High frequency of new recombinant forms in HIV-1 transmission networks demonstrated by full genome sequencing

Laura HEBBERECHT[1], Virginie MORTIER[1,] Kenny DAUWE[1], Marlies SCHAUVLIEGE[1], Delfien STAELENS[1], Els DEMECHELEER[1], Karolien STOFFELS[2], Fien VANROYE[3], Marie-Luce DELFORGE[4], Ellen VANCUTSEM[5], Géraldine DESSILLY[6], Dolorès VAIRA[7], Kristel VAN LAETHEM[8 9], Chris VERHOFSTEDE[1*]

[1]Aids Reference Laboratory, Department of Diagnostic Sciences, Ghent University, Ghent, Belgium.

[2]Aids Reference Laboratory, Centre Hospitalier Universitaire St. Pierre, Brussels, Belgium.

[3]HIV/STD Reference Laboratory, Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium.

[4]Aids Reference Laboratory, Université Libre de Bruxelles, Brussels, Belgium.

[5]Aids Reference Laboratory, Vrije Universiteit Brussel , Brussels, Belgium.

[6]Aids Reference Laboratory, Medical Microbiology Unit, Université Catholique de Louvain, Brussels, Belgium

[7]Aids Reference Laboratory, Centre Hospitalier Universitaire de Liège, Liège, Belgium.

[8]Aids Reference Laboratory, University Hospital Leuven, Leuven, Belgium.

[9]Rega Institute for Medical Research, Department of Microbiology, Immunology and Transplantation, KU Leuven, Leuven, Belgium.

*Corresponding author: Chris Verhofstede (e-mail: chris.verhofstede@ugent.be)

'Declarations of interest: none'.

Abstract

The HIV-1 epidemic in Belgium is primarily driven by MSM. In this patient population subtype B predominates but an increasing presence of non-B subtypes has been reported. We aimed to define to what extent the increasing subtype heterogeneity in a high at risk population induces the formation and spread of new recombinant forms. The study focused on transmission networks that reflect the local transmission to an important extent. One hundred and five HIV-1 transmission clusters were identified after phylogenetic analysis of 2849 HIV-1 *pol* sequences generated for the purpose of baseline drug resistance testing between 2013 and 2017. Of these 105 clusters, 62 extended in size during the last two years and were therefore considered as representing ongoing transmission. These 62 clusters included 774 patients in total. From each cluster between 1 and 3 representative patients were selected for near full-length viral genome sequencing. In total, the full genome sequence of 101 patients was generated. Indications for the presence of a new recombinant form were found for 10 clusters. These 10 clusters represented 105 patients or 13.6% of the patients covered by the study. The findings clearly show that new recombinant strains highly contribute to local transmission, even in an epidemic that is largely MSM and subtype B driven. This is an evolution that needs to be monitored as reshuffling of genome fragments through recombination may influence the transmissibility of the virus and the pathology of the infection. In addition, important changes in the sequence of the viral genome may challenge the performance of tests used for diagnosis, patient monitoring and drug resistance analysis.

Keywords: HIV-1; New recombinant forms; Transmission clusters; Full genome sequencing; Subtyping

1. Introduction

HIV-1 has the ability to rapidly generate genetic diversity through mutation and recombination. This resulted in the origination of 9 subtypes, several sub-subtypes, circulating recombinant forms (CRFs) and unique recombinant forms (URFs) within HIV-1 group M. The number of CRFs is constantly growing, at least 101 have been registered today (https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html).

Since the beginning of the HIV-1 epidemic in Western and Central Europe, subtype B infections dominated the onward transmission in these regions (Gilbert et al., 2007), whereas non-B subtypes and CRF infections mainly resulted from migration (Delgado et al., 2019; Tamalet et al., 2018). This led to different numbers in prevalence and onward spread of non-B subtypes between geographical regions due to specific local circumstances (Beloukas et al., 2016). In Belgium, non-B subtypes were introduced early, mainly through migration from Sub-Saharan Africa (Fransen et al., 1996; Snoeck et al., 2004; Vinken et al., 2019). Despite this early introduction of non-B variants, the epidemic remained dichotomized. Men-having-sex-with men (MSM), mainly of Belgian or European descent, account for 49% of the diagnoses and the majority are subtype B infections. Heterosexuals, who account for the other half, are mainly Sub-Saharan Africans infected with non-B strains (Van Beckhoven et al., 2015). The vast majority of infections in MSM are acquired locally, while most of the infections in Sub-Saharan Africans are imported infections acquired abroad (Verhofstede et al., 2019). An increasing number of non-B subtypes in the high at risk MSM population has however been observed, with especially subtype F1 and CRF02_AG spreading among the native Belgian population (Dauwe et al., 2015; Vinken et al., 2019). Co-circulation of different subtypes in MSM may increase

