MOLECULAR EPIDEMIOLOGICAL ANALYSIS OF THE TRANSMISSION CLUSTERS OF THE HIV-1 CIRCULATING RECOMBINANT FORMS CRF01_AE AND CRF02_AG IN BULGARIA

I. Alexiev[1], R. Dimitrova[1],
A. Gancheva[1], A. Kostadinova[1],
L. Grigorova[1], A. Partsuneva[1],
I. Elenkov[2], N. Yancheva[2],
M. Stoycheva[3],
T. Doychinova[4],
L. Pekova[5], R. Emilova[6],
M. Nikolova[6]

[1]National Reference Confirmatory Laboratory of HIV, NCIPD, Sofia, Bulgaria
[2]Specialised Hospital for Infectious and Parasitic Diseases, Sofia, Bulgaria
[3]Medical University, Department of Infectious Diseases, Plovdiv, Bulgaria
[4]Medical University, Department of Infectious Diseases, Pleven, Bulgaria
[5]University Hospital, Clinic of Infectious diseases, Stara Zagora, Bulgaria
[6]National Reference Laboratory of Immunology, NCIPD, Sofia, Bulgaria

ABSTRACT

Background

The purpose of this study was to analyse the underlying HIV transmission clusters of individuals from different vulnerable groups infected with the recombinant forms of HIV-1 – CRF01_AE and CRF02_AG, between 1986 and 2011 using sequencing and phylogenetic analysis.

Material and methods

Blood samples from randomly selected 242 individuals diagnosed with HIV-1 CRF01_AE and CRF02_AG in Bulgaria were analysed. HIV-1 *pol* gene fragment was sequenced using the ViroSeq HIV-1 Genotyping Test (Abbott) and/or TruGene DNA Sequencing System. The phylogenetic tree reconstruction was performed with the IQ-TREE program. Identification of the phylogenetic clusters was performed using the ClusterPicker program with two parameters to identify recent and distant infection.

Results

Two main local independent epidemics confined to different geographical regions of the country were caused by HIV-1 CRF01_AE and CRF02_AG. The various viral strains circulate predominantly in two major separate regions – CRF01_AE in Sofia and CRF02_AG in Plovdiv. Most of the individuals infected with viral strains analysed in this study were people who inject drugs (PWID) or heterosexuals and only a few were men who have sex with men. The phylogenetic analysis revealed transmission clusters in both recombinant forms, few in number when confined within a short period of time and multiple clusters over an extended timeframe.

Conclusions

The introduction and rapid spread of two different strains of HIV-1 into geographically distant groups of PWID triggered local epidemic outbreaks. The phylogenetic analysis indicated the accelerated transmission of HIV, which is a characteristic of the spread through injection practices. Our study demonstrated that transmission cluster monitoring is important for better understanding of the development of epidemic and could be used as a tool for the identification of risk indicator populations.

ADDRESS FOR CORRESPONDENCE:

Ivailo Alexiev
National Reference Confirmatory Laboratory of HIV
National Centre of Infectious and Parasitic Diseases
44a Gen. N. Stoletov Blvd., 1233 Sofia, Bulgaria
e-mail: ivoalexiev@yahoo.com

INTRODUCTION

HIV-1 is a result of cross-species transition from SIV in chimpanzees to HIV in humans (1, 2). HIV-1 has several major groups: M, H, O and P. Group M is the most significant for the current pandemic and contains several phylogenetically distinct subtypes (A, B, C, D, F, G, H, J and K), circulating recombinant forms (CRFs) and numerous unique recombinant forms (URFs) (3).

HIV-1 subtypes and CRFs are unevenly distributed in the world. This phenomenon is the result of different founder effects followed by local spread of a specific subtype within certain socioeconomic

environment and circulation within specific vulnerable groups (3, 4).

Subtype B is dominant in North America and Western Europe, and subtype A – in some countries of Eastern Europe and Central Asia, including Russia (3). Subtype C is the most abundant HIV-1 subtype in the world and is prevalent mostly in South and Eastern Africa and Southeast Asia (3). CRFs and URFs are widely distributed in Africa and in countries where different subtypes co-circulate (3, 5).

In Bulgaria, multiple HIV-1 subtypes and recombinant forms have been introduced from different countries of the world. Subsequently, the introduced strains were disseminated unequally among individuals from different transmission groups, including heterosexuals (HET), men who have sex with men (MSM) and people who inject drugs (PWID) (4, 6, 7, 8). Due to random events, certain HIV-1 strains have a chance of spreading sharply after being introduced into vulnerable groups forming expanding transmission clusters representing local outbreaks. Such events have been observed since 2005 when two different CRFs (CRF01_AE and CRF02_AG) were independently introduced and rapidly disseminated among two geographically distinct subgroups of PWID, resulting in local HIV-1outbreaks (4, 9).

