UNIVERSITY OF BELGRADE

MEDICAL FACULTY

Luka D. Jovanović

ANALYSIS OF EVOLUTIONARY AND TRANSMISSION DYNAMICS OF HUMAN IMMUNODEFICIENCY VIRUS IN SERBIA AND THE BALKANS

Doctoral Dissertation

Belgrade, 2022

UNIVERZITET U BEOGRADU

MEDICINSKI FAKULTET

Luka D. Jovanović

ANALIZA EVOLUCIONE I TRANSMISIONE DINAMIKE INFEKCIJE VIRUSOM HUMANE IMUNODEFICIJENCIJE U SRBIJI I NA BALKANU

Doktorska disertacija

Beograd, 2022.

INFORMATION ON MENTOR AND DISSERTATION COMMITTEE

MENTOR: Maja Stanojević, MD, PhD, full professor at Medical Faculty University of Belgrade

DISSERTATION COMMITTEE:

Aleksandra Knežević, MD, PhD, full professor at Medical Faculty University of Belgrade

Dimitrios Paraskevis, PhD, Associate Professor at Medical School, National and Kapodistrian University of Athens

Marina Šiljić, PhD, Research Associate, Medical Faculty University of Belgrade

Defence date:

MENTOR: Prof. dr Maja Stanojević, redovni profesor Medicinskog fakulteta, Univerziteta u Beogradu

ČLANOVI KOMISIJE:

Prof. dr Aleksandra Knežević, redovni profesor Medicinskog fakulteta Univerziteta u Beogradu

Prof. dr Dimitrios Paraskevis, vanredni profesor Medicinskog fakulteta, Univeriteta u Atini

N.Sar. dr sci. med. Marina Šiljić, naučni saradnik Instituta za mikrobiologiju i imunologiju Medicinskog fakulteta Univerziteta u Beogradu

Datum odbrane:

Izjava zahvalnosti / Acknowledgements

Posebnu zahvalnost dugujem svim zaposlenima na Institutu za mikrobiologiju i imunologiju, kao i lekarima odeljenja za HIV/AIDS UKCS-a bez čije posvećenosti zdravstvenoj zaštiti HIV-om inficiranih pacijenta ovaj rad ne bi bio moguć.

Takođe veliku zahvalonst dugujem svojoj mentorki profesorki Maji Stanojević, na stručnoj podršci i neumornoj motivaciji, koju mi je pružala tokom izrade doktorske disertacije, kao i na bezrezervnoj pomoći kad god mi je bila potrebna, bez koje sve ovo ne bi bilo moguće.

Veliku zahvalnost dugujem i svojim koleginicama i prijateljicama dr Marini Šiljić i dr Valentini Ćirković, na nesebičnom trudu, sjajnim idejama i iskustvu koje nikada nisu štedele te su time značajno doprinele ovom istraživanju.

Zahvaljujem se svim članovima komisije, koji su svojim korisnim savetima značajno doprineli poboljšanju konačne verzije ove doktorske disertacije.

Special thanks go to professor Dimitrios Parasekvis, who is a great college and friend of our laboratory, always ready to lift our research with his ideas, original insights, takes and experience. It is a great honour to have you on the dissertation committee.

Special thanks also need to go to our colleagues from Slovenia professor Mario Poljak and dr Maja Lunar, for their engagement and critical review of the analysis of HIV, spread in Balkan. Moreover, many thanks also go to our dear colleges from Romania, Greece, Bulgaria, Croatia, and Montenegro for helping me to analyse the HIV epidemic on the Balkan level.

Takođe, zahvaljujem se dragim kolegama i naučnicima Instituta za onkologiju i radiologiju Srbije, a posebno prof. dr Marini Nikitović, prof. dr Vesni Plešinac Karapandžić, n.sar. prim. dr sci. med. Vesni Stanković i ass. dr sci. med. Jeleni Bokun bez čije podrške i vere u moj razvojni put moj angažman na ovom projektu ne bi bio moguć.

Veliku zahvalnost dugujem i svojim dragim prijateljima i kolegama, a posebno dr Tamari Marinković, dr Jovani Stević, dr Mladenu Marinkoviću, dr Predragu Filipoviću, dr Miljani Čović i dr Marini Grbić na svesrdnoj podršci na ovom putu.

Najveću zahvalnost dugujem svojoj majci i ocu na neizmernoj ljubavi i vaspitanju, kao i svom bratu na neizmernoj podršci i ljubavi bez kojih ni jedna doktorska disertacija ne može da se završi.

POSVEĆENO,

Mojim roditeljima i bratu

„Te 1985. godine taj isti Ivica otišao je da se testira na neku čudnu bolest koja se pojavila u Beogradu. Vratio se i celom gradu je rekao da je zaražen AIDS-om. Ovo je jedan ružan i mali grad – okrenuo mu je leđa. A on je vrlo ponosno, dostojanstveno i skromno u svom balon mantilu još godinu i po dana svirao bubnjeve".

Sonja Savić, o svom prijatelju i legendarnom muzičaru Novog talasa Ivici Vdoviću, Vdu, navodno prvoj registrovanoj HIV pozitivnoj osobi u Jugoslaviji.

Abstract:

Aims of the Study: To investigate evolutionary and transmission dynamics of the human immunodeficiency virus (HIV) epidemic in Serbia, in general and key populations, as well as to explore cross-border transmission clusters and spread of HIV subtype B in the Balkans.

Methods:
Advanced phylogenetic approaches, including phylodynamic and phylogeographic analyses, using methods based on maximum likelihood algorithms and Bayesian statistics, implemented in different softwares such as MEGA, PAUP, MrBayes and BEAST, mathematical modeling and latent class analysis (LCA) were applied on two study datasets: a dataset of 385 Serbian HIV sequences collected during, the period from 1997 to 2015, and a dataset of 2,415 HIV subtype B sequences from seven Balkan countries, from the period 1999-2019, both of the pol genetic region and linked to demographic, epidemiological and clinical data. Maximum likelihood phylogenetic analysis implemented in IQ tree was used to identify transmission clusters, based on defined criteria sets. Long-term analysis of the HIV epidemic growth trend, in relation to the mode of transmission, was performed using two approaches: logistic modelling of the HIV epidemic growth in Serbia using data for the period 1985-2015 and join-point regression analysis of HIV incidence for 7 Balkan countries for the period 2004-2019, using the reported epidemiological data.

Results:
Phylogenetic analysis of the Serbian dataset revealed the presence of 19 transmission clusters that accomplished all predefined criteria sets along with one expanded transmission network, comprising 44% of the dataset (170/385), all within subtype B. The majority of clusters, 15/19 (79%), contained sequences from men who have sex with men (MSM) as reported transmission risk. Subtype C sequences formed a highly supported monophyletic group that did not meet the cluster definition criteria, however, consisted solely of HIV sequences related to heterosexual transmission risk. Phylodynamic analysis implied continuous expansion of the three largest transmission clusters during the entire study period, associated with subtype B and the MSM transmission group. In contrast, analysis of heterosexual monophyletic subtypes B and C clades indicated stagnation of cluster growth. Regarding HIV subtype B epidemic in the Balkans, proportion of sequences within transmission clusters was 68% (1,642/2,415), forming 93 clusters. Four of these clusters fulfilled criteria for cross-border clusters, containing 11% of the total number of studied sequences (264/2,415). Phylogeographic analysis indicated that cross-border spread of HIV subtype B in the Balkans took place between the countries of former Yugoslavia and Romania, creating complex reciprocal patterns of spread between Serbia, Slovenia, Croatia and Montenegro starting in the mid-80s, while the spread to Romania started in the mid-90s.
A trend in the number of new HIV diagnoses in Serbia indicated an early exponential phase of growth in connection with MSM transmission risk, continuously until 2030. Joinpoint regression analysis implied heterogeneous trends regarding incidence of new HIV diagnoses in the countries of the Balkans, with a decrease seen in Romania, Greece and Slovenia and an increase in Serbia, Bulgaria, Croatia and Montenegro, mainly linked to MSM transmission risk throughout the region.

Conclusion:

HIV epidemic in Serbia remains in the exponential growth phase, in particular related to MSM transmission risk, confirming young MSM with concomitant sexually transmitted diseases as the

main target group for planning structured preventive policies.
HIV subtype B dispersal pattern within the Balkans is driven by local transmission clusters, with cross-border clusters observed mainly between countries of former Yugoslavia. that has continued well after country break-up. The spread of subtype B through multiple introductions to Romania resonates with the changing pattern of migration upon European integration of the Balkan countries, which started in the early 2000s.

Keywords: HIV, Serbia, Balkans, HIV subtypes, phylogenetic analyses, transmission clusters, Phylodinamic analysis, Phylogeographic analysis, latent class analysis, join-point regression, logistic growth modeling

Scientific field: Medicine

Scientific subfield: Microbes and infection

UDK No:

Apstrakt:

Ciljevi studije: Ispitati evolucionu i transmisionu dinamiku epidemije virusom humane imunodeficijencije (HIV) u Srbiji, u opštoj i u populacijama u riziku, kao i da se ispita širenje HIV podtipa B na Balkanu i zastupljenost prekograničnih transmisionih klastera.

Metode: Napredni filogenetski pristupi, uključujući filodinamičku i filogeografsku analizu, korišćenjem metoda zasnovanih na algoritmu maksimalne verovatnoće i Bajesove statistike, implementiranih u softverima MEGA, PAUP, MrBayes i BEAST, matematičko modelovanje i analiza latentnih klasa (latent class analysis - LCA) primenjeni su na dva seta podataka: 385 sekvenci iz Srbije iz perioda od 1997. do 2015. i 2.415 HIV podtip B sekvenci iz sedam balkanskih zemalja, iz perioda 1999-2019, sa povezanim demografskim, epidemiološkim i kliničkim podacima. Filogenetska analiza maksimalne verovatnoće implementirana u softver IQ tree korišćena je za identifikaciju transmisionih klastera, uz primenu definisanih skupova kriterijuma. Za analizu dugoročnog trenda rasta HIV epidemije primenjena su dva pristupa: logističko modelovanje rasta HIV epidemije u Srbiji na onovu podataka za period 1985-2015 i „join-point" regresione analize incidencije novodijagnostikovanih HIV slučajeva u 7 balkanskih zemalja, za period 2004-2019.

Rezultati: Filogenetskom analizom HIV sekvenci iz Srbije identifikovano je postojanje 19 transmisionih klastera, koji su ispunili sve unapred definisane kriterijume, kao i jedna proširena transmisiona mreža, što čini 44% ispitivanog skupa podataka (170/385), sve unutar podtipa B. Većina klastera, 15/19 (79%), sadržala je sekvence muškaraca koji imaju seks sa muškarcima (MSM) kao prijavljeni rizik za HIV transmisiju. Sekvence podtipa C formirale su monofiletsku grupu sa visokom statističkom podrškom koja nije ispunjavala kriterijume definicije klastera, međutim, sastojala se isključivo od HIV sekvenci povezanih sa heteroseksualnim kontaktom kao rizikom za prenos. Filodinamička analiza je ukazala na kontinuiranu ekspanziju tri najveća transmisiona klastera tokom čitavog perioda istraživanja, povezana sa podtipom B i grupom transmisije MSM. Nasuprot tome, analiza heteroseksualnih monofiletskih klada podtipa B i C ukazala je na stagnaciju rasta. Što se tiče epidemije HIV podtipa B na Balkanu, udeo sekvenci unutar transmisionih klastera bio je 68% (1.642/2.415), unutar 93 klastera. Četiri klastera, sa 11% od ukupnog broja proučavanih sekvenci (264/2.415), ispunila su kriterijume za prekogranične klastere . Filogeografska analiza je pokazala da se prekogranično širenje HIV podtipa B na Balkanu odvijalo između zemalja bivše Jugoslavije i Rumunije, stvarajući složene recipročne obrasce širenja između Srbije, Slovenije, Hrvatske i Crne Gore počev od sredine 80-ih godina, dok je širenje u Rumuniju počelo sredinom 90-ih.

Trend broja novodijagnostikovanih HIV slučajeva u Srbiji ukazao je na ranu eksponencijalnu fazu rasta povezanu sa MSM rizikom od transmisije, u kontinuitetu do 2030. godine. „Join-point" regresiona analiza je ukazala na različite trendove u pogledu incidencije novodijagnostikovanih slučajeva HIV infekcije u zemljama Balkana, pad je uočen u Rumuniji, Grčkoj i Sloveniji i povećanje u Srbiji, Bugarskoj, Hrvatskoj i Crnoj Gori, uglavnom povezano sa rizikom od prenošenja MSM-a u celom regionu.

Zaključak: HIV epidemija u Srbiji se nalazi u fazi eksponencijalnog rasta, pre svega vezano za MSM transmisioni rizik, i pokazateljima da su mladi muškarci sa pratećim seksualno prenosivim infekcijama glavna ciljna grupa za planiranje strukturisanih preventivnih politika. Ključnu ulogu u širenju HIV podtipa B na Balkanu ima formiranje lokalnih transmisionih klastera, dok se prekogranični klasteri javljaju uglavnom između zemalja bivše Jugoslavije, sa nastankom i nakon raspada te zemlje. Višekratni unos HIV podtipa B u Rumuniju koincidira sa promenama u intenzitetu i obrascima migracija stanovništva nakon otvaranja i evropske integracije pojedinih balkanskih zemalja ranih 2000-ih.

Ključne reči: HIV, Srbija, Balkan, podtipovi HIV-a, filogenetička analiza, transmisioni klasteri, filodinamička analiza, filogeografska analiza, analiza latentnih klasa, joinpoint regresija, logističko modelovanje

Naučna oblast: Medicina

Uža naučna oblast: Mikrobi i infekcija

UDK br:

TABLE OF CONTENT

1 Introduction

1.1 Early history of human immunodeficiency virus and acquired immunodeficiency syndrome

Upon its first recognition in the 1980s acquired immunodeficiency syndrome (AIDS), a terminal stage of infection by human immunodeficiency virus (HIV) has become the most common causes of acquired immunodeficiency in the world (1). As one of the leading health, social and economic problems, HIV infection and AIDS have become a subject of intensive research and global surveillance, to find the best preventive measures and novel treatment options to decrease the numbers of new infections as well as the progression of HIV infection to AIDS and to reduce AIDS-related deaths (2, 3, 4). Incredible progress has been made in answering major challenges of AIDS including: HIV virus detection and identification; pathogenesis and routes of transmission are described; reliable point of care as well as laboratory based diagnostic tests became available and highly active antiretroviral therapy (HAART) was developed. However, HIV pandemic remains among the greatest public health challenges globally (2, 3, 4).

First cases of AIDS were observed in 1981 as clusters of pneumonia caused by *Pneumocystis jirovecii* and Kaposi sarcoma among seemingly healthy and young men who have sex with men (MSM), exposed to multiple sexual partners as well as with history of prior sexually transmitted infections (STIs) (5-10). Initially, the novel entity got a name of gay-related immune deficiency (GIRD) connecting the disease with brake-down of the immune system associated with MSM (9,10). Nevertheless, soon afterwards additional risk groups have been observed, including people who inject drugs (PWID), blood recipients and hemophiliacs, newborns of affected mothers, thus the GIRD was officially renamed as AIDS in 1982, but the stigma of MSM related STI remained until this day (11-14).

In 1983, scientists around Luc Montagnier and Françoise Barré-Sinoussi and their research group at the Institute Pasteur in Paris, France, were the first who isolated the causative virus of AIDS, a discovery for which they were awarded the Nobel Prize for Physiology or Medicine twenty-five years later (15). Barre-Sinoussi and co-workers recovered a retrovirus from the lymph node of an individual suffering from lymphadenopathy associated syndrome (LAS), an AIDS-associated condition (16). At the same time the National Institute of Health, Bethesda, USA, announced a similar discovery, the retrovirus HTLV-III. The two viruses have subsequently been recognized as identical and the International Committee on Virus Taxonomy announced that the virus that causes AIDS will be known as Human Immunodeficiency Virus (HIV) rather than HTLV III/LAV (17, 18). The initial HIV virus isolate, from 1983, became the prototype of what was later designated as HIV type 1 (HIV-1) group M (HIV-M) and is largely the virus type responsible for the current pandemic.

Generally speaking, the early response to HIV spread was marked with noble and brave examples of activism, previously unseen, however, organized national and global response was lagging behind (2). Due to the accelerating spread of AIDS, this disease was the subject of a session of the United Nations in 1987 and on December 1, 1988, the first World AIDS Day, was promoted. The goal of World AIDS Day has been to promote the need for continuous availability of medicines, the importance of testing, as well as changes in consciousness about people living with HIV and the public's attitude towards them. However,

somewhat delayed global response has been described as *"a story of wasted time and opportunities, of failure of leadership, of denial and discrimination"* (2).

Of note, although the first cases of HIV infection to be registered and described in scientific literature came from urban centers in the USA, making the early history of HIV spread a "Western problem", it is important to highlight that developing nations have been and still are regions with the highest rates of HIV/AIDS morbidity and death, especially in Sub-Saharan Africa and Eastern Europe and Central Asia (3).

1.2 HIV Pandemic

According to the latest estimates of the United Nations AIDS Program (UNAIDS), worldwide, around 40 million people have been living with HIV infection in 2020 and estimated 1.5 million individuals in the world became newly infected with HIV in 2020, with the majority of them aged 15 and older (19). Over the last decade, there has been a gradual reduction in the number of new HIV infections and steady global reduction of AIDS-related deaths, depicting overall progress in fight against the HIV pandemic (3, 19, 20) (Figure 1). Globally, the vast majority of people who live with HIV (PLWH) are in low- and middle-income countries and in 2020, estimated 60% of new infections have been in sub-Saharan Africa. Prevailing risk factors for HIV infection differ between the regions. Globally, recent incidence trends of HIV infections are painting a picture of two types of HIV epidemics: in sub-Saharan Africa generalized epidemics prevail, with young females and heterosexual contacts as the main driving forces, whereas compartmentalized epidemics within key populations such as MSM, PWID, sex workers etc. predominate in the rest of the world (3, 19). Furthermore, in some societies, men are less likely to seek HIV related services due to stigma and marginalization of key populations where men are over-represented, or due to traditional gender norms of male stoicism as seen in Africa, where young females are majority of registered HIV cases (3, 19). Therefore, key populations tailored preventive policies seem to be effective in reducing the number of new HIV diagnoses. Important lessons could be learned from the experiences of several African countries, where settings with good HAART coverage in infected girls and women, combined with preventive circumcision in men resulted in a decline in new HIV infections in recent years, with initial steeper decline in HIV incidence in men (3). In this case, men seem to be the indirect benefactors of reduced population viral load in women as well as adverse effects of circumcision to transmission (3).

In the last decade the highest reduction in number of new HIV infections of around 43% has been achieved in sub-Saharan Africa, whereas the highest increase in the number of new HIV infections of 43% between 2010 and 2020 has been observed in Eastern Europe and Central Asia (20). In the period 2010-2020 a reduction of 47% in AIDS-related mortality has been observed globally, with a reduction of 50% in AIDS-related deaths in sub-Saharan Africa and an increase in AIDS related deaths of 32% in Eastern Europe and Central Asia. (20). Generally speaking, stigmatization from the environment and social exclusion as well as discrimination from the health care system delay measures of secondary prevention (testing) as well as tertiary prevention (linkage to HAART). Those two processes are pillars of treatment as prevention (TASP) as well as pre-exposure prophylaxis (PrEP), since they are highly dependent on high testing rates in key populations as well as level of trust between key populations and care providers (3).

To end the HIV pandemic UNAIDS introduced an ambitious initiative in 2014 called the 90-90-90 initiative, intending to achieve specific metrics concerning the testing treatment

cascade by 2020. Namely, 90% of PLWH should know their HIV status of which 90% should be on HAART with a 90% of virological suppression rate (20). In spite of significant progress achieved in reducing AIDS-related mortality and, somewhat encouraging global trends regarding "90-90-90" initiative goals in 2020: i) 84% knew their HIV status; ii) 87% of all people who knew their status were accessing ART; iii) 90% of all people receiving HAART had viral suppression, elimination of HIV infection as a public health problem by 2030 most likely will not be achieved (19) (Figure 1). Nevertheless, taking the available data into consideration, UNAIDS introduced even more ambitious initiative regarding the testing treatment cascade called 95-95-95 to end the HIV pandemic by 2030, warranting a further need for investment in testing and HAART coverage, especially in key populations (21).

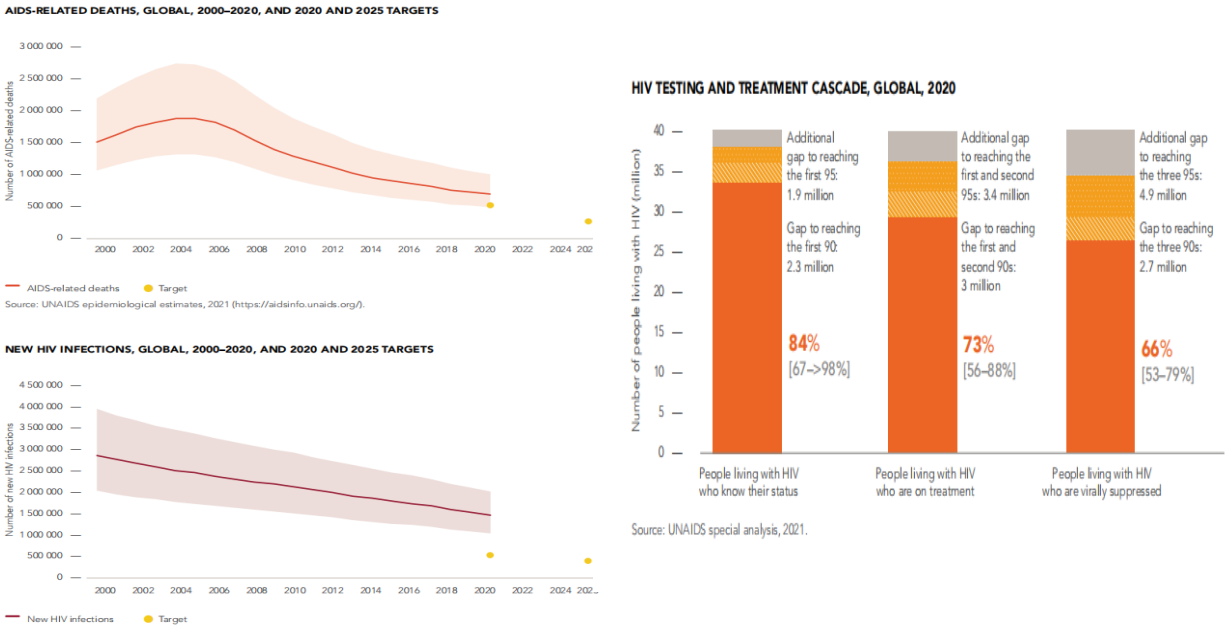


Figure 1. Global trends in AIDS related deaths and HIV infections for 2000-2020 (left), and 90-90-90 testing and treatment cascade (right).
* Yellow dots in upper graphs depict projected goals in AIDS related mortality and new HIV infections for 2020 and 2025.
Source: UNAIDS 2021 (19)

1.2.1 HIV Epidemic in Europe

In Europe, over 2 million people are living with HIV and the number of newly diagnosed cases of HIV infection registered in Europe in 2020 was around 104.000, according to the World Health Organization (WHO) and the European Center for Disease Prevention and Control (ECDC) (22). HIV/AIDS surveillance in Europe is performed in view of WHO European regions: West, Centre and East, encompassing also Central Asia (22).

The majority of new HIV cases in Europe continue to be registered in the East, with an incidence of about 32.6 per 100,000 inhabitants. In Western Europe, the incidence was about 3.7 and the lowest was in the Central Europe, 2.3 cases of newly diagnosed HIV infection per 100,000 inhabitants (22). The dominant transmission mode in West and Center was sex between men whereas heterosexual transmission and PWID were dominant mode of transmission in the East of the continent (22).

As mentioned previously, according to the latest UNAIDS data, the number of new HIV infections in Eastern Europe and Central Asia has increased for 43% between 2010-2020. Newly infected individuals mostly come from the key populations that are disproportionately more affected by disease and often lack access to the HIV services they need (19). However, when comparing the data on HIV incidence and AIDS related deaths in 2019 and 2020, the number of registered new cases seem to decrease, yet, paired to higher AIDS mortality rates (3, 19). This discrepancy indicates that HIV related care suffered dramatically in the pandemic year. These data are worrisome since HIV/AIDS response in Eastern Europe has been considered insufficient already before 2020 and it has been lagging behind, when compared with the rest of the continent and global statistics (19, 22). This lagging has been explained by compartmentalization of the epidemic within key populations, who are facing stigma and discrimination even from the health care system. Indeed, in 2020 as well, the majority of new HIV infections happened in PWID and sex workers and their clients (75% of infections). Furthermore, 16% of infections are among MSM and 9% of infections happen within the general population, however when compared with gender distribution to the western part of the continent the proportion of men is lower and currently 51% of infected individuals are men (51%) (19, 22).

As a consequence, the "90-90-90" goals has not been achieved, , it has been estimated that in Eastern Europe and Central Asia about 70% of people living with HIV knew their serostatus, of which 52% were on treatment with a 50% response rate, according to the UNAIDS report for 2020, implying even further deterioration compared to the preceding one (3, 19). The main problem in the region seems to be an insufficient linkage of key populations with stigma-free, sustainable and high-quality HIV related services, from screenings to the treatment initiation and regular follow-up as well as over-dependence of preventive programs on international donors. (3).

Regarding the high-income European West, a reduction of 11% in new HIV infections with a 30% reduction in AIDS related mortality has been observed. Nevertheless, even in spite of being closer to meeting the "90-90-90" targets, data showing that among people living with HIV 90% knew their serostatus, 84% of them were on treatment with an average viral suppression attained in 75%, which is an improvement compared to the previous report (3, 19). Vast majority of new infections (96%) were happening within key populations, 75% among MSM (19).

There are stark differences between the central and western parts of Europe .Namely, 23 countries in the western part of Europe had observed a 30% decline in newly diagnosed HIV infections, while annual HIV infections in 14 countries in central Europe had seen a rise by 45% over the same period (3). Central Europe is considered as a low incidence region, nevertheless, with the propensity to change, since national epidemics in Central Europe have a potential for significant further growth and increase in incidence (2-4, 18). Moreover, in Western and Central Europe a relatively high rate of late presenters was detected and it was highest in heterosexual men, women and PWID: 62%, 54%, and 55% respectively and lowest in MSM (41%) (23).

1.2.2 HIV Epidemic in Serbia and the Balkans

Duration of the HIV epidemic in Serbia is similar to the one in Western Europe, with the first case registered in 1985. According to 2020 data from the Institute of Public Health of Serbia, a total of over 4300 HIV cases infection has been registered in the country and since 2000s the incidence rate has been about 1.5 per 100,000 inhabitants. Since 2015 an increase in incidence has been noted, until 2020 when a reduction in registered new HIV diagnosis has been observed and incidence rate fell to 1.2 per 100,000 inhabitants, most probably linked to the overload of the healthcare system due to COVID-19 pandemic (22). Initially, HIV epidemic in Serbia spread mostly among PWID, until the mid-1990s, when heterosexual transmission route became more prevalent. In recent decades another shift in the dominant transmission route was observed, leading to the current situation where HIV epidemic is driven by transmission among MSM, comprising the majority of new HIV diagnoses, with the yet ongoing presence of heterosexual transmission (24).

Generally, the HIV epidemic in Serbia shows compartmentalization as described in the rest of Europe and globally with sub epidemics within the national HIV epidemic (19). In Serbia MSM transmission accounts for approximately 80% of new HIV diagnoses and previous molecular studies in Serbia identified transmission clusters related to HIV transmission among MSM, linked to a high prevalence of concomitant STIs (25, 26).

In Serbia, according to UNAIDS estimates, 88% of persons living with HIV know their status and 76% of them receive treatment with a 90% viral suppression rate (3). However, the situation seems much different when data are stratified by key populations, where 53.5% of MSM were covered by testing and know their sero-status and 64.2% of sex workers (19). It has been estimated that PWID are the only key population with sufficient testing coverage of 98.8% (19). Avoidance of health services due to stigma has been reported in 19% of MSM population and 8.4% of sex workers (19).

Within the MSM population, several vulnerable categories can be found: young men (especially underage boys), MSM involved in sex work (which is illegal), and bisexual men (in particular men who define themselves as heterosexual but who have sex with other men). This latter group may also be less likely to be informed (27).

Located in southeast Europe, the Balkan peninsula represents a crossroad between Central-East Europe, the Mediterranean and the Middle East. Balkan states differ in population size, demographic and economic parameters, and HIV epidemics in the Balkans have been described as very heterogeneous, regarding epidemic size, incidence rate, HIV subtypes distribution and predominant transmission mode (28). Nevertheless, the region involves common social and cultural ties with substantial cross-border interaction and mobility.

Romania has the highest number of registered cases in the region; with around 25,590 people living with HIV and the incidence rate is 2.3 per 100,000 inhabitants (22). Even today in Romania around 7000 active cases of HIV infection can be directly connected to an iatrogenic transmission in the late 1980s and early 1990s, and since 2007 an increased number of HIV infections among the adult population has been seen. It is estimated that the most frequent route of infection since 2009 in Romania was MSM (around 40%), but since 2011 an outbreak of HIV infections has been seen among people who inject drugs (PWID) (29, 30).

In Greece, according to data from 2020, around 16,762 people live with HIV and the incidence is 5.6 per 100,000 inhabitants (22). The dominant transmission path in Greece is sexual contact, predominantly among men who have sex with men (MSM), with an increase in the number of cases of infection through intravenous drug use since 2011(32-34).

The epidemiological situation in Croatia and Slovenia is similar to the one in Serbia with the total numbers of people living with HIV of 1810 and 955 respectively and the estimated incidence rates per 100,000 inhabitants are 1.9 and 1.3 respectively (22). The dominant modes of transmission is sexual contact predominantly among MSM (35, 36). In Montenegro cumulative number of HIV cases was 318 with an incidence rate of 2.4 per 100,000 inhabitants. Similarly, as in Serbia, the dominant transmission mode is sexual mainly MSM contact (22,).

The number of HIV infections in Bulgaria is 3,515 with an incidence rate of 2.9 per 100,000 inhabitants (22). Similarly, to national epidemics in Serbia, Croatia, Slovenia and Montenegro dominant transmission route is sexual contact between men as well as imported infections acquired through heterosexual contact mainly through sex work (37, 38).

When it comes to the "90-90-90" testing and treatment cascade all studied Balkan countries seem to have similar results with an exception of Montenegro where a significantly lover testing rate is observed (3, 19, 39). Namely, in Croatia, 79% of people who live with HIV know their status of which 75% are on treatment with a 90% suppression rate and in Bulgaria 85% knows its HIV status and of them, 59% are on treatment with 86% suppression rate. A similar situation is found in Romania, Greece and Slovenia with 86%, 83% and 68% of people who live with HIV know their status and 84%, 69% and 80% are on the treatment respectively and with about 80-99% suppression rates (3, 19).

Contrariwise, in Montenegro, an about 60% of HIV positive individuals know their status, significantly lower than the European average, which could be partially explained by low testing coverage and 50% of diagnosed patients received therapy with about 50% of response rate (19).

1.3 Biology of HIV infection

1.3.1 HIV structure and replication

Belonging to the Retroviridae family, Orthoretrovirinae subfamily, and Lentivirus genus (17), HIV shares common features to other lentiviruses, including a fundamental genome organization, life cycle and capacity to circumvent host immunity, contributing to the slowly evolving chronic infections. HIV genome is composed of two identical single-stranded positive sense RNA molecules. A central part of HIV life cycle is the process of reverse transcription of viral RNA into double standed proviral DNA, with further integration into the host genome, establishing latent infection.

HIV particles are enveloped, spherical, of around 100 nm in diameter. Viral membrane contains heterodimeric glycoprotein complex, composed of an envelope glycoprotein gp120 (SU) and a transmembrane glycoprotein gp41 (TM), and lined on the inner side by the matrix protein p17. (40). Viral capsid is conical and composed of the major structural protein p24. Inside of it, there are two identical copies of positive sense single-stranded viral RNA, a p9 nucleoprotein, and viral enzymes: reverse transcriptase (RT), integrase (INT), and protease (PR) (Figure 3) (40).

HIV genome is of 9.8 kb in size, it is characterized by the presence of structural genes *gag, pol* and *env* as well as 6 loci that encode regulatory proteins, with repetitive regulatory sequences located at both ends of the HIV genome (5' and 3') (Figure 2).