the chance for dual infection with different subtypes and constitute a melting pot where recombination can swiftly and substantially change viral traits. This may have important implications on epidemiology, pathogenicity, diagnosis and immune and therapy response. In support, we previously showed that the prevalence of co- or superinfection in Belgian MSM is high, estimated to be at least 6% (Hebberecht et al., 2018). The eventual generation and transmission of new recombinants therefore requires close monitoring.

Phylogenetic analysis of HIV-1 *pol* sequences generated through standard of care baseline drug resistance analysis can be used to define the evolutionary relationship between viruses and reconstitute transmission clusters. In a previous study, we identified individuals who were most likely infected locally and recently using information on the stage of the infection at diagnosis and the position of their viral sequence in the phylogenetic tree (Verhofstede et al., 2019). This clearly revealed an association between being part of a molecular transmission cluster and being infected locally. The present study intends to characterize the viral strains that are spread in these local transmission clusters. For that purpose active molecular transmission clusters were identified and representative patients were selected for near full-length sequencing of the viral genome. A total of 62 transmission clusters were examined and the viral genome of 101 patients was characterized. The results confirmed the presence and onward transmission of several new recombinants.

2. Materials and methods

2.1. Ethical approval

The study was approved by the Ethical Committee of all participating laboratories, with the Ethical Committee of Ghent University Hospital as the leading investigator center (reference number 2019/0278). Depending on the local practices in the different participating institutions, patients signed informed consent for participation in the study or were included on the base of an opt-out procedure. All samples were pseudonymized before processing with no tracing back to the original identification at any time.

2.2. Phylogenetics

Phylogenetic analysis of the Belgian HIV-1 epidemic was performed using the partial HIV-1 *pol* sequences, obtained through standard baseline drug resistance testing, from 2849 patients newly diagnosed with HIV-1 infection in Belgium between 2013 and 2017. Transmission clusters were identified as described before (Verhofstede et al., 2019). Briefly, a maximum likelihood (ML) phylogenetic tree of all the available *pol* sequences was constructed using PhyML 3.0 with automatic selection of the best fit evolutionary model of DNA substitution (GTR + G + I) using the Akaike information criterion (AIC). Branch support was obtained by approximate likelihood-ratio test (aLRT, SH-like). Sequences linked with at least one other sequence with an LRT of ≥0.97 and mean pairwise distance of ≤0.015, were considered as genetically clustered. All clusters of at least 4 members that increased in size with at least 2 members over the last years of sample collection (2016 and 2017), were considered as active transmission clusters and included in the study.

2.3. Subtyping

HIV-1 subtyping was performed based on the available *pol* sequence using COMET HIV-1 (Struck et al., 2014) and Rega v3.0 (Pineda-Pena et al., 2013). The subtype was allocated in case of a concordant result and considered as “undefined” (UD) otherwise. Subtyping based on the different genome regions obtained after full-length genome sequencing was done using the subtyping tool of Smartgene IDNS (Smartgene, Zug, Switzerland) and COMET HIV-1 (Struck et al., 2014). Transmitted drug resistance mutations were identified using Smartgene IDNS (Smartgene, Zug, Switzerland).

2.4. Characterization of the study population

Data on patient origin and mode of transmission, if available, were collected within the framework of mandatory HIV surveillance and linked to the individual patients using a pseudonymized identifier. For mode of transmission, the categories were MSM, heterosexual, and ‘other’ (including intravenous drug use, blood transfusion, and mother-to-child transmission). Between 1 and 3 representative patients from each active cluster were selected for full genome sequencing with exception of one of the subtype F1 clusters of which 5 patients were tested. For those clusters with indications for transmission of a new recombinant, one or two additional patients were selected to confirm the observations. Care was taken to avoid selection of very closely linked patients and to select preferentially patients diagnosed in different HIV reference centers that are geographically spread to reduce the risk of selecting directly linked individuals. If possible, preference was given to patients diagnosed early after infection to minimize the intra-patient variability.

