Spatio-Temporal Dynamics of HIV-I Subtype B and Circulating Recombinant Form 01_AE Clades in Iran: A Phylogeographic Approach

Sana Eybpoosh,[1] Abbas Bahrampour,[2] Kayhan Azadmanesh,[3] Ehsan Mostafavi,[4,5] Ali Akbar Haghdoost,[6,*] and Farzaneh Zolala[2]

[1]Regional Knowledge Hub, and WHO Collaborating Centre for HIV/AIDS Surveillance, Institute for Future Studies in Health, Kerman University of Medical Sciences, Kerman, IR Iran
[2]Faculty of public health, Department of Epidemiology and biostatistics, Kerman University of Medical Sciences, Kerman, IR Iran
[3]Department of Virology, Pasteur Institute of Iran, Tehran, IR Iran
[4]Department of Epidemiology, Pasteur Institute of Iran, Tehran, IR Iran
[5]Research Centre for Emerging and Reemerging Infectious Diseases, Pasteur Institute of Iran, Hamadan, IR Iran
[6]Research Centre for Modeling in Health, Institute for Future Studies in Health, Kerman University of Medical Sciences, Kerman, IR Iran

[*]*Corresponding author*: Ali Akbar Haghdoost, Research Centre for Modeling in Health, Institute for Future Studies in Health, Kerman University of Medical Sciences, Haft Bagh Ave., Kerman, IR Iran. Tel/Fax: +98-3413205405, E-mail: ahaghdoost@kmu.ac.ir

Abstract

Background: The origins and spatio-temporal dynamics of the HIV-1 subtype B and CRF01_AE clades are largely unknown in Iran.
Objectives: This study investigates the origins and spatio-temporal dynamics of HIV-1 subtype B and CRF01_AE clades in Iran using the most recent phylogeographic methods.
Methods: This Bayesian phylogeographic study examined HIV-1 sequence data available in the Los Alamos HIV databases. For the purposes of this study, we retrieved all the eligible time-stamped and geo-referenced HIV-1 subtype B (pol p66 and p51 RT) and CRF01_AE (gp160) sequences from Iran. The majority of these data were obtained from studies using convenience and snowball sampling. Evolutionary and phylogeographic parameters were jointly estimated for each clade using BEAST software v.1.8.1. A Bayesian stochastic search variable selection method, with a threshold of 3.0 for the Bayes factors test, was used to obtain significant migration rates of the virus.
Results: Our results suggest that the subtype B epidemic was introduced to Iran from Southern China, Germany and the USA. Although the migration rate from the USA to Iran was found to be statistically insignificant, the results suggested that the USA was the earliest exporter of the subtype B to Iran (i.e. since 1974, 95% highest posterior density (HPD): 1966-1974). Moreover, based on our findings, the CRF01_AE clade was first introduced to Iran from Thailand (1988, 95% HPD: 1986 - 1990). Among the factors contributing to the transmission of subtype B to Iran from Germany is the transit of goods and long-distance truck driving. However, the factors behind the transmission of this clade from the USA and Southern China remain unknown. Similar to many countries, the introduction of CRF01_AE from Thailand might be attributed to sex tourism.
Conclusions: There is evidence to suggest that subtype B and CRF01_AE were introduced to Iran through multiple pathways and origins, including Germany, Southern China, the USA, and Thailand. Factors such as international trade, the transit of goods, and tourism are among the most likely methods for transmission of these clades. However, the existing preventive strategies in Iran have been mostly focused on other risk groups, including sex workers and intravenous drug users. The findings of our study highlight the need for the further development of targeted preventive strategies, such as educational interventions, aimed at tourists, truck drivers, and people involved in international trade and transactions in order to prevent the onward transmission of disease.

Keywords: HIV-1, Phylogenetic Biogeography, Iran

1. Background

Pandemic HIV-1 group M infection in humans originated in Africa and was caused by the cross-species transmission of the simian immunodeficiency virus that infects chimpanzees (SIVcpz) (1). Among the subtypes/recombinant forms derived from group M, subtype B and circulating recombinant form (CRF) 01_AE are predominant clades in most parts of the world (2). Subtype B originated in Haiti and spread to the USA around the 1960s (3). CRF01_AE originated in Central Africa and spread to Thailand during the 1970s. The CRF01_AE epidemic in Thailand was first recognized as affecting female sex workers in 1989 (2, 4). Subsequently, CRF01_AE caused a large-scale epidemic in Southeast Asia, with a prevalence of 79%. The worldwide spread of CRF01_AE began immediately after this and was most likely transmitted through heterosexual activities (2). In Iran, both subtype B and CRF01_AE

have been detected (5, 6), however, their epidemic history (i.e. the source and the date of origin) is still unknown.

Typical epidemiological methods have not proved successful in identifying the roots of infectious disease epidemics, including the infection source, the date of origin, and the transition pathways of the pathogen (7). However, recent advances in molecular biology and biostatistics have made it possible to track the pathways of virus diffusion, over place and time, using the information accumulated in the virus' genome. Such methods are commonly referred to as phylogeography (8). Phylogeography-based methods reconstruct phylogenetic trees (8-11) to infer evolutionary (i.e. ancestor-descendent) relationships between living species, based upon the similarities/differences in their physical and/or genetic traits (12). Interestingly, the evolutionary relationships between viruses isolated from a particular epidemic can be inferred as viral "transmission" events within that epidemic. By integrating this in-

formation with sampling sites of viral isolates and with the virus' evolutionary rate (i.e. its molecular clock), one can estimate both "spatial" and "temporal" aspects of the epidemic using a Bayesian phylogeographic framework. The important aspects of the spatio-temporal process include: (i) the time and location of the most recent common ancestor (MRCA) or the "founder" virus; (ii) locations of ancestral lineages (viruses) descended from the MRCA; (iii) the time at which each ancestor-to-descendant transmission occurred; and (iv) the pathway of virus migration (8, 11).

Since the introduction of Bayesian phylogeographic methods in 2009 (8), the number of studies investigating the spatio-temporal spread of viral infections has increased dramatically. For example, Lemey et al. (2009) employed this method to investigate the global spread of influenza A (H5N1) and the rabies virus (8). Several studies have also estimated the spatio-temporal spread of some HIV-1 clades in Central (13) Eastern (14) and Western (15) Africa, Europe (16), America (3), China (17), Vietnam (18), and Singapore (19). However, in Iran, the spatio-temporal domains of the HIV-1 epidemic have not been investigated. Increasing our knowledge of the factors that led the virus to transcend Iran's national borders can inform the design of medical interventions to prevent onward transmissions (8, 20).

2. Objectives

This study investigates the epidemic roots of HIV-1 subtype B and CRF01_AE clades in Iran, including the epidemic source(s), the date of emergence of these clades in the country, and their spatio-temporal dynamics. We have also investigated the factors that have possibly contributed to the initial introduction of these clades to the country.

3. Methods

In this study, we take a Bayesian phylogeographic approach to analyze a set of HIV-1 sequences available in the Los Alamos HIV (LAHIV) databases (21).