Pol region encodes the enzyme machinery of the virus. Namely, the product of the *pol* genetic region is a product which upon splicing makes for RT, PR and INT, playing a central role in the virus replication (41). As RNA/DNA polymerase, RT enzyme plays a vital role in the life cycle of HIV. RT uses both viral RNA and pro-viral DNA as a template and also acts as ribonuclease H, it can dissolve the hybrid DNA/RNA helix. INT, plays a role in editing the terminal part of pro-viral DNA and its integration in the host genome, through the endonuclease activity of the enzyme – a unique feature of retroviruses. PR performs splicing of primary polyprotein translates, in order to create the final structural and functional proteins of viral particles (Figure 2).

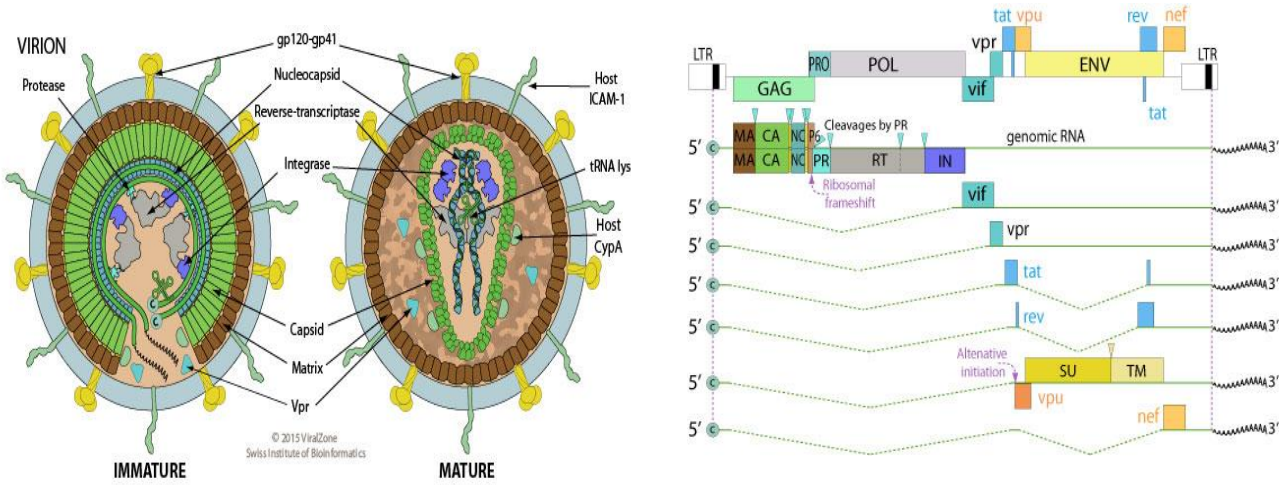


Figure 2. Diagram of HIV viral particle (Right) Diagram of the genome organization of HIV (Left) (adopted from Viral Zone available at: https://viralzone.expasy.org/7?outline=all_by_species)

Gag gene encodes polyprotein of 55 kd (p55), which makes for the following structural proteins: matrix antigen (MA) p 17, capsid antigen (CA) p24 and nucleopeptides (NC) p6 and p9 (41). *Env* gene encodes a molecule of 160 kd, which codes a products of viral envelope surface glycoproteins: gp 41 and gp 120. Chemical bond between gp120 and gp41 is a non-covalent one, allowing for gp120 to be spontaneously released and could be detected in the local environment, in serum and lymphoid tissues of HIV-infected patients (41).

The six HIV accessory genes are *tat, rev, nef, vif, vpu* and *vpr*, encoding regulatory proteins. Protein Tat interacts with the promoter in the long terminal repeat (LTR) sequence to stimulate proviral DNA transcription and is important in early phases of viral replication cycle (42). Through interactions with Rev responsive elements (RRE) contained in viral RNA Rev protein stabilizes viral mRNAs, facilitating their transport through the nucleus (43). Nef gene encodes a protein with two main effects in infected lymphocytes: T cell activation and establishment of a persistent infection. Nef is an inhibitory transcription factor with a role in promoting the survival of infected cells. Its role in HIV persistence is also based on the ability of Nef to down-modulate the surface levels of important molecules at the immune synapse such as major histocompatibility complex I (MHC I) and (MHC II) present on antigen-presenting cells (APCs) and target cells, and CD4 and CD28 present on helper T cells, a common immunity avoidance strategy of viruses (44, 45). Viral infectivity factor (Vif) helps to counteract Apolipoprotein B mRNA editing enzyme catalytic polypeptide like 3G (APOBEC3G, A3G). The APOBEC family of proteins act as editing enzymes and cause Cytosine to Uracil editing, leading to the accumulation of Guanine to Adenine mutations in the proviral sense cDNA strand. Vif also interacts with Gag polyprotein to modulate the PR mediated proteolytic processing (46). Viral protein R (Vpr) is responsible for nuclear transport of the HIV1 preintegration complex (PIC) and plays an important role in the extracellular release of viral particles (46). Viral protein U (Vpu) is a membrane-bound helper protein with two major roles. Firstly it downregulates the CD4 receptors within the endoplasmatic reticulum to prevent the interaction between CD4 and gp120 and secondly, it promotes the release of viral particles (46)

Initial step in HIV replication involves interaction between viral gp 120 and cell receptors (the main HIV receptor is CD4 molecule and co-receptors are chemokine receptors CCR5 and CXCR4). The gp 120 molecule is structured in several variable and conserved domains, of which the constant domain C3 and variable domain V3 have the most prominent role in viral replication. The C3 domain functions as a ligand for the CD4 receptor, which is expressed on the cell surface of approximately 60% of circulating T-lymphocytes, T-precursor cells in the bone marrow and thymus, monocytes/macrophages, dendritic cells, and microglial cells in the central nervous system, whereas the V3 domain V3 influences viral tropism for two main co-receptors CCR5 or CXCR4 (40).

Finally, as mentioned above gp41 also has an important role in membrane fusion by providing hydrophobic glycine-rich domains, thus facilitating the viral entry. The trimmer complex, composed of heterodimeric proteins gp120 and gp41 and CD4, is essential for virus recognition and entry into target cells and after tying gp120 with CD4 protein, the virus undergoes a structural change, exposing a co-receptor biding places such as V3 domain (47).

Then, the capsid is taken up by the endosome and its content is partially released into the cytoplasm, next RT initiates the conversion of viral RNA to DNA. This process is taking place in the cytoplasm and the first product of this reaction will be an RNA-DNA hybrid helix. At this stage RT initiates its ribonuclease activity thus releasing the single-strand DNA from the hybrid helix so the polymerization can be completed. The result of these reactions is a double-stranded DNA (dsDNA) molecule, which integrates itself into the host genome using the enzyme INT (40).

The INT removes nucleotides from both 3 'ends double helix DNA allowing their binding and making them sticky which is manifested as circular double-strand DNA. Such DNA is then transferred into the nucleus and is permanently integrated into the human genome as pro-viral DNA (48). Throughout infection, this cycle is repeated, creating numerous strains of pro-viral DNA that can be deposited into the genome making the so-called pro-viral archive of the virus (49). Expression of pro-virus requires the target cell to be in the activated state but also resident monocytes/macrophages, microglial cells and infected CD4 + T cells are important long-term cellular reservoirs of HIV, with potential for clonal expansion even during latency phase of the infection (40).

Transcription of proviral DNA into messenger RNAs occurs upon activation of the infected cell,. Firstly, regulatory genes such as tat and rev are getting transcribed. Then these regulatory proteins, namely Tat protein, interact with the TAR site on the promoter LTR at he beginning of HIV RNA in the nucleus and stimulate transcription and formation of longer secondary RNA transcripts such as *pol, gag* and *env*. Furthermore, Rev protein promotes transcription and stabilization of longer mRNAs, the expression of structural and enzyme genes, and the inhibition of regulatory proteins. Structural mRNA is then transported into the cytoplasm, where they are translated using cell ribosomes. These primary polypeptide chains are precursors of structural and functional proteins, which are spliced by the PR(41).

Finally, the assembly of new viral particles is a step-by-step process. Firstly, two viral RNA chains are linked together alongside with the enzymes, while the capsid are in the making around. Secondly, this immature particle migrates to the cell surface. New particles then pass through the cell membrane of the host, thereby acquiring a new envelope. During the bidding process, the lipid membranes of the virus are getting the gp41/gp120 proteins,

but additionally, other host receptors may become a part of a novel viral envelope as well as other essential molecules of the host cell such as cholesterol (41).

The replication cycle of HIV is depicted in the Figure 3.

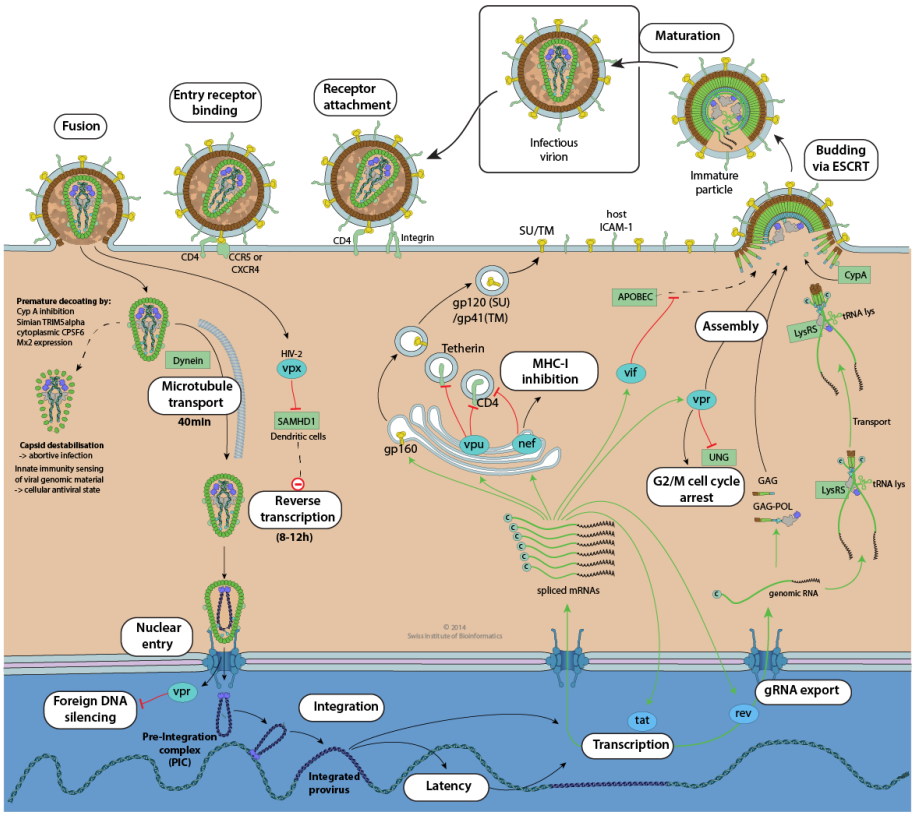


Figure 3. The replication cycle of HIV. (adopted from Viral Zone available at: https://viralzone.expasy.org/7?outline=all_by_species)

1.3.2 HIV diversity

Major feature of HIV is extremely high genetic variability. This heterogeneity is linked to a number of mechanisms: lack of proofreading function of HIV-1 reverse transcriptase, paired with a very high replication rate, leads to significant in-host variability and formation of quasispecies – a swarm of genetically closely related yet diverse strains; further on, HIV replication involves high recombination frequency. Host environment is also an important driver of HIV diversification since high variability seems to be the main strategy of both immune response and therapy evasion (50-51).

To date, two types of HIV have been identified; HIV type 1 (HIV-1) in 1983 and HIV type 2 (HIV-2) in 1986 (17). The two virus types represent a distinct viral origin - host switch of two simian immunodeficiency viruses (SIV) from different primate reservoirs: chimpanzees and sooty mangabeys (Figure 4) (52). The destiny of two viruses in human host took different paths: HIV-1 being responsible for the HIV/AIDS pandemic, while HIV-2 infection remains predominantly restricted to West African countries, such as Guinea-Bissau, Gambia, Senegal, Cape Verde, Côte d'Ivoire, Mali, Sierra Leone, and Nigeria. Notably, an increasing

number of cases have been recognized in Europe (mainly in Portugal and France) and other sites, mostly having historical ties to West Africa (53).

HIV-1 is classified into four monophyletic genetic groups or lineages: main or major (M), new or non-M, non-O (N), outlier (O) and pending (P), each of which represents independent cross-species SIV transmission event from chimpanzees and gorillas. Genetic lineage M is the most dispersed globally and the most diverse, creating distinct monophyletic clades, which are classified as subtypes (Figure 4) (52).

HIV group M is currently divided into nine "pure" subtypes or non-recombinant forms (A-D, F-H, J and K). Circulating strains of A and F subtypes are further subdivided into A1-A6 and F1/2 sub-lineages, similarly, phylogenetic analysis revealed that subtype B and D could also be seen as sub-lineages of one larger subtype based on nucleotide distance but for historical reasons B and D clades are still called subtypes (40).

Besides "pure" subtypes, recombinant forms have also been identified, by convection divided into two groups: i) circulating recombinant forms and ii) unique recombinant forms. Circulating recombinant forms are present in the population at least in two epidemiologically unrelated persons whereas unique recombinant forms are isolated from one patient and they are not present in the population (two or more epidemiologically unrelated individuals). To register novel circulating recombinants a full genome sequencing is required and thera are currently, over 100 HIV-1 recombinant forms are described (https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/crfs.comp).

The evolution of HIV is permanently divergent and rapid and to quantify this evolutionary divergence in the era of molecular phylogeny, nucleotide distance is one of the main parameters that needs to be calculated. The level of nucleotide divergenvce between HIV-1 and HIV-2 has been quantified to around 48.3%, 37.5% between HIV-1 groups, 14.7% between HIV-1 subtypes, 8.2% within HIV-1 subtypes (54).

Extreme variability of HIV and its simian precursors was necessary for host switch, to avoid immune response, but this characteristic of HIV is especially important in the era of HAART, which creates intensive selective pressure.

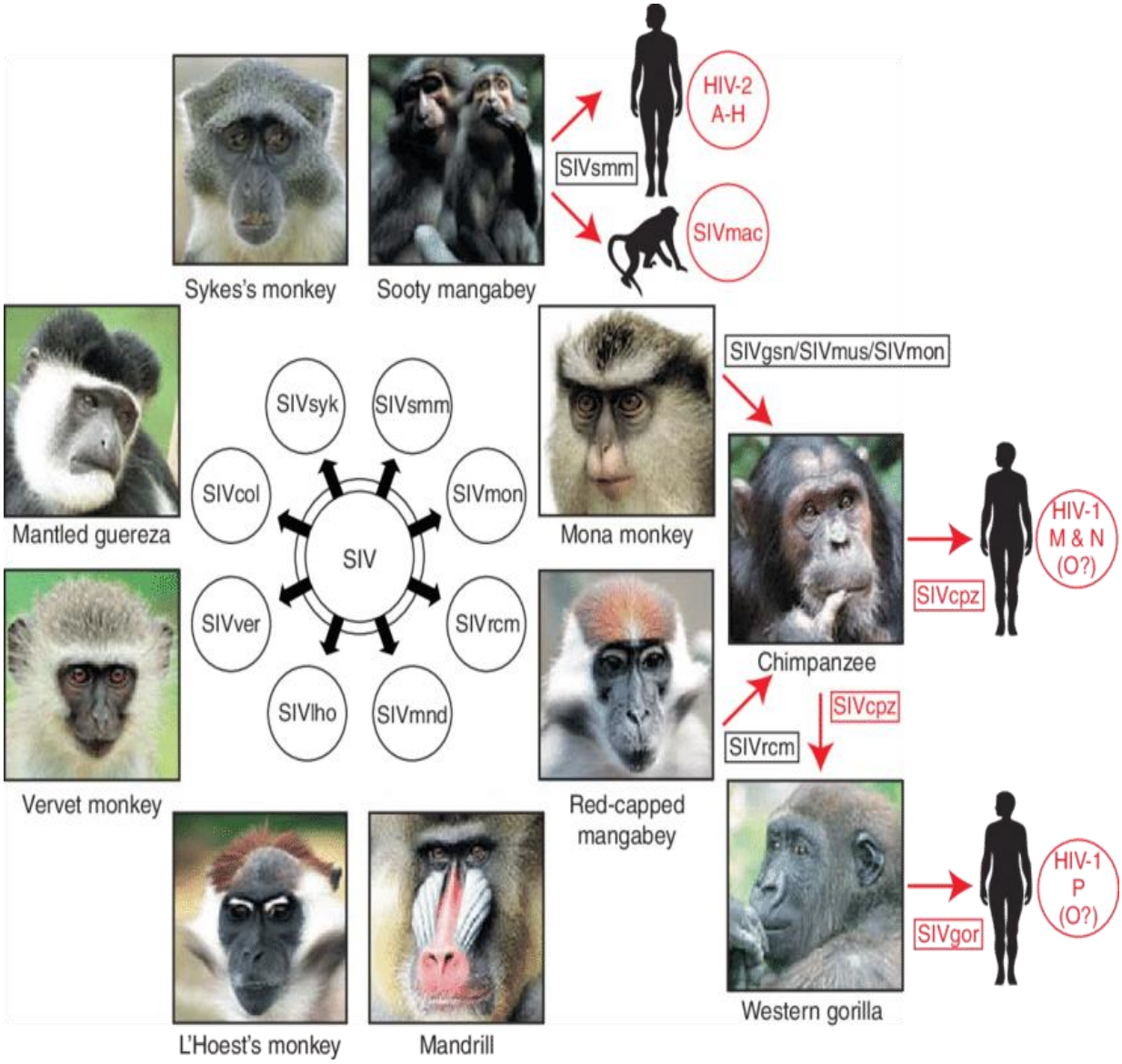


Figure 4. The early evolutionary history of HIV-from SIV ancestral strains to modern epidemic and pandemic HIV strains Red highlights indicate known instances of interspecies transfers and the associated viruses.

Source: Sharp PM, Hahn BH. 2011. (52)

1.3.3 HIV-1 Subtype Distribution

In Africa the greatest HIV diversity has been found, for example in South Africa dominant subtype is subtype C whereas in Eastern Africa subtype A is most dominant with subtype D and C co circulating as well. Increased prevalence of these subtypes is seen in Kenya, largest of the East African countries. In Uganda, however, a significant prevalence of circulating recombinant forms is found at 46%. In West and Central Africa AG recombinants predominates, which is to be expected taking into account that both subtypes A and G are present as co-dominant subtypes in Nigeria for example. Furthermore, an increase in F2 prevalence is seen in Cameron (55).

Similarly, as in Africa, great virus diversity can be found in Asia as well, and this region is marked as a melting pot of circulating recombinant forms such as BC in China and AE recombinants present in China and the rest of Asia. On the other hand, AG recombinants are common in Kyrgyzstan, whereas subtype A is the most common in the former Soviet political space of Eastern Europe and Central Asia. Similarly, subtype C is the most common in the Indian subcontinent, with reports on unique recombinant forms (55). On the other hand, Latin America, the Middle East and North Africa, Europe, USA and Australia are parts of the world where subtype B is traditionally predominant (55).

Nevertheless, mass migration including the refugee crisis as well as economic migrations, as a result, could have an impact on changing patterns of subtype distribution in some developed European countries (56-60). Furthermore, several studies analyzed the risk of HIV infection in migrant populations before and after immigration. It is noted that isolation and low integration increase the risk of HIV infection and the rate of late presenters in migrant populations (61).

Better diagnostic coverage of migrants originating from the Middle East and Africa in several western European countries provided novel data on circulating strains within the migrant population. A bigger influx of immigrants that has been present since the mid-2010s changed the HIV subtype distribution in several European countries, mainly through multiple introductions that created local transmission clusters including both migrant and local populations, as seen in Italy, Portugal, Germany, Belgium and Luxembourg (57, 58, 60). Of note is also the recent expansion of F1 subtype clusters in several European countries including Spain and Switzerland, connected to Romania as an origin (59).

Regarding the distribution of circulating HIV-1 subtypes in the Balkans, a broad diversity is seen (28). In countries of former Yugoslavia, the dominant HIV subtype is B, including Serbia, Slovenia, Croatia and Montenegro (90%, 83%, 90% and 84% respectively) (25, 26, 35, 36, 62, 63). Of note, it is found that this subtype is prone to cluster formation mainly within MSM sub-population. Furthermore, taking into the consideration that MSM transmission remains a prominent mode across these countries a dominance of HIV subtype B should not be surprising (64). Non-B subtypes were less prevalent, associated with other modes of transmission including heterosexual contact, with limited local transmission and clustering rates (25, 26, 35, 36). However, in the resent years subtype A is being more prevalent within newly diagnosed individuals and it is found in 23% in Montenegro, 8.3% in Slovenia and 4.2% in Croatia (35, 36, 62).

So far, molecular studies have shown that in Serbia subtype B is a dominant clade, present in about 91% of patients, whereas the non-B subtype accounts for about 9% of patients. In a recent study by Siljic et al most frequent non-B subtypes were subtype C with a prevalence of about 3%, whereas subtype A and subtype G accounted for about 1.5% and recombinant forms accounted for 3% (26).

In contrast in Montenegro, a recent analysis by Dragas et al revealed a changing pattern in subtype distribution in recent years with a reduced prevalence of subtype B of about 70% and an outbreak of subtype A present in 23.3% (62). Interestingly, molecular clock analysis dated the subtype B epidemic to be in the mid-80s whereas the subtype A epidemic was dated much more recently in 2014 indicating that the outbreak of subtype A in the country is a novel occurrence, connected with MSM transmission (62). In contrast subtype, C was present in a much lower prevalence of about 5% of cases (62).

In Croatia subtype B is present in about 91% of cases, similar to in Serbia and non-B subtypes are present with a 9% prevalence. Subtype B was usually connected with MSMs, male gender and local clustering whereas non-B subtypes had a lower prevalence of clustering and they were connected with female gender, heterosexual transmission (36). Of non-B clades in recent years subtype A was the most prevalent, and phylogenetic data indicate the spread of subtype A within MSM and relatively high clustering prevalence of about 40% and frequent evolutionary connections with sequences outside Croatia indicating a potential for further spread of this clade and complex origin marked with multiple introductions (36). On the other hand subtype C was the second most common non-B clade connected with multiple introductions as well as heterosexual contact and low clustering rate indicating limited spread (36).

A similar situation is found in Slovenia where subtype B was most prevalent, albeit with a lower prevalence of about 83%, whereas non-B subtypes were present in about 17% of cases. Subtype B clade was associated with local clustering, MSM, as well as frequent partner changes whereas non-B clades were connected with females, and foreigners (35). Subtype A is the most prevalent non-B clade present in 8.3% of cases and subtype C and recombinant forms were present in 2.4 % and 1.8 % respectively (35).

In Bulgaria, subtype B is present in 45-50% of infected individuals, whereas CRFs are present in about 40% of cases, especially CRF01_AE. However, other subtypes were present with lower prevalence according to 1986-2009 data: Subtype C with 3.5% prevalence, subtype H with 2% prevalence, subtype A with 2.5% prevalence and subtype F with 1% prevalence. (36, 37). On the other hand, a recent analysis included the data of newly diagnosed individuals from 2012-2017 and it revealed a changing pattern in subtype distribution where subtype A and subtype F increased in prevalence equaling 14% and 7.4% respectively (65). However recent reports revealed that subtype B is still the most prevalent subtype in Bulgaria connected to the MSM and multiple routes of introduction from different Western European countries and Israel since the late 80s (66). CRF 01_AE is the second most common clade in Bulgaria connected with PWID and heterosexual transmission. Of note is the low prevalence of MSM in this epidemic except when MSM/PWID risk is reported (67). On the other hand subtype, C is significantly less frequent although the HIV epidemic in Bulgaria started with heterosexual transmission and subtype C. This spread, however, was limited in the population, which is to be expected due to the mode of transmission (68).

In Greece, traditionally, two dominant HIV clades are subtype B and subtype A. Initially subtype B was found in about 60% of HIV cases but its prevalence was diminishing and according to the most recent data is about 40% (34, 56). The prevalence of subtype A is about 30% and it was introduced from Sub-Saharan Africa in the late 1970s as the consequence of a single founder event, which is much earlier than circulating A clades found in other Balkan countries. Subtype C and CRFs were less frequent with 2.8% and 15% prevalence respectively. The rest of the cases were unclassified (34, 56).

On the other hand, according to the available data, in Romania subtype B accounts for about 8% of newly diagnosed HIV infections and therefore it is an emerging sub-type in the country, with subtype F as the most common, mainly connected to the iatrogenic transmission of HIV in the 80s and 90s (69).

Although there is a high HIV-1 diversity in different Balkan countries and although this diversity is ever evolving subject, the omnipresence of HIV subtype B stays a prominent future of HIV epidemic in the Balkan region. It is the prevailing one in some Balkan countries (Serbia, Bulgaria, Croatia, Slovenia and Montenegro), whereas in others it is still an emerging HIV subtype (e,g, Romania), instigating the question of relatedness and clustering of the epidemics across different Balkan countries (Figure 5).

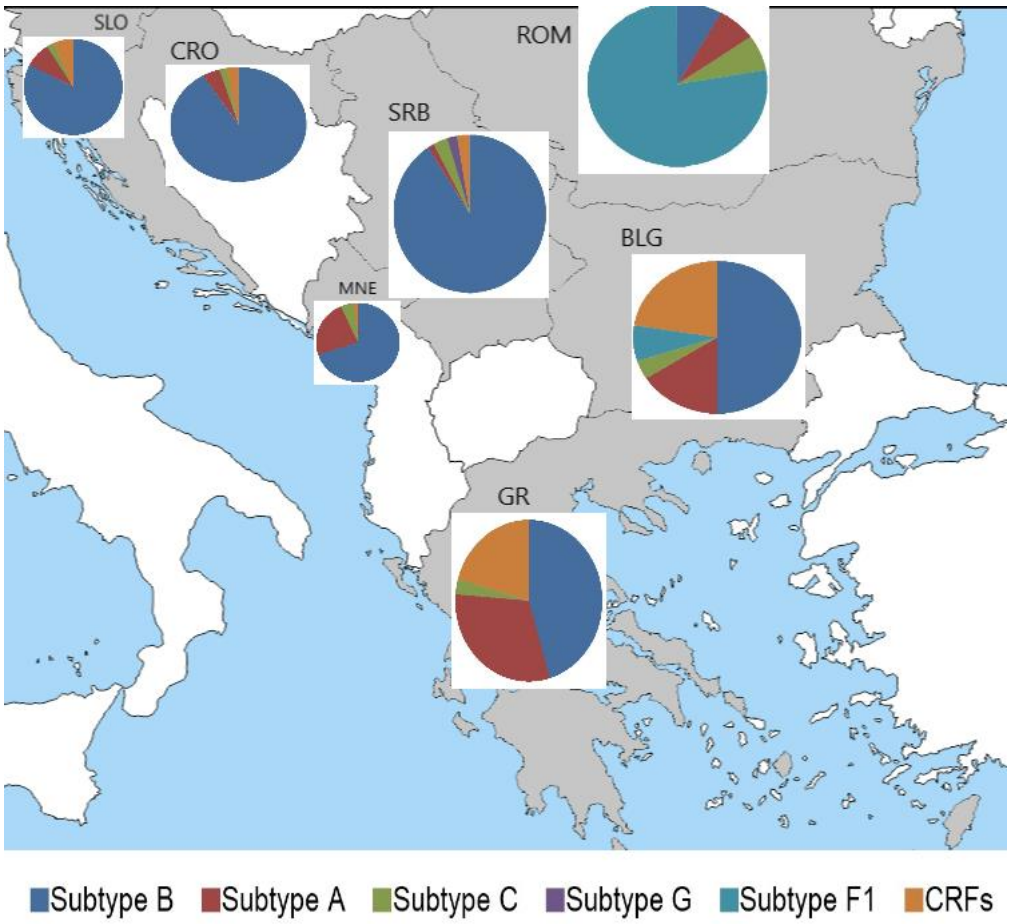


Figure 5. Distribution of HIV subtypes in the Balkans. Of note is high diversity of HIV subtype distribution, based on the literature data (25, 26, 34, 35, 36, 62, 63,36, 37, 56, 65, 69) Abbreviations for the presented Balkan countries: Romania (ROM), Greece (GR), Bulgaria (BLG), Serbia (SRB), Croatia (CRO), Slovenia (SLO), Montenegro (MNE).

1.4 Phylogenetic analyses in HIV research

Phylogenetic analysis played a crucial role in unraveling origins of HIV. Evolution of HIV as a human pathogen may be divided into several periods: cross-species transmission, early spread in Central Africa, global cryptic spread of HIV-1 group M, detection of first cases of HIV in the 1980s, local transmission clusters formation and establishment of national HIV/AIDS epidemics (70).

Among the major milestones was detection of the most recent common ancestor of HIV. Namely the analysis of SIVs obtained from different primate species and of HIV-1 sequences from Central African countries revealed that multiple different cross-species transmission events resulted in current HIV-1 diversity as described above, with specific populations of gorillas and chimpanzees endemic to Cameroon and other surrounding countries, marked as the initial reservoirs. Since then, HIV-1 diverged into several monophyletic groups (M, N, O and P) with the M group marked as the pandemic group which is further divided into subtypes as well as different circulating recombinant forms CRFs as mentioned above (52).

Further molecular clock analysis as well as phylodynamic analysis improved the understanding of the cryptic spread of HIV-1 M subtypes throughout Central Africa as well as spread into the USA. Namely, although first cases of AIDS were detected in the early 80s in the USA, the natural course of HIV infections marked with latency begged the question of the real beginning and origin of the HIV pandemic. Namely, genetic evidence indicated that HIV-1 group M was present in the circulation in The Democratic Republic of the Congo (DRC) in the late 1950s, tracing the most recent common ancestor back to the beginning of 20s century in DRC as well (52). Then during the 30s and 40s, the HIV-1 M spread throughout DRC as well as Brazzaville (the major marine port as well as the capital of the Republic of the Congo) (71).

Today it is understood that mainly social factors connected with trade and migration had the major influence on the worldwide subtype distribution of HIV-1 M, but other studies pointed out that subtype-specific differences in virulence and transmissibility had a role on this distribution as well (72).

Due to the absence of proofreading activity of RT and fast replication rate, mutations accumulate relatively fast in the HIV genome. This means that research of local transmission patterns within national HIV epidemics can be effectively studied. Numerous studies, mainly located within high-income countries, have already studied their viral transmission networks, which describes the history of infections at the resolution of individual cases (73). The important premise for such research is that the sampling depth is sufficient, meaning that every single new case of HIV is sequenced. This premise is partially present in the developed countries where the epidemic is smaller and affects specific Key populations. In this sense, most of the studies have used databases compiled by national services for screening of drug-resistance mutations (genotyping), which generate HIV partial genome sequences for virtually every individual entering/failing therapy (74). Then upon phylogenetic tree reconstruction, several approaches can be used when defining transmission clusters.

Phylogenetic research assumes that if two viral strains share common ancestor, they will be a member of local clusters, thus alleviating the problem of information bias, especially when sequence data are paired with clinical data and public health determinants. This

combination of phylogenetic and epidemiological analysis is a powerful method in analyzing the spread and risk patterns of HIV. The results of such studies could further empower and inform local communities and support groups to advocate for resource allocation as well as novel preventive measures (75). But caution is required when interpreting phylogenetic clusters for epidemiological purposes. This is because clusters are typically inferred from partially sampled transmission chains and some infected individuals (intermediary links as well as common ancestors) were not sampled (76) .

Furthermore, by molecular clock algorithms and implementation of various population growth models, phylogenetic analysis got additional roles in HIV research (apart from molecular epidemiology and risk detection) such as: i) the studies of growth/decline of the specific HIV populations (phylodynamic) (77), or ii) the impact of migration on HIV spread and the identification of hubs of transmission (phylogeography) (78).

The phylodynamic analysis uses temporal data and sequence data in order to put the static picture of phylogenetic tree into temporal context, which is made possible by implementing the molecular clock theory. According to this hypothesis, DNA/RNA and protein sequences develop at a generally consistent pace across time and across organisms. The genetic divergence between any two species is proportionate to the period since they last had a common ancestor as a result of this consistency. Therefore, if the molecular clock concept is correct, it can be a very valuable tool for calculating evolutionary timelines (79). Moreover, the temporal context of paraphiletic relations between two patients can indicate the probable route of infection, although the problem of partial transmission chain sampling remains, so vigorous epidemiological investigation is needed to support direct connections between patients (80).

