Основан 1881 Founded
www.ncipd.org

NATIONAL CENTRE FOR INFECTIOUS AND PARASITIC DISEASES

Lyubomira Svilenova Grigorova

Molecular virological analysis of transmission clusters of introduced and disseminated HIV-1 subtypes in Bulgaria

Abstract
of a dissertation

for the award of the educational and scientific degree of ‘Doctor’ in the scientific field of Virology
Field of Education: 4. Natural Sciences, Mathematics and Informatics;
Professional field: 4.3. Biological Sciences.

Scientific supervisor:
Assoc. Prof. Ivaylo Alexiev Ivanov, PhD

Sofia, 2025

The thesis was developed at the Department of Virology, National Reference Confirmation Laboratory for HIV/AIDS (NRCL for HIV/AIDS), National Centre for Infectious and Parasitic Diseases (NCIPD), Sofia. The studies were funded by the National Centre for Infectious and Parasitic Diseases, Sofia, Bulgaria, the Ministry of Health of the Republic of Bulgaria and Project BG05M2OP001-1.002-0001 ‘Fundamental, translational and clinical research in the field of infections and infectious immunology’, funded by OP ‘Science and Education for Smart Growth 2014-2020’ and the European Regional Development Fund.

The internal defence took place before an extended panel of the Virology Department on 22 May 2025 at 13:00.

The doctoral thesis consists of a list of abbreviations used, a list of figures used, a list of tables used, an introduction, a literature review, objectives and tasks, materials and methods, results and discussion, conclusions, contributions, a list of references used and appendices. The thesis contains 203 pages, 42 figures and 19 tables. The bibliography lists 198 literature sources, of which 2 are in Bulgarian and 196 are in English.

The doctoral thesis is scheduled for public defence on 30.07.2025 at 13:00 in the auditorium of the NCIPD, 26 Yanko Sakazov Blvd., Sofia, at an open meeting of the scientific jury appointed by order №196/30.06.2025 of the Director of the NCIPD, in accordance with the Implementing rules of the Law on the Development of Academic Staff of the NCIPD and the Law on the Development of Academic Staff.

Table of Contents

List of abbreviations used

Съкращение	**Пояснение към направеното съкращение**
ART	Antiretroviral therapy
BEAST	Bayesian Evolutionary Analysis by Sampling Trees
BLAST	Basic Local Alignment Search Tool
cDNA	Copy of the DNA molecule
COMET	COntext-based Modeling for Expeditious Typing
CRF	Circulating recombinant form
DNA	Deoxyribonucleic acid
HET	Heterosexuals
HIV	Human immunodeficiency virus
IDU	Injecting drug users
jpHMM	Jumping Profile Hidden Markov Model
Kb	Kilobase
MAFFT	Multiple Alignment using Fast Fourier Transform
MSM	Men who have sex with men
NCIPD	National Centre for Infectious and Parasitic Diseases
NNRTI	Non-nucleoside reverse transcriptase inhibitors
NRCL of HIV/AIDS	National Reference Confirmation Laboratory for HIV / AIDS
NRTI	Nucleoside reverse transcriptase inhibitors
ORF	Open reading frame
PCR	Polymerase Chain Reaction
PI	Protease inhibitors
RIP	Recombinant Identification Program
RNA	Ribonucleic acid
URF	Unique Recombinant Form

List of figures used

List of tables used

1. Introduction

The first case of human immunodeficiency virus (HIV) disease was described by the Centre for Disease Control and Prevention in 1981. It described a patient with Kaposi's syndrome and Pneumocystis carinii pneumonia. The patient belonged to a group of men who have sex with men (MSM). HIV-1, the virus that causes AIDS, was discovered in 1983 by Françoise Barré-Sinoussi from Luc Montagnier's laboratory in France.

In the early years of knowledge about the virus, scientists began to study its spread and attempted to reproduce the transmission of the infection. The first spreader of the virus is believed to be a man named Gaetan Dugas, who is referred to in the literature as ‘patient zero’. Scientists described the sexual relationships of patient zero, placing him at the centre of their cluster. The cluster contained more than 40 MSM who lived in over 10 major cities in the United States. The study continued, and since patient zero was a flight attendant, many more infected people were discovered. In 1987, the New York Post magazine called him ‘the man who gave us AIDS’. This is how he is remembered in history. This first study of the HIV-1 cluster led to subsequent research and the development of testing methods, initially serological and later molecular. However, it later became clear that the initial assumptions were disproven and that HIV had emerged much earlier, as far back as the early 20th century in Central Africa. Molecular and phylogenetic analyses revealed that the virus had been circulating among humans for decades before its official identification, and that Gaëtan Dugas was not the origin of the epidemic, as had long been believed.

Today, the HIV epidemic is a global public health problem. In 2023, 39.9 million people are infected and living with HIV. 630,000 people have died from AIDS-related illnesses. 1.3 million people are newly infected with HIV. Approximately 88.4 million [71.3-112.8 million] people have been infected with HIV since the beginning of the epidemic, and 42.3 million [35.7-51.1 million] have died from AIDS-related illnesses [https://www.unaids.org/en/resources/fact-sheet]. In Bulgaria, a total of 4,039 people have been diagnosed with HIV-1 by 2023.

HIV-1 is known for its extremely high genetic diversity. There are four groups, of which group M is the most widespread and is responsible for the global epidemic. Group M contains 10 subtypes, sub-subtypes, circulating recombinant forms (CRFs), and unique recombinant forms (URFs). The most prevalent worldwide subtype is subtype C, which infects South Africa and India, where the highest number of infections is observed. In the Western world, subtype B is the predominant genetic form of the virus. Subtypes A, CRF02_AG, and CRF01_AE are also widespread. Subtype B is dominant in Bulgaria.

HIV transmission occurs through several main routes of transmission: blood, sexual, and vertical, from mother to newborn. In the past, it was also possible for the virus to be transmitted through blood transfusions and blood products, but today the probability of transmission through this route has been reduced to zero, as modern tests for the detection of sexually transmitted infections are used. Transmission from mother to child is possible, but it is rare. The most common route of transmission is sexual. Health authorities use campaigns to distribute free needles, syringes, and condoms to limit the spread of the virus among risk groups.

Despite all measures and efforts to reduce the spread of the virus, HIV-1 remains a serious global health problem. For this reason, it is necessary to thoroughly study the transmission clusters formed in different genetic forms of the virus. The multiple links formed between risk groups show scientists worldwide that a detailed phylogenetic study is necessary. With the help of molecular biological methods, modern science can reconstruct phylogenetic trees, which facilitate the study of the mechanism of virus transmission.

Cluster analysis is a key element in the development of modern studies on viral spread. The knowledge accumulated about the formation and enlargement of HIV-1 clusters enables scientists and health authorities to focus their efforts on specific areas of public health to reduce the spread of HIV-1 worldwide.

2. Aim

The aim of this dissertational thesis is to analyse the transmission phylogenetic clusters of the different HIV-1 subtypes introduced and spread in Bulgaria through sequencing and phylogenetic analysis of a fragment of the HIV-1 *pol* gene.

2.1. Tasks

1. Sequencing of a fragment of the HIV-1 *pol* gene from selected plasma samples.
2. Preparation of databases for analysis by combining genomic sequences, demographic and epidemiological data.
3. Subtyping, recombinant and phylogenetic analysis of the obtained sequences.
4. Identification of transmission phylogenetic clusters.
5. Molecular epidemiological analysis of transmission clusters and their relationship to introduced and circulating HIV-1 subtypes, as well as to the demographic and epidemiological characteristics of the studied population.

3. Materials and methods

3.1. Tested Blood Samples

This study includes plasma samples from patients diagnosed with HIV at the NRCL of HIV/AIDS between 1986 and 2020, divided into three groups. The first group includes 663 sequences from patients diagnosed with HIV at the NNRCL of HIV/AIDS, the NCIPD, Sofia, Bulgaria, between 1986 and 2018 with established subtype B. The second group included 270 sequences isolated from plasma samples of patients diagnosed in NRCL of HIV/AIDS between 1995 and 2019, infected with HIV-1 CRF01_AE. The third group includes 1 053 sequences from näive patients diagnosed between 2012 and 2020. In order to obtain demographic and virological information, the self-assessment questionnaires completed by patients during the diagnostic process were analysed. Laboratory tests include a set of serological and molecular biological tests performed during the diagnostic process and virological monitoring of HIV patients in Bulgaria. Blood samples are linked to demographic and clinical data through an anonymous digital code by the ethical standards of the Republic of Bulgaria [**Alexiev I., 2013**].

3.2. Sequencing of the Viral Genome

In 2002, genotyping was performed for the first time in Bulgaria by sequencing a section of the viral genome at NRCL of HIV/AIDS. Genotyping of HIV-1 in Bulgaria began with a single-capillary ABI 310 sequencer and a Viroseq test kit. This genotyping method was integrated into the virological monitoring process for patients diagnosed with HIV who were on antiretroviral therapy (ART).

In 2005, the NRL for HIV introduced a second type of sequencer, which uses 12-lane gel electrophoresis and the Trugene test kit. The new method offers greater capacity and faster sample processing, which is necessary to cope with the increasing number of patients and the laboratory workload. This new approach ensures faster genotyping while allowing effective tracking of viral evolution in the Bulgarian population of people living with HIV.

In 2002, only 4 studies were conducted, while in 2005 the number of genotyped samples increased to 42. The subsequent expansion of the laboratory's capacity allowed the number of genotyped samples to reach 195 in 2022. These changes are related to the growing number of diagnosed patients, as well as the increase in the number of patients starting ART and requiring monitoring of viral resistance.

3.2.1. Technology for Sequencing and Analysis of HIV-1

In NRCL of HIV/AIDS, the main genetic material analysed is the HIV-1 *pol* gene, including the PR and RT genes. These two genes are responsible for key viral enzymes that are targeted by therapeutic agents used in the treatment of HIV infection. Two primary tests are used for *pol* gene sequencing: the TRUGENE® HIV-1 Genotyping Kit and the Viroseq HIV-1 Genotyping Test. Both tests target a fragment of the *pol* gene that encodes protease and reverse transcriptase, the main target enzymes for three major classes of ART: nucleoside reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs) , and protease inhibitors (PIs).

3.2.2. Sequencing Method Using a Capillary Sequencer and Viroseq Kit

The ViroSeq HIV-1 Genotyping System is an advanced method for sequencing HIV-1, which allows the detection of mutations in the HIV-1 *pol* gene and the analysis of their relationship to ART resistance. The method uses amplification of approximately 1.8 kilobases (kb) of DNA, covering the entire PR gene and approximately two-thirds of the RT gene within the *pol* open reading frame (ORF). After amplification, seven primers are used to segment the amplified DNA to generate a 1.2 kb consensus sequence.

The goal of genotyping by sequencing the *pol* gene is not only to identify mutations that lead to resistance to therapy, but also to determine the genetic affiliation of the virus to different subtypes and recombinant forms of HIV-1. This is critical because different subtypes may have different mutations that affect the effectiveness of therapies. The ViroSeq HIV-1 Genotyping System processes samples with viral loads ranging from 2,000 to 750,000 copies/ml, making the system highly useful for routine diagnosis and monitoring.

3.2.3. Steps in the Genotyping Process

The process of genotyping HIV-1 with the ViroSeq™ HIV-1 Genotyping System includes five main steps that ensure accurate and efficient genotyping (Figure 1):

1. Extraction and purification of viral RNA from the plasma sample. This step includes separating the virus from cellular components and preparing it for subsequent amplification.
2. Reverse transcription of the HIV-1 genome into cDNA, which is necessary because HIV is an RNA virus.
3. Polymerase chain reaction (PCR) amplification of the PR and RT genes from cDNA. During this step, multiple copies of specific regions of the HIV-1 genome are created, and these will be sequenced.
4. Direct sequencing of the PCR-amplified product by adding six or seven primers to the reaction mixture. This stage allows for more detailed coverage and higher accuracy.
5. Identification and analysis of the raw data obtained using specialised ViroSeq™ HIV-1 Genotyping System software. At this stage, the obtained 1.2 kb consensus sequence is assembled and edited, and mutations associated with ART resistance, as well as newly emerging mutations that have not been detected, are identified. The software also compares the obtained sequence with the pNL4-3 reference sequence recommended by the AIDS Clinical Trials Group to determine the presence of known mutations.

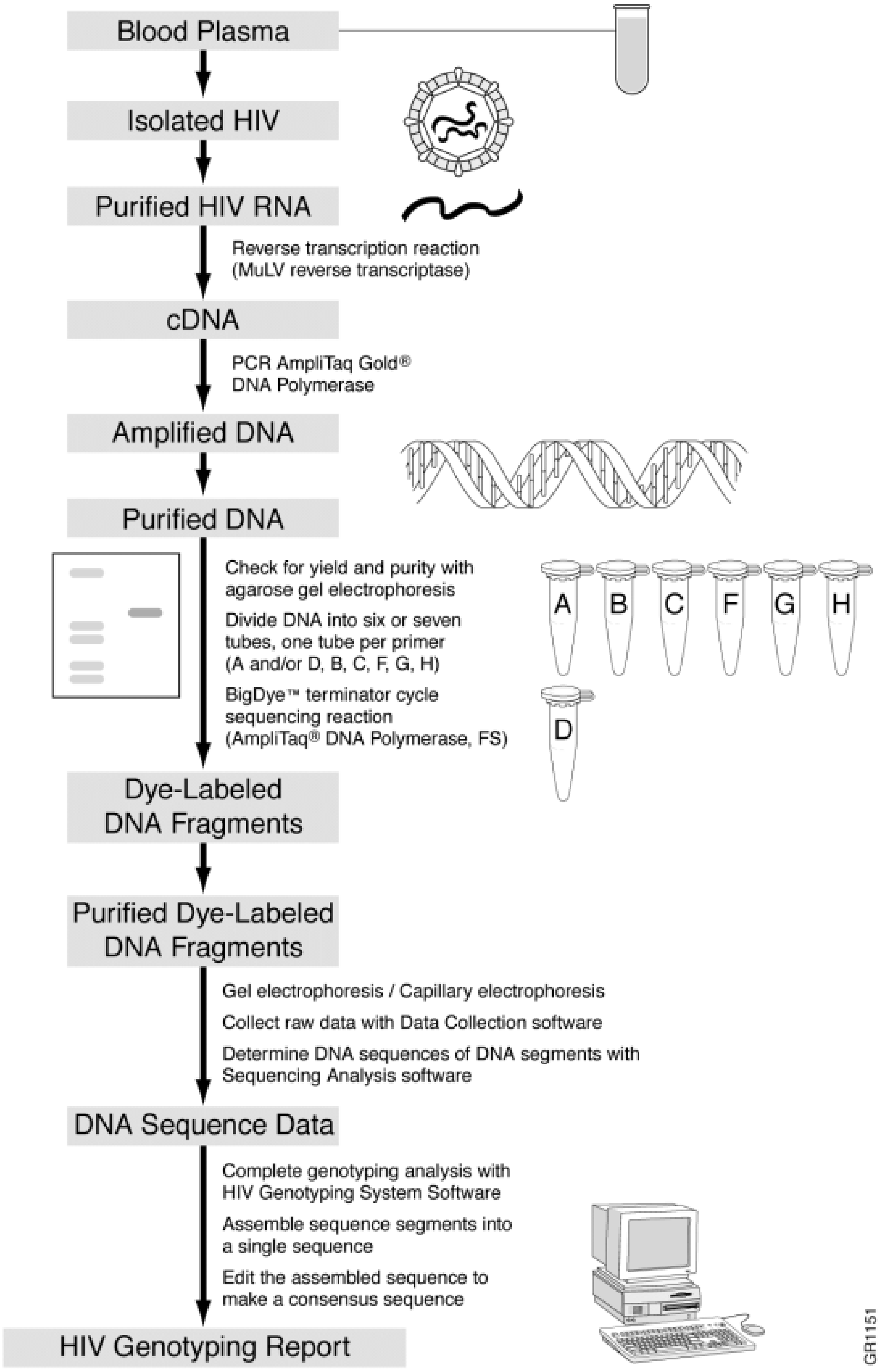


Fig. 1 Summary of the sequencing process with the ViroSeq™ HIV-1 Genotyping System. Source: ViroSeq™ HIV-1 Genotyping System, Version 2, User's Manual.

3.3. A Phylogenetic Analysis of the HIV-1 Genome

Phylogenetic analysis of viral genomes is a key tool for revealing the evolutionary history of viruses and tracking their genetic progress in the population. In the context of HIV-1, a broader understanding of its genetic diversification is essential for the diagnosis, treatment, and control of the virus. Genetic polymorphism in the HIV-1 genome is a major cause of the virus's diversity, leading to the formation of different subtypes and recombinant forms. In addition, mutations in the HIV-1 genome play a key role in the development of ART resistance, which complicates the treatment of patients.

3.3.1. Methods Used for Phylogenetic Analysis of HIV-1

In our study, we used various techniques and specialised bioinformatics analysis programmes that allow accurate determination of viral subtypes, identification of recombinant forms, and phylogenetic analysis of HIV-1. Internet-based tools and manual methods were used in the analysis to ensure high accuracy of the results.

In our study, we used several basic tools for subtyping, recombinant analysis, and phylogenetic alignment: REGA HIV-1 Subtyping Tool - Version 3.0, COMET HIV-1 Version 2.4, BLAST, RIP, jpHMM, MicrobeTrace, MAFFT, IQTree, FastTree Version 2.1.10, BEAST, ClusterPicker, AliView, and manual phylogenetic analysis.

4. Results

4.1. Demographic and Epidemiological Characteristics of Patients Infected with HIV-1 Subtype B in Bulgaria

The studied samples cover the period from 1986 to 2018 and include samples from various geographical regions of the Republic of Bulgaria. The samples cover all major population groups at risk for HIV-1 infection. After successfully sequencing the samples, we analyzed the results using internet-based subtyping tools and manual phylogenetic analysis. This allowed us to identify and categorize phylogenetic transmission clusters of HIV-1 subtype B.

Using a global phylogenetic tree, we compared the 663 Bulgarian HIV *pol* isolates with international sequences that met the criteria for pure HIV-1 subtype B.