2.5. Near full-length genome sequencing

Full genome sequencing was performed using an in-house protocol described before (Hebberecht et al., 2019). Briefly, the HIV-1 genome from nucleotide 551 to nucleotide 9582 was covered by two overlapping amplicons, a 5' amplicon containing the region between nucleotide 551 and 5061 and a 3' amplicon containing the region between nucleotide 4956 and 9582. Both amplicons were generated by nested PCR. Sequencing was performed using 36 primers in total enabling to cover the whole amplified region. Sequences were proofread and manually edited if needed with Smartgene IDNS (Smartgene, Zug, Switzerland) and assembled and aligned in BioEdit (Hall, 1999). Each individual full-length genome sequence was submitted to jpHMM (jumping profile Hidden Markov Model) to visualize the subtype structure and detect potential recombination (Schultz et al., 2012). A maximum likelihood phylogenetic tree of the generated full genome sequences was constructed as described above.

All 101 full genome sequences were submitted to GenBank (GenBank numbers MT417737-MT417772, MN989924-MN989927, MN485971-MN485981, MN485983-MN485989, MN485991-MN485996, MN486000-MN486008, MN486010, MN486012, MN486013, MN486015-MN486019, MN486021-MN486038, MN486040 and MN486041). A multiple alignment of these sequences and subtype reference sequences is available on request.

3. Results

3.1. Characteristics of the study population

The phylogenetic analysis of HIV-1 partial *pol* sequences from 2849 patients newly diagnosed with HIV-1 in Belgium between 2013 and 2017 revealed the presence of 105 clusters that fulfilled the criteria for being considered as a transmission cluster. Fifty-nine percent of these clusters (62/105) were considered as eligible for the current study because they were composed of at least 4 members and they expanded with at least 2 new members during 2016 and 2017. These 62 active clusters represented a total of 775 patients (27%).

Twenty-seven clusters were classified as MSM-only (44%), one was classified as heterosexual-only (2%) and 32 (52%) consisted of MSM as well as members reporting other transmission risks. For 14 of the latter however, MSM was reported by the majority of the members and 12 comprised only men. Information on transmission route was missing for two clusters.

Based upon the subtype of the partial *pol* gene, 44 clusters were classified as subtype B (71%) and 13 as non-B (21%). The latter included four CRF02_AG clusters, three subtype A clusters, three F1 clusters, two CRF01_AE clusters and one subtype C cluster. For four clusters (6%) the results of the subtyping tools suggested presence of a new recombinant, composed respectively of fragments of B and F1, F1 and D, A and CRF03_AB and B and CRF02_AG. For one cluster the *pol* subtype remained undefined (2%).

3.2. Results of the full genome sequencing

Full genome sequencing was enabled by the generation of two PCR amplicons, each spanning half of the HIV genome. Overall, generation of the 5' amplicon failed for seven samples (6%; 7/119), generation of the 3' amplicon for three (3%) and generation of both

amplifications failed for four (3%). Subsequent sequencing was only initiated for the 105 samples for which both amplicons were successfully obtained. Coverage of the whole region with sequences of sufficient quality failed for four, leaving 101 individual full genome sequences (85%), at least one from each of the 62 clusters, for further analysis.

Subtype analysis of the full genome sequences showed that a pure subtype B infection was responsible for 43 of the 62 selected clusters (69%). Additionally, clusters were classified as resulting from a pure CRF02_AG infection (n=3), pure subtype F1 infection (n=3), pure subtype A infection (n=1), pure subtype C infection (n=1) and pure CRF01_AE infection (n=1). For these 52 clusters the subtype classification based on the whole genome confirmed the subtype classification based on the *pol* sequence.

For 10 clusters (16%), there are indications that the infection is caused by a recombinant form. The jpHMM profiles are shown in **Figure 1**. For four, the presence of a recombinant was also concluded after subtype attributed based on the partial *pol* sequence. For one, subtyping based on *pol* was inconclusive which also suggested the presence of a rare variant.

