Title: Tracing the Impact of Public Health Interventions on HIV-1 Transmission in Portugal Using Molecular Epidemiology

Tetyana I Vasylyeva[1,2], Louis du Plessis[1], Andrea C Pineda-Peña[3,4], Denise Kühnert[5], Philippe Lemey[6], Anne-Mieke Vandamme[3,6], Perpétua Gomes[7,8], Ricardo J Camacho[6,*], Oliver G Pybus[1], Ana B Abecasis[3], Nuno R Faria[1]

[1]Department of Zoology, University of Oxford, Oxford, 19 d, UK;

[2] New College, University of Oxford, Oxford, UK;

[3]Center for Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisbon, Portugal;

[4]Molecular Biology and Immunology Department, Fundación Instituto de Inmunología de Colombia (FIDIC); Basic Sciences Department, Universidad del Rosario, Bogotá, Colombia;

[5]Max Planck Institute for the Science of Human History, Jena, Germany;

[6]KU Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Clinical and Epidemiological Virology, Leuven, Belgium;

[7]Laboratory of Molecular Biology, LMCBM, SPC, HEM – CHLO, Lisbon, Portugal;

[8]Centre for Interdisciplinary Research Egas Moniz, CiiEM, Almada, Portugal;

* deceased.

To whom correspondence should be addressed:

Dr. Nuno Rodrigues Faria (nuno.faria@zoo.ox.ac.uk)

Department of Zoology, University of Oxford, OX1 3SY, Oxford, United Kingdom

Dr. Tetyana I Vasylyeva (tetyana.vasylyeva@zoo.ox.ac.uk)

Department of Zoology, University of Oxford, OX1 3SY, Oxford, United Kingdom

Summary:

Birth-death phylodynamic models uncover how HIV transmission dynamics was decelerated by harm reduction programs in the 1990s in Portugal. After 2000, transmissions in heterosexuals and people who inject drugs were driving the Portuguese HIV-1 subtype B and G epidemic, respectively.

Conflict of interest: The authors declare no conflict of interest.

Funding: T.I.V. is supported by the Juliana Cuyler Matthews Junior Research Fellowship from New College of the University of Oxford. N.R.F. is funded by a Royal Society and Wellcome Trust Sir Henry Dale Fellowship (grant 204311/Z/16/Z). OGP and LdP were supported by the European Research Council under the European Commission Seventh Framework Programme (FP7/2007-2013)/ European Research Council grant agreement 614725-PATHPHYLODYN. DK is supported by the Max Planck Society. This study was partially supported by European Funds through grant 'Bio-Molecular and Epidemiological Surveillance of HIV Transmitted Drug Resistance, Hepatitis Co-Infections and Ongoing Transmission Patterns in Europe - BEST HOPE - (project funded through HIVERA: Harmonizing Integrating Vitalizing European Research on HIV/Aids, grant 249697)'; by L'Oréal Portugal Medals of Honor for Women in Science 2012 (financed through L'Oréal Portugal, Comissão Nacional da Unesco and Fundação para a Ciência e Tecnologia, FCT); by FCT for funds to GHTM-UID/Multi/04413/2013; by the MigrantHIV project (financed by FCT: PTDC/DTP-EPI/7066/2014); and by Gilead Génese HIVLatePresenters. This study was supported in part by the Research Foundation Flanders (FWO G.0692.14), and by the VIROGENESIS project that receives funding from the European Union's Horizon 2020 Research and Innovation Programme (under Grant Agreement no. 634650).

Parts of this work were previously presented at:

The 24th International HIV dynamics & evolution conference, May 23 - 26, 2017

The 2nd Virus Genomics & Evolution Conference, June 18-20, 2018

Abstract (word count = 198)

Background

Estimation of temporal changes in HIV transmission patterns can help to elucidate the impact of preventative strategies and public health policies.

Methods

Portuguese HIV-1 subtype B and G *pol* genetic sequences were appended to global reference datasets to identify country-specific transmission clades. Bayesian birth-death models were used to estimate subtype-specific effective reproductive numbers (R_e). Discrete trait analysis (DTA) was used to quantify mixing among transmission groups.

Results

We identified five subtype B Portuguese clades (*n*=26-79 sequences) and a large monophyletic subtype G Portuguese clade (*n*=236). We estimated that major shifts in HIV-1 transmission occurred around 1999 (95% Bayesian credible interval 1998-2000) and 2000 (1998-2001) for subtypes B and G, respectively. For subtype B, R_e dropped from 1.91 (1.73-2.09) to 0.62 (0.52-0.72). For subtype G, R_e decreased from 1.49 (1.39-1.59) to 0.72 (0.63-0.8). The DTA suggests that people who inject drugs (PWID) and heterosexuals were at the source of most (>80%) virus lineage transitions for subtypes G and B, respectively.

Conclusions

The estimated declines in R_e coincide with the introduction of highly active antiretroviral therapy and the scale-up of harm reduction for PWID. Inferred transmission events across transmission groups emphasize the importance of prevention efforts for bridging populations.

Key words: HIV, Portugal, phylogenetics, epidemiology, reproductive number, transmission groups, harm reduction.

Background

HIV-1 subtype B dominates the HIV epidemic in Western European countries (1). Since the 1980s, the Portuguese HIV epidemic was concentrated in heterosexual (HET) populations and subtype B was the most common strain (2). However, by 2001 nearly two-thirds of reported AIDS cases in Portugal were linked to people who inject drugs (PWID), and the country had the highest HIV incidence in PWID in the European Union (3, 4). The proportion of PWID in the newly diagnosed patients in Portugal stands at 35% overall, but this number declined from 45% before 2004 to 15% after (5). Although the HIV Portuguese epidemic in PWID was initially associated with subtype B transmission, subtype G and CRF14_BG infections grew rapidly in this group (6). Currently, HIV molecular profile in Portugal is very diverse, with a variety of HIV-1 subtypes (A, C, D, H, J, F) and recombinant forms, as well as a higher proportion of HIV-2 infections than in other European settings (1, 2, 6-9).

