



RECOMBINATION ANALYSIS FOR SUBTYPING UNCLASSIFIED HIV-1 SEQUENCES FROM ROMANIA

Paraschiv Simona¹, Batan Ionelia¹, Banica Leontina¹, Fratila Mihaela¹,
Doana Angelica¹, Vizitiu I.¹ and Otelea D.¹

1. *Molecular Diagnostics Laboratory, 'Prof. Dr. Matei Bals' National Institute for Infectious Diseases*

Abstract. The human immunodeficiency virus (HIV) has a high rate of evolution generated by mutations and recombinations in association with the high replication rate of this virus into the susceptible cells. The diversity of HIV can be easily observed in HIV-1 group M which comprises 9 subtypes, more than 50 inter-subtype circulating recombinant forms (CRFs) and several unique recombinant forms (URFs). Subtype distribution in Europe is geographically influenced; subtype B is highly prevalent in the Western part of the continent and subtype A is mainly responsible for the infections in the Eastern countries. From the beginning of the epidemic in Romania, a particular subtype has been reported as being highly prevalent, the F1 subtype. Previous data suggested that the HIV-1 epidemiology in Romania has changed over time, other subtypes than F1 (subtype B, C and A1) being more and more encountered. 32 of 2132 sequences corresponding to HIV-1 strains isolated during 2003 and 2011 in the reference laboratory from the National Institute for Infectious Diseases 'Matei Bals' were unassigned for the HIV-1 subtype. We have analysed these sequences using four different recombination algorithms. Half of them were classified as being CRFs (7 were CRF14_BG and 8 were CRF02_AG) and subtype F1 (1 sequence). The other 16 sequences were recombinants with more complex patterns of recombination. Therefore, we have improved the figure of HIV-1 subtype distribution in Romania, showing how using of more different bioinformatic tools can improve the subtype assignment.

Keywords: HIV-1 subtype analysis, recombination break point, bioinformatic tools

Introduction

HIV is known as one of the viruses with the highest evolution rate among all the species described so far (Gojobori T et al, 1994, Onafuwa-Nuga and Telesnitsky, 2009). The genetic diversity of HIV-1 strains is generated by complex mechanisms leading to substitutions, insertions, deletions and recombinations; this has important implications in diagnosis, drug treatment and vaccine development. From the beginning of the epidemic in humans, which dates back to the beginning of the 20th century from simian immunodeficiency virus (SIVcpz), HIV-1 strains have rapidly evolved and are currently classified in 4 groups: M, O, N and P. The great diversity of HIV-1 strains can be easily observed in group M which has nine subtypes, more than fifty inter-subtype

circulating recombinant forms (CRFs) and several unique recombinant forms (URFs) (Paraskevis D and Hatzakis A, 1999, <http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>).

Resistance genotype testing of HIV strains from patients' plasma is currently performed in clinical settings to detect the mutations responsible for resistance to antiretroviral treatment. The second advantage of resistance genotyping is that it allows subtyping of the infective strain, making possible epidemiological surveillance in different populations and/or geographical regions.

HIV-1 epidemiology in Europe varies mainly geographically: in Western and Central Europe the most common clade is subtype B, being identified especially in men who have sex with men (MSM) and intravenous drug users (IVDUs). (Strathdee et al., 2010, Platt et al., 2011). In the Eastern part of Europe subtype A is the most encountered one, being reported in countries such as Russia, Belarus, Ukraine, Bulgaria and the former Yugoslavia (Saad et al., 2006, Lazouskaya et al., 2005, Stanojevic M

Simona Paraschiv

1 Calistrat Grozovici, Str., sector 2
021105 Bucharest, Romania
email: mona@mateibals.ro

et al, 2012). From the beginning of the epidemic in Romania, a particular subtype, F1, has been identified as highly prevalent among the infected patients. Previous studies showed that the HIV-1 epidemiology in Romania has changed over time: subtype F1 that was involved in the paediatric and adult cases at the beginning of the epidemic (Hersh et al., 1993, Op de Coul et al., 2000) is now replaced by other subtypes (Paraschiv et al, 2008). Superinfections in adults can lead to recombinant forms that can be difficult to identify by the currently used bioinformatic tools. Therefore, the purpose of this study was to analyse the unclassified sequences coming from HIV-1 strains genotyped in the last nine years by our group, using different specific recombination algorithms.

