

Phylogenetics applied to the human immunodeficiency virus type 1 (HIV-1): from the cross-species transmissions to the contact network inferences

Tiago Gräf^{1/+}, Edson Delatorre^{2/+}, Gonzalo Bello^{3/+}

¹Fundação Oswaldo Cruz-Fiocruz, Instituto Gonçalo Moniz, Salvador, BA, Brasil

²Universidade Federal do Espírito Santo, Centro de Ciências Exatas, Naturais e da Saúde, Departamento de Biologia, Alegre, ES, Brasil

³Fundação Oswaldo Cruz-Fiocruz, Instituto Oswaldo Cruz, Laboratório de AIDS e Imunologia Molecular, Rio de Janeiro, RJ, Brasil

Phylogenetic analyses were crucial to elucidate the origin and spread of the pandemic human immunodeficiency virus type 1 (HIV-1) group M virus, both during the pre-epidemic period of cryptic dissemination in human populations as well as during the epidemic phase of spread. The use of phylogenetics and phylodynamics approaches has provided important insights to track the founder events that resulted in the spread of HIV-1 strains across vast geographic areas, specific countries and within geographically restricted communities. In the recent years, the use of phylogenetic analysis combined with the huge availability of HIV sequences has become an increasingly important approach to reconstruct HIV transmission networks and understand transmission dynamics in concentrated and generalised epidemics. Significant efforts to obtain viral sequences from newly HIV-infected individuals could certainly contribute to detect rapidly expanding HIV-1 lineages, identify key populations at high-risk and understand what public health interventions should be prioritised in different scenarios.

Key words: HIV-1 - phylogenetics - phylogeography - transmission clusters

Phylogenetics and the pre-epidemic phase of HIV-1 spread

The human immunodeficiency virus type 1 (HIV-1) has been spreading in human populations over the last 100 years and is responsible for most of the global HIV/AIDS pandemic. Phylogenetic analyses of simian immunodeficiency viruses (SIV) sequences from different species of non-human primates and of HIV-1 sequences from Central African countries were decisive to elucidate the origin of this pandemic human virus. Those analyses revealed that the four HIV-1 phylogenetic clades termed groups (M, N, O and P) have resulted from different cross-species transmission events of SIV from chimpanzees (*Pan troglodytes*) and gorillas (*Gorilla gorilla*) to humans.^(1,2) Specific populations of the subspecies *Pan troglodytes troglodytes* and *Gorilla gorilla gorilla*, endemic from Cameroon, were pointed as the sources of the zoonotic transmissions to humans, originating the groups M and N, and O and P, respectively [Figure (A)].^(1,2) While the HIV-1 groups N, O and P remained mostly confined to Cameroon and neighboring countries, the group M spread out of Central Africa and currently infects around 40 million people worldwide.

Phylogenetics has also greatly improved our understanding of the early spread of the HIV-1 group M, particularly during the period of cryptic dissemination in human populations. Circulation of the HIV-1 group M in humans was first detected in the United States of America and Europe in the early 1980s, shortly after recognition of the AIDS epidemic.⁽³⁾ Genetic evidence, however, revealed that this HIV-1 group was already present in the Democratic Republic of Congo (DRC) by the late 1950s, tracing the most recent common ancestor of all HIV-1 group M strains back to a human host that probably lived in Kinshasa (capital of the DRC) in the beginning of the twentieth century.⁽⁴⁾ The most recent phylogeographic study supports that during the pre-epidemic phase the HIV-1 group M primary spread from Kinshasa, reaching the neighboring city of Brazzaville (capital of the Republic of Congo) and southern DRC locations (Lubumbashi and Mbuji-Mayi) by the late 1930s, and central (Bwamanda) and northern (Kisangani) DRC locations by the middle 1940s and the early 1950s, respectively.⁽⁵⁾ Coalescent analyses suggest that during this pre-pandemic phase the HIV-1 group M underwent a relatively slow growth, until about 1960.⁽⁵⁾

Phylogenetics and the epidemic phase of HIV-1 spread

Around 1960, HIV-1 group M transitioned to a second phase of faster exponential growth that coincides with the geographical expansion of group M out of the DRC.⁽⁵⁾ The introduction and subsequent genetic diversification of some group M strains into new geographic regions resulted in strongly supported phylogenetic clades within the group M diversity, referred to as “subtypes” (A to D, F to H, J, K and L) [Figure (B)].⁽⁶⁾ In addition, some inter-subtype recombinant strains named as “Circulating Recombinant Forms” (CRFs) were also spread

doi: 10.1590/0074-02760190461

Financial support: CNPq, FAPERJ.