Furthermore, through implementation of Coalescent and Birth-Death models population dynamics can be described by means of effective population or reproductive number changes over time (81). Generally speaking, Birth Death models accounts for both Births and Deaths in the population, on the other hand Coalescent models do not account for deaths in the population. By using these two approaches different skyline plots are generated illustrating changes in reproductive number over time (Re plot) or effective population over time (Ne plot) respectively. The resulting Re plot is a polygonal line which can be above 1 (exponential population growth) and below 1 (slower linear growth or stagnation). On the other hand Ne plot can only grow or to stagnate since Coalescent models do not accounts for deaths in the population. The resulting Ne plot is a sigmoid curve of population growth with initial exponential growth and plateau, where exponential phase corresponds to Re values above 1 and plateau corresponds to the Re values below 1. Therefore both analysis have similar goal in describing phylodynamic pattern of the observed data set and can be used as a way to assess reproducibility of the proposed growth scenario.

The phylogeographic analysis is the other computational method of ancestral reconstruction that utilizes temporal and geographical metadata to infer trajectories of viral spread across different geographical points (82). For example a recent study that inferred global HIV spread from large scale data sets (5133 sequences) proposed insightful evolutionary theories on global spread history for several major HIV subtypes and CRFs (83). The results from this study implied that the current spread of subtype B is fueled by a reciprocal spread pattern between the USA, Europe and Australia, whereas the spread of this subtype into Latin America was unilateral from the USA making the USA a putative exporter. The major caveat of this study was the failure of the analysis to reconstruct the cryptic spread

of this subtype from Africa to Caribbean, which is explained due to the lack of historical sequences from the African region in the data set therefore USA was seen as the current exporter and not as the place of origin for subtype B. On the other hand simulations for C and A revealed the origins in South Africa and Uganda respectively, whereas CRF01_AE originated from Central Africa from where the spread happened to Thailand and South-East Asia where it remained more or less endemic.

Appropriately designed phylogenetic study can reveal undisclosed epidemiological linkages and when done correctly it can have great potential to inform and to empower local communities and key populations to advocate for better prevention strategies. Furthermore, the results of such analysis can be used in courts as a part of legitimate forensic procedure but, at the same time, the results of phylogenetic studies could be misused in order to oppress and to brutalize key populations (75).

2 Aims of the Study

The overarching aim of this research was to investigate evolutionary and transmission dynamics of HIV epidemic in Serbia as well as to analyze cross border transmission clusters and dispersal routes of HIV subtype B epidemic in the Balkans by combining several approaches: mathematical modeling, joinpoint incidence trend analysis, advanced phylogenetic analysis, including phylodynamics and phylogeographics.

Specific aims of the study were:

- To investigate transmission dynamics and phylodynamic patterns of the HIV epidemic in Serbia, generally and in subpopulations related to transmission mode and HIV subtype.
- To reconstruct the mathematical model of the current course and future HIV incidence trends in Serbia.
- To explore the evolutionary dynamics and phylogeographic spread patterns of the HIV subtype B epidemic in the Balkans.

3 Materials and Methods

3.1 Study design and ethical approval

The study was designed as cross-sectional phylogenetic investigation of HIV sequnce data from Serbia and from six of the Balkan countries: Bulgaria, Croatia, Greece, Montenegro, Romania and Slovenia. Additionally, a long term trend analysis of growth of HIV epidemic in Serbia and the Balkans was performed using relevant epidemiological survellance data.

The study was conducted at the Institute of Microbiology and Immunology, University of Belgrade Medical Faculty, in colaboration with HIV national reference centers from participating countries: Depatment for HIV/AIDS, Clinic for infectious and tropical diseasses, University Clinical Centre of Serbia; National Institute for Infectious Diseases "Matei Bals" in Romania; Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens in Greece; National Reference Laboratory of HIV, National Center of Infectious and Parasitic Diseases in Bulgaria; University Hospital for Infectious Diseases "Dr. Fran Mihaljević" and University of Zagreb School of Medicine in Croatia; Institute for microbiology and immuology, Medical faculty, University of Ljubljana in Slovenia; and Infectious Diseases Hospital, Clinical Center of Montenegro.

The study was approved by the Ethical commitee of University of Belgrade Medical Faculty, Decision No 9/VI-3-1.

3.2 Phylogenetic analyses

3.2.1 Study population and sequence datasets

3.2.1.1 Dataset for the study of HIV epidemic in Serbia

The earliest viral sequences from Serbian epidemic date from 1997. For the purpose of this study, data and sequences from the period 1997 to 2011 were gathered from the database of the National reference laboratory for HIV/AIDS at the Institute of Microbiology and Immunology, Medical Faculty University of Belgrade. Complete dataset of these sequences is available at the NCBI database, accession numbers are given in Annex 1.

Additionally, this study included HIV sequences obtained from blood samples collected from 2012 to 2015, from consenting HIV seropositive adults, both therapy naïve and therapy experienced, followed at the Centre for HIV/AIDS, University Hospital for Infectious and Tropical Diseases in Belgrade. Blood samples from HIV-infected patients were sent for drug resistance testing as part of patients' routine follow-up. The molecular procedure of generating HIV sequences is described in the chapter 3.2.2.

Reference sequences of different subtypes, known to be present locally in Serbia and in the Balkan region, encompassing subtypes B, C, G, A, F, as well as circulating recombinant

forms (CRFs), CRF01_AE and CRF02-AG, were downloaded from the Los Alamos database (https://www.hiv.lanl.gov/content/index).

Additionally, NCBI BLAST search was performed for each sequence found to belong to a transmission cluster (according to criteria described in detail further). For each query sequence, five sequences were included in the tree reconstruction based on the highest similarity score as obtained by NCBI BLAST search. In total, 175 sequences were included in the analyses after BLAST search. Furthermore, a group of control/background sequences was also included, in the context of the origin and spread of HIV from the isolates of North America, West Europe, and Balkan, with the clear defined subtype and the time of sampling available at the NCBI database.

The final dataset for the study of HIV epidemic in Serbia included a total of 385 local sequences.

3.2.1.2 Dataset for the study of HIV epidemic in the Balkans

The study dataset consisted of HIV-1 subtype B sequences of the *pol* region obtained through collaboration effort of HIV national reference centers from participating countries. Anonymized sequence data were collected, with the number of included sequences per country aimed to cover minimally 10% of the cumulative number of HIV subtype B infected individuals from each country registered by 2019.

For the purpose of phylogenetic analysis of HIV subtype B spread in the Balkans 53 subtype B sequnces from 2016-2019 were generated from consenting patients on HAART, according to predefined inclusion criteria (HIV-1 seropositive adults (men and non-pregnant women aged 18 years or above). The molecular procedure of generating HIV sequences is described in the chapter 3.2.2. Moreover additonal 358 Serbian subtype B sequences from the database of the National Reference Laboratory for HIV/AIDS at the Institute of Microbiology and Immunology, Medical Faculty University of Belgrade for time period 1999-2019 were included.

In total, 2,415 subtype B partial *pol* region sequences from 1999-2019 were analyzed: 220 from Romania, 770 from Greece, 411 sequences from Serbia, 251 from Bulgaria, 365 from Croatia, 333 from Slovenia and 65 from Montenegro.

Furthermore, reference sequences of different subtypes, known to be present in Balkan obtained from the Los Alamos database were used as well, which is in detail explained in chapter 3.2.1.1.

3.2.2 Molecular procedures of generating HIV sequences

3.2.2.1 RNA extraction from plasma samples

Blood samples from HIV-infected patients referred for drug resistance testing were manipulated at the Institute of Microbiology and Immunology. Upon sample receipt, peripheral blood mononuclear cells (PBMC) and plasma were separated by centrifugation at 5000 rpm (~ 370 x g) for 15 minutes and preserved at −80 ◦C for further molecular analyses.

Prior to RNA extraction, each 1.5 mL plasma sample was centrifuged for 1.30 hours at 4 C at 22,000 g. The pellet was re-suspended and utilized for RNA extraction by QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer instructions.

The samples were then mixed with 560 µl of RNA virus lysis solution (Buffer AVL) containing carrier RNA, vortexed for 15 seconds, and incubated at room temperature for 10 minutes.

Then, 560 µl of ethanol was added and samples were vortexed.

In the next step 630 µl of the solution was put to the QIAamp spin column without wetting the rim and centrifuged for 1 minute at 8000x rpm. The filtrate-filled collection tube was discarded and step was repeated until all of the sample is applied on the filter membrane of the column, thus washing and elution step of the extraction protocol can be done.

500 µl of wash buffer 1 (Buffer AW1) was added to the column, which was centrifuged for 1 minute at 8000x rpm, discarding the collection tube.

And, after 500 µl of wash buffer 2 (Buffer AW2) was added, the lid was closed, and the centrifuge was run at full velocity (14500x rpm) for 3 minutes in order to drain the washing buffer residuals, potentially contaminating the RNA solution.

Then, the QIAamp spin column was placed in a clean 1.5 ml micro-centrifuge tube and 60 µl of elution buffer was added. Upon incubation at the room temperature for 1 minute sample was centrifuged at 8000x rpm for 1 minute to elute the RNA. The viral RNA was extracted and stored at -80°C for subsequent use.

3.2.2.2 Nested polymerase chain reaction (PCR):

Prior to PCR, lyophilized primers were re-suspended in sterile distilled water to a stock concentration of 100 µmol/µl. Primers used for amplifications were further diluted to working concentrations of 20 µmol/µl. Primers were stored at -20 °C until required (Table 1). The region of interest in this study was *pol* genetic region including whole PR (99 codons) and partial RT region of at least 220 codons.

All of the reactions were done in the UV-treated laminar flow hood and RT-PCR reactions were set up on ice. The reaction mix and cycle parameters were usually set according to the manufacturer's instructions, though annealing temperature tuning was occasionally required. For PCR analyses in this research, annealing temperature, for both primer pairs encompassing prot and rt genetic region, was 50 °C.

Breifly, the first step upon the RNA extraction was the first (outer) cycle of nested PCR, one step reverse transcription+PCR (RT-PCR) using PrimeScript One Step RT-PCR Kit (TAKARA, Japan). For inner PCR reaction Thermo Scientific Dreamtaq PCR master mix (2X) (Thermo Fisher Scientific, USA) was used.

Table 1. The list of primers used in the study (https://hivfrenchresistance.org/).

Primers (5'- 3')	*Pol* Region
Outer primers	PR-TAA TTT TTT AGG GAA GAT CTG GCC TTC C RT-AGT AGG ACC TAC ACC TGT CA
Inner primers	PR-TCA GAG CAG ACC AGA GCC AAC AGC CCC A RT-TTG GTT GCA CTT TAA ATT TTC CCA TTA GTC CTA TT

General protocol for One-Step RT PCR was:

1. 12.5µl 2X 1 Step Buffer
2. 7.5 µl of nuclease free water
3. 1µl of Enzyme mix
4. 1 µl of Fw+Rw primers
5. 2µl RNA up to a total volume of 25 µl.

Samples were reverse transcribed at 50°C for 30 minutes and subsequently denatured at 94°C for 2 minutes in the thermal cycler "Eppendorf Mastercycler ep gradient S." After a brief denaturation process at 94°C for 30 seconds, primers were annealed for 45 seconds followed by the polymerization step, which lasted for 2 minutes at 72°C. This cycle was carried out 35 to 40 times.

The outer PCR products were then subjected to subsequent PCR reaction (inner PCR), using additional set of primers (Table 2) using Thermo Scientific Dreamtaq PCR master mix (2X) (Thermo Fisher Scientific, USA), thus finishing the nested PCR protocol.

General protocol for inner PCR was :

1. 25 µl of DreamTaq PCR Master Mix (2X)
2. 2 µl of Fw primer in 0.6 µM final concentration (Invitrogen by Life Technologies, Carlsbad, California, USA)
3. 2 µl of Rev primer in 0.6 µM final concentration (Invitrogen by Life Technologies, Carlsbad, California, USA)
4. 16 µl of nuclease free water
5. 5 µl of outer DNA template

Final reaction volume was 50µl.

In the thermal cycler "Eppendorf Mastercycler ep gradient S" after an initial denaturation step on 95°C for 3 min, the cycling begins with a short denaturation at 95°C for 30 sec, the primers were annealed for 45 sec at an appropriate temperature, according to the melting temperature of the primers and the template (Table2.), followed by an extension time of 3 minute at 72 ºC. This cycle was repeated 40 times and followed by a final elongation step of 10 minutes at 72 ºC.

3.2.2.3 Agarose gel electrophoresis

All PCR products were analyzed using traditional agarose gel electrophoresis, and a 2 percent agarose gel was prepared. For visualization of the DNA bands, all agarose gels were stained with a cyber green fluorescent dye (SYBRTM Thermo Fisher Scientific, USA). TAE (Tris-acetate-EDTA) buffer was used as both a running buffer and in the agarose gel. TAE buffer is commonly prepared as a 50X stock solution for laboratory use. This stock solution can be diluted 50:1 with destylled water to make a 1X working solution. This 1X solution will contain 40mM Tris, 20mM acetic acid, and 1mM EDTA.
Samples were mixed with 2 µl of the gel loading dye (0.125% Bromophenol blue, 40% Sucrose) and approximately 8 µl were loaded per lane on the agarose gel. The GeneRuler™ 1 kb DNA ladder (DNA Standard 100bp - Serva Electrophoresis GmBH, Heidelberg,) was used as a molecular length size marker. Electrophoresis reactions were run at 120 V. The DNA bands were visualized on a UV transilluminator under a UV light with a wavelength between 280 and 320 nm.

3.2.2.4 PCR products purification and cycle sequencing reaction

Prior to sequencing, excess dNTPs, enzymes, and buffers were removed using the Gene JET pcr purification kit (Thermo Fisher Scientific, USA) according to the manufacturer's instructions. The silica membrane spin purification protocol was used, which allows nucleic acids to bind to a silica membrane inside a spin column.

The following was the general protocol:

1. We added 1:1 of Buffer PB (50µl) into the inner PCR product (50 µl) and 50 µl product of nested PCR reaction, then after vortexing, the samples were applied to the MinElute column and centrifuged for 1 min at 14000 x g. Flow-through was discarded without changing the collection tube. Of note, is the color of the sample soultion that need to be yellow indicating optimal pH value fro DNA biding.

2. 700 µl of Wash Buffer was added to the MinElute column and centrifuged for 1 min at 14000 x g. Flow-through was discarded and the MinElute column was put back into the same collection tube for an additional 1 minute centrifuge at the maximum velosity.

3. The MinElute column was, then, placed in a clean 1.5 ml microfuge tube and the elution was done with 50 µl of Elution Buffer. After incubation for 1 min, the samples were centrifuged for the last time for 1 min at the maximum velosity.

The sequencing of both the sense (forward) and antisense (reverse) strands of purified PCR products was performed using the Sanger cycle sequencing method. Sanger sequencing is based on DNA polymerase's selective incorporation of chain-terminating dideoxy nucleotides during in vitro DNA replication (84). Chain termination sequencing reactions were performed in-house using the BigDye Terminator v3.1. Cycle Sequencing kit (Applied Biosystems, USA) in accordance with the manufacturer's instructions. In DNA sequencing reactions, fluorescently labeled dyes are attached to ACGT extension products. The dyes are available in four different colors: red (Thymidine base), blue (Cytosine), black (Guanine), and green (Adenine). The dyes

are incorporated into the DNA using either 5'-dye labeled primers or 3'-dye labeled dideoxy nucleotide terminators.

General protocol was as follows:

1. 1µl each primer 5 µM working concentration
2. 1µl of purified RT-PCR product,
3. 2µl 5× Cycle sequencing dilution buffer
4. 2µl BigDye
5. 4µl ultrapure water

Final reaction volume was 10µl.

The sequencing PCR included 40 cycles of 96°C for 30 s, 50 ◦C for 7 s and 60◦C for 4 min.

After running the sequencing reaction, non-incorporated dideoxynucleoside triphosphates were removed by 75% isopropyl alcohol precipitation and the pellet was resuspended in 20µl High Density Formamide (Applied Biosystems Incorporated, Foster City, CA) for denaturation.

Detection was done in an ABI Prism 310- Genetic Analyzer capillary electrophoresis system (Applied Biosystem, Foster City, CA, USA). The Sequencing analysis software v.5.2 was used to analyze the results (Applied Biosystem, Foster City, USA).

3.2.3 Phylogenetic methods

3.2.3.1 Subtyping

The subtyping of Serbian dataset as well as confirmatory subtyping of HIV subtype B dataset from six Balkan countries were done using two approaches: i) COMET HIV-1 tool (Luxembourg institute of Helath) (https://comet.lih.lu/) and ii) REGA HIV-1 subtyping tool version 3 (http://regatools.med.kuleuven.be/typing/v3/hiv/typingtool/).

3.2.3.2 Phylogenetic trees reconstruction

The alignment of Serbian dataset was aligned using the Clustal W algorithm in MEGA v 10 (https://www.megasoftware.net/). To alleviate potential differences in selective pressures between two groups of sequences (treatment-naive vs treatment-experienced), a codon-stripped sequence alignment was created by removing drug resistance-associated positions identified by the Stanford Drug Resistance Database (85).

The alignment of the Balkan dataset had to be aligned with available web servers through the MAFT algorithm available at: https://mafft.cbrc.jp/alignment/server/. Similarly as mentioned above manual editing of sequences was done in MEGA v 10 to edit out resistance-associated codons (85).

Evolutionary models depict various probabilities of nucleotide changes along the sequences, hence adopting different phylogenetic hypotheses to construct the phylogenetic tree. All phylogenetic and downstream phylodynamic and phylogeographic analysis used general time-reversible (GTR) model, with gamma distribution and proportion of invariable sites (GTR+G+I), as selected by jModelTest (an algorithm that uses multiple statistical approaches to select the model for a study alignment) (86).

Phylogenetic relations of Serbian dataset were first determined using PAUP* version 4.0 software and then a bootstrap analysis with 1000 replicates was used to estimate branch supports. Phylogenetic trees were visualized in FigTree program version 1.3.1. Then Bayesian phylogenetic tree was reconstructed through MrBayes software and a Markov Chain Monte Carlo (MCMC) search was made for 10 $x10^6$ generations using tree sampling every 100^{th} generation and a burn-in fraction of 20%. Statistical support for specific clades was obtained by calculating the posterior probability of each monophyletic clade, and a posterior consensus tree was generated after a 25% burn-in (87, 88). For the definition of HIV clusters, we performed several analyses and followed the strategy of "statistical support of clades plus similarity" that proved to be suitable in our previous investigation of transmission chains in Serbia (25, 26).

Generally local transmission clusters in Serbian data were characterized by the following criteria: i) monophyly, ii) distances below a given threshold, iii) the bootstrap support and Bayesian probability value of splits as statistical support of monophyletic clades, iv) subsistence of local clusters after inclusion in tree reconstruction of sequences with the highest similarity score, based on BLAST analyses.

Namely, using this strategy two sets of criteria were selected: i) according to the first, strict set of criteria, transmission clusters were monophyletic clades consisting of three or more sequences, fulfilling the conditions of the genetic distance of 1.5% or less, with minimal

bootstrap support of 90%, and the Bayesian posterior probability higher than 0.9. The second relaxed criteria set, included those clades with bootstrap support of over 75% and with a genetic distance of less than 4%. For each sequence found within a cluster according to the second criteria set, the 5 most similar sequences were identified using BLAST analysis generating the alignment of similar sequences in order to prove cluster sustenance.

On the other hand, different approach was used to anlayse the aligment of Balkan sequneces, due to size of the alignement as well as computational expensiveness and rigidity of conventional phylogenetic tests. Generally speaking, methods for detection of transmission clusters differ in statistical approach, software tools, analytical parameters, topological and distance thresholds used for results interpretation (89). The use of a range of criteria with more relaxed thresholds has been shown to be more appropriate for evolutionary analysis of an HIV-1 epidemic over longer time periods, to account for specifics of sampling and intra-host diversification of viral quasi-species over time (89).

Therefore, taking into the account the size and structure of Balkan alignment (over 2000 sequences from seven countries collected over two decades), a Maximum Likelihood phylogenetic analysis with two complementary tests was done as implemented in IQTree web server: Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT) and ultra-fast Bootstrap (UFB) with 1,000 replicates (90).

In summary transmission clusters in Balkan dataset were evidenced based on the following criteria: i) monophyly (cluster sharing a common ancestor); ii) SH-aLRT value of >0.9 and iii) UFB support >95%. The identified clusters were further classified based on the prevailing sequence origin as either local (over 75% of sequences are from the same country of origin) or cross-border clusters (more than 25% of sequences from different countries) (91). To further confirm the obtained cross-border clusters: i) a consensus sequence for each cluster was obtained by consensus maker tool available at EMBOSS Cons tool (92); ii) pearvise nucleotide distance between every cross-border cluster sequence and the corresponding consensus sequence was calculated in MEGA v X in order to express the range of homologie as average pearvise nucleotide distance; iii) for each consensus sequence, a BLAST search was done and 100 most similar sequences (other than those already included in the study) were added into analysis, using SH-aLRT and the UFB to prove cluster sustenance (91).

3.2.3.3 Bayesian molecular-evolution analyses

Detailed molecular clock and phylodynamic analyses were performed on detcted local transmission network and the two most expanded transmission clusters (composed of 15 and 11 sequences) associated with the MSM and subtype B. Additionally, two heterosexual monophyletic clades were analyzed as well, one each of subtypes B and C, in order to analyze heterosexual sub-epidemic as well.

Detailed molecular clock, phylodinamic and phylogeographic analysis were done on sequnces belonging to cross-border clusters in order to analyse spacio-temporal patterns of cross-border spread and to evaluate the role of this spread in evolution of HIV epidemic in the Balkan.

Phylodinamic analysis, can be done using two statistical approaches: i) The Coalescent approach (estimating time of the most recent common ancestor (tMRCA) and effective population growth over time (Ne)) and ii) The Birth-Death approach (estimating the effective

reproductive number over time (Re)). Both analyses are aimed at describing phylodynamic pattern of a cluster and can be used to assess the reproducibility of the proposed growth scenario. However, coalescent models do not account for deaths in the population creating a sigmoid Ne plot; whereas, by accounting for deaths in the population, Birth-Deaths models estimate trends of Re with the strict cut-off value of 1. The exponential growth phase of Ne corresponds to Re values above 1-the growing cluster and plateaued phase of the Ne curve corresponds to Re values below 1 indicating stagnation in growth.

To estimate tMRCA and Ne for transsmission clusters mentioned above Bayesian Skyline Plot was used as a model of population growth as implemented in BEAST v 1.10.4 (93, 94). To estimate Re, considering large time span of sampling dates, we applied the Birth-Death Skyline Serial model (BDSKY) as a model for viral transmission, as implemented in BEAST2 v 2.6.5, with literature informed set of parameters (95, 96).

The value of Re was set as a log-normal prior with a mean value (M) of 0.0 and a variance (S) of 1.25, whereas the number of dimensions was set to 4. The rate of becoming uninfectious was set as a log-normal prior with M = 1.5 and S = 0.4 (95% CI 1.7-10.2 corresponding to the maximal natural infection period of 10 years in HIV patients), as reported previously. A prior beta (1.0, 9999) was used to estimate the sampling probability, which corresponded to a minority of cases sampled (about 10% of cases in both datasets). The epidemic's origin prior was estimated using a log-normal distribution and the M and S parameters were manipulated in order to reflect corresponding values of 95% intervals of tMRCA. Namely, if M = 3 and S = 0.15, the median will be set in the 1990s and the 97 % quantile in the early 1990s, which closely matched the tMRCA ranges of the majority of clusters analyzed.

Each analysis in bout BEAST 1 and 2 was performed using the uncorrelated log-normal relaxed clock as a clock prior, previously found to be the most suitable for HIV datasets as well as GTR G+I model of nucleotide substitution. MCMC chains were run for 5×107 generations for each data set, with a burn-in of 10%. The convergence of parameters and tMRCA was assessed through the ESS>200 after excluding an initial 10% for each run. A graphical representation of the effective number of infections through time was generated by TRACER v1.7. Additionally, the BEAST 2 .log output files were plotted using the “bdskytools” package in the R studio software to visualize Re trends.

To describe the geographic dispersal of HIV subtype B multiple phylogeographic analyses were done in BEAST v1.10.4. using Coalescent Bayesian Skyline as a tree prior and Bayesian Stochastic Search Variable selection (BSSVS) method with asymmetrical substitution models for location parameters (97). In short, this dispersal was studied on two levels: (i) an alignment including all sequences from the countries that contributed to cross-border clusters; (ii) separate sub-alignments corresponding to each of the identified cross-border clusters. The location annotated maximal clade credibility (MCC) trees were visualized using Fig Tree software and analyzed further in SPREAD3 program; statistical significance of spread routes was assessed by calculating Bayes Factor (BF) using BSSVS network.log files. Routes with BF>3 and posterior probability >90% were described as significant (98). Countries with the minimum of two outbound connections were marked as putative exporters (99).

3.3 Statistical analysis

3.3.1 Logistic growth modeling of HIV epidemic growth in Serbia

Logistic growth modeling was performed based on data about new HIV cases, as well as the basic related demographic data in the period 1984 - 2016, available from the Institute of Public Health of Serbia "Dr Milan Jovanovic Batut "and the annual HIV reports of the European Center for Disease Control (100). The data on transmission route for Serbia became available since 2004 according to ECDC, so the growth modeling of new HIV cases in main Key populations (Hetero, MSM and PWID) was done for 2004-2016 time period (101). Fitting of the logistic curve of trends for cumulative annual number of new HIV cases in 1984-2016, in the general population and MSM was done by NLREG (ver. 6.6) software (http://www.nlreg.com).

The logistic regression model assumes the existence of a maximum population size which a given environment can accommodate (carrying capacity alias K) and a biphasic growth rate creating two stages of growth: i) an early exponential phase and ii) a late phase of the plateau (102). Following logistic growth model was used: y = K / (1 + exp*(a + b*x)), where K represents carrying capacity, a and b represent parameters that shape and scale the function. NLREG provides an estimation of parameters K, a and b with the strongest statistical back-up. It also provides several diagnostic statistical tests and variables which describe the statistical strength of the proposed model such as: i) prob.t, the probability that the estimated parameter is 0. Prob.t less than 0.05 is considered as a good estimation of parameter significance for the model; ii) prob.f is the probability that all of the regression parameters are 0. It ranges from 0 to 1. Prob.f less than 0.05 makes the proposed model statistically significant. The model obtained was used to predict future trends of HIV epidemic in the general population, MSM, PWID and Heterosexuals for the time period 2017-2030 (102).

3.3.2 Latent class analysis

Latent class analysis (LCA) is a statistical tool that explores underlying patterns of covariance in the data structure to identify subgroups or 'discrete classes' of participants' epidemiological, behavioral and clinical profiles. Class membership is inferred based on an individual's pattern of responses across a set of variables but not directly observed and therefore classes are considered to be latent. In this study, LCA was conducted to examine and identify participants' risk profiles regarding 11 types of latent class indicators (categorical latent variable). Latent class indicators included: gender, transmission risk, age, residence, level of education, HBV and HCV co infections, presence of other STI, CDC stage at the time of diagnosis, time-period of diagnoses, sequences within clusters/network. These classification variables were selected to represent a range of established HIV risk factors in order to identify comprehensive HIV risk profiles. The covariates were separated into ordinal categories where each category was assigned a nominal value of 1 to 4.

LCA was performed in R software with a polytomous variable LCA (poLCA) software package. PoLCA is a user-friendly package for the estimation of latent class models and latent class regression models in R available from the Comprehensive R Archive Network, at http://CRAN.R-project.org/package=poLCA (103, 104, 105). Based on the fact that this algorithm may locate a local, rather than global maximum, nrep was set to 10, which increased the probability that the global maximum log-likelihood would be located.

Parameters used to select the optimal number of latent classes in LCA included the Akaike information criteria (AIC) and the Bayesian information criteria (BIC), two most widely used parsimony measures (106,). We began with a 1-class model and increased the number of classes in each subsequent model seeking to minimize both the BIC and the AIC value, before these values increased with the addition of another class.

3.3.3 Join-point regression analysis of HIV incidence trends

To assess the time trends, the number of new HIV diagnoses in general and in key populations and transmission-related subgroups (men who have sex with men (MSM), people who inject drugs (PWID) and heterosexuals (Hetero)) was obtained from the ECDC HIV/AIDS surveillance report for participating countries: Bulgaria, Croatia, Greece, Montenegro, Romania, Serbia and Slovenia (22). Data on countries' population size as well as age and gender distribution were obtained from a reference data website (https://www.worldometers.info). The size estimates of the main key and sub-populations were obtained from UNAIDS Fact sheets (https://www.unaids.org/en).

To compare the state of the HIV epidemic in seven Balkan countries time trends of HIV incidence were calculated using joinpoint regression analysis as implemented in the Joinpoint Regression Program, Version 4.8.0.1 (107). The result of this analysis is the annual percent change (APC) of HIV incidence in general and main key populations. The relative character of APC makes it easier to compare incidence trends between states with vastly different populations in size, numbers of registered HIV cases and prevention strategies. Furthermore the software, by using the Monte Carlo Permutation method, creates the simplest regression model that the data allows with the maximal number of joinpoints. Each joinpoint indicates a statistically significant change in trends (increase or decrease of incidence) and each of those patterns is described by an APC. Furthermore, statistically insignificant regression models were reported as well, indicating a stable trend, with no statistically significant change of incidence.

4 Results

4.1. Phylogenetic and phylodynamic analyses of HIV epidemic in Serbia

In this study, a total of 385 sequences were analyzed, collected between 1997 and 2015 from both treatment-naïve and treatment-experienced HIV-1 positive subjects. Of these, the majority were male (81%), with a mean age of 37.5 years and reporting MSM contact (83.4%) as the most probable route of HIV acquisition. Epidemiological, demographic and clinical data are shown in Table 2.

Table 2. Clinical and epidemiological characteristics of Serbian patients

Category	N	%
Gender	*385*	*100*
Male	312	81
Female	73	19
Transmission route	*385*	*100*
MSM	302	78.4
Hetero	68	17.6
PWID	15	3.8
Other		
Place of residance	*385*	*100*
Urban Area	315	81.8
Rural	70	18.8
Other		
CDC disease stage	*385*	*100*
A	175	45.5
B	93	24.1
C	117	30.4
Other		
OSTI	*385*	*100*
Present	87	22.6
Absent	298	77.4
Other		
Median age		
Male	34.2	
Female	39.5	

CDC, Center for Disease Control OSTI other sexually transmitted diseases
MSM, men who have sex with men
PWID, people who inject drugs

Both maximum likelihood and Bayesian approach gave congruent results, revealing that 27.7% (107/385) of sequences grouped within 19 transmission clusters, accomplishing predefined criteria, all within subtype B (Figure 6; labeled in dark blue). Among non-B subtype, subtype C clade was found with the smallest distance (6.5%) among non-B clade, with bootstrap support of 100, composed of sequences isolated from heterosexual patients. The observed subtype C genetic distance exceeded the pre-defined criteria for transmission clusters; hence it was not classified as one, however, downstream phylodynamic analyses were still performed, in order to be able to make comparisons to the relevant subtype B clades.

As previously, transmission network was identified, as a phylogenetic clade with high bootstrap support as well as posterior probability value but with higher genetic distance of 8.2% than the predefined cut off of 4.5%. This network contained sequences sampled in the time frame of 19 years and comprised 63⁄385 sequences, encompassing 16.3% 'of the total number (Figure 6). Taken together, the number of sequences within transmission clusters/network totaled to 170/385 (44%) Importantly, only 4 of 19 (21%) transmission clusters contained sequences isolated from heterosexual individuals.

Among non-B subtypes, ther was no transmission clusters as defined by predefined criteria, however, subtype C clade was found with the smallest distance (6.5%) among non-B clade, with bootstrap support of 100, composed of sequences isolated from heterosexual patients. The observed subtype C genetic distance exceeded the pre-defined criteria for transmission clusters; hence it was not classified as one, however, downstream phylodynamic analyses were still performed, in order to be able to make comparisons to the relevant subtype B clades.

Sequences contributing to transmission clusters⁄network were from individuals of a mean age of 35.3 years (SD=4.8) and 82.3% were men. Specifically, large transmission networks encompassed sequences from older patients of a mean age of 39.4 years, in contrast to MSM clusters with members mean age of 32.6 (SD=3.2) with particular emphasis on the most expanded transmission cluster of 15 sequences with a mean age of 29.2 (SD=2.8). Statistical evaluation revealed that significantly fewer patients found in MSM clusters were diagnosed prior to 2006 ($p= 0.0184$) while a significantly higher number of them was younger than 35 ($p = 0.0388$).