Subtype B is the most widespread in Bulgaria but found in less than half of the HIV-1+ individuals in the country. The remaining over 50% of the introduced strains are non-B subtypes, CRFs and URFs, of which CRF01_AE and CRF02_AG are the most prevalent. Although various subtypes were initially introduced in the country, subtype B, CRF01_AE and CRF02_AG are the three major HIV-1 strains in the country. HIV-1 CRF01_AE and CRF02_AG have been found to be the most prevalent among the most vulnerable groups, such as PWID. In addition, according to our epidemiological data, vulnerable individuals represent a significant proportion of the current HIV+ population in Bulgaria and since 2005 there has been a sharp increase in the incidence of HIV among PWIDs, leading to an outbreak with significant involvement of the CRF01_AE and CRF02_AG viruses (4, 10, 11).

The purpose of this national representative study was to analyse the underlying HIV-1 transmission clusters of the two most widespread recombinant forms of HIV-1 CRF01_AE and CRF02_AG in Bulgaria using the most up-to-date methods for phylogenetic analysis. Using the ClusterPicker program, we analysed the transmission clusters of the CRF01_AE and CRF02_AG and the participation of different transmission groups in the accelerated spread of these viruses in Bulgaria.

MATERIAL AND METHODS
Study design and specimen preparation
Blood samples from all individuals diagnosed with HIV-1 CRF01_AE and CRF02_AG between 1986 and 2011 were analysed during a clinical follow-up at the National Reference Confirmatory Laboratory of HIV in Sofia, Bulgaria. Plasma samples were linked to epidemiological data through an anonymous numerical code according to the established ethical standards of Bulgaria as previously described (7).

Phylogenetic analysis and cluster identification
The HIV-1 *pol* gene fragment was sequenced using the ViroSeq HIV-1 Genotyping Test (Abbott) and/ or TruGene DNA Sequencing System (Siemens Healthcare) and either the Applied Biosystems 3130xl genetic analyser or an OpenGene DNA sequencing system following the manufacturer's protocol (4).

HIV-1 CRF01_AE and CRF02_AG of the analysed sequences was determined using the automated subtype identification tool COMET v2.2 (12) and REGA HIV-1 subtyping tool version 3.0 (13).

Sequence alignments were performed using the MUSCLE algorithm implemented in AliView version 1.23 (14, 15). Additional quality control of the subtype purity and possible presence of gaps in the sequence was performed. After the clean-up and preliminary quality analysis, the complete dataset contained 141 CRF01_AE sequences comprised of 901 nucleotides in length and 101 CRF02_AG sequences comprised of 918 nucleotides in length.

The phylogenetic tree reconstruction was performed with IQ-TREE v1.6.11 program and was built through the construction of an initial parsimony tree by the phylogenetic likelihood library and search for the best model among 88 DNA models included in the program. The best-fit model was found to be: TPM3 + F + I + G4 according to the Bayesian information criterion (16, 17, 18). Verification of the topology of the phylogenetic tree was performed with generating 1000 samples for ultrafast bootstrap (18). The phylogenetic tree was rooted with midpoint root and was used for further analysis of the phylogenetic clusters. The tree was visualised using FigTree v1.4.4.

Identification of phylogenetic clusters was performed using the ClusterPicker program at a genetic distance of 0.5% and 1.5% corresponding to 0.005 and 0.015 nucleotide substitutions/site, respectively.

RESULTS

Study population demographics

In this study, we generated and analysed 242 HIV-1 *pol* gene sequences from the second and third most prevalent HIV strains in Bulgaria – CRF01_AE and CRF02_AG (Table 1). Men were 74.4% and women

25.6%, most of them in the age group 20-29. According to the permanent address, patients were focused in three major regions: Sofia, Plovdiv and the town of Peshtera. Sofia and Plovdiv were the dominant regions with almost identical proportion of infected individuals (38.4% and 37.6%, respectively). It is curious that completely different viral strains circulated in these two major regions: CRF01_AE in Sofia and CRF02_AG in Plovdiv. In addition, although Peshtera is close to Plovdiv, CRF01_AE that is typical for the more remote region of Sofia has been distributed in Peshtera.

Table 1. Study population.