GB is a recipient of CNPq fellowship for Productivity in Technological Development and Innovative Extension (Grant 302317/2017-1) and is funded by grant from FAPERJ (Grant E-26/202.896/2018).

+ Corresponding authors: tiago.graf@fiocruz.br; edson.delatorre@ufes.br; gbello@ioc.com.br

https://orcid.org/0000-0003-4921-7975; https://orcid.org/0000-0002-5746-0820; https://orcid.org/0000-0002-2724-2793

Received 05 December 2019

Accepted 12 February 2020



across different individuals and branched as well-supported clades within the group M phylogenetic tree.⁽⁶⁾ While some subtypes (A to F) and CRFs (CRF01_AE and CRF02_AG) are globally disseminated and should thus be defined as pandemic variants, others are mostly restricted to central Africa and/or to a single country and are regarded as endemic-subtypes and local CRFs.

The combined use of phylogenetic and phylogeographic approaches were crucial to track the founder events that resulted in the geographic spread of pandemic HIV-1 strains.^(5,6) Those analyses revealed that most HIV-1 pandemic lineages first spread from the Congo basin to neighboring regions in southern, eastern, and west African regions, before being disseminated outside Africa.⁽⁵⁾ The subtype B and the CRF01_AE, by contrast, did not go through a phase of wide dissemination within the African continent, but moved directly from Central Africa to the Americas or Southeast Asia, respectively, from where they later spread worldwide.^(5,6)

Some studies point that variance in the dispersion routes of HIV-1 subtypes and CRFs was probably shaped by spatial and geopolitical factors that affected human activities in the second half of the 20th century, such as migration, tourism and trade.^(7,8) Other studies indicate that differences in the worldwide prevalence of HIV-1 strains might be also shaped by subtype specific differences in virulence and transmissibility.⁽⁹⁾ Future studies comparing the evolutionary and population dynamics of different HIV-1 subtypes/CRFs spreading in the same populations would be necessary to fully understand the factors that shape the contrasting epidemic success of different HIV-1 clades.

Phylogenetics and the local spread of HIV-1

The global dissemination of the pandemic-subtypes/CRFs resulted in local epidemics that usually differ in epidemic size and geographic range. Since temporal changes in the spatial dispersion and population size of HIV-infected individuals leave an imprint on HIV genetic diversity and phylogenetic patterns, we can use model-based phylodynamic inference methods to track the phylogeographic and demographic history of the HIV-1 epidemic within a defined area [Figure (C)].⁽¹⁰⁾

Phylodynamics could provide important epidemiological insights about HIV-1 epidemics affecting a vast geographic area. This approach, for example, was used to resolve the spatiotemporal dynamics of the HIV-1 subtype A variant that dominates the epidemic in the former Soviet Union (A_{FSU}) and of the non-pandemic subtype B variants prevalent in the Caribbean region (B_{CAR}).^(11,12) Diez-Fuertes et al.⁽¹¹⁾ estimated that the A_{FSU} clade resulted from exportation of a subtype A lineage from the DRC to Ukraine in the 1980s, where it initially spread via heterosexual transmission for about a decade before its explosive dissemination among intravenous drug users (IDU). Cabello et al.⁽¹²⁾ revealed that the B_{CAR} epidemic resulted from early viral transmissions from Hispaniola to Trinidad and Tobago and to Jamaica between the late 1960s and the early 1970s and from Hispaniola and Trinidad and Tobago to other Lesser Antilles islands at later times.