Besides the introduction of anti-retroviral therapy (ART) in 1996, several public health approaches have been used to reduce HIV incidence in Portugal. Since the early 1990s Portugal has been at the forefront of the implementation of harm reduction strategies (10). Methadone opioid substitution treatment (OST) has been available to PWID since 1977, while buprenorphine OST became available in 1999 (11). In 2004 OST was extended to pharmacy-based provision, and the first syringe-exchange program (SEP) was established in 1993. In 2001, Portugal effectively decriminalised drug possession (possession of a small amount of drugs for self-consumption was made an administrative, rather than a criminal offense), which reportedly resulted in a reduction in the number of drug-related deaths and possibly a drop in HIV incidence (5, 12).

The effective reproductive number (R_e) is defined as the average number of secondary infections caused by an infected individual at a particular time point during an epidemic, when susceptibility in the population has decreased, e.g. due to immunity or

intervention measures. It is often used to describe transmission dynamics over the course of an epidemic. Recent developments in phylodynamic modelling enable genetic and epidemiological modelling to be combined in order to quantify R_e in HIV populations (13-15). Yet, R_e remains to be estimated for the main subtypes and transmission groups associated with the Portuguese HIV epidemic.

Here we collate 680 HIV-1 subtype B and G *pol* genetic sequences with known transmission group information from the Portuguese HIV drug resistance database. We use several phylodynamic approaches to investigate temporal changes in R_e of the predominant subtypes and circulating clades, and describe the viral lineages mixing among different transmission groups. Our findings highlight the effect of preventative interventions on the HIV-1 dynamics in Portugal.

Methods

Genetic data

HIV-1 subtype B and G *pol* genetic sequences were obtained from the Portuguese HIV database. For the purpose of this analysis we used all the sequences collected between 2001-2013 with available information regarding transmission risk group (self-reported by patients). Reference HIV-1 subtype B sequences were obtained from the SPREAD database (SPREADdb) created by the European Society for Translational Antiviral research (ESAR) (http://www.esar-society.eu/). SPREADdb contains annotated sequences generated from newly diagnosed patients during 2001-2013 from 26 European countries (16). For each Portuguese subtype B sequence, we extracted the 10 most similar sequences from SPREADdb. For HIV-1 subtype G, we used all publicly available subtype G sequences from the HIV Los Alamos database (LANLdb) (17) as reference sequences (Suppl. Table 1). We first codon-aligned all sequences using Clustal Omega (18) and subsequently edited the alignment manually in MEGA7 (19).

Initial phylogenetic analyses

We used the combined subtype B and G datasets (Portuguese, SPREADdb and LANLdb) to generate maximum likelihood (ML) phylogenies for each subtype (20). We used the Hasegawa-Kishino-Yano nucleotide substitution model with gamma-distributed rate variation among sites (HKY+G). For subtype B, we used Cluster Picker (21) to identify HIV-1 circulating clades; these were defined as the deepest Portuguese-only clades with >10 sequences, which exhibited within-clade genetic distance <6%, and SH-like statistical support >90%. Given that most Portuguese subtype G sequences belonged to a single well-supported clade (SH-like support 90%), we analysed all sequences (n=236) as a single phylogenetic clade. For each clade we estimated a ML tree using RAxML (22), and quantified temporal

signal by regressing the root-to-tip genetic distance of each sequence in the tree with its sampling date, as implemented in Tempest (23).

Estimating reproductive numbers for HIV lineages

We used birth-death models (BDM) implemented in BEASTv2.4 (24) to estimate time-varying rates of epidemic spread, measured as changes in the effective reproductive number, $R_e(t)$. In this model, lineages are added to the tree upon infection, at rate λ (birth rate), and removed at rate δ (death rate), upon becoming non-infectious (either through treatment, isolation or death). We make the assumption that individuals become non-infectious upon sampling and are removed from the pool of transmitters, since upon sampling they are provided treatment, and their viral load can be reduced to undetectable levels. However, it is important to note that not everyone who is diagnosed will be virally suppressed, and, therefore R_e, which is here defined by λ/δ (where $1/\delta$ is the average infectious period) represents a lower bound on the true value of R_e. We also estimate the sampling proportion and the "origin date" of the epidemic (which is older than the TMRCA of the tree reconstructed from the sampled sequences)(25). In addition, R_e, δ and the sampling proportion are allowed to instantaneously change at certain time-points (thus these parameters are piecewise-constant functions of time).

We used 3 different variations of BDM in our analyses. In Model 1 (*constant rate birth-death*) R_e was assumed to be constant through time, resulting in an average R_e over the course of the epidemic (13). Model 2 (*birth-death skyline, BDSKY*) assumes a piecewise constant R_e over 10 equidistant time-intervals, between the TMRCA and the most recent sample (26). Model 3 (*constant-shift-constant*) assumes a piecewise constant R_e, but with only two-epochs, in analogy with a previously described coalescent model (27). In addition to estimating the average R_e during an early and a late epoch, Model 3 is able to estimate directly from sequence data the timing of the transition (shift) in epidemic growth rate

between the two epochs, including for hierarchical model scenarios with multiple trees. We used TreeSlicer package in BEAST2 for these analyses (https://github.com/laduplessis/skylinetools).

R_e estimates were calculated separately for the two subtypes. For the subtype B clades, we used a hierarchical phylogenetic approach across all subtype B clades, in which the molecular clock, nucleotide substitution, and birth-death tree prior model parameters were all linked across clades (28, 29). We also analysed all five HIV-1 subtype B clades as a single phylogenetic tree (see Supplementary text). For all BEAST analyses, mixing of the Markov Chain Monte Carlo (MCMC) was visually inspected in Tracer (30). Analysis were typically run for 100-200 million MCMC steps, and effective sample sizes were >200 for all parameters. We used bdskytools package in R (https://github.com/laduplessis/bdskytools) to plot the results of BDSKY analyses (Model 2).

Discrete trait and association index analyses

To characterise population structure according to transmission groups, we estimated the phylogenetically-based association index (AI) statistic that quantifies the association between a phylogeny and trait values associated with each taxon. We quantified AI, which ranges from 0 to 1 (0 represents maximum population structure and 1 indicates panmixia), using three independent approaches implemented in BEAST, BaTS, and HyPhy (31-33) (see Supplementary Text).

