networks, in contrast to the African CRF02_AG clades that are mainly disseminated through heterosexual networks. The exponential demographic model, however, also provide a good fit to the demographic signal in the $CRF02_{FSU}$ and $CRF02_{BG\text{-}DE}$ clades and supports lower epidemic expansion rates than those estimated by the logistic model. It is possible that the logistic

pattern of the $CRF02_{FSU}$ and $CRF02_{BG\text{-}DE}$ clades was more difficult to capture due to its recent stabilization and/or low number of sequences, although we cannot ruled out that those epidemics are still growing exponentially.

In summary, this study reveals that the current CRF02_AG epidemics in West and West-Central African countries resulted from the dissemination of a few founder strains out of Central Africa between the late 1960s and the middle 1980s. The CRF02_AG strain introduced into the West African region ($CRF02_{WA}$) showed a broader geographic dissemination than any other African lineage. Spread of the $CRF02_{WA}$ clade outside Africa lead to the emergence of local transmission networks in Asia and Europe between the late 1990s and the early 2000s. The epidemic outcome of the different CRF02_AG lineages was probably shaped by several factors including: time of origin, spatial accessibility at the epicenter, risk groups transmission dynamics, and viral transmissibility properties.

4.6 References

1. N. R. Faria et al., "HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations," Science (80-.)., vol. 346, no. 6205, pp. 56–61, 2014.

2. J. Hemelaar, E. Gouws, P. D. Ghys, and S. Osmanov, "Global trends in molecular epidemiology of HIV-1 during 2000-2007.," AIDS, vol. 25, no. 5, pp. 679–689, 2011.

3. R. Lihana, D. Ssemwanga, A. Abimiku, and N. Ndembi, "Update on HIV-1 Diversity in Africa: A Decade in Review," AIDS Rev, vol. 14, pp. 83–100, 2012.

4. M. Peeters, C. Toure-Kane, and J. N. Nkengasong, "Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials.," AIDS, vol. 17, no. 18, pp. 2547–2560, 2003.

5. European Centre for Disease Prevention and Control (ECDC), "Assessing the burden of key infectious diseases affecting migrant populations in the EU/EEA," 2014.

6. V. Hernando et al., "HIV infection in migrant populations in the European Union and European Economic Area in 2007-2012; an epidemic on the move," J Acquir Immune Defic Syndr, vol. 70, no. 2, pp. 204–211, 2015.

7. A. B. Abecasis et al., "HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics," Retrovirology, vol. 10, no. 7, p. 1, 2013.

8. M. T. Pyne, J. Hackett, V. Holzmayer, and D. R. Hillyard, "Large-scale analysis of the prevalence and geographic distribution of HIV-1 non-B variants in the United States," J. Clin. Microbiol., vol. 51, no. 8, pp. 2662–2669, 2013.

9. P. B. Baryshev, V. V. Bogachev, and N. M. Gashnikova, "Genetic characterization of an isolate of HIV type 1 AG recombinant form circulating in Siberia, Russia," Arch. Virol., vol. 157, no. 12, pp. 2335–2341, 2012.

10. J. K. Carr et al., "Outbreak of a West African Recombinant of HIV-1 in Tashkent, Uzbekistan," J Acquir Immune Defic Syndr, vol. 39, pp. 570–575, 2005.

11. L. M. Eyzaguirre et al., “Genetic Characterization of HIV-1 Strains Circulating in Kazakhstan,” J. Acquir. Immune Defic. Syndr., vol. 46, no. 1, pp. 19–23, 2007.

12. E. Kazennova et al., “HIV-1 Genetic Variants in the Russian Far East,” AIDS Res. Hum. Retroviruses, vol. 30, no. 8, 2014.

13. I. Lapovok et al., “Short communication: molecular epidemiology of HIV type 1 infection in Kazakhstan: CRF02_AG prevalence is increasing in the southeastern provinces.,” AIDS Res. Hum. Retroviruses, vol. 30, no. 8, pp. 769–774, 2014.

14. V. Laga et al., “The Genetic Variability of HIV-1 in Kyrgyzstan: The Spread of CRF02 _ AG and Subtype A1 Recombinants.,” J. HIV AIDS, vol. 1.2, 2015.

15. E. Delatorre, G. Bello, W. A. C.-F. S. L. Eyer-Silva, M. G. Morgado, and J. C. Couto-Fernandez, “Evidence of Multiple Introductions and Autochthonous Transmission of the HIV Type 1 CRF02_AG Clade in Brazil,” AIDS Res. Hum. Retroviruses, vol. 28, no. 10, pp. 1369–72, 2012.

16. E. Delatorre, C. A. Velasco-De-Castro, J. H. Pilotto, J. C. Couto-Fernandez, G. Bello, and M. G. Morgado, “Reassessing the Origin of the HIV-1 CRF02_AG Lineages Circulating in Brazil,” AIDS Res. Hum. Retroviruses, vol. 31, 2015.

17. W. a Eyer-Silva and M. G. Morgado, “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, vol. 102, no. 7, pp. 809–15, 2007.

18. C. Tamalet et al., “Emergence of Clusters of CRF02_AG and B Human Immunodeficiency Viral Strains Among Men Having Sex With Men Exhibiting HIV Primary Infection in Southeastern France,” J. Med. Virol. J. Med. Virol, vol. 87, no. 87, 2015.

19. D. Brand et al., “Characteristics of patients recently infected with HIV-1 non-B subtypes in France: a nested study within the mandatory notification system for new HIV diagnoses.,” J. Clin. Microbiol., vol. 52, no. 11, pp. 4010–6, Nov. 2014.

20. K. Dauwe et al., "Characteristics and spread to the native population of HIV-1 non-B subtypes in two European countries with high migration rate," BMC Infect. Dis., 2015.

21. V. Von Wyl et al., "The role of migration and domestic transmission in the spread of HIV-1 non-B subtypes in Switzerland," J. Infect. Dis., vol. 204, no. 7, pp. 1095–1103, 2011.

22. N. R. Faria et al., "Phylodynamics of the HIV-1 CRF02_AG clade in Cameroon," Infect. Genet. Evol., vol. 12, no. 2, pp. 453–460, 2012.

23. J. Esbjörnsson, M. Mild, F. Månsson, H. Norrgren, and P. Medstrand, "HIV-1 molecular epidemiology in Guinea-Bissau, West Africa: Origin, demography and migrations," PLoS One, vol. 6, no. 2, 2011.

24. N. M. C. Véras et al., "Molecular Epidemiology of HIV Type 1 CRF02_AG in Cameroon and African Patients Living in Italy," AIDS Res. Hum. Retroviruses, vol. 27, no. 11, pp. 1173–1182, 2011.

25. D. Struck, G. Lawyer, A.-M. Ternes, J.-C. Schmit, and D. P. Bercoff, "COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification.," Nucleic Acids Res., vol. 42, no. 18, p. e144, Oct. 2014.

26. T. de Oliveira et al., "An automated genotyping system for analysis of HIV-1 and other microbial sequences.," Bioinformatics, vol. 21, no. 19, pp. 3797–800, Oct. 2005.

27. K. S. Lole et al., "Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination.," J. Virol., vol. 73, no. 1, pp. 152–60, Jan. 1999.

28. S. Guindon, J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel, "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0," Syst. Biol., vol. 59, no. 3, pp. 307–321, 2010.

29. D. Posada and K. A. Crandall, “Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1).,” Mol. Biol. Evol., vol. 18, no. 6, pp. 897–906, Jun. 2001.

30. M. Anisimova and O. Gascuel, “Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative.,” Syst. Biol., vol. 55, no. 4, pp. 539–52, Aug. 2006.

31. A. J. Drummond and A. Rambaut, “BEAST: Bayesian evolutionary analysis by sampling trees.,” BMC Evol. Biol., vol. 7, p. 214, 2007.

32. M. A. Suchard and A. Rambaut, “Many-core algorithms for statistical phylogenetics.,” Bioinformatics, vol. 25, no. 11, pp. 1370–6, Jun. 2009.

33. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, and A. Rambaut, “Relaxed phylogenetics and dating with confidence.,” PLoS Biol., vol. 4, no. 5, p. e88, May 2006.

34. A. B. Abecasis, A.-M. Vandamme, and P. Lemey, “Quantifying differences in the tempo of human immunodeficiency virus type 1 subtype evolution.,” J. Virol., vol. 83, no. 24, pp. 12917–24, Dec. 2009.

35. A. J. Drummond, A. Rambaut, B. Shapiro, and O. G. Pybus, “Bayesian coalescent inference of past population dynamics from molecular sequences.,” Mol. Biol. Evol., vol. 22, no. 5, pp. 1185–92, May 2005.

36. G. Baele, P. Lemey, T. Bedford, A. Rambaut, M. A. Suchard, and A. V Alekseyenko, “Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty.,” Mol. Biol. Evol., vol. 29, no. 9, pp. 2157–67, Sep. 2012.

37. N. S. Shcherbakova et al., “Short communication: Molecular epidemiology, phylogeny, and phylodynamics of CRF63_02A1, a recently originated HIV-1 circulating recombinant form spreading in Siberia.,” AIDS Res. Hum. Retroviruses, vol. 30, no. 9, pp. 912–9, 2014.

38. N. M. Gashnikova et al., "A Rapid Expansion of HIV-1 CRF63_02A1 Among Newly Diagnosed HIV-Infected Individuals in the Tomsk Region, Russia," AIDS Res. Hum. Retroviruses, vol. 31, no. 4, pp. 456–460, 2015.

39. A. J. Tatem, J. Hemelaar, and R. R. Gray, "Spatial accessibility and the spread of HIV-1 subtypes and recombinants in sub-Saharan Africa," no. July, pp. 1–11, 2012.

40. F. Charrièrre and M. Fresia, "West Africa as a Migration and Protection area," UN High Commissioner for Refugees (UNHCR), 2008. Online.. Available: http://www.refworld.org/docid/4a277db82.html. Accessed: 14-Jan-2016..

41. D. Gnisci and M. Trémolieres, West African Studies, Regional Atlas on West Africa. 2009.

42. E. Delatorre, D. Mir, and G. Bello, "Spatiotemporal dynamics of the HIV-1 subtype G epidemic in West and Central Africa.," PLoS One, vol. 9, no. 2, p. e98908, Jan. 2014.

43. E. Delatorre and G. Bello, "Spatiotemporal dynamics of the HIV-1 CRF06_cpx epidemic in western Africa," PLoS One, vol. 9, no. 6, 2014.

44. I. A. Ivanov et al., "Detailed molecular epidemiologic characterization of HIV-1 infection in Bulgaria reveals broad diversity and evolving phylodynamics.," PLoS One, vol. 8, no. 3, p. e59666, 2013.

45. UNODC, "Illicit Drug Trends in Central Asia," 2008.

46. F. Díez-Fuertes, M. Cabello, and M. M. Thomson, "Bayesian phylogeographic analyses clarify the origin of the HIV-1 subtype A variant circulating in former Soviet Union's countries," Infect. Genet. Evol., vol. 33, pp. 197–205, 2015.

4.7 Supplementary Information

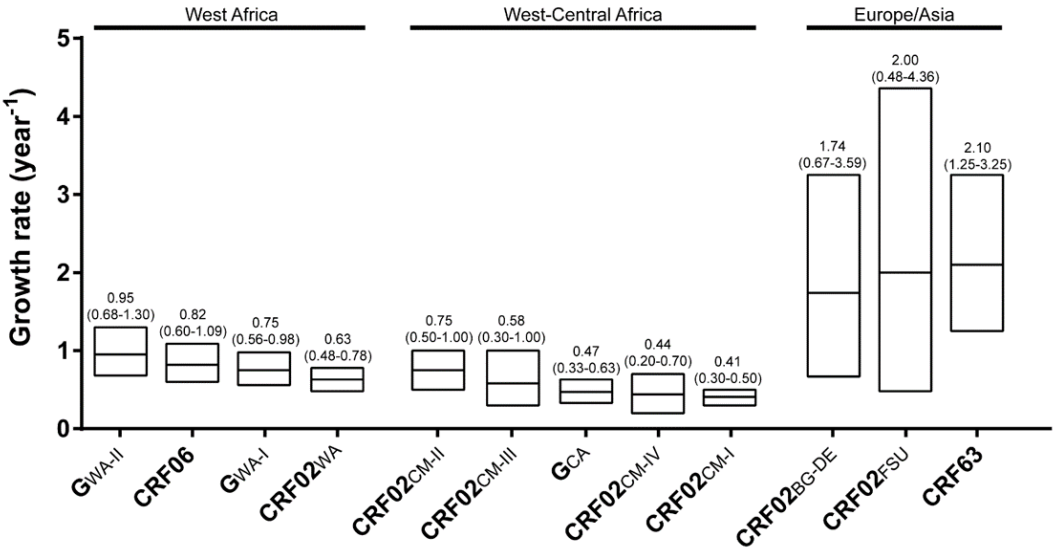


Figure S 4.1: Coalescent estimates of epidemic growth rate of major HIV-1 CRF02_AG, subtype G, CRF06_cpx and CRF63_02A1 clades identified. The vertical lines represent the median growth rates (years^{-1}) and the corresponding 95% HPD intervals of the posterior distributions estimated under the best-fitted growth coalescent model for each clade. Epidemic growth rates of subtype G (G_{WA-I}, G_{WA-II} and G_{CA}) and CRF06_cpx clades were estimated previously [42,43].

Table S 4.1: HIV-1 CRF02_AG-like and CRF63_02A1-like *pol* dataset.

	Region	Country	N	Sampling interval
CRF02_AG-like	West Africa	Benin (BJ)	92	2004-2009
		Burkina Faso (BF)	132	2001-2009
		Cape Verde (CV)	38	2010-2011
		Cote D'Ivoire (CI)	55	1997-2008
		Gambia (GM)	1	2005
		Ghana (GH)	184	1990-2008
		Guinea (GN)	1	2007
		Guinea Bissau (GW)	5	2004-2009
		Liberia (LR)	48	1994-2013
		Mali (ML)	160	2005-2009
		Mauritania (MR)	36	2009
		Nigeria (NG)	173	1999-2010
		Senegal (SN)	272	1998-2011
		Togo (TG)	116	2006-2008
	West-Central Africa	Cameroon (CM)	827	1996-2012
		Chad (TD)	2	2006
		Equatorial Guinea (GQ)	21	2008
		Gabon (GA)	58	2000-2008
	Central Africa	Angola (AO)	8	2001-2010
		DRC (CD)	17	1997-2008
	America	Brazil (BR)	15	1998-2012
		Colombia (CO)	1	2003
		Cuba (CU)	3	2003
		Ecuador (EC)	8	2005-2008
		United States (US)	142	1998-2011
	Europe	Austria (AT)	8	2003-2006
		Belgium (BE)	39	1999-2013
		Bulgaria (BG)	56	2008-2012
		Byelorussian (BY)	1	-
		Cyprus (CY)	10	2007-2011
		Czech Republic (CZ)	21	2001-2007
		Denmark (DK)	17	2000-2007
		Estonia (EE)	4	2008-2010
		Finland (FI)	3	2003-2006
		France (FR)	21	1991-2003
		Germany (DE)	58	1998-2012
		Greece (GR)	1	2007
		Italy (IT)	37	1997-2009
		Luxembourg (LU)	2	2005-2006
		Netherlands (NL)	9	2003-2010
		Norway (NO)	7	2004-2006
		Poland (PL)	7	1989-2013
		Portugal (PT)	31	2003-2012
		Romania (RO)	4	2007-2013
		Serbia (RS)	1	2011
		Slovakia (SK)	1	2005
		Slovenia (SI)	8	1998-2005
		Spain (ES)	160	1995-2012
		Sweden (SE)	25	2003-2011
		Switzerland (CH)	14	2003-2009
		Ukraine (UA)	1	2010
		United Kingdom (GB)	27	2005
	Asia/Oceania	Armenia (AM)	1	2009
		Australia (AU)	10	2004-2010
		China (CN)	12	2005-2009
		Japan (JP)	10	2003-2011
		Kazakhstan (KZ)	9	2012
		Pakistan (PK)	3	2009
		Philippines (PH)	3	2010-2011
		Republic of South Korea, (KR)	10	2003-2013
		Russian Federation (RU)	108	2008-2013
		Thailand (TH)	8	2007-2013
		Uzbekistan (UZ)	8	2002
CRF63_02A1-like	Asia	Russian Federation (RU)	38	2009-2013

Table S 4.2: Relative prevalence (%) of major HIV-1 CRF02_AG African clades.

Region	Country	CRF02WEST	CRF02CM-I	CRF02CM-II	CRF02CM-III	CRF02CM-IV
West Africa	Benin (BJ)	96.0	2.0	2.0	-	-
	Burkina Faso (BF)	100	-	-	-	-
	Cape Verde (CV)	94.7	-	2.6	2.6	-
	Cote D'Ivoire (CI)	100	-	-	-	-
	Gambia (GM)	100	-	-	-	-
	Ghana (GH)	100	-	-	-	-
	Guinea (GN)	100	-	-	-	-
	Guinea Bissau (GW)	100	-	-	-	-
	Liberia (LR)	100	-	-	-	-
	Mali (ML)	96.3	1.3	2.5	-	-
	Mauritania (MR)	100	-	-	-	-
	Nigeria (NG)	97.7	1.7	0.6	-	-
	Senegal (SN)	97.1	0.4	2.2	-	0.4
	Togo (TG)	100	-	-	-	-
West-Central Africa	Cameroon (CM)	19.0	48.6	22.8	5.9	3.7
	Chad (TD)	50.0	50.0	-	-	-
	Equatorial Guinea (GQ)	85.7	9.5	4.8	-	-
	Gabon (GA)	62.1	24.1	13.4	-	-
Central Africa	Angola (AO)	83.3	16.7	-	-	-
	DRC (CD)	25.0	25.0	-	-	50.0
America	Brazil (BR)	100	-	-	-	-
	Colombia (CO)	100	-	-	-	-
	Cuba (CU)	66.7	-	33.3	-	-
	Ecuador (EC)	87.5	12.5	-	-	-
	United States (US)	85.9	8.5	5.6	-	-
Europe	Austria (AT)	100	-	-	-	-
	Belgium (BE)	64.1	12.8	20.5	-	2.6
	Bulgaria (BG)	98.2	1.8	-	-	-
	Byelorussian SSR (BY)	-	100	-	-	-
	Cyprus (CY)	40.0	20.0	40.0	-	-
	Czech Republic (CZ)	66.7	23.9	9.5	-	-
	Denmark (DK)	70.6	17.6	5.9	5.9	0
	Estonia (EE)	100	-	-	-	-
	Finland (FI)	100	-	-	-	-
	France (FR)	81.0	4.8	14.3	-	-
	Germany (DE)	81.0	6.9	12.1	-	-
	Greece (GR)	100	-	-	-	-
	Italy (IT)	100	-	-	-	-
	Luxembourg (LU)	100	-	-	-	-
	Netherlands (NL)	66.7	22.2	11.1	-	-
	Norway (NO)	71.4	28.6	-	-	-
	Poland (PL)	85.7	-	14.3	-	-
	Portugal (PT)	96.8	3.2	-	-	-
	Romania (RO)	75.0	25.0	-	-	-
	Serbia (RS)	100	-	-	-	-
	Slovakia (SK)	100	-	-	-	-
	Slovenia (SI)	100	-	-	-	-
	Spain (ES)	86.3	8.1	5.6	-	-
	Sweden (SE)	80.0	11.5	3.8	-	3.8
	Switzerland (CH)	78.6	7.1	14.3	-	-
	Ukraine (UA)	100	-	-	-	-
	United Kingdom (GB)	74.1	25.9	-	-	-
Asia/Oceania	Armenia (AM)	100	-	-	-	-
	Australia (AU)	100	-	-	-	-
	China (CN)	83.3	8.3	8.3	-	-
	Japan (JP)	90.0	10.0	-	-	-
	Kazakhstan (KZ)	100	-	-	-	-
	Pakistan (PK)	100	-	-	-	-
	Philippines (PH)	-	33.3	66.7	-	-
	Republic of South Korea (KR)	100	-	-	-	-
	Russian Federation (RU)	99.1	0.9	-	-	-
	Thailand (TH)	100	-	-	-	-
	Uzbekistan (UZ)	100	-	-	-	-

Table S 4.3: Evolutionary and demographic parameters estimated for $CRF02_{WA}$ subsets.

Subset	N	SamplingInterval	Substitution rate (x 10^{-3})	Coefficient of variation	tMRCA	Logistic growth-rate
1	429	1994-2013	1.7(1.5-2.0)	0.20(0.17-0.24)	1963(1957-1971)	0.57 (0.44-0.72)
2	432	1994-2013	1.8 (1.5-2.2)	0.25(0.21-0.28)	1968(1960-1975)	0.61(0.46-0.76)
3	433	1990-2013	1.6 (1.5-1.8)	0.23(0.19-0.26)	1970(1966-1975)	0.66(0.50-0.83)
4	434	1993-2013	2.1 (1.7-2.6)	0.22(0.19-0.26)	1971(1963-1978)	0.70(0.52-0.89)
5	433	1993-2013	1.7(1.5-2.0)	0.23(0.20-0.27)	1965 (1958-1973)	0.62(0.49-0.77)
6	431	1994-2013	1.7(1.5-2.0)	0.22(0.18-0.25)	1966 (1959-1974)	0.58(0.46-0.71)
Mean	-	-	1.8(1.5-2.1)	0.23(0.19-0.26)	1967(1961-1974)	0.63(0.48-0.78)

Table S 4.4: Best fit demographic model for major HIV-1 CRF02_AG clades and CRF63_02A1 clades.