Methods

The infective strains originating from 2132 HIV-1 infected patients, originating from all the geographic regions of Romania were genotyped during 2003-2011 in the reference laboratory from the National Institute for Infectious Diseases 'Matei Bals', as part of drug resistance testing in Romania. The plasma samples were obtained from EDTA-whole blood (centrifugation at 3000 rpm for 15 min) and stored at -80°C prior to genotyping testing. HIV genotyping was performed using the Sanger sequencing method in the commercial kit Viroseq™ HIV-1 Genotyping System (Celera Diagnostics, Alameda, CA). HIV-1 RNA molecules were extracted from plasma either manually, by using the Sample extraction module of the Viroseq kit, or by using the automated NucleiSens easyMAG nucleic acid extraction system (BioMerieux) and specific magnetic extraction reagents, according to the manufacturer's instructions. A 1.8 kb RT-PCR product, representing the first part the *pol* gene (coding for protease and 2/3 of reverse transcriptase), was purified and sequenced bidirectionally on the ABI Prism 3100-Avant Genetic Analyzer (Applied Biosystems) or the 3500 Genetic Analyser (Applied Biosystems). The raw analysis of the sequences was made using Sequencing Analysis Software Version 3.7 (Applied Biosystems); they were then assembled with ViroSeq 2.5/2.7/2.8 HIV-1 Genotyping System Software (Celera Diagnostics, Alameda, CA) generating a consensus sequence of about 1300bp long. The sequences in Fasta format were used for subtyping purposes by comparing them with reference sequences of all known HIV-1 subtypes using the publicly available algorithm REGA HIV-1&2 Automated subtyping tool version 2.0 (www.jose.med.kuleuven.be/genotypetool/html/indexhiv.html).

Recombination analysis

A number of 32 sequences were unassigned for

HIV-1 subtype or CRFs by the REGA algorithm. We have further analyzed these sequences with four different recombination algorithms. We have performed on-line submission for three of them: jpHMM web-based algorithm (http://jphmm.gobics.de/submission_hiv.html), NCBI Genotyping tool (<http://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi>) and Scueal recombination detection algorithm (http://www.datamonkey.org/dataupload_scueal.php). We have also used Simplot v3.5.1 software (<http://sray.med.som.jhmi.edu/SCSoftware/simplot/>) with sliding window: 400-nt, T:t ratio=2.0, model of evolution: Kimura two-parameter, bootstrap: 1000 replicates to identify the recombination breakpoints. To verify the robustness of our findings we used other reference datasets that were downloaded from LosAlamos HIV-1 sequence database. We used HIV sequence Alignments tool for Subtype Reference Alignments and we selected: subtype reference for alignment type, year 2008, genomic region POL, all M group (A-K + recombinants) for subtype option.

All sequences were screened for hypermutation using the Hypermut 2.0 (<http://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html>).

Results and discussions

The subtype analysis of the HIV-1 strains genotyped during 2003-2011 in Romania showed, as expected, that the majority of Romanian patients were infected with subtype F1 strains (n=1936, 91%). The results are in concordance with previous studies showing that subtype F1 is still the most prevalent clade in Romania being mainly present at nosocomial infected children (Veras MN, 2012). However, in recent years the number of strains other than F1 has increased; subtypes B, C, A1 and CRFs were especially seen in recently diagnosed/infected patients (Florea et al, 2010). In our analysis, subtype B has encountered 4 % (94 patients) of the study population, while subtype C was seen in 54 patients (3%), but the proportion of these subtypes in newly diagnosed patients is higher (data not shown). Subtype B was previously described to be mainly associated with MSM transmission (Paraschiv S et al., 2012), while subtype C was associated with heterosexual contact (Paraschiv S et al, 2011). A number of 32 (1.5%) of the analysed HIV-1 sequences could not be assigned for a specific subtype or CRF using the algorithm REGA HIV-1&2, suggesting possible recombination breakpoints in the analyzed region. Currently, the algorithm publicly available from Rega is widely used for subtyping HIV strains, but other tools have been described to assess the subtype of the infective viral strain as well. Simplot software is a special designed application to identify recom-

bination events. Another tool frequently used for subtyping the HIV-1 sequences is available on the NCBI site, the NCBI Genotyping tool. We used these tools and other two algorithms in order to characterize in more details these 32 HIV-1 unassigned sequences isolated in Romania.