To describe and quantify viral movement between different transmission risk groups, we performed a discrete trait diffusion analysis (DTA) using transmission group as a discrete trait (34) for clades with significant transmission group structure. Initially, each sequence was annotated with one of the five known transmission groups, specifically: “MSM”, “PWID”, “HET”, “MTC” (mother-to-child transmission), or “Other” (e.g. blood transfusion). To count the expected number of virus lineage movements among transmission groups we used a

robust counting (Markov jumps) approach (35, 36). Based on preliminary analyses, we next considered three transmission groups (with the two transmission groups that contributed the highest proportion of transmission group transition events in a given dataset and with all other transmission groups included in a new category called "Other"). For both analyses we used a Bayesian stochastic search variable selection procedure to estimate most relevant transition pathways among transmission groups. Statistical support was measured using Bayes Factors (BF) (37) and summarised using SpreaD3 (38). We assumed BF>10 is strong evidence for a well-supported viral pathway between two transmission groups (37).

Additional methodological details can be found in the Supplementary Text. All XMLs are available from GitHub repository https://github.com/HIVMolEpi/HIV-1-Portugal.

Results

Genetic sequences

A total of 3462 HIV-1 subtype B and 2678 subtype G *pol* genetic sequences were available from the Portuguese HIV database. Of these, self-reported transmission group information was available for 444 subtype B and 236 subtype G sequences. The transmission risk group distribution of sequences in this analysis is very similar to that reported in the Portuguese HIV/AIDS annual report, apart from the proportion of MTC sequences (13% in our analysis versus 1% in the report, see Suppl. Table 2) (12).

Additionally, we analysed 2165 unique reference subtype B sequences from SPREADdb and 1008 unique subtype G reference sequences from LANLdb. Datasets that include both Portuguese and reference sequences will be referred to as "combined" datasets. The combined datasets therefore comprised 2609 subtype B sequences (2165 from SPREADdb and 444 from the Portuguese database) and 1244 subtype G sequences (1008 from LANLdb and 236 from the Portuguese database) (Fig. 1).

Phylogenetic clade analysis

The ML phylogenetic trees estimated for the subtype B and G combined datasets are shown in Fig. 2. For subtype B, we identified 394 clades, of which 90 (representing 334 patients) included Portuguese sequences only. Five Portuguese circulating clades contained >10 sequences each and are denoted here as clades 1 (N=26), 2 (N=79), 3 (N=20), 4 (N=40), and 5 (N=42) (Fig. 2, Table 1). Clades were located across the global diversity of subtype B, indicating that subtype B has been introduced multiple times to Portugal.

Estimates of the effective reproductive number using genetic data

To estimate temporal changes in the effective reproductive number we used a suite of birth-death models (BDM) that take into account shared ancestry and phylogenetic uncertainty. The constant-rate BDM (Model 1) suggested that the number of secondary infections was on average marginally above one over the course of the epidemic (Fig. 3, A and C). For example, assuming that an HIV-infected person remains infectious for a period of 2 years, after which this person is either virally suppressed or removed from a population (relocation or death), the median posterior estimate of R_e was 1.15 (95% Bayesian credible interval, BCI: 1.09 to 1.2) for subtype B and 1.23 (95% BCI: 1.2 to 1.27) for subtype G (Fig. 3 A and C). A sensitivity analysis specifying various infectious periods (6 months to 5 years, removal rate 0.2-2.0) shows that R_e estimates decrease with shorter infectious periods, but remain between 1 and 2 for all the specified values (Fig. 3, A and C).

When using a BDSKY model (Model 2), which allows changes in the rate of transmission through time, we find that subtype B's R_e reached its peak in or around 2001 (R_e = 2.17, 95% BCI 1.53 to 2.82), after which R_e dropped below 1 after 2004 (95% BCI 2002.5-2008) (Fig. 3B). For subtype G, the maximum value of R_e was observed in 1996 (R_e = 1.6, 95% BCI 0.98 to 2.2) (Fig. 3D). The mean R_e of subtype G dropped below 1 from 2000 onwards (R_e = 0.51, 95% BCI 0.29 to 0.75).

Finally, the two-epoch BDM (Model 3), revealed for both subtypes an initial phase of epidemic growth characterised by high R_e values (well above R_e=1), followed by a transition to a later phase of slower transmission. For subtype B, around the year 1999 (95% BCI 1998 to 2000) R_e decreased from a median estimate of 1.92 (95% BCI 1.75 to 2.1) to 0.62 (95% BCI 0.52 to 0.72) (Fig. 4). For subtype G, around the year 2000 (95% BCI 1998 to 2001), R_e decreased from a median estimate of 1.49 (95% BCI 1.39 to 1.59) to 0.71 (95% BCI 0.63 to 0.8) (Fig 4). Notably, Model 3 indicates that the major transitions in the transmission dynamics of subtypes B and G occurred around the same point in time.

To investigate whether the same transmission dynamics patterns were observed in PWID populations, we repeated the subtype G analysis including only sequences from PWID group. This procedure resulted in similar R_e patterns for Models 2 and 3 as for the full subtype G dataset (Suppl. Fig. 1).

Inferring mixture of transmission behavior

To quantify the level of phylogenetic clustering by transmission behavior we used three distinct approaches to measure the association index (AI) (Table 2), and two different discretization schemes, in which sequences were assigned to 5 or 3 different transmission risk groups. For the 5-group discretization scheme, BaTS indicated no significant transmission group structure in the subtype B clades, in contrast to the significant structure detected for subtype G (p-value <0.01). The AI statistic suggests evidence of compartmentalization in subtype B clade 5 (AI=0.77, bootstrap support 0.97) and in the Subtype G clade (AI=0.84, bootstrap support 0.98). Similar patterns for both subtypes were observed for the 3-transmission group discretization scheme.