Phylodynamics studies were also employed to resolve country-level dissemination dynamics of new and established HIV lineages. Illustrating this approach, two studies explored the origin of HIV-1 CRFs lineages recently detected in Brazil that mostly circulate in West (CRF02_AG) and Central (CRF45_cpx) Africa.^(13,14) The estimated onset dates of the CRF02_AG and CRF45_cpx local clades indicated that these CRFs were circulating in Brazil for about 20-30 years before their detection by the public health surveillance system. Those studies revealed that the CRF02_AG and CRF45_cpx Brazilian epidemics did not result from rapid expansion of recently introduced viruses, but from slow dissemination of ancient viral introductions combined with delayed detection by the public health system.

Phylodynamics analyses also helped to elucidate the origin and populations' dynamics of HIV-1 lineages spreading within small communities. One example was the characterisation of an HIV outbreak in children from a Libyan hospital that occurred in 1998, suspected to have originated from the malicious intervention of foreign medical staff.⁽¹⁵⁾ The authors found that the CRF02_AG outbreak affecting Libyan children arose from a single viral introduction from West Africa before March 1998 and that many of the HIV infections already occurred before the foreign medical staff arrived, excluding their participation in the initial transmissions. Another example was the characterisation of a spatially localised iatrogenic outbreak occurred in rural Cambodia in 2014-2015, suspected to be caused by an unlicensed health practitioner.⁽¹⁶⁾ The iatrogenic hypothesis was confirmed by the phylodynamics analysis that date the origin of the outbreak to September 2013 and estimated that the transmission reached a peak of 15 new HIV infections per day one year later, declining thereafter, coinciding with the date of arrest of the practitioner by the police.

Phylodynamics analyses combined with birth-death models could be a useful tool to elucidate the impact of preventive or therapeutic strategies on localised HIV epidemics. A recent study dated back the origin of a Belgian HIV-1 subtype F1 epidemic among men having sex with men (MSM) to the early 2000s and suggested that its extensive growth was controlled about 10 years later, most likely due to highly active antiretroviral therapy (ART) as prevention.⁽¹⁷⁾ Another study estimated that major shifts in HIV-1 transmission for subtypes B and G Portuguese clades occurred around the late 1990s and early 2000s, also coinciding with the introduction of ART and the scale-up of harm reduction for IDU.⁽¹⁸⁾ Analyses of the HIV-1 subtype C epidemic in heterosexual population from southern Brazil support that major changes in viral transmission dynamics (transient epidemic stabilisation and resume epidemic increase) coincides with people's behavioral changes driven by implementation of prevention efforts and perception of risk for HIV transmission.⁽¹⁹⁾ Phylodynamics analyses of viral sequences recovered from newly HIV-infected individuals will certainly contribute to detect rapidly expanding HIV-1 lineages and to assess the impact of public health interventions on localised and country-wide epidemics on real time.

Phylogenetics and the study of HIV-1 transmission networks

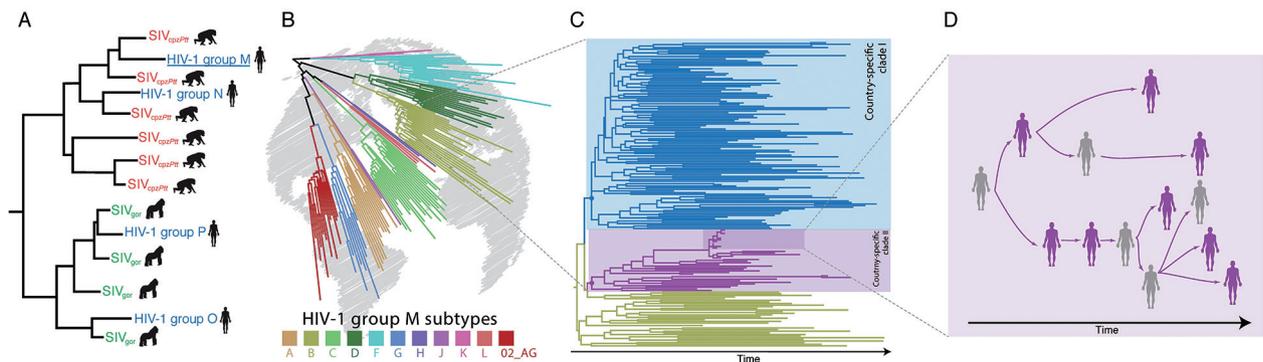
The absence of proofreading activity of the viral reverse transcriptase and fast replication rate make mutations to accumulate in HIV genomes in an epidemiological timescale.⁽⁶⁾ This not only means that we can trace HIV dissemination throughout wide territories (as continent or countries) along the decades, but also that the genetic diversity accumulates faster enough to reconstruct the viral transmission network, which describes the history of infections at the resolution of individual cases. The basic assumption is that closely related viruses in a phylogenetic tree is an indication that the hosts are connected by a common source, a direct or a short chain of transmissions [Figure (D)]. In the recent years, the use of phylogenetic analysis combined with the huge availability of HIV sequences has become an increasingly important research area to reconstruct HIV transmission networks and understand transmission dynamics in concentrated and generalised epidemics.