Clade	Model	PSLog ML	Models compared	Log BF	SSLog ML	Models compared	Log BF
CRF02WA1	Log	-44895.6	-	-	-44901.8	-	-
	Expo	-45982.5	Log/Expo	1086.9	-45999.8	Log/Expo	1098.8
	Expa	-45994.2	Log/Expa	1098.6	-46005.1	Log/Expa	1103.3
CRF02WA2	Log	-46121.0	-	-	-46139.4	-	-
	Expo	-46361.3	Log/Expo	240.3	-46378.2	Log/Expo	238.8
	Expa	-46377.2	Log/Expa	256.2	-46386.4	Log/Expa	247.0
CRF02WA3	Log	-46177.7	-	-	-46198.9	-	-
	Expo	-46439.0	Log/Expo	261.3	-46454.7	Log/Expo	255.8
	Expa	-46440.9	Log/Expa	263.2	-46457.7	Log/Expa	258.8
CRF02WA4	Log	-46387.6	-	-	-46407.7	-	-
	Expo	-46623.2	Log/Expo	235.6	-46639.3	Log/Expo	231.6
	Expa	-46652.1	Log/Expa	264.5	-46670.3	Log/Expa	262.6
CRF02WA5	Log	-46095.0	-	-	-46115.6		-
	Expo	-46370.5	Log/Expo	275.5	-46387.8	Log/Expo	272.2
	Expa	-46389.7	Log/Expa	294.7	-46407.9	Log/Expa	292.3
CRF02WA6	Log	-45316.5	-	-	-45335.4	-	-
	Expo	-45544.3	Log/Expo	227.8	-45559.1	Log/Expo	223.7
	Expa	-45557.7	Log/Expa	241.2	-45574.4	Log/Expa	239.0
CRF02CM-I	Log	-38204.0	-	-	-38228.5	-	-
	Expo	-38427.1	Log/Expo	223.1	-38450.1	Log/Expo	221.6
	Expa	-38424.9	Log/Expa	220.9	-38442.7	Log/Expa	214.2
CRF02CM-II	Log	-18919.3	-	-	-18923.3	-	-
	Expo	-19039.6	Log/Expo	120.4	-19044.0	Log/Expo	120.7
	Expa	-19061.3	Log/Expa	142.0	-19065.1	Log/Expa	141.8
CRF02CM-III	Log	-5924.9	-	-	-5925.2	-	-
	Expo	-5946.0	Log/Expo	21.1	-5946.3	Log/Expo	21.1
	Expa	-5950.9	Log/Expa	26.0	-5951.1	Log/Expa	25.9
CRF02CM-IV	Log	-4493.3	-	-	-4493.5	-	-
	Expo	-4530.4	Log/Expo	37.1	-4530.1	Log/Expo	36.6
	Expa	-4531.2	Log/Expa	37.9	-4531.4	Log/Expa	37.9
CRF02BG-DE	Log	-2723.2	-	-	-2723.3	-	-
	Expo	-2722.9	Log/Expo	-0.3	-2722.9	Expo/Log	-0.4
	Expa	-2727.1	Log/Expa	3.9	-2727.3	Log/Expa	4.0
CRF02FSU	Log	-3003.7	-	-	-3003.8	-	-
	Expo	-3005.4	Log/Expo	1.7	-3005.4	Log/Expo	1.6
	Expa	-3011.8	Log/Expa	8.1	-3011.8	Log/Expa	8.0
CRF63_02A1	Log	-6114.4	-		-6115.8		
	Expo	-6126.9	Log/Expo	12.5	-6127.3	Log/Expo	11.5
	Expa	-6140.7	Log/Expa	26.3	-6141.4	Log/Expa	25.6

Chapter 5

Inferring population dynamics of HIV-1 subtype C epidemics in Eastern Africa and Southern Brazil applying different Bayesian phylodynamics approaches.

Article published in:

Daiana Mir, Tiago Gräf, Sabrina Esteves de Matos Almeida, Aguinaldo Roberto Pinto, Edson Delatorre and Gonzalo Bello. *Scientific Reports* 2018; 8(1):8778.

5.1 Abstract

The subtype C Eastern Africa clade (C_{EA}), a particularly successful HIV-1 subtype C lineage, has seeded several sub-epidemics in Eastern African countries and Southern Brazil during the 1960s and 1970s. Here, we characterized the past population dynamics of the major C_{EA} sub-epidemics in Eastern Africa and Brazil by using Bayesian phylodynamic approaches based on coalescent and birth- death models. All phylodynamic models support similar epidemic dynamics and exponential growth rates until roughly the mid-1980s for all the C_{EA} sub-epidemics. Divergent growth patterns, however, were supported afterwards. The Bayesian skygrid coalescent model (BSKG) and the birth-death skyline model (BDSKY) supported longer exponential growth phases than the Bayesian skyline coalescent model (BSKL). The BDSKY model uncovers patterns of a recent decline for the C_{EA} sub-epidemics in Burundi/Rwanda and Tanzania (R_e < 1) and a recent growth for Southern Brazil (R_e > 1); whereas coalescent models infer an epidemic stabilization. To the contrary, the BSKG model captured a decline of Ethiopian C_{EA} sub-epidemic between the mid-1990s and mid-2000s that was not uncovered by the BDSKY model. These results underscore that the joint use of different phylodynamic approaches may yield complementary insights into the past HIV population dynamics.

5.2 Introduction

The human immunodeficiency virus type 1 (HIV-1) subtype C accounts for approximately 48% of all people living with HIV, representing the most prevalent HIV-1 subtype in the world [1]. The high global prevalence of the C subtype results from its predominance in regions with the highest rates of HIV-1 infection and with large populations, such as Southern and Eastern Africa, India and Southern Brazil [1, 2]. The origin of HIV-1 subtype C was recently traced to the Katanga region of the Democratic Republic of Congo (DRC) in the late 1930s [3] from where it spread independently to Eastern and Southern Africa, leading to a phylogeographic subdivision between the HIV-1 subtype C strains circulating in those two African regions [4, 5].

The expansion of HIV-1 subtype C inside Eastern Africa gave rise to the C East African clade (C_{EA}), whose most probable epicenter of dissemination was in Burundi around the early 1960s. During the 1970s, this country acted as ignition point of several local C_{EA} sub- epidemics in other Eastern African countries [5] and also in Southern Brazil [6] where the C_{EA} sub-epidemic was fueled from a single founder event [7]. The C_{EA} clade currently predominates among subtype C strains from Eastern African countries and Brazil, and accounts for almost 100% of subtype C sequences from Burundi and Brazil, 97% from Uganda, 64% from Kenya, 61% from Ethiopia and 49% from Tanzania [5, 8]. The evolutionary analyses of the C_{EA} sub-epidemics performed so far mostly address questions about the place and timing of outbreaks onset, focusing on the reconstruction of the geographic dissemination pathways of this viral clade [2, 5, 6, 8, 9, 10, 11, 12]. Studies on the population dynamics of C_{EA} sub-epidemics, while key to understand their historical epidemic growth trends, epidemic potential and ecological processes shaping their evolution, have been much less frequent [13, 14].

Key epidemiological and population parameters, most notably the effective number of infections (*Ne*), the epidemic growth rate (r) and the basic (R_0) and effective (R_e) reproductive number, can be estimated from viral sequence data by using Bayesian phylodynamic approaches based on coalescent [15] and birth-death [16] models. These models have very different mathematical grounds as well as particular strengths and limitations. The coalescent is appropriate only if the number of sampled infected individuals is small compared with the size of the total infected population [16], despite certain robustness to violation of this requirement has been demonstrated [17]. The birth- death model, meanwhile, explicitly models the sampling process and can thus be used for sparse or densely sampled viral populations [16], although estimates may be biased if the model of the sampling process is misspecified. The coalescent allows the inference of the R_0 (key epidemiological parameter indicator of increment [$R_0 > 1$], decline [$R_0 < 1$] or stabilization [$R_0 = 1$] in the number of new cases [18]) only through modeling the population dynamics under a deterministic assumption, which represents a limitation for populations undergoing complex dynamics [19], and require an independent estimate of the average duration of infectiousness. The birth–death model has the advantage of accounting for stochasticity of the demographic process and provides an estimate for Re e changes over time using only sequence data [16, 20]. A potential disadvantage of the birth-death model is that credibility intervals grow wider the further we go into the past, which is not the case for the coalescent-based models [21]; although simulation studies showed that the coalescent might not capture the true*r*because of the narrow credibility intervals around the median estimate attributed to its assumption of deterministic changes in the population size [20].

The present work aims to shed light on the past population dynamics of the major HIV-1 C_{EA} sub-epidemics established in Burundi, Rwanda, Ethiopia, Tanzania and Brazil by analyzing viral *pol* gene sequences sampled between 1990 and 2014 with Bayesian phylodynamic methods based on coalescent and birth-death models.

5.3 Materials and Methods

5.3.1 Sequence dataset compilation

A reference dataset of HIV-1 subtype C *pol* sequences belonging to the east, southern and central African lineages was selected from a previous study [5] and combined with: 1) more recent east African subtype C *pol* sequences with known sampling dates available in Los Alamos HIV Database (http://www.hiv.lanl.gov) by August 2017, and 2) subtype C *pol* sequences with known sampling dates isolated from heterosexual populations living in the two southernmost Brazilian states (Rio Grande do Sul and Santa Catarina) previously described [11]. The option "One sequence/patient" was selected from Los Alamos HIV database to exclude multiple sequences from the same subject. The subtype assignment of all sequences was confirmed using the REGA HIV-1 subtyping tool v.3.0. Given the two genetically distinct subtype C clades (C and C') co-circulating in Ethiopia [22], linked to subtype C viruses of eastern and southern African origin respectively, putative intrasubtype C/C' recombinant sequences (n = 99) were identified by Bootscanning using Simplot v3.5.1 [23] as described previously [5] and removed from further analyses. This resulted in a final dataset of 1,147 HIV-1 subtype C *pol* sequences (Table 5.1) covering the complete protease (PR) and the first part

of the reverse transcriptase (RT) regions (nucleotides 2,253 to 3,272 relative to HXB2 genome).

5.3.2 Identification of dominant country-specific HIV-1 C_{EA} subclades

To identify major country-specific clades within the C_{EA} radiation, HIV-1 subtype C *pol* sequences from eastern Africa and southern Brazil were first aligned with reference subtype C sequences belonging to the eastern, southern and central African clades using the CLUSTAL X program [24] and subjected to maximum likelihood (ML) phylogenetic analysis. ML trees were inferred with the PhyML program [25], using an online web server [26], under the general time-reversible model of nucleotide substitution plus invariant sites and four discrete gamma rate categories (GTR+I+Γ4) selected with jModeltest program [27] and the subtree pruning and regrafting (SPR) branch- swapping algorithm of heuristic tree search. The reliability of the phylogenies was estimated with the approximate likelihood-ratio test based on a Shimodaira–Hasegawa-like procedure (SH-aLRT) [25]. Basal HIV-1 C_{EA} sequences from Burundi and Rwanda and major ($n \geq 50$ sequences) country-specific (>90% of sequences from a single country) monophyletic groups with high support (SH-aLRT $\geq$0.85) nested within the C_{EA} clade radiation were selected for demographic analyses. Reference sequences of HIV-1 subtypes A1 and D from the Los Alamos HIV Database were used as outgroups. Final trees were visualized in FigTree v1.4.2.

5.3.3 Estimation of phylodynamic parameters

Epidemiological and evolutionary parameters of the defined C_{EA} subclades were estimated via Bayesian Markov Chain Monte Carlo (MCMC)

phylogenetic inference using coalescent and birth-death tree priors as implemented in BEAST v1.8 [28] and BEAST v.2.4 [29] software packages, respectively. Changes in *Ne* using the coalescent tree prior were first assessed using the non-parametric Bayesian skyline (BSKL) [30] and Bayesian Skygrid (BSKG) [31] models and estimates of the *r* were subsequently obtained using the parametric model that provided the best fit to the demographic signal contained in each dataset. Comparison between demographic models (logistic, exponential, or expansion) was performed using the log marginal likelihood estimation (MLE) based on path sampling (PS) and stepping-stone sampling (SS) methods [32]. The cumulative number of lineages through time (LTT) was calculated from the combined posterior distribution of sampled coalescent tree topologies by using TRACER v1.6 program [33]. A special case of the birth-death tree prior, namely the birth-death skyline (BDSKY) was applied to model viral transmissions through time [21]. The sampling rate (δ) was set to zero for the period prior to the oldest sample and estimated from the data afterwards. The R_e was estimated in a piece-wise manner over three different equidistant intervals using a lognormal prior distribution (R_e : mean = 0, standard deviation = 1). Bayesian analyses for each transmission clade employed the GTR+I+Γ4 model of nucleotide substitution selected using the jModelTest program [27] and a relaxed uncorrelated lognormal molecular clock model [34]. Because linear regression analysis of root-to-tip distances as function of sampling time obtained by TempEst v1.5 [35] revealed low temporal signal in the datasets, an informative normal prior distribution on the time to the most recent common ancestor (tMRCA) was applied based on previous estimates [5, 6]. MCMC chains were run for sufficiently long to ensure stationarity (constant mean and variance of trace plots) and good mixing (Effective Sample Size >200) for all parameter estimates, as diagnosed by TRACER v1.6 program [33].

5.4 Results

5.4.1 Identification of major subclades within the HIV-1 C_{EA} clade radiation

To obtain a more updated picture of the HIV-1 C_{EA} clade radiation, subtype C *pol* sequences from Eastern Africa and Southern Brazil deposited in Los Alamos HIV sequence database between 2013 and 2016 were combined with C_{EA} *pol* sequences from those regions previously characterized [5, 6]. The reconstructed ML phylogeny showed that most (79%) subtype C sequences from Eastern Africa and all sequences from Southern Brazil sampled at most recent time (2013–2016) branched within the highly supported (SH-aLRT = 0.96) HIV-1 C_{EA} clade (Fig. S5.1). As expected, sequences from Burundi and Rwanda were highly intermixed among each other and occupied the most basal positions of the C_{EA} clade radiation; while sequences from other Eastern African countries and Brazil were nested within Burundian and Rwandan C_{EA} sequences.

Most sequences from Kenya and Uganda appeared as sporadic (non-clustered) lineages or clustered in monophyletic subclades of small sizes ($n < 50$) (Fig. S5.1). All Brazilian sequences and most sequences from Ethiopia (67%) and Tanzania (66%), by contrast, branched within four country-specific C_{EA} subclades of large size ($n > 50$) (Fig. S5.1) that were more clearly visualized after pruning of non-clustered C_{EA} sequences and C_{EA} sequences within monophyletic subclades of small sizes (Fig. 5.1). The four identified C_{EA} subclades ($C_{EA/BR}$, $C_{EA/ET\text{-}1}$, $C_{EA/ET\text{-}2}$ and $C_{EA/TZ}$) together with sequences from Burundi and Rwanda ($C_{EA/BI\text{-}RW}$) comprise 76% ($n = 616$) of all the C_{EA} sequences analyzed here; thus confirming the epidemiological relevance of the selected subclades.

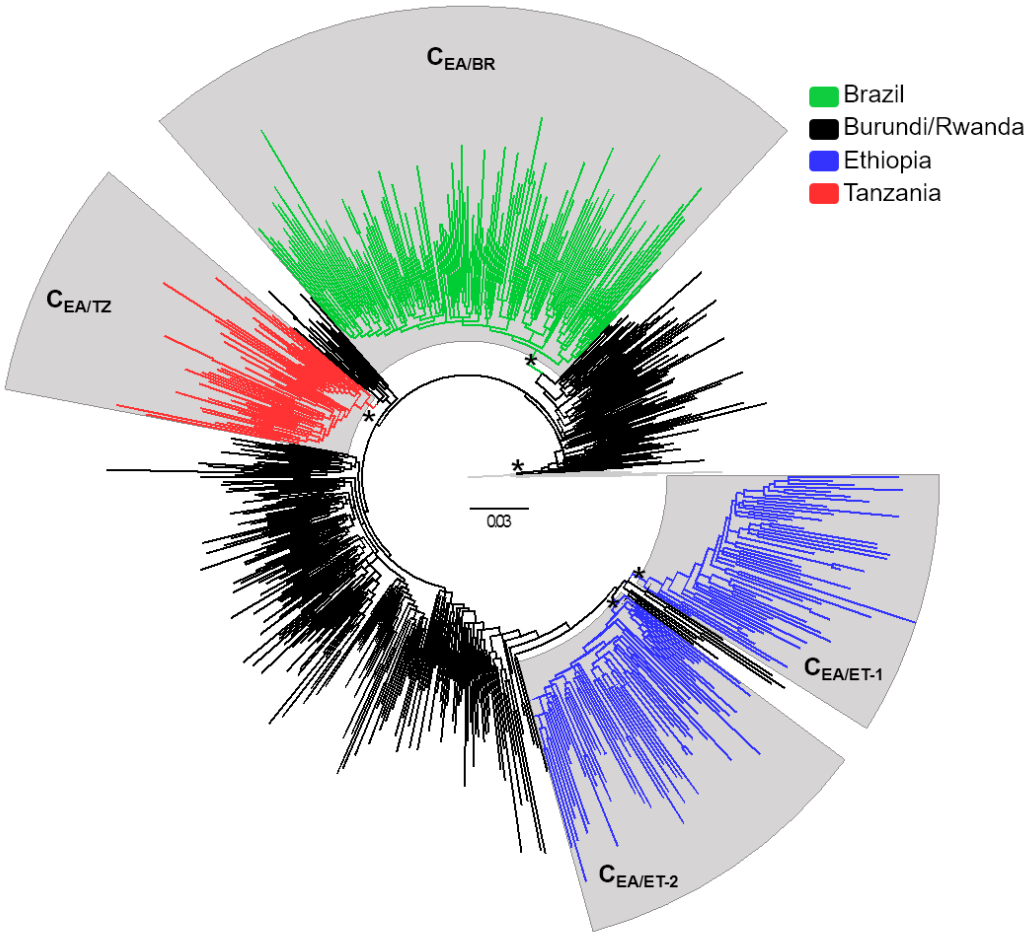


Figure 5.1: ML phylogenetic tree of HIV-1 C_{EA} *pol* PR/RT sequences (∼1,000 nt) from eastern Africa and southern Brazil. Branches are colored according to the geographic origin of sequences as indicated in the legend (upper right). Gray shaded boxes indicate the positions of major C_{EA} lineages. Asterisks point to key nodes with high support (SH-aLRT > 0.85). The tree was rooted using HIV-1 subtypes A1 and D reference sequences and the branch lengths are drawn to scale with the bar at the center indicating nucleotide substitutions per site.

5.4.2 Bayesian population dynamics inference in a coalescent framework

Bayesian MCMC coalescent-based analyses under the BSKL model suggest that all C_{EA} subclades presents roughly comparable demographic histories, with an initial phase of fast exponential growth followed by a stabilization of the *Ne* at some time between the late 1970s and the late 1980s that persisted until the most recent sampling date of each of them (Figs 2a,e, 3a,e and 4a).

The observed stabilization in the *Ne* of African C_{EA} sub-clades occurs before the respective stabilization of the HIV incidence in corresponding countries estimated around the mid-1990s according to the UNAIDS (Figs. 5.2 d,h and 5.3 d,h). The UNAIDS data also supports a significant reduction of the HIV incidence in Burundi/Rwanda, Ethiopia and Tanzania between the mid-1990s and the mid-2000s that was not captured by the BSKL inference. The stabilization of the $C_{EA/BR}$ *Ne* around the early 1990s is consistent with the stabilization of the new HIV cases in Rio Grande do Sul and Santa Catarina at around the same time (considering a lag time of eight years between HIV infection and new AIDS cases reported by the Brazilian AIDS cases databank for those Brazilian states); but fails to capture a recent increase in the number of new HIV cases from 2010 onwards (Fig. 5.4 d).

The BSKG demographic reconstructions point to a longer period of exponential growth for all C_{EA} subclades that extends up to between the early and the mid-1990s, in agreement with incidence data. For the Ethiopian C_{EA} sub-epidemics, the BSKG points to subsequent decline of the median *Ne* until mid-2000s and a final plateau until 2010, in agreement with epidemiological data (Fig. 5.2 b,d,f,h). The median estimated *Ne* for the $C_{EA/BI\text{-}RW}$, $C_{EA/TZ}$ and $C_{EA/BR}$ sub-epidemics reach a plateau that persisted until the most recent sampling time which differs markedly from the HIV incidence pattern (Figs 5.3 b,d,f,h and 5.4 b,d). The large 95% highest probability density (HPD) interval of the *Ne* estimates inferred by the BSKG model, however, may accommodate different demographic patterns in the latter stages making it difficult to draw solid conclusions on the consistency (or the lack thereof) between these estimates and the HIV incidence temporal pattern.

