Molecular epidemiology of HIV-1 infection in Europe: an overview

Apostolos Beloukas\textsuperscript{1,2,*}, Alexandros Psarris\textsuperscript{1*}, Polina Giannelou\textsuperscript{1}, Evangelia Kostaki\textsuperscript{1}, Angelos Hatzakis\textsuperscript{1}, Dimitrios Paraskevis\textsuperscript{1}

\textsuperscript{1}Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, Athens, Greece,

\textsuperscript{2}Institute of Infection & Global Health, University of Liverpool, Liverpool, United Kingdom.

* A. Beloukas and A. Psarris contributed equally to this work and they will jointly share the first authorship.

\textbf{Words: 5, 699} / \textbf{Abstract: 297} / \textbf{Figures: 3} / \textbf{References: 127}

\textbf{Correspondence to:} Dimitrios Paraskevis, Assistant Professor of Epidemiology and Preventive Medicine at the Department of Hygiene, Epidemiology and Medical Statistics of the University of Athens Medical School; 75 Mikras Asias str. PC: 11527, Athens, Greece

E-mail: dparask@med.uoa.gr

\textbf{Abbreviations}\textsuperscript{1}

\textsuperscript{1} AIDS, Acquired immunodeficiency syndrome; CRF, Circulating recombinant form; DRC, Democratic Republic of Congo; EU, European Union; EEA, European Economic Area, ECDC, European Centre for Disease Prevention and Control; ESAR, European Society for Translational Antiviral Research; HIV, Human immunodeficiency virus; MSM, Men who have sex with men; MSW, Men having sex with women; PR, Protease; PWID, People Who Inject Drugs; RT, Reverse transcriptase; SIV, Simian Immunodeficiency Viruses; SPREAD, Strategy to Control Spread of HIV Drug Resistance; tMRCA, time to Most Recent Common Ancestor, URF, Unique recombinant form; WHO, World Health Organization
Highlights

- Non-B subtypes increased their prevalence across Western and Central Europe
- In Eastern European countries, except Russia, subtypes’ distribution remains stable
- Most prevalent non-B subtypes are A, C, F and G and the recombinants CRF01, CRF02
- Migration from high prevalent areas is the reason for introducing divergent strains
- Non-prevalent clades are more frequent circulate amongst immigrant populations
Abstract (297 words)

Human Immunodeficiency Virus type 1 (HIV-1) is characterised by vast genetic diversity. Globally circulating HIV-1 viruses are classified into distinct phylogenetic strains (subtypes, sub-subtypes) and several recombinant forms. Here we describe the characteristics and evolution of European HIV-1 epidemic over time through a review of published literature and updated queries of existing HIV-1 sequence databases. HIV-1 in Western and Central Europe was introduced in the early-1980s in the form of subtype B, which is still the predominant clade. However, in Eastern Europe (Former Soviet Union (FSU) countries and Russia) the predominant strain, introduced into Ukraine in the mid-1990s, is subtype A (A$_{FSU}$) with transmission mostly occurring in people who inject drugs (PWID). In recent years, the epidemic is evolving towards a complex tapestry with an increase in the prevalence of non-B subtypes and recombinants in Western and Central Europe. Non-B epidemics are mainly associated with immigrants, heterosexuals and females but more recently, non-B clades have also spread amongst groups where non-B strains were previously absent - non-immigrant European populations and amongst men having sex with men (MSM). In some countries, non-B clades have spread amongst the native population, for example subtype G in Portugal and subtype A in Greece, Albania and Cyprus. Romania provides a unique case where sub-subtype F1 has predominated throughout the epidemic. In contrast, HIV-1 epidemic in FSU countries remains more homogeneous with A$_{FSU}$ clade predominating in all countries. The differences between the evolution of the Western epidemic and the Eastern epidemic may be attributable to differences in transmission risk behaviours, lifestyle and the patterns of human mobility. The study of HIV-1 epidemic diversity provides a useful tool by which we can understand the history of the pandemic in addition to allowing us to monitor the spread and growth of the epidemic over time.
42 **Key Words:** HIV-1; subtypes; molecular epidemiology; European HIV-1 epidemic;
43 phylogeny; genetic diversity
1. Introduction

The Human immunodeficiency virus type 1 (HIV-1) epidemic is the most devastating in human history and remains a global public health problem with an estimated 2.5 million people living with HIV in the WHO European area in 2014 ((ECDC), 2015; Tebit and Arts, 2011; UNAIDS, 2013). Around half of these people are undiagnosed making identification of transmission networks important for targeted public health intervention programmes.