Cluster ID	#Patients	#FGS	Gender	Subtype PR/RT	Subtype FGS	Risk
42	18	3	M	B/CRF02_AG	B/CRF02_AG	MSM
41	17	2	M/F	D/F1	D/F1	Mixed
43	13	2	M	A	A/C	Mixed
23	12	2	M/F	A	A/B	Mixed
33	11	2	M	B/F1	B/F1	MSM

21	10	3	M	A/CRF03_AB	A/CRF03_AB	Mixed
3	9	2	M	CRF02_AG	G/CRF02_AG	MSM
17	6	2	M	B	B/CRF50_A1D	Mixed
45	5	2	M	CRF01_AE	B/CRF01_AE	MSM
59	4	2	M/F	UD	F2/CRF22_01A1	Mixed

Table 1. Overview of the 10 clusters with new recombinant forms. #FGS = number of samples subjected to near full-length genome sequencing; M= Male; F= Female; Risk= Transmission route

In all of the latter, a breakpoint of recombination was observed in the *pol* region (**Fig 1**). The D/F1 recombinant (cluster ID 41) showed a region in the *envelope* gene as well as in the *protease* gene that was classified as subtype D. This was confirmed by Smartgene and COMET but was not seen when subjecting the whole genome sequence to jpHMM. When visualizing only the *protease* region in jpHMM however, a part was clearly identified as subtype D. For the G/CRF02_AG recombinant, jpHMM identified a small part of the accessory genes as CRF01_AE. Subtyping of this region with the COMET and REGA subtyping tools however assigned this region as CRF02_AG. For the five recombinants without a breakpoint in *pol*, breakpoints were located in *env* or in the part of *pol* that is not routinely sequenced.

A phylogenetic tree representing the 101 near full-length sequences and reference sequences of the subtypes most prevalent in our populationis shown in **Figure 2**.

3.3. Transmission risk and patient origin

Of the 43 subtype B clusters, 18 were MSM-only (42%), 22 were mixed transmission risk clusters, one was heterosexual-only and for two the transmission route was unknown. Of the 9 pure non-B clusters, 5 (56%) were MSM-only (two CRF02_AG, one F1, one A and one CRF_AE) and four were mixed transmission risk clusters (two F1, one CRF02_AG and one C). Of the 10 clusters in which a new recombinant was identified, four (40%) were MSM-only (clusters B/CRF02_AG, B/F1, G/CRF02_AG and B/CRF01_AE) and six were of mixed transmission risk (A/B, A/C, A/CRF03_AB, D/F1, B/CRF50_A1D and F2/CRF22_01A1) (**Table 1**).

Data on patient origin were incomplete but the presence of at least one member of non-European origin could be confirmed in 39 clusters including 6 of the 10 clusters with a new recombinant.

3.4. Drug resistance

Transmission of virus with mutations associated with drug resistance was documented in 11 subtype B clusters, 2 subtype F1 clusters and the D/F1 recombinant cluster. In the majority of these clusters only minor mutations or mutations at polymorphic positions were detected, including the integrase mutations E157Q (in 4 subtype B clusters, representing 40 patients) and T97A (D/F1 cluster; 17 patients) and the NNRTI mutations A98G (one large subtype F cluster; 147 patients), E138A (3 clusters; 18 patients) and V106I (one subtype B cluster; 4 patients). Major drug resistance was observed in 2 subtype B clusters (representing 52 patients) with a revertant mutation at position 215 of the RT gene and 2 subtype B and one subtype F1 cluster with the NNRTI mutation K103N (representing 63 patients).

4. Discussion

Near full-length HIV-1 genome sequencing of representative patients selected from 62 active HIV-1 transmission clusters revealed the frequent presence of new intersubtype recombinants. Ten (16%) of 62 active Belgian transmission clusters resulted from the transmission of a new recombinant virus. These 10 clusters comprised 105 patients or 13.5% of all patients in the 62 transmission clusters analyzed. In comparison, subtype B, the most prevalent subtype, was isolated from 43 clusters representing 470 (60.7%) patients, followed by CRF02_AG (3 clusters, 16 patients (2.1%)), subtype F1 (3 clusters, 165 patients (21.3%)), A (1 cluster, 8 patients (1.0%)), CRF01_AE (1 cluster, 6 patients (0.8%)) and C (1 cluster, 5 patients (0.6%)). The detection of new intersubtype recombinants was not completely unexpected. We knew from previous work that non-B subtypes were increasingly found in the local MSM population that is traditionally dominated by subtype B (10) and we previously demonstrated the occurrence of co- or superinfection in this population (11). Moreover, in several of the MSM transmission clusters in Belgium at least one of the members originated from abroad (Dauwe et al., 2015; Hebberecht et al., 2018; Verhofstede et al., 2018), potentially increasing the chance for subtype mixture in this high at risk population thereby providing the ideal conditions for the generation of intersubtype recombinants.

