Journal of Global Infectious Diseases
Translating Science from Benchside to Bedside and Beyond
Official Publishing of International Infectiologists Network
Home | Current issue | Instructions | Submit article

J Glob Infect Dis. 2016 Jul-Sep; 8(3): 108–114.

doi: 10.4103/0974-777X.188592

PMCID: PMC4997794

Phylogenetic Analysis of Human Immunodeficiency Virus Type 2 Group B

Eleonora Cella,[1,2] Alessandra Lo Presti,[1] Marta Giovanetti,[1,3] Carla Veo,[4] Alessia Lai,[4] Giordano Dicuonzo,[5] Silvia Angeletti,[5] Marco Ciotti,[6] Gianguglielmo Zehender,[4] and Massimo Ciccozzi[1,7]

Address for correspondence: Dr. Massimo Ciccozzi, E-mail: ti.ssi@izzoccic

Abstract

INTRODUCTION

Human immunodeficiency virus types 1 and 2 (HIV-1 and HIV-2) are distinct human lentiviruses that cause acquired immunodeficiency syndrome. SIVcpz from chimpanzees (*pan troglodytes troglodytes*) and SIVsmm from sooty mangabeys (*Cercocebus atys atys*) are known to have crossed the species barrier, generating HIV-1 and HIV-2, respectively.[1] The major clinical difference between the two infections is that progression to immunodeficiency occurs more slowly in HIV-2 infection compared with HIV-1; hence, long-term nonprogressors are observed more frequently in HIV-2. Because of the high degree of replication in vivo and the error-prone process of replication, lentiviruses undergo rapid evolution in an infected host.[2] HIV-2 was the first described and isolated in 1985 and 1986, respectively,[3] in West Africa. Worldwide, about 1-2 million of those living with HIV are infected with HIV-2.[4] HIV-2 infections are mainly restricted to West Africa, including Guinea-Bissau, Gambia, Senegal, Cape Verde, Cote d'Ivoire, Mali, Sierra Leone, and Nigeria. In addition to West Africa, countries with past socioeconomical links with Portugal including southwest India[5,6,7] have a significant numbers of HIV-2 infections.[8,9] In the recent years, the prevalence of HIV-2 is a growing concern in some European countries.

In Europe, HIV-2 is mainly reported in Portugal and France countries that have socioeconomic relationship with some Western African regions. In France, HIV-2 represents 1.8% of new HIV infections.[9]

HIV-2 can be divided into epidemic Groups (A and B) and nonepidemic Groups (C-G).[9,10]

HIV-2 Groups A and B have established human-to-human transmission chains, whereas the six other HIV-2 lineages do not appear to be transmissible among humans.[11] Although the rate of sexual transmission of HIV-2 is very low relative to that of HIV-1, this virus can be transmitted by the same routes as HIV-1;[12] moreover, molecular evidence of homosexual transmission of HIV-2 in Spain has been reported.[13]

By analyzing viral sequences obtained over several decades and calibrating a molecular clock based on observed nucleotide changes, a reliable rate of sequence evolution can be inferred. In addition, recently developed probabilistic methods that integrate phylogeographical with demographic and molecular clock

models supported by a firm statistical basis, allowed the simultaneous reconstruction of viral epidemics on spatial and temporal scales.[14]

Previous studies dated the introduction of HIV-2 Groups A and B into the human population estimated the epidemic history and the origin in Guinea-Bissau using a local population.[15,16]

This work was based on the analysis of a large *pol* gene dataset, including all the sequences available in HIV database (www.hiv.lanl.gov) until May 2014 to estimate the origin, the population dynamics, and the global phylogeography of HIV-2 Group B. The study was focused only on Group B because Group A was already studied,[15,17] whereas for the other groups, there are not enough sequences available in the HIV database to study the spatial–temporal diffusion.

MATERIALS AND METHODS

Dataset

All available HIV-2 Group B *pol* gene sequences obtained from the HIV Los Alamos HIV sequence database (www.hiv.lanl.gov) were retrieved until May 2014. It was chosen to study the *pol* gene region because of the large number of sequences available in HIV database.