The logistic growth model was the best-fit parametric model of population growth for all the C_{EA} subclades (Table 5.2). The median *Ne* trajectories

obtained by the logistic growth model during the initial exponential phase closely matched the corresponding trajectories obtained with the non-parametric approaches, particularly those obtained with the BSKG (blue line Figs 5.2a,b,e,f, 5.3 a,b,e,f and 5.4 a,f).

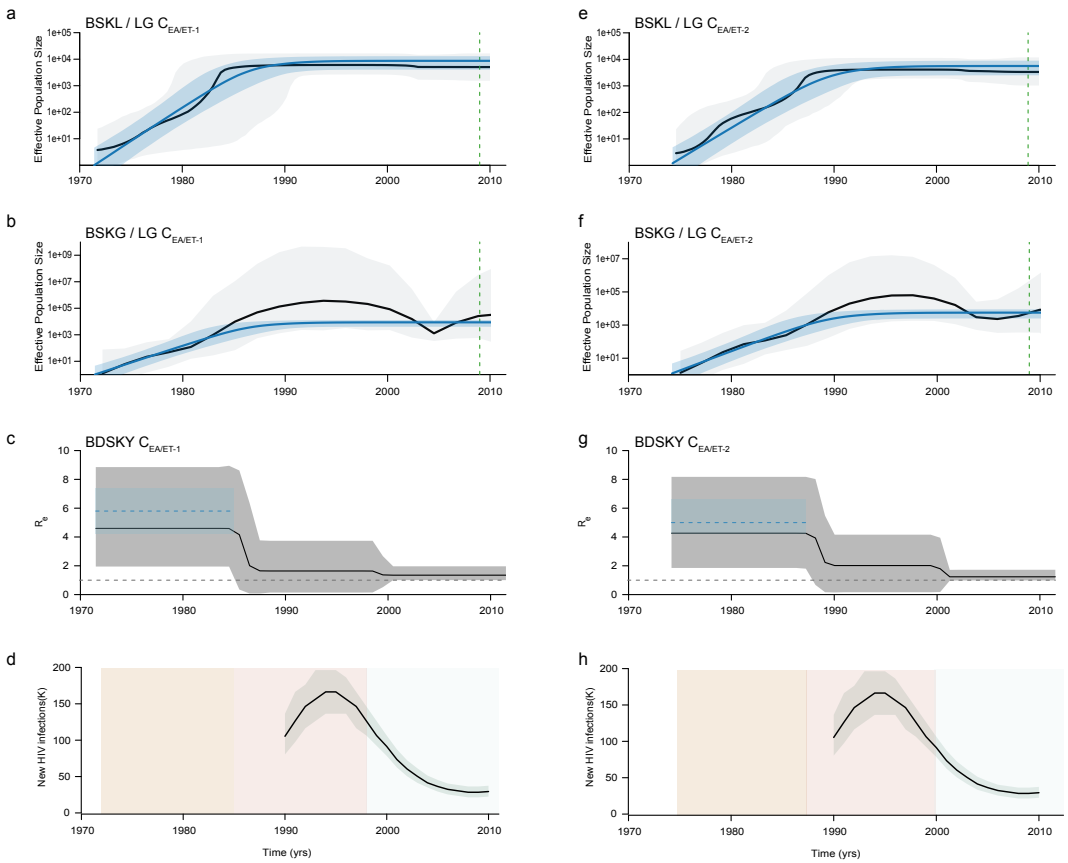


Figure 5.2: Epidemiological and population dynamics of the C_{EA} sub-epidemics in Ethiopia. Median estimates of the effective number of infections (*Ne*) using the Bayesian skyline or skygrid models (black lines) together with their 95% highest probability density (HPD) intervals (gray areas), co-plotted together with the median *Ne* estimates using the logistic coalescent-based parametric model (blue lines) and its 95% HPD intervals (blue areas). The green dashed lines indicate the time of the last coalescent event reported by the lineages-through-time (LTT) (a,b,e,f). Temporal fluctuation of the effective reproductive number (R_e) of the $C_{EA/ET\text{-}1}$ and $C_{EA/ET\text{-}2}$ sub-epidemics estimated using the Bayesian birth-death approach (c and g). For an easier visualization, the median coalescent-based R_0 estimate (blue dashed lines) inferred for each sub-epidemic and its 95% HPD intervals (blue area) were added. The gray dashed lines indicate R_e=1 (c and g). Plots representing the number of new HIV cases in Ethiopia as obtained from UNAIDS website http://aidsinfo.unaids.org/ (d and h). The yellow, pink and gray intervals denote the time spanned for the birth-death-based R_e-initial, R_e-middle and R_e-final estimates of each C_{EA} sub-epidemic.

Table 5.1: Evolutionary and demographic parameters estimated for HIV-1 C_{EA} subclades

Clade	N	Sampling interval	Method	Substitution rate (10^{-3})	TMRCA	R0/R_e-initial	R_e-middle	R_e-final
BI-RW	303	2002-2012	coalescent	1.5(1.3-1.6)	1958(1951-1964)	5.0 (4.2-5.8)	-	-
			birth-death	1.5(1.4-1.6)	1957(1952-1962)	3.8 (1.9 -7.1)	2.3(1.2-3.8)	0.5(0.3-0.7)
ET-1	63	2003-2011	coalescent	1.3 (1.0-1.7)	1971 (1969-1973)	5.8(4.2-7.4)	-	-
			birth-death	1.3 (1.0-1.5)	1972 (1970-1974)	4.6 (2.0-8.8)	1.6 (0.1-3.7)	1.3(1.0-2.0)
ET-2	56	2003-2012	coalescent	1.1 (0.9-1.4)	1974 (1972-1976)	5.0 (4.2-6.6)	-	-
			birth-death	1.1 (1.0-1.3)	1974 (1973-1976)	4.3 (1.9-8.1)	2.0 (0.2-4.2)	1.2 (1.0-1.7)
TZ	50	2004-2014	coalescent	1.1 (0.7-1.6)	1972(1960-1983)	5.8 (3.4-8.2)	-	-
			birth-death	1.1 (0.9-1.3)	1971(1966-1977)	4.2 (1.8-8.2)	2.8(1.2-5.4)	0.3(0.1 -0.6)
BR	144	1992-2014	coalescent	1.6(1.2-2.0)	1974(1966-1982)	5.9(4.4-7.6)	-	-
			birth-death	1.5(1.3-1.7)	1978(1974-1981)	4.9 (2.2-9.2)	1.0(0.2-2.1)	2.4(1.4-3.7)

The mean R_0 values derived from the logistic growth model using the formula R_0 = rD + 1 [15] (where D is average duration of infectiousness herein considered of eight years) was very similar for all C_{EA} subclades, ranging between 5.0 and 5.9 (Table 5.1). The median *Ne* trajectories of the $C_{EA/BI\text{-}RW}$, $C_{EA/TZ}$ and $C_{EA/BR}$ at the plateau phase obtained with parametric and non-parametric approaches were very similar. As the logistic growth model is unable to capture non-constant trends on recent time period estimates, obviously do not reproduced the declining phase pointed by the BSKG for the $C_{EA/ET\text{-}1}$ and $C_{EA/ET\text{-}2}$ sub-clades, however, the parametric and non-parametric models converge to similar median *Ne* values at the late plateau phase.

5.4.3 Bayesian population dynamics inference under birth-death model

The BDSKY model was applied allowing the R_e to change in a piece-wise manner over three different time intervals that allow us to estimate the epidemic potential before identification of HIV (first time interval, 1960–1970 to 1980–1984) and observe the potential impact of prevention (second time interval, 1981–1985 to 1995–1999) or therapy (third time interval, 1996–2000 to 2012–2014) measures on epidemic dynamics (Figs. 5.2 c, g, 5.3 c, g, 5.4 c).

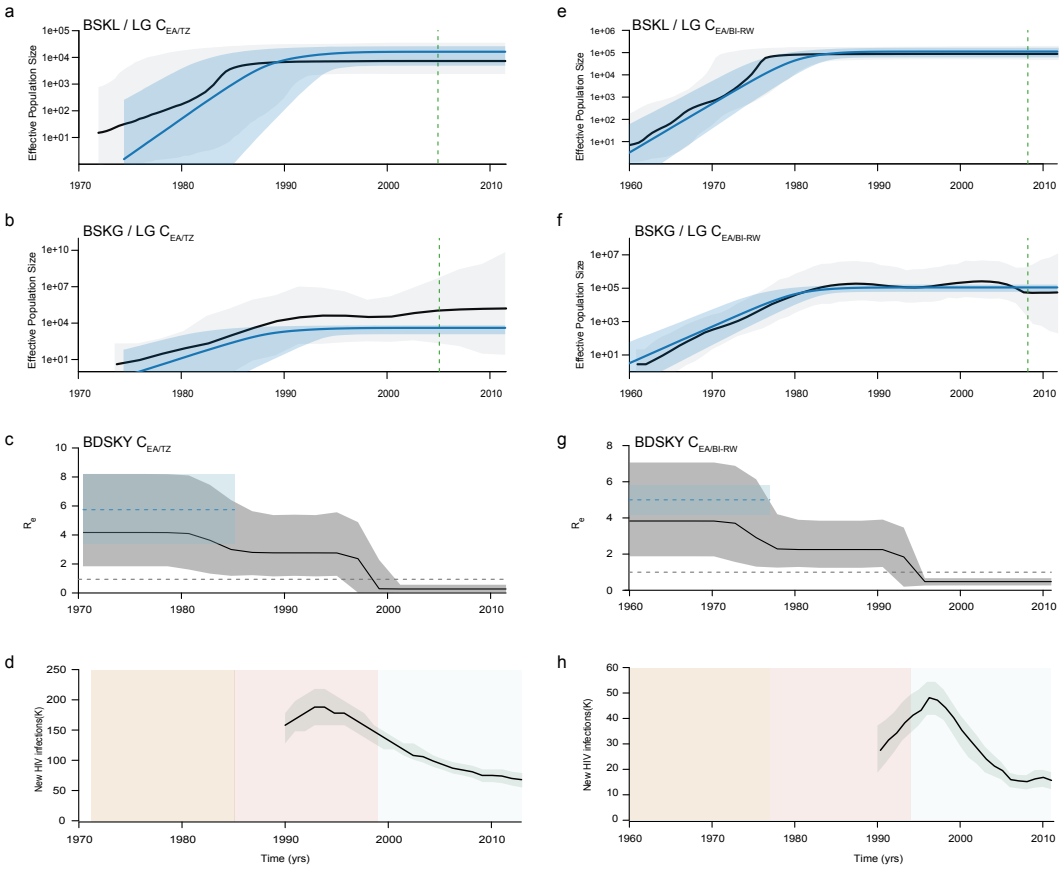


Figure 5.3: Epidemiological and population dynamics of the C_{EA} sub-epidemics in Tanzania and Burundi/Rwanda. Median estimates of the effective number of infections (*Ne*) using the Bayesian skyline or skygrid models (black lines) together with their 95% highest probability density (HPD) intervals (gray areas), co-plotted together with the median *Ne* estimates using the logistic coalescent-based parametric model (blue lines) and its 95% HPD intervals (blue areas). The green dashed lines indicate the time of the last coalescent event reported by the lineages-through-time (LTT) (a,b,e,f). Temporal fluctuation of the effective reproductive number (R_e) of the $C_{EA/TZ}$ and $C_{EA/BI\text{-}RW}$ sub-epidemics estimated using the Bayesian birth-death approach (c and g). For an easier visualization, the median coalescent-based R_0 estimate (blue dashed lines) inferred for each sub-epidemic and its 95% HPD intervals (blue area) were added. The gray dashed lines indicate R_e=1 (c and g). Plots representing the number of new HIV cases in Tanzania and Burundi/Rwanda as obtained from UNAIDS website `http://aidsinfo.unaids.org/` (d and h). The yellow, pink and gray intervals denote the time spanned for the birth-death-based R_e-initial, R_e-middle and R_e-final estimates of each C_{EA} sub-epidemic.

The BDSKY model support initial exponential growth dynamics fully consistent with those estimated using coalescent models. Although the mean R_0 values (5.0–5.9) were slightly higher than the estimated mean R_e-initial (3.8–4.9), the uncertainty on the R_0 estimates was always contained within the broader 95% HPD intervals of R_e-initial (Table 1).

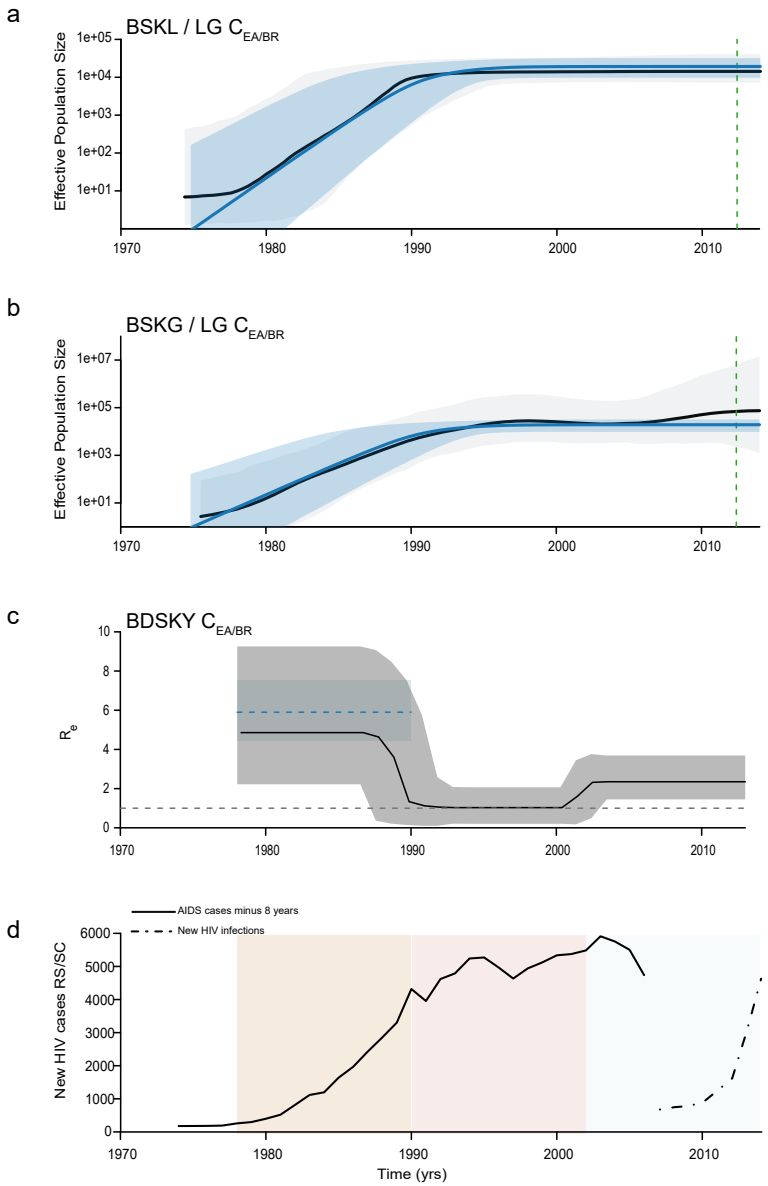


Figure 5.4: Epidemiological and population dynamics of the C_{EA} sub-epidemic in southern Brazil. Median estimates of the effective number of infections (*Ne*) using Bayesian skyline or skygrid models (black lines) together with their 95% highest probability density (HPD) intervals (gray areas), co-plotted together with the median *Ne* estimates using the logistic coalescent-based parametric model (blue lines) and its 95% HPD intervals (blue areas). The green dashed lines indicate the time of the last coalescent event reported by the lineages-through-time (LTT) (a and b). Temporal fluctuation of the effective reproductive number (R_e) of the $C_{EA/BR}$ sub-epidemic estimated using the Bayesian birth-death approach (c). For an easier visualization, the median coalescent-based R_0 estimate (blue dashed line) inferred for the C EA/BR subclade and its 95% HPD intervals (blue area) were added. The gray dashed line indicate R_e=1 (c). Plot representing the number of new HIV cases in the Southern Brazilian states of Rio Grande do Sul (RS) and Santa Catarina (SC). AIDS cases reported by the Brazilian AIDS cases databank (SINAN = SIM = SISCEL: http://www.portalsinan.saude.gov.br/dados-epidemiologicos-sinan), minus eight years, was used as an approximation for new HIV infections (solid black line). From 2007 onward, Brazilian Ministry of Health started to report HIV new infections (http://www.aids.gov.br/pt-br/pub/2016/boletim-epidemiologico-de-aids-2016), represented here as a dashed black line (d). The yellow, pink and gray intervals denote the time spanned for the birth-death-based R_e-initial, R_e-middle and R_e-final estimates of the $C_{EA/BR}$ sub-epidemic.

The BDSKY model supports a progressive reduction of the mean R_e in the second (mean R_e-initial >mean R_e-middle) and third (mean R_e-middle >mean R_e-final) time intervals for all African subclades. Although the mean R_e estimates for the Ethiopian clades remained above one during all time intervals, which clearly did not match with the declining and subsequent stabilization of the HIV incidence in Ethiopia from the mid-1990s onwards (Fig.5.2 d,h), those R_e estimates should be interpreted with caution because the extremely wide 95% HPD intervals (Table 1). For the $C_{EA/BI\text{-}RW}$ and $C_{EA/TZ}$ subclades, the R_e-middle was above one, while the R_e-final was below one, in agreement with the increasing HIV incidence in Burundi/Rwanda and Tanzania up to the mid-1990s and the subsequent declining from the mid-1990s onwards (Fig.5.3 d,h). For the $C_{EA/BR}$ subclade, the BDSKY model supports a reduction of the mean R_e in the second interval (R_e-initial $\sim$1) and a new increase in the third one (R_e-final >1). This is consistent with the HIV incidence trends in Rio Grande do Sul and Santa Catarina Brazilian states, that supports an epidemic stabilization during the 1990s and a new epidemic increase during the 2000s (Fig.5.4 c,d)

5.5 Discussion

In this study, we characterized key features of the epidemic dynamics of major HIV-1 subtype C_{EA} lineages circulating in East Africa and Southern Brazil through the use of different phylodynamic frameworks based in coalescent and birth-death process. The different coalescent models capture very similar epidemic dynamics over the earlier decades of the C_{EA} lineages dissemination; but point to quite different epidemic dynamics from the mid-1980s onwards.

Both phylodynamic approaches suggest an initial stage of fast exponential growth of all the C_{EA} sub-epidemics during the period of cryptic transmission of HIV in human populations. These initial phases of exponential growth herein inferred correlate with retrospective serological-based studies and simulations that indicated that an explosive epidemic was already sweeping the Eastern African region [36, 37, 38, 39, 40, 41] by 1981, when the AIDS was first recognized. The exponential growth phase inferred for the Brazilian C_{EA} sub-epidemic during the 1970s and 1980s is also fully consistent with the sharp increase in the number of new HIV cases detected in the southernmost Brazilian states during the 1980s [42]. The mean R_0 (5.0–5.9) and R_e-initial (3.8–4.9) values here estimated for all C_{EA} sub- clades were roughly comparable. The 95% HPD intervals of R_0 were always smaller and contained within the broader 95% HPD intervals of R_e -initial, consistent with previous empirical and simulated data [20,21]. The coalescent-based logistic growth model is expected to provide narrow HPD intervals because it considers a deterministic population trajectory, while the BDSKY model incorporates stochasticity in population size. Additionally, the HPD interval of R_e estimates from the BDSKY model grows wider the further we go towards the past, that is not the case for the coalescent-based logistic growth model estimates.

Importantly, the mean R_0 and R_e-initial values inferred for the African C_{EA} sub-epidemics were fully consistent with those estimated through analyses of HIV prevalence rate and life expectancy in Eastern African countries [43]. This suggests that both phylodynamic frameworks were able to recover the true early growth rates of HIV-1 C_{EA} sub-epidemics.

Factors like gender inequality [44, 45, 46], civil and ethnic conflicts [47,48, 49, 50], conflict-induced displacement, and increasing urbanization [51, 52] have shaped the early HIV epidemic dynamics across all Eastern African

countries, consistent with the similar epidemic growth rates of African C_{EA} sub-epidemics. Notably, the epidemic growth rate inferred for the Brazilian C_{EA} sub-epidemic was very similar to those obtained for the African C_{EA} sub-epidemics despite significantly distinctive history of human conflicts in those regions. The exponential growth phase of the $C_{EA/BR}$ sub-epidemic matches with a period in which public health system was unaware about the severity of the epidemic [53] and the $C_{EA/BR}$ subclade was efficiently disseminated in Southern Brazil through heterosexual (HET) networks [2, 13, 54], similar to that observed in Eastern Africa. This suggests that the absence of prevention efforts and the predominant viral transmission through HET route may have been the common driving forces of the early dynamics of the C_{EA} sub- epidemics in Eastern African and Southern Brazil.

The BSKL model supports that African C_{EA} sub-epidemics grew exponentially until between the late 1970s and late 1980s, after which there occurs a plateau in the *Ne* until the most recent sampling time of each of them. The stabilization of the *Ne* trajectories occurs around 10 years before the last coalescent event (Figs 2a,b,e,f and 3a,b,e,f), thus supporting that the inferred plateau of the *Ne* is not due to a paucity of coalescent events after the early 1990s [55]. More important, such stabilization occurred before implementation of prevention campaigns during the 1990s [56, 57, 58, 59, 60] and introduction of universal access to antiretroviral (ARV) therapy during the 2000s [61, 62] in Eastern Africa. The overall *Ne* trajectories inferred by the BSKL after the mid-1980s, however, differ markedly from the data of the United Nations Joint Program on HIV/AIDS (UNAIDS) [63] according to which the HIV incidence in Burundi/Rwanda, Ethiopia and Tanzania reached a peak around the mid-1990s (rather than during the 1980s), and was followed by a sharp decline (rather than a plateau) until the mid-2000s, before stabilize.