The origins of HIV can be traced to multiple zoonotic infections with Simian Immunodeficiency Viruses (SIV) from African non-human primates with the first HIV transmissions identified as occurring in the Democratic Republic of Congo (DRC) in the early 1920s (Faria et al., 2014). There two major types of HIV, HIV type 1 and HIV type 2 (HIV-2), with HIV-2 differing genetically by more than 55% from HIV-1 and being far less widespread. HIV-1 is characterised by its high genetic variability caused by its high replication rate, recombination and error-prone replication due to lack of proof reading activity of the reverse transcriptase enzyme resulting in high substitution rate. (Peeters and Sharp, 2000; Seillier-Moiseiwitsch et al., 1994). It is classified into four distinct groups: M (major), O (outlier), N (non-M, non-O), and P that was more recently identified, each corresponding to independent cross-species transmissions of SIVs from chimpanzees of the Central subspecies (Pan troglodytes troglodytes) and in Western lowland gorillas (Gorilla gorilla gorilla) (D'Arc et al., 2015; Gao et al., 1999; Peeters et al., 1989; Plantier et al., 2009; Simon et al., 1998; Van Heuverswyn et al., 2006) (Figure 1). Groups N and P viruses are geographically restricted to Central-Western Africa, notably Cameroon (Ayouba et al., 2000; Peeters et al., 1997; Vallari et al., 2011).
Group M viruses are responsible for the HIV-1 global pandemic, spreading out from Central Africa and are further classified in multiple phylogenetically distinct subtypes (A-D, F-H, J and K), sub-subtypes (A1, A2, F1 and F2) and recombinant forms. Existing clades (subtypes, sub-subtypes and recombinants) can further recombine resulting in new mosaic forms (Figure 2). Recombinants include an expanding list of 79 circulating recombinant forms (CRFs) and multiple unique recombinant forms (URFs) (www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html).

The HIV-1 subtypes/sub-subtypes global distribution is highly heterogeneous and varies geographically (Hemelaar, 2012). In 2004, subtype C was the most prevalent subtype globally, accounting for up to 50% of all infections, followed by A and B, with 12% and 10% respectively. However, subtype A viruses predominate in Central and Eastern African countries (Kenya, Uganda, Tanzania, and Rwanda), as well as in Eastern European countries formerly constituting the Soviet Union (FSU countries). **Subtype B is the main HIV-1 clade in Western and Central Europe**, North America (including USA, Canada, Mexico), as well as in several countries in Central and South America, Caribbean, Australia, Northern Africa, and in the Middle East (Buonaguro et al., 2007).

Based on the earliest AIDS diagnoses and molecular phylogenetic analyses of early strains, the introduction of HIV-1 in Europe dates back to early-1980s (Brunet et al., 1984; Glauser and Francioli, 1984; Melbye et al., 1984; Robbins et al., 2003). At that time, most infections were due to subtype B viruses and mainly associated with sexual transmissions among men having sex with other men (MSM) or men having sex with women (MSW), transfusions (haemophiliacs), and People Who Inject Drugs (PWID) (Brunet et al., 1984; Melbye et al., 1984). **Over the last 15 years, the molecular epidemiology of different HIV-1 subtypes and CRFs, referred to hereafter as HIV-1 clades, in Europe has**

---

2 Classification of European countries intro regions was according to the ECDC/WHO criteria
significantly changed becoming more heterogeneous (Abecasis et al., 2013; Hemelaar, 2011). The increasing complexity raises plausible questions about potential issues in diagnosis, clinical management and even in pathogenesis (Camacho, 2006; Chaix et al., 2013; Easterbrook et al., 2010; Geretti et al., 2009; Hemelaar, 2013; Paraskevis et al., 2013b; Santoro and Perno, 2013; Scherrer et al., 2011; Siemieniuk et al., 2013; Touloumi et al., 2013). Investigation of the European HIV-1 epidemic by means of molecular epidemiology provides valuable information to investigate and monitor the epidemic’s dynamics over time (Frost and Pillay, 2015; Kuhnert et al., 2014; Pybus and Rambaut, 2009; Stadler and Bonhoeffer, 2013; Stadler et al., 2012). Interestingly, knowledge of the global HIV-1 clades’ distribution along with population distribution within a given area can, by itself, provide insights into transmission dynamics.
2. HIV-1 epidemiology in Europe

The cumulative number of diagnosed HIV-1 infections in the European continent (i.e. European Union, European Economic Area (EU/EEA), Russia and FSU countries) reached 1,840,136 by the end of 2014 with 49% of these diagnosed in Russia, as reported by The European Surveillance System (TESSy), a joint ECDC/WHO database for HIV/AIDS surveillance (ECDC, 2015). In 2014, there were 142,197 newly diagnosed HIV-1 infected individuals in EU/EEA and Russia comprising the highest number since reporting started in the 1980s (ECDC, 2015; AIDS, 2015). Of these new infections, 77% were diagnosed in the East, 19.2% in the West and 3.5% in the Centre of the European continent (using the TESSy/WHO criteria for the geographical division). The incidence was found to be more than 7 times higher in the Eastern than in the Western European countries (43.2 vs 6.4 per 100,000 people) and considerably lower in the Central ones (2.6 per 100,000) (ECDC, 2015). Despite public health intervention strategies and extensive prevention programmes to eliminate new HIV-1 infections implemented in the last 10 years in Europe, the rate of new infection has remained rather stable from 6.7 in 2005 to 6.4 per 100,000 in 2014 (both adjusted for reporting delay). Interestingly, about one third of newly acquired HIV-1 infections are among immigrants (including foreign born individuals), whilst two thirds are among natives. The predominant route of HIV-1 transmission is sex between men, with a considerable increase in the rate of new HIV-1 infections among MSM from 30% in 2005 to 42% in 2014 (ECDC, 2015). In Eastern European countries HIV-1 transmission through heterosexual intercourse is considered the main reason for the increased rate of new infections, while transmission through PWID networks also remains high.
3. Overview of HIV-1 diversity in Europe