In an attempt to reconstruct the HIV transmission events in a population, the better the sampling coverage, the closer to reality is the inferred network. In this sense, most of the studies in this field have used databases compiled by national services for screening of drug-resistance mutations (genotyping), which generate HIV partial genome sequences for virtually every individual entering/failing therapy. Focusing on highly supported phylogenetic clusters mostly compose by local (from a single country) individuals, Hué et al.⁽²⁰⁾ reported the importance of multiple subtype B lineages circulating in separate transmission networks in the United Kingdom (UK) and also how the growth rate decrease of these networks was more likely to be correlated to behavior changes than to the introduction of ART. Still in UK, Hughes et al.⁽²¹⁾ estimated that clustering rates and the number of transmissions happening in the acute phase were much smaller among heterosexual than in the men who have sex with men (MSM) group. In Switzerland, Koyous et

al.⁽²²⁾ revealed the importance of IDU in spreading HIV to heterosexual individuals during the 1980's and the diminishing role of this relationship over time.

To identify current and still active transmission networks, many studies apply a genetic cut-off to define clusters.⁽²³⁾ When analysing factors correlated with cluster membership, typical features like high viral load and CD4⁺ T cell counts, not on ART and not aware about the HIV serostatus were found, underscoring the importance of acute phase of infection in the HIV transmission dynamics. Despite the high HIV transmissibility during acute phase, Volz et al.⁽²⁴⁾ revealed that time since infection is the main explanatory variable driving clustering in a phylogenetic tree. Poon et al.⁽²⁵⁾ expanded these findings by modeling heterogeneity of transmission and sampling rates among HIV infected subpopulations, reporting that phylogenetic clusters tend to be enriched with individuals sampled soon after the infection. In other words, transmission networks, as defined by a phylogenetic cluster with high statistical support and small within clade genetic distance, might better represent individuals highly engaged in accessing primary care, instead of vulnerable subpopulations burdened by high transmission rates. Thus, caution is needed when interpreting and analysing phylogenetic-based HIV-1 transmission networks.

When a dense body of clinical and demographic data from patients is available, the source attribution methods could be used to find potential transmission pairs and resolve the timing and direction of the infection. Ratmann et al.⁽²⁶⁾ used this approach to study the HIV transmission dynamics among MSM in a Dutch cohort and found that 71% of the transmissions have origin in undiagnosed man. This highlights that the prevention potential of immediate ART is limited without an approach that includes intensification of HIV testing and also points out the importance of pre-exposure prophylaxis (PrEP) as an effective intervention among this population. Maybe the most ambitious application of phylogenetics and molecular epidemiology on HIV is the



Phylogenetics applied to the human immunodeficiency virus type 1 (HIV-1) molecular epidemiology investigation revealed the history of the virus dissemination among humans. (A) Phylogenetic tree of HIV-1 and related simian immunodeficiency viruses (SIV) sequences reveals the multiple virus cross-species transmission events that originated HIV-1 groups. (B) The successful global spread of HIV-1 group M and fast diversification in multiple phylogenetic lineages. Figure depicts the current 10 subtypes and the most widely spread recombinant form CRF02_AG. (C) Phylogenetic tree representing the virus dispersion of a given HIV-1 subtype in space and time, which can reveal transmission routes between and within geographic locations represented by different branch colours. (D) When sampling coverage is high, phylogenies are informative to investigate transmissions in the social space, also known as HIV-1 transmission network reconstruction. The figure shows a hypothetical contact network, where sampled individuals (in purple) formed a closely related cluster in the tree.

real-time (or near-real-time) monitoring of HIV transmission networks growth. In British Columbia, Canada, an automated HIV monitoring system is helping to track the virus dissemination and was shown to be effective in reducing transmitted HIV drug resistance.⁽²⁷⁾ Implementation of such approach requires integration across clinical, laboratory, sequence analysis and public health teams (which will ultimately conduct interventions).