The demographic and epidemiological characteristics of patients with HIV-1 subtype B in Bulgaria were established by analyzing data that the patients provided during the diagnostic process. This information was collected via self-identification questionnaires regarding belonging to vulnerable groups and engaging in risky practices. Demographic data included information on gender, age at diagnosis, country of origin, probable country of infection, region of permanent residence, and transmission category. The summarized results are presented in Table 1, which shows the distribution of patients according to these characteristics.

Table 1. Demographic data for patients with HIV-1 subtype B included in this dissertation

Demographic data	**Subtype B**	
	number	**(%)**
Total number	663	100
Gender		
Male	593	89,4
Female	70	10,6
Age		
≤19	23	3,5
20-29	253	38,2
30-39	256	38,6
40-49	92	13,9
≥50	39	5,9
Country of origin		
Bulgaria	651	98,2
Other	12	1,8
Probable country of infection		
Bulgaria	585	88,2
Other	78	11,8
Region in Bulgaria by permanent residence		
Sofia	368	55,5
Plovdiv	44	6,6
Varna	48	7,3
Other	203	30,6
Transmission category		
HET	256	38,6
MSM	376	56,7
IDU	22	3,3
MSM + IDU	1	0,2
Blood transfusion (in the early years of the epidemic)	4	0,6
MTC	4	0,6

Legend: HET – Heterosexual individuals; MSM – Men who have sex with men; IDU – Intravenous drug users; MSM + IDU – Persons belonging to two transmission groups (men who have sex with men and intravenous drug users); MTC – Vertical transmission from mother to child.

In the early years of the HIV epidemic in Bulgaria, the MSM group represented the largest share of registered cases of HIV infection. In the context of escalating global integration and shifting socio-economic dynamics within nations, patterns of HIV transmission are undergoing notable transformations. The prevalence of individuals infected with HIV within this particular transmission category has exhibited an upward trend over time. The MSM group has exhibited a substantial increase since 2005, with this trend persisting and growing exponentially. This phenomenon aligns with observations in other European countries, indicative of heightened awareness and reporting of sexual practices within these communities. The emergence of HIV infection in the IDU group was first documented in 2004, with a subsequent peak in cases among this vulnerable population recorded in 2009 [**Alexiev I., 2016**].

An analysis of demographic and epidemiological data for patients diagnosed with HIV at the National Reference Laboratory for HIV for the period 1986-2018 reveals that 593 (89.4%) of the patients are male, while 70 (10.6%) are female (Figure 2).

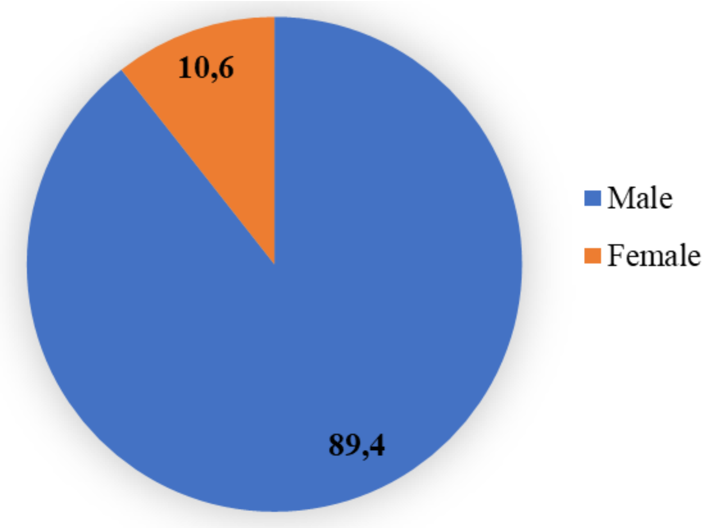


Fig. 2. Percentage distribution by gender for patients with HIV-1 subtype B included in this dissertational thesis.

The average age of patients at diagnosis is 33 years, with the age of patients ranging from 1 to 73 years. The 20-39 age group accounts for 76.8% of the studied patients. The smallest proportion of patients is in the under 9 age group (3.5%), while 5.9% of patients are over 50 years of age. The 40-49 age group is represented by 13.9% (Figure 3).

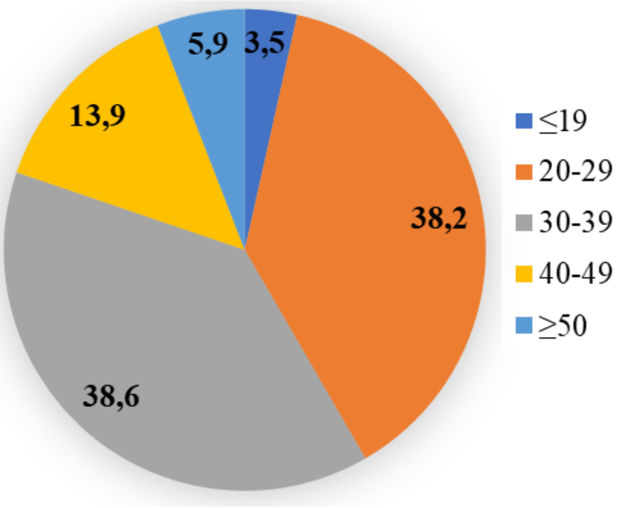


Fig. 3. Percentage distribution by age at diagnosis for patients with HIV-1 subtype B included in this dissertational thesis.

Epidemiological data show that 651 (98.2%) of the patients are Bulgarian citizens, while 12 (1.8%) are foreigners (Figure 4).

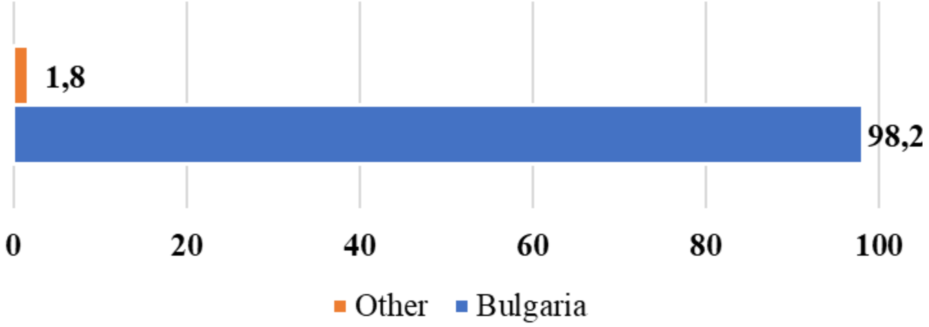


Fig. 4. Percentage distribution by country of origin for patients with HIV-1 subtype B included in this dissertational thesis.

Most of them (585 or 88.2%) state that they acquired the infection in Bulgaria, while 78 (11.8%) claim that they were infected outside the country, with 63 (9.5%) of them infected in the European Union (Figure 5). This shows that Bulgaria remains the main region for the spread of HIV among patients with subtype B, although migration and international travel may play a significant role in the transmission of the infection.

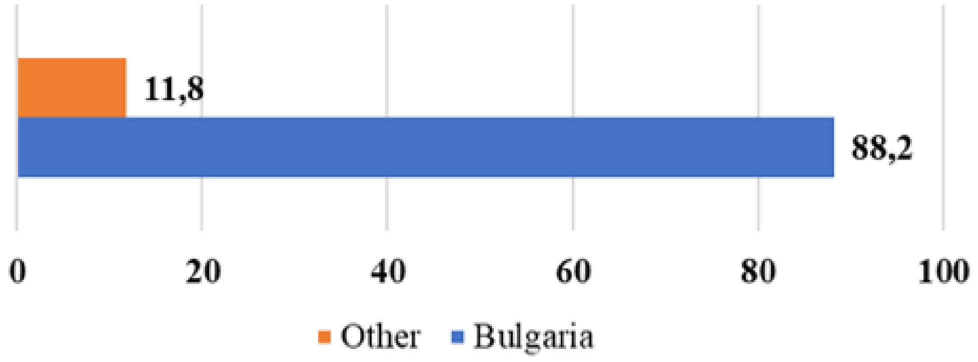


Fig. 5. Percentage distribution by country of infection for patients with HIV-1 subtype B included in this dissertational thesis.

Geographical analysis shows that more than half of the samples belong to patients with permanent residence in Sofia (55.5%). The next largest regions are Varna (7.3%) and Plovdiv (6.6%) (Figure 11), while the remaining regions of the country are less represented. This reflects not only demographic trends but also access to medical care and diagnostics in larger urban areas.

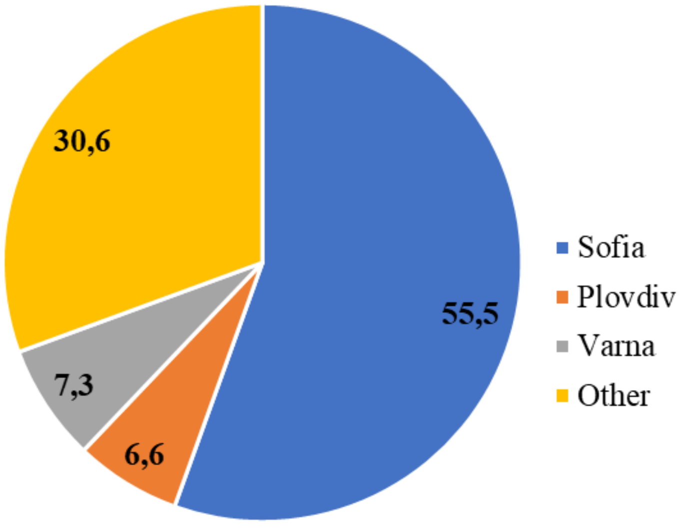


Fig. 6. Percentage distribution by region in Bulgaria by permanent address for patients with HIV-1 subtype B included in this dissertational thesis.

The analysis by transmission categories reveals the dominance of the MSM group, which accounts for 56.7% of patients, followed by heterosexuals (38.6%) and intravenous drug users (3.3%). Other vulnerable groups are underrepresented, with percentages below 1%

(Figure 7). These data highlight the need for targeted preventive measures that cover specific risk groups and adapt to changing social and epidemiological conditions.

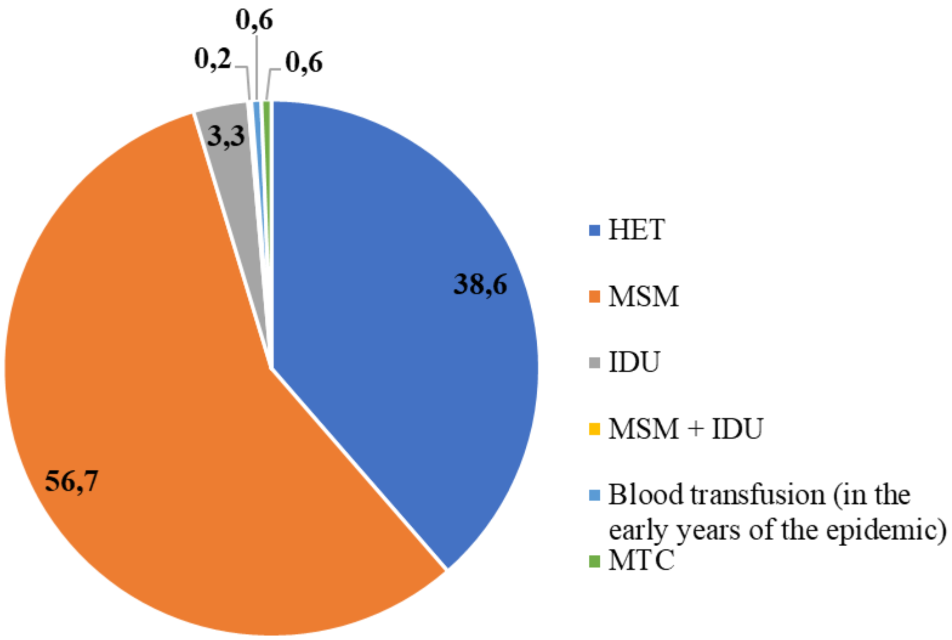


Fig. 7. Percentage distribution by transmission category for patients with HIV-1 subtype B included in this dissertational thesis.

4.1.1. Subtyping and Recombinant Analysis with Internet-Based Tools

One of the main steps in the process of genetic analysis of HIV-1 is the subtyping of isolated sequences, which provides important information about the epidemiological dynamics and molecular characteristics of the infection. The subtyping process usually involves the use of advanced Internet-based tools that allow large volumes of data to be analysed in a relatively short time. In this thesis, sequences from various independent online tools were used to ensure accurate and precise determination of the subtypes of the isolated HIV-1 sequences.

The first tool used was the REGA HIV-1 Subtyping Tool version 3.0 (http://dbpartners.stanford.edu:8080/RegaSubtyping/stanford-hiv/typingtool/), developed at Leland Stanford Junior University in Stanford, California. This tool uses powerful algorithms for subtyping through phylogenetic analysis and bootscanning, which are suitable for processing large data sets. The main advantage of the REGA tool is that it builds multiple phylogenetic trees for each individual sequence, thus optimising the accuracy of determining the corresponding HIV-1 subtype. Through this approach, the tool enables the most accurate classification of viral isolates at the subtype level.

After the isolated HIV-1 sequences underwent the subtyping process with the REGA tool, they were analysed with a second tool, COMET HIV-1 version 2.4 (https://comet.lih.lu/), developed by the Luxembourg Institute of Health. COMET provides a context-based assessment of viral sequences by comparing them to reference sequences, allowing for a deeper understanding of the affiliation of each nucleotide sequence to a specific subtype, CRF (classical recombinant forms) or URF (unidentified recombinant forms). This allows scientists to identify not only standard subtypes, but also recombinant forms that may be of epidemiological significance.

The use of these tools in a sequence of steps ensures high accuracy in subtyping and guarantees the correctness of the results. The methodology was applied to the three main datasets used in this study, which allowed for consistency and reliability of the results obtained.

4.1.2. Results From the Analysis of the Spread of HIV-1 Subtype B in Bulgaria

Our study shows that in 1988, the first patient infected with HIV-1 subtype B was diagnosed in Bulgaria [**Alexiev I., 2013**]. During the initial period of the epidemic, the viral subtype was not particularly widespread, but since 1996, more than half of the newly registered HIV cases in the country have been associated with subtype B. This shows that from this period onwards, subtype B began to dominate the epidemiological data for Bulgaria.

In 2001, 62.5% of all people diagnosed with HIV in Bulgaria were carriers of HIV-1 subtype B. However, there was a subsequent decline in the percentage of people infected with subtype B, falling to 20% in 2008. This change can be explained by changes in the behaviour of risk groups and the spread of other subtypes, as well as the development of preventive measures against HIV.

In the following period from 2009 to 2018, 558 new cases of HIV-1 infection with subtype B were registered in Bulgaria, representing 48.3% of all cases diagnosed with HIV-1 subtype B for the period 1986-2018 in the country. This highlights the continIDUg importance of subtype B in the epidemiological picture in Bulgaria, despite the competition from other variants of the virus and new forms of infection that were observed during the years studied.

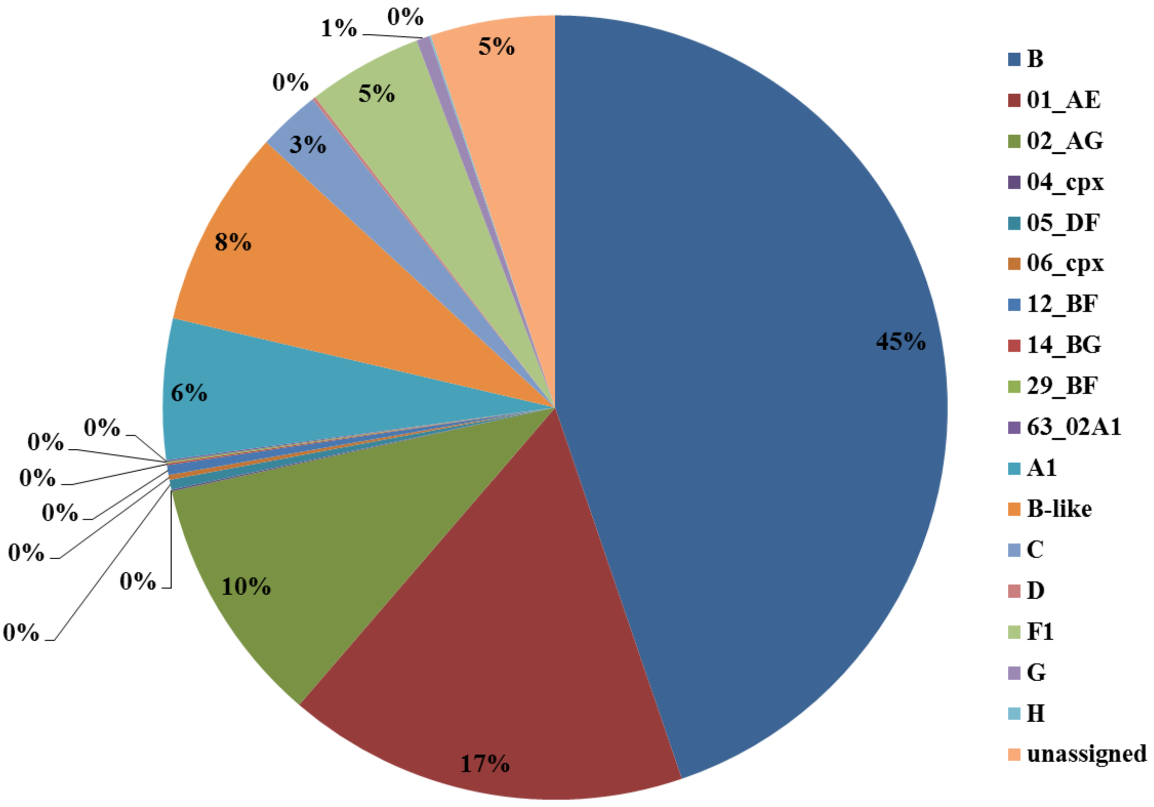


Fig. 8. Percentage distribution of HIV-1 subtypes, CRF (classical recombinant forms) and URF (unidentified recombinant forms) in Bulgaria for the period 1986-2018 based on data from the NRCL of HIV/AIDS, NCIPD. The graph shows the percentage shares of the different subtypes and recombinant forms, with each category marked in a different colour, making it easy to observe changes in the distribution of viral forms.