The geographical distribution and prevalence of HIV-1 clades in the European continent are highly heterogeneous. Briefly, since their introduction subtype B clade has predominated in most Western and Central European countries (Western-type epidemic), whilst in Eastern Europe the epidemic has been dominated by subtype A (A_{FSU}) (Eastern-type epidemic) (Abecasis et al., 2013; Hemelaar, 2011). Interestingly though, the molecular epidemiology of the epidemic in West and Central Europe has significantly changed over the last 15 years. Non-B and/or CRF clades have been introduced mainly through waves of migration from areas where are predominant and have been spread through population mobility between European countries. Subsequently their prevalence in the Central and Western European region have been increased, because of dispersal through European and mainly regionally restricted MSM, PWID and heterosexual transmission networks increasing the complexity of the epidemic (Abecasis et al., 2013; de Oliveira et al., 2010; Fabeni et al., 2015; Fox et al., 2010; Hemelaar, 2011; Hoenigl et al., 2016; Simonetti et al., 2014; Tamalet et al., 2015). Subtype B clade still predominates in Western and Central European countries, but each country has a unique pattern shaped by different regional circumstances and high prevalent local transmission routes.

Conversely, in Eastern Europe the A_{FSU} clade has predominated since the mid-1990s, when Russia and other FSUs had a very low incidence of HIV-1. Transmissions usually occurred through sexual intercourse and there were a large number of circulating strains, including B clade (Bobkova, 2013). At this stage the absolute number of cases was small and cases were largely confined to MSM. An explosive spread of an HIV-1 subtype A with very low genetic diversity was noted first in Ukraine in 1994 (Novitsky et al., 1998) and thereafter in Russia and Belarus (Bobkova, 2013), Azerbaijan (Saad et al., 2006a), Georgia (Zarandia et al., 2006)
and Armenia (Laga et al., 2015) amongst PWID. This subtype has been variably termed subtype IDU-A or A_{FSU} (subtype A associated with states from the former Soviet Union) and is largely confined to PWID (Bobkov et al., 1997; Bobkov et al., 2004; Bobkova, 2013).

Phylogenetic investigation of HIV-1 strains from Kiev, Crimea, Donetsk, Poltava and Odessa revealed that subtypes A and B were simultaneous introduced into Ukraine (Bobkova, 2013), however A_{FSU} strains spread successfully across the FSU territory evolving into one of the fastest growing epidemics in the world, (Bobkova, 2013). HIV-1 subtype B clade infections remained stable in the area circulating mostly among MSMs (Thomson et al., 2009).
4. Prevalence of HIV-1 subtypes in Europe

To describe the evolution of the European epidemic over time, including the changes in the prevalence of different clades, we used data from the SPREAD cohort (Strategy to Control Spread of HIV Drug Resistance); part of the ESAR collaboration (European Society for Translational Antiviral Research) (Abecasis et al., 2013), and the published review on the global prevalence of different HIV-1 subtypes by Hemelaar et al. (Hemelaar, 2011). The SPREAD/ESAR cohort enrolls newly diagnosed patients from 20 European countries and Israel, while the Hemelaar et al review reported estimates in Europe between 2004-2007 (Hemelaar, 2011). The SPREAD cohort database reveals B clade as predominant (70.2%) in newly HIV-1 diagnosed patients, after adjusting for oversampling in some countries, followed by C, CRF02_AG, G and A, with 5.0%, 4.9%, 4.8% and 3.6%, respectively. However, there are countries, such as Portugal, Cyprus, Sweden and Greece, where subtype B viruses are less prevalent in new infections (<50%), whilst in the Czech Republic, Germany, Spain, Slovenia and Poland the prevalence of B clade exceeds 80%. Non-B and CRF clades have mainly been associated with immigrants, heterosexual transmission and male gender. Conversely, there is evidence for regional dispersal among native population of subtype A in Greece and subtype G in Portugal (Carvalho et al., 2015; Esteves et al., 2003; Paraskevis et al., 2007). Notably and unlike any other European country, in Romania the F clade (sub-subtype F1) has predominated since the begging of the epidemic in the late-1980s with as little as 5% of characterised HIV-1 strains belonging to non-F1 subtypes (Apetrei et al., 1998; Stanojevic et al., 2012). However, by 2007 the prevalence of non-F1 strains in treatment naïve individuals showed an increase up to 23% and included clades C, B and A (Paraschiv et al., 2007).

Aligned with the SPREAD cohort data, Hemelaar et al, reported the B clade accounted for 85.2% of the total infections in Western and Central European countries between 2004-2007.
Among the non-B clades, CRF02_AG was the most prevalent (4.5%) followed by C and A (1.91% and 1.76%, respectively). Notably, 9.3% of the total infections were due to CRFs and/or URFs (Hemelaar, 2011).