Among all the available HIV-2 Group B sequences in the database, only sequences with known location and sampling date were considered. It was performed an alignment to identify and select the pol gene sequences. Then, an alignment including 236 sequences was obtained, but due to the over-representation of the United Kingdom sequences, a randomization was performed to restore an equitable distribution among different countries.

The final dataset of Group B included 52 sequences from Belgium (n = 1), Cote D'Ivoire (n = 1), France (n = 3), Ghana (n = 4), Japan (n = 1), Luxembourg (n = 1), Mali (n = 4), Senegal (n = 5), and the United Kingdom (n = 32). The sequences were collected from 1986 to 2009.

The sequences were aligned using Clustal X and manually edited by Bioedit software version 7.2.5 as already described.[18] Modeltest 3.7 was used[19] to select the simplest evolutionary model that fitted adequately the sequence data.

Likelihood-mapping

To obtain an overall impression of the phylogenetic signals in the HIV-2 *pol* gene Group B dataset, a likelihood-mapping analysis, undertaken of 10,000 random quartets generated using Tree Puzzle was carried out as already described.[20] A likelihood map consists of an equilateral triangle containing dots representing the likelihoods of the three possible un-rooted trees for a set of four sequences (quartets) randomly selected from the dataset–the dots close to the corners or at the sides represent tree-like (fully resolved phylogenies in which one tree is clearly better than the others) or network-like phylogenetic signals (three regions in which it is not possible to decide between two topologies). The central area of the map represents a star-like signal (the region in which the star tree is optimal). When using this strategy, if more than 30% of the dots fall into the center of the triangle, the data are considered unreliable for the purposes of phylogenetic inference.

Evolutionary rate estimate and demographic history

The evolutionary rates for HIV-2 dataset were estimated using a Bayesian Monte Carlo Markov Chain (MCMC) approach[21,22] (BEAST version 1.7.4, http://beast.bio.ed.ac.uk) implementing the model selected with Modeltest using both a strict and an uncorrelated log-normal relaxed clock model.

As coalescent priors, three parametric demographic models of population growth (constant size, exponential, and expansion growth) and a Bayesian skyline plot (BSP, a nonparametric piecewise-constant model) were compared. The best fitting models were selected using a Bayes factor (BF with marginal likelihoods).

Chains were conducted for 50×10^6 generations and sampled every 5000 steps.

In accordance with Kass and Raftery,[23] the strength of the evidence against H0 (null hypothesis) was evaluated as follows: 2lnBF <2, no evidence; 2-6, weak evidence; 6-10, strong evidence; and >10, very strong evidence. A negative value indicates evidence in favor of H0. Only values of ≥6 were considered statistically significant. Convergence of the MCMC was assessed by calculating the effective sampling size (ESS) for each parameter. Only parameter estimates with ESS's of >200 were accepted.

The Bayesian tree of the first dataset was summarized in a target tree by Tree Annotator program implemented in the Beast package by choosing the tree with the maximum product of posterior probabilities (maximum clade credibility [MCC]) after a 10% burn-in.

The demographic history was estimated on HIV-2 dataset on an individual basis by comparing four coalescent models (constant population size, exponential growth, expansion growth, and BSP) under a lognormal relaxed molecular clock model.

Phylogeographic analysis

The phylogeographic analysis was performed on the HIV-2 dataset using the evolutionary rate estimated previously. The continuous-time Markov Chain (CTMC) process over discrete sampling locations implemented in BEAST 1.7.4[24] was used for the geographical analysis, implementing the Bayesian Stochastic Search Variable Selection model that allows the diffusion rates to be zero with a positive prior probability. Comparison of the posterior and prior probabilities of the individual rates being zero provided a formal BF for testing the significance of the linkage between locations.

Viral gene flow analysis

The MacClade version 4 program (Sinauer Associates, Sunderland, MA) was used to count viral gene out/in flow in dataset among different locations; this was tested with a modified version of the Slatkin and Maddison test.[24]

One-character data matrix was obtained from the original dataset by assigning to each taxon in the tree, a one-letter code indicating its own sampling location. Then, the putative origin of each ancestral sequence (i.e., internal node) in the tree is inferred by finding the most parsimonious reconstruction of the ancestral character. The final tree length, i.e., the number of observed migrations in the genealogy, can easily be computed and compared to the tree-length distribution of 10,000 trees obtained by random joining-splitting. Observed genealogies significantly shorter than random trees indicate the presence of subdivided population with restricted gene flow.