While phylogenetics has been increasingly deployed to study HIV transmission dynamics in developed countries with concentrated epidemics, where viral sequences are routinely sampled, few studies were able to apply these tools in resource-limited locations, where the epidemic is generalised and HIV has the greatest burden. The main limitation is the lack of a good sampling coverage of these locations, for example, African countries has more than 60% of the HIV infection cases in the world, but less than 30% of the public available sequences. Yet, cohort studies like the HIV Incidence Provincial Surveillance System (HIPSS) and the Rakai Community Cohort Study (RCCS) has enlighten important features of HIV dissemination in geographic settings with huge epidemics like South Africa and Uganda, respectively.^(28,29) De Oliveira et al.⁽²⁸⁾ has shown that age-disparate sexual partnering may be driving HIV transmission towards young women. Grabowsky et al.⁽²⁹⁾ reveals a complex dynamics of HIV transmissions among several communities, with constant introduction of new viral lineages. In the concept of treatment as prevention (TasP), both studies provided valuable data for better implementation strategies.

Phylogenetics analysis has been successful in identifying and analysing HIV-1 transmission networks, especially when sampling covers a great proportion of the infected population and when combined with detailed epidemiological and clinical data. Consequently, most studies describe transmission dynamics in developed countries, while significant sequencing efforts are still needed to phylogenetic methods be able to capture HIV transmission networks where the epidemic is worst. In an intermediate position, middle income countries might find in the phylogenetic approach a way to design better public health campaigns to halt HIV dissemination. Brazil, for instance, has one of the biggest public health systems in the world and access to ART is free and universal, with genotyping service deployed for those who fail therapy. This could be a good source of data to investigate HIV transmission dynamics in the country and complement the scarce literature on this topic.⁽³⁰⁾ The ultimate goal is to improve strategies of HIV prevention in the Brazilian population, identifying key populations and understanding what interventions should be prioritised.

AUTHORS' CONTRIBUTION

All authors contributed equally to this work. Conceived and design the study - TG, ED and GB; collected de data - TG, ED and GB; wrote the first draft and approved the final manuscript - TG, ED and GB.