Using the Los Alamos HIV Sequences Database, we investigated the prevalence of different clades over the last ten years (www.hiv.lanl.gov) in order to obtain a more recent picture of the molecular epidemiology of HIV-1 in Europe. The query resulted in a total of 30,996 HIV-1 sequences across Western, Central and Eastern European countries with sampling dates since 2005. Overall, the results for Western and Central Europe (N=26,758) were in concordance with estimates from the SPREAD/ESAR cohort and specifically B clade was found to be the predominant (69.4%), followed by C (7.0%), A (3.5%), CRF02_AG (3.2%), F (3.0%), G (2.9%), CRF06_cpx (2.8%) and CRF01_AE (1.7%). The countries with the highest prevalence for non-B clades were Romania (85.4%), Ireland (68.2%), Luxembourg (67.2%), Portugal (63.0%), Bulgaria (62.5%), Cyprus (61.1%), Finland (57.1%), Greece (54.0%), United Kingdom (50.4%) and Sweden (49.2%) (Figure 3). In Eastern Europe (N=4,238), there is a trend towards increasing prevalence of non-A clades. Specifically in the Russian Federation the predominant clade is A followed by B (6.5%), CRF63_02A1 (8.9%), CRF02_AG (4.0%) and C clade (1.4%). In the rest of the FSU countries having adequate sampling, we found that subtype A (A_{FSU}) remains the most prevalent clade and there have been no spill over of non-A clades from Russia or other Western and/ or Central European countries to these countries until to date. The only exception was Estonia where CRF06_cpx remains the predominant clade (Zetterberg et al., 2004).

According to the above reviewed data in Western and Central European countries, A clade was the most prevalent among the non-Bs in Slovenia, Czech Republic, Poland, Greece and Cyprus, C predominated in the United Kingdom and Denmark, and subtype G in Portugal. Multiple clades were found at high prevalence in Switzerland (A, C, CRF01_AE,
CRF02_AG), Italy (C, F, G and CRF02_AG), Sweden (C and CRF01_AE), France (A, G, CRF01_AE and CRF02_AG) and Spain (A, C, F, G and CRF02_AG) (Figure 3).

The approaches described above have several limitations. Specifically the SPREAD/ESAR cohort enrolled only newly diagnosed individuals. Similarly, Hemelaar et al reviewed HIV-1 strains sampled between 2004 and 2007. Furthermore, the most recent published data for some of the countries are relatively old. On the other hand, figures available on the Los Alamos HIV sequences database are more recent, but the data may not reflect actual population frequencies and should be regarded as a rough indication of the subtype distributions. Therefore, maps drawn using the latest approach should be interpreted cautiously keeping in mind these limitations.
5. Origin of non-predominant HIV-1 clades in European countries

**Western Europe**

HIV-1 subtype B has been responsible for what is often called the ‘Western epidemic’ in Europe and has remained the predominant clade despite the introduction of non-B clades from later migrating populations. However, the prevalence of non-B subtypes has been increasing linked to migration and later dispersal through transmission networks with patterns varying between individual countries within the region.

In the United Kingdom, HIV-1 was first identified during the 1980s among MSM with all identified strains belonging to B clade (Brown et al., 1997; I., 1988; Wade et al., 1998). However, the pattern in heterosexual groups is changing with a steady increase in the number of non-B clades since the 1990s (Hughes et al., 2009) and has mainly been associated with sub-Saharan African and South American immigration (de Oliveira et al., 2010; Fox et al., 2010; Resistance, 2014). Molecular epidemiological analyses revealed that non-B sequences among heterosexuals in the UK were initially strictly linked with strains from sub-Saharan Africa (Hughes et al., 2009), while the C clade viruses have been associated with South America (de Oliveira et al., 2010; Resistance, 2014).

Nevertheless, a recently published study reported that the prevalence of non-B clades among MSM increased by more than 3 times between 2002 and 2010, and, despite the increase in non-B in heterosexual transmission networks, MSM and PWID are still at high risk for non-B infections (Ragonnet-Cronin et al., 2016). In Ireland a similar pattern of increasing prevalence of non-B subtypes acquired through heterosexual exposure has been observed (De Gascun et al., 2012) and the only subtype identified in Iceland was B up until 1993 when the introduction of non-B subtypes was linked to immigration (Del Amo et al.,
2011; Löve et al., 2000). Although non-Bs were introduced into Belgium somewhat earlier, in the mid-1980s, initial prevalence was relative low but increased over time, from 0% in 1983 (as reported in a small two-clinical sites study) to 57% in 2001. (Fransen et al., 1996; Snoeck et al., 2004). Additionally in Belgium and Luxembourg 55.8% of non-B infections have been detected in individuals originated from Africa, but 30.5% of non-B clades were also found among native population (Dauwe et al., 2015). In the Netherlands the vast majority of non-Bs were linked to sub-Saharan Africa, in addition to single cases from the Caribbean, South America, Thailand, Russia and Italy (Bezemer et al., 2004; Op de Coul et al., 1998). Immigrants from sub-Saharan Africa also introduced non-B clades in France increasing the prevalence of non-B clades from 4% in the 1980s to more than 20% in just a decade, mostly spreading in MSM transmission networks (Barin et al., 1997; Chaix et al., 2013; Tamalet et al., 2015). Notably in a recent study, Brand et al, found that non-B clade infections have spread among individuals of French origin and especially MSM (Brand et al., 2014). Similarly Chaix et al found that a considerable proportion of French heterosexuals (37%) with a primary infection were infected with non-B clades (Chaix et al., 2013). France is probably the only West European country where a much higher proportion than in other countries reaching up to 23% of French Africans citizens are infected with subtype B, suggesting that the regional sub-epidemics in native and immigrant populations are linked (Chaix et al., 2013; Tamalet et al., 2015).