To perform this analysis, the HIV-2 Group B sequences were assigned to three groups such as EU1 (the United Kingdom), EU2 (France, Belgium, and Luxembourg), and AF (Africa).

Sequence from Japan (for the genotype B) was not included due to their limited representativeness.

The null hypothesis of panmixia (i.e., no population subdivision or complete intermixing of sequences from different geographic areas) was rejected by the randomization test ($P < 0.0001$).[24]

RESULTS

Likelihood-mapping

The phylogenetic noise of dataset was investigated by means of likelihood-mapping. The percentage of dots falling in the central area of the triangles was 13.5% for HIV-2 Group B [Supplementary Figure 1]. As the dataset did not show more than 30% of noise, it contained sufficient phylogenetic signal.

Supplementary Figure 1

Likelihood-mapping of HIV-2 Group B pol gene. The dots inside the triangles represent the likelihood of the possible unrooted topologies for each quartet. Numbers indicate the percentage of dots in the center of the triangle corresponding to phylogenetic noise (star-like trees).

Evolutionary rate estimate and demographic history

The mean evolutionary rate of HIV-2 Group B *pol* gene was estimated using a Bayesian MCMC approach, implementing the TN93+I+G nucleotide substitution model. BF analysis showed that the relaxed clock fitted the data significantly better than strict clock (2lnBF between the strict and relaxed clock was 26.9 in favor of the second).

Under the relaxed clock, BF analysis showed that BSP was better than the other models (2lnBF > 24). The estimated mean value of HIV-2 Group B *pol* gene evolutionary rate was 1.8×10^{-3} substitution/site/year (95% highest posterior density [HPD]: $7.21 \times 10^{-4} - 3.06 \times 10^{-3}$).

The root of the Bayesian MCC tree of HIV-2 Group B *pol* gene sequences [Figure 1] dated back to 1957 (95% HPD: 1927–1980).

Figure 2 shows the population dynamics of the HIV-2 Group B *pol* gene.

The demographic history of HIV-2 Group B showed that the epidemic remained constant up to 1970 when started an exponential growth. From 1985 to early 2000s, the epidemic reached a plateau; then it was characterized by two bottlenecks and a new plateau at the end of 2000s.

Phylogeographic analysis

Figure 1 shows the phylogeographic analysis of HIV-2 Group B pol gene. Phylogeographic reconstruction showed that the most probable location for the root of the tree was Ghana (state probability, sp = 0.46).

Two statistically supported (posterior probability = 1) clades were found a main clade (A) that dated back to 1967 (HPD 95% 1945-1983) and a minor clade (B), which dated back to 1973 (HPD 95% 1955-1985) and both originated in Ghana (sp = 0.45 for A and sp = 0.62 for B).

Within the A clade, several statistically supported clusters (A1, A2, and A3) were identified.

The most probable origin for the cluster A2, which dated back to 1975 (95% HPD: 1963-1984) and is the oldest one among the three clusters, was Ghana (sp = 0.60).

Particularly, within cluster A2, two sequences from Ghana sampled in 1986 (95% HPD: 1985-1986) represented the out group of this cluster; all the strains from the United Kingdom grouped in a statistically supported cluster dating back to 2000 (95% HPD: 1995-2005) and representing the youngest in the tree.

Cluster A3 included sequences from Senegal and Mali; the most probable origin was in Mali in 1978 (95% HPD: 1957-1995).

Cluster A1 included only two strains, one from Belgium and the other from Japan, and originated in Japan.

The minor clade (B) included two statistically supported cluster named B1 and B2. Cluster B1 included sequences from Ghana, Cote d'Ivoire, Luxemburg, Mali, and France and dated in 1976 (95% HPD: 1963-1986); cluster B2 involved only two sequences from France and Mali and originated in Ghana.