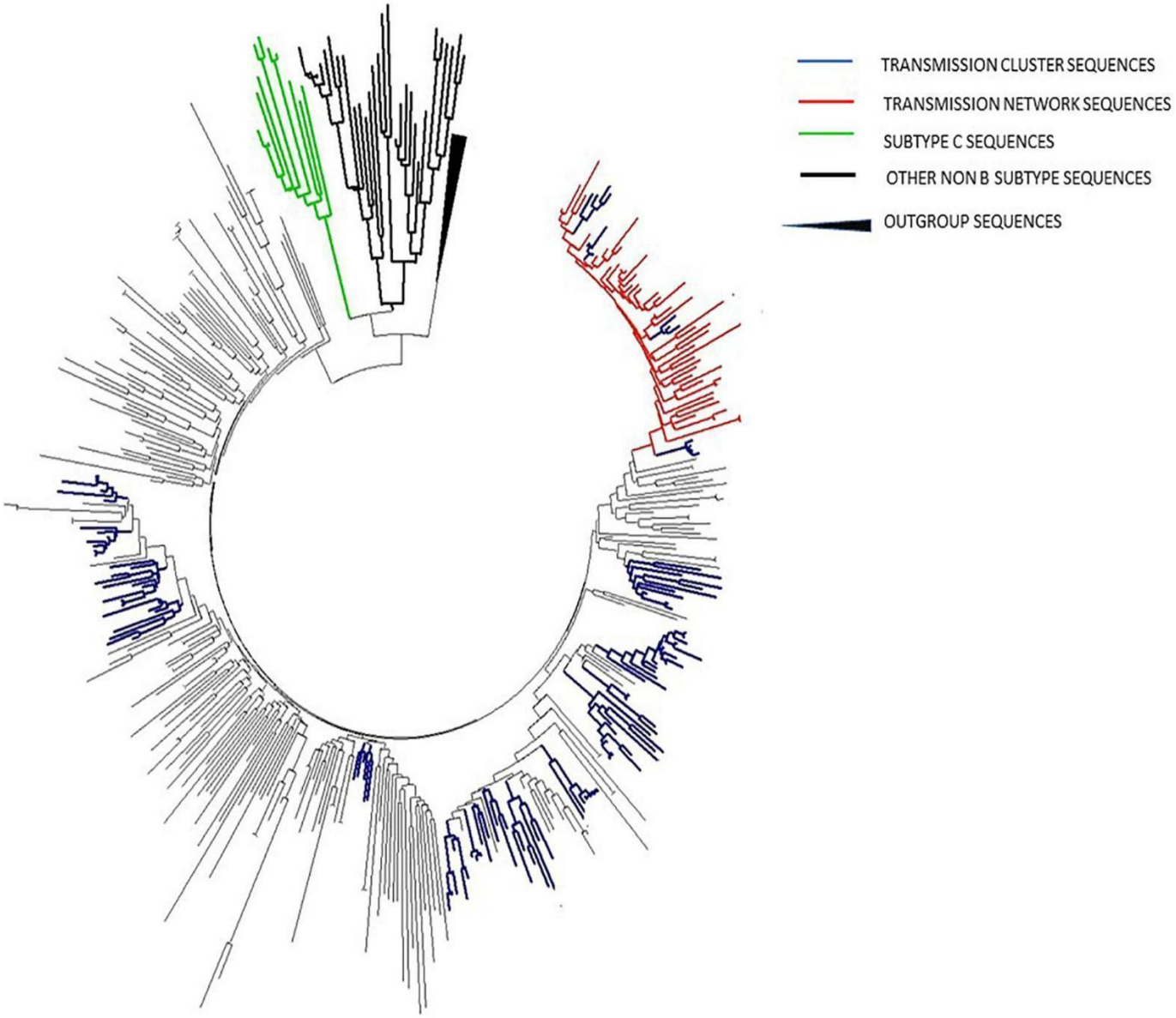


Figure. 7. ML phylogenetic tree constructed in MEGA software including all sequences analyzed in the current study together with reference sequences and background sequences from the NCBI database.

The time of the most recent common ancestor for transmission network (composed of 63 sequences) was dated to the beginning of nineties (1993 HPD: 1988-1995), while for the two MSM clusters of 15 and 11 sequences it was dated to 2008 (HPD: 2003-2013) and 2005 (HPD: 2000-2010), respectively. Dating analysis of two heterosexual clades suggested that the most recent common ancestor of the subtype B and subtype C clusters was present approximately in the 1990 (HPD: 1984-1996) and 1989 (HPD: 1985-1993), respectively.

A Bayesian Skyline plot and logistic growth analyses were performed for 3 major MSM clades and 2 heterosexual clades, of subtype B and subtype C sequences. The highest exponential growth in almost 2 logs was identified for the most expanded transmission cluster, that was also the most recent one. A demographic reconstruction of transmission network and transmission cluster of 11 sequences, showed slightly lower population growth, still with initial exponential growth in over one log which stabilized in the mid- decade of the 2000s, as reflected in a stationary phase approaching the present time (Figure 7). The obtained growth for two heterosexual clades (subtype C clade and subtype B clade) was below 1 log until the late nineties, followed by stationary phase afterwards (Figure 8).

Estimation of the Re trend over study period by BDSKY analyses showed significant differences among two sub-epidemics in Serbia, MSM monophyletic clades and heterosexual clades.

Of note, for all three investigated MSM clades Re over 1 was present during the whole analyzed period suggesting that further spread of HIV in Serbia could happen through transmission chains within the MSM key population. Especially interesting is the Re trend of the transmission network that suggests further spread of an old cluster with tMRCA estimated in the 90s.

Although all analysed MSM-associated clades were classified as active during the study period the values of Re were relatively low ranging up to 1.2. Namely, higher values of Re were found in two MSM transmission clusters with more recent tMRCA, estimated in the 00s whereas lower Re values were observed in transmission network illustrating the “ageing” of the clade (Figure 9).

On the other hand, birth-death skyline plot analyses for heterosexual subtype B monophyletic clade showed a Re value below 1 in the 2005-2015 time period, suggesting inactivity of the clade for about 10 years indicating the plateau of growth that happened after the initial period of relatively limited transmission estimated to be about 1.1. A similar situation is found in the subtype C heterosexual clade, with the obtained Re trends below one in the last years of the study period (Figure 10).

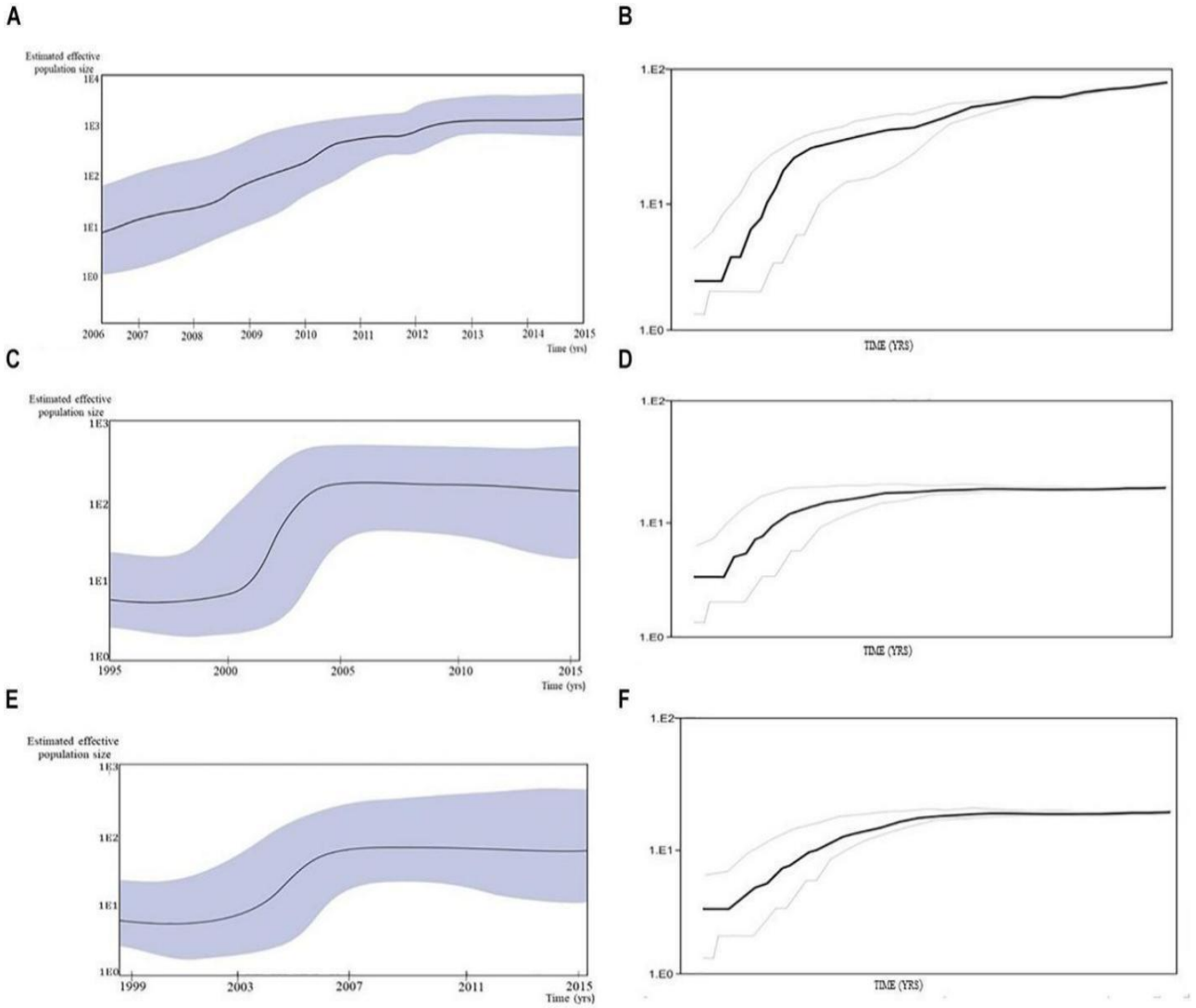


Figure 7. Population growth and the cumulative number of lineages (infections) in a logarithmic scale over time for the **(A, B)** transmission cluster of 15 sequences; **(C, D)** transmission network; **(E, F)** transmission cluster of 11 sequences. The median estimate of the effective number of infections (solid line) and 95% confidence limits of the estimate (dashed lines) are shown in each graphic.

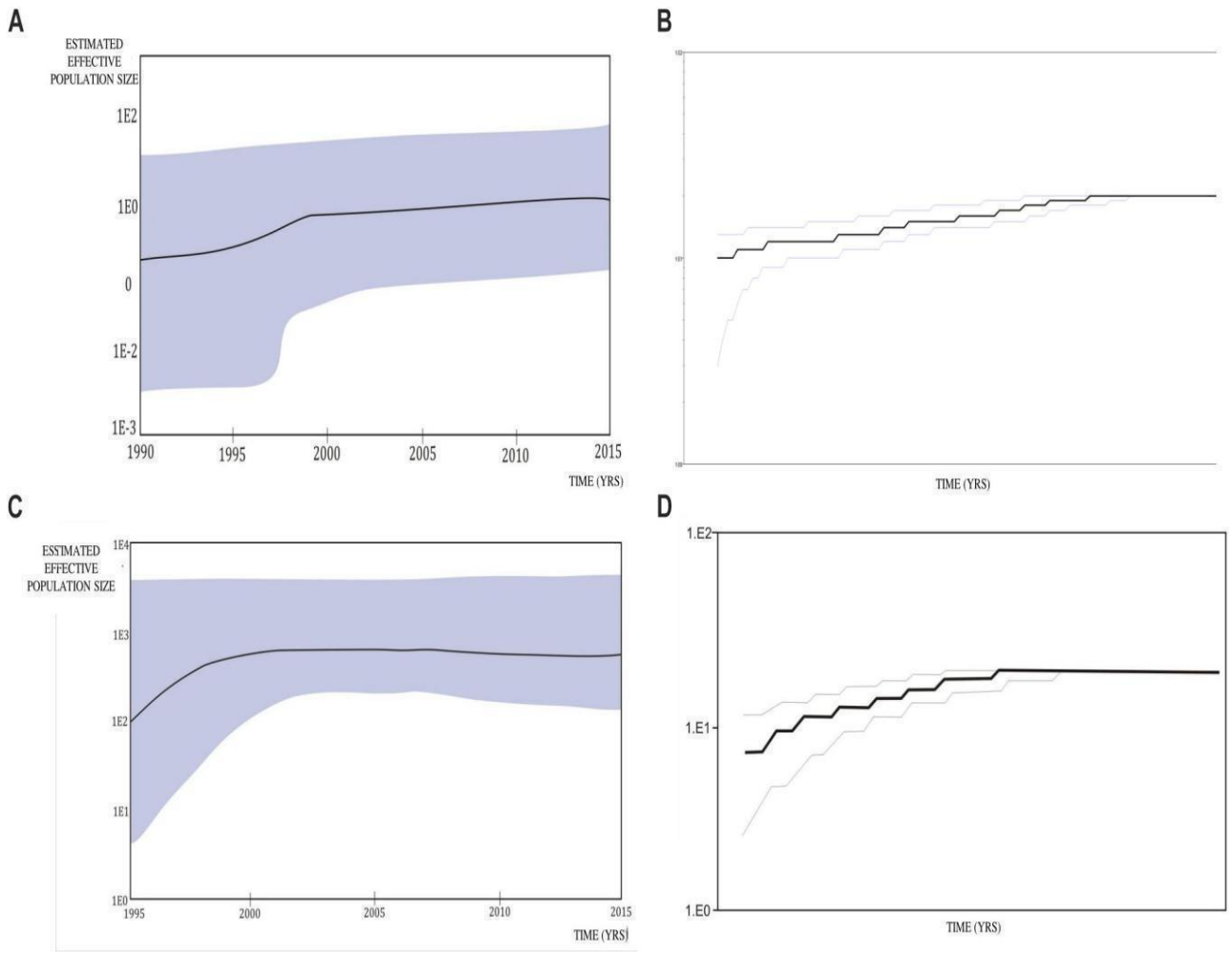
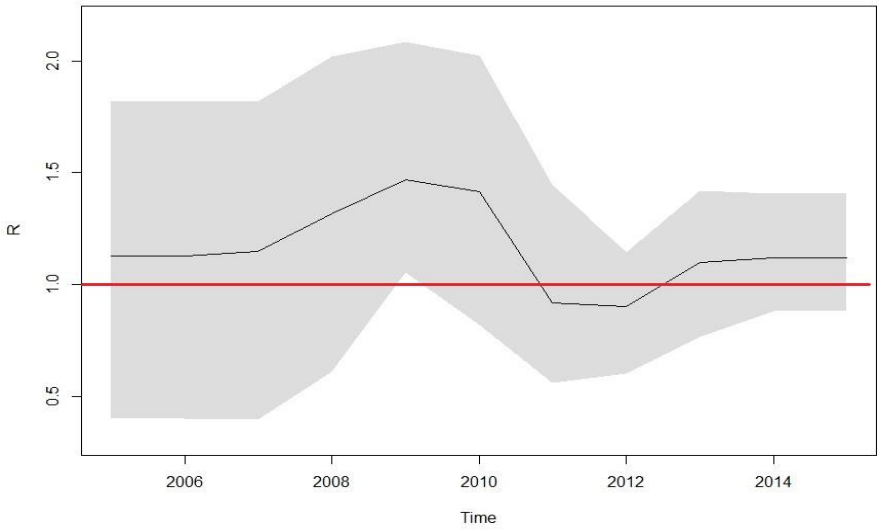
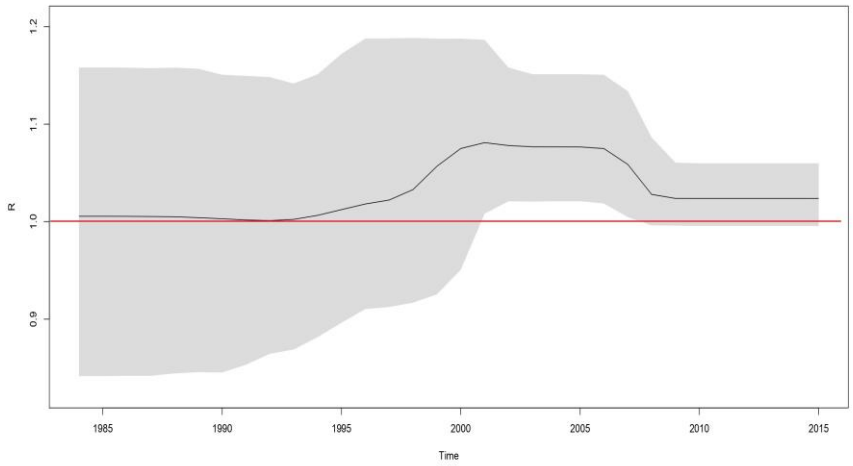


Figure 8. Bayesian Skyline plot and logistic growth analyses performed in BEAST 1 presenting population growth and the cumulative number of lineages (infections) in a logarithmic scale over time for **(A)** and **(B)** the subtype B clade composed of sequences from heterosexuals; **(C)** and **(D)** the subtype C clade composed of sequences from heterosexuals.

A)



B)



C)

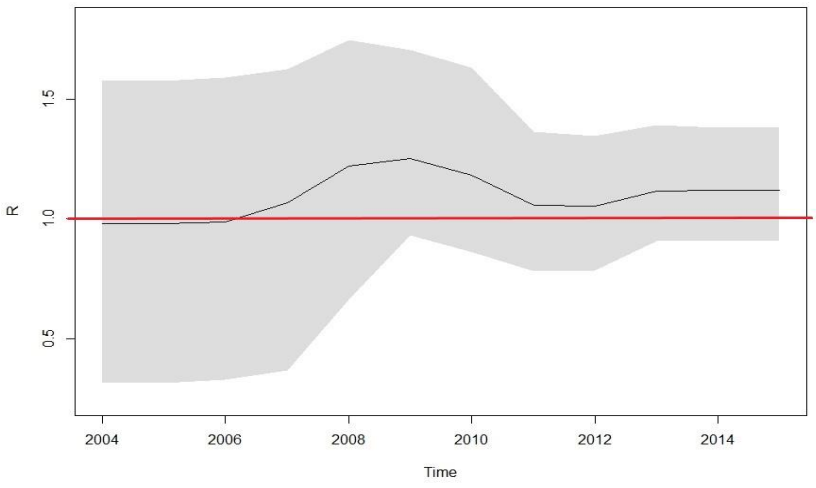
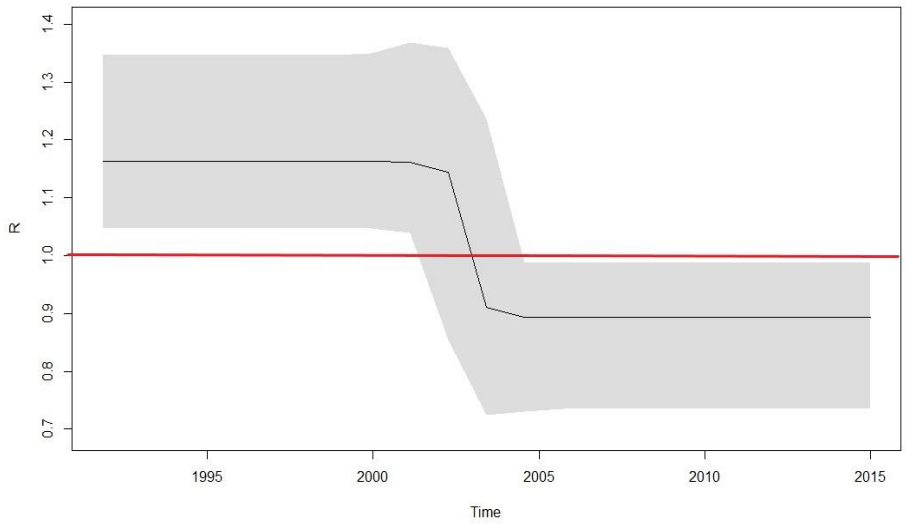


Figure 9. The estimated birth-death skyline serial models by BEAST 2 presenting the effective reproductive number (Re) over time for the (A) transmission cluster of 15 sequences (B) transmission network (C) transmission cluster of 11 sequences.

A)



B)

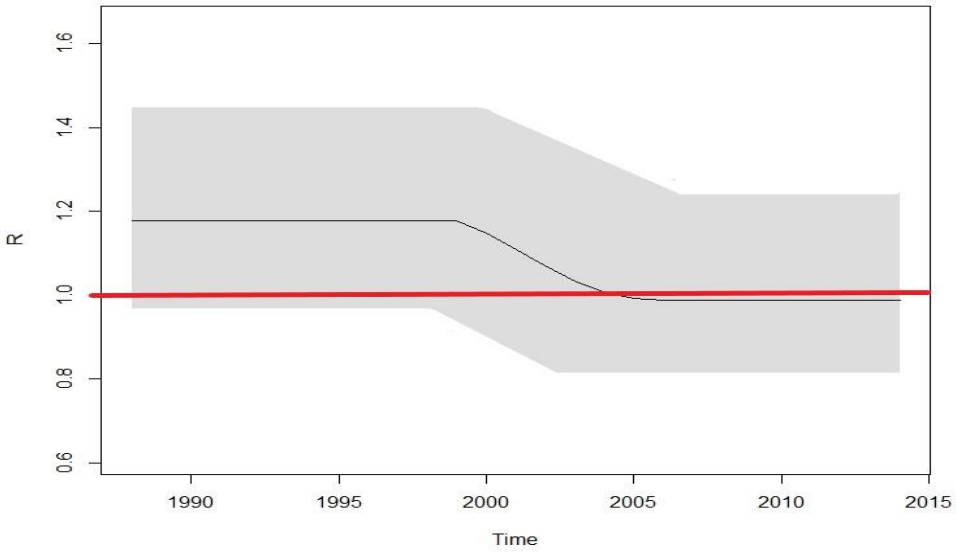


Figure 10. The estimated birth-death skyline serial models by BEAST 2 presenting the effective reproductive number (Re) over time for the (A) subtype B clade composed of sequences from heterosexuals (B) subtype C clade composed of sequences from heterosexuals.

4.2 Latent Classes Analysis

LCA analysis identified five latent classes, based on the lowest BIC and AIC values. The grouping of patients into five latent classes based on distribution of different characteristics (Figure 11) . Transmission clusters were dominantly present in 2/5 classes.
First class included young MSM subjects up to 25 years old, living in Belgrade, with HIV infection diagnosed within the last three years of the study period, with 73% of coinfection with HBV and/or other STIs and the highest percentage (53%) of grouping within transmission clusters.

The second class encompassed MSM subjects up to 45 years of age with high education, HIV diagnosis in the period from mid-decade of the 2000s onward, with less than 10% of coinfection with HBV and other STIs and around 20% of sequences found grouped in transmission clusters.

In the further three classes no tendency to group in clusters was shown: the third latent class included equal percentage of MSM and heterosexual male subjects, diagnosed during the whole period of the HIV epidemic, from Belgrade, with high percentage of HBV but no other STIs; the fourth latent class encompassed heterosexual subjects of both sexes, aged over 25 , with the date of HIV diagnosis predominantly mid-decade of the 2000s, mostly with secondary education, living in rural areas; the fifth class included female subjects reporting intravenous drug use as a risk for HIV infection, living in Belgrade, with very high percentage of HCV coinfection and diagnosed during whole period of HIV epidemic in Serbia.

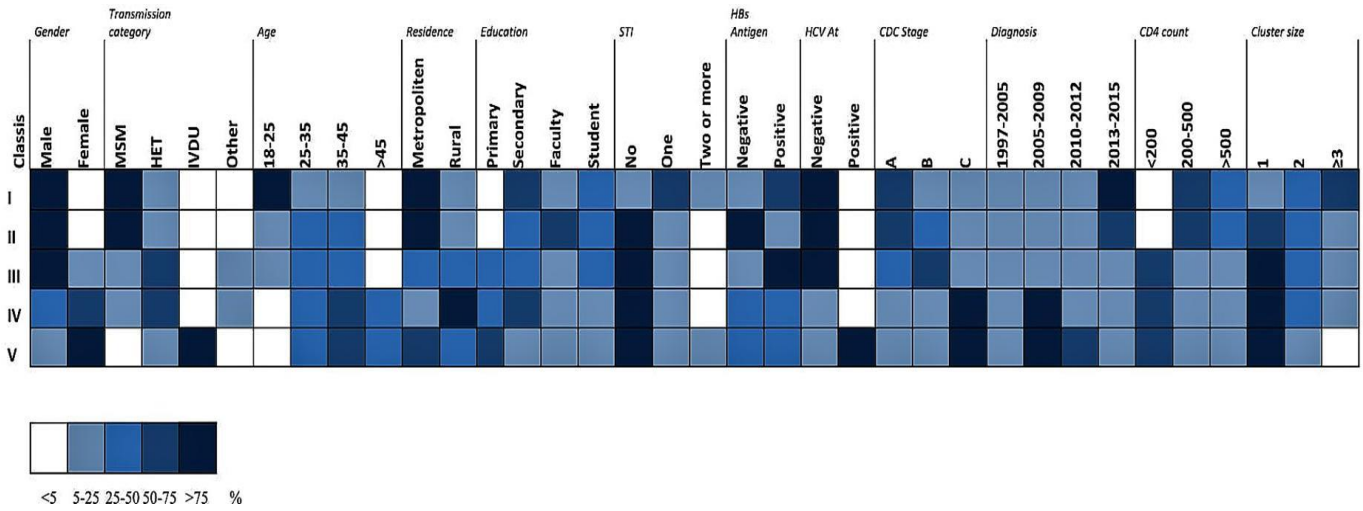


Figure 11. LCA performed in R software with a poLCA software package on data of HIV-1 infected individuals included in the current research. The distribution of participant characteristics within the latent class and predicted proportions in each cluster size by latent class are shown. Darker colors represent higher proportion as shown in the legend.

4.3 Logistic growth modeling

In the period 1984 -2016 a total cumulative number of new HIV cases was 3590. The applied logistic growth model fitted accurately the growth trend in this time period (Figure 12). Software estimated parameters K, a, b equaling 4227±271, 2.61±0.07, -0.12±0.008, respectively were found, with prob. t parameter 0.00001. Overall statistical significance was illustrated with prob.f = 0.00001.

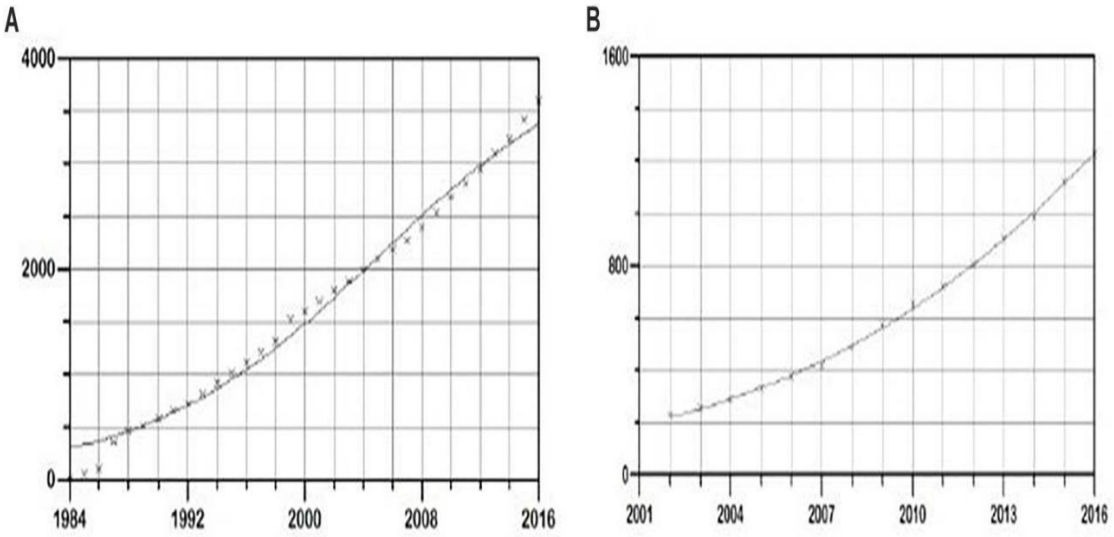


Figure 12. Logistic growth model of new HIV infections in **(A)** general population and **(B)** MSM population in Serbia, in the period 1985–2016. On *y*-axis cumulative number of new HIV cases is shown, and on the *x* time period in years is shown.

Data on transmission routes became available in 2004, hence, the prevalence of new HIV infections through MSM transmission routes in the period 2004-2016 were used in further analysis. In this time period a total of 2785 new cases of HIV in MSMs were detected. Again, the logistic growth model fitted the growth trend accurately (Figure 13). The estimated function parameters K, a, b was found to be 3377±455, 3.12±0.11, -0.15±0.005, respectively, with prob.t = 0.00001. Overall statistical significance illustrated with prob.f of 0.00001 showed strong statistical support for the proposed model.

The fitted model was used to create trends of HIV epidemic for period 2017-2030 in general population and MSMs. The modeled trend predicts approaching plateau in new HIV infections by 2030 in the general population, whereas in MSMs an exponential like curve with no signs of plateau in the studied time period was obtained (Figure 13).

Furthermore, growth trends of heterosexual transmission group as well as PWID transmission group were made in order to assess potential confounding in growth pattern of national HIV epidemic (Figures 14 and 15). The analysis showed less steep growth curve of heterosexual and PWID transmission compared with MSM transmission growth with evident plateau by the year 2030.

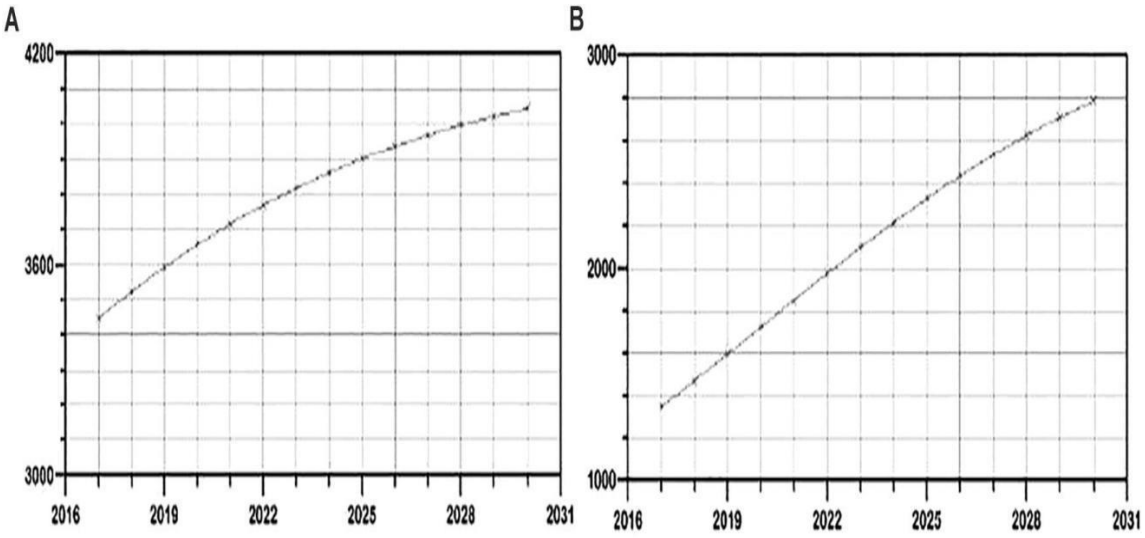


Figure 13. Evaluation of the growth trends for the period 2017–2030 of new HIV infections in **(A)** general and **(B)** MSM population in Serbia.

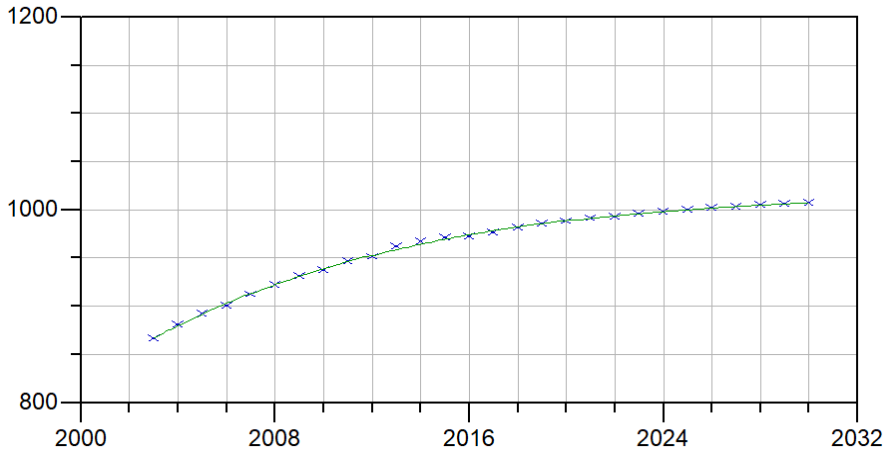


Figure 14. Logistic growth model of new HIV infections in PWID transmission group in Serbia in the period 2004-2030. On the y-axis cumulative number of new HIV cases is shown, and on the x time period in years is shown.