Characteristics	CRF01_AE		CRF02_AG		Total	
	(n)	(%)	(n)	(%)	(n)	(%)
Total	141	100.0	101	100.0	242	100.0
Men	96	68.1	84	83.2	180	74.4
Women	45	31.9	17	16.8	62	25.6
Age (years)						
≤19	16	11.3	14	13.9	30	12.4
20-29	69	48.9	58	57.4	127	52.5
30-39	41	29.1	18	17.8	59	24.4
40-49	12	8.5	6	5.9	18	7.4
≥50	3	2.1	5	5.0	8	3.3
Region						
Permanent address – Sofia	83	58.9	10	9.9	93	38.4
Permanent address – Plovdiv	8	5.7	83	82.2	91	37.6
Permanent address – Peshtera	18	12.8	0	0.0	18	7.4
Permanent address – other	32	22.7	8	7.9	40	16.5
Transmission category						
HET	63	44.7	23	22.8	86	35.5
MSM	2	1.4	0	0.0	2	0.8
MSM+PWID	3	2.1	1	1.0	4	1.7
PWID	69	48.9	77	76.2	146	60.3
MTCT	4	2.8	0	0.0	4	1.7

HET – heterosexuals; MSM – men who have sex with men; PWID – people who inject drugs; MTCT – mother-to-child transmission.

Epidemiological data in combination with results from the phylogenetic analysis, reviewed that the HIV-1 strain CRF02_AG was highly concentrated in the region of Plovdiv and only 17.8% of these viruses were dispersed outside this region. In contrast, CRF01_AE was more evenly distributed in the country including Sofia (58.9%), Peshtera (12.8%) and other regions (28.4%) – Fig. 1 and Table 1.
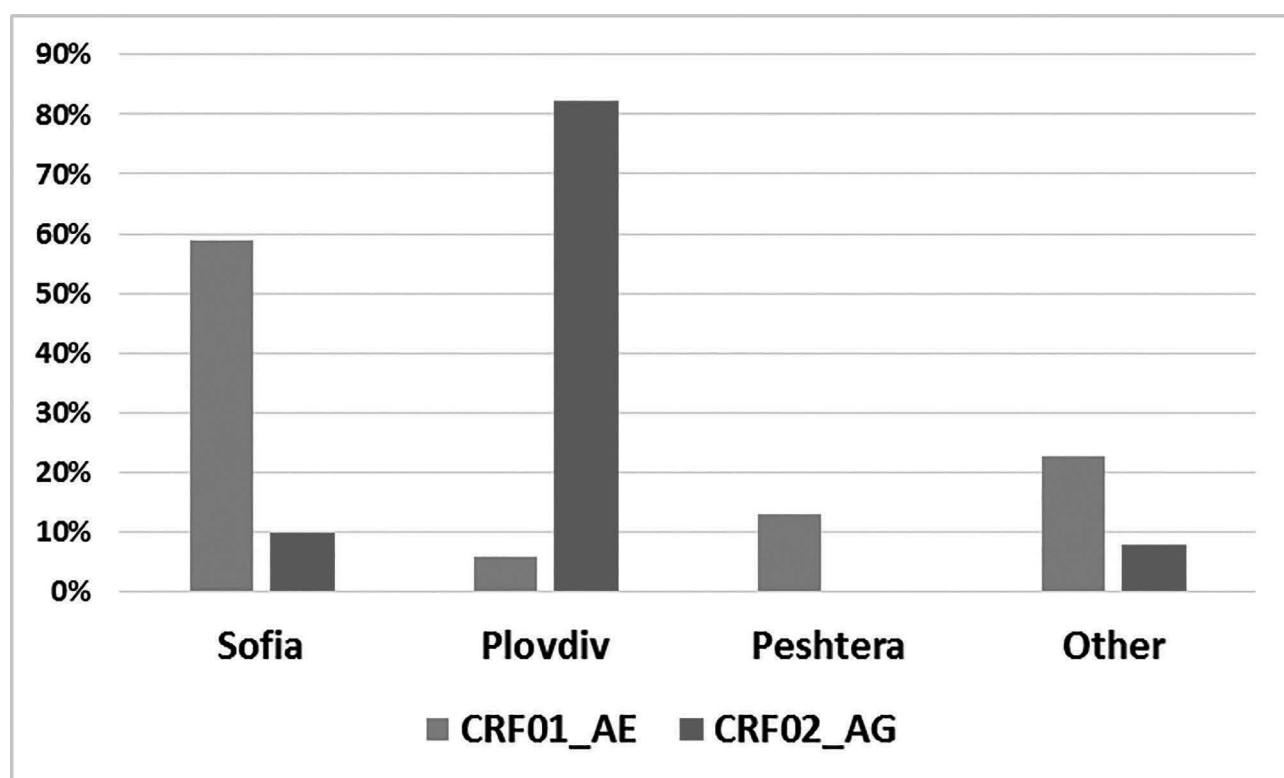
Figure 1. Distribution of CRF01_AE and CRF02_AG in different regions of the country.