Non-B clades were introduced relatively early in Switzerland with 28.2% of all infections as non-B by the mid-1990s (Böni et al., 1999). These infections are mainly associated with people of African origin (95%), heterosexual transmission (44%) and being female (43%). Conversely, subtype B clade was predominant in European, American and Asian immigrants, with particularly high frequencies in homosexuals (mostly MSM) (97%) and PWID (94%) (Böni et al., 1999). In a study reported by von Wyl et al, subtype C and CRF02_AG were
associated with being of African origin, whilst subtype A was found at similar proportions in western Europeans and Africans. However, CRF01_AE was detected more frequently among Western Europeans than South East Asians. All these non-B clades were mostly associated with heterosexual transmissions (von Wyl et al., 2011).

In Austria, African immigrants were identified in most cases with non-Bs suggesting Africa as the putative origin of non-B infections that subsequently spread within MSM networks (Falkensammer et al., 2007; Hoenigl et al., 2016). In Germany, non-B clades have been detected at a 20% prevalence and were linked with migration from Sub-Saharan Africa (subtype A and CRF02_AG), Eastern Europe (A_FSU) and South Eastern Asia (CRF01_AE) (personal communication with EIDB curators) (EIDB, 2016).

In Austria, African immigrants were identified in most cases with non-Bs suggesting Africa as the putative origin of non-B infections that subsequently spread within MSM networks (Falkensammer et al., 2007; Hoenigl et al., 2016). In Germany, non-B clades have been detected at a 20% prevalence and were linked with migration from Sub-Saharan Africa (subtype A and CRF02_AG), Eastern Europe (A_FSU) and South Eastern Asia (CRF01_AE) (personal communication with EIDB curators) (EIDB, 2016).

In Austria, African immigrants were identified in most cases with non-Bs suggesting Africa as the putative origin of non-B infections that subsequently spread within MSM networks (Falkensammer et al., 2007; Hoenigl et al., 2016). In Germany, non-B clades have been detected at a 20% prevalence and were linked with migration from Sub-Saharan Africa (subtype A and CRF02_AG), Eastern Europe (A_FSU) and South Eastern Asia (CRF01_AE) (personal communication with EIDB curators) (EIDB, 2016).

In Finland between 1988-1994 non-B clades were mainly transmitted heterosexually through direct or indirect contact with African or Southeast Asian populations. (Liitsola, 2000). Later, in 1998, Finland experienced an HIV-1 outbreak among PWID involving infections with CRF01_AE, which was probably introduced from South Eastern Asia (Angelis et al., 2015; Liitsola et al., 2000). In Sweden CRF01_AE strains circulating amongst PWID were introduced from Finland (Skar et al., 2011). The rest of non-B strains including all major subtypes and many different recombinants were introduced from different African regions (subtype C, CRF02_AG), South Eastern Asia (CRF01_AE), but also from Eastern Europe (CRF06_cpx) (Neogi et al., 2014). For Denmark the origin of non-B subtypes has not been described in detail.

During the 1990s, non-B clades were also detected in Portugal, again linked with immigrant populations mainly African. Interestingly, G clade was also reported with high prevalence among PWID (34.1%, 1997-2001) (Esteves et al., 2003), unlike any other neighbouring or not country in the European continent. Although polyphyletic analyses suggested multiple and old introductions of the B and G clades, as expected, additional non-B and non-G clades
have established local epidemics among native individuals, with the dates of the most recent common ancestor estimated to be in the early 2000s (Carvalho et al., 2015). As in Portugal, the introduction of non-B clades in neighbouring Spain, was also documented in the mid to late-1990s, mainly among immigrants (Soriano et al., 1997). Yebra et al in a study from the Spanish ART Naïve cohort suggested that non-B strains were introduced by immigrants and subsequently circulated among natives in Spain (Yebra et al., 2012). In a study from Madrid, non-B transmissions were associated with people of African origin and heterosexuals (González-Alba et al., 2011). Specifically, CRF02_AG and subtype A were more frequently found among Africans, BG recombinants infected mainly PWID and BF recombinants circulate d only in South Americans and Spaniards (González-Alba et al., 2011).

Of particular note is Galicia in Northwest Spain where PWID viruses, the G and CRF14_BG clades were found to have originated from Portugal (Thomson et al., 2001; Thomson and Najera, 2007). Furthermore, an outbreak of subtype F1 was detected in northwest Spain (mostly in Galicia) including sequences from other Western European countries that were found to have originated from Southern America (Delgado et al., 2015; Paraskevis et al., 2015a).

In Italy, Baldanti et al, showed that African ethnicity, heterosexual transmission route of infection and having a recent diagnosis (2000-2006) were independently associated with non-B infections (Baldanti et al., 2008). The latter was also demonstrated by Lai et al, who studied the prevalence of HIV-1 subtypes in 3,670 individuals from 50 centres in Italy between 1980 and 2008 (Lai et al., 2010). Results were very similar to the study reported by Baldanti et al, showing that the prevalence of non-B clades increased from 2.6% in 1980-1992 to 18.9% in 1993-2008, affecting mostly heterosexuals (77.2%) and people of African origin (94.8% of African people carried a non-B strain) (Lai et al., 2010). Nevertheless,
MSM transmission networks have been reported recently to drive the expansion of non-B Italian regional sub-epidemics (Fabeni et al., 2015; Simonetti et al., 2014).