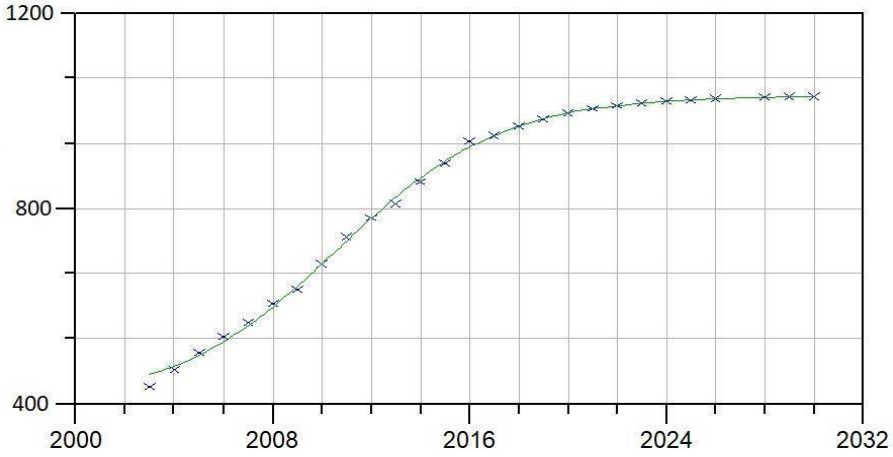


Figure 15. Logistic growth model of new HIV infections in heterosexual transmission group in Serbia in the period 2004-2030. On the y-axis cumulative number of new HIV cases is shown, and on the x time period in years is shown.

4.4 Phylogenetic Analysis of HIV-1 Subtype B in the Balknas

Phylogenetic tree reconstruction revealed that 68% (1642/2415) of sequences included in the study were grouped within transmission clusters of three or more sequences. The proportion of clustered subtype B sequences in different Balkan countries ranged from 39% in Romania to 80% seen in Croatia and Montenegro, as shown in Table 4. In total, 93 transmission clusters, accomplishing predefined criteria were detected in the study population of which 89 clusters met criteria for national clusters and four cross-border transmission clusters were identified (Figure 16). These cross-border clusters consisted of 264 sequences representing 11% of the study population (264/2,415), and 16% of the total clustered sequences (264/1,642). The identified cross-border clusters were further marked as Clusters 1, 2 (containing sequences from Slovenia, Serbia, Croatia and Montenegro), 3 (consisting of sequences from Slovenia, Croatia and Romania) and 4 (comprising sequences from Serbia and Montenegro only), ordered by size, comprising 120, 72, 38 and 34 sequences, respectively.

The contribution of different countries to cross-border clusters was markedly different: despite similar total prevalence of clustered sequences found in Croatia and Montenegro (≈ 80%), the percentage of sequences from the two countries found in cross-border clusters was 5.4% and 58%, respectively. Both from Serbia and Slovenia, around 40% of clustered sequences contributed to cross-border clusters, whereas from Romania, a much lower percentage of 6% of clustered sequences contributed to the cross-border cluster formation. (Table 3). Of note, none of the clustered sequences from Bulgaria and Greece contributed to the cross-border cluster formation (Table 3).

The ranges of homologies between cluster sequences and their consensus sequences were expressed as an average pairwise nucleotide distance (95% CI). The results are presented in relation to clusters: i) for Cluster 1 average pairwise distance between sequences and consensus sequence was 1.7% (1.5%-1.8%); ii) for Cluster 2 the average pairwise distance was 1.6% (1.4-1.9%); iii) for Cluster 3 average pairwise distance was 1.5% (1.3-1.7) and iv) for Cluster 4 average pairwise distance was 1.7% (1.6-1.8%). Generally speaking, the observed nucleotide distances were

sufficiently low, therefore consensus sequences were representative of the clusters and used in the BLAST search.

Confirmatory phylogenetic analysis with sequences obtained through BLAST search showed a small proportion of intermixed sequences (33/264, 12%), of which very few did not originate from ex-Yugoslav countries (8/264, 3%). BLAST identified sequences from the Balkans were included in the downstream phylodynamic and phylogeographic analysis.

Table 3. General information on the national HIV epidemics, dataset and cross-border transmission clusters

Country	No. of PLHIV*	Estimated No. of subtype B infections (%)**	No. of subtype B sequences included in the study***	No. of clustered sequences (%)	No. of clustered sequences in cross-border clusters (%)	Published national studies on HIV subtype distribution
Romania	24,990	1,999 (8.0)	220 (11)	85 (39)	5 (6)	Paraschiv et al. 2012 (107)
Greece	16,171	6,468 (40)	770 (12)	539 (70)	0 (0)	Kostaki et al. 2020 (106) Beloukas et al. 2016 (94)
Serbia	4,192	3,772 (90)	411 (11)	277 (67)	120 (43)	Siljic et al. 2017 (24) Siljic et al. 2013 (23)
Bulgaria	3,316	1,492 (45)	251 (17)	173 (70)	0 (0)	Billings et al. 2019 (35) Alexiev et al 2015 (34)
Croatia	1,734	1,560 (90)	365 (23)	295 (81)	16 (5.4)	Oroz et al 2019 (33)
Slovenia	923	766 (83)	333 (43)	221 (66)	93 (42)	Lunar et al. 2018 (32)
Montenegro	303	254 (84)	65 (26)	52 (80)	30 (58)	Beloukas et al 2016 (94) Stanojevic et al 2012 (26)

* PLHIV-People living with HIV in 2019
** Estimation made using the data from ECDC HIV surveillance reports and HIV molecular surveillance studies listed in the last column to the right.
***% of the total estimated No. of subtype B infections per country

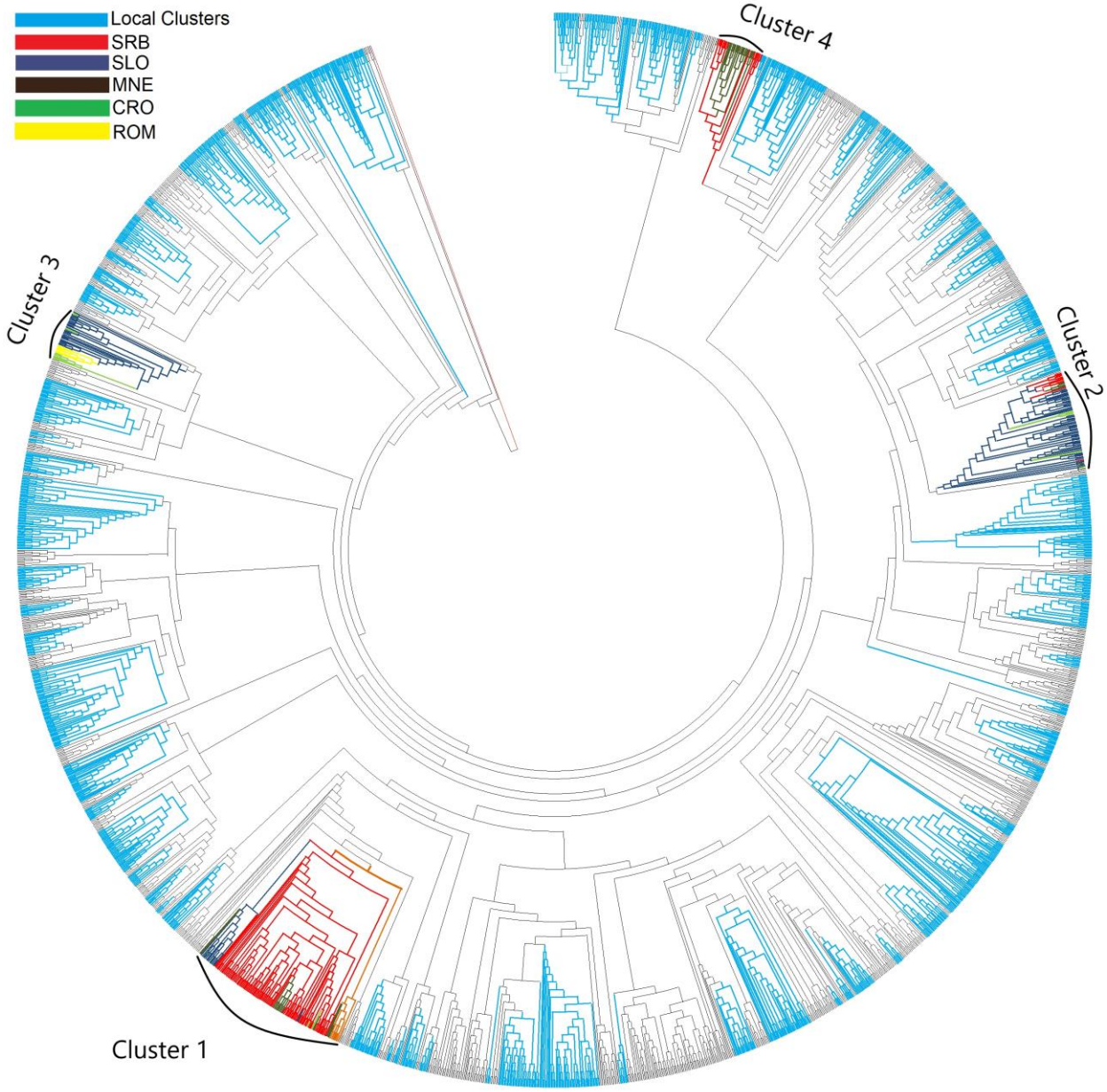


Figure 16. ML phylogenetic tree constructed in IQ Tree phylogenetic software including 2415 subtype B sequences of *pol* genetic region collected from participating Balkan countries during the period from 1999-2019. Local transmission clusters (93 in total encompassing 68% of studied sequences) are marked in blue. Sequences in cross-border clusters are color coded as follows: sequences from Serbia (SRB) in red; Slovenia (SLO) in violet; Montenegro (MNE) in brown; Romania (ROM) in yellow; Croatia (CRO) green.

4.5 Phylodynamic, Molecular Clock and Phylogeographic Analyses of Cross-border Spread of HIV-1 Subtype B in the Balkans

The years of tMRCA for the Clusters 1, 2, 3 and 4 were estimated at 1996 (95% HPD 1990–2003), 1997 (95% HPD 1994–2000), 1996 (95% HPD 1992–2001) and 2005 (95% HPD 2000–2007), respectively. Of note, both phylodynamic inference approaches yielded similar patterns of growth: (i) Bayesian Skyline plot analysis (Ne), as well as (ii) birth-death Skyline serial analysis (Re), (Figure 17 and 18).

Namely, Re value above 1 was observed in all cross-border clusters since their inception, coinciding with Ne values reflecting population growth. For Clusters 1-3, gradual decline in Re values started in early and mid-2000s, whereas for Cluster 4 constant activity was observed up to 2010. For all the clusters Re values reached around or below 1 around 2015, corresponding to plateau growth phase of Ne. Similarly, as in the Serbian dataset, the observed Re values were relatively low ranging up to 1.2.

Phylogeographic analysis of the alignment that included all study sequences from the countries contributing to cross-border clusters (total of 1,394 sequences from Romania, Serbia, Croatia, Slovenia and Montenegro), implied complex spread patterns between included countries of former Yugoslavia throughout the duration of the epidemics, with sporadic introductions to Romania occuring through the mid-1990s and 2000s (Figure 19). Serbia, Slovenia and Montenegro have been implied as putative exporters, mostly leading to multiple reciprocal introductions between the same countries, except for a single introduction from Slovenia to Croatia. On the other hand, Romania emerged as putative importer of subtype B sequences from former Yugoslavia.

Furthermore, the origin of Cluster 1 is estimated to be in Serbia from which multiple introductions to surrounding countries (Croatia, Slovenia, Bosnia and Montenegro) is seen which happened in the 2003-2010 period corresponding with peak activity of the Caster (Figures 20). On the other hand, Clusters 2 and 3 originated from Slovenia, but similarly as in Cluster 1 multiple introductions were made in Croatia and Serbia that happened in 2000-2003 time period and in Romania, which happened in 2008, again, corresponding with the peak activity of Clusters (Figures 21 and 22). On the other hand, for Cluster 4 origin is estimated to be in Serbia and branching at the origin into two distinct Serbian and Montenegrin sub-clusters is observed corresponding to the initial growth observed at Ne and Re over time plots (Figures 23).

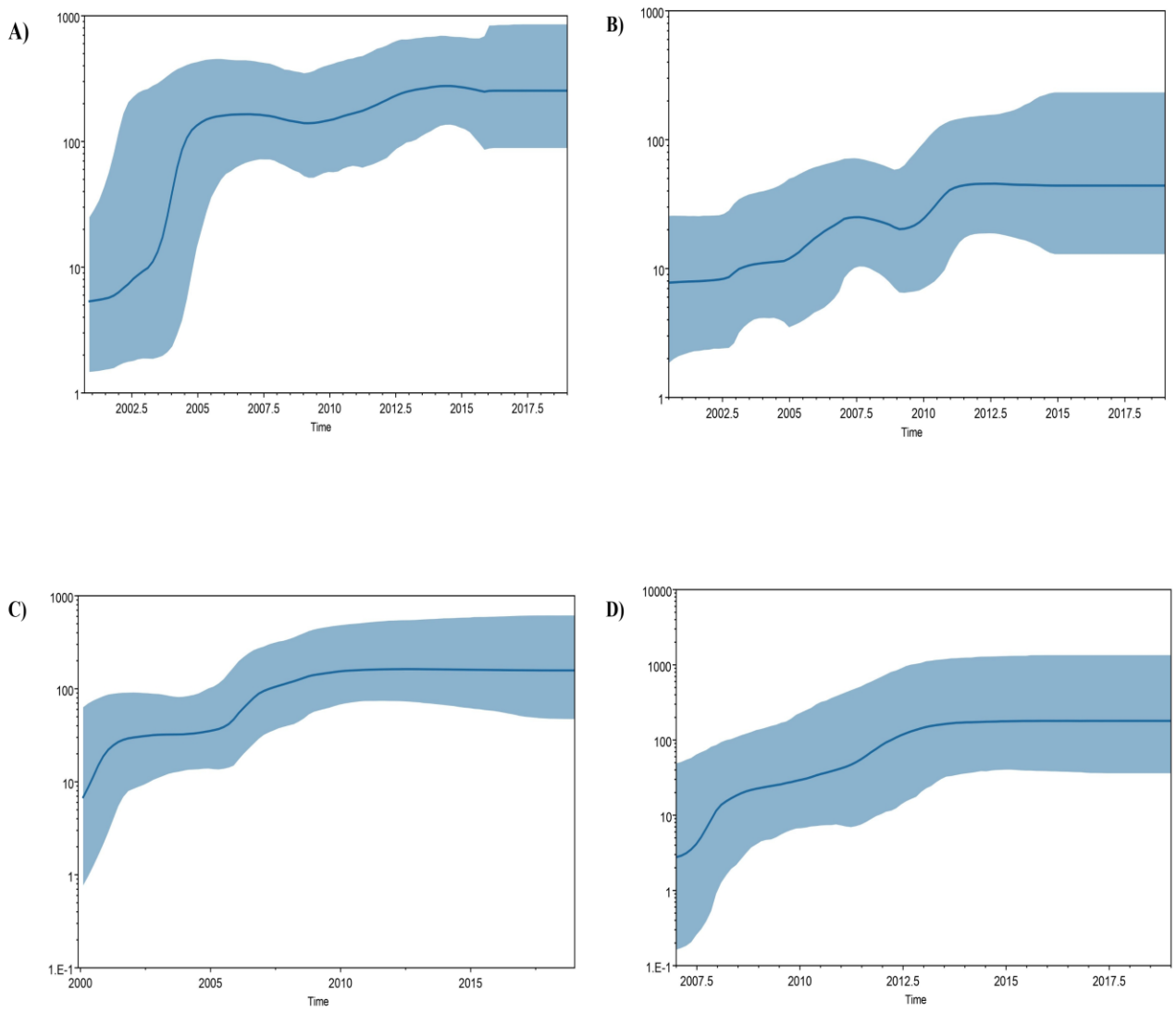


Figure 17. Bayesian Skyline plot analyses performed for cross-border clusters in BEAST v1.10.4 presenting population growth through Ne estimates (Y-axis) in a logarithmic scale over time (X-axis): Cluster 1 (A), Cluster 2 (B), Cluster 3 (C), Cluster 4 (D).

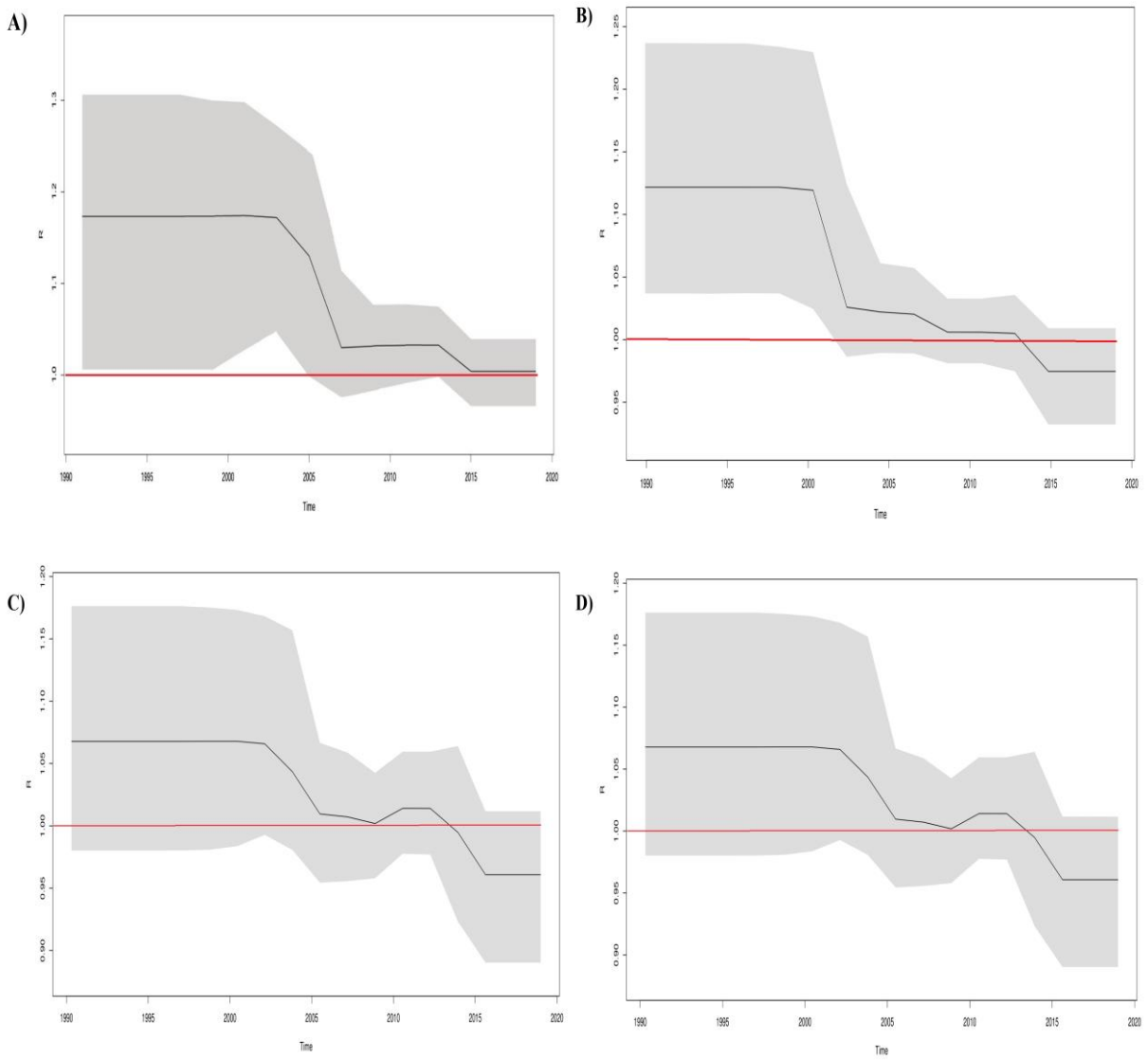


Figure 18. Estimates of effective reproductive number (Re) over time for cross-border clusters conducted in BEAST 2.6.5 under the Birth-Death Skyline Serial model: Cluster 1 (A), Cluster 2 (B), Cluster 3 (C), Cluster 4 (D). Re estimates are presented on Y-axis and time in year is presented in X-axis. Red line demonstrates the cut-off value of Re of 1.

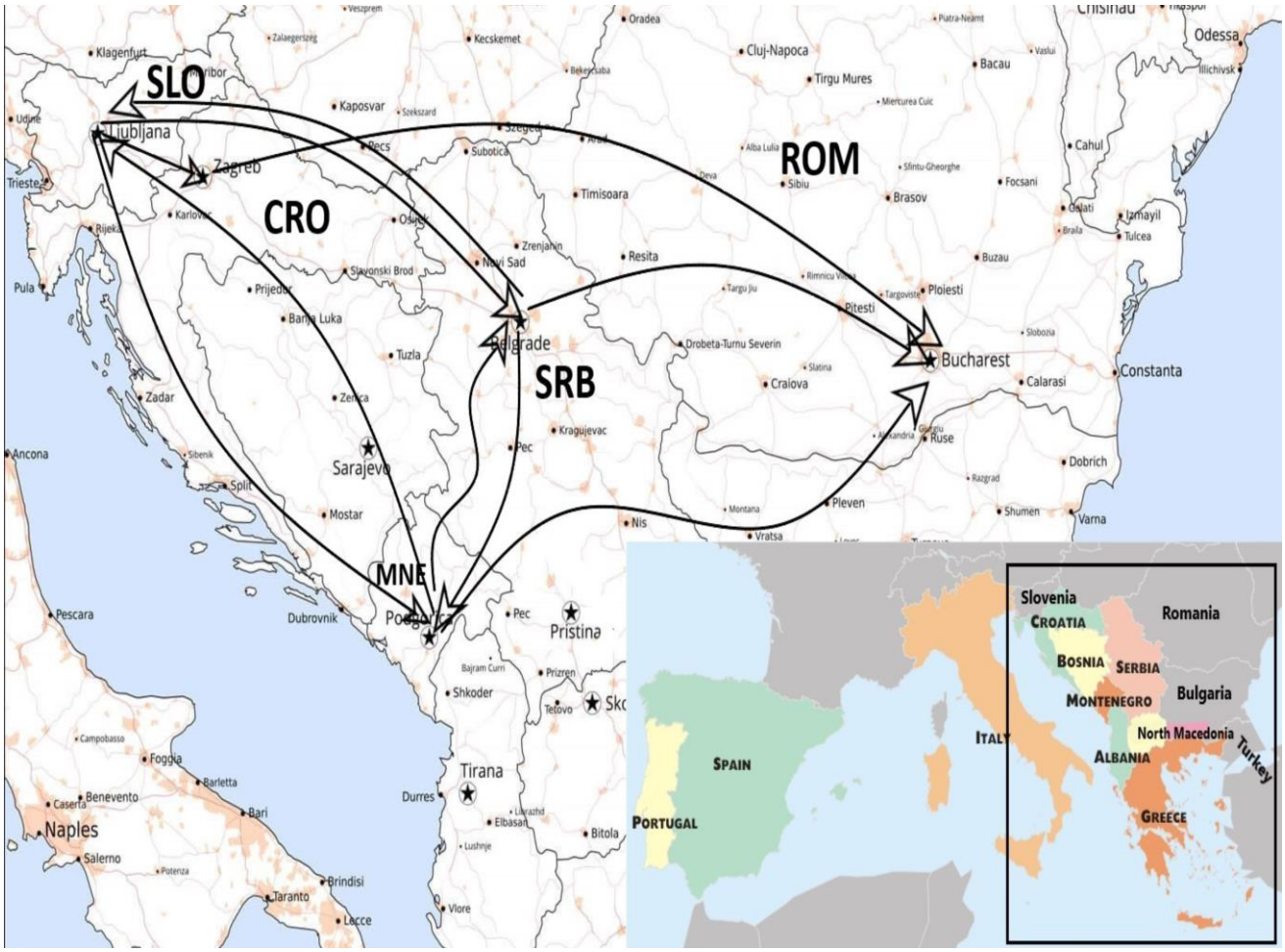


Figure 29. Phylogeographic analysis of viral dispersal among the countries that participated in cross-border cluster formation: Serbia, Croatia, Slovenia, Montenegro (Ex-Yugoslavia region) and Romania. Migration patterns supported by BF>3 and posterior probability > 0.9 are shown.
Abbreviations of participating Balkan countries: Romania (ROM), Greece (GR), Bulgaria (BLG), Serbia (SRB), Croatia (CRO), Slovenia (SLO), Montenegro (MNE).

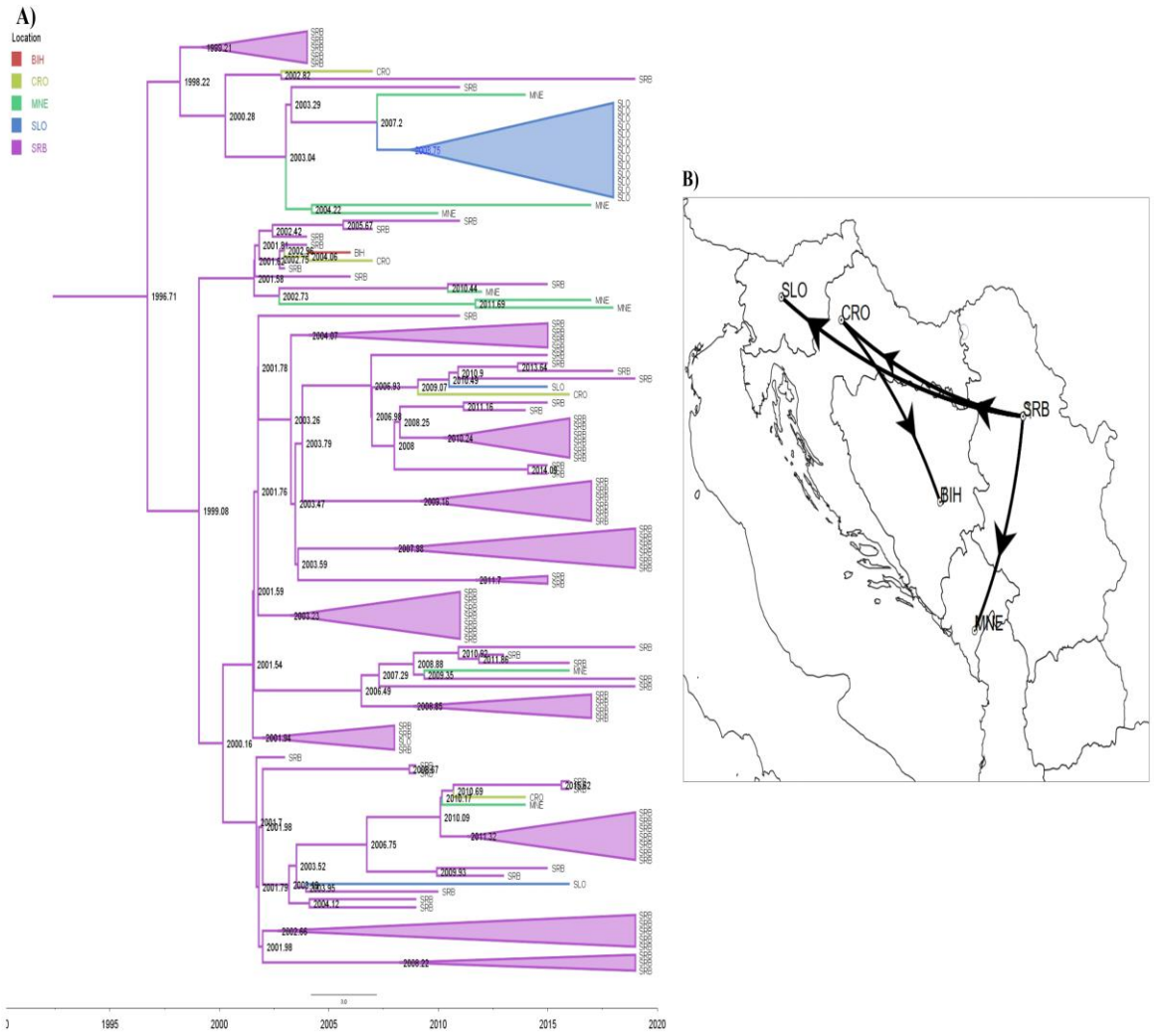


Figure 20. The location annotated MCC tree (A) and geographic representation of cross-border spread (B) of Cluster 1. The origin of the cluster is estimated to be from Serbia. Of note are multiple outward routes of transmission from Serbia to the surrounding countries of Serbia, Croatia and Montenegro as well as connection observed between Croatia and Bosnia. Migration patterns supported by BF>3 and posterior probability > 0.9 are shown.

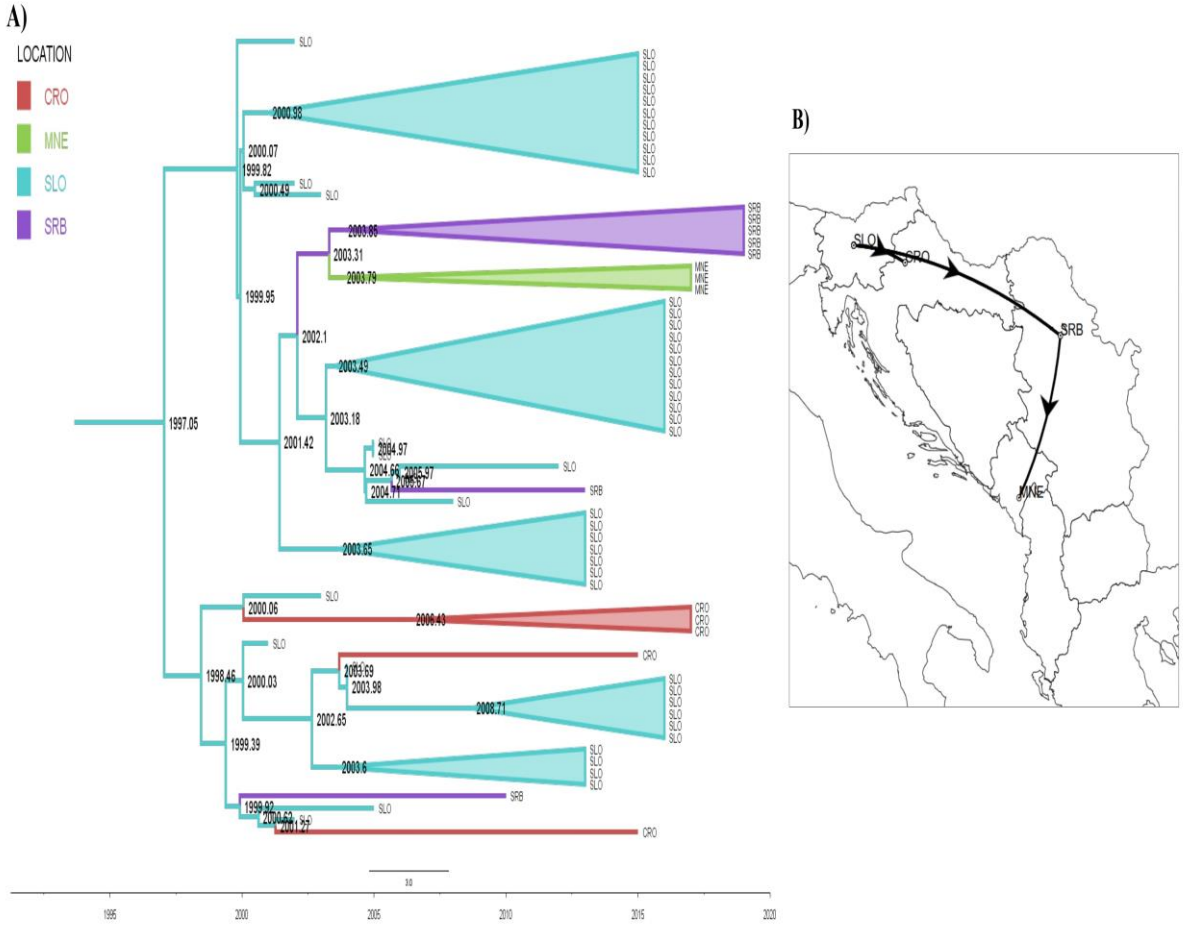


Figure 21. The location annotated MCC tree (A) and geographic representation of cross-border spread (B) Cluster 2. The origin of the cluster is estimated to be from Slovenia. Migration patterns supported by BF>3 and posterior probability > 0.9 are shown.