There was also a significant difference in the spread of CRF01_AE and CRF02_AG among different transmission groups of the population. CRF01_AE was more evenly distributed between PWID (48.9%) and HET (47.7%), whereas CRF02_AG was found predominantly in PWID (76.2%), and much less in HET (22.8%) – Fig. 2 and Table 1.
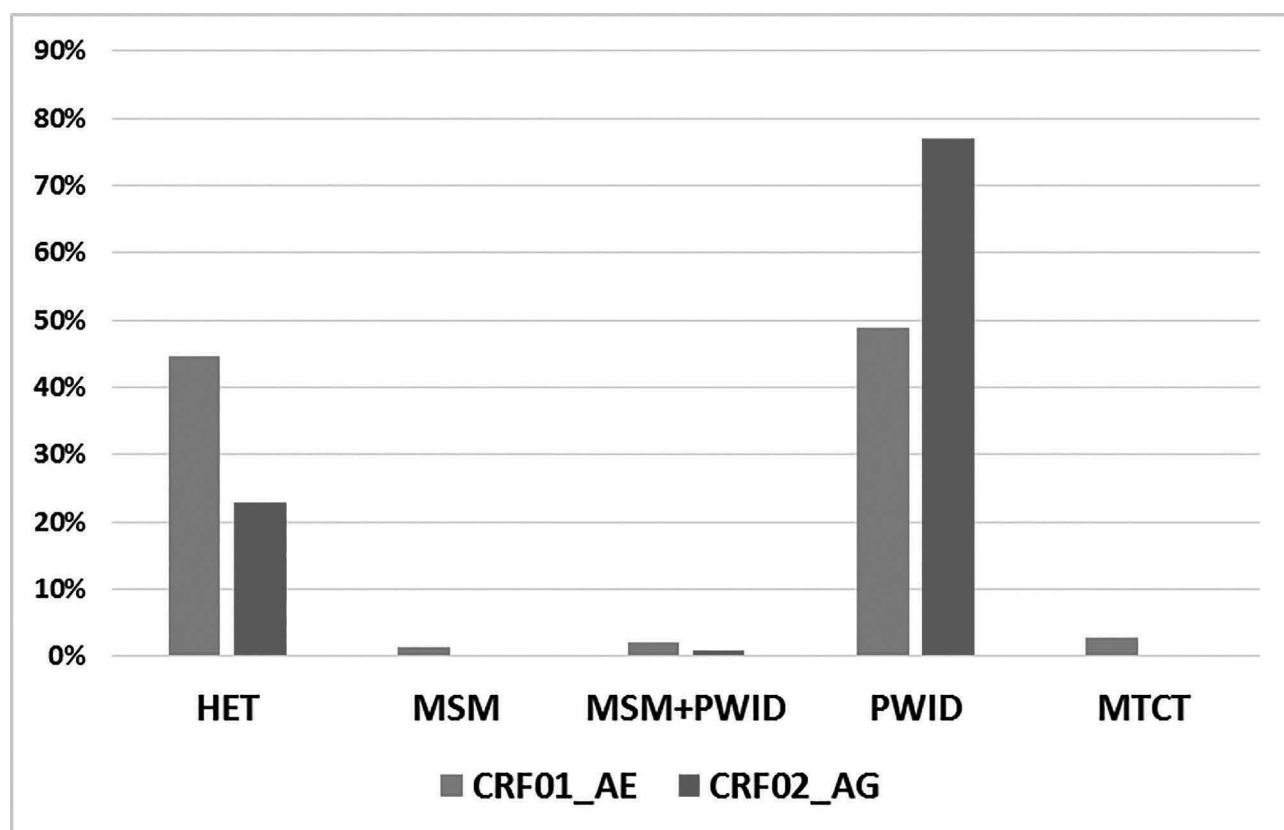

Figure 2. Distribution of CRF01_AE and CRF02_AG in different transmission groups.

Phylogenetic analysis and cluster definition

ClusterPicker software was implemented to analyse the phylogenetic clusters of the two viral strains, CRF01_AE and CRF02_AG. The phylogenetic clusters were defined with two parameters for genetic distance: 0.5% and 1.5% corresponding to 0.005 and 0.015 nucleotide substitutions/site, indicating recent and more remote transmission events. The initial cluster support threshold was defined to have bootstrap values >0.9. Two different options were defined for sequence positions on the topology of the phylogenetic tree: single sequences and phylogenetic clusters. Clusters with threshold size of 10 or more sequences were defined as large clusters.