After its Ghanaian origin, HIV-2 followed the statistically significant routes involving Japan, Mali, and the UK (sp = 0.26, sp = 0.29, and sp = 0.99, respectively).

Viral gene flow analysis

Figure 3 shows the viral gene flow of HIV-2 genotype B *pol* gene sequences. The null hypothesis of panmixia (i.e., no population subdivision or complete intermixing of sequences from different geographic areas) was rejected by the randomization test ($P < 0.0001$). The only observed viral gene flow was from Africa to France, Belgium, and Luxembourg (75%, EU2).

DISCUSSION

About 1-2 million of subjects living with HIV are infected with HIV-2 worldwide.[2] In the recent years, the prevalence of HIV-2 is a growing concern in some European countries[10] such as Portugal and France.[25,26]

In the present study, an accurate molecular clock strategy to investigate the epidemic history of HIV-2 Group B, to estimate the timescale of the epidemic, and to infer the demographic history of the virus was performed.[27] The demographic history of HIV-2 Group B is characterized by a first period of low endemicity followed by an exponentially increasing number of infections. Our phylodynamic analysis of HIV-2 Group B grew of 1-log from 1970 to 1985 when reached a plateau. In the early 2000s, the epidemic showed two bottlenecks and a new phase of plateau at the end of 2000s. Our demographic estimates suggest that some environmental and social events enabled HIV-2 Group B to switch the epidemic growth around 1970-1985 1985 whereas in 2000s there was a decreasing phase and bottlenecks.

The independence war from Portugal in Guinea-Bissau, a former Portuguese colony that occurred between 1963 and 1974 probably played an important role as route of transmission of Group B as already postulate by Lemey *et al*.[15] This statement is also supported by the fact that the first reported cases of HIV-2 in Europe were Portuguese veterans who had served in the army during the independence war.[27,28] Moreover, commercial sex work in growing cities (after 50s years) played an important role in maintaining the epidemic. Similarly, the virus spread to other colonial countries such as French and England. The number of infected persons had an exponential amplification in the early 1950s probably because the facility of injection administered mostly for treatment of presumed syphilis.[29] The phylogeographic routes of HIV-2 Group B migration highlighted by the present study are probably supported by social-political events such as wars, colonialism, population movements that favor the diffusion, and spread of the epidemic.

The most probable geographic origin of HIV-2 Group B was in Ghana; from there, a close inspection of viral flow revealed a scenario of multiple introductions resulting in different paths of transmission clusters involving African and European countries. As expected, the main phylogeographic path was within African countries with the recent European and Asiatic introduction involving the UK and Japan. Excluding colonialism and wars, because the route of the virus entrance in the UK dated back 2000 year, we can hypotheses that this event can be due at an important migration flow (i.e., refugees). Indeed, in Western Europe, the year 2000 was characterized by three major interrelated developments with a direct bearing on UNHCR's protection activities. The first was the acceleration of the European integration process, particularly with regard to the enlargement of the European Union. The second was the very noticeable return of legal immigration to the Western European political agenda, with discussion focusing almost exclusively on the demographics of labor, i.e., economic necessity. The third was the commitment of Member States of the European Union to the development of a common asylum system. Western European governments showed increasing readiness to co-operate with UNHCR on subjects of common interest, such as asylum and migration, the root causes of humanitarian crises, the provision of humanitarian aid, and solutions to protracted refugee situations. Totally, 452,350 asylum applicants in Western Europe was registered in 2000 (roughly 20,000 fewer than in 1999); Northern Europe remaining the main destination for asylum-seekers. The largest number of asylum claims was received in Germany (including re-opened applications), the United Kingdom, the Netherlands, and France (www.unhcr.org).

Regarding the viral gene flow of HIV-2 Group B, the only viral gene flow observed (75%) was from Africa to European countries such as France, Belgium, and Luxembourg as consequence of wars, colonialism, and movements of people contributing to the diffusion of the epidemic.