This graph not only visualises the results of our study, but also highlights the trend towards an increase in recombinant forms and the emergence of new variations of the virus, which is of key importance for the development of strategies for the prevention and treatment of HIV infection in Bulgaria.

4.1.3. Phylogenetic Transmission Analysis of HIV-1 Subtype B

4.1.3.1. Sequence Alignment for Phylogenetic Analysis

The preliminary step in constructing a phylogenetic tree is sequence alignment, which is essential for subsequent phylogenetic analysis. In this study, the three databases of HIV-1 subtype B sequences were aligned using the MUSCLE algorithm, using the bioinformatics programme AliView v.1.23. The MUSCLE algorithm is recognised as one of the most effective methods for multiple sequence alignment and provides high accuracy in the processing of viral isolates, which is critical for the correct interpretation of genetic information.

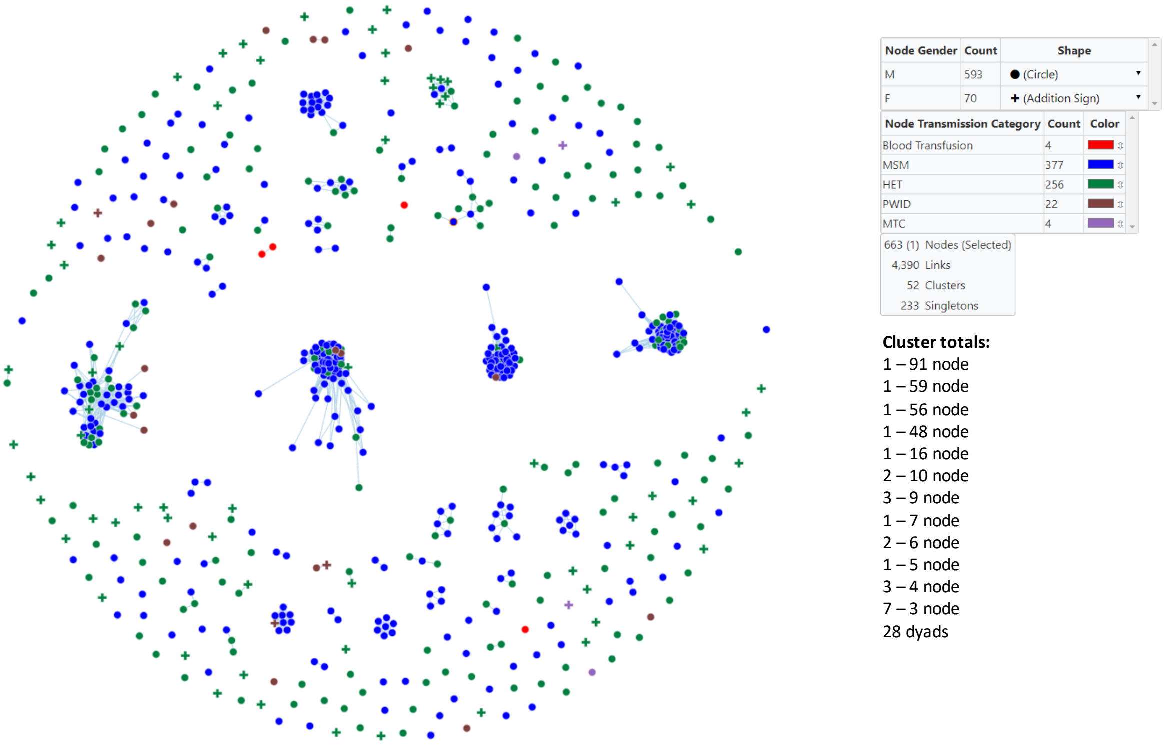
In addition to automated alignment, some of the sequences underwent manual alignment, which was performed to ensure additional precision in the nucleotide sequence alignment process. Manual alignment is important when automatic algorithms cannot resolve specific discrepancies, especially in areas of high variability in the virus genome, which is common in HIV-1.

4.1.3.2. Identification of HIV-1 subtype B transmission clusters with MicrobeTrace

After the sequences were aligned, an analysis was performed to identify HIV-1 subtype B transmission clusters using the bioinformatics programme MicrobeTrace. The Tamura-Nei (d) algorithm for genetic distance was used, as it also takes into account the well-known mutation rate of HIV-1. For the purposes of our study, a genetic distance threshold of 1.5% was selected, which is optimal for the analysis of viral isolates corresponding to the current epidemiological situation and allows for a clear distinction between transmission clusters.

The MicrobeTrace programme graphically visualises transmission clusters by connecting nodes with lines, with each node representing an individual patient sequence. The lines between the nodes indicate genetic proximity and cluster formation. Two or more sequences that are connected to each other are considered to form a cluster. In our study, we found 52 transmission clusters, of which 24 clusters consisted of 3 or more sequences, and 28 clusters were pairs of sequences formed by closely related viruses (Figure 9).

The distribution of the sequences participating in the clusters by gender showed that 68.8% of the analysed sequences belonged to men and 31.2% to women. It is interesting to note that in three large clusters containing 56, 59 and 91 sequences, respectively, the main transmission group was MSM (men who have sex with men), with percentages of 55.4%, 73% and 80% for the three clusters. The next largest group was HET (heterosexual), accounting for 39.3%, 27.1% and 17.6% of the respective clusters. The IDU (intravenous drug users) group was also represented in two of the clusters examined, with an interesting result being that in the smallest cluster (with 56 sequences), the percentage of IDUs was highest – 5.4%, while in the largest cluster the percentage of IDUs was only 2.2%.



Node Gender	Count	Shape
M	593	● (Circle)
F	70	+ (Addition Sign)

Node Transmission Category	Count	Color
Blood Transfusion	4	🟥
MSM	377	🟦
HET	256	🟩
PWID	22	🟫
MTC	4	🟪

663 (1) Nodes (Selected)
4,390 Links
52 Clusters
233 Singletons

Cluster totals:
1 – 91 node
1 – 59 node
1 – 56 node
1 – 48 node
1 – 16 node
2 – 10 node
3 – 9 node
1 – 7 node
2 – 6 node
1 – 5 node
3 – 4 node
7 – 3 node
28 dyads

Fig. 9. Visualisation of HIV-1 subtype B transmission clusters identified using the MicrobeTrace programme at a genetic distance of 1.5%. The gender of patients is indicated by a circle for men and a '+' sign for women, and the transmission category is shown by different colours: red – blood transfusion, blue – MSM, green – HET, brown – IDU, purple – MTC. The visualisation also contains the total number of links between sequences, the number of clusters and the number of unique sequences.

4.1.3.3. Extended Analysis of HIV-1 Subtype B Transmission Clusters

To facilitate examination of the formed transmission clusters, they were assigned conditional numbers. This not only facilitated classification, but also allowed their structure and epidemiological significance to be distinguished more clearly. As cluster numbers increase, their size tends to decrease, with larger clusters containing more sequences and smaller clusters containing fewer (Table 2).

Cluster 1 consists of 91 sequences, the majority of which (90.1%) are from males, with only one sequence from a female. A wide range of transmission groups was found in this cluster. 73 men reported HIV transmission through MSM, while 15 men and one woman fell into the HET group and two more men were identified as IDU. Analysis of this cluster revealed that the infection was transmitted between the three vulnerable groups, with the MSM group being the most affected. This emphasises the importance of MSM as the primary route of HIV-1 transmission in Bulgaria during this period.

Cluster 2 contains 59 sequences, all of which were isolated from men. Of these, 43 men reported belonging to the MSM group and 16 to the HET group. Phylogenetic analysis revealed a transmission bridge between the MSM and HET groups in this cluster, suggesting that the infection was transmitted along the MSM-HET axis. This may reflect the dynamics of transmission in the context of different sexual and social interactions.

Cluster 3 includes 56 sequences, 52 of which are from men and 4 from women. Analysis shows that, within this cluster, 31 men belong to the MSM group, 22 to the HET group and three to the IDU group. This cluster exhibits an intrigIDUg transmission pattern involving a triple pathway: MSM → HET → IDU. This demonstrates that virus transmission can occur through various pathways involving interaction.

Cluster 4 comprises 48 sequences, of which 47 are from men and one is from a woman. This cluster broadly represents the MSM group, with 43 sequences coming from this group, four from the HET group and one from the IDU group. Interestingly, this cluster also establishes a transmission bridge between the MSM, HET, and IDU groups. This confirms multi-group transmission of the virus, with different vulnerable categories actively participating in spreading the infection.

Cluster 5 contains 16 viral nucleotide sequences isolated entirely from men. Of these, 15 men belong to the MSM group and one man belongs to the HET group. Phylogenetic analysis shows that the virus was probably transmitted from the MSM group to the HET group, establishing direct transfer between these two main transmission categories.

Cluster 6, consisting of ten sequences, is unique among the analysed clusters in that it mainly involves women. This cluster includes six women and four men, two of whom reported infection through MSM contact and the remaining eight through HET. This suggests that, in this cluster, the infection may have spread from HET to MSM, which, although unusual, is important for a broader understanding of the dynamics of HIV-1 transmission among different gender and social groups.

Clusters 7, 9, 10, 12, 14 and 15 predominantly contain participants from the MSM and HET groups, demonstrating a predominant trend of transmission along the MSM → HET axis. This indicates that the highest infection prevalence in Bulgaria is associated with these two main transmission routes.

Clusters 8, 11, 13, 16 and 17 demonstrate an even wider spread of the virus within the MSM group. This further proves that MSM are the most vulnerable group and the main channel for HIV-1 transmission in the country.

4.1.3.4. Transmission Triads and Dyads

When analysing the triads, with a genetic distance of 1.5%, we found that there were 7 triads containing a total of 21 sequences isolated entirely from men. In these triads, it was found that MSM and HET transmission categories were represented relatively equally, indicating that the virus is actively transmitted between these two vulnerable groups along the HET-MSM axis.

Detailed analysis of the dyads or closely related pairs revealed that they included 56 patients, of whom 47 were men and 9 were women. In these dyads, MSM and HET categories were equally represented, and 5 patients from the IDU group and 2 patients who were infected through blood transfusion in the early years of the epidemic, when control over transfused blood was weaker, were also identified. This highlights the importance of older routes of infection, such as blood transfusions, which may still be significant in certain regions.

4.1.3.5. Single Virus Introductions

When analysing the single sequences, we found that patients from the HET group dominated, representing 138 isolated HIV-1 subtype B *pol* sequences. The MSM group is represented by 79 patients, while the IDU group includes 10 sequences. It is noteworthy that four single cases of infection were also found in the MTC vertical transmission group. An interesting result was that these cases did not fall into the clusters or closely related pairs formed, indicating that they remain isolated from the main transmission network of the virus.

4.1.3.6 Regional Distribution of Single Sequences

The single sequences show an identical regional distribution, with 4 of the 12 cases located in Sofia and the rest distributed across different regions of the country. This highlights the importance of the geographical distribution and localisation of HIV-1 epidemiological foci in Bulgaria.

Table 2. General characteristics of 663 HIV-1 *pol* sequences with subtype B in Bulgaria for the period 1998-2018 at a genetic distance of 1.5%. Presents the overall picture of the demographic and epidemiological spread of the virus in the country.

№ of clusters	**Number sequences**	**Male**		**Female**		**MSM**		**HET**		**IDU**		**Blood transfusion**		**MTC**		**Country of origin - Bulgaria**		**Country of origin - Other**	
		n	**(%)**	**N**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**
1	91	90	98,9	1	1,1	73	80,2	16	17,6	2	2,2	0	0	0	0	83	91,2	8	8,8
2	59	59	100	0	0	43	72,9	16	27,1	0	0	0	0	0	0	54	91,5	5	8,5
3	56	52	92,9	4	7,1	31	55,4	22	39,3	3	5,4	0	0	0	0	51	91,1	5	8,9
4	48	47	97,9	1	2,1	43	89,6	4	8,3	1	2,1	0	0	0	0	46	95,8	2	4,2
5	16	16	100	0	0	15	93,8	1	6,3	0	0	0	0	0	0	15	93,8	1	6,3
6	10	4	40	6	60	2	20	8	80	0	0	0	0	0	0	9	90	1	10,0
7	10	9	90	1	10	4	40	6	60	0	0	0	0	0	0	10	100	0	0,0
8	9	8	88,9	1	11,1	8	88,9	0	0	1	11,1	0	0	0	0	9	100	0	0,0
9	9	9	100	0	0	7	77,8	2	22,2	0	0	0	0	0	0	6	66,7	3	33,3
10	9	9	100	0	0	4	44,4	5	55,6	0	0	0	0	0	0	8	88,9	1	11,1
11	7	7	100	0	0	7	100	0	0	0	0	0	0	0	0	7	100	0	0,0
12	6	6	100	0	0	4	66,7	2	33,3	0	0	0	0	0	0	5	83,3	1	16,7
13	6	6	100	0	0	6	100	0	0	0	0	0	0	0	0	4	66,7	2	33,3
14	5	5	100	0	0	4	80	1	20	0	0	0	0	0	0	5	100	0	0,0
15	4	4	100	0	0	3	75	1	25	0	0	0	0	0	0	4	100	0	0,0
16	4	4	100	0	0	4	100	0	0	0	0	0	0	0	0	3	75	1	25,0
17	4	4	100	0	0	4	100	0	0	0	0	0	0	0	0	3	75	1	25,0
7 triads (21 seq.)		21	100	0	0	12	57,1	9	42,9	0	0	0	0	0	0	18	75	3	14,3
28 dyad (56 seq.)		47	83,9	9	16,1	24	42,9	25	44,6	5	8,9	2	3,6	0	0	51	91,1	5	8,9
singles (233 seq.)		185	79,4	48	20,6	79	33,9	138	59,2	10	4,3	2	0,9	4	1,7	194	83,3	39	16,7
Total number		593	89,4	70	10,6	377	56,9	256	38,6	22	3,3	4	0,6	4	0,6	585	88,4	78	11,8

Legend:

1. Clusters smaller than 3 and single sequences are grouped together.
2. Abbreviations are defined as follows: MSM – men who have sex with men; HET – heterosexual individuals; IDU – intravenous drug users and MTC – mother-to-child transmission.

4.2. Demographic and Epidemiological Characteristics of Patients Infected with HIV-1 CRF01_AE in Bulgaria

The second dataset included in this dissertation covers all individuals diagnosed in Bulgaria between 1995 and 2019 with the established HIV-1 CRF01_AE subtype. For this set of patients, data from 270 patients falling into different transmission categories were used, with the main affected group being IDUs (intravenous drug users). To identify the HIV-1 CRF01_AE subtype, a sequenced *pol* nucleotide sequence of the virus was used, extracted from patient plasma samples collected and analysed at the National Reference Laboratory for HIV. For successful sequencing of patient samples, they must have a viral load ≥ 1,000 copies/mL, and the generated sequences undergo a series of analyses, as described above, to accurately determine their subtype. After successful sequencing and subtyping of these 270 *pol* sequences, they were included in an analysis conducted with the MicrobeTrace programme to identify transmission clusters and bridges of virus transmission between different vulnerable groups.

4.2.1. Demographic and Epidemiological Characteristics

The demographic and epidemiological characteristics of patients with HIV-1 CRF01_AE subtype diagnosed in Bulgaria between 1995 and 2019 were analysed based on data collected through questionnaires completed by the patients themselves. The questionnaires are completed during the diagnostic process and include questions about self-assessment of risk practices and opportunities for infection. The summary results of these data are presented in Table 3, which shows the main demographic and epidemiological characteristics of the patients considered.

Table 3. Demographic data for patients with HIV-1 CRF01_AE in Bulgaria (1995-2019).

Demographic data	**CRF01_AE**	
	number	**(%)**
Total number	270	100
Gender		
Male	187	69,3
Female	83	30,7
Age		
≤19	28	10,4
20-29	108	40
30-39	90	33,3
40-49	35	13
≥50	9	3,3
Country of origin		
Bulgaria	269	99,6
Other	1	0,4
Probable country of infection		
Bulgaria	251	93
Other	19	7
Region in Bulgaria by permanent residence		
Sofia	164	60,7
Peshtera	20	7,4
Other	86	31,9
Transmission category		
HET	101	37,4
MSM	13	4,8
IDU	141	52,2
Other	15	5,6

Legend:

HET – Heterosexual individuals;

MSM – Men who have sex with men;

IDU – Intravenous drug users.

4.2.2. Demographic Structure of Patients

Data analysis shows that the first patient with HIV-1 CRF01_AE in Bulgaria was diagnosed in 1995 and was a heterosexual man. Four years later, in 1999, the first newborn infected vertically with this subtype was diagnosed. In 2002, CRF01_AE was introduced into the IDUs, which was a key year for the spread of this subtype among risk groups in Bulgaria.

In 2009, seven years after the introduction of CRF01_AE into the IDU group, an epidemic outbreak was detected in Sofia, involving 26 newly diagnosed patients with the CRF01_AE subtype, 69.2% of whom (18 patients) self-identified as IDU. Two years later, in 2011, a new outbreak of the virus was detected, involving 31 patients with CRF01_AE, of whom 54.8% were from the IDU group, 38.7% were HET and 6.5% were MSM patients. This shows the active spread of the virus among various vulnerable groups, with the highest percentage of infected people in the IDU group [**L. Grigorova, 2021**].

4.2.2.1. Age Structure of Patients

The average age of patients at diagnosis is 30 years, with the highest prevalence of infection observed in the 20-29 and 30-39 age groups, which account for a total of 73.3% of all patients with HIV-1 CRF01_AE (Figure 10). The least affected group is that of patients aged ≥50 years, which accounts for only 3.3% of all cases. The youngest patient diagnosed with HIV-1 CRF01_AE is a newborn boy, and the oldest is a 63-year-old heterosexual man from Silistra.