When analysing the 141 CRF01_AE sequences with genetic distance of 0.5%, two clusters (representing very recent transmission) were identified with two sequences each (Fig. 3 A). At genetic distance of 1.5%, 13 clusters (representing remote in time transmission) with 2 or more sequences were identified, of which 9 clusters were composed of 2 sequences, 2 clusters of 3 sequences, and 2 clusters of 4 sequences (Fig. 3 B).
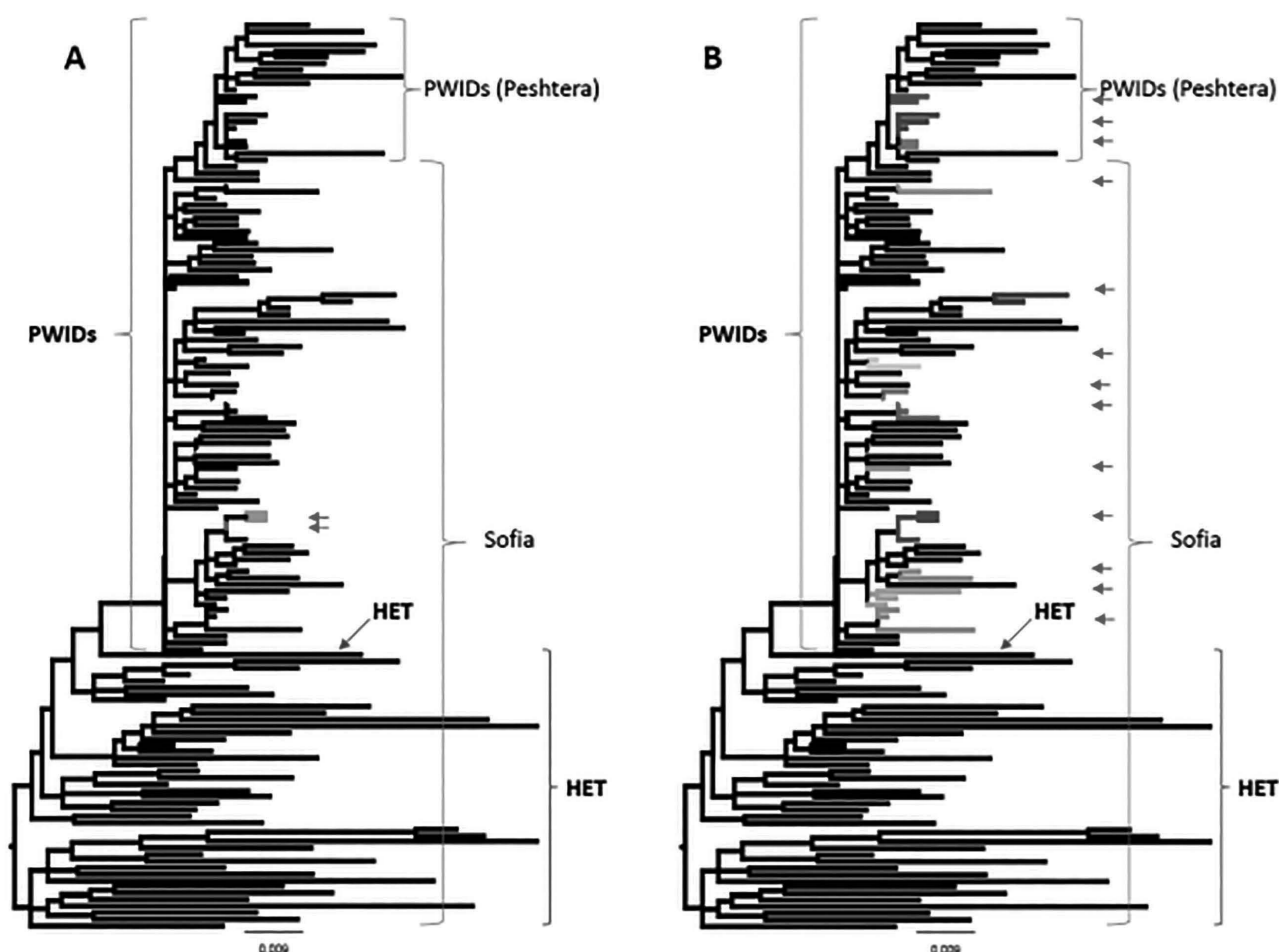
Figure 3 A and B. Phylogenetic tree of CRF01_AE strains reconstructed with ML algorithm using IQ-tree program. **A.** ClusterPicker analysis with genetic distance of 0.005 nucleotide substitutions/site. **B.** ClusterPicker analysis with genetic distance of 0.015 nucleotide substitutions/site. The red arrows indicate the location of the phylogenetic clusters with the corresponding genetic distance. One sequence isolated from HET indicated with blue arrow is at the basis of the sequences isolated from the group of the PWIDs. The size and digital expression of nucleotide substitutions/site are shown at the bottom of the phylogenetic tree.