Data from viral gene flow and phylogeographic analysis seem to suggest that HIV-2 Group B originated in Ghana moving toward European countries during the year 1970s. Recently, HIV-2 Group B was isolated more frequently in the UK probably as consequence of migration flows; in fact, the phylogeographic analysis dated the UK sequences in the year 2000, the youngest cluster in the tree.

A possible limitation of our study can be the limited availability of sequences with known sampling date and isolation country sometimes. To optimize the analysis, we tried to overtake any possible bias randomizing the samples to have the greater homogeneity in our heterogeneity. Maybe the dissemination of HIV-2 Group B needs to be further investigated using a large dataset.

In conclusion, the study gives insights into the origin, history, and phylogeography of the HIV-2 Group B epidemic. Despite the low virus transmissibility of HIV-2, the growing number of infections worldwide indicates the need of strengthening surveillance. In particular in the European countries is important to monitor the diffusion of HIV-2 infection from endemic countries.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

Acknowledgments

The authors would like to thank Dr. Valerio Ciccozzi for the English revision of the manuscript.

REFERENCES

1. Peeters M, Courgnaud V, Abela B. Genetic diversity of lentiviruses in non-human primates. AIDS Rev. 2001;3:ss3–10.

2. Nyamweya S, Hegedus A, Jaye A, Rowland-Jones S, Flanagan KL, Macallan DC. Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis. Rev Med Virol. 2013;23:221–40. [PubMed: 23444290]

3. Barin F, M'Boup S, Denis F, Kanki P, Allan JS, Lee TH, et al. Serological evidence for virus related to simian T-lymphotropic retrovirus III in residents of West Africa. Lancet. 1985;2:1387–9. [PubMed: 2867393]

4. Clavel F, Guétard D, Brun-Vézinet F, Chamaret S, Rey MA, Santos-Ferreira MO, et al. Isolation of a new human retrovirus from West African patients with AIDS. Science. 1986;233:343–6. [PubMed: 2425430]

5. Babu PG, Saraswathi NK, Devapriya F, John TJ. The detection of HIV-2 infection in southern India. Indian J Med Res. 1993;97:49–52. [PubMed: 8505073]

6. Kulkarni S, Thakar M, Rodrigues J, Banerjee K. HIV-2 antibodies in serum samples from Maharashtra state. Indian J Med Res. 1992;95:213–5. [PubMed: 1478722]

7. Rübsamen-Waigmann H, Briesen HV, Maniar JK, Rao PK, Scholz C, Pfützner A. Spread of HIV-2 in India. Lancet. 1991;337:550–1. [PubMed: 1671908]

8. Schim van der Loeff MF, Aaby P. Towards a better understanding of the epidemiology of HIV-2. AIDS. 1999;13(Suppl A):S69–84. [PubMed: 10885765]

9. Valadas E, França L, Sousa S, Antunes F. 20 years of HIV-2 infection in Portugal: Trends and changes in epidemiology. Clin Infect Dis. 2009;48:1166–7.[PubMed: 19292644]

10. Ciccozzi M, Babakir-Mina M, Cella E, Bertoli A, Lo Presti A, Maniar JK, et al. A case of Italian HIV type 2 infection: A genetic analysis. AIDS Res Hum Retroviruses. 2011;27:1333–5. [PubMed: 21453133]

11. De Cock KM, Adjorlolo G, Ekpini E, Sibailly T, Kouadio J, Maran M, et al. Epidemiology and transmission of HIV-2. Why there is no HIV-2 pandemic. JAMA. 1993;270:2083–6. [PubMed: 8147962]

12. Kanki PJ, Travers KU, MBoup S, Hsieh CC, Marlink RG, Gueye-NDiaye A, et al. Slower heterosexual spread of HIV-2 than HIV-1. Lancet. 1994;343:943–6.[PubMed: 7909009]

13. Cilla G, Rodés B, Pérez-Trallero E, Arrizabalaga J, Soriano V. Molecular evidence of homosexual transmission of HIV type 2 in Spain. AIDS Res Hum Retroviruses. 2001;17:417–22. [PubMed: 11282010]

14. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009;5:e1000520. [PMCID: PMC2740835][PubMed: 19779555]

15. Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M, Vandamme AM. Tracing the origin and history of the HIV-2 epidemic. Proc Natl Acad Sci U S A. 2003;100:6588–92. [PMCID: PMC164491] [PubMed: 12743376]

16. Wertheim JO, Worobey M. Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. PLoS Comput Biol. 2009;5:e1000377. [PMCID: PMC2669881][PubMed: 19412344]

17. Faria NR, Hodges-Mameletzis I, Silva JC, Rodés B, Erasmus S, Paolucci S, et al. Phylogeographical footprint of colonial history in the global dispersal of human immunodeficiency virus type 2 group A. J Gen Virol. 2012;93(Pt 4):889–99. [PMCID: PMC3542711] [PubMed: 22190015]

18. Ciccozzi M, Callegaro A, Lo Presti A, Cella E, Giovanetti M, Salpini R, et al. When phylogenetic analysis complements the epidemiological investigation: A case of HIV-2 infection, Italy. New Microbiol. 2013;36:93–6. [PubMed: 23435822]

19. Posada D, Buckley TR. Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst Biol. 2004;53:793–808. [PubMed: 15545256]

20. Ciccozzi M, Ciccaglione AR, Lo Presti A, Equestre M, Cella E, Ebranati E, et al. Evolutionary dynamics of HBV-D1 genotype epidemic in Turkey. J Med Virol. 2014;86:109–16. [PubMed: 24243521]

21. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005;22:1185–92. [PubMed: 15703244]

22. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007;7:214. [PMCID: PMC2247476] [PubMed: 17996036]

23. Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc. 1995;90:773–95.

24. Slatkin M, Maddison WP. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics. 1989;123:603–13. [PMCID: PMC1203833][PubMed: 2599370]

25. Yusim K, Peeters M, Pybus OG, Bhattacharya T, Delaporte E, Mulanga C, et al. Using human immunodeficiency virus type 1 sequences to infer historical features of the acquired immune deficiency syndrome epidemic and human immunodeficiency virus evolution. Philos Trans R Soc Lond B Biol Sci. 2001;356:855–66.[PMCID: PMC1088479] [PubMed: 11405933]

26. Cazein F, Pillonel J, Le Strat Y, Lot F, Pinget R, David D, et al. Surveillance de l'infection à VIH-sida en France, 2007. Bull Epidemiol Hebd. 2008;45-46:434–42.

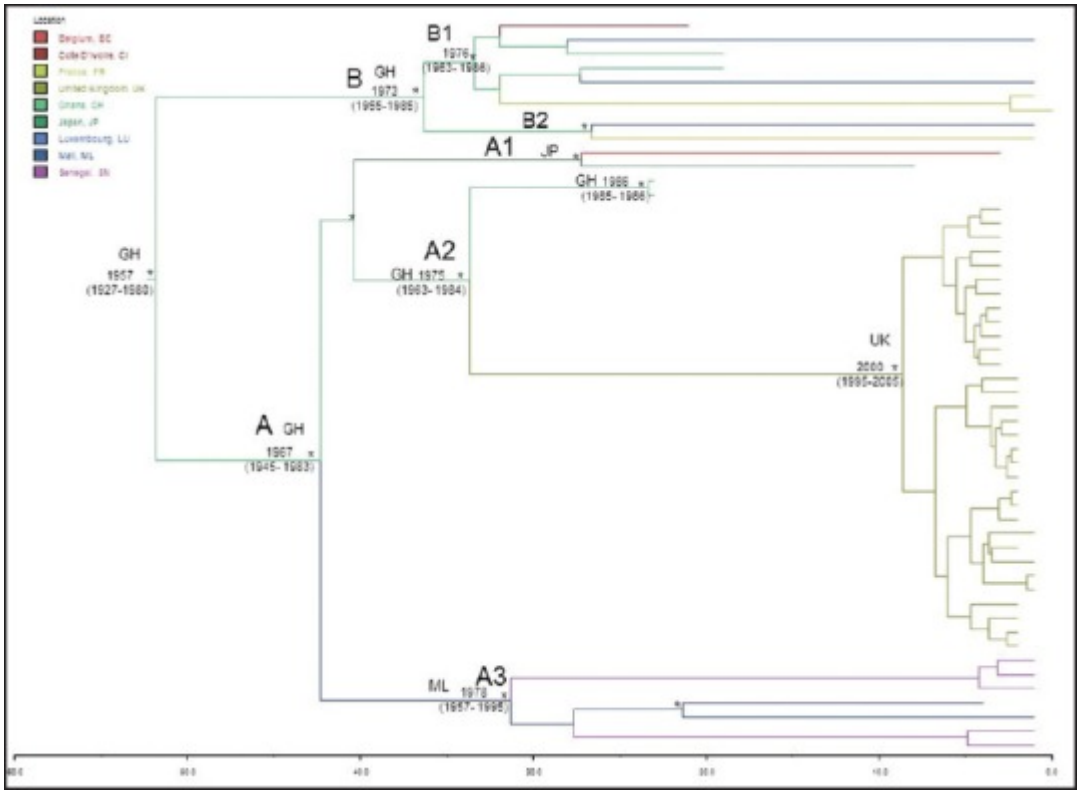
27. Poulsen AG, Kvinesdal B, Aaby P, Mølbak K, Frederiksen K, Dias F, et al. Prevalence of and mortality from human immunodeficiency virus type 2 in Bissau, West Africa. Lancet. 1989;1:827–31. [PubMed: 2564911]