The overall epidemic dynamics inferred by the BSKG model from the mid-1980s to the mid-1990s are more consistent with the HIV incidence data than those inferred by the BSKL, although some divergences were also detected at later times. The BSKG model points that C_{EA} African sub-epidemics grew exponentially until the early/mid-1990s and further supports a declining *Ne* of $C_{EA/ET}$ subclades between the mid-1990s and the mid-2000s, consistent with epidemiological data. This model, however, failed to capture a similar decline of *Ne* for the $C_{EA/BI\text{-}RW}$ and $C_{EA/TZ}$ sub-epidemics. These results indicate that the BSKG model can correctly predict epidemic decline in some situations, as demonstrated here for the $C_{EA/ET}$ subclades and previously for the CRF02_AG epidemic in Cameroon [31, 64]; but not in others.

Although the BDSKY model also supports a progressive reduction of epidemic growth over time, an R_e >1 was estimated at the second time interval (that roughly covers the period between early/mid-1980s and mid/late 1990s) for all African C_{EA} subclades, consistent with a continuous increase of HIV incidence up to the mid-1990s. The BDSKY model capture an R_e <1 for the $C_{EA/BI\text{-}RW}$ and $C_{EA/TZ}$ sub- epidemics at the most recent time interval (after the mid-1990s); but supported an $R_e \geq 1$ for the $C_{EA/ET}$ sub-epidemics in the same time interval. These results confirms that the BDSKY model can correctly uncover a signature of a recent declining epidemic not reflected in the coalescent plots, as previously seen in the HIV-1 subtype B epidemic in the UK [21]; but also reveals that it may fail to capture such trend in some other datasets.

It is interesting to note that the BSKG failed to capture a decline of *Ne* for the $C_{EA/BI\text{-}RW}$ and $C_{EA/TZ}$ sub-epidemics since the middle-late 1990s onwards, while the BDSKY model failed to capture a $R_e \leq 1$ for the $C_{EA/ET}$ sub-epidemics at the same time interval, suggesting that the performance of

different phylodynamics approaches could be affected by different factors. The BSKG model requires strongly informative data to prevent erroneous estimates of *Ne* stabilization as pointed by a recent study [65]. The lower proportion of HIV-1 subtype C sequences sampled at recent times (since 2008 onwards) in the $C_{EA/BI\text{-}RW}$ (6%) and $C_{EA/TZ}$ (30%) datasets compared with the $C_{EA/ET}$ datasets ($\geq$70%) may have reduce the ability of this the coalescent model to capture changes in *Ne* for the $C_{EA/BI\text{-}RW}$ and $C_{EA/TZ}$ sub-epidemics at most recent times. The BDSKY model could be more robust to the paucity of coalescent events at most recent time; but its performance could be limited by the number of time intervals (changes in R_e) specified. Increasing the number of R_e-changes may allow a better fit of the R_e trajectories to the epidemiological data for the $C_{EA/ET}$ sub-epidemics. This strategy, however, resulted in a lack of parameter convergence and huge 95% HPD intervals, indicating that accurate R_e estimations at more time intervals would require a larger number of $C_{EA/ET}$ sequences than those used in the present our study.

The BSKL, BSKG and BDSKY models support quite consistent epidemic dynamics for the $C_{EA/BR}$ sub-epidemic until the late 1990s. According to the coalescent models, the *Ne* of the $C_{EA/BR}$ subclade growth exponentially until the early (BSKL) or mid-1990s (BSKG) and then reached a plateau. In agreement, the BDSKY model supports an expanding epidemic (R_e >1) in the first time interval ($\sim$ mid-1970s to late 1980s) and a transient epidemic stabilization ($R_e \sim 1$) in the second time interval ($\sim$ late 1980s to early 2000s). The stabilization of the $C_{EA/BR}$ incidence since the early/mid-1990s is in line with the reported trend toward stability of the new HIV cases in Rio Grande do Sul and Santa Catarina states since the mid-1990s [42], probably due to the implementation of prevention efforts that acted as the driven-force of people's behavioral changes [53, 66]. While coalescent models support a roughly constant *Ne* for the $C_{EA/BR}$ sub-epidemic until the most recent

sampling time, the BDSKY model uncovers a new epidemic increase (R_e >1) at the last time interval. This matches with an upward trend of new HIV diagnoses in Rio Grande do Sul and Santa Catarina states since 2007 [67]. Such epidemiological changes are probably too recent to be fully captured by coalescent models.

A recent study using BSKG to analyze the population dynamics of the $C_{EA/BR}$ sub-epidemic from *pol* and *env* sequences from HET and men having sex with men (MSM) individuals reported a continuous increase in the *Ne* until mid to late 2000's that was associated with the recent expansion of subtype C throughout the MSM group [13]. Interestingly, universal access to free fully suppressive ARV therapy is available in Brazil since the late 1990s [61, 62] and an association between ARV treatment availability and increases in sexual risk behavior (and consequent rise in HIV incidence) have been previously reported among MSM from developed countries [68, 69, 70, 71]. Our BDSKY analyses of sequences from HET individuals support that the recent expansion of the $C_{EA/BR}$ sub-epidemic is probably not restricted to a specific group, but also occurred among HET individuals. Increases in sexual risk behavior among HET individuals fully agrees with the sustained increase of HIV [67] and other sexually transmitted disease observed in Southern Brazil since 2010 [72].

A drawback to consider about the highlighted agreements and disagreements between the available epidemiological data and our phylodynamic modeling is that while the former characterizes the HIV epidemic of each of the countries/regions as a whole, the C_{EA} clade herein analyzed is not the only prevalent HIV lineage in all of them [5]. Then, it is possible that trends in the number of new HIV cases belonging to the C_{EA} sub-epidemics do not fully correspond with those of the overall HIV epidemic. Besides, a more homogeneous and dense sampling of each C_{EA}

sub-epidemic over time as well as the use of sequence data from multiple genetic loci [31] and the incorporation of covariates into the demographic inference framework [64] may improve the performance of phylodynamics methods to recover true population trajectories.

Overall, this study supports that major HIV-1 C_{EA} lineages circulating in Eastern Africa and Southern Brazil seem to have had an exponential spread with very similar growth rates until the early/mid-1990s. The overall agreement of the R_0 and R_e-initial values here estimated from genetic sequences with those previously obtained from classical epidemiological data strengthen the utility of coalescent and the birth- death phylodynamic approaches to infer relevant epidemiological information of HIV epidemics at the earlier stages. Our data supports that introduction of universal access to ARV therapy during the late 1990s and early 2000s coincides with a declining epidemic in Eastern Africa, but with an upward trend of new HIV diagnoses in Southern Brazil. Our results also underscore the importance of the joint use of both coalescent and birth-death phylodynamic approaches for the analyses of HIV population dynamics given the its apparent differential sensitivity for recovering changes in population dynamics at most recent times in different datasets.

5.6 References

1. Hemelaar, J., Gouws, E., Ghys, P. D. Osmanov, S. Global trends in molecular epidemiology of HIV-1 during 2000–2007. AIDS 25, 679–689 (2011).

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

3. Faria, N. R. et al. The early spread and epidemic ignition of HIV-1 in human populations. Science 346, 56–61 (2014).

4. Thomson, M. M. Fernández-García, A. Phylogenetic structure in African HIV-1 subtype C revealed by selective sequential pruning. Virology 415, 30–8 (2011).

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

6. Delatorre, E. et al. Tracing the Origin and Northward Dissemination Dynamics of HIV-1 Subtype C in Brazil. PLoS One 8, e74072 (2013).

7. Soares, M. A. et al. A specific subtype C of human immunodeficiency virus type 1 circulates in Brazil. AIDS 17, 11–21 (2003).

8. Bello, G. et al. Origin and evolutionary history of HIV-1 subtype C in Brazil. AIDS 22, 1993–2000 (2008).

9. Fontella, R., Soares, M. A. Schrago, C. G. On the origin of HIV-1 subtype C in South America. AIDS 22, 2001–2011 (2008).

10. de Oliveira, T., Pillay, D., Gifford, R. J. UK Collaborative Group on HIV Drug Resistance. The HIV-1 subtype C epidemic in South America is linked to the United Kingdom. PLoS One 5, e9311 (2010).

11. Gräf, T. et al. Contribution of Epidemiological Predictors in Unraveling the Phylogeographic History of HIV-1 Subtype C in Brazil. J. Virol. 89, 12341–12348 (2015).

12. Véras, N. M. C., Gray, R. R., Brígido, L. F. D. M., Rodrigues, R. Salemi, M. High-resolution phylogenetics and phylogeography of human immunodeficiency virus type 1 subtype C epidemic in South America. J. Gen. Virol. 92, 1698–1709 (2011).

13. Gräf, T. et al. Comprehensive Characterization of HIV-1 Molecular Epidemiology and Demographic History in the Brazilian Region Most Heavily Affected by AIDS. J. Virol. 90, 8160–8 (2016).

14. Bello, G. et al. Short communication: Evidences of recent decline in the expansion rate of the HIV type 1 subtype C and CRF31$_B$ *Cepidemicsinsouthern Brazil*. *AIDS Res. Hum. Retroviruses* 25, 1065˘1069(2009).

15. Pybus, O. G. et al. The epidemic behavior of the hepatitis C virus. Science 292, 2323–5 (2001).

16. Stadler, T. et al. Estimating the basic reproductive number from viral sequence data. Mol. Biol. Evol. 29, 347–357 (2012).

17. Fu, Y.-X. Exact coalescent for the Wright-Fisher model. Theor. Popul. Biol. 69, 385–94 (2006).

18. Diekmann, O., Heesterbeek, J. A. P. Metz, J. A. J. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. J. Math. Biol. 28, 365–382 (1990).

19. Rasmussen, D. A., Ratmann, O. Koelle, K. Inference for nonlinear epidemiological models using genealogies and time series. PLoS Comput. Biol. 7, e1002136 (2011).

20. Boskova, V., Bonhoeffer, S. Stadler, T. Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. PLoS Comput. Biol. 10, e1003913 (2014).

21. Stadler, T., Kuhnert, D., Bonhoeffer, S. Drummond, a. J. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc. Natl. Acad. Sci. 110, 228–233 (2013).

22. Pollakis, G. et al. 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 (2003).

23. Lole, K. S. et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73, 152–60 (1999).

24. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. Higgins, D. G. The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882 (1997).

25. Guindon, S. et al. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst. Biol. 59, 307–321 (2010).

26. Guindon, S., Lethiec, F., Duroux, P. Gascuel, O. PHYML Online - A web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 33, 557–559 (2005).

27. Posada, D. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–6 (2008).

28. Drummond, A. J. Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214 (2007).

29. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537 (2014).

30. Drummond, A. J., Rambaut, A., Shapiro, B. Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22, 1185–92 (2005).

31. Gill, M. S. et al. Improving bayesian population dynamics inference: A coalescent-based model for multiple loci. Mol. Biol. Evol. 30, 713–724 (2013).

32. Baele, G. et al. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 29, 2157–67 (2012).

33. Rambaut, A., Suchard, M. Drummond, A. Tracer v1.6. (2013). Available at: http://tree.bio.ed.ac.uk/software/tracer/.

34. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. Rambaut, A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88 (2006).

35. Rambaut, A., Lam, T. T., Max Carvalho, L. Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2, vew007 (2016).

36. Mhalu, F. S. et al. Some Aspects on the Epidemiology of AIDS and Infections with the Human Immunodeficiency Virus in the United Republic of Tanzania. in AIDS and Associated Cancers in Africa 50–60 (S. Karger AG, 1988). doi:10.1159/000415521

37. Serwadda, D. et al. Slim Disease: A new disease in Uganda and its association with HTLV-III infection. Lancet 326, 849–852 (1985).

38. Bayley, A. C. et al. HTLV-III serology distinguishes atypical and endemic Kaposi´s sarcoma in Africa. Lancet 325, 359–361 (1985).

39. Melbye, M. The natural history of human T lymphotropic virus-III infection: the cause of AIDS. Br. Med. J. 292, 5–12 (1986).

40. Jonckheer, T. et al. Cluster of HTLV-III/LAV infection in an African family. Lancet (London, England) 1, 400–1 (1985).

41. Stoneburner, R., Carballo, M., Bernstein, R. Saidel, T. Simulation of HIV incidence dynamics in the Rakai population-based cohort, Uganda. AIDS 12, 226–8 (1998).

42. Notification Aggregation Information System (SINAN) [in Portuguese]. Available at: http://www.portalsinan.saude.gov.br/dados-epidemiologicos-sinan. (Accessed: 15th November 2017)

43. Williams, B. G. Gouws, E. R0 and the elimination of HIV in Africa: Will 90-90-90 be sufficient? arXiv:1304.3720 Epub, 5–7 (2014).

44. Richardson, E. T. et al. Gender inequality and HIV transmission: a global analysis. J Int AIDS Soc 17, 19035 (2014).

45. Ramjee, G. Daniels, B. Women and HIV in Sub-Saharan Africa. AIDS Res. Ther. 10, 30 (2013).

46. Nyindo, M. Complementary factors contributing to the rapid spread of HIV-I in sub-Saharan Africa: a review. East Afr. Med. J. 82, 40–6 (2005).

47. Buvé, A., Bishikwabo-Nsarhaza, K. Mutangadura, G. The spread and effect of HIV-1 infection in sub-Saharan Africa. Lancet 359, 2011–2017 (2002).

48. UNAIDS. HIV/AIDS and conflict. (2003). Available at: http://data.unaids.org/topics/security/fs_conflict_en.pdf. (Accessed: 12th November 2017)

49. Spiegel, P. HIV/AIDS among conflict affected and displaced populations: dispelling myths and taking action. Disasters 28, 322–339 (2004).

50. Omare, D. Kanekar, A. Determinants of HIV/AIDS in armed conflict populations. J. Public Health Africa 2, e9 (2011).

51. Iliffe, J. The African AIDS epidemic: a history. (Oxford: James Currey, 2006).

52. Terminski, B. Development-Induced Displacement and Resettlement: Theoretical Frameworks and Current Challenges. Development 10, 101 (2013).

53. Levi, G. C. Vitória, M. A. a. Fighting against AIDS: the Brazilian experience. AIDS 16, 2373–2383 (2002).

54. Gräf, T. 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. 51, 186–191 (2011).

55. de Silva, E., Ferguson, N. M. Fraser, C. Inferring pandemic growth rates from sequence data. J. R. Soc. Interface 9, 1797–1808 (2012).

56. Hearst, N. Chen, S. Condom promotion for AIDS prevention in the developing world: Is it working? Stud. Fam. Plann. 35, 39–47 (2004).

57. Lugalla, J. et al. Social, cultural and sexual behavioral determinants of observed decline in HIV infection trends: Lessons from the Kagera Region, Tanzania. Soc. Sci. Med. 59, 185–198 (2004).

58. Ng'weshemi, J. Z. et al. Changes in male sexual behaviour in response to the AIDS epidemic: evidence from a cohort study in urban Tanzania. AIDS 10, 1415–20 (1996).

59. Macintyre, K., Brown, L. Sosler, S. 'It's not what you know, but who you knew': examining the relationship between behavior change and AIDS mortality in Africa. AIDS Educ. Prev. 13, 160–74 (2001).

60. Williams, B. G. HIV and TB in Eastern and Southern Africa: Evidence for behaviour change and the impact of ART. arXiv:1406.6912 1–9 (2014).

61. World Health Organization. Countries offering free access to HIV treatment. 1–2 (2005). Available at: http://www.who.int/hiv/countries_freeaccess.pdf. (Accessed: 21st November 2017)

62. '3 by 5' country information. WHO (2011). Available at: http://www.who.int/3by5/countryprofiles/en/. (Accessed: 21st November 2017)

63. United Nations Joint Program on HIV/AIDS (UNAIDS). AIDSinfo. Available at: http://aidsinfo.unaids.org/. (Accessed: 1st December 2017)

64. Gill, M. S., Lemey, P., Bennett, S. N., Biek, R. Suchard, M. A. Understanding Past Population Dynamics: Bayesian Coalescent-Based Modeling with Covariates. Syst. Biol. 65, 1041–1056 (2016).

65. Volz, E. M. Didelot, X. Modeling the Growth and Decline of Pathogen Effective Population Size Provides Insight into Epidemic Dynamics and Drivers of Antimicrobial Resistance. Syst. Biol. e-pub ahead print. (2018). doi:10.1093/sysbio/syy007

66. Berkman, A., Garcia, J., Muñoz-Laboy, M., Paiva, V. Parker, R. A Critical Analysis of the Brazilian Response to HIV/AIDS: Lessons Learned for Controlling and Mitigating the Epidemic in Developing Countries. Am. J. Public Health 9595, 1162–1172 (2005).

67. Brazilian Ministry of Health. AIDS Epidemiological Bulletin [in Portuguese]. (2016). Available at: http://www.aids.gov.br/pt-br/pub/2016/boletim-epidemiologico-de-aids-2016.

68. Bezemer, D. et al. A resurgent HIV-1 epidemic among men who have sex with men in the era of potent antiretroviral therapy. AIDS 1071–1077 (2008).

69. Le Vu, S. et al. Population-based HIV-1 incidence in France, 2003–08: a modelling analysis. Lancet Infect. Dis. 10, 682–687 (2010).

70. The Kirby Institute/Sydney University of New South Wales. HIV, viral hepatitis and sexually transmissible infections in Australia: Annual Surveillance Report. (2011). Available at: https://kirby.unsw.edu.au/sites/default/files/kirby/report/SERP_2011-Annual-Surveillance-Report.pdf. (Accessed: 29th November 2017)

71. White, P. J., Ward, H. Garnett, G. P. Is HIV out of control in the UK? An example of analysing patterns of HIV spreading using incidence-to-prevalence ratios. AIDS 20, 1898–1901 (2006).

72. Brazilian Ministry of Health. Syphilis Epidemiological Bulletin [in Portuguese]. (2017). Available at: http://www.aids.gov.br/pt-br/pub/2017/boletim-epidemiologico-de-sifilis-2017.

5.7 Supplementary Information

Table S 5.1: HIV-1 subtype C *pol* sequences from Africa and America used for ML phylogenetic analyses.

Region	Country	Total	C_{EA}	Sampling Time
East Africa	Burundi	284	281	2002-2012
	Ethiopia	233	179	2002-2013
	Kenya	55	44	2004-2012
	Rwanda	22	22	2005-2012
	Tanzania	177	76	2003-2014
	Uganda	62	58	1990-2010
South Africa	Zambia	148	0	1998-2008
Central Africa	Democratic Republic of the Congo	22	4	2002-2007
America	Brazil	144	144	1992-2014

Table S 5.2: Best fit demographic model for major country-specific HIV-1 C_{EA} subclades.

Clade	Model	PSLog ML	Models compared	Log BF	SSLog ML	Models compared	Log BF
BI-RW	**Log**	**-37755.2**	-	-	**-37766.1**	-	-
	Expo	-37912.4	Log/Expo	157.2	-37917.5	Log/Expo	151.4
	Expa	-37930.9	Log/Expa	175.7	-37930.9	Log/Expa	164.8
ET1	**Log**	**-9237.7**	-	-	**-9238**	-	-
	Expo	-9328.4	Log/Expo	90.7	-9328.8	Log/Expo	90.8
	Expa	-9336.2	Log/Expa	98.5	-9336.5	Log/Expa	98.5
ET2	**Log**	**-8111.9**	-	-	**-8112.2**	-	-
	Expo	-8176.5	Log/Expo	64.6	-8176.7	Log/Expo	64.5
	Expa	-8184.8	Log/Expa	72.9	-8185.2	Log/Expa	73
TZ	**Log**	**-7878.2**	-	-	**-7878.5**	-	-
	Expo	-7901	Log/Expo	22.8	-7901.3	Log/Expo	22.8
	Expa	-7910.3	Log/Expa	32.1	-7910.9	Log/Expa	32.4
BR HET-S	**Log**	**-19564.8**	-		-19568.3	-	
	Expo	-19651.6	Log/Expo	86.8	-19654.2	Log/Expo	85.9
	Expa	-19654.1	Log/Expa	89.3	-19656.8	Log/Expa	88.5

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 > 1 indicates that model H1 is more strongly supported by the data than model H0.