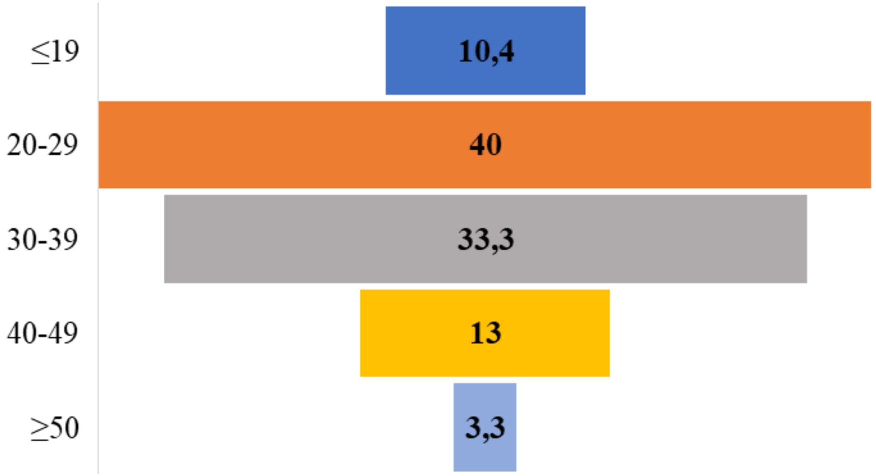


Fig. 10. Percentage distribution by age at diagnosis for patients with HIV-1 CRF01_AE included in this dissertational thesis.

4.2.2.2. Geographical Distribution

Sofia is the most affected region, with 60.7% of patients with HIV-1 CRF01_AE having a permanent address in the capital. The next largest region is Peshtera, with 7.4% of patients (Figure 11). The other regional cities are less represented, and their distribution is uneven.

These results show that the epidemic is mainly concentrated in the large urban areas of Bulgaria, where risk groups are also concentrated.

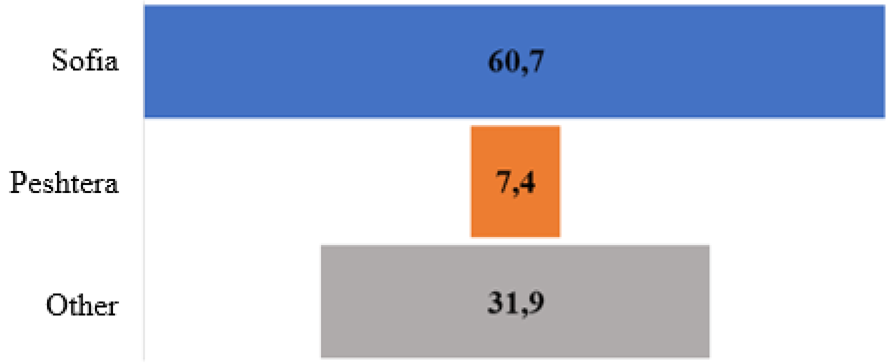


Fig. 11. Percentage distribution by region in Bulgaria by permanent address for patients with HIV-1 CRF01_AE included in this dissertational thesis.

4.2.2.3. Transmission Groups

Observations from the analysis of transmission groups show that IDU is the most affected group among patients with HIV-1 CRF01_AE, accounting for 52.2% of cases, which is significantly higher than the percentage of those infected with other subtypes. The HET group has a similar prevalence rate to non-CRF01_AE infections, accounting for 37.4% of cases, while MSM are the least affected with 4.8% of all patients (Figure 12).

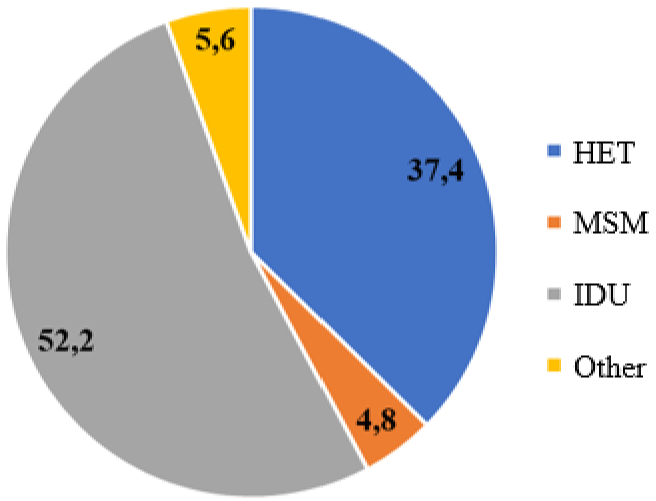


Fig. 12. Percentage distribution by transmission category for patients with HIV-1 CRF01_AE included in this dissertational thesis.

This analysis shows that the HIV-1 CRF01_AE subtype has different epidemiological characteristics in Bulgaria compared to other HIV-1 subtypes. The spread of this subtype is strongly associated with the group of intravenous drug users (IDUs), with the highest prevalence observed in large cities such as Sofia. This highlights the need for specific interventions targeting the most affected groups in the context of the epidemic.

4.2.3. Phylogenetic Transmission Analysis of HIV-1 CRF01_AE

4.2.3.1. Identification of Transmission Clusters of HIV-1 CRF01_AE with MicrobeTrace

In this study, the Internet-based programme MicrobeTrace was used to identify the transmission clusters of HIV-1 CRF01_AE. A genetic distance threshold of 1.5% was selected for the phylogenetic analysis, and the Tamura-Nei (d) algorithm was used to identify and visualise the transmission clusters (Figure 13). This approach enabled accurate mapping of phylogenetic relationships between viral isolates and revealed the potential for HIV-1 transmission in different vulnerable groups.

4.2.3.2. Results of the MicrobeTrace Analysis

Analysis of the data obtained with MicrobeTrace revealed that six clusters were identified in the study population, including three larger transmission clusters consisting of five or more sequences and three closely related virus pairs (Figure 13). The remaining 98 sequences, not related to others, were classified as single virus introductions, suggesting individual cases of infection that do not form cluster groups.

Of the total number of participants in the transmission clusters, 130 (75.6%) were men and 42 (24.4%) were women. The main transmission category in the clusters is the IDU group, which includes 119 (69.2%) of all sequences, followed by HET with 37 (21.5%) sequences. MSM and other vulnerable groups are significantly less represented, reflecting their lower burden in the spread of HIV-1 CRF01_AE in Bulgaria.

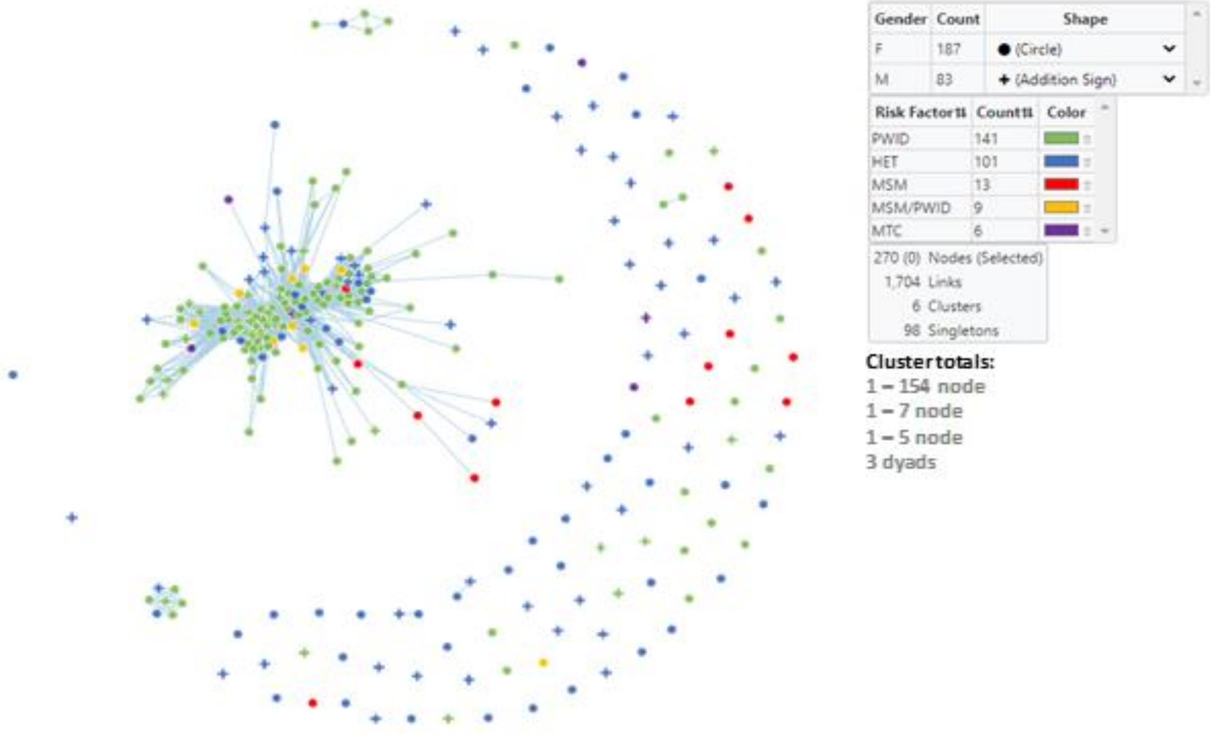


Fig. 13. Visualisation of MicrobeTrace clusters in HIV-1 CRF01_AE in Bulgaria. Analysis using a genetic distance of 1.5%. 6 clusters were identified. Gender is indicated by a plus sign '+' (men) and a circle (women). The transmission category is indicated by colour: green - people who use injection drugs (IDUs); blue - heterosexuals (HET); red - men who have sex with men (MSM); yellow - MSM+IDUs; purple - vertical transmission from mother to child (MTC). The total number of links between sequences, the number of clusters and single sequences are mentioned.

4.2.3.3. Analysis of Transmission Clusters

Cluster 1 is the largest and contains 154 sequences, of which 116 (75.3%) are male and 38 (24.7%) are female. The IDU group predominates in this cluster, accounting for 70.1% of the sequences, followed by HET (19.5%) and MSM (3.2%). An interesting finding in this cluster is that it also includes three newborns infected vertically from mother to child. 92.9% of participants in this cluster report infection within the Republic of Bulgaria, while 7.1% indicate abroad as the likely country of infection. This shows that CRF01_AE is widespread in Bulgaria, especially among vulnerable groups.

Cluster 2 is significantly smaller, consisting of 7 sequences. Five of them belong to male, and the other two are female. A transmission bridge between IDUs, and HETs was found in this cluster, indicating the possibility of virus transfer between these two vulnerable groups. Interestingly, one of the patients reported infection abroad, suggesting that the virus may have been brought to Bulgaria.

Cluster 3 contains 5 sequences, all of which are from men. Four of these men belong to the IDU group and one to the HET group. Despite the small size of the cluster, transmission of the virus along the IDU → HET axis was established, suggesting that the virus passes between these two transmission categories.

Clusters 4, 5, and 6 are dyads (closely related pairs) composed of viral isolates that also show transmission between the IDU and HET groups, further confirming the formation of HIV-1 transmission bridges between these categories.

4.2.3.4. Single Virus Introductions

When analysing single virus introductions, we found a relatively even distribution by gender, with 65.3% of these cases being HET patients and 22.4% being IDU patients. However, single sequences reveal interesting observations, including 3 newborns who do not form closely related pairs with their parents, suggesting that these cases are independent and not part of existing clusters or transmission networks.

Table 4. General characteristics of all 270 HIV-1 *pol* sequences CRF01_AE in Bulgaria for the period 1995-2019.

№ of clusters	**Number**	**Male**		**Female**		**MSM**		**HET**		**IDU**		**MTC**		**Country of origin - Bulgaria**		**Country of origin - Other**	
		n	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**
1	154	116	75	38	25	5	3,2	30	19,5	108	70,1	3	1,9	143	92,9	11	7,1
2	7	5	71	2	29	0	0	2	28,6	5	71,4	0	0	6	85,7	1	14,3
3	5	5	100	0	0	0	0	1	20	4	80	0	0	5	100	0	0
3 dyad (6 seq.)		4	67	2	33	0	0	4	66,7	2	33,3	0	0	5	83,3	1	16,7
singles (98 seq.)		57	58	41	42	1	1	64	65,3	22	22,4	3	3,1	92	93,9	6	6,1
Total number	270	187	69	83	31	9	3,3	101	37,4	141	52,2	6	2,2	251	93	19	7

Legend:

1. Clusters smaller than 3 and single sequences are grouped together.
2. Abbreviations are defined as follows: MSM – men who have sex with men; HET – heterosexual individuals; IDU – intravenous drug users; and MTC – mother-to-child transmission.

Analysis of genetic data with MicrobeTrace clearly shows the formation of transmission clusters among HIV-1 CRF01_AE patients in Bulgaria. Although the main transmission group is IDUs, the virus also circulates actively between the HET and MSM groups. This demonstrates the complexity and dynamics of the epidemic in Bulgaria and the need for targeted preventive measures for different risk groups.

4.3. Demographic and epidemiological characteristics of näive patients diagnosed with HIV-1 in Bulgaria (2012–2020)

The third set of data included in this dissertation covers the period from 2012 to 2020, with the study including only patients diagnosed with HIV-1 who meet the criterion of ART näivety. Näivety is defined as the absence of previous treatment with antiretroviral drugs, which is critical for studying the epidemiology of new cases of infection. This study covers 1 053 näive patients diagnosed in Bulgaria during this period and includes generated HIV-1 *pol* sequences collected from the five infectious disease departments in the country that have HIV treatment units. The plasma samples used are from patients living in different geographical regions of the country and cover all transmission categories, including MSM, HET, and IDU.

To be included in the study, patients must have a viral load ≥ 1,000 copies/mL, which is the minimum threshold for successful sequencing in the NRCLof HIV/AIDS. The sequences obtained underwent a series of web-based tools and manual analysis to successfully determine their subtype. Once the subtype of each of the sequences studied had been determined, phylogenetic trees were constructed to define the transmission clusters formed and possible bridges of transmission between different vulnerable groups. These phylogenetic structures allowed us to trace the evolution of HIV-1 infection in Bulgaria and reveal how different social and behavioural factors influence the spread of the virus.

4.3.1. Demographic and Epidemiological Data

The demographic and epidemiological characteristics of 1 053 näive patients diagnosed with HIV-1 in Bulgaria between 2012 and 2020 were analysed based on data provided by the patients themselves. Each patient completes a questionnaire during the diagnostic process, which aims to identify risk practices and help the patient self-identify as belonging to the relevant vulnerable group. The data from the questionnaires were used to categorise patients by gender, age, transmission group, and probable country of infection. Table 5 shows the summary data for these patients and provides information on their gender, age at diagnosis, country of origin, probable country of infection, region of permanent residence in Bulgaria, and transmission group.

Table 5. Demographic data for näive patients diagnosed with HIV-1 in Bulgaria for the period 2012-2020.

Demographic data	**Näive patients diagnosed with HIV-1 in Bulgaria for the period 2012-2020**	
	number	**(%)**
Total number	1053	100
Gender		
Male	905	85,9
Female	148	14,1
Age		
≤19	25	2,4
20-29	345	32,8
30-39	418	39,7
40-49	182	17,3
≥50	83	7,9
Country of origin		
Bulgaria	851	80,8
Other	202	19,2
Probable country of infection		
Bulgaria	870	82,6
Other	183	17,4
Region in Bulgaria by permanent residence		
Sofia	514	48,8
Plovdiv	113	10,7
Varna	46	4,4
Pazardzhik	40	3,8
Other	340	32,3
Transmission category		
HET	401	38,1
MSM	534	50,7
IDU	101	9,6
MSM + IDU	17	1,6

Legend:

HET - Heterosexual individuals

MSM - Men who have sex with men

IDU - Intravenous drug users

MSM+IDU - Individuals belonging to two transmission groups, namely men who have sex with men and intravenous drug users

This study included 1 053 patients who were diagnosed with the infection between 2012 and 2020. These patients met the criteria of antiretroviral therapy näivety, meaning that none of them had previously received treatment with antiretroviral drugs. HIV-1 *pol* sequences were collected for all included patients and underwent a series of genetic and phylogenetic analyses to accurately determine the virus subtype and analyse possible transmission clusters. Data were collected from the five main infectious disease departments in the country that diagnose and treat HIV, with patients coming from different geographical regions of Bulgaria.

The first demographic criterion used to divide patients was gender. The results show a predominance of males, with 905 male patients (85.9%) and 148 female patients (14.1%) (Figure 14). These results correspond to the global trend, in which men, especially in risk groups such as MSM, are more frequently affected by HIV infection.

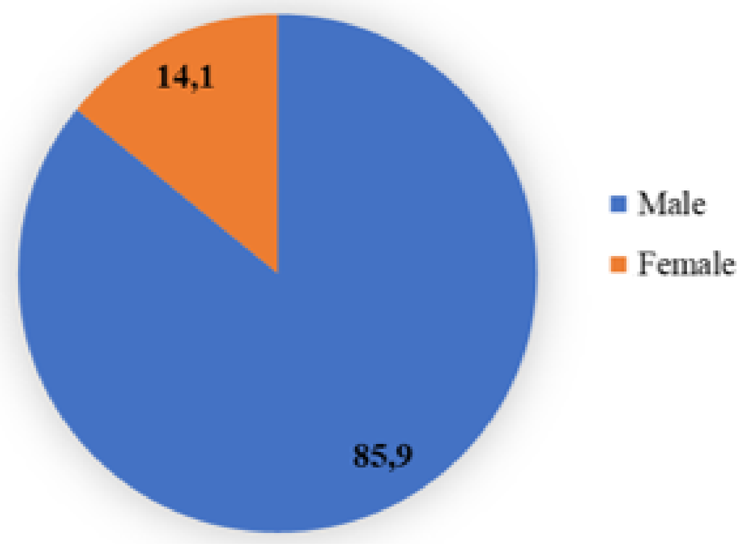


Fig. 14. Percentage distribution by gender of näive patients diagnosed with HIV-1 for the period 2012-2020 in Bulgaria.

The age distribution of participants shows that the age of patients ranges from 13 to 78 years, with the average age of all näive patients being 34.4 years. A relatively small proportion of patients are aged ≤19 years, with only 25 (2.4%) of all patients falling into this age group (Figure 15). There were no newborn patients diagnosed with vertical transmission, which demonstrates the effectiveness of vertical transmission prevention in the country.