When analysing 101 CRF02_AG sequences with genetic distance of 0.5%, four clusters (representing very recent transmission) with two sequences each, were identified (Fig. 4 A). At genetic distance of 1.5%, 12 clusters (representing remote in time transmission) with 2 to 5 sequences were identified. Of these, 9 clusters were composed of 2 sequences, 1 cluster of 3 sequences and 2 clusters of 5 sequences (Fig. 4 B).
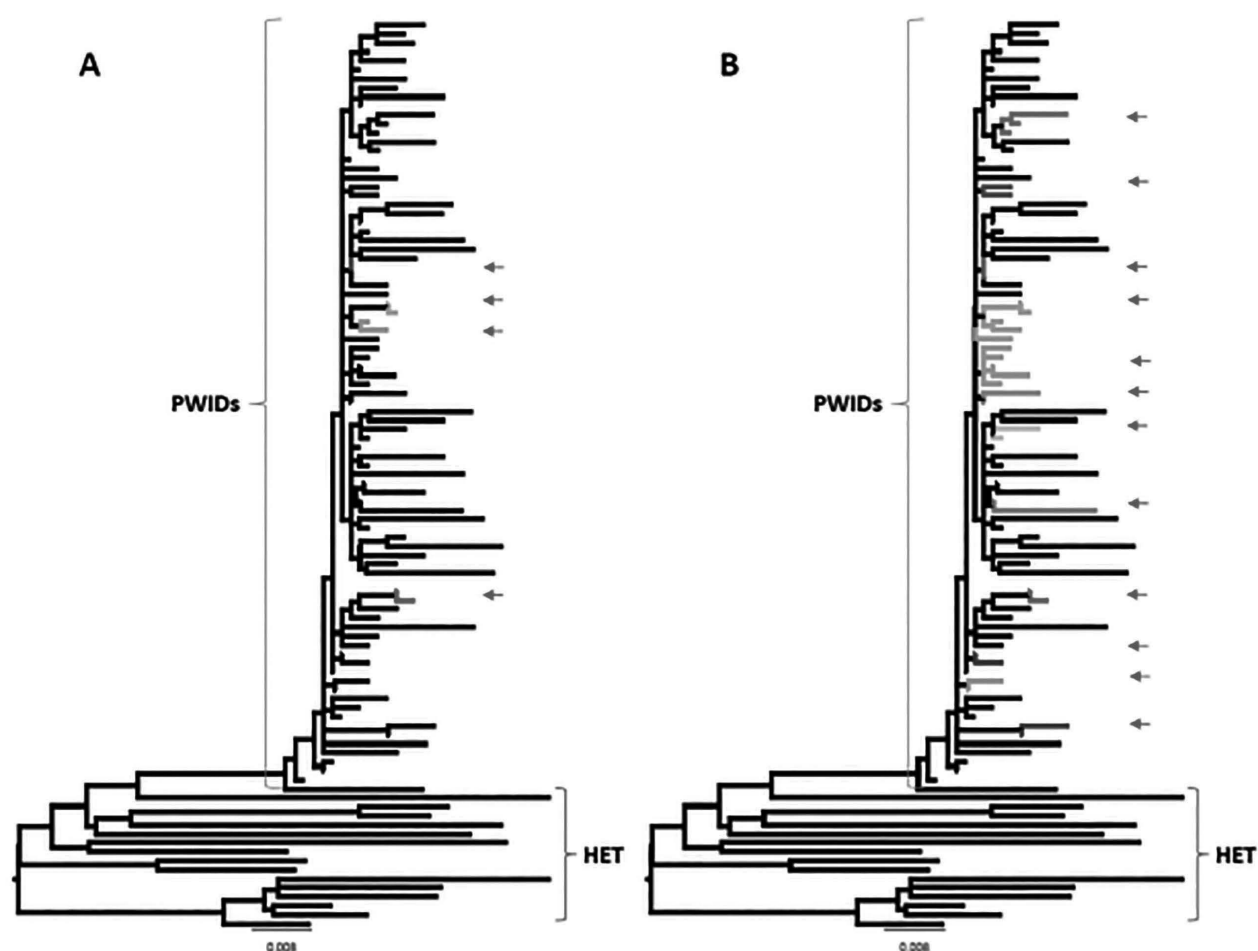
Figure 4 A and B. Phylogenetic tree of CRF02_AG strains reconstructed with ML algorithm using IQ-tree program. **A.** ClusterPicker analysis with genetic distance of 0.005 nucleotide substitutions/site. **B.** ClusterPicker analysis with genetic distance of 0.015 nucleotide substitutions/site. The red arrows indicate the location of the phylogenetic clusters with the corresponding genetic distance. The size and digital expression of nucleotide substitutions/site are shown at the bottom of the phylogenetic tree.