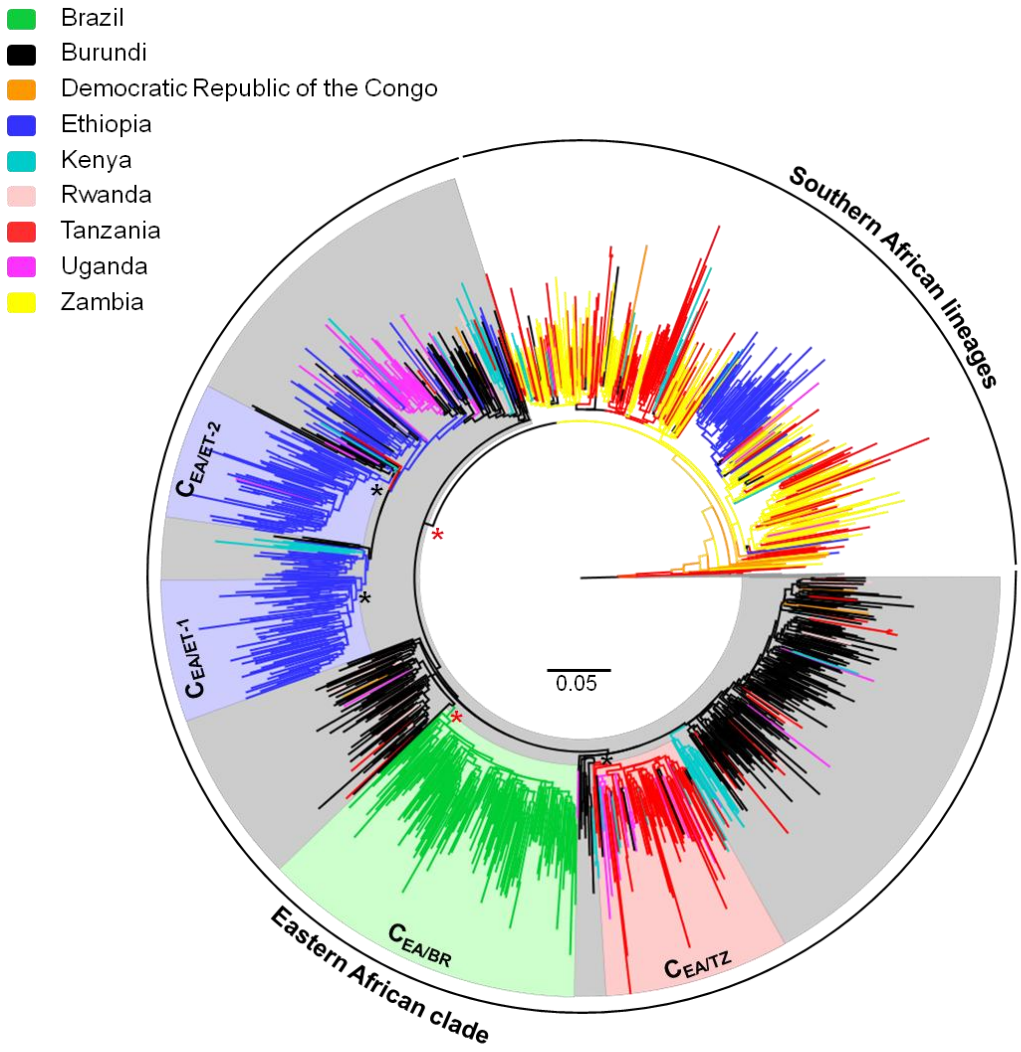


Figure S 5.1: ML phylogenetic tree of HIV-1 C *pol* PR/RT sequences (∼1,000 nt) from east, southern and central Africa and southern Brazil. Branches are colored according to the geographic origin of sequences as indicated in the legend (upper left). Gray shaded box indicate the positions of the Eastern African clade (C_{EA}) and coloured boxes indicate the position of major C_{EA} lineages. Red asterisks point to key nodes with SH-aLRT $\geq$ 0.90 and black asterisks point to key nodes with SH-aLRT > 0.85. The tree was rooted using HIV-1 subtypes A1 and D reference sequences and the branch lengths are drawn to scale with the bar at the center indicating nucleotide substitutions per site.

Chapter 6

General Discussion and Conclusions

There has been a long way in the development of epidemiological models to study infectious epidemics trends. From the mathematical field, perhaps, the first model of transmission disease used to assess interventions strategies was the one developed by Ronald Ross at the beginning of the 20th century aimed to the eradication of malaria [1], [2]. Following up the work of Ross, Kermack and McKendrick introduced the compartmental epidemic models based in the track of the host flow between Susceptible/Infected/Recovered-Removed compartments over the time of an epidemic [3]–[5]. To this end, a set of ordinary differential equations was defined by the model through which is possible to estimate a central parameter in epidemic theory, namely: the basic reproductive number (R_0). The R_0 is defined as the expected number of secondary cases produced by an infected individual during its entire period of infectiousness in a completely susceptible population and its magnitude is a useful indicator of both the risk of an emerging epidemic (when $R_0 > 1$) and the potential end of an established epidemic (when $R_0 < 1$) [6]. The R_0 is also a useful indicator of the effort required to control an infection.

It is on this basis, that since the past century onwards, mathematical epidemiological models have been extremely useful for the estimation of key epidemic parameters that in turn allow better-targeted public health policies [7], [8]. Despite the great value of the standard mathematical models for the analysis of infectious disease dynamics, their reliance on surveillance data constitutes a major shortcoming. Surveillance data is expensive and difficult to collect, and the estimates based on them are highly error prone due to variations of the reporting rate and the intensity of surveillance [9]. More importantly, the exclusive use of surveillance data for the analyses of epidemiological dynamics conspicuously omit the large amount of information that genetic sequences can provide [10].

Through the Chapters 3 , 4 and 5 that make up this thesis we characterized the evolutionary dynamics of some of the most epidemiologically relevant HIV-1M clades through a Bayesian phylodynamic approach based on genetic sequence data. In a Bayesian context, the likelihood of the population and epidemiological parameters for a given genealogy is the tree prior. There are two main strategies to model the tree prior within a Bayesian phylodynamic analyses: by using the coalescent [11]–[13] or the birth–death [14], [15] models.

In this way, key population and epidemiological parameters which would be very difficult to assess through classic surveillance means, were estimated for three major HIV-1M lineages, namely: subtype B (the most geographically disseminated HIV clade worldwide), subtype C (the most prevalent HIV clade worldwide), and the CRF02_AG (the most prevalent recombinant HIV clade worldwide). The R_0 values for the HIV-1 $B_{PANDEMIC}$ and CRF02_AG clades described in Chapters 3 and 4, respectively, were derived from the coalescent-based epidemic growth rate values using the formula $R_0 = rD + 1$ [16] (where D is the average duration

of infectiousness). The R_0 and the R_e for the HIV-1 C_{EA} clades analyzed in Chapter 5 were estimated using both coalescent and birth-death models.

6.1 The R_0 of major HIV-1 $B_{PANDEMIC}$ lineages circulating in Latin America

Through the coalescent-based phylodynamic analyses of HIV-1 subtype B *pol* sequences retrieved from viruses circulating in Latin America presented in Chapter 3, we characterized 12 major country-specific (or region-specific) $B_{PANDEMIC}$ clades with particular successful epidemic outcomes that together comprise 36% of all subtype B sequences from the region under analysis. Their median tMRCA estimates embraces a period of two decades between the late 1960s and the late 1980s and the mean R_0 estimates related to their initial phases of epidemic exponential growth ranged from 5.0 to 8.4 (Fig. 6.1). Such results triggered us to questions about the processes that shaped the early dynamics of the major Latin American $B_{PANDEMIC}$ clades and why in most Latin American countries analyzed the $B_{PANDEMIC}$ lineages were much more successful disseminated than the $B_{NON\text{-}PANDEMIC}$ clades. Comparing our results with those of previously characterized $B_{PANDEMIC}$ [17], [18] and $B_{NON\text{-}PANDEMIC}$ [19] clades from Latin America and the Caribbean we conclude that: 1) because variations in the R_0 took place for the same regional $B_{PANDEMIC}$ clade circulating in different countries, viral-specific genetic properties are certainly not the answer for the different epidemic potentials; and 2) given the temporal overlap in the introduction of $B_{PANDEMIC}$ and $B_{NON\text{-}PANDEMIC}$ clades in Latin America, the greater epidemic success of the $B_{PANDEMIC}$ lineages in this region is probably not a consequence of having reached the susceptible population before the $B_{NON\text{-}PANDEMIC}$ clades (Fig. 6.2).

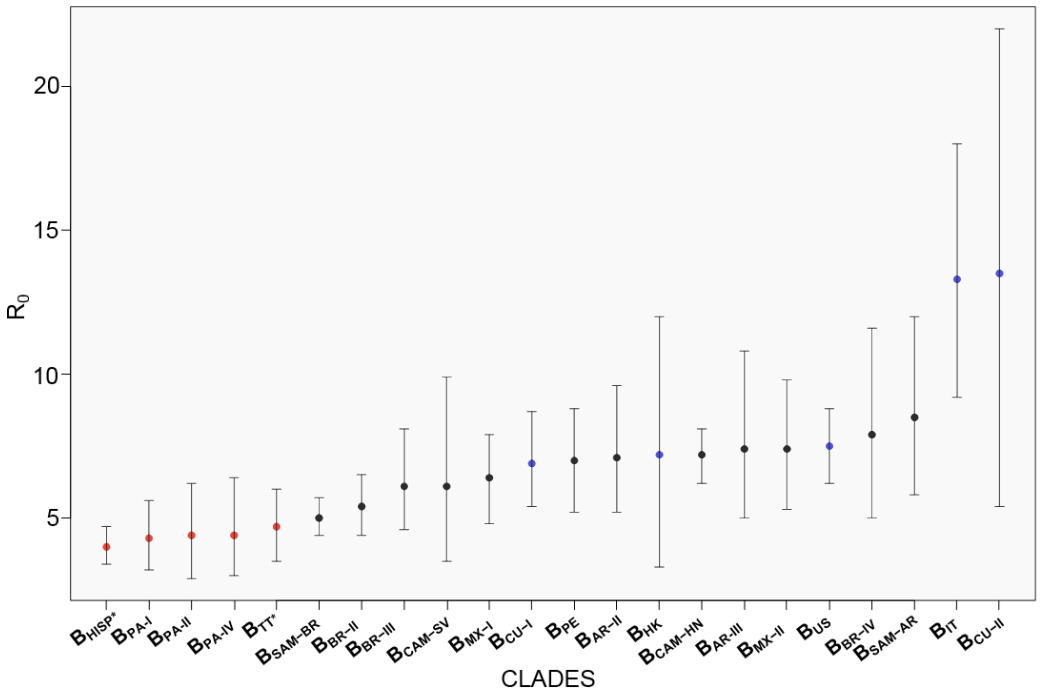


Figure 6.1: Coalescent R_0 estimates of major worldwide characterized HIV1 $B_{PANDEMIC}$ and $B_{NON\text{-}PANDEMIC}$ (indicated with an asterisk) clades. The circle and the vertical lines represent the median R_0 and the corresponding 95% HPD intervals of the posterior distributions estimated under the logistic growth coalescent model for each clade. Subtype B clades mostly associated to heterosexual (red circles), MSM (blue circles) or unknown (black circles) transmission networks are indicated. Epidemic growth rates of subtype B clades circulating in Cuba (CU), Hispaniola (HISP), Jamaica (JM), Trinidad and Tobago (TT), Panama (PA), United States (US), Italy (IT) and Hong-Kong (HK) were estimated previously in Delatorre et al. 2013, Cabello et al. 2014, Mendoza et al. 2014, Robbins et al. 2003, Worobey et al. 2016, Zehender et al. 2010 and Chen et al. 2011.

Notably, subtype B clades mainly associated to an heterosexual (HET) mode of transmission like the B_{CAR} and the Panamanian $B_{PANDEMIC}$ clades, are those for which the lowest R_0 values were estimated (B_{HISP} = 4.0, $B_{PA\text{-}I}$ = 4.3, $B_{PA\text{-}II}$ = 4.4, $B_{PA\text{-}IV}$ = 4.4 and B_{TT} = 4.7) (Fig. 6.1). Much higher mean R_0 values were reported for $B_{PANDEMIC}$ lineages associated to MSM populations from Cuba (R_0 = 6.9-13.5) [18], the US (R_0 = 7.5-7.7) [20], [21], Italy (R_0 = 13.3) [22] and Hong-Kong (R_0 = 7.2) [23] (Fig. 6.1). These results support the idea that differences in the transmission potential during the early epidemic exponential growth phase among subtype B clades may reflect differences in the epidemiological characteristics of the transmission networks. This hypothesis suggests that those Latin American $B_{PANDEMIC}$

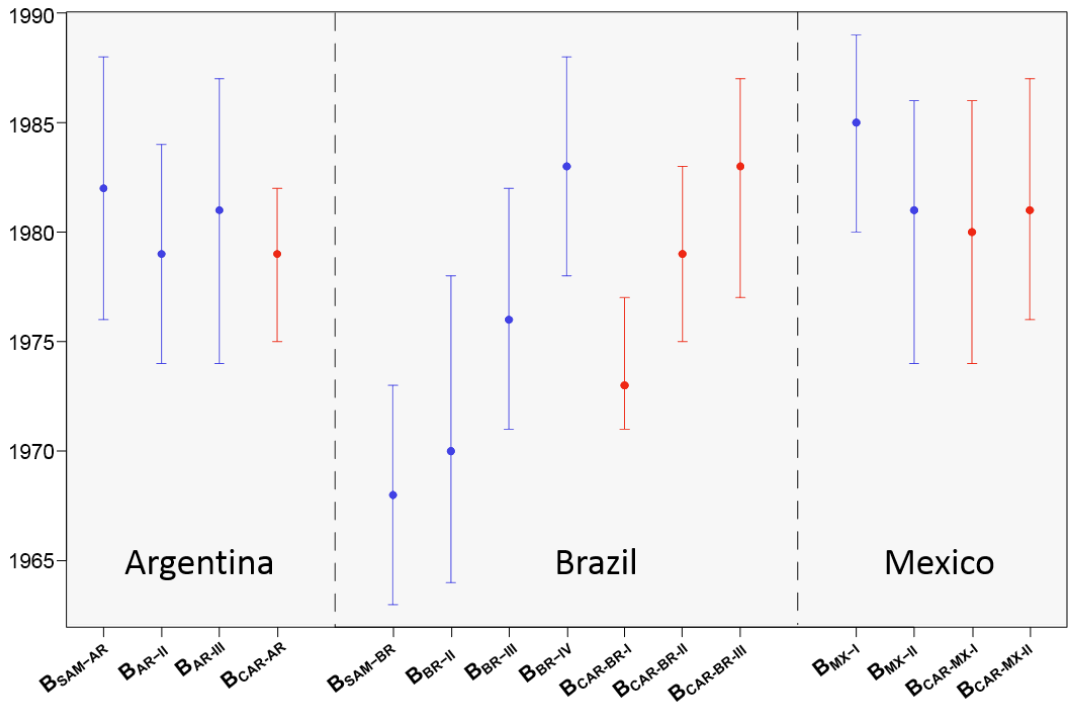


Figure 6.2: tMRCA estimates of major characterized B_{CAR} and $B_{PANDEMIC}$ clades circulating in Argentina, Brazil and Mexico. Epidemic growth rates of B_{CAR} clades were estimated previously in Cabello et al. 2015

lineages with higher R_0 values (B_{PE} = 7.0, $B_{AR\text{-}II}$ = 7.1, $B_{CAM\text{-}HN}$ = 7.2, $B_{AR\text{-}III}$ = 7.4, $B_{MX\text{-}II}$ = 7.4, $B_{BR\text{-}IV}$ = 7.9 and $B_{SAM\text{-}AR}$ = 8.5) probably followed their early epidemic trends into fully connected networks predominantly associated with populations of MSM or IDUs; while those with lower inferred R_0 values ($B_{SAM\text{-}BR}$ = 5.0, $B_{BR\text{-}II}$ = 5.4, $B_{BR\text{-}III}$ = 6.1, $B_{CAM\text{-}SV}$ = 6.1 and $B_{MX\text{-}I}$ = 6.4) were likely mainly disseminated across transmission networks with greater proportions of HET infections. Unfortunately, the scarce availability of epidemiological data linking each of the analyzed Latin American $B_{PANDEMIC}$ clades to a prevailing risk group does not allow us to test this hypothesis.

6.2 The R_0 of major HIV-1 CRF02_AG lineages circulating in Western Africa

The phylodynamic analysis performed on HIV-1 CRF02_AG *pol* sequences from viruses circulating in West and West-Central African countries presented in Chapter 4 aimed to identify the major regional clades of this prevalent HIV-1 lineage and characterize their epidemic dynamics in that sub-Saharan region. We identified five major African clades within the radiation of the CRF02_AG of which $CRF02_{CM\text{-}I}$, $CRF02_{CM\text{-}II}$, $CRF02_{CM\text{-}III}$ and $CRF02_{CM\text{-}IV}$ mostly circulate in Cameroon whereas the clade $CRF02_{WA}$ was the most prevalent one in West Africa and at a global level. Their median tMRCA estimates embraces a period of two decades between the late 1960s and the middle 1980s and the mean R_0 estimates related to their initial phases of epidemic exponential growth ranged from 4.3 to 7.0 (Fig. 6.3). To explain the biological processes that shaped the variability of the mean R_0 estimates for the major CRF02_AG clades circulating in West and West-Central Africa is a complex task.

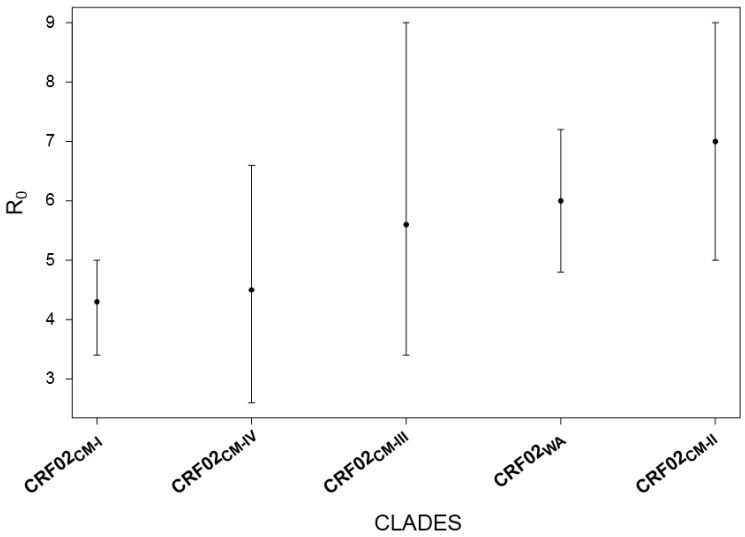


Figure 6.3: Coalescent R_0 estimates of major HIV-1 CRF02_AG lineages circulating in Western Africa. The circle and the vertical lines represent the median R_0 and the corresponding 95% HPD intervals of the posterior distributions estimated under the logistic growth coalescent model for each clade.

Spatial accessibility [24] and human mobility [25] (sometimes forced) between key population centers were pointed as key driving forces of the HIV spread in the African continent. The large variability of R_0 estimates obtained for CRF02_AG lineages co-circulating in the same country (Cameroon), however, supports that the expansion rate of the different CRF02_AG lineages circulating in West and West-Central Africa was also influenced by factors other than spatial accessibility. We also found no evidence to support the hypothesis that those CRF02_AG clades emerging earliest (tMRCA around late 1960s) displayed faster growth rates than those clades emerged at most recent times (tMRCA between the late 1970s and the middle 1980s). Differences in the proportion of HET and MSM individuals across different transmission networks are probably not a major cause for the growth rate variances observed because most HIV transmissions in Africa occur through unprotected HET contacts [26]–[28]. We cannot discard, however, the possible existence of a compartmentalized epidemic, particularly in Cameroon, within which the CRF02_AG lineages infected different risk or ethnic groups and/or the existence of clade-specific genetic factors that resulted in different viral transmissibility properties.

6.3 The R_0 of major HIV-1 CRF02_AG lineages circulating in Europe and Asia

Within the $CRF02_{WA}$ radiation analyzed in Chapter 4, we characterize two autochthonous transmission networks out of Africa: the sub-clade $CRF02_{BG\text{-}DE}$ including sequences sampled from Bulgaria and Germany and the sub-clade $CRF02_{FSU}$ composed by sequences from the former Soviet Union (FSU) countries (Russia, Armenia, Kazakhstan, Uzbekistan and Ukraine) which probably arose at around the early 2000s and the late 1990s,

respectively. The mean R_0 estimated for these two clades circulating in European and FSU countries were very similar among each other (14.6-17.0) and three to five times greater than those estimated for African CRF02_AG clades (Fig. 6.3). It is, however, not surprising to obtain such a high R_0 values given the epidemiological link between the CRF02_AG lineage and highly connected IDUs transmission networks reported in Bulgaria [29] as well as in Kazakhstan [30], [31], Uzbekistan [32], Kyrgyzstan [33], and the Russian Federation [34]. Notably, the origin and explosive growth of the $CRF_{BG\text{-}DE}$ and $CRF02_{FSU}$ clades, may be associated to the establishment of two main international heroin traffic routes linking Afghanistan (the world's largest opium producer) to the markets of the Russian Federation and Western Europe, namely: the Balkan trafficking route, which starts in Turkey and goes via Bulgaria, Romania, the former Yugoslavia, Hungary, Slovakia and the Czech Republic, finally reaching Austria and Germany [35] and the Northern traffic route that covers the territory of Central Asia, the Caucasus, Russia, Ukraine and Poland [35].