The most commonly affected age groups are 20-29 years (32.8%) and 30-39 years (39.7%). This is expected, as these age groups fall within the period of sexual maturity and active social behaviour, which leads to a higher risk of sexual transmission of the virus. A total of 72.5% of all näive patients diagnosed with HIV-1 in Bulgaria between 2012 and 2020 are between the ages of 20 and 39, indicating that this is the main age group susceptible to infection. The 40-49 and ≥ 50 age groups are represented in a lower percentage – 17.3% and 7.9% respectively, which is logical given the lower sexual activity and risky practices in these age categories.

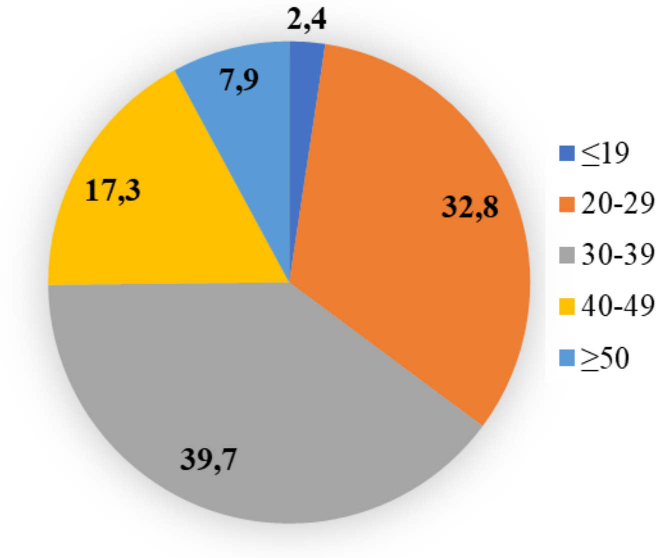


Fig. 15. Percentage distribution by age of näive patients diagnosed with HIV-1 for the period 2012-2020 in Bulgaria.

4.3.2. Epidemiological Data on Origin and Probable Country of Infection

Epidemiological data show that 80.8% of all registered patients in Bulgaria for the period 2012-2020 are citizens of the Republic of Bulgaria, while 19.2% are foreign citizens who were diagnosed in Bulgaria. This highlights the international nature of the HIV epidemic and the importance of migration and population mobility. Among foreign citizens, 5 patients from the Russian Federation, 4 from Ukraine, 3 from Cameroon, and 3 from Italy were diagnosed, with other countries being less represented (Figure 16).

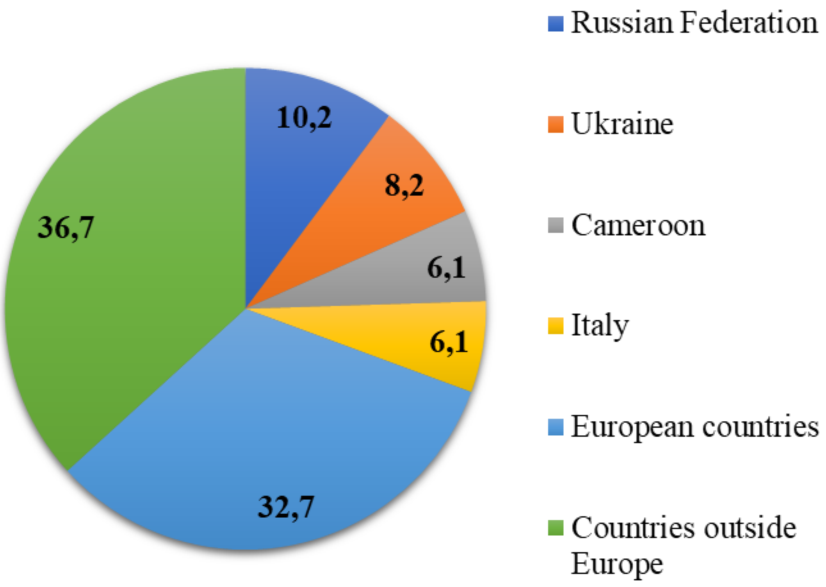


Fig. 16. Percentage distribution of näive patients diagnosed with HIV-1 for the period 2012-2020 in Bulgaria who are foreign citizens.

82.6% of patients were infected in Bulgaria, while 17.4% were infected abroad. It is extremely important to note that 132 (72.1%) of the 183 näive patients who reported that the infection was acquired outside Bulgaria (Figure 17) indicate that they were infected in Europe. The infection was most commonly acquired in Germany (14.2%) and Spain (13.7%), followed by the United Kingdom (9.8%), Greece (8.2%), France, and Italy (5.5% of cases each). This confirms the trend of HIV-1 transmission within Europe, with these countries being the main sources of infection for Bulgaria.

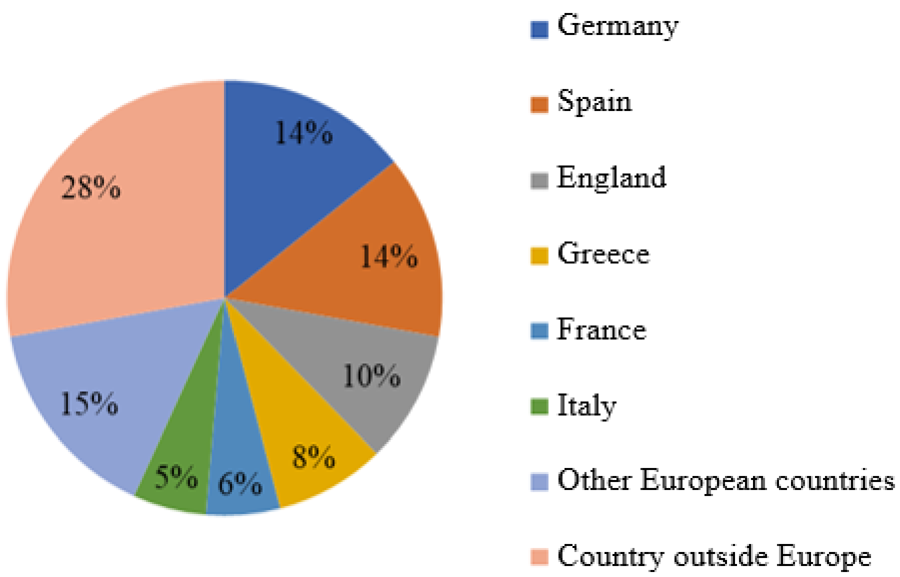


Fig. 17. Percentage distribution of näive patients diagnosed with HIV-1 for the period 2012-2020 in Bulgaria who report infection in a country other than Bulgaria.

In addition, patients from countries outside Europe are also well represented, but to a much lesser extent. For example, the United States (3.3%), the Russian Federation (2.7%), and Turkey (2.2%) are the main countries outside Europe from which infected patients come. Patients from countries such as Cameroon, Côte d'Ivoire, China, Australia, Israel, and others are rare, indicating that the spread of the virus is most concentrated in Europe, with migration flows from these regions playing an important role in the spread of HIV.

4.3.3. Geographic Distribution of Näive Patients

Geographic analysis shows that Sofia is the most affected region in Bulgaria, with 514 patients (48.8%), which is almost half of all diagnosed cases during the period under review (Figure 18). This confirms that the capital is the main focus of HIV-1 spread in the country, probably due to the concentration of migrants, tourists, and better access to health services. The next largest region is Plovdiv with 113 patients (10.7%), with a difference of almost 40% compared to Sofia. The remaining regions of the country have a lower percentage of diagnosed patients, accounting for 40.5% of all näive infections for the period 2012-2020. The spread of the infection in smaller regions may be influenced by lower mobility and fewer migration flows.

The lowest prevalence of HIV-1 infection is observed in areas such as Dobrich, Gabrovo, Razgrad, Shumen, and Kardzhali, where diagnosed patients account for a small proportion of the total number (only 4.1%). This shows that the spread of HIV-1 in Bulgaria is not uniform but concentrated in certain geographical regions, which may be related to socio-economic factors as well as migration trends and the level of public health in different areas of the country.

The study shows that HIV-1 infection in Bulgaria is highly concentrated in the male population and the active age group between 20 and 39 years, with these patients accounting for 72.5% of all diagnosed cases. Migration and international mobile flows play an important role in the spread of the infection, with European countries being the main sources of infection and countries outside Europe being less significant. The capital city of Sofia is the main region with a high concentration of newly diagnosed cases, while smaller regions of the country show lower incidence rates. These results highlight the need for targeted health interventions and preventive measures aimed at risk groups, especially in urbanised and migration-active areas.

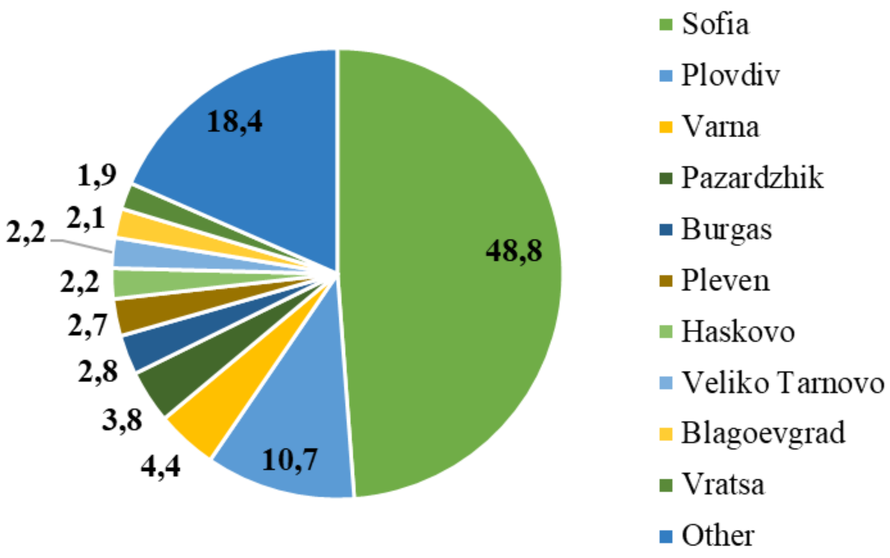


Fig. 18. Percentage distribution of näive patients diagnosed with HIV-1 for the period 2012-2020 in Bulgaria by region in the country according to permanent address.

Within the scope of the study on the geographical distribution of patients diagnosed with HIV-1 in Bulgaria for the period 2012-2020, we found significant differences in the spread of the infection across regions. The study showed that almost half of all cases of HIV-1 infection during this period were diagnosed in the region of the capital Sofia, with 514 patients (48.8%) coming from this region. This represents almost half of all newly diagnosed cases of HIV-1 in the country, highlighting the concentration of infected people in Bulgaria's largest urbanised and economically active area. The difference with the next most affected region – Plovdiv, which accounts for 113 patients (10.7%) – is significant and amounts to almost 40% of the total number of cases diagnosed in the country during the period under review.

These data highlight that the infection is more prevalent in large cities, which may be related to factors such as high population mobility, better access to health services, and greater social interaction, which increase the chances of transmission of the virus. At the same time, smaller regions of Bulgaria show lower levels of infection, with significantly fewer cases of HIV-1 infection in these areas. Overall, regions outside Sofia and Plovdiv account for 40.5% of all cases diagnosed in the country during the period 2012-2020. This means that the spread of HIV-1 in these areas is much less pronounced, which may be due to factors such as lower social mobility, lower concentration of risk groups, and less access to specialised medical care.

Some of the regions in Bulgaria that are less affected by HIV-1 infection include Varna with 46 patients (4.4%), Pazardzhik with 40 patients (3.8%), Burgas with 30 patients (2.8%), Pleven with 28 patients (2.7%), as well as smaller percentages for Haskovo and Veliko Tarnovo, where there are 23 patients (2.2%) in each of these regions.

Among the less affected regions, we observe Blagoevgrad with 22 patients (2.1%), Vratsa with 1.9%, as well as areas such as Ruse, Stara Zagora, Pernik, Vidin, Kyustendil, Montana, Lovech, Sliven, Yambol and Smolyan, which have fewer than 20 diagnosed näive patients for the entire period. These regions account for a total of 14.3% of all diagnosed HIV-1 patients in the country, which shows that they are in the second category in terms of the spread of the infection.

The fewest HIV-1 infected patients were diagnosed in Dobrich, Gabrovo, Razgrad, Shumen, Kardzhali, Silistra, Targovishte, and Peshtera. These areas account for only 4.1% of all diagnosed cases in the country for the period 2012-2020. This lower prevalence can be explained by the lower mobility of the population, the smaller concentration of risk groups, and possibly less access to health services in these less urbanised and less developed regions.

An analysis of the geographical distribution of HIV-1 infection in Bulgaria clearly shows that the infection is highly concentrated in large urbanised regions such as Sofia, with important factors such as population mobility and access to health services playing a significant role in the spread of the virus. However, HIV-1 infection is also present in smaller regions of the country, with the number of diagnosed cases remaining significant in these areas, albeit at lower levels. The distribution of cases highlights the need for targeted preventive measures and efforts to raise awareness among the population, especially in the most affected and least affected regions of the country.

4.3.4. Subtyping and Recombinant Analysis with Internet-Based Tools

The 1 053 isolated HIV-1 sequences were subjected to a subtyping process using various Internet-based tools. This is one of the main steps in the preliminary analysis of genetic data, as these tools have a significant capacity to process large volumes of data. The use of such platforms allows for the rapid and efficient retrieval of reliable information about the subtype of the virus and the presence of genetic mutations in each sequence, which is key to the subsequent analysis and interpretation of the results.

At the same time, subtyping and recombinant analysis of the sequences were performed using the methods and sequences described above. These analyses allowed the diagnosed näive patients to be divided into two main groups – those infected with HIV-1 subtype B and those infected with other subtypes and recombinant forms of the HIV-1 virus. This division was essential due to the differences in the dominant transmission categories between subtype B and the other genetic forms of the virus, which is key to understanding epidemiological trends and risk groups for the spread of infection.

When analysing the subtypes of all 1 053 sequences generated from diagnosed näive HIV-1 patients for the period 2012-2020, we found that the most widespread subtype in Bulgaria is subtype B, with 60.4% of those infected having this subtype (Figure 19). The next most prevalent subtypes are CRF01_AE with 11.4%, F1 with 6.9% and CRF02_AG with 5.2%. The other subtypes and recombinant forms of HIV-1 have lower values, not exceeding 5%. These results clearly show the dominance of subtype B in the country and its predominant spread among new infections.

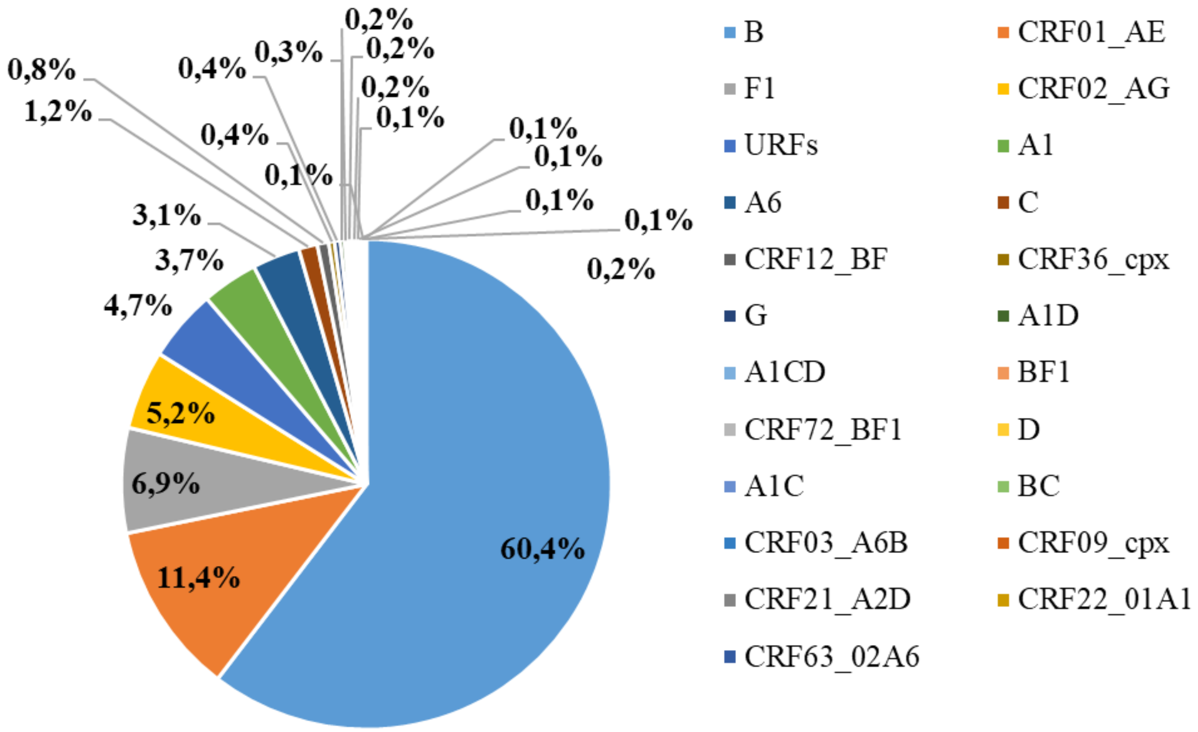


Fig. 19. Percentage distribution of HIV-1 subtypes, CRFs and URFs in Bulgaria for the period 2012-2020 among näive patients diagnosed at the NRCLof HIV/AIDS, NCIPD.

4.3.5. Analysis of Transmission Categories in Different Subtypes

When analysing transmission categories in different subtypes (Figure 20), we found that subtype B is dominated by the MSM transmission category. The distribution of transmission categories in subtype B is as follows: MSM – 63.4%, HET – 34.1%, WID – 2% and MSM+IDU – 0.5%. This emphasises that MSM represent the main risk group for the spread of subtype B in Bulgaria, which is in line with international trends, where this group is mainly affected by HIV infection.