The results of this analysis by these four distinct recombination algorithms were similar in most of the cases. There were situations when discordant results were obtained, especially when analysed with jpHMM and Scueal algorithms. Recombination events were observed in all the sequences studied in the analysed region (about 1300 bp) except for

recombination events between two subtypes/CRFs. It is the case of sequence 1247vb_2008 (CRF06 and CRF02), 1250vb_2008 (F1 and B), 385_2011 (CRF12 and B), 4204_2011 (CRF14 and F1), 455_2009 (F1 and CRF05), 5162_2008 (B and F1), 5197_2008 (B and CRF15). For some other sequences, the recombination pattern is more complex. The sequence 111_2008 as well as 2966_2010 resulted from recombination between CRF06_cpx, CRF02_AG and CRF06_cpx. The sequence 2579_2005 is a mosaic formed by CRF02_AG, subtype A1, subtype G, CRF02_AG and subtype A1. 2870_2009 resulted from recombination of CRF02_AG with

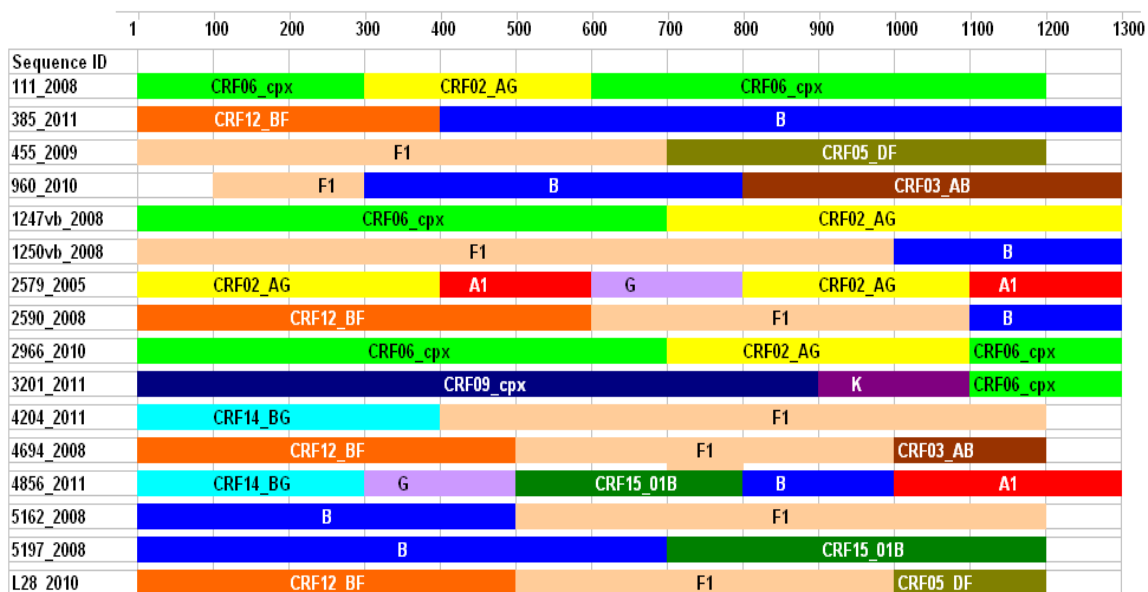


Figure 1. Recombination analysis results and break point recombination positions for the analysed genomic region (the first 1300bp of HIV pol gene)

one sequence (subtype F1). The best concordance between recombination results was obtained using NCBI Genotyping tool and Simplot software. Half of the analysed sequences (16) were assigned by these two algorithms as belonging to a specific CRF or subtype F1. CRF14_BG was assigned in 7 cases and 8 sequences were classified as CRF02_AG. Most of these strains were isolated from patients who were diagnosed with HIV infection during the last 5 years. The rest of the sequences (16) are recombinants with more complex patterns: the polymerase gene is a mosaic resulted from a recombination event between certain subtypes and/or CRFs.