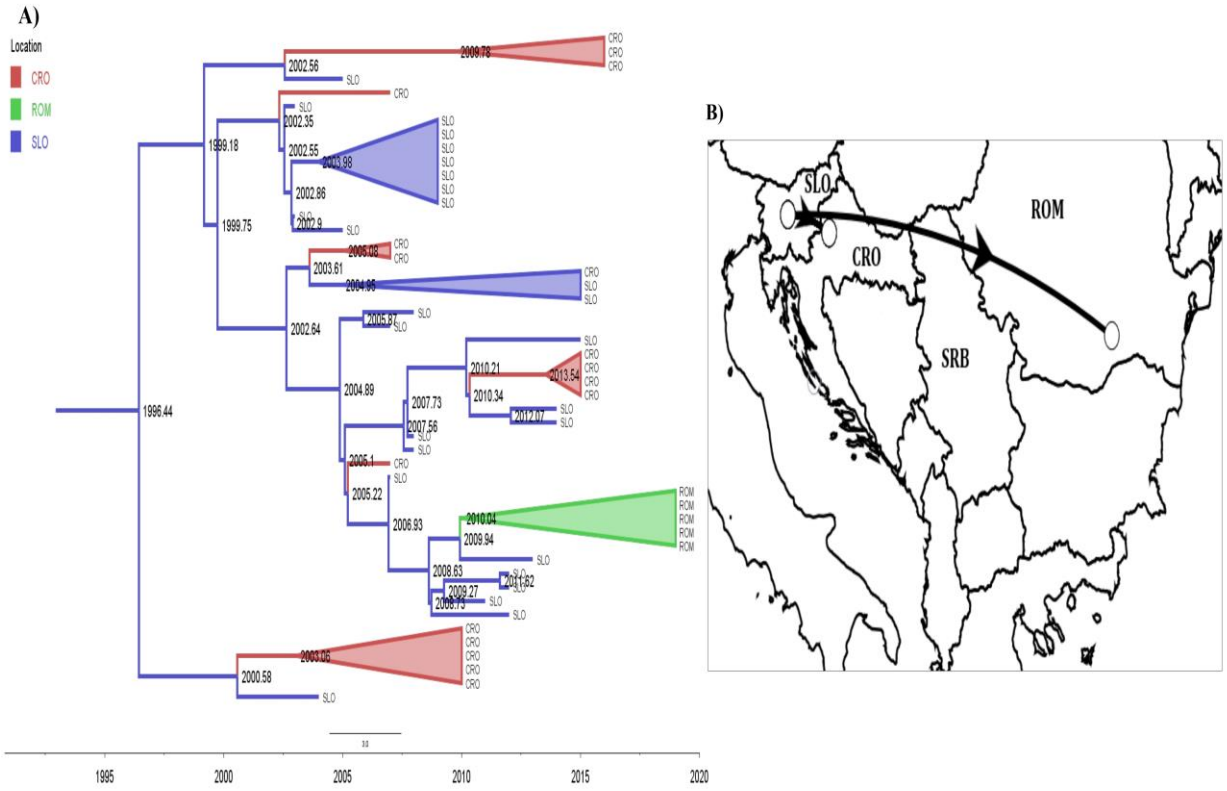


Figure 22. The location annotated MCC tree (A) and geographic representation of cross-border spread (B) of Cluster 3. The origin of the cluster is estimated to be from Slovenia. Migration patterns supported by BF>3 and posterior probability > 0.9 are shown.

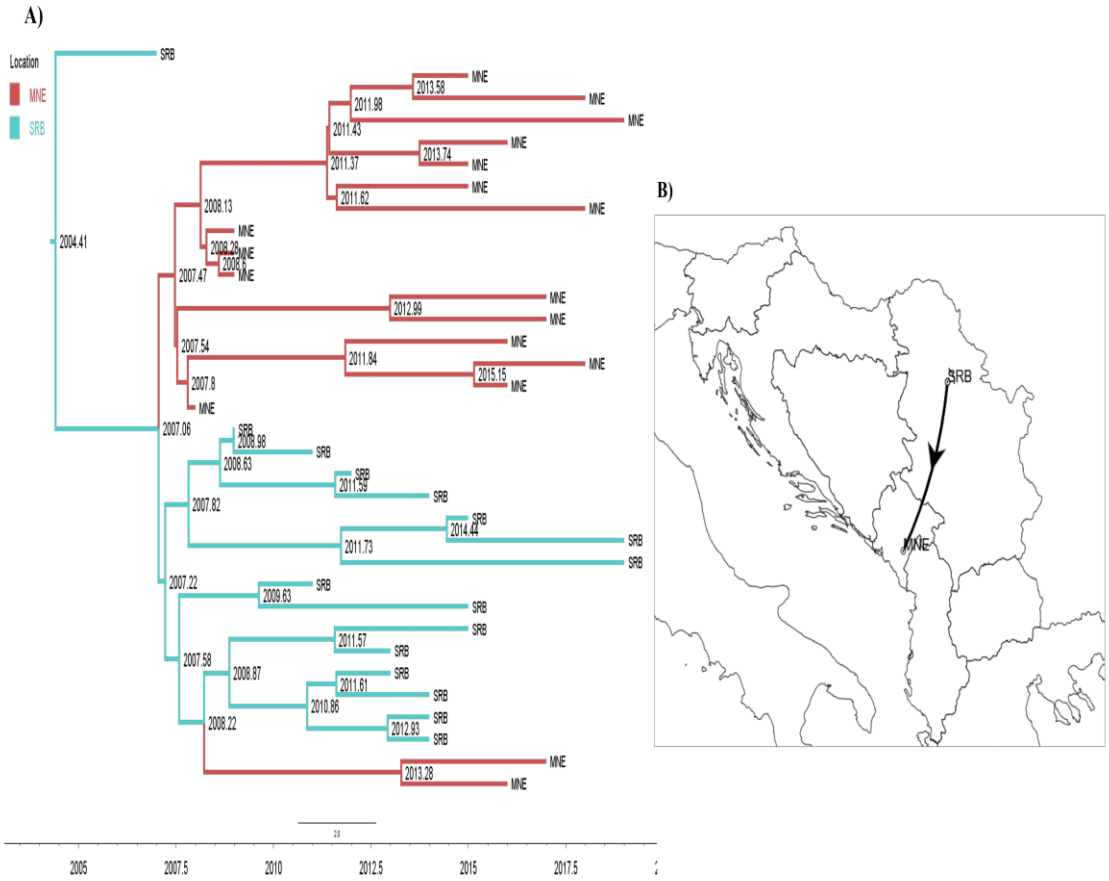


Figure 23. The location annotated MCC tree (A) and geographic representation of cross-border spread (B) of Cluster 4. The origin of the cluster is estimated to be from Serbia (direction of spread marked with the arrows). Migration patterns supported by BF>3 and posterior probability > 0.9 are shown.

4.6 Join-point Regression Analysis

Throughout the studied period (2004-2019), new HIV diagnoses in Serbia, Croatia and Montenegro showed a constant trend of increase, mainly due to MSM transmission where the highest values in APC were seen albeit an APC reduction is seen in MSM in Croatia it the last 3 years of the study period (Table 4, Figures 24, 25 and 26). Stable trends for the heterosexual population in Serbia and Montenegro were observed, with a reduction seen in Croatia (Figure 24, 25 and 26).

In Romania, there has been a reduction in the overall HIV new diagnoses since 2012 due to lower HIV incidence among PWID, nevertheless an increasing HIV trend related to both heterosexual and MSM transmission was observed (Figure 27). A similar situation was found

in Bulgaria, where the reduction in new diagnoses was seen only in PWID since 2008, with an overall increase in HIV new diagnoses (Table 4, Figure 28).

In Greece, negative values of APC implicated a reduction of HIV new diagnoses in both PWID and MSM. Greece experienced a large outbreak of HIV among PWID that peaked in the number of new HIV diagnoses in 2012 and decreased thereafter (Table 4, Figure 29).

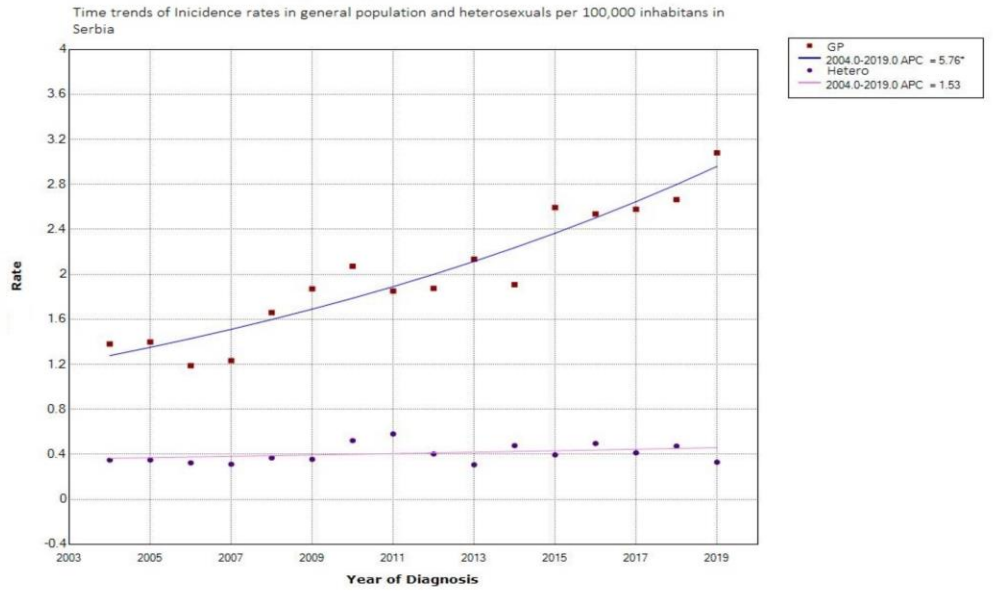
In Slovenia, likewise, a trend of reduction of new HIV diagnoses was observed since 2017, in the general population, as well as in all three studied sub-populations (MSM, heterosexuals and PWID) (Table 4, Figure 30).

Table 4. Per country time trends of the annual percentage change (APC) of new HIV diagnosis in the general population and in the main key and sub-populations in the period 2004-2019.

	APCs over time and joinpoints 2004-2019			
Country#	General population	Heterosexuals	MSM	PWID
Romania	2004-2012= 21.47 **2013-2019= -4.44***	2004-2009= 19.06 *2010-2019= 3.73**	2004-2010= 51.35 *2011-2019= 7.88*	2004-2012= 137.49 **2013-2019= - 22.98**
Greece	2004-2012= 10.47 **2013-2019= - 5.99**	0.84*	2004-2009= 14.08 *2010-2015= 1.99** **2016-2019= - 12.55**	2004-2012= 74.41 **2013-2019= - 26.42**
Serbia	5.76	1.53*	9.55	2004-2009= 36.77* **2010-2019= - 11.63***
Bulgaria	2004-2009= 22.56 *2010-2019= 6.33*	5.61	2004-2014= 30.06 *2015-2019= 8.38**	2004-2008= 60.71 **2009-2019= - 7.53**
Croatia	5.55	**-4.27**	2004-2016= 11.07 **2017-2019= - 6.85***	**-6.04**
Slovenia	2004-2016= 4.35 **2017-2019= -13.73***	2004-2016= 4.35 **2017-2019= -13.73***	2004-2016= 4.04 **2017-2019= -18.36***	2004-2016= 4.04 **2017-2019= -18.36***
Montenegro	11.63	2.52*	13.16	0.15*

Annual percent changes (APC) in HIV infection rates per 100.000 inhabitants for 2004 – 2019; in case of a shift in APC, the values for time periods prior and after the year of change („joinpoint") are shown, along with indicated time periods; negative APCs correspond to the reduction of infection rate and they are written in bold. Decreased positive APC values are shown in italic; APCs that didn't reach statistical significance are marked with asterisk (*). Abbreviations: MSM - men who have sex with men; PWID- people who inject drugs. #countries are listed in descending order, by total number of HIV infections

A)



B)

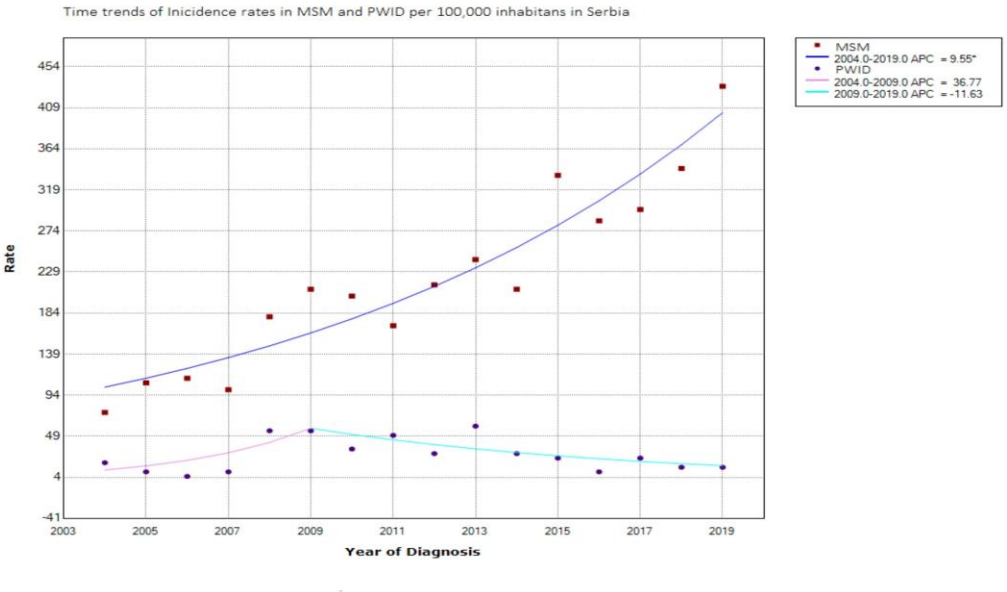
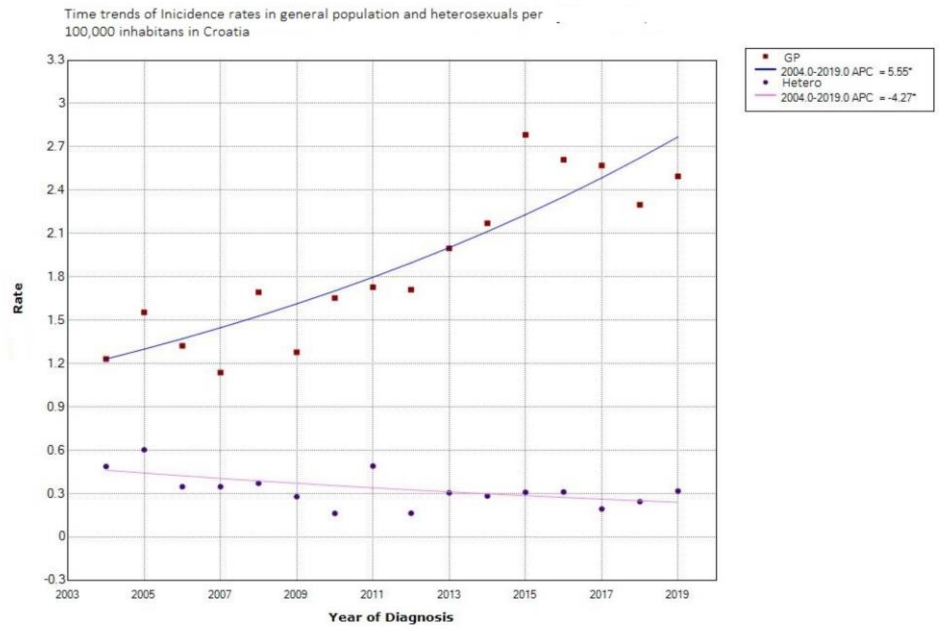


Figure 24. Time trends of incidence rates per 100.00 inhabitants in Serbia: in general and heterosexual populations (A) and MSM and PWID (B)

A)



B)

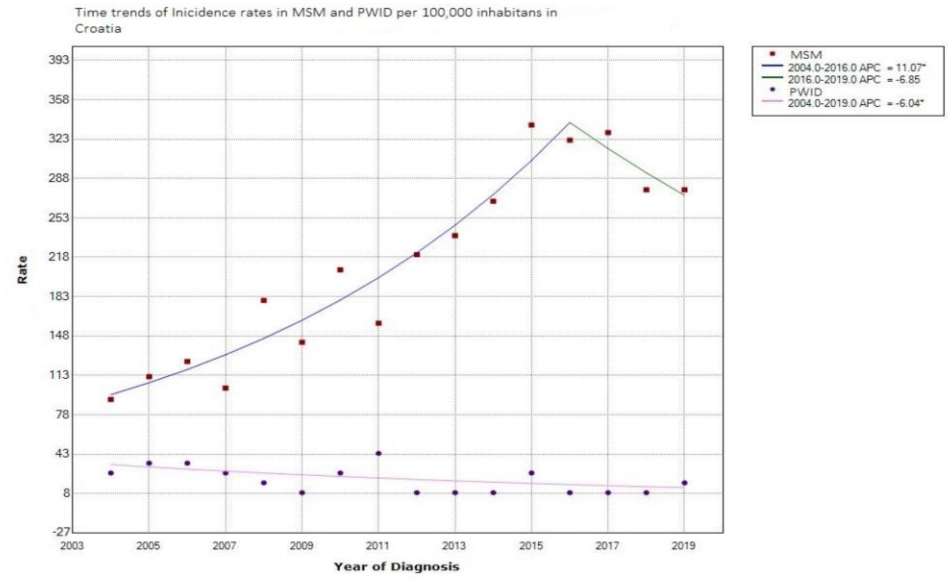
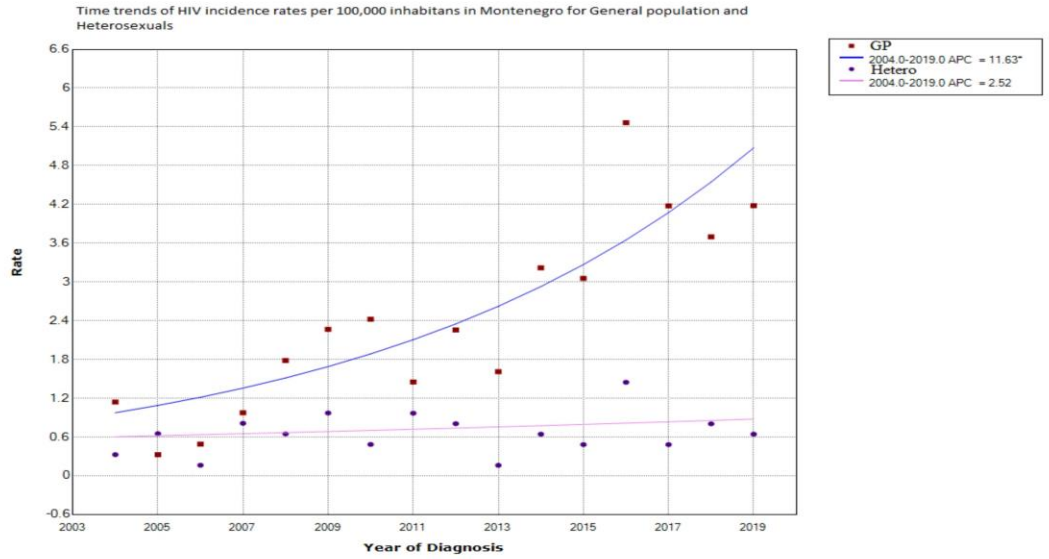


Figure 25. Time trends of incidence rates per 100.00 inhabitants in Croatia: in general and heterosexual populations (A) and MSM and PWID (B)

A)



B)

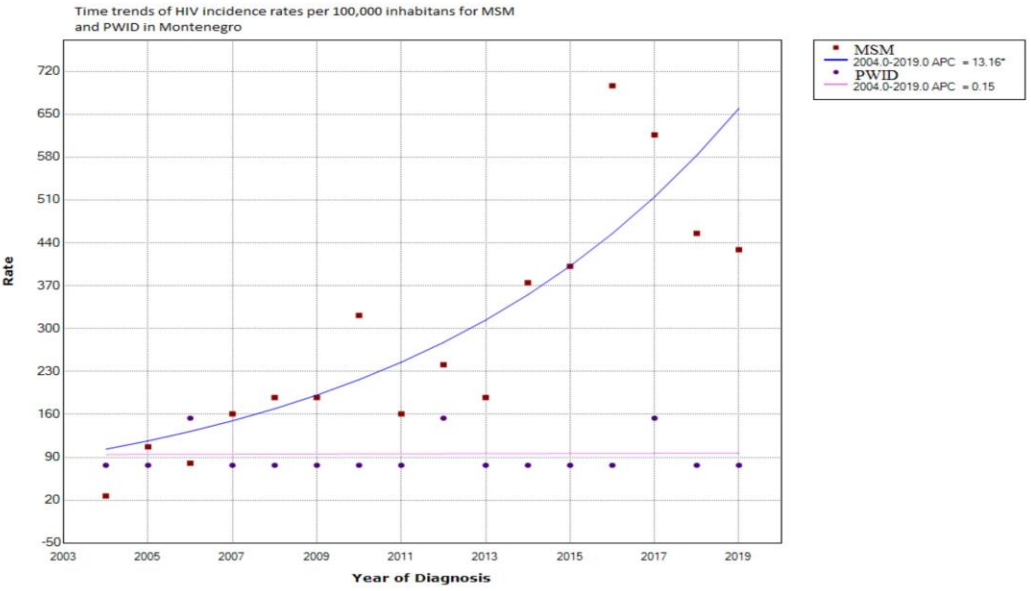
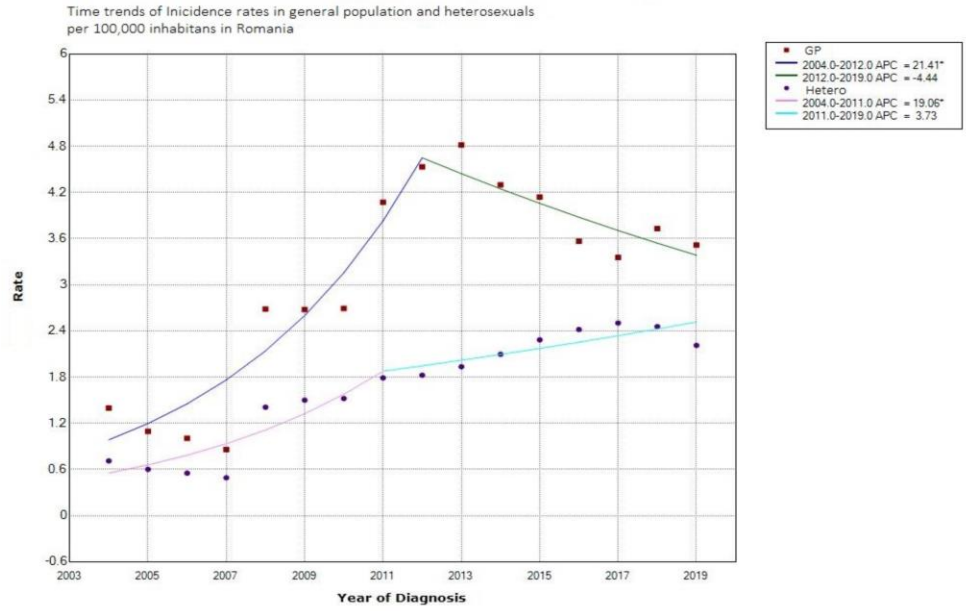


Figure 26. Time trends of incidence rates per 100.00 inhabitants in Montenegro: in general and heterosexual populations (A) and MSM and PWID (B)

A)



B)

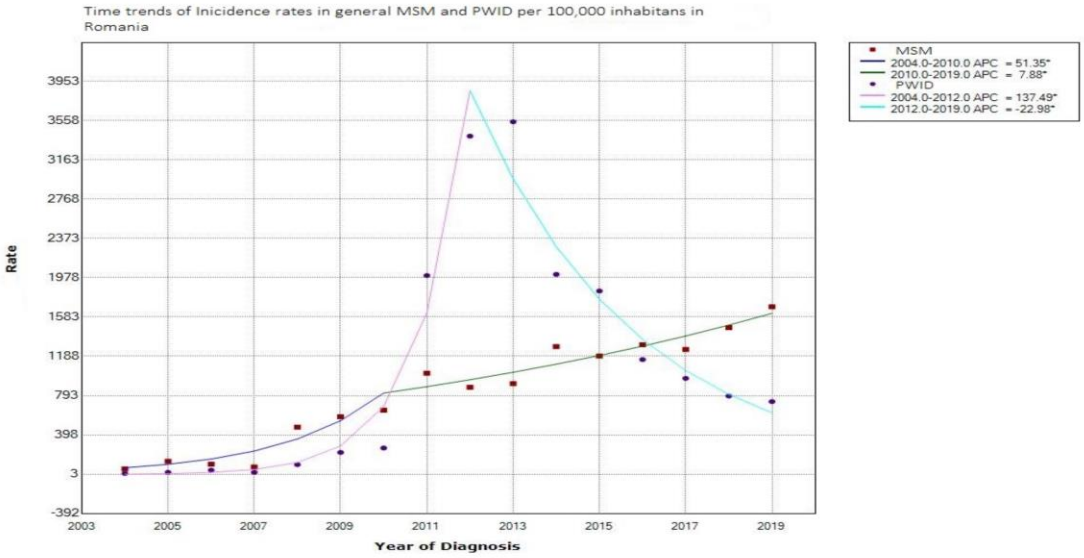
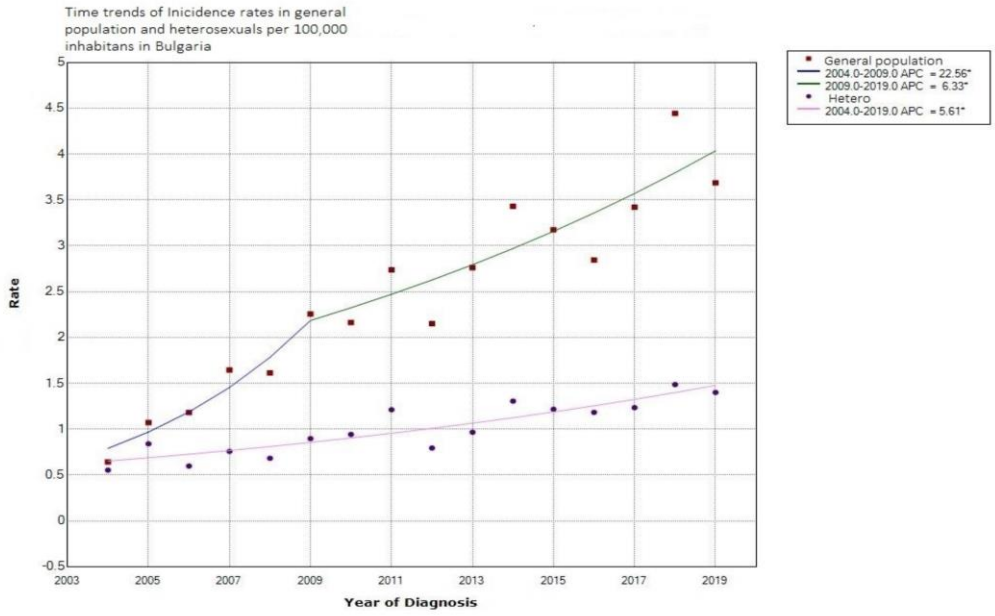


Figure 27. Time trends of incidence rates per 100.00 inhabitants in Romania: in general and heterosexual populations (A) and MSM and PWID (B)

A)



B)

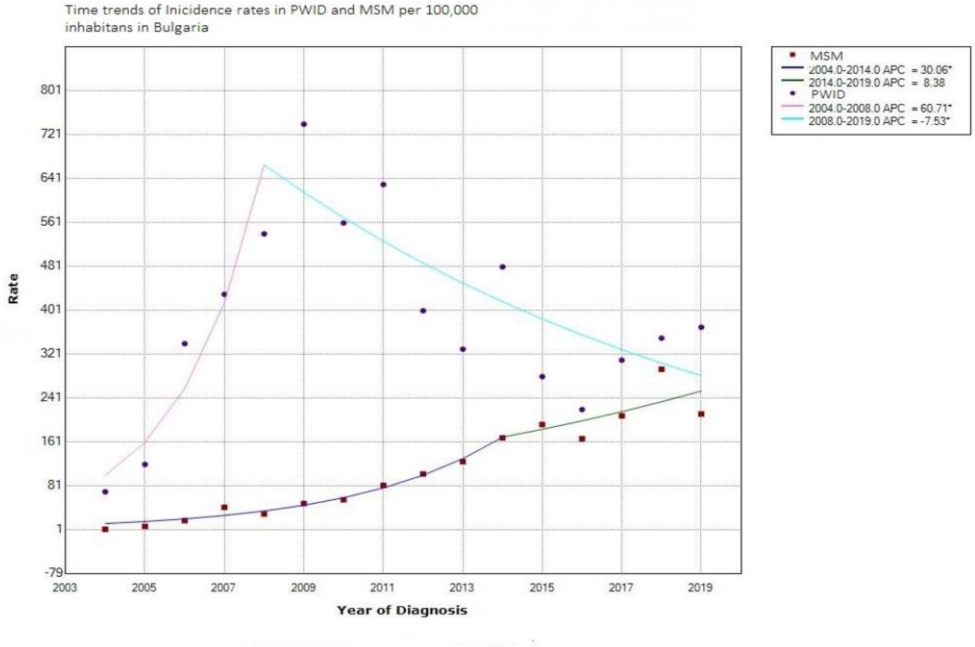
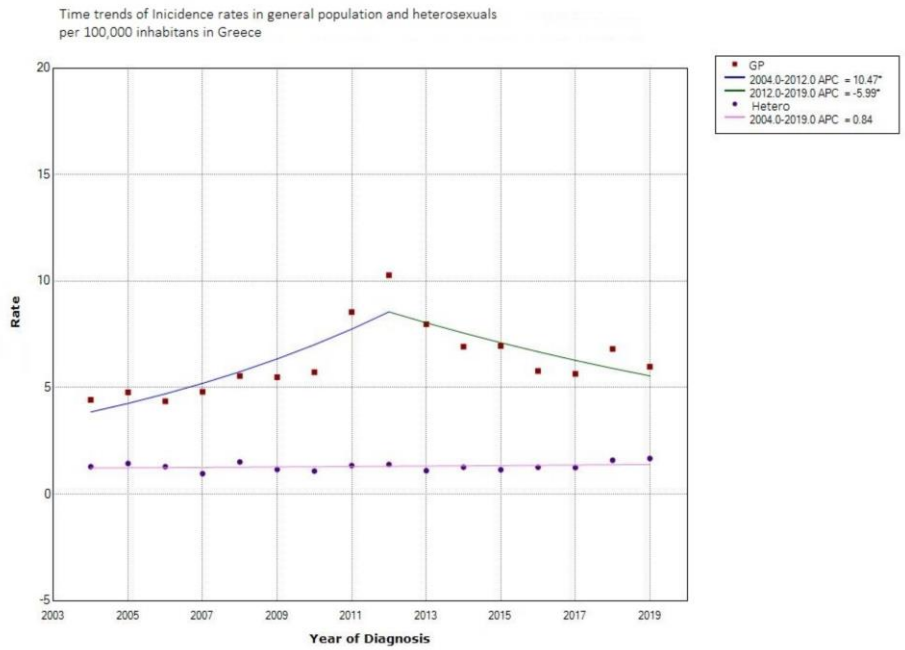


Figure 28. Time trends of incidence rates per 100.00 inhabitants in Bulgaria: in general and heterosexual populations (A) and MSM and PWID (B)

A)



B)