6.4 The R_0 of major HIV-1 subtype C lineages circulating in East Africa and Brazil

In sharp contrast to the wide variation in R_0 values estimates for major subtype B clades circulating in the Americas and CRF02_AG clades circulating in West and West-Central Africa, the mean R_0 values (5.0-5.8) inferred in Chapter 5 of this thesis for the major HIV-1 subtype C sub-epidemics in Burundi, Rwanda, Ethiopia Tanzania and Brazil (Fig 6.4) were roughly similar. It is not surprising that African C_{EA} sub-epidemics display very similar epidemic growth rates given that factors like gender inequality [36]–[38], ethnic conflicts [39]–[42], conflict-induced

displacements and labor migration [43], [44] were common cultural and socioeconomic features that shaped the early HIV epidemic dynamics across all Eastern African countries.

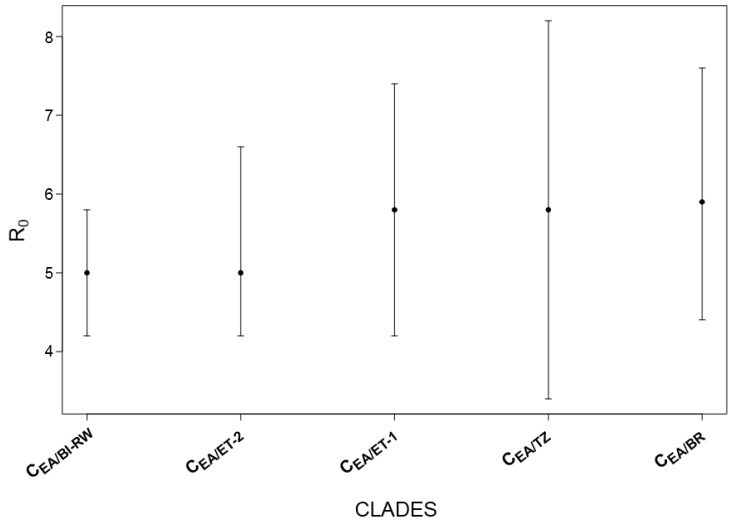


Figure 6.4: Coalescent R_0 estimates of major HIV-1 C_{EA} sub-epidemics. The circle and the vertical lines represent the median R_0 and the corresponding 95% HPD intervals of the posterior distributions estimated under the logistic growth coalescent model for each clade.

Much more surprisingly, however, were the very similar R_0 values of African and Brazilian C_{EA} sub-epidemics despite the significantly distinctive history and epidemiological context of these regions (Fig 6.4). Some common structural factors, including the insufficient of knowledge about HIV transmission and the predominant dissemination through HET transmissions, may be the key determinants of these similar mean R_0 estimates inferred for the C_{EA} sub-epidemics in both geographic regions. The inferred timeframe in which the Eastern African and Southern Brazilian C_{EA} sub-epidemics grew exponentially (late 1960s to late 1980s) matches with a period in which public health government agencies were unaware about the severity of the epidemic, therefore prevention efforts were not yet abundant. Furthermore, the epidemic success of the Southern Brazilian C_{EA} subclade in Southern Brazil, as well as of those circulating in Eastern Africa, was linked to its efficient dissemination in an expanding HET network [45]–[47].

6.5 The R_0 of HIV-1 in HET transmission networks

The mean R_0 values estimated in this thesis and in previous studies [48], [49] for the major HIV-1 lineages mainly spreading among HET populations from the Americas, West Africa and Eastern Africa were widely variable. In general, the most prevalent HIV-1 clades circulating in HET populations from West Africa (CRF02_AG, CRF06_cpx and subtype G) tend to spread at faster rate (R_0 = 6.0-8.6) than HIV-1 clades circulating in HET populations from Eastern Africa (subtype C, R_0 = 5.0-5.8) and from Latin American and the Caribbean (subtypes B and C, R_0 = 4.0-5.9) (Fig. 6.5). Of note, the relative different transmission potentials inferred for the distinct HIV-1 clades cannot be associated to any factor related to their time scale of dissemination given the great overlap of their mean tMRCA estimates.

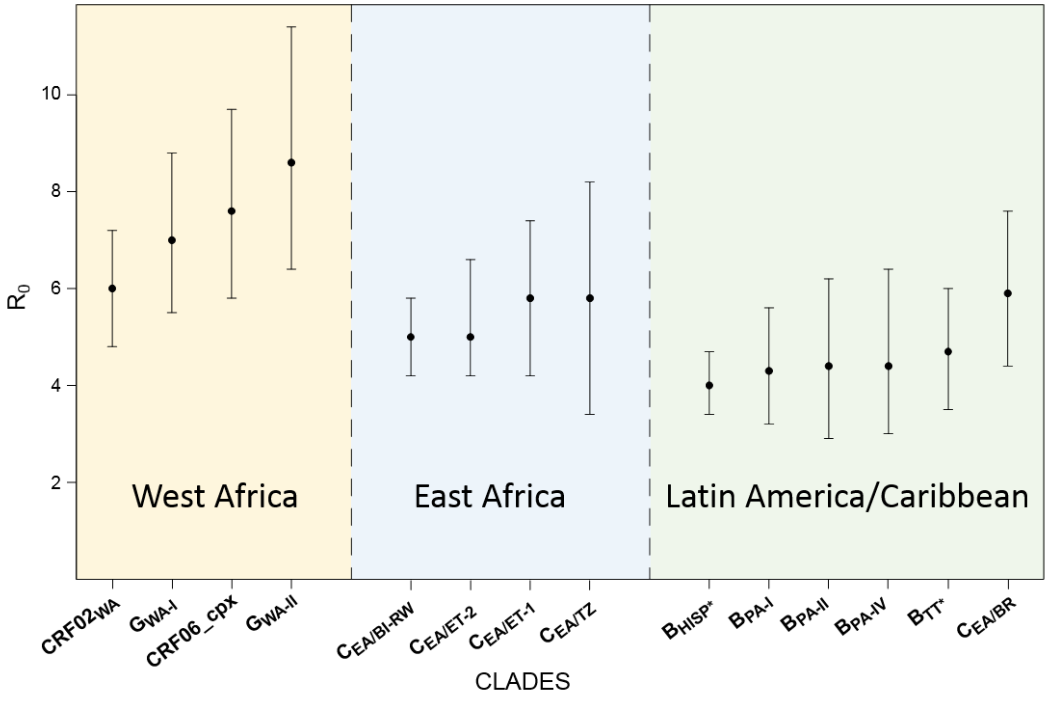


Figure 6.5: Coalescent R_0 estimates in HET transmission networks spreading in the Americas, West and East Africa The circle and the vertical lines represent the median R_0 and the corresponding 95% HPD intervals of the posterior distributions estimated under the logistic growth coalescent model for each clade. Epidemic growth rates of G_{WA-I}, G_{WA-II}, CRF06_cpx, B_{HISP*}, B_{PA-I}, B_{PA-II}, B_{PA-IV} and B_{TT*} clades were estimated previously in: Delatorre et al. 2014a, Delatorre et al. 2014b, Mendoza et al. 2014 and Cabello et al. 2014.

Religious and cultural practices have been proven to have a great impact in HIV transmission (both positive and negative), and thus in the rate of HIV-1 spread, among HET populations from different regions. Circumcision in sub-Saharan Africa has been associated with traditional religious practices of indigenous ethnic groups [50]–[52] and has been also identified as a protective factor against HIV infection by reducing the susceptibility to other communicable diseases that are known to act as co-factors for the HIV infection (e.g. syphilis and herpes) [53], [54]. Several studies also point to a negative association between Islam and HIV prevalence rate fueled by the conservative social norms (mandatory circumcision, reduced alcohol consumption and criminalization of adultery) of this religion [28], [55]–[57]. The higher proportion of circumcised male population and/or Muslim people in Western Africa compared to Eastern Africa, Latin America and the Caribbean regions [27], [58]–[60], however, do not support any negative relationship between those religious and cultural practices and the rate of HIV-1 expansion. These results, in turn, agree with available researches that downplay the linkage of HIV to religion, attributing to it a modest role in the promotion of abstinence, faithfulness and condom use [61]–[66].

Viral genetic characteristics may have also a great impact in HIV transmission dynamics. Several studies reported a less virulence (less pathogenic fitness) of the HIV-1 subtype C variant in comparison with other HIV-1 group M subtypes [67]–[70]. These findings support the hypothesis that the dominance of subtype C in the global epidemic may be related to its lower pathogenicity, which would increase the asymptomatic period and allow a greater number of secondary transmissions. However, the mean R_0 values estimated for the C_{EA} subclades were slightly lower than those estimated for highly prevalent HIV-1 lineages circulating in Western Africa and comparable to those estimated for the more virulent subtype B. Thus,

our estimations suggest that despite its lower virulence, the HIV-1 subtype C seems not to have spread in Eastern Africa and Southern Brazil at a faster rate than other prevalent HIV-1 group M clades circulating in the Americas and Western Africa.

6.6 The stabilization of HIV-1 epidemics

A common finding in all the American and African HIV-1 epidemic dynamics inferred through the chapters of this thesis is their good fit to a logistic model of epidemic growth. The transition between the initial exponential growth phase and the plateau state of the Ne suggested by the coalescent-based BSP and logistic models supposes an imminent stabilization in the incidence rate of all viral dynamics under analysis.

Despite the inferred slow-down in the incidence rate of the $CRF02_{WA}$ since early 1980s, according to the United Nations Joint Program on HIV/AIDS (UNAIDS) the number of new HIV infections reached a peak and started to decline only between 1990 to 1999 in most of the Western African countries where this clade disseminates [71] (Fig. 6.6). Similarly, according to the UNAIDS data there was a sustained increase in the number of new HIV infection in Cameroon until the year 2000 [71] (Fig. 6.6), while our inferences support an epidemic stabilization of the Cameroonian CRF02_AG clades between 1985 and 1995. Such time discrepancies are probably related to the fact that the analyzed CRF02 clades are not the only HIV-1 lineages of epidemiological relevance in those regions. The epidemiological data in West Africa likely reflects as well, the incidence rate dynamic of more recently introduced HIV-1 clades such as the subtype G and the CRF06_cpx [48], [49]; while epidemiological data in Cameroon also comprises a wide range of subtypes, CRFs and URFs [72]–[78]. The high prevalence of HIV-1C

and of the C_{EA} lineages among subtype C strains in Burundi, Ethiopia and Tanzania [79]–[88] is probably the key factor of the time consistency between the onset of decline in number of new HIV cases in mid 1990s according to the UNAIDS data (Fig. 6.7) and the start of the epidemic stabilization inferred by our coalescent-based phylodynamic analyses for all the analyzed African C_{EA} sub-epidemics.

Subtype B is the predominant HIV-1 lineage reported in Mexico [89]–[91], El Salvador [92]–[94], Honduras [95]–[100] and Peru [101]–[105], accounting for $\geq$ 98% of the HIV infections in those countries. Although the lack of epidemiological data from Peru makes it impossible to highlight any agreements or disagreements between the inferred dynamic for the B_{PE} clade and empirical data, the onset of the epidemic stabilization inferred around mid 1990s for the Mexican, Salvadoran and Honduran $B_{PANDEMIC}$ clades ($B_{MX\text{-}I}$, $B_{MX\text{-}II}$, B_{CAMSV}, B_{CAMHN}), coincides with the timing of maximum growth in the number of new HIV cases in such countries [71] (Fig. 6.8). These results are consistent with the fact that the concordance between the epidemiological data and the inferred viral dynamics depends on the prevalence of the specific HIV clade in the context of the overall epidemic of the country/region from which it belongs.

Despite CRF12_BF and related BF recombinants account for about 50% of the HIV infections in Argentina, we also observed a good agreement between the inferred dynamic for the B_{AR} clades and the epidemiological data (Fig. 6.8) in that country. The synchronization of the population growth dynamics previously inferred for the CRF12_BF [107] and the ones inferred in Chapter 3 of this thesis for the Argentinean $B_{PANDEMIC}$ clades, is probably the main cause for the consistency between the epidemiological data and the phylodynamic inferences in the indicated decline of the HIV epidemic growth rate in this country since early 1990s.

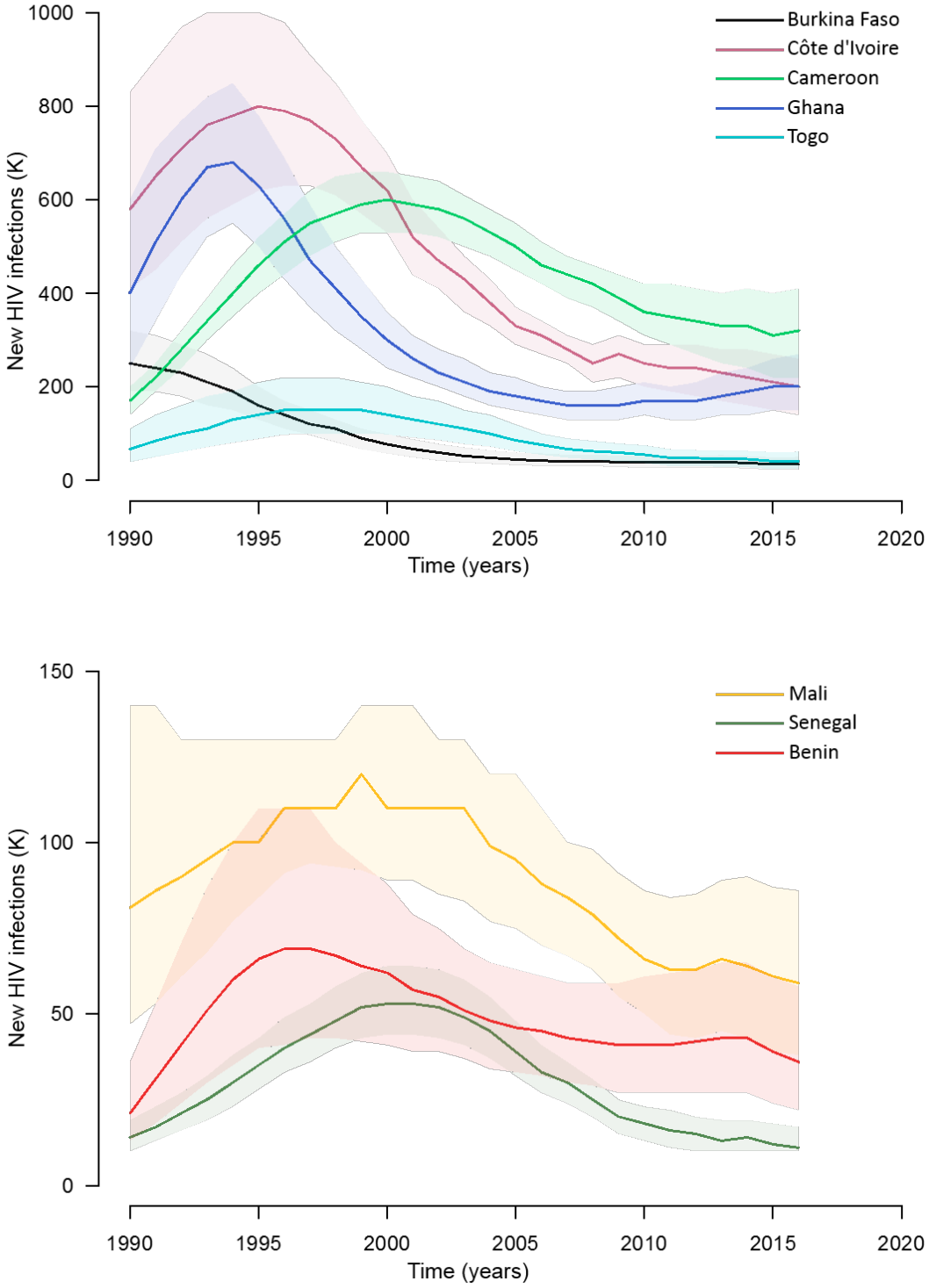


Figure 6.6: Number of new HIV cases in Western African countries where CRF02$_{WA}$ disseminates. Data was obtained from UNAIDS website: http://aidsinfo.unaids.org/

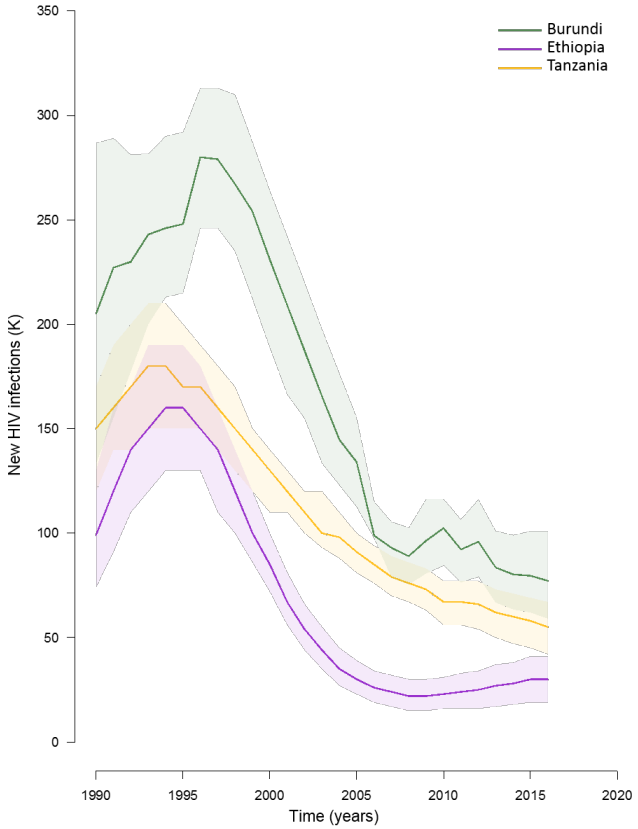


Figure 6.7: Number of new HIV cases in East African countries where C_{EA} clades disseminates. Data was obtained from UNAIDS website: http://aidsinfo.unaids.org/

In Brazil, the phylodynamic modeling points to a decline of the incidence rate between 1985 and 1995 for the $B_{PANDEMIC}$ and $C_{EA/BR}$ clades, which, assuming a time-lag of ∼8-10 years between HIV and AIDS reported cases, is concordant with the stabilization in the number of new AIDS cases reported by the Brazilian Ministry of Health since the early 2000s in the Southeast and Southern Brazilian regions, where most sequences from our datasets came from (Fig. 6.9).

Taking into account the time scale in which the switch between the phase of exponential growth and the stabilization of the epidemic growth took place across all analyzed HIV-1 clades, it is evident that the introduction of ART was not the driven factor in this crucial change of their epidemic dynamics. While in Latin American countries the free and universal access to

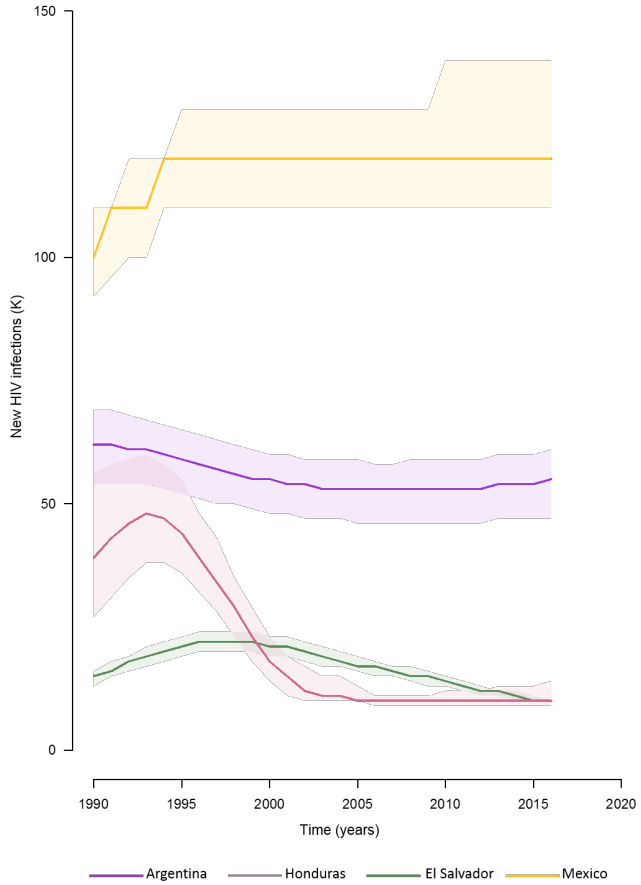


Figure 6.8: Number of new HIV cases in Latin American countries where $B_{PANDEMIC}$ clades disseminates. Data was obtained from UNAIDS website: http://aidsinfo.unaids.org/

ART started approximately between the mid and late 1990s [98], [108]–[111], in East and West African countries as well as in Central Africa when it is available, did not begin until the early 2000s [108], [111]–[113]. Consequently, factors such as intrinsic networks dynamics and/or changes in population risk behaviors, must be responsible for the reported and inferred declines in the HIV incidence rate of the analyzed Latin American HIV clades between late 1980s and mid 1990s and of the African CRF02_AG and C_{EA} analyzed sub-epidemics between early 1980s and early 2000s.