28. Piedade J, Venenno T, Prieto E, Albuquerque R, Esteves A, Parreira R, et al. Longstanding presence of HIV-2 infection in Guinea-Bissau (West Africa) Acta Trop. 2000;76:119–24. [PubMed: 10936570]

29. Pépin J. The origins of AIDS: From patient zero to ground zero. J Epidemiol Community Health. 2013;67:473–5. [PubMed: 23322854]

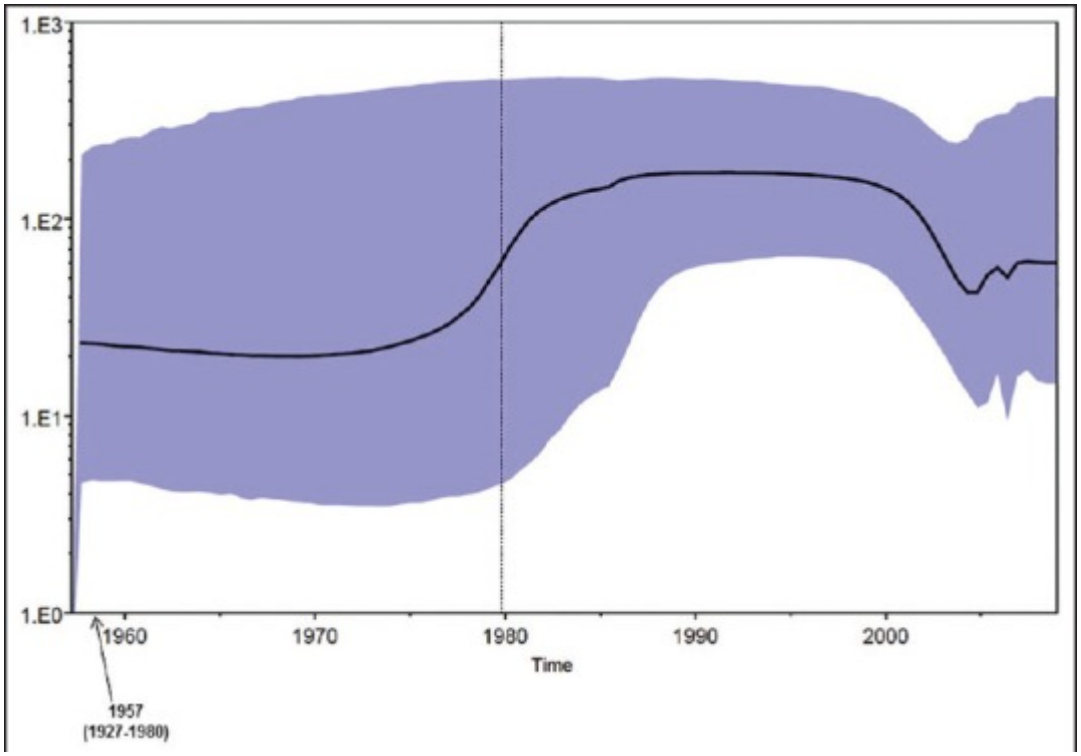
Figures and Tables

Figure 1



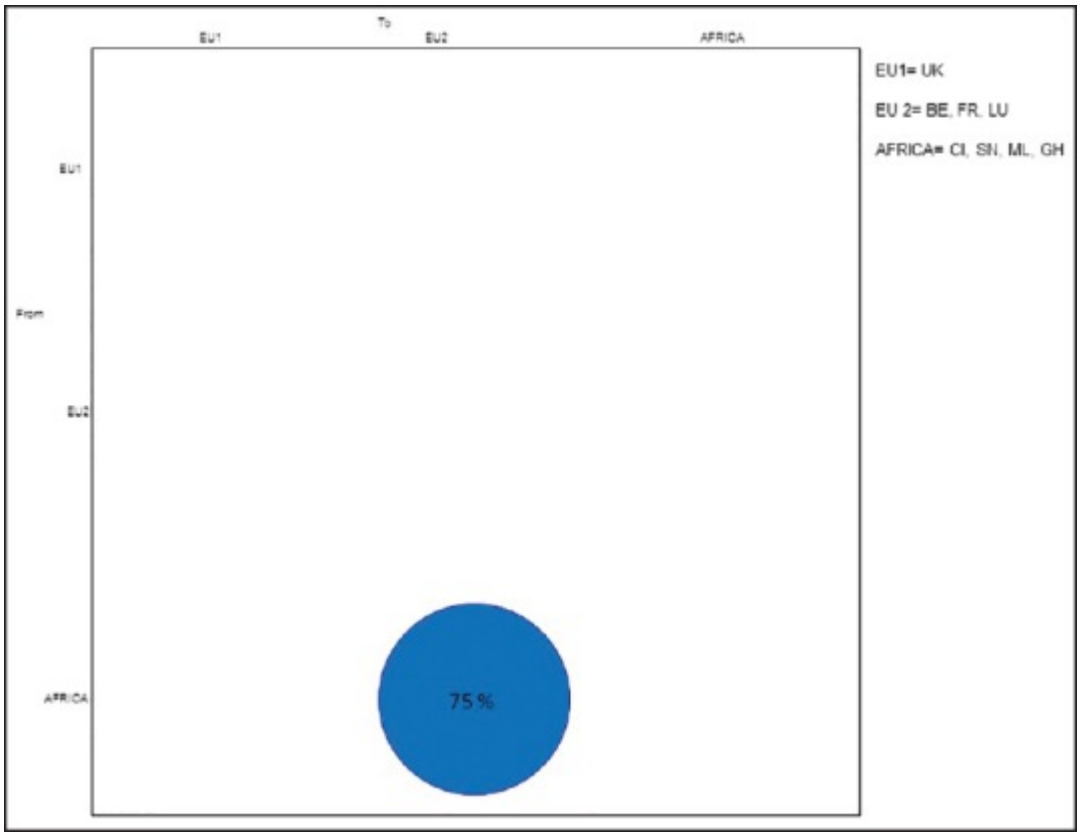
Bayesian phylogeographic tree of HIV-2 Group B *pol* gene sequences. The branches are colored on the basis of the most probable state location of the descendent nodes. States are reported on the left corner of the figure. The scale at the bottom of the tree represents time in years. Years are reported for the main nodes.

Figure 2



Bayesian skyline plot of the HIV-2 Group B *pol* gene. The effective number of infections is reported on the Y-axis. Time is reported on the X-axis. The colored area corresponds to the credibility interval based on 95% highest posterior density interval.

Figure 3



Maximum parsimony migration patterns of the HIV-2 Group B *pol* gene to and from different areas. The bubblegram shows the frequency of gene flow (migrations) to and from different areas. The surface of each circle is proportional to the percentage of observed migrations given within the circle. Migrations were inferred with a modified version of the Slatkin and Maddison algorithm.