We next performed discrete trait diffusion analyses (DTA) for subtype B clade 5 and for the large subtype G clade. The DTA results for subtype B clade 5 suggest that HET individuals disproportionally contributed to viral dissemination among transmission groups, as this group represented the inferred source of 82% of viral transmission group transition events between the transmission groups (N=17, 95% BCI 15-19 out of a total N=21, 95% BCI 19-24), while representing only 45% of the sequences (Fig. 5, Table 1). We next used a Bayesian stochastic search variable selection approach (BSSVS) to identify the statistical support for virus movement among transmission risk groups (37). We find that transmission group transitions with the highest support were: HET to MSM, HET to PWID, HET to MTC, and MSM to OTH (BF = 976, 136, 48, and 22, respectively) (Fig. 5). When only 3 transmission groups were considered (HET, MSM, or "Other"), the estimated proportion of

transmission group transition events from HET to other transmission groups was 90% (N=16, 95% BCI: 14-19, out of total N=18, 95% BCI: 15-21) and the pathways with the highest BF support were HET to MSM and HET to OTH (BF=1105 and 550, respectively).

Discrete diffusion analyses of subtype G suggest that PWID populations were the main contributor to viral transitions among transmission groups, accounting for 81% of transmission group transitions (N=82, 95% BCI: 60-96, out of total N=101, 95% BCI: 90-110), while representing 47% of the sequences (Fig. 5, Table 1). HET accounted for 17% (N=18, 95% BCI: 5-32) of transmission group transitions in the same dataset while comprising 35% of the sequences. All other groups accounted for <1% of transmission group transition events. Pathways with the highest support were: PWID to HET, PWID to MTC, and HET to MTC, with BF = 29353, 29353, and 116, respectively. Again, the same patterns were observed when only PWID, HET, and "Other" were considered.

Discussion

We analysed HIV-1 genetic sequences from Portugal to better understand the dynamics of subtype B and G transmission in the country. Remarkably, our analyses of genetic data reveal that the rate of epidemic spread dropped significantly between 1998-2001. Moreover, we found little evidence for phylogenetic clustering by risk behaviour, suggesting frequent mixing among transmission risk groups.

The estimated effective reproductive number of the Portuguese HIV-1 epidemic was on average above the epidemiological threshold of R_e=1 throughout the history of subtype B and G epidemics in the country. When we estimated changes in R_e over time, we found that the growth rates of both subtypes were well above 1 for more than 10 years. R_e started to decline around 2001 for subtype B and 1996 for subtype G. Importantly, the timing of this decline coincided with several major public health interventions around the mid- to late 1990s, in particular the introduction of highly active antiretroviral treatment (HAART) in 1996. Interestingly, the estimated reproductive number of subtype G, which is dominated by transmissions in PWID, started to drop long before that of subtype B. This might be explained by a combined effect of HAART and harm reduction programs introduced at the same time, as both SEPs and OST have been shown to slow down PWID-driven epidemics (39). Furthermore, drug decriminalization in 2001 had a major effect on drug-related mortality in Portugal, with the number of cases decreasing from 80 to 16 between 2001-2012 (40). Finally, the timing of a major decline in viral transmission dynamics for both subtypes, with a particularly rapid decrease in R_e for the PWID-driven subtype G, compared to subtype B, coincided with the decriminalization of drugs (Fig. 3). This estimated decline in epidemic growth also coincides with a decline in the annual number of new HIV cases (5).

We used a novel hierarchical approach, implemented in a Bayesian framework, which allowed us to combine information across five different smaller phylogenetic trees representing five subtype B clades, and estimate epidemiological parameters for this subtype

without forcing all clades to belong to one phylogenetic tree. If all five clades were combined into one large phylogeny, then estimation of the deep branches in the tree (i.e. the among-clade branches connecting the five clades) would add to statistical uncertainty, in part due to sequence saturation. Furthermore, the approach we used here allowed us to reliably estimate R_e even for the years prior to the sampling date of the oldest sequence in the analysis. This is an important advantage compared to methods that are based solely on the sequences' sampling proportion, which might overestimate R_e for the year of the oldest samples by assuming that all cases diagnosed that year were infected the same year (15).

For subtype B specifically, it is important to remember that the findings of this study describe transmission dynamics of the main circulating subtype B clades, but do not necessarily represent the Portuguese HIV-1 subtype B population as a whole. Even though the proportion of sequences from different transmission groups in our analysis is similar to that reported for the Portuguese epidemic before[12] (see Suppl. Table 2), we cannot be sure that the transmission dynamic parameters inferred for the large clades would be identical for smaller clades.

We find phylogenetic evidence of frequent transmissions among different transmission groups, suggesting that interventions need to target epidemiologically linked groups simultaneously. Mathematical modelling suggests that halting new infections within "bridging" transmission groups may have the greatest impact in further reducing HIV incidence (41). Historically, subtype B has been associated mostly with sexual HIV transmission in Portugal, while subtype G and CRF14_BG are more frequently reported in PWID (42). Our results support the idea that PWID disproportionately contributed to the subtype G epidemic. After becoming established in PWID populations, subtype G spread to the more generalized HET population. For subtype B, no transmission group structure was observed for most of the clades, with only subtype B clade 5 showing a particularly high viral flow from HET to MSM. Notably, in our analysis, sequences from MSM grouped closely with sequences from HET and PWID, similarly to the pattern previously reported from Minho

province in the north of Portugal (7). In other European countries and the United States, however, sequences sampled from MSM tend to form isolated clades that are poorly mixed with other transmission groups (43-47). Because we rely on self-reported transmission risk group information, it is however possible that the virus lineage movement between HET and MSM results, at least in part, from the misclassification of HET infections and underreporting of MSM contacts, as previously shown in the UK and Nordic countries (48, 49).

Portugal has a well-developed surveillance system that enables timely identification of new HIV infections, large-scale ART provision, and routine antiretroviral drug resistance testing, which forms the basis of a large national database of protease and reverse transcriptase (PR/RT) region sequences. Unfortunately, the majority of the sequences from the HIV Portuguese database are unlinked to additional epidemiological information such as transmission risk group, stage of infection, etc. Analysis of between- and within-risk group transmission is critical to the design of optimal strategies that can directly impact the progress of an HIV epidemic in a country (44). Improving the linkage between patient records and sequence data would allow larger scale studies to be undertaken that could directly impact public health strategies aimed at reducing HIV transmission in real-time, similarly to what have recently been done in Canada (50).