3.1. HIV-1 Subtype B and CRF01_AE Data Sets

In order to investigate the spatio-temporal dynamics of subtype B and CRF01_AE in Iran, HIV-1 DNA sequences from Iran, as well as other countries from around the world, were required. However, it was not feasible to collect a global sample hoping that all would be subtype B and CRF01_AE. Therefore, the data was retrieved from the LAHIV sequence database where the strain of each sample was already defined (see Box 1 for a brief description of the database) (21). Convenience and snowball sampling were

the most commonly used sampling methods in the majority of the primary studies whose data were submitted to the LAHIV database and were used in our study.

The selection of all subtype B and CRF01_AE sequences covering partial pol (HXB2 numbering: 2619 - 3219 nts, 600 bp) and env (gp160HXB2 numbering: 6958 - 7376 nts, 418 bp) regions resulted in two data sets, each consisting of more than 1000 sequences. In order to reduce the size of these data sets and decrease the computational burden, we followed a filtering procedure to identify and discard the "redundant" sequences from non-neighboring countries. In this regard, sequences with high genetic similarity (i.e. genetic distance < 5% (13) from each non-neighboring country were clustered using the Cluster Database at High Identity with Tolerance (CD-HIT) program (24), accessible via an online web server (25), and only one sequence per cluster was retained in each data set. Since the number of sequences from Iran and its neighboring countries was small, we did not reduce the number of these data.

HIV-1 sequences were selected according to the following inclusion criteria: (i) only HIV-1 sequences generated by a polymerase chain reaction (PCR) technique and the standard sequencing methods (i.e. pyrosequencing or Sanger sequencing) were included. This was to ensure the inclusion of high quality data and to minimize between-laboratory and between-method variability (7); (ii) only one sequence per patient was selected randomly; (iii) only the viral isolates, whose sampling dates and locations were known, were included. It is worth noting that these are the most important variables in reconstructing spatio-temporal processes; (iv) only the pol sequences from therapy-naive subjects and/or those with no major drug resistance mutations were selected. This was to prevent convergent evolution due to drug therapy (26); and (v) only the sequences with no frameshift and/or stop-codon mutations were included. Applying these criteria, a total of 37 sequences were excluded from the initially retrieved data (26 and 11 sequences from subtype B and CRF01_AE data sets, respectively). The resulting data sets were used for phylogenetic analysis (see methods) in order to find sequences that were closely related to Iranian lineages. These sequences were retained in the final data sets for subsequent phylogeographic analyses.

3.2. Statistical Analyses

3.2.1. Likelihood Mapping Analysis

We employed this analysis in order to assess whether subtype B and CRF01_AE data sets contain enough information for phylogenetic analysis. If the data sets lack adequate information, the evolutionary relationships cannot be fully resolved in the phylogenetic tree. In such a case, more than 30% of trees will be "unresolved" (see (27, 28) for

Box 1. A Brief Description of the Los Alamos HIV Databases (21-23)

Los Alamos HIV Description
The Los Alamos HIV databases represent an international collaboration dedicated to covering HIV molecular data and include: (i) an HIV genome sequence database; (ii) HIV molecular immunology data; (iii) HIV drug resistance mutations; and (iv) nonhuman primate HIV vaccine trials (22). The Los Alamos HIV genome sequence databases (LAHIV) are population-based online databases that were created to manage the enormous amount of genetic data generated by studies of HIV variation. The databases are defined as stored collections of genetic samples in the form of DNA sequences, and as of December 2015, they have indexed more than 590,000 DNA sequences (21).
The LAHIV databases are fed biweekly with HIV sequences from the GenBank database that contains most published HIV sequences (23). Upon receipt of a sequence submission, the GenBank staff examines the originality of the data and assigns an accession number to the sequence and performs quality assurance checks using standard quality control guidelines for genetic data (23). The submissions are then released to the public database. Upon updating data from the GenBank, the LAHIV staff annotates the sequences with supplemental information, not included in the GenBank, from the literature and from direct communication with the authors.
Therefore, the DNA sequences in the LAHIV databases are annotated with extra information, including medical (e.g. CD4 counts, disease stage), temporal (e.g. sampling date), geographical (e.g. sampling country) and lifestyle (e.g. risky behaviors, infection route) characteristics. With their advanced search and filtering options, the LAHIV databases allow for the retrieval of information on specific genetic regions and strain-types of the virus. It is also possible to retrieve HIV sequences belonging to a particular sampling date, sampling location, or risk group (such as heterosexuals or people who inject drugs [PWID]). Furthermore, the databases allow for excluding problematic sequences and multiple sequences per patient (21, 22). These features have made the LAHIV databases suitable for conducting various types of sequence analyses, including phylogeographic studies (for a comprehensive review of the databases, see (21-23)).

more methodological details). For subtype B and CRF01_-AE data sets used in this study, these values were 27.3% and 20.8%, respectively, indicating the adequacy of the data.

3.2.2. Phylogenetic Analysis

In order to reduce the computational burden of phylogeographic analyses, we selected sequences that were significantly related to Iranian lineages using the maximum likelihood (ML) phylogenetic technique. ML trees were reconstructed under the general time-reversible (GTR) model of DNA evolution (29). An heuristic tree search was performed by employing a subtree-pruning-regrafting (SPR) branch-swapping algorithm. The SPR is a tree rearrangement algorithm that searches for the optimal tree topology (i.e. branching pattern of the phylogenetic tree (30)). This search is performed by selecting and removing a subtree from the main tree and reinserting it elsewhere to create a new node (27). Reliability of the tree topology was estimated using a bootstrap resampling technique, with 1000 replicates. Only the countries that were clustered with Iranian lineages in the ML tree, with bootstrap values greater than 50% (31), were selected for subsequent phylogeographic analyses (Figure 1).

3.2.3. Bayesian Phylogeographic Analysis

Spatio-temporal diffusion patterns of subtype B and CRF01_AE were estimated using a Bayesian Markov Chain Monte Carlo (MCMC) approach (8). Analyses were performed employing the Hasegawa-Kishino-Yano (HKY) model of nucleotide substitutions (32) while considering a proportion of nucleotide sites to be invariable (i.e. experiencing no mutations). We assumed that the rate of nucleotide substitutions across the genome varies under a discrete gamma distribution with four discrete categories. For both the subtype B and CRF01_AE data sets, the best fitted clock model was found to be the uncorrelated lognormal relaxed-clock (UCLD). This model allows the evolutionary rate to vary independently throughout the tree, with a lognormal distribution (30).

Migration events were reconstructed using a discrete phylogeography model that considers each country to be a discrete variable and with a reversible transition matrix that assumes equal probability for reciprocal migration rates of the virus between two countries. The significant migration rates were obtained using the Bayesian stochastic search variable selection (BSSVS) method. In this method, a parsimonious set of significant non-zero rates is identified by conducting a Bayes factors (BF) test. The BF value is the ratio of the likelihood of the model with a particular parameter to the likelihood of the model without that parameter. The parameter-rich model will be chosen if the BF value is greater than three (BF > 3) (8).