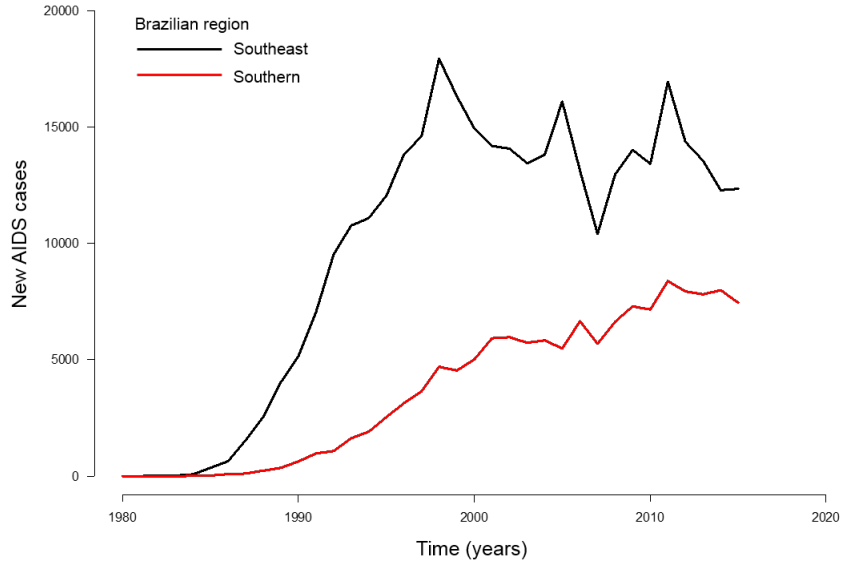


Figure 6.9: Number of new AIDS cases reported by the Brazilian Ministry of Health since early 2000s in the Southeast and Southern Brazilian regions. http://www.portalsinan.saude.gov.br/dados-epidemiologicos-sinan

6.7 Recent changes in HIV-1 epidemics

The fit of the population dynamics of all the HIV-1 clades analyzed under the frame of a logistic model as the coalescent-based phylodynamic analysis suggests, implies that since their initial stabilization, their incidence rates have remained stable up to the date of the most recent genetic sequence included in their datasets. Nevertheless, the epidemiological data support some recent changes in the growth rates of HIV-1 epidemics from Western Africa, Eastern Africa and Latin America that the coalescent-based phylodynamic model was unable to capture. Epidemiological data support a decline in the number of new HIV infections in some Western (Burkina Faso, Ghana, Mali, Togo and Senegal) and Eastern (Burundi, Rwanda, Ethiopia, and Tanzania) African countries since the late 1990s and early 2000s, as well in Honduras and El Salvador since the mid 1990s (Figs. 6.6 - 6.8). By contrast, the Brazilian Ministry of Health reported an upward trend of new HIV diagnoses since 2007 in Southern states where the $C_{EA/BR}$ clade has been widely disseminated (Fig. 6.9) [114]. Recent declines in HIV incidence

support the relevance of prevention efforts that acted as the driven-force of people's behavioral changes in several countries. At the same time, the recent upward trend of new $C_{EA/BR}$ infections herein inferred in Southern Brazil coincides with an association between ART availability and relapses into high-risk sexual behavior, reported among vulnerable population groups from developed countries [115]–[118] and emphasizes the importance for the urgent implementation of prevention strategies in those groups.

The sensitivity shown by the birth-death phylodynamic model to capture changes advanced by the epidemiological data in the most recent stages of the C_{EA} sub-epidemics, denotes that the lack of sensitivity of the coalescent-based approach to capture such changes is likely due to intrinsic assumptions of the model and not to sampling bias and/or the relative prevalence of the HIV-1 lineage under study. The greater precision of the coalescent-based parameter estimates for the oldest epidemic period and the apparent higher sensitivity of the birth-death model for recover changes in population dynamics at most recent times, leads us to recognize the great importance of the joint use of both approaches for the comprehensive reconstruction of past population dynamics of HIV epidemics and highlights the need for undertake a birth-death-based phylodynamic analysis on the $B_{PANDEMIC}$ and CRF02_AG clades characterized in Chapters 3 and 4 of this thesis for which this approach was not applied. It will be also of paramount importance to test whether other HIV-1 major lineages circulating in Brazil also follow the same recent upward trend in infections herein inferred for the $C_{EA/BR}$ clade.

6.8 Conclusions

- One-third of HIV-1 subtype B infections in Latin America resulted from the expansion of some Bpandemic clades likely to be introduced at a very early stage of the HIV epidemic in the Americas.
- When the AIDS epidemic was recognized in the early 1980s, subtype B had already arrived and established itself in several countries in Latin America.
- The average R identified for the characterized $B_{PANDEMIC}$ clades varies between different countries and different lineages in the same country, supporting the leading role of transmission networks instead of country-specific prevention programs or viral determinants in the rate of propagation of this HIV-1 clades in America.
- Current HIV-1 CRF02_AG epidemics in West-Central and West African countries resulted from the spread of some founding strains outside Central Africa between the 1960s and the mid-1980s.
- The founder strain CRF02_AG introduced in West Africa showed faster growth and a broader geographical spread epidemic than those introduced in West Central Africa.
- The spread of the $CRF02_{WA}$ clade outside Africa led to the emergence of local transmission networks in Asia and Europe between the late 1990s and the early 2000s.
- The outcome of the epidemic of different lineages of the CRF02_AG variant was probably shaped by several factors, including: origin time, spatial accessibility, transmission dynamics and viral determinants.

- Major HIV-1 C_{EA} lineages circulating in Eastern Africa and Southern Brazil seem to have had an exponential spread with very similar growth rates until roughly the mid 1990s, when a stabilization of the HIV incidence rate took place in some of them, and a decline was noted in others, which is likely associated with intrinsic networks dynamics and/or changes in population risk behaviors.
- The overall agreement of the R estimated from genetic sequences with those previously obtained from classical epidemiological data strength the utility of coalescent and the birth-death phylodynamic approaches to infer relevant epidemiological information of HIV epidemics.

6.9 References

1. R. Ross, “Report on the prevention of malaria in Mauritius,” 1908.

2. R. Ross, “Some Quantitative Studies in Epidemiology,” Nature, vol. 87, no. 2188, pp. 466–467, Oct. 1911.

3. W. O. Kermack and A. G. McKendrick, “A Contribution to the Mathematical Theory of Epidemics,” Proc. R. Soc. A Math. Phys. Eng. Sci., vol. 115, no. 772, pp. 700–721, Aug. 1927.

4. W. O. Kermack and A. G. McKendrick, “Contributions to the Mathematical Theory of Epidemics. II. The Problem of Endemicity,” Proc. R. Soc. A Math. Phys. Eng. Sci., vol. 138, no. 834, pp. 55–83, Oct. 1932.

5. W. O. Kermack and A. G. McKendrick, “Contributions to the Mathematical Theory of Epidemics. III. Further Studies of the Problem of Endemicity,” Proc. R. Soc. A Math. Phys. Eng. Sci., vol. 141, no. 843, pp. 94–122, Jul. 1933.

6. O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, “On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations,” J. Math. Biol., vol. 28, no. 4, pp. 365–382, Jun. 1990.

7. R. M. Anderson and R. M. May, “Population biology of infectious diseases: Part I,” Nature, vol. 280, no. 5721, pp. 361–367, Aug. 1979.

8. R. M. May and R. M. Anderson, “Population biology of infectious diseases: Part II,” Nature, vol. 280, no. 5722, pp. 455–461, Aug. 1979.

9. E. M. Volz, K. Koelle, and T. Bedford, “Viral phylodynamics.,” PLoS Comput. Biol., vol. 9, no. 3, p. e1002947, Jan. 2013.

10. L. Du Plessis and T. Stadler, “Getting to the root of epidemic spread with phylodynamic analysis of genomic data,” Trends Microbiol., vol. 23, no. 7, pp. 383–386, 2015.

11. J. F. . Kingman, “Exchangeability and the Evolution of Large Populations,” Exch. Probab. Stat., pp. 97–112, 1982.

12. J. F. C. Kingman, “The coalescent,” Stoch. Process. their Appl., vol. 13, no. 3, pp. 235–248, 1982.

13. J. F. C. Kingman, “On the Genealogy of Large Populations,” J. Appl. Probab., vol. 19, no. 1982, p. 27, 1982.

14. D. G. Kendall, “On Some Modes of Population Growth Leading to R. A. Fisher’s Logarithmic Series Distribution.,” Biometrika, vol. 35, no. 1/2, pp. 6–15, 1948.

15. D. G. Kendall, “On the Generalized ‘Birth-and-Death’ Process.,” Ann. Math. Stat., vol. 19, no. 1, pp. 1–15, 1948.

16. O. G. Pybus, M. A. Charleston, S. Gupta, A. Rambaut, E. C. Holmes, and P. H. Harvey, “The epidemic behavior of the hepatitis C virus.,” Science (80-.)., vol. 292, pp. 2323–5, 2001.

17. Y. Mendoza et al., “Human immunodeficiency virus type 1 (HIV-1) subtype B epidemic in Panama is mainly driven by dissemination of country-specific clades.,” PLoS One, vol. 9, no. 4, p. e95360, 2014.

18. E. Delatorre and G. Bello, “Phylodynamics of the HIV-1 Epidemic in Cuba.,” PLoS One, vol. 8, p. e72448, 2013.

19. M. Cabello, D. M. Junqueira, and G. Bello, “Dissemination of nonpandemic Caribbean HIV-1 subtype B clades in Latin America.,” AIDS, vol. 29, no. 4, pp. 483–92, 2015.

20. K. E. Robbins et al., “U.S. Human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains.,” J. Virol., vol. 77, no. 11, pp. 6359–66, Jun. 2003.

21. M. Worobey et al., “1970s and ‘Patient 0’ HIV-1 genomes illuminate early HIV/AIDS history in North America.,” Nature, vol. 539, no. 7627, pp. 98–101, 2016.

22. G. Zehender et al., "Population Dynamics of HIV-1 Subtype B in a Cohort of Men-Having-Sex-With-Men in Rome, Italy," JAIDS J. Acquir. Immune Defic. Syndr., vol. 55, no. 2, pp. 156–160, Oct. 2010.

23. J. H.-K. Chen, K.-H. Wong, K. C.-W. Chan, S. W.-C. To, Z. Chen, and W.-C. Yam, "Phylodynamics of HIV-1 Subtype B among the Men-Having-Sex-with-Men (MSM) Population in Hong Kong," PLoS One, vol. 6, no. 9, p. e25286, Sep. 2011.

24. A. J. Tatem, J. Hemelaar, R. R. Gray, and M. Salemi, "Spatial accessibility and the spread of HIV-1 subtypes and recombinants.," AIDS, vol. 26, no. 18, pp. 2351–60, 2012.

25. D. Gnisci, M. Trémolières, and (SWAC/OECD), "Charpter 4. Migration," in West African Studies. Regional Atlas on West Africa. Population and settlement., OECD/SWAC, 2009, p. 284.

26. P. R. Walker, O. G. Pybus, A. Rambaut, and E. C. Holmes, "Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic growth," Infect. Genet. Evol., vol. 5, no. 3 SPEC. ISS., pp. 199–208, 2005.

27. S. Sovran, "Understanding culture and HIV/AIDS in sub-Saharan Africa.," SAHARA J J. Soc. Asp. HIV/AIDS Res. Alliance, vol. 10, no. 1, pp. 32–41, Mar. 2013.

28. A.-A. Velayati et al., "Religious and Cultural Traits in HIV/AIDS Epidemics in Sub-Saharan Africa," Arch Iran. Med Arch. Iran. Med. Arch. Iran. Med., vol. 10, no. 4, pp. 486–497, 2007.

29. I. A. Ivanov et al., "Detailed molecular epidemiologic characterization of HIV-1 infection in Bulgaria reveals broad diversity and evolving phylodynamics.," PLoS One, vol. 8, no. 3, p. e59666, 2013.

30. L. M. Eyzaguirre et al., "Genetic Characterization of HIV-1 Strains Circulating in Kazakhstan," J. Acquir. Immune Defic. Syndr., vol. 46, no. 1, pp. 19–23, 2007.

31. I. Lapovok et al., “Short communication: molecular epidemiology of HIV type 1 infection in Kazakhstan: CRF02_AG prevalence is increasing in the southeastern provinces.,” AIDS Res. Hum. Retroviruses, vol. 30, no. 8, pp. 769–774, 2014.

32. J. K. Carr et al., “Outbreak of a West African Recombinant of HIV-1 in Tashkent, Uzbekistan,” J Acquir Immune Defic Syndr, vol. 39, pp. 570–575, 2005.

33. V. Laga et al., “The Genetic Variability of HIV-1 in Kyrgyzstan: The Spread of CRF02_ AG and Subtype A1 Recombinants.,” J. HIV AIDS, vol. 1.2, 2015.

34. E. Kazennova et al., “HIV-1 Genetic Variants in the Russian Far East,” AIDS Res. Hum. Retroviruses, vol. 30, no. 8, 2014.

35. UNODC, “Illicit Drug Trends in Central Asia,” 2008.

36. E. T. Richardson et al., “Gender inequality and HIV transmission: a global analysis,” J Int AIDS Soc, vol. 17, p. 19035, 2014.

37. G. Ramjee and B. Daniels, “Women and HIV in Sub-Saharan Africa.,” AIDS Res. Ther., vol. 10, no. 1, p. 30, Dec. 2013.

38. M. Nyindo, “Complementary Factors Contributing To The Rapid Spread Of HIV-I In Sub-Saharan Africa: A Review,” East Afr. Med. J., vol. 82, no. 1, Jul. 2005.

39. A. Buvé, K. Bishikwabo-Nsarhaza, and G. Mutangadura, “The spread and effect of HIV-1 infection in sub-Saharan Africa,” Lancet, vol. 359, no. 9322, pp. 2011–2017, 2002.

40. UNAIDS, “HIV/AIDS and conflict,” 2003.

41. P. Spiegel, “HIV/AIDS among conflict affected and displaced populations: dispelling myths and taking action,” Disasters, vol. 28, no. 3, pp. 322–339, 2004.

42. D. Omare and A. Kanekar, “Determinants of HIV/AIDS in armed conflict populations.,” J. Public Health Africa, vol. 2, no. 1, p. e9, Mar. 2011.

43. J. Iliffe, The African AIDS epidemic: a history. 2006.

44. B. Terminski, "Development-Induced Displacement and Resettlement: Theoretical Frameworks and Current Challenges," Development, no. May, 2013.

45. T. Gräf and A. R. Pinto, "The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent," Virology. 2013.

46. T. Gräf, H. M. Fritsch, R. M. de Medeiros, D. M. Junqueira, S. E. de M. Almeida, and A. R. Pinto, "Comprehensive characterization of the HIV-1 molecular epidemiology and demographic history in the Brazilian region most heavily affected by AIDS," J. Virol., vol. 90, no. 18, p. JVI.00363-16, 2016.

47. T. Gräf 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., vol. 51, no. 3, pp. 186–191, Jul. 2011.

48. E. Delatorre, D. Mir, and G. Bello, "Spatiotemporal dynamics of the HIV-1 subtype G epidemic in West and Central Africa.," PLoS One, vol. 9, no. 2, p. e98908, Jan. 2014.

49. E. Delatorre and G. Bello, "Spatiotemporal dynamics of the HIV-1 CRF06_cpx epidemic in western Africa," PLoS One, vol. 9, no. 6, 2014.

50. B. Auvert, D. Taljaard, E. Lagarde, J. Sobngwi-Tambekou, R. Sitta, and A. Puren, "Randomized, Controlled Intervention Trial of Male Circumcision for Reduction of HIV Infection Risk: The ANRS 1265 Trial," PLoS Med., vol. 2, no. 11, p. e298, Oct. 2005.

51. J. Bongaarts, P. Reining, P. Way, and F. Conant, "The relationship between male circumcision and HIV infection in African populations.," AIDS, vol. 3, no. 6, pp. 373–7, Jun. 1989.

52. H. A. Weiss, M. A. Quigley, and R. J. Hayes, "Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis.," AIDS, vol. 14, no. 15, pp. 2361–70, Oct. 2000.

53. D. T. Halperin and R. C. Bailey, "Male circumcision and HIV infection: 10 years and counting," Lancet, vol. 354, no. 9192, pp. 1813–1815, Nov. 1999.

54. J. C. Caldwell and P. Caldwell, "Toward an Epidemiological Model of AIDS in Sub-Saharan Africa," in Social Science History 4, vol. 4, no. winter, 1996.

55. C. T. Temah, What Drives HIV/AIDS Epidemic in Sub-Saharan Africa?, vol. 17, no. 5. 2009.

56. J. Zou, Y. Yamanaka, M. John, M. Watt, J. Ostermann, and N. Thielman, "Religion and HIV in Tanzania: influence of religious beliefs on HIV stigma, disclosure, and treatment attitudes.," BMC Public Health, vol. 9, p. 75, Mar. 2009.

57. P. B. Gray, "HIV and Islam: is HIV prevalence lower among Muslims?," Soc. Sci. Med., vol. 58, no. 9, pp. 1751–1756, May 2004.

58. S. Moses, J. E. Bradley, N. J. Nagelkerke, A. R. Ronald, J. O. Ndinya-Achola, and F. A. Plummer, "Geographical patterns of male circumcision practices in Africa: association with HIV seroprevalence.," Int. J. Epidemiol., vol. 19, no. 3, pp. 693–7, Sep. 1990.

59. N. D. Duncan, "Male circumcision and the Caribbean HIV epidemic.," West Indian Med. J., vol. 59, no. 4, pp. 348–350, 2010.

60. "The Islamic world."
Available: http://www.nature.com/news/specials/islamandscience/map/islam-map.html Accessed: 12-Jan-2018.

61. E. Lagarde et al., "Religion and protective behaviours towards AIDS in rural Senegal.," AIDS, vol. 14, no. 13, pp. 2027–33, Sep. 2000.

62. B. K. Takyi, "Religion and women's health in Ghana: insights into HIV/AIDs preventive and protective behavior.," Soc. Sci. Med., vol. 56, no. 6, pp. 1221–34, Mar. 2003.

63. S. Agha, P. Hutchinson, and T. Kusanthan, "The effects of religious affiliation on sexual initiation and condom use in Zambia," J. Adolesc. Heal., vol. 38, no. 5, pp. 550–555, May 2006.

64. J. Trinitapoli, "Religious teachings and influences on the ABCs of HIV prevention in Malawi," Soc. Sci. Med., vol. 69, no. 2, pp. 199–209, Jul. 2009.

65. A. S. Muula, J. C. Thomas, A. E. Pettifor, R. P. Strauss, C. M. Suchindran, and S. R. Meshnick, "Religion is not associated with HIV infection among women in Malawi," Int. J. Disabil. Hum. Dev., vol. 11, no. 2, pp. 121–131, Jan. 2012.

66. A. S. Muula, "Marriage, not Religion, is Associated with HIV Infection Among Women in Rural Malawi," AIDS Behav., vol. 14, no. 1, pp. 125–131, Feb. 2010.

67. S. C. Ball et al., "Comparing the ex vivo fitness of CCR5-tropic human immunodeficiency virus type 1 isolates of subtypes B and C.," J. Virol., vol. 77, no. 2, pp. 1021–38, Jan. 2003.

68. K. K. Ariën, G. Vanham, and E. J. Arts, "Is HIV-1 evolving to a less virulent form in humans?," Nat. Rev. Microbiol., vol. 5, no. 2, pp. 141–151, 2007.

69. K. K. Ariën, A. Abraha, M. E. Quiñones-Mateu, L. Kestens, G. Vanham, and E. J. Arts, "The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates.," J. Virol., vol. 79, no. 14, pp. 8979–90, Jul. 2005.

70. C. M. Venner et al., "Infecting HIV-1 Subtype Predicts Disease Progression in Women of Sub-Saharan Africa," EBioMedicine, vol. 13, pp. 305–314, 2016.

71. United Nations Joint Program on HIV/AIDS (UNAIDS), "AIDSinfo." Available: http://aidsinfo.unaids.org/. Accessed: 01-Dec-2017.

72. N. Ndembi et al., "Genetic diversity of HIV type 1 in rural eastern Cameroon.," J. Acquir. Immune Defic. Syndr., vol. 37, no. 5, pp. 1641–1650, 2004.

73. R. Powell, D. Barengolts, L. Mayr, and P. Nyambi, "The evolution of HIV-1 diversity in rural Cameroon and its implications in vaccine design and trials," Viruses, vol. 2, no. 2, pp. 639–654, 2010.

74. N. Ndembi et al., "Molecular characterization of human immunodeficiency virus type 1 (HIV-1) and HIV-2 in Yaoundé, Cameroon: Evidence of major drug resistance

mutations in newly diagnosed patients infected with subtypes other than subtype B," J. Clin. Microbiol., vol. 46, no. 1, pp. 177–184, 2008.

75. E. a. Soares et al., "Molecular diversity and polymerase gene genotypes of HIV-1 among treatment-naïve Cameroonian subjects with advanced disease," J. Clin. Virol., vol. 48, no. 3, pp. 173–179, 2010.

76. C. B. Ndongmo et al., "HIV genetic diversity in Cameroon: possible public health importance.," AIDS Res. Hum. Retroviruses, vol. 22, no. 8, pp. 812–816, 2006.

77. J. N. Torimiro et al., "Human immunodeficiency virus type 1 intersubtype recombinants predominate in the AIDS epidemic in Cameroon," New Microbiol., vol. 32, no. 4, pp. 325–332, 2009.

78. F. a J. Konings et al., "Genetic analysis of HIV-1 strains in rural eastern Cameroon indicates the evolution of second-generation recombinants to circulating recombinant forms.," J. Acquir. Immune Defic. Syndr., vol. 42, no. 3, pp. 331–341, 2006.