In Greece, subtype A, the most prevalent among the non-Bs, was found to spread among the native population. It was introduced from sub-Saharan Africa, as the result of a single founder event in the late-1970s (MRSA 1977.9; 95% highest posterior density interval, 1973.7-1981.9). The other non-A non-B clades transmissions mostly occurred amongst heterosexual or immigrant population (Paraskevis et al., 2007). The origin of clades associated with PWID differs to the sexually transmitted epidemics in Greece. The origin of the four different clades circulating among PWID during an outbreak in Athens (detected in early 2011), was Afghanistan/Iran (CRF35_AD), Romania (CRF14_BG) and Greece (subtypes A and B) (Niculescu et al., 2015; Paraskevis et al., 2013a; Paraskevis et al., 2015b).

Whilst Israel is a West Asian country, data from here is included in the SPREAD/ESAR cohort, and many other European studies, so is relevant for inclusions in the description of the epidemiology of HIV-1 in Europe. Here, non-B clades were introduced via two major routes; C clade viruses originated from Ethiopia and infected mainly heterosexuals and A_{FSU} clade was introduced from FSU countries and circulated and expanded mostly among PWID transmission networks (Grossman et al., 2015).

Central Europe

The HIV-1 epidemic in Central European countries also has the Western epidemic pattern. Up until 1999, only B clade was detected in Polish HIV-1 infected individuals (Stańczak et al., 2010) and mainly occurred in MSM (42%) and PWID (35%). Non-B clades (namely A1, C, D and F1) were introduced later and detected more frequently among heterosexuals and females (Parczewski et al., 2012; Parczewski et al., 2010;
Similarly in the Czech Republic and Slovakia, non-B strains were found almost exclusively in heterosexuals (Chabadova et al., 2014; Linka et al., 2008; Reinis et al., 2001). Between 2008 and 2010 in Hungary, 96.6% of the patients were infected with subtype B and 3.3% with subtype A (Mezei et al., 2011), whereas subtype C was detected during the late 90s in an isolated case, where the virus was contracted in Africa (Mezei et al., 2000).

In the Balkan states (Bosnia & Herzegovina, Bulgaria, Croatia, Former Yugoslav Republic of Macedonia (FYROM), Montenegro, and Slovenia but excluding Albania, Romania, and Greece), both the epidemiology and prevalence of the various HIV-1 clades are similar to Central Europe and subtype B predominates in MSM (Mezei et al., 2006; Siljic et al., 2013). Other subtypes, particularly recombinant subtypes, have been increasing their prevalence since the beginning of the 21st century. Non-B transmissions in Slovenia were also associated with heterosexuals in contrast to Bs, which circulated among MSM (Stanojevic et al., 2012). Notably the majority of non-B infected individuals (86%) were of Slovenian nationality (Lunar et al., 2013; Stanojevic et al., 2012). Interestingly, the introduction of non-B viruses in Croatia was often reported to have occurred via heterosexual contact with seamen rather than immigrants from highly prevalent areas as is commonly found elsewhere (Ramirez-Piedad et al., 2009).

In Albania, data from the last decade showed that the local HIV-1 epidemic was characterised by a high prevalence of non-B infections (65.2%) (Ciccozzi et al., 2005). Specifically, A clade spread as a result of a founder effect from the A clade epidemic in neighbouring Greece (Paraskevis et al., 2007; Salemi et al., 2008a). In Bulgaria, there are several HIV-1 clades circulating and as it has been shown clades B and A1 were introduced by at least three or four independent sources in last 25 years (Salemi et al., 2008b). Although B clade still predominates, with higher prevalence among women and PWID, there are
several clades (A1, B, C, F1 and H) and CRFs (namely CRFs; 01_AE, 02_AG, 04_cpx, 05_DF, 14_BG, and 36_cpx) circulating among MSM and PWID, increasing the HIV-1 epidemic heterogeneity (Alexiev et al., 2015; Ivanov et al., 2013). In Serbia non-B clades (i.e G, C, A, F, CRF01 and CRF02) has mainly been transmitted heterosexually (Siljic et al., 2013; Stanojevic et al., 2002). In Montenegro there is a low prevalence of subtype A and C viruses. However, the origin of the infections have not yet been identified (Ciccozzi et al., 2011).

In Romania, the HIV-1 epidemic is unique as the globally-rare subtype F1 predominates and any non-F1 subtypes are referred to as divergent strains. Here, significant numbers of mainly institutionalised children were infected in the late 1980s via transfusion of infected blood products or unsafe parenteral treatments (Apetrei et al., 1997). Estimates suggest that as many as 10,000 children were infected (Lucking et al., 2013). However, B clade has been found amongst MSM and heterosexuals at approximately similar rates and have in fact been found to have originated from Western Europe (Paraschiv et al., 2012). There are other non-F1 and non-B clades also circulating with C clade mainly being associated with heterosexuals infected abroad (Paraschiv et al., 2011). Since 2010, an increasing trend of HIV-1 infections amongst PWID has been observed, largely centred in Bucharest. Although F1 clade still predominates in the PWID epidemic, other clades, including CRF14_BG, have also been found (Niculescu et al., 2015). The outbreak of CRF4_BG strains among PWID was found to have originated from Spain, whilst the two subtype F1 sub-outbreaks originated from regionally prevalent Romanian strains (Paraskevis et al., 2015b).