Two independent MCMC chains were run for at least 50 million generations, burning the initial 20% of the samples. The convergence and stationary state of each parameter was visually evaluated after combining independent chains using the lag plots. In the posterior distribution, all of the parameters showed convergence with an effective sample size (ESS) of greater than 200. Since the MCMC chains result in a distribution of phylogenetic trees, we needed to summarize them in order to obtain one "consensus" tree. For this purpose, the maximum clade credibility (MCC) trees were constructed. In this method, each tree in the posterior distribution is evaluated and each clade of the tree is given a score based on the fraction of times it appears in the posterior sample (known as posterior clade credibility). The product of these scores is taken as the total score of the tree. The tree with the highest score is selected as the MCC tree (30). In this study, the MCC tree was used to check the timing and the pathways of the virus migration
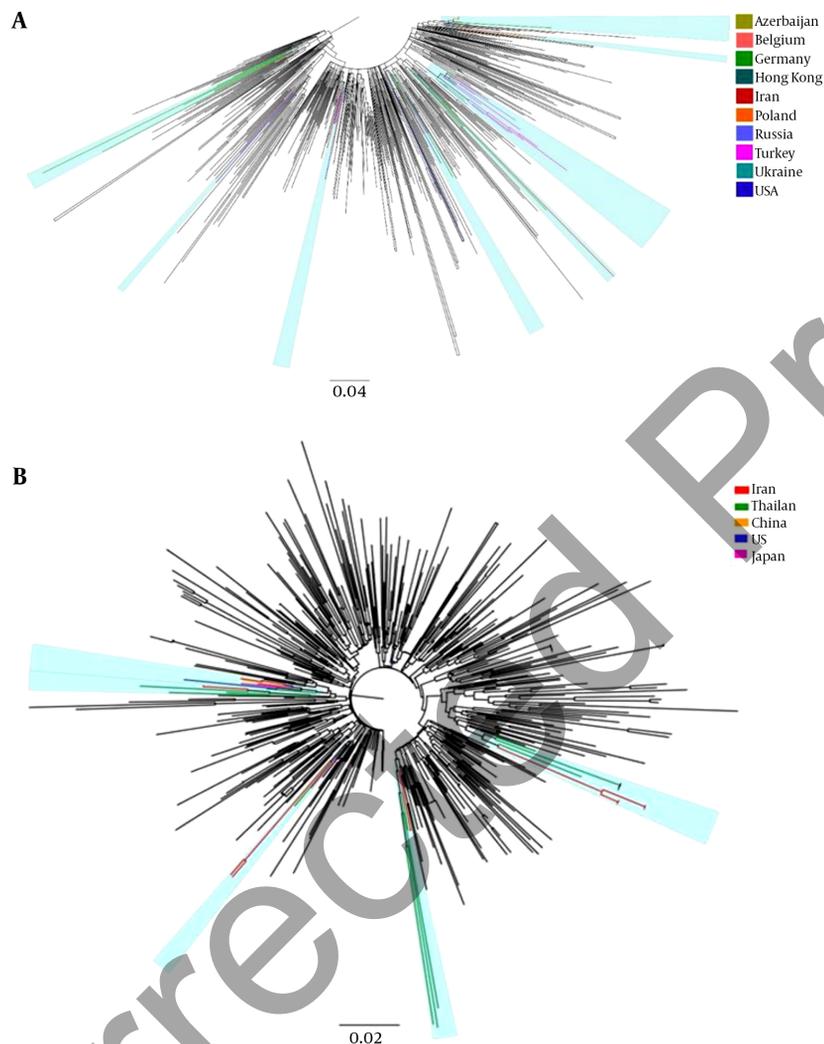
Figure 1. Midpoint rooted maximum likelihood tree of subtype B (Part A) and CRF01_AE (Part B). A set of non-redundant HIV-1 subtype B and CRF01_AE sequences were recruited from countries around the world and used to reconstruct maximum likelihood trees. Only those clusters including one or more Iranian lineages are highlighted in blue. Countries that showed significant genetic similarity with Iranian subtype B lineages (bootstrap > 50%) include Belgium, China, Germany, Poland, Turkey, Ukraine, and the USA. Countries that showed significant genetic similarity with Iranian CRF01_AE sequences (bootstrap > 50%) include Belgium, Cameroon, Central African Republic, China, Italy, Japan, Myanmar, Poland, Singapore, Taiwan, Thailand, United Kingdom, USA, and Vietnam.

between the sampling locations.

The main parameters of the Bayesian phylogeographic model included the evolutionary rate and the date of subtype B and CRF01_AE origins in Iran, and the parameters of the viral migration rate (overall migration rate and number of non-zero rates). Previous estimates of the evolutionary rate for the pol (1.0 - 3.0 × 10^{-3} substitutions/site/year) (13, 33) and env (3.0 - 6.0 × 10^{-3} substitutions/site/year) (34, 35) genes of the HIV-1 group M, under relaxed molecular clocks, were used as uniform priors for evolutionary rate estimation. The date at which the first case of HIV was reported in Iran (1984) (5, 36) was used a priori to estimate

the date of subtype B and CRF01_AE origins in Iran. In this regard, a normal prior with a mean of 27 years (i.e. the time interval between 1984 and 2011 - the latest sampling date in our study) and a standard deviation of three years was used. Migration rate parameters were estimated with uninformative priors because no previous estimate was available.

3.3. Statistical Software

All sequences were aligned using MEGA software v.6 (37) and were manually edited in AliView v.1.17.1 (38). Likelihood mapping analysis was performed using the Tree-

Puzzle program [(28, 39). ML trees were reconstructed using MEGA software v.6 (37). Bayesian phylogeographic analyses were performed in BEAST v.1.8.1 (40) while using the BEAGLE library (41) to improve the computational speed. Multiple independent Bayesian MCMC chains were combined using LogCombiner v.1.8.1 (42). The convergence and stationary state of each parameter was visually checked using lag plots in Tracer v.1.6 (43). MCC trees were reconstructed using TreeAnnotator v.1.8.1 (44) and visualized (same as ML trees) using FigTree v.1.4.2 (45). Significant migratory events were computed using SPREAD v.1.0.7 software (46) and were visualized using Google Earth.

3.4. Ethical Consideration

The ethics committee of Kerman University of Medical Sciences (KUMS) approved the study protocol in 2015 via ethical code 'IR.KMU.REC.1394.181'. Accession numbers of the data used in this study are provided in Appendix 1 in supplementary file.

4. Results

4.1. HIV-1 Sequence Data

The following data sets were used for estimating the subtype B and CRF01_AE spatio-temporal dynamics in Iran:

(i) Subtype B: Contained a total of 366 sequences, with a time-span of 31 years (1983 - 2014). The samples covered part of the pol p66 and p51 RT regions (HXB2 numbering: 2619 - 3219 nts, 600 bp). All the Iranian samples were from Iran and were sampled between 2009 and 2011 (Table 1).

(ii) CRF01_AE: Contained 101 sequences, with a time-span of 19 years (1992 - 2011). Samples covered part of the gp160 region (HXB2 numbering: 6958-7376 nts, 418 bp). All the Iranian samples were from Iran and were sampled between 2006 and 2011 (Table 1).

4.2. Spatio-Temporal Dispersion Pattern

4.2.1. Subtype B

Subtype B sequences from 10 countries around the world were selected for phylogeographic analysis. These countries were included because they were either Iran's neighbors or they showed genetic similarities to the Iranian lineages in the ML tree (see Figure 1A), thus they could be a potential origin for the subtype B epidemic in Iran. These countries included Azerbaijan, Belgium, Germany, Southern China, Poland, Russia, Turkey, Ukraine, and the USA (Table 1).