Comparison of our findings with the results of other studies is hampered by the fact that in most studies the characterization of circulating HIV variants is based on short sequences of a single region in the viral genome. This approach most likely results in an

underestimation of the number of recombinants. In our study, only 4 of the 10 recombinant forms were identified based on the *pol* sequence generated for the purpose of drug resistance testing.

Using HIV *pol* sequences, Lai et al. evaluated the prevalence and distribution of non-B subtypes in a large HIV-1-infected cohort in Italy (Lai et al., 2010). They reported an overall prevalence of non-B clades of 11.4% (417 of 3670 patients). Among these 417 non-B infections, a new recombinant form was found in 39 (9%). In a follow up study, the same investigators focused on the further spread and evolution of one of these recombinants, the CRF60_BCD. They were able to document novel, second generation recombinants (Lai et al., 2019) supporting the belief that the probability for generation of recombinants is high and the impact thereof on the virus variability considerable. Dunn et al. found a frequency of 9.9% novel recombinants among key populations living with HIV-1 in the UK (Dunn et al., 2014). Here too, the analysis was restricted to the *pol* sequence. A study in Cameroon, where CRF02_AG is the predominant strain, revealed 18 (26%) new recombinants in a selection of 73 patients. In this study, partial sequences of *gag*, *pol* and the *env gp41* were analyzed (Ragupathy et al., 2011).

Our recently developed protocol for full genome sequencing (Hebberecht et al., 2019) is sensitive and reliable for a large variety of subtypes and very well suited for a more intensive investigation of recombinant generation. A challenge however is the reliable assignment of the subtype of these long sequences. HIV-1 subtyping based on the *pol* region has been well validated and several tools are available for this purpose (Rega, COMET, Scueal, Stanford, jpHMM) (Desire et al., 2018; Fabeni et al., 2017; Pineda-Pena et al., 2013). Subtype attribution based on sequences from other regions however is much

more challenging because of the limited availability of data and tools. Scueal and Stanford do not offer the possibility to subtype outside of *pol* (Pineda-Pena et al., 2013). For the full genome sequences we compared the results obtained with the subtyping tools of COMET, Smartgene and jpHMM. The only difficulty encountered was with the D/F1 recombinant (cluster ID41). Smartgene and COMET classified the *protease* part of *pol* as well as nearly the whole *env* gene as subtype D but in jpHMM only the *env* gene was recognized as subtype D. When the *protease* sequence alone was subjected to jpHMM however, it was assigned subtype D.

The fact that 4 of the 10 recombinant viruses were also identified as recombinants based only on the *pol* sequence generated for the purpose of drug resistance testing may suggest that this region contains favorite recombination breakpoints, but the position of the cross over differed for all 4.

Based on the known subtype prevalence in the Belgian MSM population primarily recombinants of the subtypes B, F1 and CRF02_AG were expected. Others have shown that the recombinants detected most frequently contained parts of the most prevalent subtypes in the region, like the CRF02_AG and F2 recombinants isolated in Cameroon (Banin et al., 2019) and the B and F1 recombinants isolated in Brazil (Reis et al., 2019). Foster et al. detected a second generation recombinant within an MSM population in the UK, composed of the predominant B subtype and the already circulating newly identified CRF50_A1D (Foster et al., 2014).