In conclusion, our study shows how genetic data can help us to address challenging epidemiological questions and support evidence for public health decision-making. Further improvements both in statistical approaches and in genomic surveillance in health care settings can facilitate and encourage the application of phylogenetics to eliminate new infections among distinct transmission groups.

Acknowledgements

We would like to thank all the patients for the data provided. We thank the European Society for translational Antiretroviral Research for providing us with genetic sequence data for this study. We also thank Anne Cori of the Imperial College London and Christophe Fraser of the University of Oxford for early discussions. Additional information: Co-author Ricardo J Camacho passed away on July 4th, 2018.

References

1. Abecasis AB, Wensing AM, Paraskevis D, Vercauteren J, Theys K, Van de Vijver DA, et al. HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. Retrovirology. 2013;10:7.

2. Esteves A, Parreira R, Venenno T, Franco M, Piedade J, Germano De Sousa J, et al. Molecular epidemiology of HIV type 1 infection in Portugal: high prevalence of non-B subtypes. AIDS Res Hum Retroviruses. 2002;18(5):313-25.

3. Amaral JA, Pereira EP, Paixao MT. Data and projections of HIV/AIDS cases in Portugal: An unstoppable epidemic? J Appl Stat. 2005;32(2):127-40.

4. European Centre for Disease Prevention and Control WROfE. HIV/AIDS surveillance in Europe: Surveillance report. Stockholm: ECDC; 2015.

5. INSA DdDaId. Infeção VIH SIDA: a situação em Portugal a 31 de dezembro de 2014. Lisboa: Instituto Nacional de Saúde Doutor Ricardo Jorge; 2015.

6. Esteves A, Parreira R, Piedade J, Venenno T, Franco M, de Sousa JG, et al. Spreading of HIV-1 subtype G and envB/gagG recombinant strains among injecting drug users in Lisbon, Portugal. Aids Res Hum Retrov. 2003;19(6):511-7.

7. Carvalho A, Costa P, Triunfante V, Branca F, Rodrigues F, Santos CL, et al. Analysis of a local HIV-1 epidemic in portugal highlights established transmission of non-B and non-G subtypes. J Clin Microbiol. 2015;53(5):1506-14.

8. Bartolo I, Abecasis AB, Borrego P, Barroso H, McCutchan F, Gomes P, et al. Origin and epidemiological history of HIV-1 CRF14_BG. PLoS One. 2011;6(9):e24130.

9. Palma AC, Araujo F, Duque V, Borges F, Paixao MT, Camacho R, et al. Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. Infect Genet Evol. 2007;7(3):391-8.

10. MacArthur GJ, van Velzen E, Palmateer N, Kimber J, Pharris A, Hope V, et al. Interventions to prevent HIV and Hepatitis C in people who inject drugs: A review of reviews to assess evidence of effectiveness. Int J Drug Policy. 2014;25(1):34-52.

11. Drug Treatment Profiles: Portugal [Internet]. 2019. Available from: http://www.emcdda.europa.eu/html.cfm/index35987EN.html.

12. Loo MV, Van Beusekom I, Kahan JP. Decriminalization of drug use in Portugal: The development of a policy. Ann Am Acad Polit Ss. 2002;582:49-63.

13. Stadler T, Kouyos R, von Wyl V, Yerly S, Boni J, Burgisser P, et al. Estimating the basic reproductive number from viral sequence data. Mol Biol Evol. 2012;29(1):347-57.

14. Vasylyeva TI, Friedman SR, Lourenco J, Gupta S, Hatzakis A, Pybus OG, et al. Reducing HIV infection in people who inject drugs is impossible without targeting recently-infected subjects. AIDS. 2016;30(18):2885-90.

15. Bezemer D, Cori A, Ratmann O, van Sighem A, Hermanides HS, Dutilh BE, et al. Dispersion of the HIV-1 Epidemic in Men Who Have Sex with Men in the Netherlands: A Combined Mathematical Model and Phylogenetic Analysis. PLoS Med. 2015;12(11):e1001898; discussion e.

16. Hofstra LM, Sauvageot N, Albert J, Alexiev I, Garcia F, Struck D, et al. Transmission of HIV Drug Resistance and the Predicted Effect on Current First-line Regimens in Europe. Clin Infect Dis. 2016;62(5):655-63.

17. Kuiken C, Korber B, Shafer RW. HIV sequence databases. AIDS Rev. 2003;5(1):52-61.

18. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539.

19. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870-4.

20. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26(7):1641-50.

21. Ragonnet-Cronin M, Hodcroft E, Hue S, Fearnhill E, Delpech V, Brown AJ, et al. Automated analysis of phylogenetic clusters. BMC Bioinformatics. 2013;14:317.

22. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3.

23. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016;2(1):vew007.

24. Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. Plos Comput Biol. 2014;10(4):e1003537.

25. Boskova V, Stadler T, Magnus C. The influence of phylodynamic model specifications on parameter estimates of the Zika virus epidemic. Virus Evol. 2018;4(1):vex044.

26. Stadler T, Kuhnert D, Bonhoeffer S, Drummond AJ. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc Natl Acad Sci U S A. 2013;110(1):228-33.

27. Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, et al. Rise and fall of the Beringian steppe bison. Science. 2004;306(5701):1561-5.

28. Edo-Matas D, Lemey P, Tom JA, Serna-Bolea C, van den Blink AE, van 't Wout AB, et al. Impact of CCR5delta32 host genetic background and disease progression on HIV-1 intrahost evolutionary processes: efficient hypothesis testing through hierarchical phylogenetic models. Mol Biol Evol. 2011;28(5):1605-16.

29. Kuhnert D, Kouyos R, Shirreff G, Pecerska J, Scherrer AU, Bonl J, et al. Quantifying the fitness cost of HIV-1 drug resistance mutations through phylodynamics. Plos Pathogens. 2018;14(2).

30. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Systematic Biology. 2018;67(5):901-4.

31. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005;21(5):676-9.

32. Parker J, Rambaut A, Pybus OG. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. Infect Genet Evol. 2008;8(3):239-46.

33. Wang TH, Donaldson YK, Brettle RP, Bell JE, Simmonds P. Identification of shared populations of human immunodeficiency virus type 1 infecting microglia and tissue macrophages outside the central nervous system. J Virol. 2001;75(23):11686-99.

34. Bezemer D, Faria NR, Hassan A, Hamers RL, Mutua G, Anzala O, et al. HIV Type 1 Transmission Networks Among Men Having Sex with Men and Heterosexuals in Kenya. Aids Res Hum Retrov. 2014;30(2):118-26.

35. Minin VN, Suchard MA. Counting labeled transitions in continuous-time Markov models of evolution. J Math Biol. 2008;56(3):391-412.

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

37. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian Phylogeography Finds Its Roots. Plos Comput Biol. 2009;5(9).

38. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. Mol Biol Evol. 2016;33(8):2167-9.

39. Fernandes RM, Cary M, Duarte G, Jesus G, Alarcao J, Torre C, et al. Effectiveness of needle and syringe Programmes in people who inject drugs - An overview of systematic reviews. BMC Public Health. 2017;17(1):309.

40. Transford: Getting drugs under control [Internet]2014. Available from: http://www.tdpf.org.uk/blog/drug-decriminalisation-portugal-setting-record-straight.

41. Gupta S, Anderson RM, May RM. Networks of sexual contacts: implications for the pattern of spread of HIV. AIDS. 1989;3(12):807-17.

42. Esteves A, Parreira R, Piedade J, Venenno T, Franco M, Germano de Sousa J, et al. Spreading of HIV-1 subtype G and envB/gagG recombinant strains among injecting drug users in Lisbon, Portugal. AIDS Res Hum Retroviruses. 2003;19(6):511-7.

43. Patino-Galindo JA, Torres-Puente M, Bracho MA, Alastrue I, Juan A, Navarro D, et al. Identification of a large, fast-expanding HIV-1 subtype B transmission cluster among MSM in Valencia, Spain. PLoS One. 2017;12(2):e0171062.

44. Chaillon A, Essat A, Frange P, Smith DM, Delaugerre C, Barin F, et al. Spatiotemporal dynamics of HIV-1 transmission in France (1999-2014) and impact of targeted prevention strategies. Retrovirology. 2017;14(1):15.

45. Parczewski M, Leszczyszyn-Pynka M, Witak-Jedra M, Szetela B, Gasiorowski J, Knysz B, et al. Expanding HIV-1 subtype B transmission networks among men who have sex with men in Poland. PLoS One. 2017;12(2):e0172473.

46. Chalmet K, Staelens D, Blot S, Dinakis S, Pelgrom J, Plum J, et al. Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. BMC Infect Dis. 2010;10:262.

47. Chan PA, Hogan JW, Huang A, DeLong A, Salemi M, Mayer KH, et al. Phylogenetic Investigation of a Statewide HIV-1 Epidemic Reveals Ongoing and Active Transmission Networks Among Men Who Have Sex With Men. J Acquir Immune Defic Syndr. 2015;70(4):428-35.

48. Hue S, Brown AE, Ragonnet-Cronin M, Lycett SJ, Dunn DT, Fearnhill E, et al. Phylogenetic analyses reveal HIV-1 infections between men misclassified as heterosexual transmissions. AIDS. 2014;28(13):1967-75.

49. Esbjornsson J, Mild M, Audelin A, Fonager J, Skar H, Bruun Jorgensen L, et al. HIV-1 transmission between MSM and heterosexuals, and increasing proportions of circulating recombinant forms in the Nordic Countries. Virus Evol. 2016;2(1):vew010.

50. Poon AF, 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.

Fig 1. Diagram describing the process of selecting subtype B and G sequences for the analysis.

Fig 2. Top: Maximum Likelihood (ML) phylogenetic trees reconstructed from the combined datasets: subtype B (left) and subtype G (right). The black colour indicates Portuguese sequences and is used to highlight their position on the global diversity trees. The highlighted clades are the clades of at least ten sequences with a within-clade genetic distance <6% and SH-like statistical support >90%. Bottom: ML phylogenetic trees of the identified phylogenetic clades. The colours of the tips represent the transmission group of a patient from whom the sequence was sampled.

Fig 3. Panels A and B represent the R_e estimated using the BD Model 1 (constant R_e) under different assumptions of infectious period (6 months – 5 years, becoming uninfectious rate 2.0 – 0.2) for subtypes B (panel A) and G (panel B). The vertical lines in the middle of the violin plots represent the interquartile range. The filled with colour violin plots represent the value of R_e assuming the infectious period of 2 years. The horizontal red dotted line represents the epidemiological threshold (R_e=1). Panels C and D represent the R_e estimates using the BD Model 2 (BDSKY) over 10 equidistant time-intervals between the tree height and the most recent tip in the analysis for subtype B (panel C) and G (panel D), assuming the infectious period of 2 years. The shaded area represents Bayesian credible interval. The horizontal red dotted line represents the epidemiological threshold (R_e=1). Vertical lines correspond to the time of introduction of major HIV preventive interventions.

Fig 4. R_e estimates obtained using the BD Model 3 (2 time-intervals) and the estimated time of change in the rate of epidemic spread for both subtypes. The violin plots of R_e distribution before and after the time of estimated change in viral spread rate are illustrated for both subtypes. The shaded area represents Bayesian credible interval. The horizontal red dotted line represents the epidemiological threshold (R_e=1).

Fig 5. Results of the discrete trait analysis (DTA) for both subtypes. Top: Molecular clock phylogenetic trees reconstructed for the subtype B clade 5 (left) and the subtype G clade (right). The colour of the branches corresponds to the transmission group state; the width of the branches corresponds to the posterior probability support for the indicated transmission group state. Bottom: Results of the DTA for both subtypes considering 5 and 3 transmission group (k=5 and k=3). The colours represent the transmission group of a patient from whom the sequence was sampled.