Subtype assortments as well as the break point recombination positions are presented in figure 1. To simplify the picture we have presented in figure 1 only the recombination results indicated by the NCBI genotyping tool. As it can be observed from the figure, some sequences are the results of

CRF09_cpx and CRF02_AG. This strain was isolated from a female patient originated from Congo. The sequence 3201_2011 is a recombination between CRF09_cpx with K and CRF06_cpx. Another interesting recombination form can be seen in sequence 4856_2011 a mosaic between CRF14_BG, subtype G, CRF15_01B, subtype B and subtype A1, corresponding to the viral strain isolated from a patient with high risk of super infection.

Using specialized algorithms to identify recombination events in the HIV-1 polymerase gene we could improve the subtype distribution picture in Romanian patients - the analysis on the strains genotyped between 2003 and 2011 presented in figure 2. Of 32 unassigned sequences we could subtype half of the sequences as being CRFs or F1 subtype, while the rest of 16 were recombinants consisting in a mixture of two or more different subtypes or CRFs. The number of patients infected with CRFs

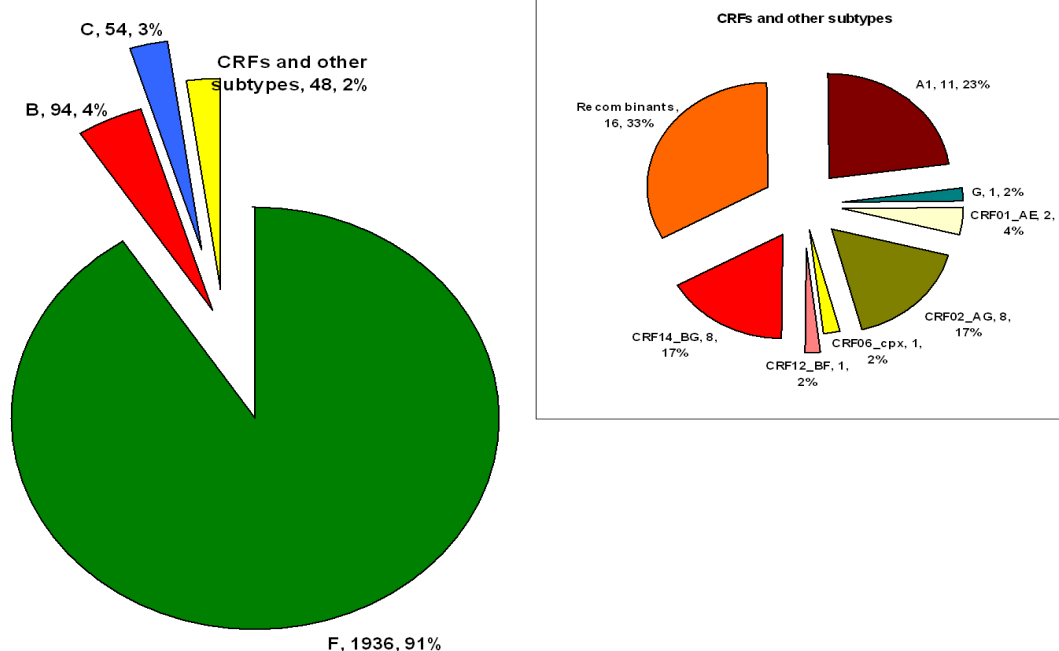


Figure 2. Subtype distribution in HIV-1 infected patients in Romania
The corresponding strains were genotyped during 2003-2011

reached 20 and other 16 are infected with other recombinant forms. Further investigations on these recombinants, such as full genome sequencing followed by subtyping analysis, are needed to classify the strains as unique recombinant forms.

Conclusion

The analysis of unclassified HIV-1 pol sequences using specialized bioinformatic tools for detection of recombination improved the subtype assignment and the picture of HIV-1 subtype distribution in Romanian patients.