As regards CRF01_AE, the transmissible category of IDUs dominates with 53.8%, followed by HET with 32.8%, MSM with 8.4% and MSM+IDU with 5%. This distribution shows that IDUs are the main risk category for this recombinant virus, while also demonstrating the differences in the epidemiological picture for the different genetic forms of HIV-1.

In subtype F1, we observe a relatively even distribution between MSM (47.3%) and HET (51.4%), with the percentage of MSM+IDUbeing relatively small – 1.3%. This can be explained by differences in social and behavioural patterns in the different transmission groups in the country.

The last subtype considered, CRF02_AG, is dominated by IDUs (52.8%), followed by WID (28.3%), MSM (7.6%) and MSM+IDUs (11.3%). This distribution shows that HET play a greater role in the spread of the virus, which may be related to specific local social and behavioural factors.

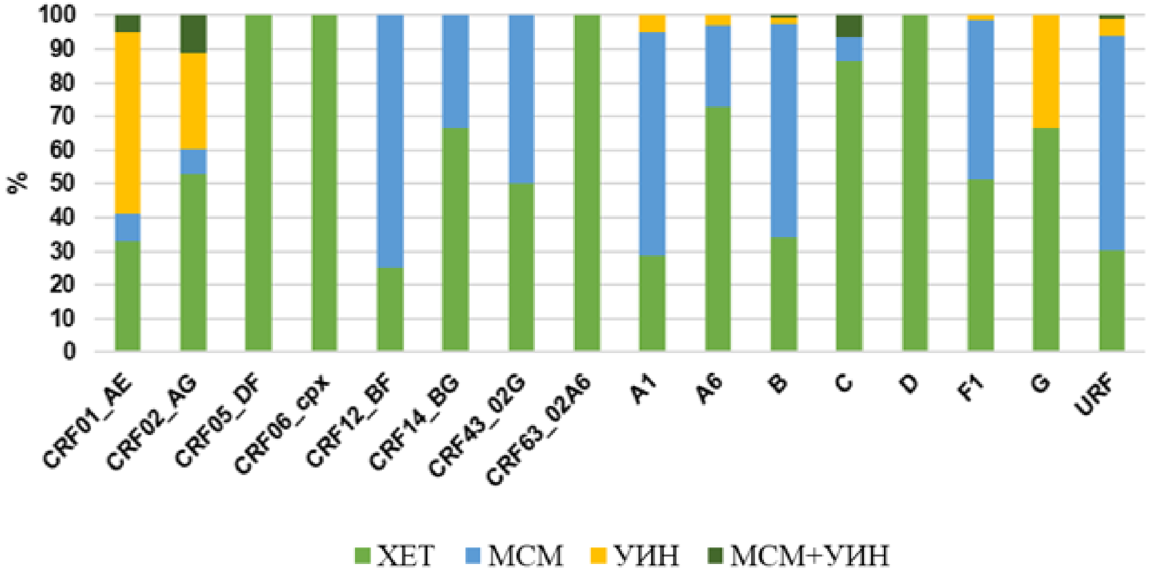


Fig. 20. Percentage distribution by transmission category in HIV-1 subtypes, CRFs and URFs in Bulgaria for the period 2012-2020 in näive patients diagnosed at the NRCL of HIV/AIDS, NCIPD.

4.3.5.1. Predominance of the Transmission Category of MSM in Recombinant Forms

We observed a prevalence of infected MSM in all recombinant forms of subtype B, demonstrating the rapid spread of subtype B in this vulnerable group. This trend is an important indicator of the need for targeted health interventions targeting MSM to limit the spread of HIV-1.

4.3.6. Analysis of Transmission Categories by Year

We performed a year-by-year analysis of the distribution of transmission categories (Figure 21) and found that the percentage of naïve MSM diagnosed was relatively constant throughout the study period, averaging 49.7% of all diagnosed patients. However, we observed a decrease in the percentage of diagnosed MSM in 2013, which coincides with the early years of the study period when stigma and social discrimination were major barriers to diagnosis and treatment of HIV infection. We observed three peaks of infected MSM in 2015 with 56.9%, and in 2017 and 2018 with 57.5% and 57.7%, respectively. This shows a steady increase in new cases in these years, suggesting that social awareness and efforts to remove stigma about HIV infection in Bulgaria have played a role in increasing diagnosis and treatment of this group. The next largest transmission category was HET, with an average rate of 39.2% throughout the period. The percentage distribution of HETs is similar to that of MSM, and in years with peaks in MSM, we see declines in HETs and vice versa. The average percentage of diagnosed HETs for the period 2012-2020 was 9.6%, and for MSM+HETs was 1.6%.

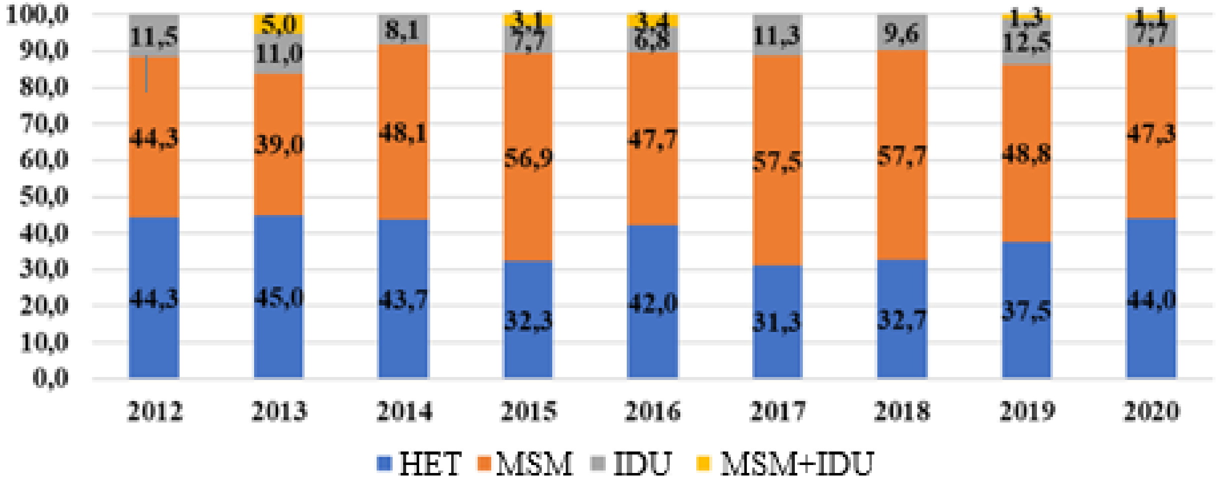


Fig. 21. Distribution of transmission categories by year in Bulgaria for the period 2012-2020 in HIV-1 näive patients diagnosed at the NNRCL of HIV/AIDS, NCIPD.

The analysis of HIV-1 subtypes and recombinant forms in Bulgaria shows a clear dominance of subtype B, as this subtype is mainly prevalent among MSM transmission category. Thus, targeted health interventions to reduce HIV infection should focus on this vulnerable group, with attention to other at-risk groups, such as IDUs and HET, who also have a significant prevalence of infection.

4.3.7. Reconstructing a Phylogenetic Tree

The first phylogenetic tree we reconstructed included all 1053 sequences of naïve patients diagnosed with HIV-1 in Bulgaria for the period 2012-2020 (Figure 22). This phylogenetic tree was an essential tool in our study, as it provided valuable information on the genetic relationships and evolutionary pathways of the virus in the population of interest. In constructing the tree, we found that the sequences with subtype B split into two branches, with the branches of subtype F1, CRF12_BF and URFs falling in between. CRF01_AE, subtype A6, subtype A1, CRF02_AG, CRF14_BG and subtype C were separated into a separate and rather distant branch. This separation clearly shows the differences in the genetic characteristics of the virus in Bulgaria, which is important for assessing the epidemiological situation.

In a more detailed analysis using the Cluster Picker program, we found the presence of multiple clusters. 34 clusters were detected in the branch containing subtype B more distant from the root of the tree, while 61 clusters were observed in the branch closer to the root. For subtype F1, 13 clusters were detected, for CRF12_BF only 1 cluster, and for URFs we detected 3 clusters. In the lower branch of the tree, where the remaining HIV-1 genetic forms are located, we found 13 clusters for CRF01_AE, 2 for subtype A6, 6 for subtype A1, 4 for CRF02_AG, 1 for CRF14_BG and 1 for subtype C.

This division of the largest subtype B into two distinct branches necessitated splitting the sequences into two categories: sequences with subtype B and sequences with non-B subtype. This step was necessary to better assess the transmissible clusters formed and the possible transmission bridges between different vulnerable groups in the population. For this reason, two separate phylogenetic trees were constructed, each focusing on the relevant transmissive categories.

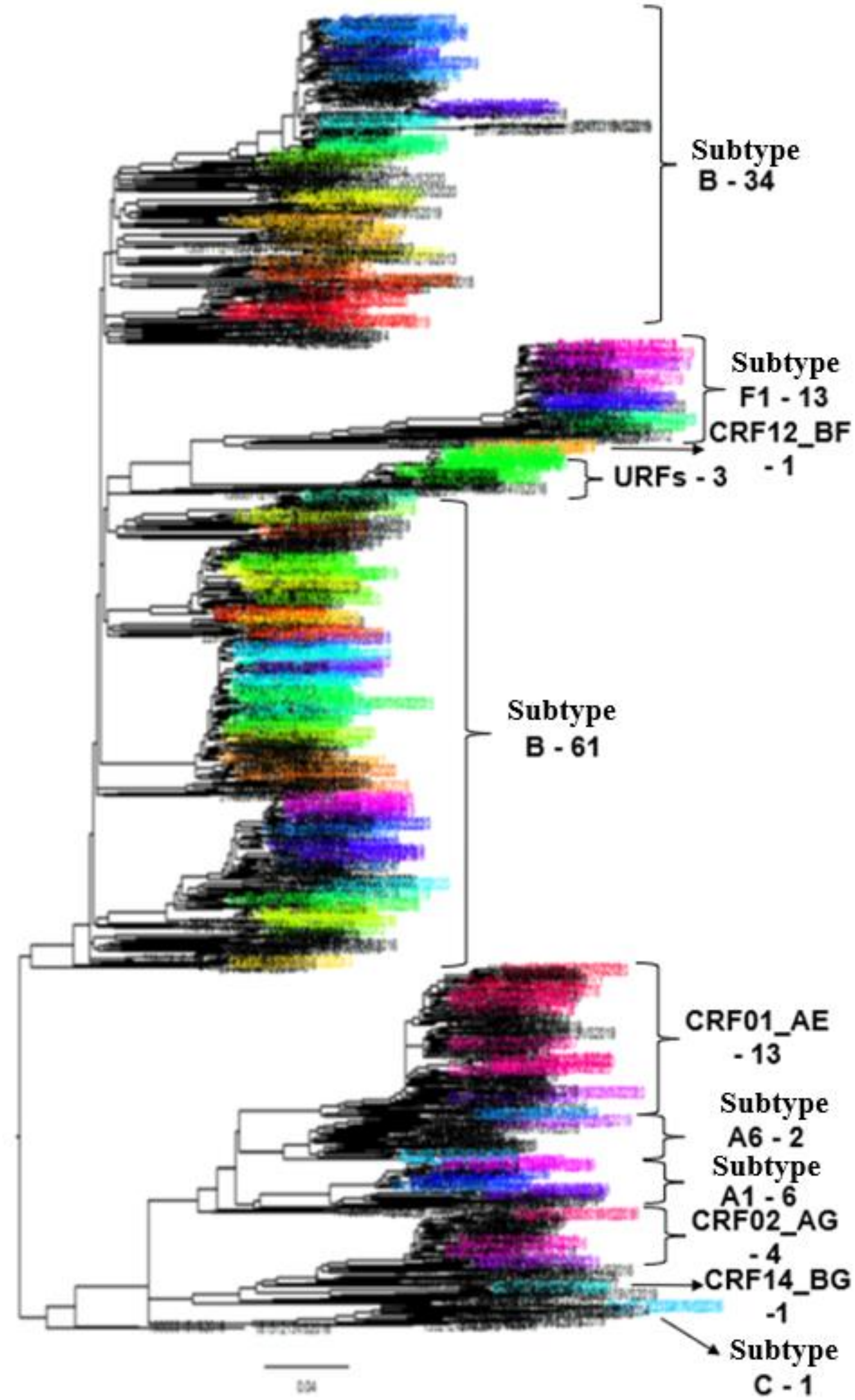


Fig. 22. Phylogenetic tree containing 1053 sequences from patients diagnosed with HIV-1 at the NRCL of HIV/AIDS for the period 2012-2020. The phylogenetic tree with brackets and arrows shows the subtypes of HIV-1. The number of clusters formed is indicated after the name of each subtype with a dash. Each cluster is coloured with a colour other than black.

4.3.8. Reconstruction of a Phylogenetic Tree of HIV-1 Subtype B

In the reconstruction of a phylogenetic tree for HIV-1 subtype B, 595 sequences from näive patients diagnosed with HIV-1 subtype B in Bulgaria between 2012 and 2020 were used (Figure 23). Cluster analysis was performed using the Cluster Picker program, which revealed the presence of 88 clusters. The criterion for cluster classification was set to 2 or more sequences with bootstrap ≥ 90. The majority of the observed clusters contained between 2 and 4 sequences of the same transmissive category, with MSM being the predominant transmissive group. Based on this it was decided in the present study to consider only clusters containing 5 or more nucleotide sequences, as these clusters provide better evidence for the possible formation of transmissible transmission bridges between different vulnerable groups in the population.

The first cluster we looked at was cluster 25. It contained 6 sequences, all of which belonged to males of the vulnerable MSM group. The mean age at diagnosis of the patients was 26.83 years. One of the participating men in the cluster was diagnosed at the time of seroconversion, which may be an important indicator of early detection of HIV infection. All patients were born in Bulgaria, but three reported that they probably acquired the infection abroad, in France, Israel and the United Kingdom, respectively. Three of the patients also reported the presence of other sexually transmitted infections, two suffering from syphilis and one from gonorrhoea. Due to the presence of only one transmissible category, this cluster does not demonstrate the formation of transmission bridges of HIV-1 infection in different vulnerable groups.

The next cluster that meets the criteria for 5 or more sequences is cluster 26, which contains exactly 5 sequences. Four of these sequences belong to MSM and one to a HET female. The average age of diagnosis was 27 years. All patients were Bulgarian citizens infected within the borders of the Republic of Bulgaria. One patient reported infection with another sexually transmitted infection, namely hepatitis B, which is a common coinfection with HIV. Despite the small size of the cluster, it demonstrates the possibility of transmission between different transmissible categories, with the MSM likely to have transmitted the infection to a HET woman, suggesting possible bridges of transmission between vulnerable groups.

Cluster 29 contains 9 nucleotide sequences, 8 of which belong to MSM and 1 to HET. The mean age at diagnosis was 28.4 years. Patients were born in Bulgaria and reported infection in the country. Four of the patients had a permanent address in Sofia, and the rest had addresses in different regions of the country, such as Pazardzhik, Burgas, Yambol, Blagoevgrad and

Varna. Two of the patients reported the presence of another sexually transmitted infection, namely gonorrhoea and syphilis. The examined cluster does not prove the formation of bridges between different transmissible categories, but shows possible routes of spread of infection in different regions of the country, which underlines the importance of regional differences in the spread of infection.

Cluster 33 contains 7 sequences, all of which belong to males born and infected in Bulgaria. The mean age at diagnosis was 36.7 years. Three of the patients in the cluster had a permanent address in Sofia, and the remaining four had addresses in Varna, Plovdiv, Pleven and Stara Zagora. In this cluster, a transmission bridge was observed as 3 of the sequences belonged to MSM and the remaining 4 to HET males. This demonstrates the transmission of infection between different transmissible categories, indicating a possible route of transmission from MSM to HET males.

Cluster 55 contains 5 sequences. All patients in this cluster are Bulgarian citizens, four of them live in the Sofia region and one in Pernik. All patients fell into the MSM transmission category and their mean age at diagnosis was 29.2 years. The cluster includes two patients suspected to have become infected abroad, one in the United Arab Emirates and the other in the United Kingdom. In addition to HIV infection, both patients infected abroad reported chlamydia coinfection, which is common in patients with HIV, and this may be due to a higher risk of sexually transmitted infections in these regions. Although this cluster does not demonstrate the formation of transmission bridges, it does provide interesting data on the dual introduction of genetically similar HIV-1 subtype B viruses into Bulgaria from two different countries located on different continents. This is indicative of the dynamics of virus transmission and possible international routes of spread.

Cluster 58 contains 7 nucleotide sequences, all belonging to MSM born and infected in Bulgaria and living in Sofia. The mean age at diagnosis was 30.7 years. One patient reported being infected with syphilis and gonorrhea in addition to HIV, which is also a common HIV coinfection and highlights the importance of early detection and treatment of STIs in this population. Due to the homogeneity of the cluster, with all participants being from the same transmissible category of MSM, we are unable to identify the presence of bridges of transmission between different transmissible groups, but this cluster clearly indicates the route of infection of each individual contained within it. This also highlights the importance of targeted treatment and prophylaxis in specific populations.

Cluster 60 contains 5 sequences of Bulgarian citizens, all of whom are infected in Bulgaria and live in Sofia. The mean age at diagnosis was 39.2 years. The cluster included five males, three of whom were HET. One of them reported coinfection with hepatitis B, which is also an important aspect in the treatment and management of HIV infection. The remaining two men are part of the transmissible MSM category. Like cluster 58, this cluster does not demonstrate the formation of transmission bridges between vulnerable groups, but it does provide useful information about the internal structure of the transmission group as well as the coinfections that may exist in HIV patients.

Cluster 78 contains 5 sequences, all belonging to MSM born and infected in Bulgaria. The mean age at diagnosis was 32.6 years. Four patients in the cluster have their permanent address in Sofia and one in Lovech. One patient was diagnosed early in the course of infection, during seroconversion, which may be indicative of early HIV detection. Although this cluster does not demonstrate the presence of transmission bridges, it is interesting because it represents an isolated group with uniform genetic characteristics, indicating the routes of infection in this specific group. This highlights the importance of diagnosing HIV in the early stages and controlling infection.