Bottom panels left to right: 1) Proportion of migration events attributed to a transmission group reconstructed using robust counts approach (Ns correspond to the number of sequences in the transmission group);

2) The rescaled Association index (AI);

3) Bayes Factor (BF) values for the significant virus migration pathways; the size of the circles corresponds to the number of sequences in the analysis; the width of the arrows corresponds to the number of migration events for the corresponding pathways.

Figure 1

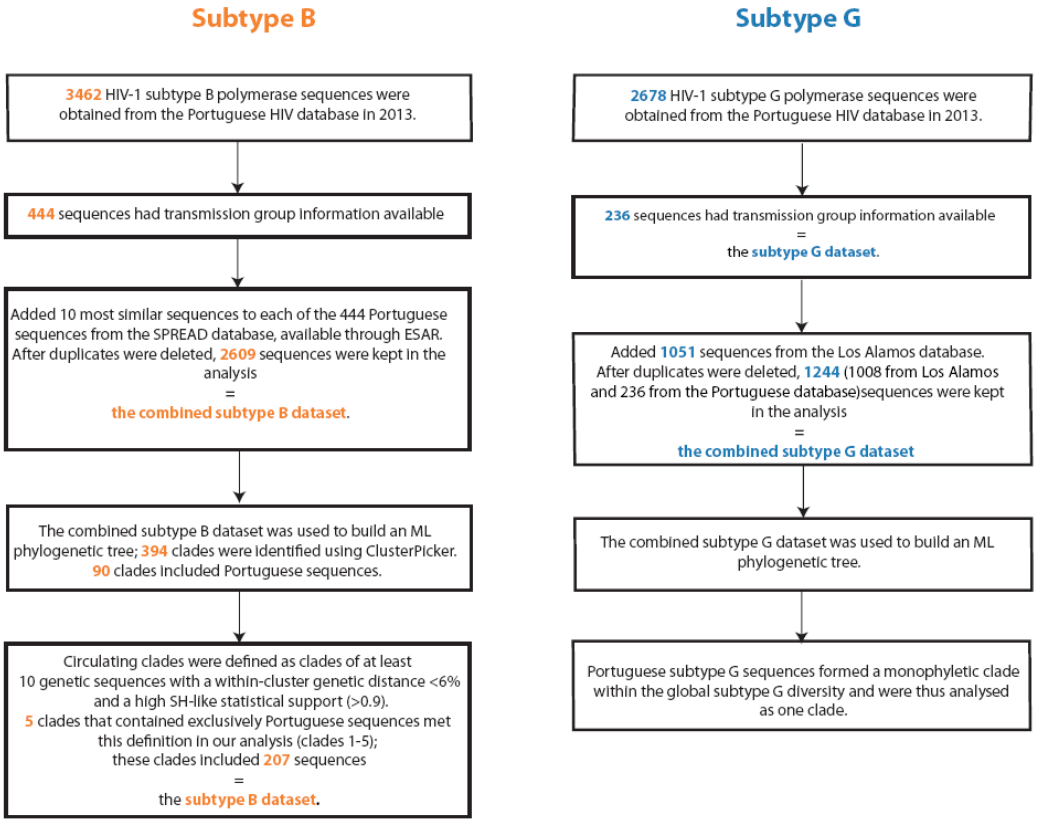

Figure 2

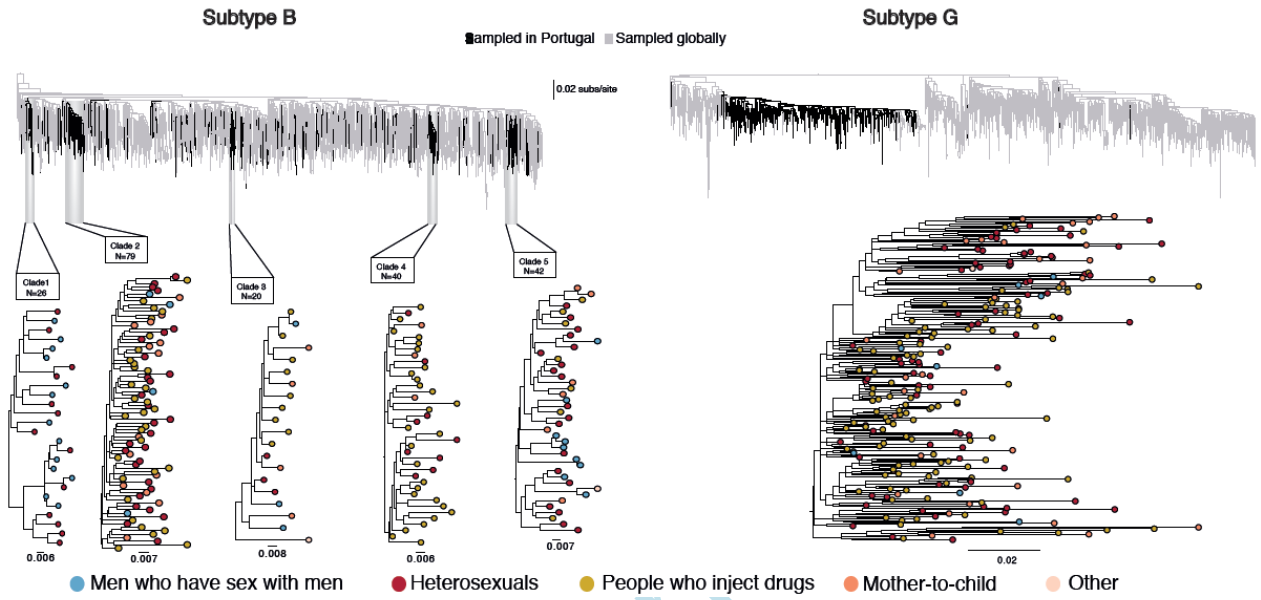

Figure 3

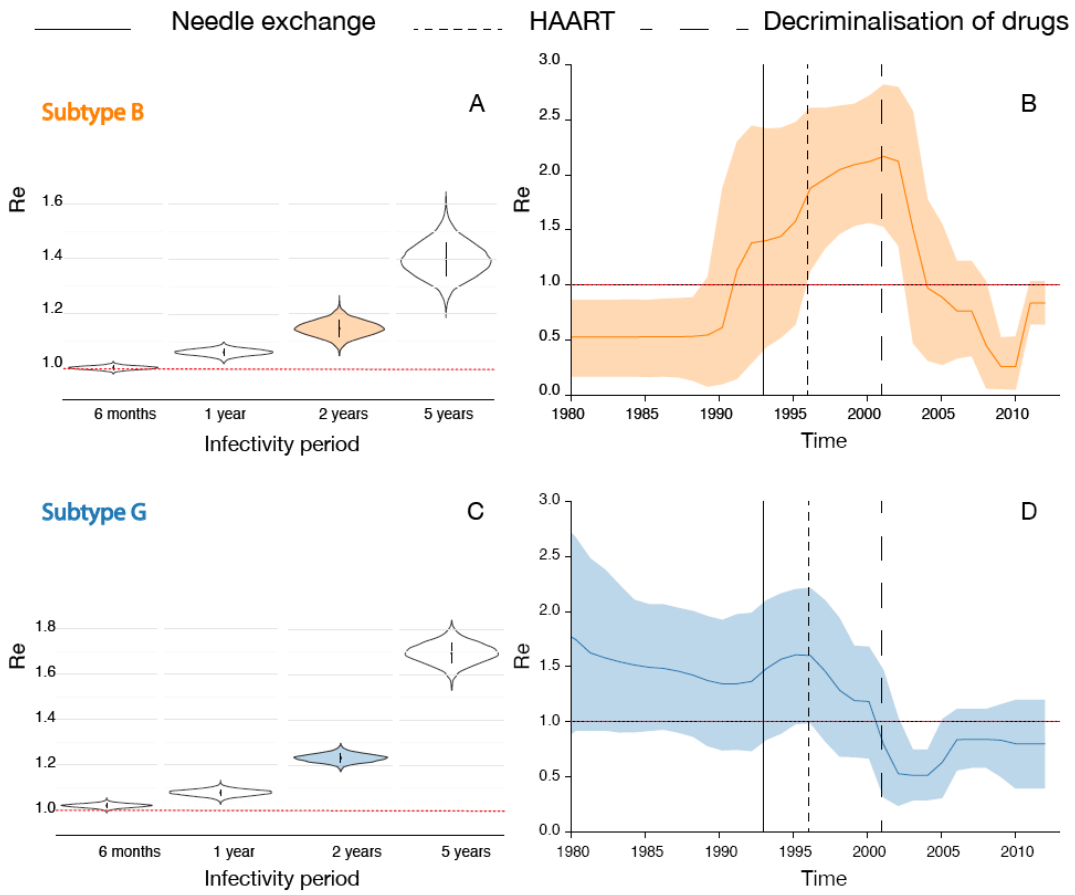

Figure 4

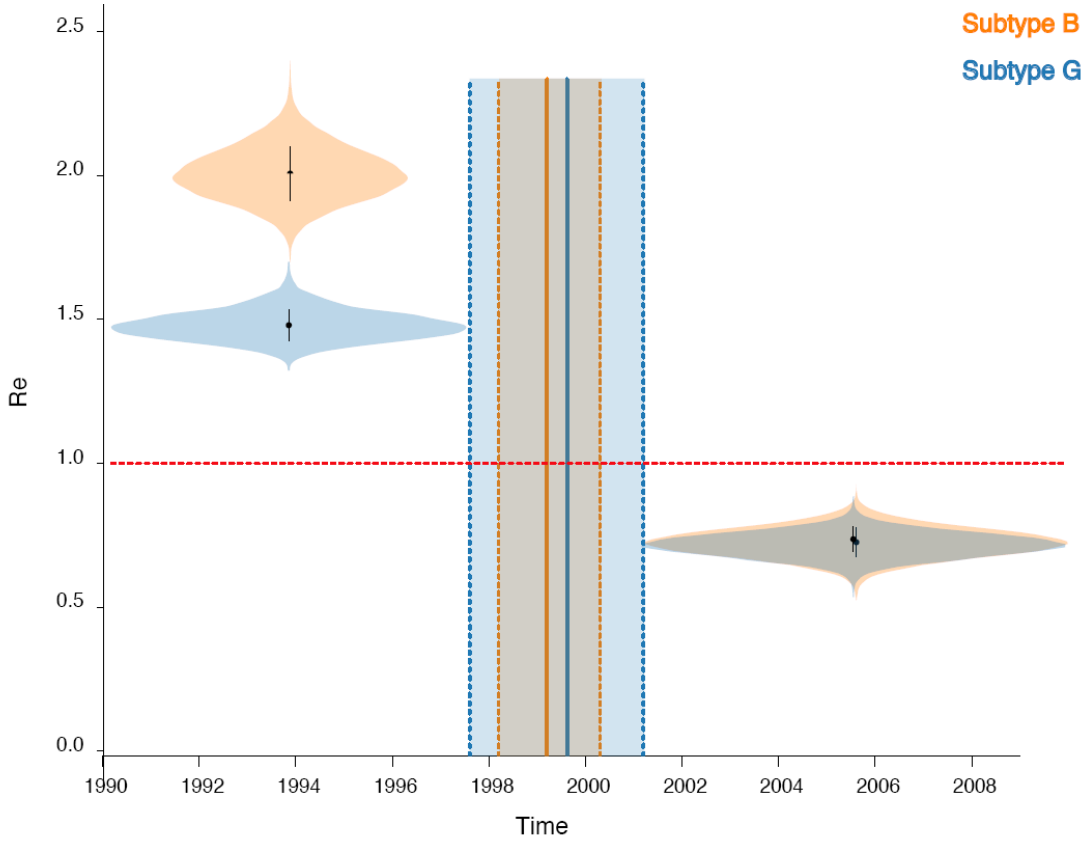

Figure 5