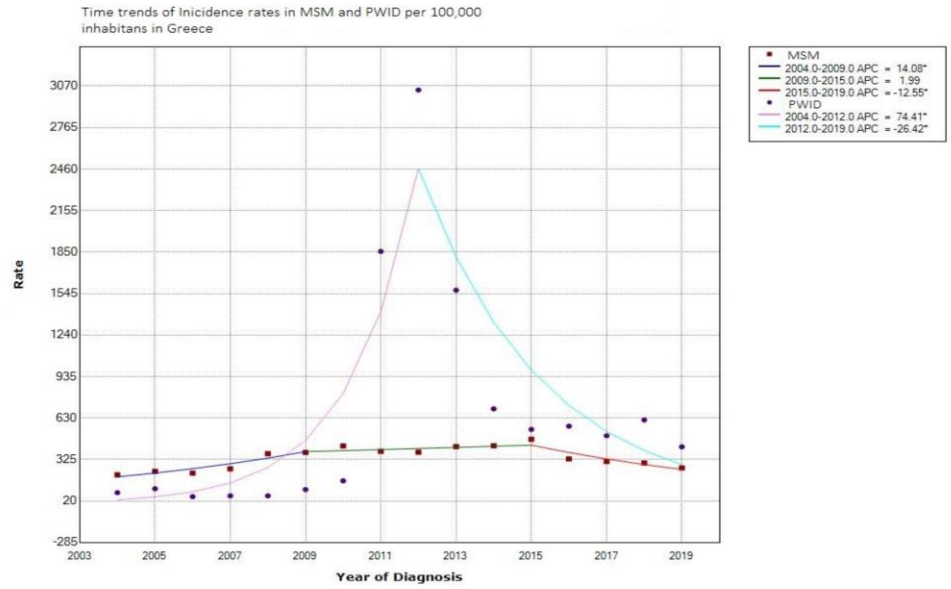
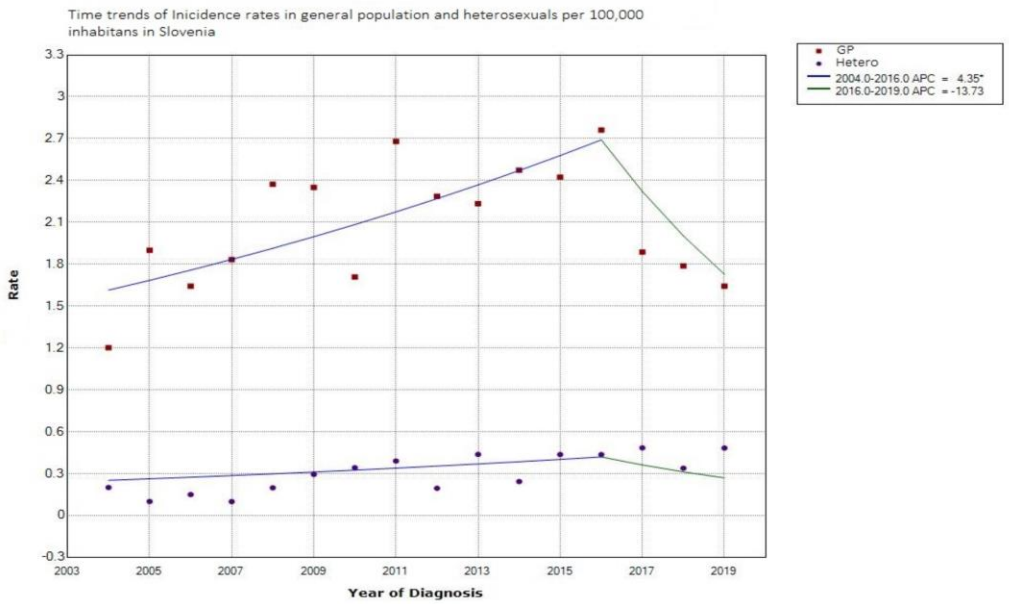


Figure 29. Time trends of incidence rates per 100.00 inhabitants in Greece: in general and heterosexual populations (A) and MSM and PWID (B)

A)



B)

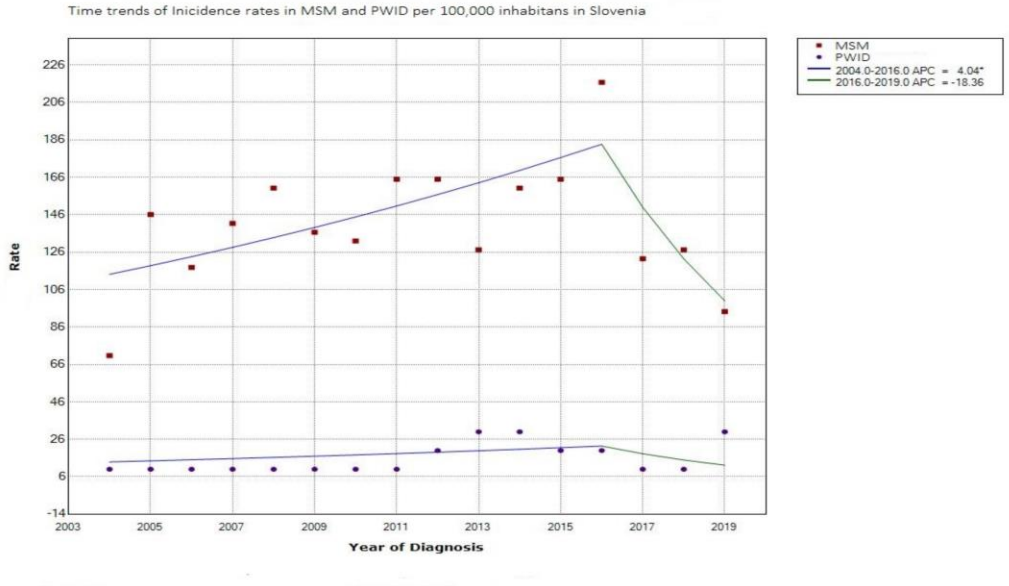


Figure 30. Time trends of incidence rates per 100.00 inhabitants in Slovenia: in general and heterosexual populations (A) and MSM and PWID (B)

5 Discussion

Monitoring of the HIV epidemic is essential for assessing the impact of effective HIV prevention interventions, determining public health priorities, and estimating current and future health care needs. The Central European region encompassing Balkan countries and is considered a low HIV incidence region, however, national epidemics in Central Europe have a potential for significant further growth and constant increase in incidence has been seen, with potential for substantial increase in incidence (18, 19). Such a situation underlines the importance of primary prevention regarding key populations such as implementation of PrEP, condom usage promotion and diverse harm reduction programs.

Lately, the need to flatten the curve has entered mainstream discourse in the era of a COVID 19 pandemic, and a reproductive number R has been widely used to illustrate the state of the epidemic The main reasons for using R number as the parameter of choice for global pandemic surveillance is that it is relative in nature, therefore allowing the results between different populations to be compared, and also it is intuitive and easy to understand: i) R above 1 indicates an active epidemic with the potential of further growth; ii) R below 1 indicates the downward trend of the outbreak (3). One major issue with R is the fact that this parameter does not take into consideration differences in population density and it cannot be representative in compartmentalized epidemics such as HIV epidemic (3). Therefore, UNIADS implied other epidemiological parameters such as incidence/prevalence ratio, defining the cut-off value of up to 3% to achieve the reduction of the epidemic and in our study none of the participating countries reached this target (3).

Therefore, in order to assess characteristics of incidence growth in general and key populations we used the APC of HIV incidence and joinpoint regression. This is further used to illustrate the nature of the HIV epidemic and effects different preventive policies implemented in different Balkan countries. The obtained results implied a downwards trend of the overall HIV incidence in some Balkan countries, namely Greece, Slovenia and Romania, whereas in others, a further increase of HIV incidence was seen, i.e. Serbia, Croatia, Montenegro and Bulgaria.

In Slovenia a reduction of HIV incidence has been seen across all key populations in the last 4 years of the study period, which could probably be linked to targeted prevention efforts implemented in the country, through implementation of PrEP, increased condom coverage in key populations and good coverage of PWID population with harm reduction programs, as stated in ECDC report on combination prevention (108). A similar situation is seen in Greece, where rapid response against HIV outbreak in PWID under the ARISTOTLE national program, as well as condom coverage in key populations on the national level may explain the reduction in the overall HIV incidence and sharp reduction in HIV incidence in PWID and MSM (31-33). On the other hand, the obtained HIV incidence trend in Romania indicates that the reduction in overall cases of HIV is mainly due to successfully combating the HIV outbreak in PWID in the 2010s, whereas sexual transmission remains a problem that needs to be addressed further.

Despite constant increase in the yearly number of new HIV diagnosis that has been seen in Serbia, Bulgaria and Croatia, these countries saw good results when it comes to 90 - 90- 90 targets, not significantly different from the average values for Western Europe,

according the UNIADS and other literature data (3, 19). On the other hand, lower testing coverage of about 60% is estimated in Montenegro, which is much lower than European average, and only 50% of them received the therapy (19).

The constant trend of increase in HIV incidence in Bulgaria, Serbia, Croatia and Montenegro, seen also in our study, is indeed worrisome. According to the ECDC report on HIV combination prevention (108), STI testing coverage is high among key populations except in Montenegro and sexual education programs are implemented mainly in the teenage population through peer education, but condom coverage in key populations remains low in these countries. This could partly explain observed trends of increased sexual transmission mainly in MSM in Serbia and Montenegro whereas early signs of reduction in HIV incidence in MSM in Croatia seen in the last 3 years of the study period could be related to national availability of PrEP and increased condom coverage in key populations reported in 2018 (108).

The incidence trends for the heterosexual population in Serbia and in Montenegro have been stable, with a reduction seen in Croatia. The reduction in HIV incidence among PWID has been seen in all three countries (Croatia, Serbia and Montenegro) delineating ongoing needle exchange programs (although deemed insufficient) and availability of methadone substitution therapy (108). In Bulgaria similarly as in Romania a reduction in HIV incidence is seen in PWID but sexual transmission (MSM and Heterosexual) is still on the rise which explains the overall increase in HIV incidence in Bulgaria.

Furthermore, results obtained from, logistic growth modeling of the epidemic in Serbia that was performed on cumulative numbers of new HIV cases for 1984-2016 time period have successfully predicted further growth of HIV epidemic in Serbia especially within MSM, where an early exponential phase of growth was seen. The results of growth simulation for 2017-2030 time period showed that plateau will not be successfully achieved by the end of the simulation period especially in the MSM population. Moreover, the incidence trends of new HIV diagnosis in Serbia obtained from an updated data set for the time period 2004-2019 for the general population and MSM are in alignment with growth simulations.

On the other hand, this growth was not as steep in heterosexual and PWID population indicating more favorable epidemiological situations in these key populations. Such results are in alignment with the incidence trends obtained by joinpoint analysis of corresponding HIV incidences in Serbia. This situation illustrates the importance of investigation of HIV epidemic within different key populations, because the burden of HIV infection is not proportionally distributed through the general population.

With the prevalence of HIV infection below 0.1%, Serbia is ranked as a low-prevalence country; however, it is of great importance to study the epidemiological situation in populations at risk such as the MSM and PWID transmission groups and among sex workers. Overall, the highest prevalence of HIV in Serbia has been estimated for MSM (8.3%), much higher when compared to the general population, as well as to other groups at risk such as sex workers (1.8%) and PWID transmission group (1.6%) (3, 109).

Potential disadvantages of the applied approach include a number of potential biases and difficulties, such as information bias, lack of data, under-reporting, lack of information on changes over time etc. But the observed trend of increase of new HIV infections are also in line with results obtained in the Region. Although the number of new HIV cases is decreasing

globally, it is increasing in the WHO European region, mostly in Eastern and Central Europe (22).

In Serbia, these findings might reflect that, though efforts have been made to decentralize preventive and advisory services, the coverage of high-risk groups with preventive services could still be insufficient. This may be due to lack of recognition of risk behaviors or fear of further stigmatization, especially in smaller communities (3).

Finally, this growth simulation study failed to take into account the COVID 19 pandemic. Even UNAIDS in its report for the year 2020 pointed out that the current epidemiological situation will negatively affect the global progress in ending HIV by the year 2030 (3). By November 2020 a total of 55 new HIV cases were registered in Serbia, which was widely cited in the news-outlets, since it is the lowest number of new HIV cases in the last 20 years (110). Complete data for 2020, however revealed 123 new HIV cases, still almost half of the number reported for 2019, with overall similar epidemiological characteristics as seen before: the high male to female ratio, congregation of new cases in Belgrade and other urban areas, and dominance of MSM transmission route (22). Similar trend of reduced HIV incidence is seen in the rest of the continent and particularly worrisome is situation in the Eastern Europe were reduction in novel HIV cases is pared with an increase in AIDS related mortality (19). Therefore, it can be hypothesized that the reduction in HIV registration and reporting could be due to drained strategic health care resources by the COVID 19 and that, as this epidemic is ending, a surge in HIV cases could be expected. This was underlined in the report by Institute for public health of Serbia for 2021, which stated that by November 120 new cases were registered, more than 100% increase than in the same period in 2020 *(110)*. Overall epidemiological trends such as high male to female ratio and MSM predominance remained the hallmark of Serbian HIV epidemic in 2021 as well (*110*).

Similarly, recent phylogenetic study from Greece followed the active ongoing transmission among PWID following the outbreak in the 2010s and rapid response programs (111). Although Greece saw a reduction in HIV incidence across all key populations, thus slowing the HIV spread, the result from this study serves as a reminder that HIV epidemic is "chronic" and resource expensive public health problem that can exacerbate in the time of social and economic crisis such as COVID 19 pandemic or global financial crisis from 2008, in which aftermath the last large HIV outbreak among PWID from the 2010s happened.

In order to elucidate the pattern of the HIV epidemic spread in Serbia, based on HIV-1 pol sequence data generated through the antiretroviral resistance testing, we investigated the transmission clusters and phylodynamic profiles, using in-depth phylogenetic analyses that involve the maximum likelihood and Bayesian coalescence strategy. These kinds of sophisticated analyses provide an insight into the epidemic trends, patterns of evolutionary history of a certain viral population, revealing the size of transmission clusters and the dynamics of transmission within them.

Although the *pol* region is relatively conservative when compared to the *env* or *gag* region it is the most widely used gene region for phylogenetic analysis although it can create problems: i) lover mutation and recombination rate; ii) the convergent evolution due to HAART iii) pol region covers less than 20% of HIV genome instigating the precision of phylogenetic analysis (112, 113). However, despite all of these limitations pol region contains sufficient information for the majority of phylogenetic applications in HIV surveillance. Moreover, the genotype resistance testing in the last two decades resulted in large

repositories of *pol* sequences which are particularly important for phylogenetic research (114)

Moreover, in recent years a plethora of sophisticated algorithms for phylogenetic analysis emerged using different approaches and statistical tests as well as different cut-off values (89). For example, there are numerous algorithms that generate phylogenetic trees based on nucleotide distance as a measure of similarity between sequences. Such an approach is fast and suitable for analysis of related sequences and it is usually the first step in phylogenetic analysis. On the other hand, maximal likelihood methodology is the criterion-based approach. By using a specific substitution model and through bootstrap, an iteration-based analysis, multiple evolutionary scenarios are getting tested and the most likely scenario is depicted as a phylogenetic tree with statistical support values. Such an approach is used in MEGA software, one of the most popular programs for phylogenetic analysis (https://www.megasoftware.net/). Furthermore, when the Bayesian approach is used the substitution model is selected by the user through adjustment of different prior parameters. The result of such approach is a consensus tree with statistical support calculated as Bayesian posterior probability. This approach is implemented in another popular phylogenetic tool Mr Bayes (88). Combining different statistical approaches when defining clusters in local epidemic makes for robust analysis and in this study, local clusters were defined using a more rigid or more relaxed set of criteria including both i) similarity expressed through nucleotide distance and ii) statistical support based on bootstrap or Bayesian posterior probability values.

Furthermore, molecular clock and phylodynamic analysis of most expanded transmission clusters can further illustrate hot spots of transmission within a national epidemic and reconstruct the transmission history of the cluster, which could alleviate the problem of the absence of intermediary and ancestral sequences. This is of great interest when studying all epidemics that spread in compartments such as the HIV epidemic (115).

Interpretation of these results, however, may be hampered if the analyzes are not associated with other clinical and epidemiological traits, which serve as determinants of the transmission dynamics and cluster activity (116-118).

Our previous study established the role of transmission clusters/networks in HIV epidemic spread in Serbia, however phylodynamic aspects were not considered nor the impact of different socio-demographic and clinical characteristics on the clustering patterns (26, 27). In this research we found high levels of transmission clustering among Serbian MSM: nineteen HIV transmission clusters and one large transmission network have been identified. Existence of the largest identified network in our study, comprising 63 sequences in the time span of 19 years, suggests that HIV spread is still driven by local MSM transmission, as observed also in other European countries (116-119).

Similar to some other European countries, clustering sequences from MSM were found within subtype B (117). Furthermore, we found an extension of previously reported small transmission clusters related to young newly diagnosed MSM patients (26, 27). Most of the heterosexual transmissions seemed to be limited to transmission pairs and small clusters, without substantial further spread of the infection. All things considered; our results show that local HIV transmission in Serbia is mainly driven by MSM transmission clusters.

In this thesis, we tried to illustrate phylodynamic differences between two main key populations in Serbia MSM and heterosexuals. In phylodynamic analysis only expanded

clusters with more than 10 sequences were used. That was a problem, due to the fact that all clusters predefined happened within subtype B and MSM transition route dominantly. In order to dill with this problem two heterosexual monophyletic clades were studied, although they did not meet the criteria for transition clusters one of subtype B and the other of subtype C, which encompasses all subtype C sequences isolated in the study period. Such an approach allowed us to further analyse the overall epidemiological trends, but on the level of different compartments and subtypes.

Based on the Re trends estimated through birth-death plots, together with the number of infections over time, significant differences between the MSM and heterosexual clades were found. Overall, MSM clades showed mean reproductive numbers over one during the whole investigated period, and higher values of Re were observed in clusters with more recent tMRCA. Nevertheless, relatively low values of Re were observed in the studied clades ranging up to 1.2. This could be explained by the modus of transmission. Since HIV is transmitted through sexual contact or parenterally we could expect lower values of Re than in respiratory viruses where R numbers are usually above 2 and can be as high as 18 for measles outbreaks (120, 121). In a recent large-scale phylodynamic analysis of HIV clusters in Ukraine, Portugal and Senegal Re values ranged up to the maximal value of 2 which is relatively similar to Re trends obtained from analyzed local clusters and cross-border clusters (122-124).

Although the transmission network was defined to be active, with reproductive potential striking only slightly higher than 1 it also appears to be “aging”, while clusters appear to be of lower mean age; especially the most expanded cluster that was found to be “the youngest” was also found to be with the highest reproductive potential. On the other hand, particularly worrying is the fact that even though the transmission network comprised sequences from MSMs diagnosed early in infection, this does not seem to have stopped clusters from growing. Other research showed that a large part of onward transmission among MSM is related to the early phase of infection (125-127).

HIV transmission dynamics is considered to be shaped by a number of diverse constraints within the host, upon a transmission process and at the population level. These further influences viral evolution (within-host and inter-host), but also social and demographic factors can influence the spread dynamics (128). Hence, the inference of HIV-1 transmission dynamics and factors influencing epidemic spread may provide an important input for the design of efficient public health interventions.

The LCA analysis provided further insights by combining clustering patterns and a set of characteristics associated with patients such as socio-demographic, clinical, transmission risk, diagnosis date. Together, these features allow us to assess multiple determinants of the local HIV transmission, helping to understand its occurrence in a “real life” manner. This analysis has separated five subgroups (latent classes) with various combinations of analyzed characteristics that affect epidemics differently on epidemics in Serbia. Importantly, four factors were found to be significantly associated with patients belonging to clusters: very young age, from the capital city, recently diagnosed and with a high rate of STI and HBV coinfections.

At the end it should never be forgotten that clusters are epidemiological categories derived from contact tracing, which is the main tool of epidemiologists when analyzing the spread of infectious disease in the population (129). But conventional epidemiological approaches can suffer from different types of bias. The absence of bias makes a study valid.

But this validity can be oriented inward (to the study population) or outward (results can be generalized-results as universal truth). The presence of bias means that the results are not representative and in epidemiological research is divided as: i) selection bias and ii) information bias (130).

Information bias in HIV research is mainly connected to gender and transmission risk data. This problem arises from stigmatization of different Key populations (MSM or PWID) or criminalization of undisclosed HIV transmission. Such, climate prevents people who live with HIV to seek preventive measures, testing and counseling (3). The effect of stigma further creates a selection bias at the beginning of the surveillance cascade (counseling center) resulting in over-representation of negative results or over-representation of particular key populations such as "*out of the closet*" MSM if robust peer support groups are present (131). Furthermore, data on transmission risk, gender and age could be false or non-existent due to the right to anonymity creating the information bias.

Moreover, in this study, a question of dynamics of the HIV epidemic in the Balkans was approached through analysis of general HIV rate trends as well as phylogenetic, phylodynamic and phylogeographic analyses. Of note is that this is the first comparison epidemiological analysis between different national epidemics complemented with a phylogenetic analysis of HIV spread in a region where HIV is not distributed proportionally in regard to the size of national epidemics as well as subtype distribution. Moreover when we take into account the size of the sequence dataset (over 2000 sequences analysed) this part of the thesis is the first big data phylogenetic analysis regarding the Balkan region. Such a situation created a set of problems that needed to be addressed in the study design and methodology selection: i) comparison of HIV incidence rates between populations different in size and HIV prevalence: ii) the selection of relevant HIV clades; iii) selection of phylogenetic tools adapted for large datasets.

As mentioned above, the first problem was successfully addressed by using the relative nature of APC and joinpoint regression analysis (see chapter 5.1). This insight into the dynamics of the number of new HIV diagnoses over time provided epidemiological context for the complex phylogenetic analyses of HIV-1 subtype B in the Balkans. As mentoned before HIV subtype B was selected since it is most prevalent in majority of Balkan countries exscept in Romania where it is an emerging subtype, instigating the question of relatedness and clustering of this subtype across different countries.

The third problem regarded the size of the alignment as well as the limited computational power of our local servers, therefore a different maximal liklehood methodology implemented in IQTree web server needed to be used. Namely, these recent and fast measures of branch support (approximate likelihood ratio test [aLRT] and Shimodaira–Hasegawa [SH]-aLRT) provide a compelling alternative to slower conventional methods such as boostrap or maximal parsimony, offering not only speed advantages but also excellent levels of accuracy and power as stated in the comparison study by Anisimova M et al. (132). Namely, this study compared the results obtained from standard bootstrap methods (SBM), aLRT, Bayes posterior probability as well as SH-aLRT tests. The simulation studies as well as real data analysis revealed that SBM are prohibitive for large data sets and that it is rather conservative and slower than SH-aLRT which had higher statistical power as well as resistance to severe model violations, where aLRT and Bayes posterior probability gave higher proportions of false positive results (132). These disadvantages of SBM are well known in the literature and although the bootstrap value (BV) above 0.9 should be considered as

sufficient statistical support the BVs above 0.75 could be sufficient as well, if confirmatory analysis is provided (similarity and/or posterior probability) (25, 26).

Nevertheless, our own experience taught us that confirmatory analysis should be included even when SH-aLRT value over 0.9 is present, especially when unexpected paraphyletic relations are observed. In the literature SH-aLRT is usually paired with the ultrafast bootstrap method (UFBoot) as a confirmatory analysis of choice (133). This test was introduced in 2013 in simulation study by Minh B Q et al. Namely simulation study revealed that UFBoot approximation approach outperforms the RAxML rapid bootstrap in terms of the computational time, achieves more unbiased support values as SBM, and is relatively robust against moderate model violations (134). Bout SH-aLRT as well as UFBoot tests are available in IQTree web server where UFBoot tests are recommended as confirmatory analysis to SH-aLRT. UFBoot values above 0.95 are sufficient to support a transmission cluster of interest (133).

Furthermore in this study different BLAST search strategy was used to prove cross-border cluster sustenance. Namely, in our phylogenetic practice, it is common to include similar background sequences obtained through BLAST to prove cluster sustenance and to exclude falsely positive results and artifacts (25, 26). Usually, the range of homology is expressed through similarity and there is a plethora of algorithms used to search for homology including BLAST as implemented in the NCBI database. But, as mentioned above another measure of similarity is genetic distance, therefore in order to express the range of homology between consensus sequences with every single sequence from relevant clusters we calculated an average pairwise nucleotide distance. With average nucleotide distances below 2%, the consensus sequences were similar enough taking into account the structure of the dataset and the fact that clusters were not selected based on similarity but on statistical support. Such a relaxed approach is deemed suitable for large datasets collected over a long period of time and several different national epidemics (89).

The main goal of this phylogenetic analysis was to detect cross-border clusters and to further analyse its role in HIV spread in the Balkans. In general, frequently traveling and migrant populations (i.e. truck drivers, military personnel, seamen etc.) are a significant driving force of the epidemic in the regions where the HIV epidemic is generalized (135-137). In Europe, changes in subtype distribution have been linked to migration influx, yet further propagation of these subtypes was found to be dominantly local, confined geographically and by transmission mode, with occasional cross-border cluster formation (57, 58, 59, 138).

In our study, phylogenetic analysis of subtype B spread in the Balkans revealed the ongoing transmission mainly through local, within-country clusters, with an average prevalence of clustering of 68%. However, the proportion of clustered sequences in Montenegro and Croatia was found to be higher, of around 80%, with a substantial difference in the proportion of clustered sequences within cross-border clusters. This result suggests that subtype B spread in Montenegro is akin to spread in surrounding countries, mainly Serbia, whereas in Croatia local clustering predominates. Lower clustering prevalence observed in Romanian sequences (39%), might be linked to the rather recent emergence of this subtype in Romania, shown to have arisen from multiple introductions.

Although local HIV transmission is dominant in Europe, cross-border clustering has also been reported, often linked to the changing patterns of migrations, resulting in geographically dispersed clusters, or due to connections between countries that are relatively close or share the border, underlining socio-demographic influence on clustering (59,73).

Historical, sociopolitical, and even geographical context of the Balkans is notoriously complex and turbulent, laying the grounds for these interactions. In decades concomitant with HIV/AIDS epidemics, countries of the region have faced major social and political changes: the demise of socialism, transition economy; in the 1990s, violent conflict and war started the break-up of Yugoslavia into several separate states, four of which are included in the study (Slovenia, Croatia, Serbia, Montenegro); in the 2000s, along with the further dissolution of ex-Yugoslavia, some Balkan countries joined the European Union (Slovenia, Romania, Bulgaria, Croatia), whereas in mid-2010s, within European migrant crisis, Balkan nations have seen a significant inflow of refugees and migrants from the Middle East and Africa. Our phylogeographic analysis implied that the spread within ex-Yugoslavia started in the early 1980s, whereas introductions into Romania started to happen in the mid-1990s as seen earlier (69).

In the current study, four cross border clusters have been identified, comprising a relatively small fraction of the total clustered sequences (16%) and composed of sequences originating from countries of ex-Yugoslavia (Serbia, Croatia, Slovenia, Montenegro) and Romania. Estimated tMRCA for these 4 clusters was in the mid-nineties (Clusters 1-3), coinciding with the partition of Yugoslavia and 2005 (Cluster 4), along with the final dissolution of Serbia and Montenegro. Phylodynamic analysis of the observed cross-border clusters revealed their prolonged activities and cross-border spread, especially between Serbia, Croatia, Slovenia and Montenegro, well after dissolution of Yugoslavia - through the early and mid-2000s since when a decrease in Re values and stagnation in cluster growth were observed.

No sequences from Bulgaria and Greece were found within the Balkan cross-border clusters, which might reflect the possible sampling bias; however, a recent study exploring the origins of HIV subtype B epidemics in Bulgaria has indicated links to multiple countries, including Israel, Germany, Spain, UK, Russia, yet no connection to the neighboring Balkan countries emerged (66). In a study exploring HIV subtype B migratory pathways across Europe, evidence for directional viral dispersion was detected, where Greece and Serbia, the only Balkan countries included in the study, were found to act as sources of migration events towards the west Europe, albeit the number of migration events from these countries to west Europe was low (139).

Moreover, the low levels of dispersal between Greece and Bulgaria and the ex-Yugoslav area can be explained since these countries don't belong to the ex-Yugoslav space where, probably, higher population mobility exists across the borders of the current countries. In the present study, that is one of the few looking into detail the HIV dispersal pattern within the Balkan area. Serbia, Slovenia and Montenegro emerged as putative exporters of HIV subtype B, whereas Romania emerged as importer of HIV subtype B from ex-Yugoslav space (Serbia, Croatia, Slovenia and Montenegro).

6 Conclusions

According to the defined objectives and based on the obtained results the following conclusions can be reached:

- Transmission dynamics of HIV spread in Serbia is linked to transmission clusters formation, which are connected to MSM transmission and subtype B.

- Phylodynamic analysis of active clusters coupled to LCA analysis confirmed young MSM with concomitant STIs as the main target group for planning structured preventive policies.

- Presented results imply that HIV epidemic in Serbia is still in the exponential growth phase, in particular related to the MSM transmission that is estimated to retain the steep growing curve until 2030.

- HIV subtype B dispersal pattern within the Balkans is mainly driven by local transmission clusters, with cross-border clusters observed mainly between countries of former Yugoslavia, with an exception of Romania .

- Regional spread of HIV within the ex-Yugoslav space has continued after country break-up, whereas the spread of subtype B through multiple introductions to Romania resonates with the changing pattern of travel and migration due to European integration of Balkan countries in the early 2000s.

- HIV epidemic remains an important public health problem in the Balkans, in-spite of the recent downwards trend in the number of new HIV diagnoses seen in some countries, reflecting differences in preventive policies implemented in participating countries.

7 References

1. Chinen J, Shearer WT. Secondary immunodeficiencies, including HIV infection. J Allergy Clin Immunol. 2010. 125: S195-S203.
2. UN Joint Programme on HIV/AIDS (UNAIDS). UNAIDS the first 10 years. 2008.
3. UN Joint Programme on HIV/AIDS (UNAIDS). The Reference Data, 2020.
4. European Centre for Disease Prevention and Control/WHO Regional Office for Europe. HIV/AIDS Surveillance in Europe 2020 – 2019 data.
5. Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N Engl J Med. 1981; 305(24):1425-31.
6. Masur H, Michelis MA, Greene JB, Onorato I, Stouwe RA et al. An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. N Engl J Med. 1981; 305(24):1431-8.
7. Hymes KB, Cheung T, Greene JB, Prose NS, Marcus A et al. Kaposi's sarcoma in homosexual men-a report of eight cases. Lancet. 1981;2(8247):598-600.
8. Gallo RC. A reflection on HIV/AIDS research after 25 years. Retrovirology. 2006; 3:72.
9. Younai FS. Thirty years of the human immunodeficiency virus epidemic and beyond. Int J Oral Sci. 2013;5(4):191-199.
10. Masur H, Michelis MA, Wormser GP, Lewin S, Gold J et al. Opportunistic infection in previously healthy women. Initial manifestations of a community-acquired cellular immunodeficiency. Ann Intern Med. 1982; 97(4):533-9.
11. Oleske J, Minnefor A, Cooper R Jr, Thomas K, dela Cruz A et al. Immune deficiency syndrome in children. JAMA. 1983; 249(17):2345-9.
12. Bloom AL. Acquired immunodeficiency syndrome and other possible immunological disorders in European haemophiliacs. Lancet. 1984; 1(8392):1452-5.
13. Peterman TA, Jaffe HW, Feorino PM, Getchell JP, Warfield DT et al. Transfusion-associated acquired immunodeficiency syndrome in the United States. JAMA. 1985; 254(20):2913-7.
14. Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science. 1983; 220(4599):868-71.
15. Vahlne A. A historical reflection on the discovery of human retroviruses. Retrovirology. 2009; 6:40.
16. Marx JL. Strong new candidate for AIDS agent. Science. 1984; 224(4648):475-7.
17. Case K. Nomenclature: human immunodeficiency virus. Ann Intern Med. 1986; 105(1):133.
18. Gökengin D, Oprea C, Uysal S, Begovac J. The growing HIV epidemic in Central Europe: a neglected issue? J Virus Erad. 2016; 2(3):156-61.
19. UN Joint Programme on HIV/AIDS (UNAIDS). The Reference Data, 2021.
20. UNAIDS. 90-90-90 An ambitious treatment target to help end the AIDS epidemic., . Joint United Nations Programme on HIV/AIDS (UNAIDS),2014.
21. Frescura L, Godfrey-Faussett P, Feizzadeh A A, El-Sadr W, Syarif O, Ghys PD et al. Achieving the 95 95 95 targets for all: A pathway to ending AIDS. PLoS One. 2022 Aug 4;17(8):e0272405.
22. European Centre for Disease Prevention and Control/WHO Regional Office for Europe. HIV/AIDS Surveillance in Europe 2021 – 2020 data.
23. UN Joint Programme on HIV/AIDS (UNAIDS). The Reference Data, 2019.
24. Institute for Public Health of Serbia. Narrative Progress Report on HIV/AIDS Response of Republic of Serbia in 2015. 2016.