The MCC tree showed that lineages from the USA occupied the root and basal nodes of the tree while lineages from other countries were mainly nested among them. Germany showed a secondary role in virus exportation to

some countries, including Iran, Turkey, and the USA (Figure 2A, and Figure 3). The topology of the MCC tree suggests that subtype B entered Iran through multiple introduction events and the lineages were probably introduced to the country from the USA, Germany and Southern China in 1974 (Figure 2A).
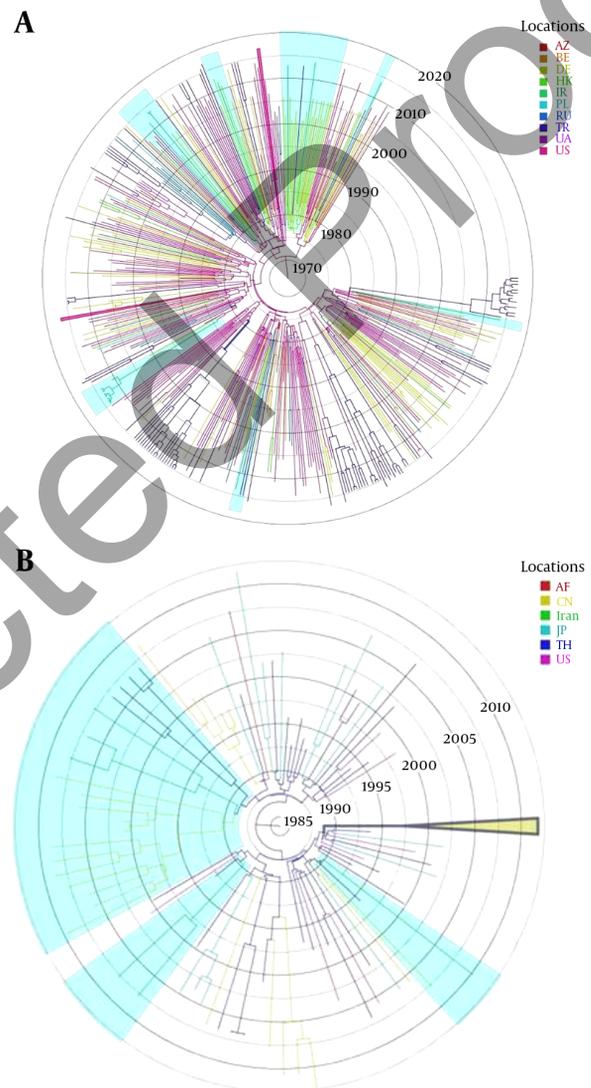

Figure 2. Time-scaled bayesian maximum clade credibility tree for HIV-1 subtype B (p51 RT) and CRF01_AE (gp160) lineages from Iran and selected countries. Branches are colored according to the most probable location state of their descendent nodes. Clusters including Iranian lineages are highlighted in blue. The monophyletic groups containing only one location were collapsed into a triangle for visual clarity. The topology of the trees shows that subtype B has been introduced from Germany, Southern China, and the USA through multiple introduction events (posterior clade credibilities were > 0.8; Part A) and the ancestral location of all CRF01_AE lineages is Thailand (posterior clade credibilities were > 0.8; Part B).

Three countries sharing a border with Iran - Azerbaijan,

Table 1. HIV-1 Subtype B and CRF01_AE Sequences used for Phylogeographic Analysis[a]

Characteristics	Subtype B, p51 RT	01_AE, gp160
	No.	No.
Sex		
Male	244	306
Female	237	265
Unknown	28	32
Sampling Region, Country		
Middle East		
Afghanistan	-	1
Azerbaijan	3	-
Iran	15	6
Asia		
China	10	29
Japan	-	17
Myanmar	-	16
Russia	16	-
Singapore	-	20
Taiwan	-	10
Thailand	-	381
Ukraine	10	-
Vietnam	-	19
Middle Africa		
Cameroon	-	13
Central African Republic	-	13
Europe		
Belgium	24	13
France	-	-
Germany	109	-
Italy	-	15
Poland	10	11
Turkey	133	-
United Kingdom	-	16
North America		
USA	165	17
Total	495	603

[a]Only countries showing genetic similarities to Iranian lineages in the maximum likelihood tree with bootstrap > 50% are presented here and used for phylogeographic analysis. All sequences from countries neighboring Iran were recruited regardless of their genetic similarity to the Iranian lineages. These countries included Afghanistan, Azerbaijan, Russia, and Turkey.

Turkey, and Russia showed independent epidemic histories. Azerbaijan probably imported the virus from Ukraine and the USA in 1976. Turkey showed evidence of multiple independent epidemics since 1972 and imported the virus from Germany, Russia, Ukraine, and the USA. Sequences from Russia only showed an epidemiological link to the
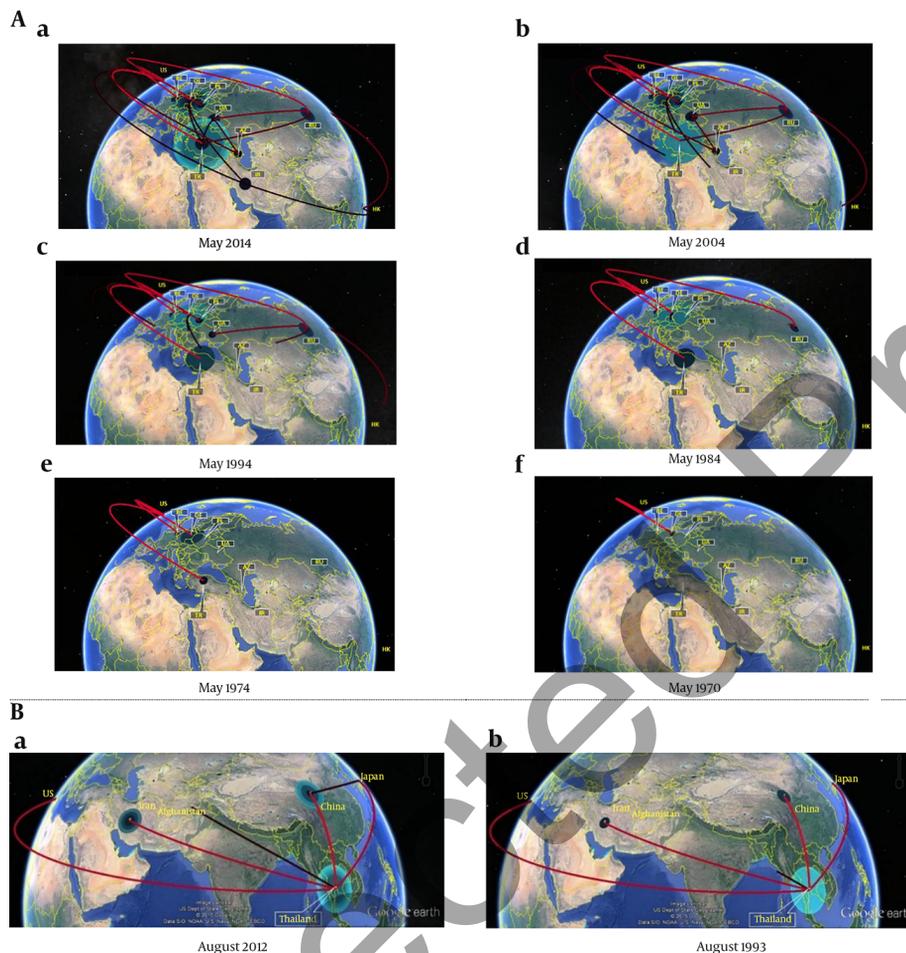
Figure 3. Temporal dynamics of HIV-1, subtype B (part A) and CRF01_AE (part B) spatial dispersion among countries selected for this study. Snapshots of the virus migration are provided for different time points. Lines in the map are branches in the time-scaled Bayesian MCC tree and show the linkage between two countries at each point in time. Location circle diameters are proportional to the square root of the number of MCC branches maintaining a particular country at each point in time. Branch color spectrum is relative to the age of the virus transition between locations with darker colors representing more recent transitions and lighter colors representing older transitions. The results in the upper panel (Part A) show that the earliest transmission of subtype B dates back to 1984 when the virus was probably imported to Iran from the USA (posterior location probability = 0.9). The results in the lower panel (Part B) show that the earliest transmission of CRF01_AE dates back to 1993 with Thailand as the sole virus exporter to Iran (posterior location probability = 0.9). Aa, May 2014; Ab, May 2004; Ac, May 1994; Ad, May 1984; Ae, May 1974; Af, May 1970; Ba, May 2012; Bb, May 1993.