Discussion

In this study, we analysed two of the most prevalent HIV-1 strains in Bulgaria – CRF01_AE and CRF02_AG. The spread of these two recombinant forms is significant both at the country and at the global levels as demonstrated by previous national and global worldwide distribution surveys (4). Of additional interest are the findings that these two strains have been introduced among the most vulnerable individuals to blood-borne infections such as PWID, where viruses have a chance of accelerated spread in limited geographical areas causing an epidemic outbreak among the affected populations. For example, in the late 1990s-early 2000s in Finland, HIV affected a marginalised population of PWID with high rates of imprisonment and homelessness and CRF01_AE, a prevalent variant in Southeast Asia that was circulating in Finland in the early

1990s, was the cause of the outbreak. CRF01_AE was imported from Helsinki, Finland to Stockholm, Sweden leading to an outbreak there among PWID that started probably in around 2003 and was detected in 2006. Similarly, in a PWID-related outbreak in the early 2000s in Northern Italy, HIV-1 diagnoses among PWID formed a monophyletic cluster of subtype G with origin in West Africa (19).

Our phylogenetic analysis uses bioinformatics programs and a range of phylogenetic proximity and genetic distance parameters to analyse isolated viral sequences and to identify phylogenetic clusters that represent transmission events occurring at different time frame.

The smaller genetic distance of 0.005 nucleotide substitutions/site allows identification of transmission events that occurred within one year. Using this time constraint, only two clusters of two

sequences each were detected in the CRF01_AE phylogenetic tree, representing transmission events within the time frame of one year (Fig. 3 A). These transmission events took place within PWIDs in Sofia. In contrast, four clusters with two sequences each were identified for the other HIV-1 strain – CRF02_AG, indicating more turbulent transmission events occurring within a short time frame among the group of PWIDs (Fig. 4 A).

At a greater genetic distance of 0.015 nucleotide substitutions/site, many phylogenetic clusters have been identified in both phylogenetic trees, representing the unfolding of broad-scale transmission events that have taken place over a longer period of time during the expansion of independent local epidemics in remote geographic regions.

Individual groupings of sequences stand out on the topology of phylogenetic trees reconstructed by both recombinant forms of HIV-1. On the phylogenetic tree of CRF01_AE sequences, PWIDs are separated into a distinct group of sequences with relatively short branches generally representing short evolutionary history. The sequences from the HET individuals are positioned closer to the root of the tree. The sequences isolated from HET individuals are at the basis of all sequences and precede all PWIDs sequences showing that these viruses were first introduced and spread within HET transmission group, and later were transferred to PWIDs. This finding is in line with our epidemiological data and previous reports indicating that HIV was introduced into the vulnerable group of PWID later as compared to HET individuals (4). In addition, there are grouped sequences with geographical division and sequences from the town of Peshtera that stand out separately from those of Sofia. Moreover, although the phylogenetic trees have not been dated, the isolated sequences from Sofia appear to be the ancestors of those from Peshtera. Furthermore, the topology of the phylogenetic tree of CRF02_AG sequences also demonstrates division into two major transmission groups: PWIDs and HET. Similarly to the CRF01_AE phylogenetic tree, in the CRF02_AG tree the sequences isolated from HET individuals are closer to the root of the tree while the distinct group of sequences from PWIDs are with relatively short branches representing short evolutionary history. Most of the patients with CRF02_AG were from Plovdiv and only a small

number of sequences have been isolated from persons living outside this region.

Both recombinant forms of HIV-1 – CRF01_AE and CRF02_AG appear to have been introduced in Bulgaria initially among HET individuals and subsequently introduced into the vulnerable group of PWIDs, where the viruses spread rapidly (4). This rapid spread of HIV among vulnerable populations has evoked urgent and widespread actions to curb the epidemic by the Ministry of Health. A campaign was launched for educational initiatives, free tests and counselling, including distribution of needles and disposable syringes among PWIDs. These initiatives led to a significant decrease of the number of newly diagnosed PWIDs with HIV.

Our study has some potential limitations. Firstly, it comprises the period between 1986 and 2011, and may not reflect the overall picture of HIV-1 CRF01_AE and CRF02_AG epidemic in Bulgaria, which is characterised by high dynamics due to the introduction of these HIV-1 strains in different transmission groups of the population. Secondly, our data sample includes only the individuals from whom the HIV-1 *pol* gene was successfully generated, and not those from whom no viral sequence has been obtained due to successful therapy and low viral load.

Conclusions

A large variety of HIV-1 subtypes and recombinant forms has been introduced in Bulgaria. The two recombinant forms CRF01_AE and CRF02_AG have been introduced into geographically distant groups of PWID and triggered local epidemic outbreaks. The phylogenetic analysis indicated the presence of accelerated transmission of HIV, which is characteristic of the spread through injection practices. The presence of a reservoir of a large number of HIV-infected individuals from vulnerable groups is of concern because there is a possibility of rapid dissemination and transmission into the general population. The monitoring of transmission clusters is important for better understanding of the epidemic and for identification of risk indicator population groups.