Acknowledgements. This paper was supported in part by CNCSIS-UEFISCSU, project number PN II-RU PD 157/2010 and by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU/89/1.5/S/64109.

References

1. Gojobori T, Yamaguchi Y, Ikeyo K, Mizokami M, 1994. Evolution of pathogenic viruses with special reference to the rates of synonymous and nonsynonymous substitutions. *Jpn J Genet.*;69(5):481-8.
2. Onafuwa-Nuga A and Telesnitsky A, 2009. The Remarkable Frequency of Human Immunodeficiency Virus Type 1 Genetic Recombination. *Microbiol. Mol. Biol. Rev.*;73:451-480
3. Paraskevis D, Hatzakis A, 1999. Molecular epidemiology of HIV-1 infection. *AIDS Rev*; 1: 238-249.
4. Strathdee SA, Hallett TB, Bobrova N, Rhodes T, Booth R, Abdool R, Hankins CA, 2010. HIV and risk environment for injecting drug users: the past, present, and future. *Lancet*; 376(9737):268-84.
5. Platt L, Grenfell P, Bonell C, Creighton S, Wellings K, Parry J, Rhodes T, 2011. Risk of sexually transmitted infections and violence among indoor-working female sex workers in London: the effect of migration from Eastern Europe. *Sex Transm Infect.*; 87(5):377-84.
6. Saad MD, Shcherbinskaya AM, Nadai Y, Kruglov YV, Antonenko SV, Lyullchuk MG, et al., 2006. Molecular epidemiology of HIV Type 1 in Ukraine: birthplace of an epidemic. *AIDS Res Hum Retroviruses*. 2006; 22(8):709-14.
7. Lazouskaya NV, Eremin VE, Adema KW, Gasich EL, Baan E, Lukashov VV, 2005. The HIV type 1 epidemic in Belarus: predominance of Eastern European subtype A strains and circulation of subtype B viruses. *AIDS Res Hum Retroviruses*; 21(9):830-3.
8. Hersh BS, Popovici, F, Jezek, Z, Satten GA, Apetrei RC, Beldescu N et al., 1993. Risk factors for HIV infection among abandoned Romanian children. *AIDS*, 7(12):1617-1624.
9. Op de Coul E, Van der Burg R, Asjo B et al, 2000. Genetic evidence of multiple transmission of HIV-1 type 1 subtype F within Romania from adult blood donors to children. *AIDS Res Hum Retroviruses*; 16 (4): 327-336.
10. Paraschiv S, Otelea D, Tinischi M, 2008. Molecular epidemiology of HIV-1 in Romania: an update. *Reviews in Antiviral Therapy*, 2: 69-70 (abstract 34).

11. Florea D., Paraschiv S., Fratila M. and Otelea D., 2010. Human Immunodeficiency Virus F1 subtype is associated with late presentation in newly diagnosed Romanian patients. *Clinical Microbiol. and Infect., Supp 1: abstract 1213.*

12. Paraschiv S, Foley B, Otelea D, 2011. Diversity of HIV-1 subtype C strains isolated in Romania. *Infect Genet Evol.* 2011;11(2):270-5

13. Paraschiv S, Otelea D, Batan I, Baicus C, Magiorkinis G, Paraskevis D, 2012. Molecular typing of the recently expanding subtype B HIV-1 epidemic in Romania: evidence

for local spread among MSMs in Bucharest area. *Infect Genet Evol.*; 12(5):1052-7.

14. Stanojevic M, Alexiev I, Beshkov D, Gökengin D, Mezei M, Minarovits J, Otelea D, Paraschiv S, Poljak M, Zidovec-Lepej S, Paraskevis D, 2012. HIV-1 molecular epidemiology in the Balkans: a melting pot for high genetic diversity. *AIDS Rev.*; 14(1):28-36.

15. Veras NM, Paraschiv S, Tudor AM, Otelea D, Salemi M, 2012. Phylogentic analysis of F1 epidemic in Romania: tracing back the origin of the epidemic. *Reviews Antiviral Therapy Infect Dis*, 2:32 (abstract P_05).