USA, with the earliest introduction of the virus dating back to 1975 (Figures 2A and 3A).

BSSVS analysis showed that the migration rates of subtype B from Germany and Southern China to Iran were statistically significant (BF ≥ 3, Table 2). Although not statistically significant, the migration rate between Iran and Azerbaijan and Turkey was greater than that of Germany and Southern China (Table 2).

The estimated median evolutionary rate under the UCLD assumption was 2.45×10^{-3} substitutions/site/year (95% HPD: $2.1 - 2.8 \times 10^{-3}$). The coefficient of variation of the evolutionary rate was 0.48 (95% HPD: 0.4 - 0.5), which significantly deviates from zero and shows the fitness of the

relaxed molecular clock assumption.

4.2.2. CRF01_AE

The CRF01_AE partial gp160 sequences from Iran, Afghanistan, China, Japan, Thailand, and the USA were considered to be six discrete locations (Table 1). These locations were recruited because they showed genetic similarities with Iranian lineages in the ML tree, thus they could be potential sources for CRF01_AE entry into Iran (Figure 1B). Afghanistan did not make a cluster with Iranian lineages in the ML tree, but was included in the analysis because it shares a common border with Iran.

The topology of the MCC tree showed evidence for mul-

Table 2. HIV-1 Subtype B and CRF01_AE Migration Rates Between Iran and Selected Countries

Countries	Migration Rates, km/y	BF
Subtype B		
Poland	1.00	0.06
Ukraine	1.00	0.05
Russia	1.00	0.05
Belgium	0.98	0.07
Azerbaijan	0.97	0.07
Turkey	0.97	0.05
USA	0.87	1.10
Germany	0.83	28.22[a]
Southern China	0.61	6.21[a]
CRF01_AE		
Japan	0.96	0.83
China	0.93	0.15
Afghanistan	0.93	0.14
USA	0.93	0.11
Thailand	0.66	3.58[a]

[a] Bayes factors values $\geq$ 3 indicate significant migration rates.

tiple introductions of CRF01_AE from Thailand into Iran and Thailand occupied the root and basal nodes of the tree (Figure 2B). The median migration rate between Thailand and Iran was 0.66 km/year (HPD 95%: 0.05, 1.06) and was statistically significant (BF = 3.58, Table 2). There was also evidence for CRF01_AE entry from Thailand to Afghanistan, but through a transmission event that was independent from Iran (Figures 2B and 3B).

The estimated date of origin of the CRF01_AE clade in Iran was 1988 (95% HPD: 1986 - 1989). More recent lineages were introduced into the country through independent transmission events in 1988 (95% HPD: 1987 - 1991) and 1989 (95% HPD: 1987 - 1991). The estimated date of origin for the Afghan lineage was 1988 (1990 - 1986).

The estimated median evolutionary rate under a log-normal relaxed-clock was 5.90×10^{-3} nucleotide substitution/site/year (95% HPD: 4.99×10^{-3} - 6.21×10^{-3}). The coefficient of variation of the evolutionary rate was 0.47 (95% HPD: 0.37 - 0.58) and shows the fitness of the relaxed molecular clock.

5. Discussion

The date of origin and migration pathways of the HIV virus between Iran and other countries has remained largely unknown. There is also a knowledge gap when it comes to the factors that might have contributed to such migratory pathways. The work presented in this paper is the first of its kind to provide information about these parameters in Iran, using a Bayesian phylogeographic approach to HIV-1 subtype B and CRF01_AE clades.

Our results suggest that the subtype B epidemic was introduced to Iran from Southern China, Germany, and the USA. The migration rates from Germany and Southern China were found to be statistically significant. Based on the results of the MCC tree, the entry of the virus from the USA and Germany occurred through multiple transmission events. It is possible that the transmission of subtype B from Southern China to Iran was sporadic, since only one sequence from Iran showed an epidemiological link with this location. Between-country trading and long-distance truck drivers might have played a role in the transmission of the virus between Iran and Germany. The engagement of long-distance truck drivers in high-risk behaviors like unprotected sex with multiple partners and drug use has been reported in several studies (47-49). It is also possible that the introduction of subtype B from Germany and Southern China occurred "indirectly" through other countries.

To investigate this possibility, we added a sample of sequences from Thailand, China, Japan, and twelve European countries (i.e. Austria, Bulgaria, Czech Republic, Denmark, France, Greece, Guyana, Italy, Spain, Sweden, the Netherlands, and the United Kingdom) to the phylogeographic analysis. It was found that the results were robust to the inclusion of these countries. That is, none of the aforementioned countries showed epidemiological links with Iran (results are not shown). Since the identification of the B subtype among some hemophiliac patients in Iran, it has been assumed that this clade was imported from France through contaminated blood products (5). Our study reveals the existence of other (previously unknown) sources of subtype B entry into Iran.

Although not statistically significant, the migration rates of the HIV-1 subtype B between Iran and Azerbaijan, as well as Iran and Turkey, were found to be greater than those of Southern China and Germany. Due to a number of factors, such as geographic proximity and ease of travel, Turkey is a major destination for Iranian tourists and tradespeople. The large volume of human traffic between Iran and Turkey means that it is crucial to prevent HIV-1 transmission between these two countries.

Our results showed that the HIV-1 subtype B was introduced to Iran in 1974 (95% HPD: 1966-1974). The first AIDS case was identified and reported to Iran's health care system in 1984 (36). Faria et al. (2012) reported a lag of about 10 years between the tMRCA of CRF02_AG clade in Cameroon

(which is analogous to the "incidence" of this clade (50, 51)) and HIV "prevalence" data (52). Based on their estimates, the 10-year lag observed between the tMRCA of subtype B in Iran (1974; the incidence time) and the date of the first "prevalent" case in Iran (1984) seems to be plausible. Moreover, we conducted a set of sensitivity analyses by using different priors for the tMRCA of the Iranian lineages. Interestingly, the estimated tMRCA did not change and was robust to the priors used (the results are not shown).