Acknowledgments

REFERENCES

1. Van Heuverswyn F, Peeters M. *The origins of HIV and implications for the global epidemic.* Curr Infect Dis Rep. 2007; 338-346.
2. Sharp PM, Hahn BH. *Origins of HIV and the AIDS pandemic.* Csh Perspect Med. 2011; 1(1):a006841.
3. Hemelaar J, Elangovan R, Yun J, Dickson-Tetteh L, Fleminger I, Kirtley S, Williams B, Gouws-Williams E, Ghys PD, Alash'le GA, Agwale S. *Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis.* Lancet Infect Dis. 2019; 19(2):143-155.
4. Alexiev I, Shankar A, Dimitrova R, Gancheva A, Kostadinova A, Teoharov P, Golkocheva E, Nikolova M, Muhtarova M, Elenkov I, Stoycheva M. *Origin and spread of HIV-1 in persons who inject drugs in Bulgaria.* Infect Genet Evol. 2016; 46:269-278.
5. Angelis K, Albert J, Mamais I, Magiorkinis G, Hatzakis A, Hamouda O, Struck D, Vercauteren J, Wensing AM, Alexiev I, Asjö B. *Global dispersal pattern of HIV-1 CRF01_AE: a genetic trace of human mobility related to heterosexual activities centralized in South-East Asia.* J Infect Dis. 2015; 211(11):1735-1744.
6. Salemi M, Goodenow MM, Montieri S, de Oliveira T, Santoro MM, Beshkov D, Alexiev I, Elenkov I, Elenkov I, Yakimova T, Varleva T. *The HIV type 1 epidemic in Bulgaria involves multiple subtypes and is sustained by continuous viral inflow from West and East European countries.* AIDS Res and Hum Retroviruses. 2008; 24(6):771-779.
7. Alexiev I, Beshkov D, Elenkov I, Stoicheva M, Nikolova D. *Molecular Epidemiology Surveillance of Pure HIV-1 Subtypes in Bulgaria.* Probl Infect Parasit Dis. 2013; 41(1):19-25.
8. Alexiev I, Shankar A, Wensing AM, Beshkov D, Elenkov I, Stoycheva M, Nikolova D, Nikolova M, Switzer WM. *Low HIV-1 transmitted drug resistance in Bulgaria against a background of high clade diversity.* J Antimicrob Chemother. 2015; 70(6):1874-1880.
9. Alexiev I, Beshkov D, Georgieva V, Varleva T. *Diversity of HIV-1 in Bulgaria Phylogenetic analysis.* Probl Infect Parasit Dis. 2009; 37(2):9-12.
10. Ganova-Raeva L, Dimitrova Z, Alexiev I, Punkova L, Sue A, Xia GL, Gancheva A, Dimitrova R, Kostadinova A, Golkocheva-Markova E, Khudyakov Y. *HCV transmission in high-risk communities in Bulgaria.* PloS one. 2019; 14(3).
11. Alexiev I, Lo Presti A, Dimitrova R, Foley B, Gancheva A, Kostadinova A, Nikolova L, Angeletti S, Cella E, Elenkov I, Stoycheva M. *Origin and spread of HIV-1 subtype B among heterosexual individuals in Bulgaria.* AIDS Res Hum Retroviruses. 2018; 34(3):244-253.
12. Struck D, Lawyer G, Ternes AM, Schmit JC, Bercoff DP. *COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification.* Nucleic acids Res. 2014; 42(18):e144.
13. Peña ACP, Faria NR, Imbrechts S, Libin P, Abecasis AB, Deforche K, Gomez A, Camacho RJ, de Oliveira T, Vandamme A-M. *Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools.* Infect Genet Evol. 2013; 19:337-348.
14. Larsson A. *AliView: a fast and lightweight alignment viewer and editor for large datasets.* Bioinformatics. 2014; 30(22):3276-3278.
15. Robert EC. *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res. 2004; 32(5):1792-1797.
16. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. *IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies.* Mol Biol Evol. 2015; 32(1):268-274.
17. Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermiin LS. *ModelFinder: fast model selection for accurate phylogenetic estimates.* Nat Methods. 2017; 14(6):587.
18. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. *UFBoot2: improving the ultrafast bootstrap approximation.* Mol Biol Evol. 2018; 35(2):518-522.
19. Nikolopoulos GK, Kostaki E-G, Paraskevis D. *Overview of HIV molecular epidemiology among people who inject drugs in Europe and Asia.* Infect Genet Evol. 2016; 46:256-268.