We found that 7 of the 10 new recombinants indeed contained genome fragments of the subtypes B, F1 or CRF02_AG and this suggests a local origination. The remaining 3 consisted of fragments of subtype A or of subtype A-containing CRFs and this suggests

an origination in Sub Saharan Africa. For many of the clusters included in this study at least one of the members was of non-European origin but we have no data to support the hypothesis that these individuals introduced the recombinant. Another shortcoming of this study is that full genome sequencing was restricted to a few representative patients per transmission cluster and the findings were extrapolated to the whole cluster. Representative patients however were carefully selected to cover the whole cluster avoiding selection of very closely linked patients. Moreover, the phylogenetic tree used to identify the clusters was based on the *pol* sequences so at least for those recombinants with a breakpoint in *pol* this provides strong evidence that all members of the cluster carry the same recombinant strain. An important note is also that the high prevalence of new recombinants found may not be extrapolated to the whole Belgian epidemic as we restricted the analysis to patients in active transmission clusters.

The formation of new recombinants may result in the generation and subsequent spread of viruses that carry new traits. Kouri et al. found an increasing trend of rapid progression associated with the spread of CRF19_cpx in Cuba. This recombinant is composed of subtype D, A and G parts and subtype D has been associated before with faster disease progression (Kouri et al., 2015). This supports the hypothesis of Ragupathy et al. that the characteristics of recombinants may mainly depend on the subtypes composing these viruses (Ragupathy et al., 2011). Lee et al. investigated 143 individuals in Uganda of whom 46% were infected with an intersubtype recombinant. They found no significant differences in clinical or therapy outcomes (Lee et al., 2017). Wertheim et al. however very recently provided evidence that individuals in transmission clusters have higher viral loads than non-clustered individuals suggestive for higher fitness and transmissibility

(Wertheim et al., 2019). Recombination may thus be part of a selection process driven by fitness and transmissibility. An interesting observation in that regard is that when examining the increase in clusters size over the two years following the closure of inclusion for this study (2018 and 2019), the 10 clusters resulting from transmission of a new CRF showed a higher expansion than the 52 clusters resulting from transmission of a pure subtype or known CRF (respectively 21.9% or an increase of 23 patients on a total of 105 compared to 15.1% or an increase with 101 patients on a total of 670; *p=0.09*)(additional observation, results not presented in the result section). Disturbed primer recognition is a known cause of misquantification in viral load assays (Alvarez et al., 2015) so changes in the genome sequence resulting from recombination may, in extreme cases, also impact the assays used for HIV viral load measurement as well as those used for drug resistance analysis and diagnosis. Problems with viral load measurement in regions with highly polymorphic HIV-1 have been reported (Bruzzone et al., 2014).

In conclusion, we demonstrated the increasing heterogeneity of HIV-1 strains transmitted in Belgium and the important contribution of new recombinants in local transmission. Reshuffling of the viral genome through recombination may result in viruses with new traits with implications on the transmissibility and pathogenesis of the infection and on the performance of laboratory assays for diagnosis and follow-up. Therefore it is an evolution that needs to be closely monitored in the future.

Acknowledgments

The Belgian AIDS Reference Laboratories are supported by the Belgian Ministry of Social Affairs through a fund within the Health Insurance System.

References

Alvarez, P., et al., 2015. HIV-1 variability and viral load technique could lead to false positive HIV-1 detection and to erroneous viral quantification in infected specimens. J Infection 71, 368-376. http://doi.org/10.1016/j.jinf.2015.05.011

Banin, A.N., et al., 2019. Near full genome characterization of HIV-1 unique recombinant forms in Cameroon reveals dominant CRF02_AG and F2 recombination patterns. Journal of the International AIDS Society 22, e25362. http://doi.org/10.1002/jia2.25362

Beloukas, A., et al., 2016. Molecular epidemiology of HIV-1 infection in Europe: An overview. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 46, 180-189. http://doi.org/10.1016/j.meegid.2016.06.033

Bruzzone, B., et al., 2014. Discordances with HIV-1 RNA quantitative determinations by three commercial assays in Pointe Noire, Republic of Congo. Journal of virological methods 203, 102-106. http://doi.org/10.1016/j.jviromet.2014.02.030

Dauwe, K., et al., 2015. Characteristics and spread to the native population of HIV-1 non-B subtypes in two European countries with high migration rate. BMC infectious diseases 15, 524. http://doi.org/10.1186/s12879-015-1217-0

Delgado, E., et al., 2019. Diverse Large HIV-1 Non-subtype B Clusters Are Spreading Among Men Who Have Sex With Men in Spain. Front Microbiol 10. http://doi.org/10.3389/fmicb.2019.00655