Our study revealed that the evolutionary rate of subtype B at the pol gene (2.45×10^{-3}, 95% HPD= 2.1-2.8 $\times 10^{-3}$ substitutions/site/year) in Iran is similar to the rates previously described for HIV-1 group M circulating in the Caribbean Region (13), Angola, Brazil, Romania (33), Uruguay (53), East Africa (14), and the United Kingdom (54). This finding shows that there is no major difference between these locations and Iran in terms of their rates of HIV evolution. This finding also accredits the estimates of the MRCA in our study. The branching pattern of the American lineages in our MCC tree, highlighting the USA as one of the main sources of subtype B dissemination in the world, is similar to that of previous studies (3). This observation also strengthens the validity of our analyses and topology estimates.

CRF01_AE strains were probably imported from Thailand into Iran through multiple transmission events during a time-span of about five years (between 1986 and 1991). There was no evidence of virus exportation from Iran to other countries. Thailand experienced a dramatic rise in the number of HIV-1 infected individuals in the early nineties (1988 - 1993) and the majority of the infections in this period were due to heterosexual transmission from commercial sex workers (2, 4). Although implementing condom-use programs decreased the number of new infections to a great extent, Thai sex workers continue to suffer from an CRF01_AE epidemic (2). Similar to Turkey, Thailand is another popular touristic and commercial destination for many Iranians. These factors might have played an important role in the transmission of CRF01_AE from Thailand to Iran. Angelis et al. (2015) conducted a phylogeographic study to infer the origins and global migration pathways of CRF01_AE. They showed that Thailand has been the source of viral dispersal to most parts of the world. Regarding Iran, the only CRF01_AE sequence that was used in their analysis showed a transmission history from Thailand, which is consistent with the results of our study. In their study, such a dispersal pattern of the virus was attributed to the popularity of Thailand as a vacation destination, mostly characterized by sex tourism, and to the emigration of Thai citizens to the western world (2).

The date of CRF01_AE entry into Iran was estimated to be 1986, which is consistent with the date of CRF01_-AE entry into Thailand (i.e. circa the 1970s) (2), with the latter date preceding the former. The median evolutionary rate of the CRF01_AE env gene was 5.90×10^{-3} substitutions/site/year (95% HPD: 4.99×10^{-3} - 6.21×10^{-3}). This is also consistent with the previous estimates for the env gene of the HIV-1 group M (34, 35). These findings accredit our tMRCA estimates and support the validity of our phylogeographic model.

Among Iran's neighbors, only Afghanistan had an available CRF01_AE sequence. The estimated date of CRF01_AE entry into Afghanistan was 1988, showing relatively the same antiquity of the epidemic as Iran. There was also evidence for direct transmission of the virus from Thailand to Afghanistan. This finding is in contrast with Mumtaz et. al.'s (2014) claim that suggested Iran was the exporter of HIV to Afghanistan (55). While inadequate samples prevent us from deriving general conclusions about the CRF01_AE transmission between Afghanistan and Iran, this finding challenges the theory that Iran has been a mere exporter of the HIV virus to Afghanistan.

While the transmission of CRF01_AE has been related to sexual activities and was probably introduced to Iran through heterosexual contacts, samples in our data set with known risk group reported intravenous drug use as their main HIV-1 transmission risk. This observation sheds light on possible episodes of HIV-1 transmission occurring between heterosexuals and people who inject drugs (PWIDs) in Iran; this has implications for public health and preventive strategies.

Despite extended preventive and educational programs conducted in Iran (56), our study suggests that the HIV-1 CRF01_AE strains continue to be transmitted episodically from Thailand to Iran. This finding highlights the need for the development of targeted preventive strategies, such as proper educational interventions for tourists, tradespeople, and truck drivers.

Unfortunately, we could not obtain a random sample of HIV-infected individuals who were infected with subtype B or CRF01_AE. Obtaining the sampling frame for HIV-infected populations is not a trivial task, since these populations are usually "hidden" and "hard-to-reach" (due to the social stigma commonly associated with the activities that may have led to their contracting of the disease, such as drug injection, unprotected sex, having multiple sexual partners, etc.), or they are "unaware" of their HIV infection. For example, in Iran, it is estimated that about two-thirds of those infected with HIV are not aware of their status (57). This makes it even more tedious, if not impossible, to gather a random sample in cases such as our study, where the goal is to recruit individuals infected with a particular strain of HIV (such as subtype B or CRF01_AE). Information regarding an individual's viral strain cannot be ob-

tained until we collect their blood specimens and identify the strain type. Therefore, typical sampling frameworks cannot be applied in such circumstances, since the characteristics of the reference population cannot be defined before the sampling is completed. Therefore, we followed the common approach in this regard and used available HIV sequence data, whose strain types were known, from the LAHIV databases (58).

We believe it is a reasonable assumption that the main conclusions about the epidemic roots will not change dramatically if we use a random sample from the epidemic that we have studied (i.e. the epidemic from which we have selected our sequences). This is because we inferred the epidemic roots from the genetic material of the HIV virus, which does not significantly differ among the lineages circulating within an epidemic. However, we acknowledge that these results cannot be generalized on a broader scale if parallel epidemics with similar strain-types but different roots exist in Iran. Moreover, due to the lack of knowledge regarding the molecular-epidemiology of HIV in Iran, the existence of such epidemics could not be verified. The lack of sufficient data and/or information about the HIV epidemic in Iran underscores the contribution of our study as it sheds light on previously unknown aspects of the epidemic in the country (i.e. date of origin, spatial dynamics, and key factors contributing to the virus' ability to transcend Iran's national borders).

Our estimations of the unknown aspects of the HIV epidemic in Iran are based on the data sampled up to 2011. It is reasonable to assume that these estimates can be generalized to more recent viral samples isolated from the epidemics that we have studied. The reason is that in a particular HIV epidemic, more recent HIV samples are more probably the "descendants" of the older samples (since the HIV virus is transmitted from one patient to another as time goes on) and share similar genetic material. Therefore, we believe it is reasonable to assume that this issue does not significantly affect the spatio-temporal inferences we have made about the epidemics from which we have sampled.

Given that the geo-referenced sequence data available for this analysis was limited, the phylogeographic inferences presented here would benefit from a more comprehensive sampling scheme and lateral behavioral information. For example, the small number of samples available from Azerbaijan, Afghanistan, and Iran may have hampered the detection of epidemiological links between Iran and these neighboring countries. We also could not evaluate the history of virus transmission between Iran and other neighboring countries, such as Iraq, Bahrain, United Arab Emirates, Kuwait, Qatar, and Oman due to the lack of HIV-1 sequence data available for these countries. This invokes the need for further research in this area as more

data becomes available in the future.

Since Bayesian phylogeography is a new approach, estimates about migration rate parameters for different phylogeographic scenarios are still scarce in the literature. For this reason, we estimated migration rate parameters using uninformative priors.

Our study highlights the use of phylogeographic studies to improve our understanding of infectious disease epidemics and factors potentially associated with their spread. Considering our limited knowledge on the HIV epidemic in Iran, any findings presenting different aspects of the epidemic are helpful. Our study highlighted the role of some "neglected" groups, such as long-distance truck drivers, tradespeople and tourists, in importing the HIV virus to Iran. While typical high-risk groups, such as PWIDs, female sex workers (FSWs), and men who have sex with men (MSMs), contribute to within-country virus transmission, the groups identified in this study facilitate between-country virus transmission resulting in the introduction of diverse, and in some cases drug-resistant, strains to the country. This poses serious challenges for the treatment and control of the HIV epidemic in Iran. Therefore, proper preventive strategies should be considered for these high-risk groups.