Turkey was included in our analysis as a close, and therefore relevant, neighbour to the European continent. Here, the prevalence of non-B clades was high and included many different subtypes and CRFs. Unfortunately, the origin of these transmissions and/or local epidemics remains unclear (Inan and Sayan, 2014; Stanojevic et al., 2002).
**Eastern Europe**

In the Eastern European sub-continent $A_{FSU}$ clade is the predominant and has been spread through a large PWID-epidemic in the mid-1990s. In the pre-PWID epidemic era, an early study (mid-1980s) from Belarus, Russia and Lithuania described the presence of HIV-1 subtype B clade in homosexually infected individuals and subtype C in heterosexually infected individuals, while A, C, D and G clades were also detected in parentally infected individuals (Lukashov et al., 1995). In the mid-1990s Ukraine experienced the begging of a large epidemic in the PWID communities which subsequently spread into the Russian Federation, Belarus, Moldova, Lithuania, Latvia, Kazakhstan, Kyrgyzstan, Turkmenistan, Georgia, Azerbaijan and Armenia (Bobkov et al., 1998; Bobkova, 2013; Saad et al., 2006b). The geographic origin of the $A_{FSU}$ PWID-epidemic was in Odessa and the tMRCA was approximately in 1993 (Diez-Fuertes et al., 2015). Moreover, the origin of the $A_{FSU}$ PWID-epidemic strain has been recently identified in Democratic Republic of Congo (DRC), and the upper limit of the dispersal time for the ancestral strain in 1970 (Diez-Fuertes et al., 2015). Besides the $A_{FSU}$ that dominates in FSU countries, subtype B transmissions have been described at low prevalence and are mainly associated with MSM, probably associated with the Western European epidemic (Bobkova, 2013).

The epidemiology of the HIV-1 epidemic in the Baltic states of Estonia, Latvia and Lithuania is similar to the neighbouring countries of Belarus and Russia, where PWID transmission networks contribute significantly to the spread of the epidemic (Avi et al., 2014; Lai et al., 2014). In 2000, the rapid spread of HIV-1 amongst PWID intensified; when an outbreak was noted amongst PWID in Estonia’s Eastern regions (Adojaan et al., 2005). This was a large outbreak resulted in a prevalence to around 1000 HIV-1 infected per million, consisting the highest in the European Union (Avi et al., 2014). It was recognized from the early start that this outbreak involved a recombinant HIV-1 subtype which differs from the
predominant A clade circulating in neighbouring countries, called CRF06_cpx (Adojaan et al., 2005). The Estonian CRF06_cpx strain was likely originated in Africa (Zetterberg et al., 2004), and although it has been reported in other European countries, it is a minority variant in all studies performed outside Estonia. Conversely, subtype A dominates in Latvia and Lithuania, where the HIV-1 epidemic is otherwise very similar and driven by injecting drug use (Andrews et al., 2013; Popa et al., 2013). More recently CRF02_AG has been increasingly detected amongst PWID in Russia and is probably linked with the CRF02_AG epidemics in Central Asia (Bobkova, 2013).

The origin of non-predominant HIV-1 clades in Europe can be described according to the following three patterns: i) Cross-continental transmissions: Overall in Western Europe non-B clades have mostly been associated with immigration and heterosexual intercourse. This pattern has been remained consistent since the earliest stage of the epidemic. Non-B transmissions were detected at different proportions amongst individuals with a non-African origin in Western Europe. South Eastern Asia and South America are also a source of non-B clades in Western European countries, but to a lesser extent than Africa. ii) Cross-border infections across Western Europe: Non-B infections circulating across Western European countries such as subtype G infections in Luxemburg and Spain which have originated in Portugal; CRF14_BG in Spain, which also originated in Portugal; CRF01_AE epidemic among PWID in Sweden, which is regarded as a spill over from Finland and the F1 clade circulating in North-West Spain and other European countries iii) Cross-border transmissions across European areas: In this group there are examples of non-B transmissions in Eastern Europe which originated from Western Europe, such as the CRF14_BG from the Iberian Peninsula (Portugal and/or Spain) and the \text{A}_{FSU} from Eastern Europe transmissions in Central and Western Europe introduced.
border spill overs have also been discovered between Central Asia and Eastern Europe (Russian Federation).
6. HIV-1 subtypes in immigrants

Population movements including migration from the African and Asian continents have transformed European countries over the past two decades and have been linked to several infectious disease outbreaks, including local HIV-1 epidemics (Kentikelenis et al., 2015). In 2013, epidemiological reports showed 39.9% of new HIV-1 infections were in immigrants. A large proportion of new HIV-1 infections were among immigrants from sub-Saharan Africa (54.3%), Latin America (12.2%), Western Europe (9.5%), and Central Europe (6%) ((ECDC), 2013). Sub-Saharan Africa was identified as the origin of 13.8% of all HIV-1 diagnoses in the EU/EEA, 35.0% of heterosexually acquired infections and 38.3% of mother-to-child transmissions (MTCT), as shown by studies from Spain and the United Kingdom (Monge-Maillo et al., 2009).