79. N. Koch, J.-B. Ndihokubwayo, N. Yahi, C. Tourres, J. Fantini, and C. Tamalet, "Genetic Analysis of HIV Type 1 Strains in Bujumbura (Burundi): Predominance of Subtype C Variant," AIDS Res. Hum. Retroviruses, vol. 17, no. 3, pp. 269–273, 2001.

80. N. Vidal et al., "HIV Type 1 Diversity and Antiretroviral Drug Resistance Mutations in Burundi," AIDS Res. Hum. Retroviruses, vol. 23, no. 1, pp. 175–180, 2007.

81. A. Ababa et al., "HIV Type 1 Subtype C in Addis Ababa, Ethiopia.," AIDS Res. Hum. Retroviruses, vol. 13, no. 12, 1997.

82. A. Abebe et al., "Identification of a Genetic Subcluster of HIV Type 1 Subtype C (C9) Widespread in Ethiopia," AIDS Res. Hum. Retroviruses, vol. 16, no. 17, pp. 1909–1914, 2000.

83. A. Kassu, M. Fujino, M. Matsuda, M. Nishizawa, F. Ota, and W. Sugiura, "Sequence Note Molecular Epidemiology of HIV Type 1 in Treatment-Naive Patients in North Ethiopia," AIDS Res. Hum. Retroviruses, vol. 23, no. 4, pp. 564–568, 2007.

84. B. Renjifo et al., "Epidemic Expansion of HIV Type 1 Subtype C and Recombinant Genotypes in Tanzania," AIDS Res. Hum. Retroviruses, vol. 14, no. 7, 1998.

85. I. E. Kiwelu et al., "HIV Type 1 Subtypes among Bar and Hotel Workers in Moshi, Tanzania," AIDS Res. Hum. Retroviruses, vol. 19, no. 1, pp. 57–64, 2003.

86. K.-H. Herbinger et al., "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, vol. 22, no. 7, pp. 599–606, 2006.

87. B. M. Nyombi, K. I. Kristiansen, G. Bjune, F. Müller, and C. Holm-Hansen, "Diversity of Human Immunodeficiency Virus Type 1 Subtypes in Kagera and Kilimanjaro Regions, Tanzania," AIDS Res. Hum. Retroviruses, vol. 24, no. 6, 2008.

88. E. O. Delatorre and G. Bello, "Phylodynamics of HIV-1 subtype C epidemic in East Africa," PLoS One, vol. 7, no. 7, pp. 1–10, 2012.

89. L. G. Rivera-Morales et al., "The molecular epidemiology of HIV type 1 of men in Mexico.," AIDS Res. Hum. Retroviruses, vol. 17, no. 1, pp. 87–92, Jan. 2001.

90. S. E. Guerra-Palomares, P. G. Hernandez-Sanchez, M. A. Esparza-Perez, J. R. Arguello, D. E. Noyola, and C. A. Garcia-Sepulveda, "Molecular Characterization of Mexican HIV-1 Vif Sequences," AIDS Res. Hum. Retroviruses, vol. 32, no. 3, pp. 290–295, 2016.

91. L. Eyzaguirre et al., "First molecular surveillance report of HIV type 1 in injecting drug users and female sex workers along the U.S.-Mexico border.," AIDS Res. Hum. Retroviruses, vol. 23, no. 2, pp. 331–4, Feb. 2007.

92. W. Murillo, I. Lorenzana de Rivera, J. Albert, M. E. Guardado, A. I. Nieto, and G. Paz-Bailey, "Prevalence of transmitted HIV-1 drug resistance among female sex workers and men who have sex with men in El Salvador, Central America," J. Med. Virol., vol. 84, no. 10, pp. 1514–1521, Oct. 2012.

93. Á. Holguín et al., "Transmitted drug-resistance in human immunodeficiency virus-infected adult population in El Salvador, Central America.," Clin. Microbiol. Infect., vol. 19, no. 12, pp. E523-32, Dec. 2013.

94. Á. Holguín et al., "Drug Resistance Prevalence in Human Immunodeficiency Virus Type 1 Infected Pediatric Populations in Honduras and El Salvador During 1989–2009," Pediatr. Infect. Dis. J., vol. 30, no. 5, pp. e82–e87, May 2011.

95. C. Lara, M. Sällberg, B. Johansson, I. L. de Rivera, and A. Sönnerborg, "The Honduran human immunodeficiency virus type 1 (HIV-1) epidemic is dominated by HIV-1 subtype B as determined by V3 domain sero- and genotyping.," J. Clin. Microbiol., vol. 35, no. 3, pp. 783–4, Mar. 1997.

96. B. Renjifo et al., "HIV-1 subtype B in Honduras.," Virus Res., vol. 60, no. 2, pp. 191–197, 1999.

97. B. Lloyd et al., "Prevalence of Resistance Mutations in HIV-1-Infected Hondurans at the Beginning of the National Antiretroviral Therapy Program," AIDS Res. Hum. Retroviruses, vol. 24, no. 4, pp. 529–535, Apr. 2008.

98. W. Murillo et al., "Prevalence of drug resistance and importance of viral load measurements in Honduran HIV-infected patients failing antiretroviral treatment," HIV Med., vol. 11, no. 2, pp. 95–103, Feb. 2010.

99. W. Murillo et al., "Transmitted drug resistance and type of infection in newly diagnosed HIV-1 individuals in Honduras.," J. Clin. Virol., vol. 49, no. 4, pp. 239–44, Dec. 2010.

100. S. Avila-Ríos et al., "HIV Drug Resistance Surveillance in Honduras after a Decade of Widespread Antiretroviral Therapy."

101. C. A. Yabar, J. Salvatierra, and E. Quijano, "Polymorphism, Recombination, and Mutations in HIV Type 1 gag- Infecting Peruvian Male Sex Workers," AIDS Res. Hum. Retroviruses, vol. 24, no. 11, pp. 1405–1413, Nov. 2008.

102. C. A. Yabar 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, vol. 28, no. 12, pp. 1712–1722, Dec. 2012.

103. A. G. Carrión et al., "Molecular Characterization of the Human Immunodeficiency Virus Type 1 among Children in Lima, Peru," AIDS Res. Hum. Retroviruses, vol. 25, no. 8, pp. 833–835, Aug. 2009.

104. J. Soria et al., "Transmitted HIV resistance to first-line antiretroviral therapy in Lima, Peru.," AIDS Res. Hum. Retroviruses, vol. 28, no. 4, pp. 333–8, Apr. 2012.

105. K. L. Russell et al., "Emerging genetic diversity of HIV-1 in South America.," AIDS, vol. 14, no. 12, pp. 1785–91, Aug. 2000.

106. J. K. Carr et al., "Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America.," AIDS, vol. 15, no. 15, pp. F41-7, Oct. 2001.

107. G. Bello et al., "Phylodynamics of HIV-1 Circulating Recombinant Forms 12_BF and 38_BF in Argentina and Uruguay."

108. World Health Organization, "Countries offering free access to HIV treatment," 2005.

109. "Countries | AIDS Healthcare Foundation." Available: https://www.aidshealth.org/#/countries/. Accessed: 05-Jan-2018.

110. C. Del Rio and J. Sepúlveda, "AIDS in Mexico: Lessons learned and implications for developing countries," Aids, vol. 16, no. 11, pp. 1445–1457, 2002.

111. "'3 by 5' country information," WHO, 2011. Available: http://www.who.int/3by5/countryprofiles/en/. Accessed: 21-Nov-2017.

112. National Agency for the Control of AIDS, "Federal Republic of Nigeria: GLOBAL AIDS RESPONSE Country Progress Report," Fed. Repub. Niger. Glob. Aids Response Ctry. Prog. Rep. Niger. GARPR 2014 Niger., pp. 1–69, 2014.

113. S. Kouanda et al., "User fees and access to ARV treatment for persons living with HIV/AIDS: Implementation and challenges in Burkina Faso, a limited-resource country," AIDS Care - Psychol. Socio-Medical Asp. AIDS/HIV, vol. 22, no. 9, pp. 1146–1152, 2010.

114. Brazilian Ministry of Health, "AIDS Epidemiological Bulletin [in Portuguese]" 2016.

115. D. Bezemer et al., "A resurgent HIV-1 epidemic among men who have sex with men in the era of potent antiretroviral therapy," AIDS, no. February, pp. 1071–1077, 2008.

116. S. Le Vu et al., "Population-based HIV-1 incidence in France, 2003–08: a modelling analysis," Lancet Infect. Dis., vol. 10, no. 10, pp. 682–687, Oct. 2010.

117. The Kirby Institute/Sydney University of New South Wales, "HIV, viral hepatitis and sexually transmissible infections in Australia: Annual Surveillance Report," 2011.

118. P. J. White, H. Ward, and G. P. Garnett, "Is HIV out of control in the UK? An example of analysing patterns of HIV spreading using incidence-to-prevalence ratios," AIDS, vol. 20, no. 14, pp. 1898–1901, Sep. 2006.

Appendixes

Appendix A

Spatiotemporal dynamics of the HIV-1 subtype G epidemic in West and Central Africa

Delatorre, E., Mir, D., Bello, G. (2014).

PloS One, 9(2), e98908. http://doi.org/10.1371/journal.pone.0098908

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 (GWA-I and GWA-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 GWA-I (0.75 year21) and GWA-II (0.95 year21) clades was about two times higher than that estimated for central African lineages (0.47 year21). Notably, the overall evolutionary and demographic history of GWA-I and GWA-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.

Appendix B

Zika virus in the Americas: Early epidemiological and genetic findings.

Faria, N. R., Azevedo, R. do S. da S., ... Vasconcelos, P. F. C. (2016).

Science (New York, N.Y.), aaf5036.

http://doi.org/10.1126/science.aaf5036

Brazil has experienced an unprecedented epidemic of Zika virus (ZIKV), with ~30,000 cases reported to date. ZIKV was first detected in Brazil in May 2015 and cases of microcephaly potentially associated with ZIKV infection were identified in November 2015. Using next generation sequencing we generated seven Brazilian ZIKV genomes, sampled from four self-limited cases, one blood donor, one fatal adult case, and one newborn with microcephaly and congenital malformations. Phylogenetic and molecular clock analyses show a single introduction of ZIKV into the Americas, estimated to have occurred between May-Dec 2013, more than 12 months prior to the detection of ZIKV in Brazil. The estimated date of origin coincides with an increase in air passengers to Brazil from ZIKV endemic areas, and with reported outbreaks in Pacific Islands. ZIKV genomes from Brazil are phylogenetically interspersed with those from other South American and Caribbean countries. Mapping mutations onto existing structural models revealed the context of viral amino acid changes present in the outbreak lineage; however no shared amino acid changes were found among the three currently available virus genomes from microcephaly cases. Municipality-level incidence data indicate that reports of suspected microcephaly in Brazil best correlate with ZIKV incidence around week 17 of pregnancy, although this does not demonstrate causation. Our genetic description and analysis of ZIKV isolates in Brazil provide a baseline for future studies of the evolution and molecular epidemiology in the Americas of this emerging virus.

Appendix C

In-depth phylogenetic analysis of hepatitis C virus subtype 1a and occurrence of 80K and associated polymorphisms in the NS3 protease.

Santos, A. F., Bello, G., Vidal, L. L., Souza, S. L., Mir, D., Soares, M. A. (2016).

Scientific Reports, 6, 31780. http://doi.org/10.1038/srep31780

HCV genetic diversity is high and impacts disease progression, treatment and drug resistance. HCV subtype 1a is divided in two clades (I and II), and the 80 K natural polymorphism in the viral NS3 protease is prevalent in clade I. Paradoxically, countries dominated by this clade have contrasting frequencies of 80 K. Over 2,000 HCV 1a NS3 sequences were retrieved from public databases representing Europe, Oceania and the Americas. Sequences were aligned with HCV reference sequences and subjected to phylogenetic analysis to investigate the relative presence of different subtype 1a clades and NS3 protease mutations. HCV-1a sequences split into clades I and II. Clade I was further structured into three subclades, IA to C. Sub-clade IA prevailed in the U.S., while subclade IC was major in Brazil. The NS3 80 K polymorphism was associated with subclade IA, but nearly absent in subclades IB and IC, a pattern similarly seen for the 91S/T compensatory mutation. Three HCV-1a-I sub-clades have been identified, with different frequencies in distinct regions. The 80 K and 91A/S mutations were associated with subclade IA, which provide an explanation for the disparities seen in simeprevir resistance profiles of countries dominated by HCV 1a-I, like the U.S. and Brazil.

Appendix D

New insights into the hepatitis E virus genotype 3 phylodynamics and evolutionary history

Mirazo, S., Mir, D., Bello, G., Ramos, N., Musto, H., Arbiza, J. (2016). Infection, Genetics and Evolution, 43, 267–273.

http://doi.org/10.1016/j.meegid.2016.06.003

Hepatitis E virus (HEV) is an emergent hepatotropic virus endemic mainly in Asia and other developing areas. However, in the last decade it has been increasingly reported in high-income countries. Human infecting HEV strains are currently classified into four genotypes (1–4). Genotype 3 (HEV-3) is the prevalent virus genotype and the mostly associated with autochthonous and sporadic cases of HEV in developed areas. The evolutionary history of HEV worldwide remains largely unknown. In this study we reconstructed the spatiotemporal and population dynamics of HEV-3 at global scale, but with particular emphasis in South America, where case reports have increased dramatically in the last years. To achieve this, we applied a Bayesian coalescent-based approach to a comprehensive data set comprising 97 GenBank HEV-3 sequences for which the location and sampling date was documented. Our phylogenetic analyses suggest that the worldwide genetic diversity of HEV-3 can be grouped into two main Clades (I and II) with a mrca dated in approximately 320 years ago (95% HPD: 420– 236 years) and that a unique independent introduction of HEV-3 seems to have occurred in Uruguay, where most of the human HEV cases in South America have been described. The phylodynamic inference indicates that the population size of this virus suffered substantial temporal variations after the second half of the 20th century. In this sense and conversely to what is postulated to date, we suggest that the worldwide effective population size of HEV-3 is not decreasing and that frequently sources of error in its estimates stem from assumptions that the analyzed sequences are derived from a single panmictic population. Novel insights on the global population dynamics of HEV are given. Additionally, this work constitutes an attempt to further describe in a Bayesian coalescent framework, the phylodynamics and evolutionary history of HEV-3 in the South American region.

Appendix E

Evolutionary history and spatiotemporal dynamics of DENV-1 genotype V in the Americas.

de Bruycker-Nogueira, F., Mir, D., Dos Santos, F. B., Bello, G. (2016).

Infection, Genetics and Evolution, 45, 454–460.

http://doi.org/10.1016/j.meegid.2016.09.025

The genotype V has been the most prevalent dengue virus type 1 (DENV-1) clade circulating in the Americas over the last 40 years. In this study, we investigate the spatiotemporal pattern of emergence and dissemination of DENV-1 lineages in the continent. We applied phylogenetic and phylogeographic approaches to a comprehensive data set of 836 DENV-1 E gene sequences of the genotype V isolated from 46 different countries around the world over a period of 50years (1962 to 2014). Our study reveals that genetic diversity of DENV-1 genotype V in the Americas resulted from two independent introductions of this genotype from India. The first genotype V strain was most probably introduced into the Lesser Antilles at around the early 1970s and this Caribbean region becomes the source population of several DENV-1 lineages that spread in the Americas during the 1970s and 1980s. Most of those lineages appear to become extinct during the 1990s, except one that persisted in Venezuela and later spread to other American countries, dominating the DENV-1 epidemics in the region from the early 2000s onwards. The second genotype V strain of Indian origin was also most probably introduced into the Lesser Antilles at around the early 1980s. This lineage remained almost undetected for nearly 15years, until it was introduced in Northern Brazil around the middle 1990s and later spread to other country regions. These results demonstrate that different geographic regions have played a role in maintaining and spreading the DENV-1 genotype V in the Americas over time. DENV-1 genotype V lineages have originated, spread and died out in the Americas with very different dynamics and the phenomenon of lineage replacement across successive DENV-1 epidemic outbreaks was a common characteristic in most American countries.

Appendix F

Detection and molecular characterization of emergent GII.P17/GII.17 Norovirus in Brazil, 2015.

Andrade, J. S. R., Fumian, T. M., Leite, J. P. G., Assis, M. R. d., Bello, G., Mir, D., Miagostovich, M. P. (2017).

Infection, Genetics and Evolution, 51, 28–32.

http://doi.org/10.1016/j.meegid.2017.03.011

A newly GII.17 Kawazaki_2014 variant strain was detected recently in Brazil. Phylogenetic analysis reveals at least four independent introduction events of this lineage into this country that took place throughout 2014, coinciding with FIFA World Cup in Brazil, 2014, and Hong Kong has been identified as themost likely source of introduction. This variant emerged in Asia causing outbreaks and replacing prevalent GII.4. Emergence of GII.P17/ GII.17 variant emphasizes the need for active laboratory surveillance for NoV including molecular epidemiology and studies on virus evolution.

Appendix G

Tracing the origin of the NS1 A188V substitution responsible for recent enhancement of Zika virus Asian genotype infectivity.

Delatorre, E., Mir, D., Bello, G. (2017).

Memórias Do Instituto Oswaldo Cruz, 4(0), 1–3.

http://doi.org/10.1590/0074-02760170299

A recent study showed that infectivity of Zika virus (ZIKV) Asian genotype was enhanced by an alanine-to- valine amino acid substitution at residue 188 of the NS1 protein, but the precise time and location of origin of this mutation were not formally estimated. Here, we applied a Bayesian coalescent-based framework to estimate the age and location of the ancestral viral strain carrying the A188V substitution. Our results support that the ancestral ZIKV strain carrying the A188V substitution arose in Southeastern Asia at the early 2000s and circulated in that region for some time (5-10 years) before being disseminated to Southern Pacific islands and the Americas.

Appendix H

Phylodynamics of Yellow Fever Virus in the Americas: new insights into the origin of the 2017 Brazilian outbreak.

Mir, D., Delatorre, E., Bonaldo, M., Lourenço-de-oliveira, R., Vicente, A. C., Bello, G. (2017).

Scientific Reports, 7, 1–9.

http://doi.org/10.1038/s41598-017-07873-7

Yellow fever virus (YFV) strains circulating in the Americas belong to two distinct genotypes (I and II) that have diversified into several concurrent enzootic lineages. Since 1999, YFV genotype I has spread outside endemic regions and its recent (2017) reemergence in non-endemic Southeastern Brazilian states fuels one of the largest epizootic of jungle Yellow Fever registered in the country. To better understand this phenomenon, we reconstructed the phylodynamics of YFV American genotypes using sequences from nine countries sampled along 60 years, including strains from Brazilian 2017 outbreak. Our analyses reveals that YFV genotypes I and II follow roughly similar evolutionary and demographic dynamics until the early 1990s, when a dramatic change in the diversification process of the genotype I occurred associated with the emergence and dissemination of a new lineage (here called modern). Trinidad and Tobago was the most likely source of the YFV modern-lineage that spread to Brazil and Venezuela around the late 1980s, where it replaced all lineages previously circulating. The modern- lineage caused all major YFV outbreaks detected in non-endemic South American regions since 2000, including the 2017 Brazilian outbreak, and its dissemination was coupled to the accumulation of several amino acid substitutions particularly within non-structural viral proteins.

Appendix I

Genomic and structural features of the yellow fever virus from the 2016–2017 Brazilian outbreak.

Gómez, M.M., de Abreu, F.V.S., Dos Santos, A.A.C., de Mello, I.S., Santos, M.P., Ribeiro, I.P., Ferreira-de-Brito, A., de Miranda, R.M., de Castro, M.G., Ribeiro, M.S., da Costa Laterrière, R., Aguiar, S.F., Meira, G.L.S., Antunes, D., Torres, P.H.M., Mir, D., Vicente, A.C.P., Guimarães, A.C.R., Caffarena, E.R., Bello, G., Lourenço-de-Oliveira, R., Bonaldo, M.C.

J. Gen. Virol., vol. 99, no. 4, 2018.

http://jgv.microbiologyresearch.org/content/journal/jgv/10.1099/jgv.0.001033#tab2

Southeastern Brazil has been suffering a rapid expansion of a severe sylvatic yellow fever virus (YFV) outbreak since late 2016, which has reached one of the most populated zones in Brazil and South America, heretofore a yellow fever-free zone for more than 70 years. In the current study, we describe the complete genome of 12 YFV samples from mosquitoes, humans and non-human primates from the Brazilian 2017 epidemic. All of the YFV sequences belong to the modern lineage (sub- lineage 1E) of South American genotype I, having been circulating for several months prior to the December 2016 detection. Our data confirm that viral strains associated with the most severe YF epidemic in South America in the last 70 years display unique amino acid substitutions that are mainly located in highly conserved positions in non-structural proteins. Our data also corroborate that YFV has spread southward into Rio de Janeiro state following two main sylvatic dispersion routes that converged at the border of the great metropolitan area comprising nearly 12 million unvaccinated inhabitants. Our original results can help public health authorities to guide the surveillance, prophylaxis and control measures required to face such a severe epidemiological problem. Finally, it will also inspire other workers to further investigate the epidemiological and biological significance of the amino acid polymorphisms detected in the Brazilian 2017 YFV strains.