25. Siljic M, Salemovic D, Jevtovic D, Pesic-Pavlovic I, Zerjav S et al. Molecular typing of the local HIV-1 epidemic in Serbia. Infect Genet Evol. 2013; 19:378-85.
26. Siljic M. Phylogenetic analysis and molecular characterization of human immunodeficiency virus in Serbia (Doctoral Dissertation). 2017. Medical Faculty, University of Belgrade.
27. Godinho J, Jaganjac N, Eckertz D, Renton A, Novotny T, Garbus L (2005). HIV/AIDS in the Western Balkans Priorities for Early Prevention in a High-Risk Environment.World Bank working paper; no. 68.
28. Stanojevic M, Alexiev I, Beshkov D, Gökengin D, Mezei M et al. HIV-1 molecular epidemiology in the Balkans: a melting pot for high genetic diversity. AIDS Rev. 2012. 14(1):28-36.
29. National Aids Committee, Department For Aids Monitoring From Romania. Statistic data on HIV/AIDS infection in Romania. CNLAS; Bucharest: 2013. Available at: http://www.cnlas.ro/date-statistice.html.
30. Pharris A, Wiessing L, Sfetcu O, Hedrich D, Botescu A et al. Human immunodeficiency virus in injecting drug users in Europe following a reported increase of cases in Greece and Romania, 2011. Euro Surveill. 2011; 16(48):20032.
31. Kostaki E, Magiorkinis G, Psichogiou M, Flampouris A, Iliopoulos P et al. Detailed Molecular Surveillance of the HIV-1 Outbreak Among People who Inject Drugs (PWID) in Athens During a Period of Four Years. Curr HIV Res. 2017. 15(6):396-404.
32. Hatzakis, A., Sypsa, V., Paraskevis, D., Nikolopoulos, G., Tsiara et al. Design and baseline findings of a large-scale rapid response to an HIV outbreak in people who inject drugs in Athens, Greece: the ARISTOTLE programme. Addiction. 2015. 110(9), 1453–1467.
33. Nikolopoulos GK, Pavlitina E, Muth SQ, Schneider J, Psichogiou M et al. A network intervention that locates and intervenes with recently HIV-infected persons: The Transmission Reduction Intervention Project (TRIP). Sci Rep. 2016; 6:38100.
34. Kostaki EG, Gova M, Adamis G, Xylomenos G, Chini M et al. A Nationwide Study about the Dispersal Patterns of the Predominant HIV-1 Subtypes A1 and B in Greece: Inference of the Molecular Transmission Clusters. Viruses. 2020 ;12(10):1183.
35. Lunar MM, Židovec Lepej S, Tomažič J, Vovko TD, Pečavar B et al. HIV-1 transmitted drug resistance in Slovenia and its impact on predicted treatment effectiveness: 2011-2016 update. PLoS One. 2018; 13(4): e0196670.
36. Oroz M, Begovac J, Planinić A, Rokić F, Lunar MM et al. Analysis of HIV-1 diversity, primary drug resistance and transmission networks in Croatia. Sci Rep. 2019; 9(1):17307.
37. Alexiev I, Shankar A, Wensing AM, Beshkov D, Elenkov I et al. Low HIV-1 transmitted drug resistance in Bulgaria against a background of high clade diversity. J Antimicrob Chemother. 2015;70(6):1874-80.
38. Billings E, Heipertz RA, Varleva T, Sanders-Buell E, O'Sullivan AM et al. HIV-1 genetic diversity and demographic characteristics in Bulgaria. PLoS One. 2019;14(5): e0217063.
39. Gokengin D, Oprea C, Begovac J, Horban A, Zeka AN et al. HIV care in Central and Eastern Europe: How close are we to the target? Int J Infect Dis. 2018; 70:121-130.
40. German Advisory Committee Blood (Arbeitskreis Blut), Subgroup 'Assessment of Pathogens Transmissible by Blood'. Human Immunodeficiency Virus (HIV). Transfus Med Hemother. 2016;43(3):203-222.
41. Zumla A. Mandell, Douglas, and Bennett's principles and practice of infectious diseases. Lancet Infect Dis. 2010 May;10(5):303–4.
42. Rice AP. The HIV-1 Tat Protein: Mechanism of Action and Target for HIV-1 Cure Strategies. *Curr Pharm Des*. 2017;23(28):4098-4102.
43. Jackson PEH, Dzhivhuho G, Rekosh D, Hammarskjold ML. Sequence and Functional Variation in the HIV-1 Rev Regulatory Axis. *Curr HIV Res*. 2020;18(2):85-98.

44. Rohr O, Marban C, Aunis D, Schaeffer E. Regulation of HIV-1 gene transcription: from lymphocytes to microglial cells. J Leukoc Biol. 2003;74(5):736-49.
45. Mann JK, Byakwaga H, Kuang XT, Le AQ, Brumme CJ, Mwimanzi P et al. Ability of HIV-1 Nef to downregulate CD4 and HLA class I differs among viral subtypes. Retrovirology. 2013; 10:100.
46. Faust TB, Binning JM, Gross JD, Frankel AD. Making Sense of Multifunctional Proteins: Human Immunodeficiency Virus Type 1 Accessory and Regulatory Proteins and Connections to Transcription. *Annu Rev Virol*. 2017;4(1):241-260.
47. Chan DC, Fass D, Berger JM, Kim PS. Core structure of gp41 from the HIV envelope glycoprotein. Cell. 1997 Apr 18;89(2):263-73.
48. Craigie R, Bushman FD. HIV DNA integration. *Cold Spring Harb Perspect Med*. 2012;2(7):a006890.
49. Maldarelli F. The role of HIV integration in viral persistence: no more whistling past the proviral graveyard. *J Clin Invest*. 2016;126(2):438-447.
50. Hu WS, Hughes SH. HIV-1 reverse transcription. Cold Spring Harb Perspect Med. 2012 Oct 1;2(10):a006882.
51. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. Nat Rev Genet. 2004 Jan;5(1):52-61.
52. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med. 2011;1(1): a006841.
53. Visseaux B, Damond F, Matheron S, Descamps D, Charpentier C. Hiv-2 molecular epidemiology. Infect Genet Evol. 2016 Dec;46:233-240.
54. Li G, Piampongsant S, Faria NR, Voet A, Pineda-Peña AC et al. An integrated map of HIV genome-wide variation from a population perspective. Retrovirology. 2015; 12:18.
55. Bbosa N, Kaleebu P, Ssemwanga D. HIV subtype diversity worldwide. Curr Opin HIV AIDS. 2019 ;14(3):153-160.
56. Beloukas A, Psarris A, Giannelou P, Kostaki E, Hatzakis A et al. Molecular epidemiology of HIV-1 infection in Europe: An overview. Infect Genet Evol. 2016; 46:180-189.
57. Hauser A, Hofmann A, Meixenberger K, Altmann B, Hanke K et al. Increasing proportions of HIV-1 non-B subtypes and of NNRTI resistance between 2013 and 2016 in Germany: Results from the national molecular surveillance of new HIV-diagnoses. PLoS One. 2018;13(11): e0206234.
58. Maggiorella MT, Sanarico N, Brindicci G, Monno L, Santoro CR et al. High HIV-1 diversity in immigrants resident in Italy (2008-2017). Sci Rep. 2020 ;10(1):3226.
59. Delgado E, Cuevas MT, Domínguez F, Vega Y, Cabello M et al. Phylogeny and Phylogeography of a Recent HIV-1 Subtype F Outbreak among Men Who Have Sex with Men in Spain Deriving from a Cluster with a Wide Geographic Circulation in Western Europe. PLoS One. 2015;10(11): e0143325.
60. Dauwe K, Mortier V, Schauvliege M, Van Den Heuvel A, Fransen K et al. Characteristics and spread to the native population of HIV-1 non-B subtypes in two European countries with high migration rate. BMC Infect Dis. 2015; 15:524.
61. Ross J, Cunningham CO, Hanna DB. HIV outcomes among migrants from low-income and middle-income countries living in high-income countries: a review of recent evidence. Curr Opin Infect Dis. 2018; 31(1):25-32.
62. Dragas S, Siljic M, Cirkovic V, Jovanovic L, Stanojevic M. Molecular characterization of HIV epidemic in Montenegro by means of phylogenetic analysis (Meeting Abstract) . HIV MEDICINE. 2021; 22 (Suppl. 3): 73.
63. Ciccozzi M, Vujošević D, Lo Presti A, Mugoša B, Vratnica Z et al. Genetic diversity of HIV type 1 in Montenegro. AIDS Res Hum Retroviruses. 2011; 27(8):921-4.
64. Jovanovic L, Siljic M, Cirkovic V, Stanojevic M. A Joinpoint Regression Analysis of Trends in

HIV Incidence in Balkan Countries in 15 Years Period. Reviews in Antiviral Therapy & Infectious Diseases. 2020; 7: 55.
65. Alexiev I, DimitrovaR, Nikolova R, Gancheva A, Grigorova L et al. HIV-1 diversity and antiretroviral resistance amongbulgarian citizens infected abroad and foreigners registeredwith HIV/AIDS in bulgaria from 2012 to 2017 (preliminary analysis). HIV MEDICINE. 2019; 20(Suppl. 9), 3–316.
66. Alexiev I, M Campbell E, Knyazev S, Pan Y, Grigorova L, Dimitrova R, et al. Molecular Epidemiology of the HIV-1 Subtype B Sub-Epidemic in Bulgaria. Viruses 2020; 12:441.
67. Alexiev I, Campbell EM, Knyazev S, Pan Y, Grigorova L et al. Molecular Epidemiological Analysis of the Origin and Transmission Dynamics of the HIV-1 CRF01_AE Sub-Epidemic in Bulgaria. Viruses. 2021 Jan 16;13(1):116.
68. Alexiev I, Mavian C, Paisie T, Ciccozzi M, Dimitrova R, Gancheva A et al. Analysis of the Origin and Dissemination of HIV-1 Subtype C in Bulgaria. Viruses. 2022; 14(2):263.
69. Paraschiv S, Otelea D, Batan I, Baicus C, Magiorkinis G et al. Molecular typing of the recently expanding subtype B HIV-1 epidemic in Romania: evidence for local spread among MSMs in Bucharest area. Infect Genet Evol. 2012; 12 (5):1052-7.
70. Gräf T, Delatorre E, Bello G. Phylogenetics applied to the human immunodeficiency virus type 1 (HIV-1): from the cross-species transmissions to the contact network inferences. Mem Inst Oswaldo Cruz. 2020; 115: e190461.
71. Worobey M, Gemmel M, Teuwen D. E, Haselkorn T, Kunstman K, Bunce, M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature.* 2008; *455*(7213): 661–664.
72. 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.
73. Ragonnet-Cronin ML, Shilaih M, Günthard HF, Hodcroft EB, Böni J, Fearnhill E et al. A Direct Comparison of Two Densely Sampled HIV Epidemics: The UK and Switzerland. Sci Rep. 2016; 6:32251.
74. Dennis AM, Herbeck JT, Brown AL, Kellam P, de Oliveira T et al. Phylogenetic studies of transmission dynamics in generalized HIV epidemics: an essential tool where the burden is greatest? J Acquir Immune Defic Syndr. 2014; 67(2):181-95.
75. Coltart CEM, Hoppe A, Parker M, Dawson L, Amon JJ et al. Ethical considerations in global HIV phylogenetic research. Lancet HIV. 2018; 5(11): e656-e666.
76. Grabowski MK, Herbeck JT, Poon AFY. Genetic Cluster Analysis for HIV Prevention. Curr HIV/AIDS Rep. 2018;15(2):182-189.
77. Nascimento FF, Baral S, Geidelberg L, Mukandavire C, Schwartz SR et al. Phylodynamic analysis of HIV-1 subtypes B, C and CRF 02_AG in Senegal. Epidemics. 2020; 30:100376.
78. Bbosa N, Ssemwanga D, Nsubuga RN, Salazar-Gonzalez JF, Salazar MG et al. Phylogeography of HIV-1 suggests that Ugandan fishing communities are a sink for, not a source of, virus from general populations. Sci Rep. 2019 ;9(1):1051.
79. Kimura M. Molecular evolutionary clock and the neutral theory. J Mol Evol. 1987;26(1-2):24-33.
80. Siljic M, Salemovic D, Cirkovic V, Pesic-Pavlovic I, Ranin J, Todorovic M et al. Forensic application of phylogenetic analyses - Exploration of suspected HIV-1 transmission case. Forensic Sci Int Genet. 2017 ;27:100-105.
81. Barido-Sottani J, Bošková V, Plessis LD, Kühnert D, Magnus C, Mitov V et al. Taming the BEAST-A Community Teaching Material Resource for BEAST 2. Syst Biol. 2018; 67(1):170-174.
82. Bloomquist EW, Lemey P, Suchard MA. Three roads diverged? Routes to phylogeographic inference. *Trends Ecol Evol*. 2010;25(11):626-632.

83. Bennedbæk M, Zhukova A, Tang ME, Bennet J, Munderi P, Ruxrungtham K et al. Phylogenetic analysis of HIV-1 shows frequent cross-country transmission and local population expansions. Virus Evol. 2021 ;7(2):veab055.
84. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74(12):5463-5467.
85. Wensing A, Calvez V, Günthard H, Johnson V, Paredes R et al. 2014 Update of the drug resistance mutations in HIV-1. Top Antivir Med. 2014; 22(3):642-50.
86. Posada D. Selection of models of DNA evolution with jModelTest. Methods Mol Biol. 2009; 537:93-112.
87. Swofford DL. PAUP* 40: phylogenetic analysis using parsimony (*and other methods), version 40b10. Sunderland, MA: Sinauer Associates Inc; 1999.
88. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003; 19(12):1572-1574.
89. Novitsky V, Steingrimsson JA, Howison M, Gillani FS, Li Y et al. Empirical comparison of analytical approaches for identifying molecular HIV-1 clusters. Sci Rep. 2020; 10(1):18547.
90. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 2016 ;44(W1): W232-5.
91. Paraskevis D, Beloukas A, Stasinos K, Pantazis N, de Mendoza C et al. HIV-1 molecular transmission clusters in nine European countries and Canada: association with demographic and clinical factors. BMC Med. 2019;17(1):4.
92. Madeira F, Park YM, Lee J, Buso N, Gur T et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019; 47(W1):W636-W641.
93. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005; 22(5):1185-92.
94. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012; 29(8):1969-73.
95. Paraskevis D, Kostaki E, Gargalianos P, Xylomenos G, Lazanas M et al. Transmission Dynamics of HIV-1 Drug Resistance among Treatment-Naïve Individuals in Greece: The Added Value of Molecular Epidemiology to Public Health. Genes. 2017; 8(11):322.
96. Veo C, Della Ventura C, Moreno A, Rovida F et al. Evolutionary Dynamics of the Lineage 2 West Nile Virus That Caused the Largest European Epidemic: Italy 2011-2018. Viruses. 2019; 11(9):814.
97. Pérez-Parra S, Chueca N, Álvarez M, Pasquau J, Omar M et al. Phylodynamic and Phylogeographic Profiles of Subtype B HIV-1 Epidemics in South Spain. PLoS One 2016; 11: e0168099.
98. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A et al. SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. Mol Biol Evol. 2016; 33:2167-9.
99. Paraskevis D, Pybus O, Magiorkinis G, Hatzakis A, Wensing AM et al. Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. Retrovirology 2009; 6:49.
100. European Centre for Disease Prevention and Control/WHO Regional Office for Europe. HIV/AIDS Surveillance in Europe 2017 – 2016 data. Available at: https://www.ecdc.europa.eu/en/publications-data/hivaids-surveillance-europe-2017-2016-data.
101. Vandermeer J. How Populations Grow: The Exponential and Logistic Equations. Nature Education Knowledge. 2010; 3(10):15.
102. Sherrod PH. NLREG Nonlinear Regression Analysis Program [reference manual] 1991. Available at: http://www.nlreg.com/DownloadDemo.htm

103. Linzer P, Tilley AM, Williamson MV. What floats a float nurse's boat? Creat Nurs. 2011; 17(3):130-8.
104. Dalgaard P.R: A language and environment for statistical computing. R Development Core Team [Computer programme]. 2010. Available at: http://www.R-project.org/.
105. Schwarz G. Estimating the Dimension of a Model. Ann Statist. 1978; 6 (2): 461-464.
106. Akaike H. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika. 1973; 60 (2):255–265.
107. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Stat Med. 2000; 19:335-351.
108. European Center for Disease Prevention and Control. HIV: Combination prevention. Monitoring implementation of the Dublin Declaration on partnership to fight HIV/AIDS in Europe and Central Asia: 2018 progress report. Stockholm: ECDC; 2020.
109. UN Joint Programme on HIV/AIDS (UNAIDS). The Reference Data, 2017. Available at: https://www.unaids.org/en/resources/documents/2017/unaids-data.
110. Institute for Public Health of Serbia. Epidemiological determinants of HIV infection and AIDS in Serbia by the end of November 2021. 2021 Available at: https://www.batut.org.rs/download/aktuelno/hivEpid.pdf
111. Roussos S, Paraskevis D, Psichogiou M, Kostaki EG, Flountzi E et al. Ongoing HIV transmission following a large outbreak among people who inject drugs in Athens, Greece (2014-2020). medRx. 2021.
112. Castro-Nallar E, Pérez-Losada M, Burton GF, Crandall KA. The evolution of HIV: inferences using phylogenetics. Mol Phylogenet Evol. 2012 Feb;62(2):777-92.
113. Brodin J, Zanini F, Thebo L, Lanz C, Bratt G, Neher RA, Albert J. Establishment and stability of the latent HIV-1 DNA reservoir. Elife. 2016 Nov 15;5:e18889.
114. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjörnsson J. Defining HIV-1 transmission clusters based on sequence data. AIDS. 2017 Jun 1;31(9):1211-1222.
115. Liu M, Han X, Zhao B, An M, He W et al. Dynamics of HIV-1 Molecular Networks Reveal Effective Control of Large Transmission Clusters in an Area Affected by an Epidemic of Multiple HIV Subtypes. Front Microbiol. 2020;11:604993.
116. Brenner B,Wainberg MA, Roger M. Phylogenetic inferences on HIV-1 transmission: implications for the design of prevention and treatment interventions. AIDS. 2013; 27(7):1045-57.
117. Grabowski MK, Redd AD. Molecular tools for studying HIV transmission in sexual networks. CurrOpin HIV AIDS. 2014; 9(2):126-33.
118. Chan PA, Hogan JW, Huang A, DeLong A, Salemi M 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.
119. van Griensven F, de Lind van Wijngaarden JW, Baral S, Grulich A. The global epidemic of HIV infection among men who have sex with men. Curr Opin HIV AIDS. 2009; 4(4):300-7.
120. Guerra FM, Bolotin S, Lim G, Heffernan J, Deeks SL, Li Y, et al. The basic reproduction number (R0) of measles: a systematic review. Lancet Infect Dis. 2017;17(12):e420-e428.
121. Reis J, Shaman J. Simulation of four respiratory viruses and inference of epidemiological parameters. Infect Dis Model. 2018; 3:23-34.
122. Nascimento FF, Baral S, Geidelberg L, Mukandavire C, Schwartz SR, Turpin G, et al. Phylodynamic analysis of HIV-1 subtypes B, C and CRF 02_AG in Senegal. Epidemics. 2020; 30:100376.
123. Vasylyeva TI, du Plessis L, Pineda-Peña AC, Kühnert D, Lemey P, Vandamme AM et al. Tracing the Impact of Public Health Interventions on HIV-1 Transmission in Portugal Using Molecular Epidemiology. J Infect Dis. 2019; 220(2):233-243.

124. Vasylyeva TI, Zarebski A, Smyrnov P, Williams LD, Korobchuk A, Liulchuk et al. Phylodynamics Helps to Evaluate the Impact of an HIV Prevention Intervention. Viruses. 2020;12(4):469.
125. Sullivan PS, Hamouda O, Delpech V, Geduld JE, Prejean J et al. Reemergence of the HIV epidemic among men who have sex with men in North America, Western Europe, and Australia, 1996–2005. Ann Epidemiol. 2009; 19: 423–1.
126. van Kesteren NM, Hospers HJ, Kok G. Sexual risk behavior among HIV-positive men who have sex with men: a literature review. Patient Educ Couns. 2007; 65(1):5-20.
127. Abecasis AB,Wensing AM, Paraskevis D, Vercauteren J, Theys K et al. HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. Retrovirology. 2013; 10:7.
128. Bezemer D, van Sighem A, Lukashov VV, van der Hoek L, Back N et al. Transmission networks of HIV-1 among men having sex with men in the Netherlands. AIDS. (2010) 24(2):271-82.
129. Danon L, Ford AP, House T, Jewell CP, Keeling MJ et al. Networks and the epidemiology of infectious disease. Interdiscip Perspect Infect Dis. 2011; 2011:284909.
130. Kumar G, Acharya A. Biases in epidemiological studies: How far are we from the truth? Indian J Med Spec. 2014;(5):29-35.
131. Golub SA, Gamarel KE. The impact of anticipated HIV stigma on delays in HIV testing behaviors: findings from a community-based sample of men who have sex with men and transgender women in New York City. AIDS Patient Care STDS. 2013;27(11):621-7.
132. Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol*. 2011;60(5):685-699.
133. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 2016 ;44(W1): W232-5.
134. Minh BQ, Nguyen MA, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30(5):1188-95.
135. Michalopoulos LM, Ncube N, Simona SJ, Kansankala B, Sinkala E, Raidoo J. A qualitative study of migrant-related stressors, psychosocial outcomes and HIV risk behaviour among truck drivers in Zambia. Afr J AIDS Res 2016; 15:219-26.
136. Botão C, Horth RZ, Frank H, Cummings B, Inguane C, Sathane I, et al. Prevalence of HIV and Associated Risk Factors Among Long Distance Truck Drivers in Inchope, Mozambique, 2012. AIDS Behav 2016; 20:811-20.
137. Jochelson K, Mothibeli M, Leger JP. Human immunodeficiency virus and migrant labor in South Africa. Int J Health Serv 1991; 21:157-73.
138. Frentz D, Wensing AM, Albert J, Paraskevis D, Abecasis AB et al. Limited cross-border infections in patients newly diagnosed with HIV in Europe. Retrovirology. 2013; 10:36.
139. Magiorkinis G, Angelis K, Mamais I, Katzourakis A, Hatzakis A et al. The global spread of HIV-1 Subtype B epidemic. Infect Genet Evol. 2016; 46:169-179.

LIST OF ABBREVIATIONS

AIDS
Acquired Immunodeficiency Syndrome
AIC
Akaike information criterion
APC
Annual percent change
BLAST
Basic Local Alignment Search Tool
BIC
Bayesian information criterion
BDSKY
Birth-Death Skyline Serial model
MCCMaximum clade credibilityBSSVS
Bayesian Stochastic Search Variable selection
CDC
Centers for Disease Control and Prevention
CD4
Cluster of Differentiation 4
CCR5
C-C chemokine receptor type 5
CXCR4
C-X-C chemokine receptor type 4
CRF
Circulating Recombinant Forms
DNA
Deoxyribonucleic Acid
DRC
Democratic Republic of Congo
dNTP
deoxyribonucleotide triphosphate
ECDC
European Centre for Disease Prevention and Control
env
Envelope
gag
group specific antigen gene
gp
glycoprotein
GRID
Gay Related Immune Deficiency
GTR
Generalised time reversible model155
HIV/AIDS RESEARCHES IN THE WORLD
GTR+G+I
GTR plus gamma (Ł) and proportion invariant model
HAART
Highly Active Antiretroviral Therapy

HHV-8
Human Herpesvirus 8
HIV
Human Immunodeficiency Virus
HIV-1
Human Immunodeficiency Virus type 1
HIV-2
Human Immunodeficiency Virus type 2
HLA
Human Leukocyte Antigen
HPD
Highest posterior density
HTLV-III
Human T-Lymphocyte Virus III
IN
Integrase
LAS
Lymphadenopathy Syndrome
LAV
Lymphadenopathy-Associated Virus
LCA
Latent Class Analysis
LTR
Long Terminal Repeats
MCMC
Markov Chain Monte Carlo
ML
Maximum Likelihood
MSM
Men who have Sex with Men
Ne
Effective Population Size
NCBI
National Center for Biotechnology Information
PBMC
Peripheral Blood Mononuclear Cells
PCR
Polymerase chain reaction
PIs
Protease Inhibitors
pol
Polymerase gene
PR
Protease
PWID
People Who Inject Drugs
RNA
Ribonucleic Acid
RT
Reverse Transcriptase

Re
Effective Reproductive Number
SIV
Simian Immunodeficiency Virus
STI
Sexually Transmitted Infection
tMRCA
time to Most Recent Common Ancestor
UNAIDS
Joint United Nations Programme on HIV/AIDS
WHO
World Health Organisation

Annex 1.

Accession numbers of Serbian sequences included in the present study, sampled from 1997 to 2019 and deposited in the NCBI database:

GQ399763.1, GQ399551.1, GQ400327.1, GQ400482.1, GQ398972.1, GQ399179.1, GQ399012.1, GQ399888.1, GQ399221.1, GQ399018.1, GQ480327.1, GQ399263.1, GQ395533.1, GQ400459.1, GQ400490.1, GQ400092.1, GQ399955.1, GQ400505.1, GQ399341.1, GQ399810.1, GQ398855.1, GQ400529.1, GQ400303.1, GQ400203.1, GQ400169.1, GQ399328.1, GQ398698.1, GQ399770.1, GQ399505.1, GQ399605.1, GQ399463.1, GQ399262.1, GQ400380.1, GQ399684.1, GQ399526.1, GQ400192.1, GQ399335.1, GQ400562.1, GQ400664.1, GQ400943.1, GQ400568.1, GQ400867.1, GQ399293.1, GQ400985.1, GQ400636.1, GQ400934.1, GQ400637.1, GQ400971.1, GQ400727.1, GQ400860.1, GQ400842.1, GQ400711.1, GQ400847.1, GQ400975.1, GQ401005.1, GQ400863.1, GQ399151.1, GQ400634.1, GQ400696.1, GQ400623.1, GQ400576.1, GQ400698.1, GQ400683.1, JX299860.1, JX300595.1, JX300466.1, JX301157.1, JX300670.1, JX300934.1, JX300963.1, JX301026.1, JX299883.1, JX299967.1, JX300342.1, JX300698.1, JX301113.1, JX299941.1, JX300732.1, KF056325, MH750236.1, MH750235.1, KF157408 - KF157549 MK253347-MK253439, OP254296-OP254343, OM751989 - OM752052

Accession numbers of reference sequences used for phylogenetic analysis were as follows:

Subtype A (AF069670); B (K03455); C (U52953); G (AF061642); CRF01_AE (U54771); CRF02_AG (AF063223); F (AJ249236, AF075703, AF07733, AF377956).

HIV-1 control/background sequences sampled across Europe, Northern America and Africa downloaded from the NCBI database (http://www.ncbi. nlm. nih.gov/nuccore) used for the phylogenetic analysis for both the second and the third dataset:

Albania: AY611666; AY611672; AY611684; AY611688

Austria: AF347214; AF347518; DQ878531; DQ878532; EJ936557

Belgium: DQ177230; DQ177232; DQ877759; FJ653084; EU248460; DQ177224; DQ177231; DQ177227

Bulgaria: EF517439; EF517457; EF517462; EF517464; EF517488; EF517472; EF517439; EF517410

Croatia: FN424300; FN424301

Cyprus: EU673375; EU673382; EU673408

Czech Republic: AY694218; AY694233; AY694364

Denmark: AJ419453; AJ582147; AM490879; DQ108366; DQ877795

France: AF487122; DQ878075; DQ877953; DQ877930

Germany: AF347190; AF347288; AF347140; AY878668; AY878677; DQ878276; DQ878304 ; FJ030769; GQ400800

Greece: DQ878544; DQ878548; DQ878559; DQ878569; DQ878595; EF563173

Greenland: AM285220; AM285242; AM285267; AM937019; AM937024

Ireland : DQ877830; DQ877832;

Italy: AY375051; AY362443; DQ348057; DQ348033; DQ345139; DQ345123; DQ345246; DQ345123; DQ345262; DQ345233; DQ345221; AF251947; AF252026; AF376547; AF493371; AF517266; AF517471; AY672455; AY352444; AY855419; AY855724; AY994341; AY995503; DQ345170; DQ345265; DQ369253; DQ672623; DQ878603; EF526205; EU019810; EU496146; FJ228037; FJ228081; FJ228038; FJ209055; FJ209061; FJ228131; FJ228123; FJ228127

Luxembourg: DQ877749; EF563190

Netherlands: AY423387; AY423383; AY877314; DQ877839; U34604; GQ399672; DQ8778

Biografija autora:

Dr Luka Jovanović rođen je 02.11.1991. godine u Beogradu. Medicinski fakultet Univerziteta u Beogradu upisao je 2010. godine na kome je diplomirao 2016. godine sa prosečnom ocenom 9,83 i stekao zvanje Doktora medicine. U oktobru 2016. godine upisao je doktorske studije na Medicinskom fakultetu Univerziteta u Beogradu na modulu Mikrobi i infekcija. Eksperimentalni deo doktorske disertacije završio je na Institutu za mikrobiologiju i imunologiju u laboratoriji za virusologiju kod prof. dr Maje Stanojević. Od avgusta 2017. stalno je zaposlen na Institutu za onkologiju i radiologiju Srbije kao klinički lekar na Klinici za radiološku onkologiju i dijagnostiku, služba radioterapije, a specijalistički ispit iz radijacione onkologije položio je aprila 2022. godine sa odličnim uspehom. Takođe od 2018. godine dr Jovanović je u istraživačkim zvanjima pri Institutu za onkologiju i radiologiju Srbije i Institutu za mikrobiologiju i imunologiju, Medicinskog fakulteta, Univerziteta u Beogradu a trenutno nosi zvanje istraživača saradnika.

Klinički staž uradio je na Klinici za kardiologiju Urgentnog centra UKCS-a 2016. godine. U periodu od 2013. do 2016. godine boravio je u Univerzitetskoj bolnici u Pragu (student na rotaciji na odeljenju nuklearne medicine), Klinici Sv. Đorđe u Lajpcigu (student na rotaciji na odeljenju neurohirurgije), Klinici Kramare u Bratislavi i Nacionalnoj referentnoj laboratoriji za streptokok Instituta za mikrobiologiju i imunologiju, Medicinskog fakulteta u Beogradu gde je još kao student otpočeo istraživački rad.

Author Biography:

Dr Luka Jovanović was born on November 2, 1991. in Belgrade. He entered the Faculty of Medicine at the University of Belgrade in 2010, where he graduated in 2016 with an average grade of 9.83 and obtained the title of Doctor of Medicine. In October 2016, he enrolled in doctoral studies at the Faculty of Medicine of the University of Belgrade on Microbes and Infection module. The experimental part of the doctoral dissertation was completed at the Institute of Microbiology and Immunology in the laboratory of virology under the mentorship of prof. Maja Stanojević. Since August 2017, he has been permanently employed at the Institute for Oncology and Radiology of Serbia as a physician at the Clinic for Radiation Oncology and Diagnostics, Department of Radiotherapy, and he passed the residency exam in radiation oncology in April 2022 with excellent results. Also, since 2018, dr Jovanović holds research positions at the Institute of Oncology and Radiology of Serbia and the Institute of Microbiology and Immunology, Faculty of Medicine, University of Belgrade, and currently holds the title of a research assistant.

He completed his clinical internship at the Cardiology Clinic of the UKCS Emergency Center in 2016. In the period from 2013 to 2016, he stayed at the University Hospital in Prague (a student on rotation at the Department of Nuclear Medicine), the Clinic of St. George in Leipzig (a student on rotation at the neurosurgery department), the Kamara Clinic in Bratislava and the National Streptococcus Reference Laboratory of the Institute of Microbiology and Immunology, Faculty of Medicine in Belgrade, where he started his research work while still a student.

образац изјаве о ауторству

Изјава о ауторству

Име и презиме аутора: Јовановић Лука

Број индекса 9058/16

Изјављујем

да је докторска дисертација под насловом

Анализа еволуционе и трансмисионе динамике инфекције вирусом хумане имунодефицијенције у Србији и на Балкану

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

Потпис аутора

У Београду, 30.09.2022.

_____

образац изјаве о истоветности штампане и електронске верзије докторског рада

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора Лука Јовановић

Број индекса 9058/16

Студијски програм микроби и инфекција

Наслов рада Анализа еволуционе и трансмисионе

динамике инфекције вирусом хумане имунодефицијенције

у Србији и на Балкану

Ментор проф. др Маја Станојевић

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла ради похрањивања у **Дигиталном репозиторијуму Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

_____

У Београду, 30.09.2022.

образац изјаве о коришћењу

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Анализа еволуционе и трансмисионе динамике инфекције

вирусом хумане имунодефицијенције у Србији и на Балкану

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство (CC BY)
2. Ауторство – некомерцијално (CC BY-NC)
3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)
4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)
5. Ауторство – без прерада (CC BY-ND)
6. Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених
лиценци. Кратак опис лиценци је саставни део ове изјаве).

Потпис аутора

У Београду, 30.09.2022.