Cluster 88 contains 7 nucleotide sequences, all of which belong to males born and infected in Bulgaria. The mean age of patients at diagnosis was 42.7 years. Six of them have an address in Sofia and belong to the transmissible MSM category, and one patient is from Pleven and belongs to the HET category. This cluster is interesting because it includes patients who reported infection during blood donation, which is a possible route for HIV spread in case of insufficient infection control measures during the early transmission period. Two patients reported the presence of sexually transmitted infections syphilis and gonorrhoea, indicating the importance of early diagnosis and treatment of sexually transmitted infections. Cluster 88 is of particular importance as it shows a bridge of transmission of HIV-1 subtype B infection from MSM to HET. This is evidence of the possibility of transmission between different transmission categories, which has serious epidemiological implications, especially in the context of prevention and control of infection.

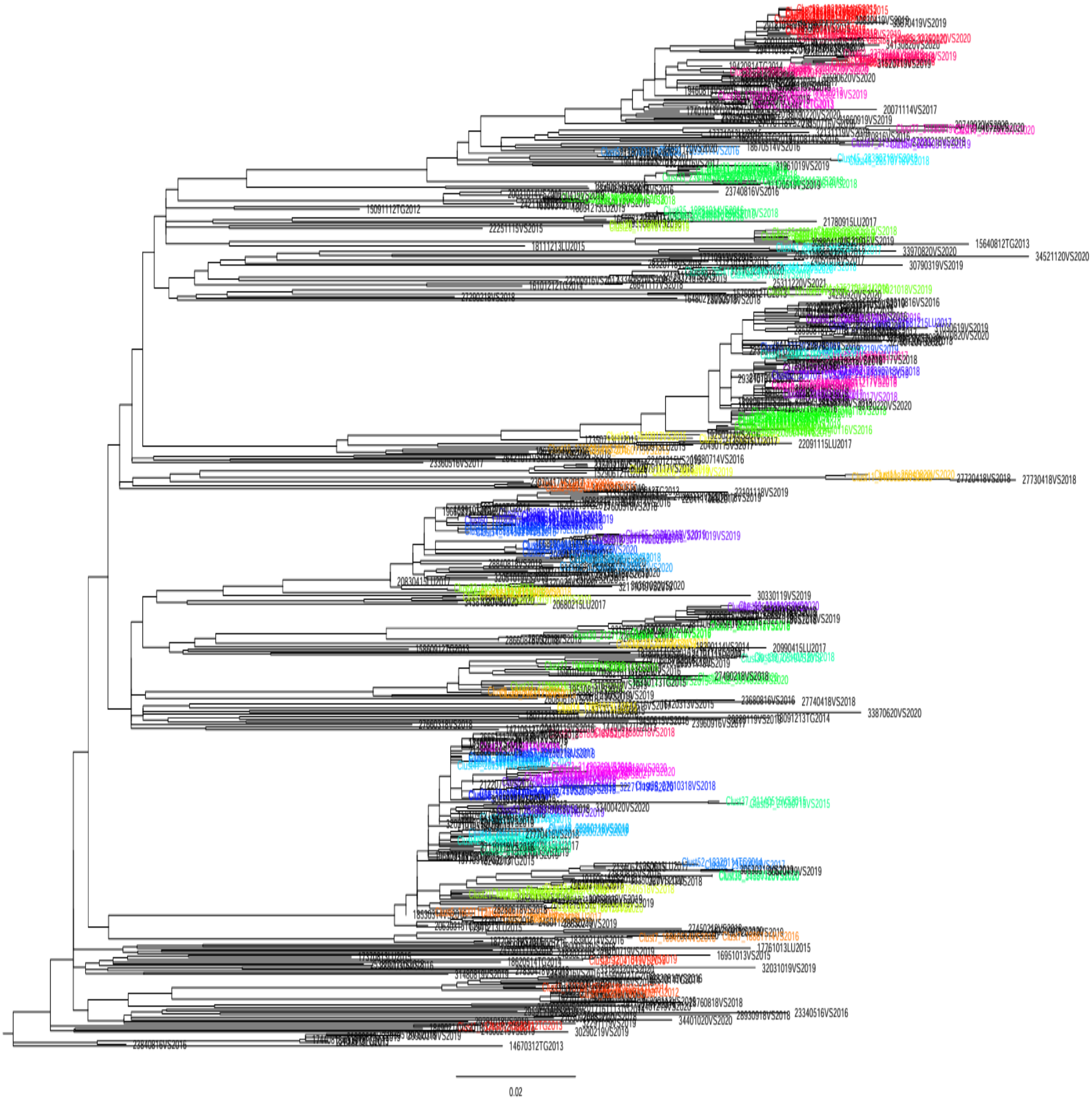


Fig. 23. Reconstructed phylogenetic tree consisting of 595 sequences isolated from näive patients diagnosed in Bulgaria for the period 2012-2020 with HIV-1 subtype B.

4.3.9. Reconstructed Phylogenetic Tree of HIV-1 non-B Subtypes

The reconstruction of the phylogenetic tree including all non-B subtypes, CRFs and URFs that are outside subtype B used 458 sequences from näive patients diagnosed in Bulgaria for the period 2012-2020 with HIV-1 non-B subtypes (Figure 24). After performing cluster analysis using the Cluster Picker program, we found the formation of 50 clusters. The cluster criterion was set for 2 or more sequences with bootstrap ≥ 90.

Data analysis showed that 14 clusters were formed in CRF01_AE, which included a total of 34 patients, of whom 4 were HET, 6 were MSM, and 25 were IDU. Cluster 22 is important, which is an inbred pair of two males with different transmission categories, 1 HET and 1 IDU. This cluster demonstrates transmission along the HET-IDU axis, showing potential routes of HIV transmission among these vulnerable groups. Cluster 27 includes 3 sequences, each falling into a different transmission category. In this cluster, we observe the formation of an HET-MSM-IDU relationship, demonstrating the possibility of transmission between different transmission groups.

CRF02_AG formed 4 clusters with a total of 9 sequences, including 2 HETs, 6 IDUs, and 1 MSM+IDU. This indicates diversity in transmission among patients with this recombinant virus and suggests the possibility of multiple routes of infection in different vulnerable groups.

For CRF12_BF, a single cluster composed of 4 sequences was identified, all of which belong to the MSM transmission category. The possible routes of infection in this cluster suggest that MSM play a central role in the spread of this recombinant virus.

CRF14_BG forms 1 cluster composed of a closely related pair that includes two males with different transmission categories, 1 HET and 1 WIN. This cluster demonstrates the presence of a transmission bridge, which is evidence of the interaction between different vulnerable groups in Bulgaria and the potential transmission routes between them.

Subtype A1 formed 6 clusters including 19 patients. Of these patients, 4 are HET, 14 are MSM and 1 is IDU. These clusters show the dominance of transmission via heterosexual and homosexual routes, with a high concentration of patients from the vulnerable MSM group. In subtype A6, 2 clusters were found. One cluster contained 2 HET patients, while the other was composed of 3 men belonging to the three main vulnerable groups, 1 HET, 1 MSM and 1 IDU. Despite the small size of this cluster, it highlights the possibility of transmission of the infection between different transmission categories.

Subtype C forms a single cluster composed of an inbred couple - a HET male and female. This cluster provides clear information on heterosexual HIV transmission within the couple and highlights the importance of prevention among heterosexual individuals.

Subtype F1 formed the largest number of clusters, with 15 clusters comprising a total of 38 patients, including 15 HET and 23 MSM. This indicates that subtype F1 has a high prevalence among MSM and HET, and these two transmission categories probably play a major role in the prevalence of this subtype.

In addition to clusters composed of subtypes and CRFs, we also found clusters in URFs. These were 6 clusters that included 33 patients, of which 4 were HET, 28 MSM, and 1 IDU. These URF clusters show a strong dominance of MSM as the major route of HIV transmission in these emerging recombinant forms.

After applying the criterion of considering clusters with 5 or more nucleotide sequences, 4 clusters that fell into different subtypes were considered and analyzed.

Cluster 8 was composed of 5 sequences of subtype A1. The mean age at diagnosis was 31.8 years. Four of the patients were Bulgarian citizens and one was a citizen of the Republic of North Macedonia, but all reported Bulgaria as the likely country of acquisition of the infection. Four of them have their permanent address in Sofia and one in Kardzhali. The patients fall into the vulnerable MSM group. Two of the patients reported the presence of other sexually transmitted infections, such as hepatitis B and Trichomonas vaginalis. Although this cluster does not demonstrate the formation of a transmission bridge across different transmission groups, it does indicate possible transmission of the virus to neighbouring countries, such as the Republic of North Macedonia, highlighting the importance of international control and cooperation to address HIV.

Cluster 16 was formed by 7 sequences with URF. The median age at diagnosis was 37 years. All sequences belonged to males born in Bulgaria. Five reported acquiring the infection in the country and two abroad, one in Spain and one in the United States of America. Four of the men had a permanent address in Sofia, and the others in Shumen, Pazardzhik and Lovech. The transmission category of six of the participants is MSM and one is HET. Three of the patients reported the presence of other sexually transmitted infections: 2 syphilis and 1 condyloma. Two of the patients were diagnosed during blood donation. The cluster analyzed provides information on possible transmission between two vulnerable groups and an MSM HET bridge formed. This indicates the potential mobility of infection among different transmission categories and the possibility of interaction between them.

Cluster 31 contains 5 nucleotide sequences of F1 subtype. These sequences belong to Bulgarian citizens infected in the country. Three of them live in Sofia and two in Plovdiv. The mean age at diagnosis was 31.6 years. Four of the participants had the transmission category MSM and 1 had HET. The cluster carries information on bridge formation between MSM and HET vulnerable groups, indicating that the infection can be transmitted between these two main transmission categories despite the small size of the cluster. This may imply that even with a small number of sequences, the virus can spread between different social groups.

Cluster 33 is the largest cluster analyzed, composed of 14 URF F1B sequences. It includes only males, with a mean age at diagnosis of 28.9 years. Thirteen of the patients were native-born, of whom 11 reported being infected within the country. The cluster also includes one Moldovan national who was infected in his home country but lives in Sofia. Two presumed to be infected in Spain and Italy. Five of the men live in Sofia and the rest in different areas of the country. Five reported another sexually transmitted infection, the most common being syphilis. Two of the participants were HET and 12 were MSM. Due to its size, cluster 33 shows a high prevalence of URF F1B across the country. Furthermore, we clearly observe transmission along the MSM-HET axis, demonstrating the possibility of transmission bridges between these two different groups, in the context of the national epidemiological situation.

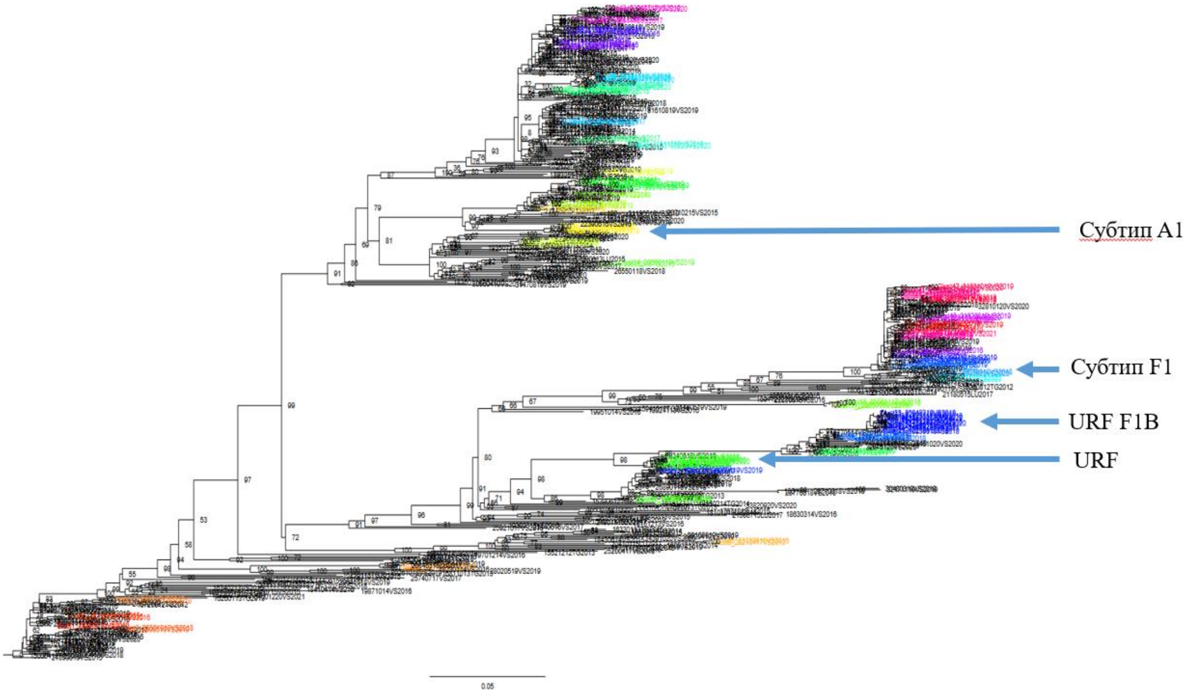


Fig. 24. Reconstructed phylogenetic tree containing 458 sequences isolated from näive patients diagnosed in Bulgaria between 2012 and 2020 with HIV-1 non-B subtypes. Blue arrows indicate the locations of large clusters containing 5 or more sequences, and the genetic forms of the sequences in the cluster are also mentioned.

These clusters and the phylogenetic analyses conducted demonstrate the importance of monitoring and thoroughly understanding HIV-1 transmission pathways, as well as HIV-1 prevalence among different vulnerable groups. The analysis also demonstrates that despite existing differences in transmission categories, infection continues to spread in smaller but significant clusters, highlighting the need for ongoing surveillance and prevention.

4. Discussion

This study was conducted on three large datasets. These include the most prevalent HIV-1 subtypes in this country, subtype B and the circulating recombinant form CRF01_AE, as well as other subtypes and recombinant forms that were analyzed within a large dataset of newly diagnosed HIV-1 naïve individuals on ART.

The study of the demographic and epidemiological characteristics of HIV-1 subtype B patients in Bulgaria, carried out between 1986 and 2018, provides an in-depth understanding of the prevalence of infection in the country, the dynamics of the epidemic and the importance of various social and biological factors that influence the development of HIV infection. The inclusion of data from different geographic regions of Bulgaria and the analysis of transmission groups at risk of infection provides a comprehensive picture of HIV-1 subtype B, which is a major contributor to the epidemic in the country. This study uses state-of-the-art bioinformatics and genetic tools for subtyping and phylogenetic analysis that also address global trends and local features in the epidemiology of HIV infection.

Observations on demographic characteristics show that men account for 89.4% of all HIV 1 subtype B patients in Bulgaria. This is not surprising, as in HIV infection globally men are overrepresented in epidemiological statistics, especially in MSM cases. Possible explanations for this imbalance between men and women include social, cultural and health factors that significantly influence the diagnosis and treatment of women with HIV in Bulgaria. In addition, men who have sex with men constitute the main vulnerable group, leading to the prevalence of infection among them. On the other hand, the relatively low proportion of women (10.6%) indicates that HIV infection is less prevalent in the female population, which should not, however, lead to an underestimation of the risk among women, especially with their increasing involvement in social and sexual networks that put them at risk. In addition, women often experience greater stigma associated with HIV infection, which may affect their willingness to be tested and seek treatment.

The distribution of age groups showed that the most infected patients were in the range of 20 to 39 years, who constituted 76.8% of the patients studied. This is consistent with global trends where young people are more susceptible to HIV infection [https://www.unaids.org/en/resources/fact-sheet]. The 20 to 39 age group typically includes individuals with active social and sexual practices that increase the likelihood of contact with the infection. The proportion of patients in the under 19 age group is lowest (3.5%), which may be interpreted as a relatively lower risk of infection in this age category. However, this does not mean that young people are completely protected. Sexual health and prevention programmes targeting the younger generation need to be strengthened to avoid future outbreaks in this age group. Also, 5.9% of patients were over the age of 50 years, which is an interesting result as it indicates possible trends of increasing infection at a later age, which may be due to changed social habits, migration or less active testing and diagnosis among the elderly.

The highest concentration of HIV-1 subtype B patients was in the capital city of Sofia (55.5%), followed by Varna (7.3%) and Plovdiv (6.6%). Although more than half of the patients studied live in Sofia, the presence of cases in less urbanized areas indicates that the spread of HIV-1 subtype B is not limited to large cities, but requires efforts to create a prevention strategy that covers all parts of the country.

A small proportion of the patients (1.8%) were foreigners, highlighting that HIV-1 subtype B is mainly localised in Bulgaria. However, migration and international travel also play an important role in the spread of infection. 11.8% of patients reported being infected outside Bulgaria, with a significant proportion of these cases in other European countries. These data show that Bulgaria remains a major source for the spread of HIV-1 infection among Bulgarians, despite the involvement of international factors. These results provide a basis for a better understanding of the global and local factors that influence the spread of HIV-1 and highlight the importance of cooperation between countries in the fight against the epidemic.

Epidemiological data show that the highest prevalence of HIV-1 subtype B is among MSM, at 56.7%. This trend is also observed in other European countries where MSM are a major risk factor for infection. As the MSM group is a major source of HIV transmission, testing and education of these individuals should be a priority in prevention strategies. It is also important to address other at-risk groups, such as HET and IDUs, who also represent significant risk categories.

Phylogenetic analysis of transmission clusters using the MicrobeTrace tool reveals important details about the genetic proximity between different HIV isolates in Bulgaria. Based on the genetic distance, 52 transmission clusters were identified, indicating the main route of HIV-1 spread in the country. Interestingly, most of these clusters were associated with the MSM group, highlighting the major importance of this group in the spread of HIV in Bulgaria.

After successful alignment and subtyping of viral sequences, the next key stage is the analysis of transmission clusters. These are groups of viral isolates that show high genetic similarity, suggesting that these isolates belong to patients who were likely infected through common routes of transmission. The main aim of this phylogenetic analysis is to establish the genetic relationship between viral isolates from different patients and to identify the routes of HIV transmission within specific risk groups.