5.1. Conclusion

Understanding the origins and migration pathways of successful HIV-1 clades over large geographic distances is invaluable for improving the characterization and control of the spread of HIV. Our study describes for the first time the emergence and dynamics of two important HIV-1 clades in Iran, highlighting the role of China, Germany, the USA, and Thailand as major viral exporters. Factors such as international trade, transit of goods, and tourism are among the likely medians for transmission of these clades. However, the existing preventive strategies in Iran have been mostly focused on other risk groups, such as sex workers and people who inject drugs. The findings of our study call for the need to develop targeted preventive strategies, such as educational interventions, for tourists, truck drivers, and people involved in international trade and transactions in order to prevent onward transmissions. Nevertheless, and given that the geo-referenced sequence data available for analysis is limited, the phylogeographic inferences presented here would benefit from a more comprehensive sampling scheme and lateral behavioral information.

Supplementary Material

Supplementary material(s) is available here.

Acknowledgments

Footnotes

References

1. Sharp PM, Hahn BH. The evolution of HIV-1 and the origin of AIDS. *Philos Trans R Soc Lond B Biol Sci.* 2010;**365**(1552):2487–94. doi: 10.1098/rstb.2010.0031. [PubMed: 20643738].

2. Angelis K, Albert J, Mamais I, Magiorkinis G, Hatzakis A, Hamouda O, et al. Global Dispersal Pattern of HIV Type 1 Subtype CRF01_AE: A Genetic Trace of Human Mobility Related to Heterosexual Sexual Activities Centralized in Southeast Asia. *J Infect Dis.* 2015;**211**(11):1735–44. doi: 10.1093/infdis/jiu666. [PubMed: 25512631].

3. Junqueira DM, de Medeiros RM, Matte MC, Araujo LA, Chies JA, Ashton-Prolla P, et al. Reviewing the history of HIV-1: spread of subtype B in the Americas. *PLoS One.* 2011;**6**(11):27489. doi: 10.1371/journal.pone.0027489. [PubMed: 22132104].

4. Carr JK, Salminen MO, Koch C, Gotte D, Artenstein AW, Hegerich PA, et al. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J Virol.* 1996;**70**(9):5935–43. [PubMed: 8709215].

5. Sarrami-Forooshani R, Das SR, Sabahi F, Adeli A, Esmaeili R, Wahren B, et al. Molecular analysis and phylogenetic characterization of HIV in Iran. *J Med Virol.* 2006;**78**(7):853–63. doi: 10.1002/jmv.20634. [PubMed: 16721846].

6. Jahanbakhsh F, Ibe S, Hattori J, Monavari SH, Matsuda M, Maejima M, et al. Molecular epidemiology of HIV type 1 infection in Iran: genomic evidence of CRF35_AD predominance and CRF01_AE infection among individuals associated with injection drug use. *AIDS Res Hum Retroviruses.* 2013;**29**(1):198–203. doi: 10.1089/AID.2012.0186. [PubMed: 22916738].

7. Foxman B. In: Molecular Tools and Infectious Disease Epidemiology. 1 ed. 1st , editor. New York: Elsevier; 2012.

8. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. *PLoS Comput Biol.* 2009;**5**(9):1000520. doi: 10.1371/journal.pcbi.1000520. [PubMed: 19779555].

9. Kuhnert D, Wu CH, Drummond AJ. Phylogenetic and epidemic modeling of rapidly evolving infectious diseases. *Infect Genet Evol.* 2011;**11**(8):1825–41. doi: 10.1016/j.meegid.2011.08.005. [PubMed: 21906695].

10. Holmes EC. The phylogeography of human viruses. *Mol Ecol.* 2004;**13**(4):745–56. [PubMed: 15012753].

11. Lemey P, Rambaut A, Welch JJ, Suchard MA. Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol.* 2010;**27**(8):1877–85. doi: 10.1093/molbev/msq067. [PubMed: 20203288].

12. Hein J, Schierup MH, Wiuf C. Gene Genealogies, Variation and Evolution. Oxford: Oxford University Press; 2005.

13. 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):106045. doi: 10.1371/journal.pone.0106045. [PubMed: 25148215].

14. Delatorre EO, Bello G. Phylodynamics of HIV-1 subtype C epidemic in east Africa. *PLoS One.* 2012;**7**(7):41904. doi: 10.1371/journal.pone.0041904. [PubMed: 22848653].

15. Delatorre E, Bello G. Spatiotemporal dynamics of the HIV-1 CRF06_cpx epidemic in Western Africa. *AIDS.* 2013;**27**(8):1313–20. doi: 10.1097/QAD.0b013e32835f1df4. [PubMed: 23343915].

16. Paraskevis D, Pybus O, Magiorkinis G, Hatzakis A, Wensing AM, van de Vijver DA, et al. Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. *Retrovirology.* 2009;**6**:49. doi: 10.1186/1742-4690-6-49. [PubMed: 19457244].

17. Abubakar YF, Meng Z, Zhang X, Xu J. Multiple independent introductions of HIV-1 CRF01_AE identified in China: what are the implications for prevention?. *PLoS One.* 2013;**8**(11):80487. doi: 10.1371/journal.pone.0080487. [PubMed: 24282546].

18. Nerurkar VR, Nguyen HT, Dashwood WM, Hoffmann PR, Yin C, Morens DM, et al. HIV type 1 subtype E in commercial sex workers and injection drug users in southern Vietnam. *AIDS Res Hum Retroviruses.* 1996;**12**(9):841–3. doi: 10.1089/aid.1996.12.841. [PubMed: 8738437].

19. Kalish ML, Korber BT, Pillai S, Robbins KE, Leo YS, Saekhou A, et al. The sequential introduction of HIV-1 subtype B and CRF01AE in Singapore by sexual transmission: accelerated V3 region evolution in a subpopulation of Asian CRF01 viruses. *Virology.* 2002;**304**(2):311–29. [PubMed: 12504572].

20. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med.* 2011;**365**(6):493–505. doi: 10.1056/NEJMoa1105243. [PubMed: 21767103].

21. Apetrei C, Hahn B, Mizrachi I, Mullins J, Rambaut A, Wolinsky S. Los Alamos Sequence database. Los Alamos HIV databases 2015. Available from: http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html.

22. Apetrei C, Hahn B, Mizrachi I, Mullins J, Rambaut A, Wolinsky S. Los Alamos HIV databases. *US.* 2015.

23. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res.* 2009;**37**(Database issue):26–31. doi: 10.1093/nar/gkn723. [PubMed: 18940867].

24. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006;**22**(13):1658–9. doi: 10.1093/bioinformatics/btl158. [PubMed: 16731699].

25. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* 2010;**26**(5):680–2. doi: 10.1093/bioinformatics/btq003. [PubMed: 20053844].

26. Delatorre E, Couto-Fernandez JC, Guimaraes ML, Vaz Cardoso LP, de Alcantara KC, Stefani MM, et al. Tracing the origin and northward dissemination dynamics of HIV-1 subtype C in Brazil. *PLoS One.* 2013;**8**(9):74072. doi: 10.1371/journal.pone.0074072. [PubMed: 24069269].