REFERENCES

- Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*. 2006; 313(5786): 523-6.
- D'Arc M, Ayouba A, Esteban A, Learn GH, Boue V, Liegeois F, et al. Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proc Natl Acad Sci USA*. 2015; 112(11): E1343-52.
- Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*. 1983; 220(4599): 868-71.
- Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature*. 2008; 455(7213): 661-4.
- Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science*. 2014; 346(6205): 56-61.
- Tebit DM, Arts EJ. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis*. 2011; 11(1): 45-56.
- Tatem AJ, Hemelaar J, Gray RR, Salemi M. Spatial accessibility and the spread of HIV-1 subtypes and recombinants in sub-Saharan Africa. *AIDS*. 2012; 26(18): 2351-60.
- Magiorkinis G, Angelis K, Mamais I, Katzourakis A, Hatzakis A, Albert J, et al. The global spread of HIV-1 subtype B epidemic. *Infect Genet Evol*. 2016; 46: 169-79.
- Arien KK, Vanham G, Arts EJ. Is HIV-1 evolving to a less virulent form in humans? *Nat Rev Microbiol*. 2007; 5(2): 141-51.
- Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*. 2004; 303(5656): 327-32.
- Diez-Fuertes F, Cabello M, Thomson MM. Bayesian phylogeographic analyses clarify the origin of the HIV-1 subtype A variant circulating in former Soviet Union's countries. *Infect Genet Evol*. 2015; 33: 197-205.
- Cabello M, Mendoza Y, Bello G. Spatiotemporal dynamics of dissemination of non-pandemic HIV-1 subtype B clades in the Caribbean region. *PLoS One*. 2014; 9(8): e106045.
- Delatorre E, de Azevedo SSD, Rodrigues-Pedro A, Velasco-de-Castro CA, Couto-Fernandez JC, Pilotto JH, et al. Tracing the origin of a singular HIV-1 CRF45_cpx clade identified in Brazil. *Infect Genet Evol*. 2016; 46: 223-32.
- Delatorre E, Velasco-De-Castro CA, Pilotto JH, Couto-Fernandez JC, Bello G, Morgado MG. Reassessing the origin of the HIV-1 CRF02_AG lineages circulating in Brazil. *AIDS Res Hum Retroviruses*. 2015; 31(12): 1230-7.
- de Oliveira T, Pybus OG, Rambaut A, Salemi M, Cassol S, Ciccozzi M, et al. Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak. *Nature*. 2006; 444(7121): 836-7.
- Rouet F, Nouhin J, Zheng DP, Roche B, Black A, Prak S, et al. Massive iatrogenic outbreak of human immunodeficiency virus type 1 in rural Cambodia, 2014-2015. *Clin Infect Dis*. 2018; 66(11): 1733-41.
- Vinken L, Fransen K, Cuypers L, Alexiev I, Balotta C, Debaisieux L, et al. Earlier initiation of antiretroviral treatment coincides with an initial control of the HIV-1 sub-subtype F1 outbreak among men-having-sex-with-men in Flanders, Belgium. *Front Microbiol*. 2019; 10: 613.
- Vasylyeva TI, du Plessis L, Pineda-Pena AC, Kuhnert D, Lemey P, Vandamme AM, et al. Tracing the impact of public health interventions on HIV-1 transmission in Portugal using molecular epidemiology. *J Infect Dis*. 2019; 220(2): 233-43.
- Mir D, Graf T, Almeida SEM, Pinto AR, Delatorre E, Bello G. Inferring population dynamics of HIV-1 subtype C epidemics in Eastern Africa and Southern Brazil applying different Bayesian phylodynamics approaches. *Sci Rep*. 2018; 8(1): 8778.

20. Hué 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 USA*. 2005; 102(12): 4425-9.
21. Hughes GJ, Fearnhill E, Dunn D, Lycett SJ, Rambaut A, Leigh Brown AJ. Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. *PLoS Pathog*. 2009; 5(9): e1000590.
22. Kouyos RD, von Wyl V, Yerly S, Böni J, Taffé P, Shah C, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. *J Infect Dis*. 2010; 201(10): 1488-97.
23. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjörnsson J. Defining HIV-1 transmission clusters based on sequence data. *AIDS*. 2017; 31(9): 1211-22.
24. Volz EM, Koopman JS, Ward MJ, Brown AL, Frost SDW. Simple epidemiological dynamics explain phylogenetic clustering of HIV from patients with recent infection. *PLoS Comput Biol*. 2012; 8(6): e1002552.
25. Poon AFY. Impacts and shortcomings of genetic clustering methods for infectious disease outbreaks. *Virus Evol*. 2016; 2(2): vew031.
26. Ratmann O, van Sighem A, Bezemer D, Gavryushkina A, Jurriaans S, Wensing A, et al. Sources of HIV infection among men having sex with men and implications for prevention. *Sci Transl Med*. 2016; 8(320): 320ra2.
27. Poon AFY, Gustafson R, Daly P, Zerr L, Demlow SE, Wong J, et al. Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. *Lancet HIV*. 2016; 3(5): e231-8.
28. de Oliveira T, Kharsany ABM, Gräf T, Cawood C, Khanyile D, Grobler A, et al. Transmission networks and risk of HIV infection in KwaZulu-Natal, South Africa: a community-wide phylogenetic study. *Lancet HIV*. 2017; 4(1): e41-e50.
29. Grabowski MK, Lessler J, Redd AD, Kagaayi J, Laeyendecker O, Ndyanabo A, et al. The role of viral introductions in sustaining community-based HIV epidemics in rural Uganda: evidence from spatial clustering, phylogenetics, and egocentric transmission models. *PLoS Med*. 2014; 11(3): e1001610.
30. Junqueira DM, de Medeiros RM, Gräf T, Almeida SEM. Short-term dynamic and local epidemiological trends in the South American HIV-1B epidemic. *PLoS One*. 2016; 11(6): e0156712.