The use of software, such as MicrobeTrace, to calculate genetic distance allows the generation of visual graphs that show how different viral isolates are related to each other, forming clusters. For the purposes of this study, a genetic distance threshold of 1.5% was used, which is optimal for separating isolates that are likely part of a single epidemiological group. This analysis resulted in the identification of 52 transmission clusters.

Each of the identified transmission clusters shows a different structure that can provide important information about HIV epidemiological networks in Bulgaria. For example, in three of the large clusters containing 56, 59 and 91 sequences, respectively, the predominant transmission group is MSM. In these clusters, the percentage of MSM was 55.4%, 73% and 80%, highlighting the major role of this group as the main route of HIV-1 transmission in Bulgaria. Also, heterosexuals are also a significant component of these clusters. For example, in the clusters with 56 and 59 sequences, the rates of HET were 17.6% and 27.1%, respectively. These results highlight the importance of heterosexual contact in HIV transmission. Of additional interest is the group of injecting drug users, which is also represented in these clusters. The possibility of HIV transmission through sharing injecting practices places IDUs as a vulnerable group for infection, while indicating that transmission may be through different routes and link different vulnerable groups.

We analyzed clusters according to their size and structural characteristics. For example, in cluster 1, which includes 91 viral isolates, the predominant group of 80.2% is MSM, and representatives of the HET and WIN groups are also found in this cluster. This indicates that the infection in this cluster has passed between different transmission groups, with MSM probably playing the major role in transmission. A similar analysis for cluster 2 with 59 sequences, in which 72.9% of the men were part of the MSM group and 27.1% of them belonged to the HET group, indicates a direct link between these two main risk groups.

After all our findings on transmission groups, we should not ignore the existence of stigma that is associated with HIV/AIDS and sexual practices.

In the present study, a detailed phylogenetic analysis of HIV-1 CRF01_AE was also performed using the MicrobeTrace program, which allows the identification of genetic clusters and the tracing of virus transmission in specific populations. CRF01_AE is one of the most prevalent recombinant forms of HIV-1, highlighting the importance of its study in the context of Bulgaria, where this viral subtype plays a significant role in the epidemic.

Using genetic distances between viral isolates of 1.5%, it was possible to uncover phylogenetic relationships between different HIV-1 CRF01_AE isolates, identify transmission clusters and assess transmission routes in different risk groups in Bulgaria. The results of the analysis show that HIV-1 CRF01_AE is actively circulating in the country, spreading not only in traditionally vulnerable groups such as IDUs, but also among HET and MSM.

Six transmission clusters were identified in the HIV-1 CRF01_AE population in Bulgaria, including three larger clusters composed of 5 or more viral sequences and three inbred pairs. This indicates that although the virus is transmitted in a relatively small number of cases (dyads), there are broader clusters in which the virus isolates involved may have common ancestors, showing that the virus can spread actively in different social networks and vulnerable groups.

Of the total number of participants in transmission clusters, 75.6% were male and 24.4% were female. This is interesting as it highlights the predominance of men in these clusters, which may be a result of the overwhelming participation of men in groups with high HIV prevalence, such as PWID and MSM. The main transmission category in these clusters is the IDU group, which accounts for 69.2% of all viral sequences. This highlights the critical role of injecting drug use in the spread of HIV-1 CRF01_AE in Bulgaria. Next in importance is the HET group, which accounts for 21.5% of sequences. A smaller percentage of isolates were found in the MSM group (3.3%), indicating that the prevalence of HIV-1 CRF01_AE in Bulgaria is more highly concentrated following intravenous drug users and heterosexuals than in homosexual communities.

Cluster 1 was the largest, containing 154 sequences, of which 75.3% were male and 24.7% were female. This cluster is dominated by the IDU group, with 70.1% of the sequences belonging to drug users. Then, in the cluster, 19.5% of the participants are heterosexual, and 3.2% are men who have sex with men. Interestingly, in this cluster, three infants were also found to be vertically infected, which may indicate local outbreaks of HIV infection and the importance of controlling vertical transmission. 92.9% of participants reported being infected in Bulgaria, highlighting that the virus is actively circulating in the country.

Cluster 2, composed of 7 sequences, had 71.4% men and 28.6% women. In this cluster, a bridge of transmission along the WIN to HET axis is detected, indicating that the virus is passing between these two groups, highlighting the dynamics of infection between different social groups. One of the patients reported being infected abroad, suggesting that the virus may be transmitted into the country through migration.

Cluster 3, which includes 5 sequences, also shows transmission along the WIN to HET axis, but with a small cluster size, reflecting a smaller number of participants and more limited transmission between these groups.

Clusters 4, 5 and 6 represent dyads or inbreeding pairs that show transmission between the UIN and HET groups, and these clusters further confirmed the existence of transmission bridges between different risk categories.

When analyzing the single sequences that did not form clusters, an even distribution by sex was found, with 65.3% of single cases in heterosexual patients and 22.4% in drug addicts. One interesting result is that three neonates were also classified as single cases without being associated with other sequences in existing clusters, indicating that they are independent cases and not part of the HIV core epidemiological network.

In addition to the most prevalent HIV-1 subtypes (subtype B and CRF01_AE), an in-depth analysis was conducted on a large number of ART-näive individuals with HIV, among whom multiple other HIV-1 subtypes and recombinant forms were identified.

In a study conducted between 2012 and 2020 in Bulgaria, 1053 patients diagnosed with HIV-1 infection who met the criteria for antiretroviral therapy (ART) naïve were analyzed. The phylogenetic analysis performed by subtyping and classifying viral isolates allowed us to elucidate the genetic profiles of viruses and identify transmission clusters that reveal the routes of transmission among different vulnerable groups.

The results of the demographic analysis showed that the majority of diagnosed patients were male, accounting for 85.9% of all cases, while females represented 14.1%. This is consistent with the global trend in which men, especially those from at-risk groups such as MSM, are more commonly affected by HIV infection [**Alexiev, 2020**]. The age distribution shows that the main age group diagnosed with HIV-1 is between 20 and 39 years, with these patients accounting for 72.5% of all cases between 2012 and 2020. This is consistent with the active social and sexual lives in these age groups, which puts them at risk for sexual transmission of the virus. The age of the patients ranged from 13 to 78 years, with the smallest percentage being under the age of 19 years.

Epidemiological data show that 80.8% of patients are Bulgarian citizens, while 19.2% are foreign nationals diagnosed in Bulgaria. This confirms the international nature of the HIV epidemic and highlights the importance of migration and population mobility in the introduction and spread of HIV. Infection is most commonly acquired in the European Union, with cases predominantly from Germany, Spain and the United Kingdom. However, there are also a significant number of patients who become infected in Bulgaria, highlighting the importance of the domestic mechanisms of HIV spread in the country.

Geographical analysis shows that Sofia is the most affected region, with almost half of all diagnosed cases. This result is not surprising given the higher mobility of the population. Plovdiv is the second largest region, but the difference with Sofia is significant. However, the prevalence of HIV-1 infection is lower in smaller regions of the country, which can be explained by lower social mobility and fewer migration flows in these areas.

When looking at the distribution of transmission categories, the results show a clear dominance of MSM, with this category comprising 50.7% of all diagnosed cases. The next largest category was HET, with 38.1%, while WIN accounted for 9.6% of cases. The presence of the MSM+UID category (persons who fall into two transmission groups) is small, indicating that despite the small numbers, this group is still vulnerable to infection due to overlapping risk practices.

The HIV-1 sequence subtyping process shows that the dominant subtype in Bulgaria is subtype B, which accounts for 60.4% of all cases. This is followed by CRF01_AE with 11.4%, F1 with 6.9% and CRF02_AG with 5.2%. The remaining subtypes and recombinant forms of HIV-1 virus have lower values, not crossing the 5% threshold. This indicates that subtype B continues to be the major form of HIV-1 infection in Bulgaria and the leading form in new cases.

Phylogenetic analysis using Cluster Picker revealed the presence of 88 clusters for subtype B, focusing on those with 5 or more sequences. This allowed to follow the evolution of the virus and to identify possible transmission bridges between different transmission categories. Clusters containing only a single transmission category, such as MSM, did not show bridge formation, whereas more complex clusters that included patients from different transmission categories, such as MSM and HET, showed possible pathways of virus spread between these groups. For example, in cluster 26, transmission of infection from MSM to a HET woman was observed, demonstrating the importance of focused prevention measures that reach different social and sexual groups.

In the analysis of non-B subtypes (including CRF01_AE, CRF02_AG, CRF12_BF, and URFs), 50 clusters were found, with the largest number of clusters associated with CRF01_AE, which is usually associated with the IDU group. The possibility of transmission of infection between different risk groups was also observed here, for example, in cluster 22 of CRF01_AE, where transfer of infection between HET and UIN was observed. This shows that despite differences in transmission categories, HIV-1 continues to spread between different social and sexual groups, necessitating targeted efforts to prevent the epidemic.

Findings from this analysis highlight the importance of targeting key risk groups such as MSM and IDUs through targeted health interventions. The distribution of HIV shows that despite a global trend towards an increase in new infections in urban areas, smaller regions are also not exempt from the spread of HIV, requiring improved access to medical services and prevention in these areas [**Alexiev, 2020; Alexiev, 2016**].

The summary results of the analysis of patients diagnosed with HIV-1 in Bulgaria highlight key aspects of the epidemiological situation, including the dominance of subtype B and the importance of MSM and HCT as major transmission groups. The concentrated prevalence in urbanised regions, such as Sofia and Plovdiv, shows geographically specific trends, and migration plays an important role in the spread of the virus, with European countries being the main sources of infection.

Phylogenetic and transmission analysis of HIV-1 in Bulgaria provides valuable evidence of the dynamics of the epidemic, with findings of transmission clusters and genetic proximity between viral isolates revealing the complexity of spread through different risk groups. The results show that HIV infection in Bulgaria is mainly spread among MSM and HET groups, with vulnerable groups such as IDUs also playing an important role in the epidemiological network.

Additionally, the analysis of HIV-1 CRF01_AE shows the complexity of the epidemiological situation, as the main transmission channel remains the IDU group, but there is also significant transmission between HET and MSM. These results underscore the need for targeted prevention interventions that focus on vulnerable groups and include strategies to prevent vertical transmission, drug use, and heterosexual relationships to reduce the spread of HIV in Bulgaria. These findings highlight the importance of adapting prevention strategies to the specific epidemiological conditions in Bulgaria and focusing on at-risk groups to effectively combat the epidemic.

6. Conclusions

1. Through careful selection during the period of this thesis, suitable HIV-1 patients were selected for the study. From their plasma samples, a fragment of the HIV-1 *pol* gene was successfully sequenced for further analysis.

2. The obtained genetic information was pooled with sequences from the NRCL of HIV/AIDS archive. A complete data set was generated using demographic and epidemiological data of the study patients required for accurate molecular epidemiological analysis.

3. Subtypes and circulating recombinant forms of the studied sequences were defined using subtyping, recombinant analysis and manual phylogenetic analysis tools. Our analysis revealed that HIV-1 subtype B was the most common subtype, followed by CRF01_AE and CRF02_AG, as well as the presence of multiple introductions of diverse genetic forms in Bulgaria from different geographic regions.

4. Through phylogenetic analysis, transmission clusters were identified in both the widespread HIV-1 subtypes and CRFs and the less common HIV-1 genetic subtypes, confirming the diversity and complexity of the epidemic process in the HIV-1 population in the country.

5. Analysis of transmission clusters showed widespread HIV-1 in the country with multiple introductions from abroad. The clusters identified were heterogeneous in terms of gender, age, behavioral risk factors, and social groups, highlighting the presence of transmission bridges between different vulnerable populations.

6. Comprehensive molecular epidemiological analysis revealed that the HIV-1 epidemic in Bulgaria is characterized by a significant diversity of different HIV-1 subtypes and recombinant forms and dynamics of spread, as a result of multiple independent introductions from other countries and the presence of internal transmission between diverse vulnerable populations. These results highlight the complexity and dynamics of the HIV-1 epidemic in the country and provide valuable information on its evolution and characteristics

7. Contributions

Original Contributions

1. First detailed molecular virological study on the transmission clusters of introduced and prevalent HIV-1 subtypes in Bulgaria was conducted.

2. A total of 1 466 sequences of HIV-1 *pol* genes isolated from patients diagnosed with HIV-1 infection in Bulgaria were analyzed. The studied samples included the main circulating genetic forms in the country: subtype B, CRF01_AE, CRF02_AG and others. Sequences from the study have been deposited in the NCBI genebank.

3. Phylogenetic trees composed of bulgarian and reference sequences have been reconstructed, showing the similarity between sequences and tracing the evolutionary history of transmission relationships in the country.

4. Numerous transmission clusters have been identified, which vary in size and heterogeneity and include representatives of diverse vulnerable population groups with possible spillover between them.

5. The identified phylogenetic clusters indicate the presence of dynamic transmission events among vulnerable populations over a short period of time.

6. The results of our study make a significant contribution to a better understanding of the HIV-1 epidemic in Bulgaria.

Contributions of a Scientific and Applicational Character

1. The identification of a wide variety of subtypes and recombinant forms, as well as the characterization of transmission clusters involving vulnerable populations, can serve as a basis for the development of more effective strategies for the prevention and control of HIV infection in the country.

2. These data could be used to guide public health in planning interventions targeting specific at-risk groups and transmission routes.

8. References

1. Alexiev, I., Beshkov, D., Shankar, A., Hanson, D.L., Paraskevis,D., Georgieva, V., Karamacheva, L., Taskov, H., Varleva,T., Elenkov, I., et al. (2013) Detailed molecular epidemiologiccharacterization of HIV-1 infection in Bulgaria reveals broaddiversity and evolving phylodynamics. PLoS ONE 2013, 8,e59666.
2. Alexiev, I., Shankar, A., Dimitrova, R., Gancheva, A., Kostadinova, A., Teoharov, P., Golkocheva, E., Nikolova, M., Muhtarova, M., Elenkov, I., Stoycheva, M., Nikolova, D., Varleva, T., & Switzer, W. M. (2016). Origin and spread of HIV-1 in persons who inject drugs in Bulgaria. Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases, 46, 269–278. https://doi.org/10.1016/j.meegid.2016.05.029
3. Alexiev, I., M Campbell, E., Knyazev, S., Pan, Y., Grigorova, L., Dimitrova, R., Partsuneva, A., Gancheva, A., Kostadinova, A., SegIDU-Devaux, C., & M Switzer, W. (2020). Molecular Epidemiology of the HIV-1 Subtype B Sub-Epidemic in Bulgaria. Viruses, 12(4), 441. https://doi.org/10.3390/v12040441
4. Global HIV & AIDS statistics — Fact sheet https://www.unaids.org/en/resources/fact-sheet accessed: 15.04.2024
5. L. Grigorova, R. Dimitrova, A. Partsuneva, A. Gancheva, A. Kostadinova, I. Elenkov, N. Yancheva, M. Stoycheva, Ts. Doychinova, L. Pekova, M. Kosmidis and I. Alexiev. Drug resistance mutations and transmission clusters of the HIV-1 CRF01_AE sub-epidemic in Bulgaria Problems of infectious and parasitic diseases, volume 49, number 3/2021.

Acknowledgements

I would like to express my gratitude towards my academic supervisor, Associate Professor Ivaylo Aleksiev, for support, assistance, useful advice and encouragement, as well as for the opportunity to learn.

I am grateful to my family and friends who have been there for me through every difficulty.

This study was funded by :

- National Centre for Infectious and Parasitic Diseases, Sofia, Bulgaria;
- Ministry of Health of the Republic of Bulgaria;
- Project BG05M2OP001-1.002-0001 ‘Fundamental, translational and clinical research in the field of infections and infectious immunology’, funded by the OP ‘Science and Education for Intellectual Growth 2014 - 2020’ and the European Regional Devel

Publications Related to the Thesis

Publications With Impact Factor

1. Molecular Epidemiological Analysis of the Origin and Transmission Dynamics of the HIV-1 CRF01_AE Sub-Epidemic in Bulgaria. Ivailo Alexiev, Ellsworth M. Campbell, Sergey Knyazev, Yi Pan, **Lyubomira Grigorova**, Reneta Dimitrova, Aleksandra Partsuneva, Anna Gancheva, Asya Kostadinova, Carole SegIDU-Devaux, Ivaylo Elenkov, Nina Yancheva and William M. Switzer. Viruses 2021, 13, 116. https://doi.org/10.3390/v13010116
   Impact Factor: 5.048
   Quartiles Q1 (2021)

2. Transmitted HIV Drug Resistance in Bulgaria Occurs in Clusters of Individuals from Different Transmission Groups and Various Subtypes (2012–2020). Ivailo Alexiev, Anupama Shankar, Yi Pan, **Lyubomira Grigorova**, Alexandra Partsuneva, Reneta Dimitrova, Anna Gancheva, Asya Kostadinova, Ivaylo Elenkov, Nina Yancheva, Rusina Grozdeva, Dimitar Strashimirov, Mariana Stoycheva, Ivan Baltadzhiev, Tsetsa Doichinova, Lilia Pekova, Minas Kosmidis, Radoslava Emilova, Maria Nikolova and William M. Switzer. Viruses 2023, 15(4), 941; https://doi.org/10.3390/v15040941
   Impact Factor: 5.818
   Quartiles Q1 (2023)

Publication Without Impact Factor

1. DRUG RESISTANCE MUTATIONS AND TRANSMISSION CLUSTERS OF THE HIV-1 CRF01_AE SUB-EPIDEMIC IN BULGARIA **Lyubomira Grigorova**, Reneta Dimitrova, Aleksandra Partsuneva, Anna Gancheva, Asya Kostadinova, Ivaylo Elenkov, Nina Yancheva, Mariyana Stoycheva, Tsetsa Doychinova, Liliya Pekova, Minas Kosmidis and Ivailo Alexiev. Problems of infectious and parasitic diseases, volume 49, number 3/2021.

Lyubomira Svilenova Grigorova

Molecular virological analysis of transmission clusters of introduced and disseminated HIV-1 subtypes in Bulgaria