Based on data from numerous molecular epidemiological studies reviewed above, non-B transmissions in Western Europe have been associated with immigrants from sub-Saharan Africa and to a lesser extent from South East Asia, South America and Eastern Europe. Phylogenetic analyses revealed that different proportions of regional transmissions occur amongst immigrants, as for example in Switzerland, where this proportion ranged between 16% and 28% for several non-B clades (von Wyl et al., 2011). Lai et al. showed that individuals from the generalised epidemic were less likely to belong within local clusters than individuals from South America and Italy (Lai et al., 2014). Notably, subtype B transmissions were found at 23% and 3% of Africans and immigrants from sub-Saharan Africa living in France and Spain, respectively, suggesting that these transmissions could have possibly occurred in Europe (Chaix et al., 2013; Rivas et al., 2013).

Based on the fact that immigrants have mostly been infected with non-B strains, a valid hypothesis is that they were infected, at least at some proportions, before migrating and...
therefore they could provide the main source of divergent strains in Europe. This picture is consistent in Western Europe where non-Bs predominate among non-Europeans. Of course there are exceptions, such as Greece and Portugal, where non-B infections have been spread within local sexual networks (Thomson and Najera, 2007). Notably, in Central Europe non-B clades are mainly linked with heterosexual route of transmission and not with non-European origin. On the other hand, non-Bs have a distinct pattern of epidemic spread in Eastern European countries, for instance in FSU countries are associated with regional dispersal in PWID and in Russian federation with heterosexual transmissions within local immigrants’ sexual networks.

It is clear that migrating populations have played an important role in shaping the genetic heterogeneity of the HIV-1 epidemic in Western Europe. This finding is consistent with socioeconomic factors indicating higher migratory rates towards Western European countries rather than FSU countries and Russia.
To conclude, the complexity of the European HIV-1 epidemic has been increasing in Western and Central Europe during recent years. The higher proportions of non-B clades and their increasing prevalence across Western and Central Europe reflect this. Conversely, in FSU countries, expect Russia, the epidemic is less complex where subtype A (A_{FSU}) still predominates across different areas. Our review of numerous recently published studies and updated database queries suggest that the distribution of different clades greatly differs across Western and Central Europe, where the most prevalent non-B clades are A, C, F and G and the CRFs 01\_AE and 02\_AG. The introduction of divergent strains occurs mostly through mobility from sub-Saharan Africa and circulation of these strains is more frequent amongst immigrants, local transmission networks, such as heterosexual, MSM and PWID (Abecasis et al., 2013). Although the origin of HIV-1 transmissions in immigrants warrants further investigation, preliminary analyses of the already published studies suggests that most of those coming from Africa, mainly sub-Saharan Africa, are infected in their country of origin (pre-migration), rather than in their hosting European country (post-migration).

Acknowledgments

AB is funded through the IKY Fellowships of Excellence for Postdoctoral Research in Greece – Siemens Program. The study was partly supported by the Hellenic Society for the study of AIDS and STDs. The authors wish to thank Ms. Rachael Jones who assisted in the proof-reading of the manuscript.
Figure Legends

Figure 1: Phylogenetic tree of full-length genomic sequences from SIV infecting different monkey species (A) and HIV-1/SIVcpz (B) shown in different colours. HIV-1 and HIV-2 groups are shown in red and stars indicate cross-species transmission events. Full-length sequence alignments were downloaded from the HIV Los Alamos sequence database.

Figure 2: HIV-1 and HIV-2 classification scheme. In addition to groups, clades and recombinants described in this figure, there have been named several monophyletic clades of viruses circulating in specific geographic regions (e.g. the A_{FSU} clade including sequences within subtype A circulating in Russia and other Former Soviet Union (FSU) countries).

Figure 3: Map of the most prevalent non-B subtypes and recombinants across Europe as explained in the colour legend. The prevalence of subtype B for each country is drawn on grey scale. Information about the prevalence of HIV-1 clades was based on a query from the Los Alamos HIV sequence database for sequences sampled later than 2005 using a single sequence per patient.
Reference list


Enhanced HIV-1 surveillance using molecular epidemiology to study and monitor HIV-1 outbreaks among intravenous drug users (IDUs) in Athens and Bucharest. Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases 35, 109-121.


Peeters, M., Gueye, A., Mboup, S., Bibollet-Ruche, F., Ekaza, E., Mulanga, C., Ouedrago, R., Gandji, R., Mpele, P., Dibanga, G., Koumare, B., Saidou, M., Esu-Williams, E., Lombart, J.P., Badombena, W., Luo,


HIV-1
- Group N
- Group P
- Group O
  - Group M
    - Unclassified strains
    - Recombinants
      - CRFs
      - URFs

Subtypes (A-D, F-H, J and K)
Sub-subtypes (A1, A2, F1 and F2)

HIV-2
- Unclassified strains
- Recombinant
  - CRF

Groups (A-H)