27. Marcus U, Schmidt AJ, Kollan C, Hamouda O. The denominator problem: estimating MSM-specific incidence of sexually transmitted infections and prevalence of HIV using population sizes of MSM derived from Internet surveys. *BMC Public Health.* 2009;**9**:181. doi: 10.1186/1471-2458-9-181. [PubMed: 19519889].

28. Strimmer K, von Haeseler A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc Natl Acad Sci U S A.* 1997;**94**(13):6815–9. [PubMed: 9192648].

29. Ackermann M, Strimmer K. A general modular framework for gene set enrichment analysis. *BMC Bioinformatics.* 2009;**10**:47. doi: 10.1186/1471-2105-10-47. [PubMed: 19192285].

30. Lemey P, Salemi M, Vandamme AM. The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. 2 ed. london: Cambridge University Press; 2009.

31. Tee KK, Pybus OG, Li XJ, Han X, Shang H, Kamarulzaman A, et al. Temporal and spatial dynamics of human immunodeficiency virus type 1 circulating recombinant forms 08_BC and 07_BC in Asia. *J Virol.* 2008;**82**(18):9206–15. doi: 10.1128/JVI.00399-08. [PubMed: 18596096].

32. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* 1985;**22**(2):160–74. [PubMed: 3934395].

33. Bello G, Afonso JM, Morgado MG. Phylodynamics of HIV-1 subtype F1 in Angola, Brazil and Romania. *Infect Genet Evol.* 2012;**12**(5):1079–86. doi: 10.1016/j.meegid.2012.03.014. [PubMed: 22484759].

34. Veras NM, Gray RR, Brigido LF, Rodrigues R, Salemi M. High-resolution phylogenetics and phylogeography of human immunodeficiency virus type 1 subtype C epidemic in South America. *J Gen Virol.* 2011;**92**(Pt 7):1698–709. doi: 10.1099/vir.0.028951-0. [PubMed: 21450946].

35. 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. doi: 10.1126/science.1256739. [PubMed: 25278604].

36. Naderi HR, Tagliamonte M, Tornesello ML, Ciccozzi M, Rezza G, Farid R, et al. Molecular and phylogenetic analysis of HIV-1 variants circulating among injecting drug users in Mashhad-Iran. *Infect Agent Cancer.* 2006;**1**:4. doi: 10.1186/1750-9378-1-4. [PubMed: 17150133].

37. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 2013;**30**(12):2725–9. doi: 10.1093/molbev/mst197. [PubMed: 24132122].

38. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics.* 2014;**30**(22):3276–8. doi: 10.1093/bioinformatics/btu531. [PubMed: 25095880].

39. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics.* 2002;**18**(3):502–4. [PubMed: 11934758].

40. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 2007;**7**:214. doi: 10.1186/1471-2148-7-214. [PubMed: 17996036].

41. Suchard MA, Rambaut A. Many-core algorithms for statistical phylogenetics. *Bioinformatics.* 2009;**25**(11):1370–6. doi: 10.1093/bioinformatics/btp244. [PubMed: 19369496].

42. Rambaut A, Drummond A. LogCombiner v1.8.1 UK: BEAST; 2002-2014. Available from: http://beast.bio.ed.ac.uk/LogCombiner.

43. Rambaut A, Drummond A. Tracer v1.4 UK: BEAST; 2007. Available from: http://beast.bio.ed.ac.uk/Tracer.

44. Rambaut A, Drummond A. TreeAnnotator v1.8.1 UK: BEAST; 2002-2014. Available from: http://beast.bio.ed.ac.uk/TreeAnnotator.

45. Rambaut A. FigTree v1.4: Tree Figure Drawing Tool UK: BEAST; 2009. Available from: http://treebioedacuk/software/figtree/.

46. Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics.* 2011;**27**(20):2910–2. doi: 10.1093/bioinformatics/btr481. [PubMed: 21911333].

47. Motevalian SA, Jahani M, Mahmoodi M. Driving under influence of opiates in heavy vehicle drivers of Iran in 2001. *Hakim Res J.* 2004;**7**(1):1–8.

48. Delany-Moretlwe S, Bello B, Kinross P, Oliff M, Chersich M, Kleinschmidt I, et al. HIV prevalence and risk in long-distance truck drivers in South Africa: a national cross-sectional survey. *Int J STD AIDS.* 2014;**25**(6):428–38. doi: 10.1177/0956462413512803. [PubMed: 24352131].

49. Baltazar CS, Horth R, Inguane C, Sathane I, Cesar F, Ricardo H, et al. HIV prevalence and risk behaviors among Mozambicans working in South African mines. *AIDS Behav.* 2015;**19 Suppl 1**:S59–67. doi: 10.1007/s10461-014-0941-6. [PubMed: 25398418].

50. Frost SD, Volz EM. Viral phylodynamics and the search for an 'effective number of infections'. *Philos Trans R Soc Lond B Biol Sci.* 2010;**365**(1548):1879–90. doi: 10.1098/rstb.2010.0060. [PubMed: 20478883].

51. Volz EM, Kosakovsky Pond SL, Ward MJ, Leigh Brown AJ, Frost SD. Phylodynamics of infectious disease epidemics. *Genetics.* 2009;**183**(4):1421–30. doi: 10.1534/genetics.109.106021. [PubMed: 19797047].

52. Faria NR, Suchard MA, Abecasis A, Sousa JD, Ndembi N, Bonfim I, et al. Phylodynamics of the HIV-1 CRF02_AG clade in Cameroon. *Infect Genet Evol.* 2012;**12**(2):453–60. doi: 10.1016/j.meegid.2011.04.028. [PubMed: 21565285].

53. Bello G, Aulicino PC, Ruchansky D, Guimaraes ML, Lopez-Galindez C, Casado C, et al. Phylodynamics of HIV-1 circulating recombinant forms 12_BF and 38_BF in Argentina and Uruguay. *Retrovirology.* 2010;**7**:22. doi: 10.1186/1742-4690-7-22. [PubMed: 20307282].

54. Foster GM, Ambrose JC, Hue S, Delpech VC, Fearnhill E, Abecasis AB, et al. Novel HIV-1 recombinants spreading across multiple risk groups in the United Kingdom: the identification and phylogeography of Circulating Recombinant Form (CRF) 50_A1D. *PLoS One.* 2014;**9**(1):83337. doi: 10.1371/journal.pone.0083337. [PubMed: 24454702].

55. Mumtaz GR, Riedner G, Abu-Raddad LJ. The emerging face of the HIV epidemic in the Middle East and North Africa. *Curr Opin HIV AIDS.* 2014;**9**(2):183–91. doi: 10.1097/COH.0000000000000038. [PubMed: 24445372].

56. UNAIDS . Islamic Republic of Iran AIDS progress report. Geneva: UN-AIDS; 2014.

57. Haghdoost AA, Mostafavi E, Mirzazadeh A, Navadeh S, Feizzadeh A, Fahimfar N. Modelling of HIV/AIDS in Iran up to 2014. *J AIDS HIV Res.* 2011;**3**(12) doi: 10.5897/jahr11.030.

58. Apetrei C, Hahn B, Mizrachi I, Mullins J, Rambaut A, Wolinsky S. HIV Circulating Recombinant Forms (CRFs) US: Los Alamos HIV databases; 2015. Available from: http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.htmlCRF01.