Desire, N., et al., 2018. Characterization update of HIV-1 M subtypes diversity and proposal for subtypes A and D sub-subtypes reclassification. Retrovirology 15, 80. http://doi.org/10.1186/s12977-018-0461-y

Dunn, D. and the UK Collaborative Group on HIV Drug Resistance, 2014. The increasing genetic diversity of HIV-1 in the UK, 2002-2010. Aids 28, 773-780. http://doi.org/10.1097/Qad.0000000000000119

Fabeni, L., et al., 2017. Comparative Evaluation of Subtyping Tools for Surveillance of Newly Emerging HIV-1 Strains. Journal of clinical microbiology 55, 2827-2837. http://doi.org/10.1128/JCM.00656-17

Foster, G.M., et al., 2014. Novel HIV-1 recombinants spreading across multiple risk groups in the United Kingdom: the identification and phylogeography of Circulating Recombinant Form (CRF) 50_A1D. PloS one 9, e83337. http://doi.org/10.1371/journal.pone.0083337

Fransen, K., et al., 1996. Longstanding presence in Belgians of multiple non-B HIV-1 subtypes. Lancet (London, England) 347, 1403. http://doi.org/10.1016/s0140-6736(96)91042-9

Gilbert, M.T., et al., 2007. The emergence of HIV/AIDS in the Americas and beyond. Proceedings of the National Academy of Sciences of the United States of America 104, 18566-18570. http://doi.org/10.1073/pnas.0705329104

Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95–8.

Hebberecht, L., et al., 2018. Frequency of occurrence of HIV-1 dual infection in a Belgian MSM population. PloS one 13, e0195679. http://doi.org/10.1371/journal.pone.0195679

Hebberecht, L., et al., 2019. Single genome sequencing of near full-length HIV-1 RNA using a limiting dilution approach. Journal of virological methods 274, 113737. http://doi.org/10.1016/j.jviromet.2019.113737

https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html, last assessed February 24th 2020.

Kouri, V., et al., 2015. CRF19_cpx is an Evolutionary fit HIV-1 Variant Strongly Associated With Rapid Progression to AIDS in Cuba. EBioMedicine 2, 244-254. http://doi.org/10.1016/j.ebiom.2015.01.015

Lai, A., et al., 2010. Changing patterns in HIV-1 non-B clade prevalence and diversity in Italy over three decades. HIV medicine 11, 593-602. http://doi.org/10.1111/j.1468-1293.2010.00832.x

Lai, A., et al., 2019. Local Epidemics Gone Viral: Evolution and Diffusion of the Italian HIV-1 Recombinant Form CRF60_BC. Front Microbiol 10, 769. http://doi.org/10.3389/fmicb.2019.00769

Lee, G.Q., et al., 2017. Prevalence and clinical impacts of HIV-1 intersubtype recombinants in Uganda revealed by near-full-genome population and deep sequencing approaches. AIDS (London, England) 31, 2345-2354. http://doi.org/10.1097/QAD.0000000000001619

Pineda-Pena, A.C., et al., 2013. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 19, 337-348. http://doi.org/10.1016/j.meegid.2013.04.032

Ragupathy, V., et al., 2011. Identification of new, emerging HIV-1 unique recombinant forms and drug resistant viruses circulating in Cameroon. Virology journal 8, 185. http://doi.org/10.1186/1743-422X-8-185

Reis, M.N.G., et al., 2019. Identification of New HIV-1 Circulating Recombinant Forms CRF81_cpx and CRF99_BF1 in Central Western Brazil and of Unique BF1 Recombinant Forms. Front Microbiol 10, 97. http://doi.org/10.3389/fmicb.2019.00097

Schultz, A.K., et al., 2012. jpHMM: recombination analysis in viruses with circular genomes such as the hepatitis B virus. Nucleic Acids Research 40, W193-W198. http://doi.org/10.1093/nar/gks414

Snoeck, J., et al., 2004. Rising prevalence of HIV-1 non-B subtypes in Belgium: 1983-2001. Journal of acquired immune deficiency syndromes (1999) 35, 279-285. http://doi.org/10.1097/00126334-200403010-00009

Struck, D., et al., 2014. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic acids research 42, e144. http://doi.org/10.1093/nar/gku739

Tamalet, C., et al., 2018. Emergence of uncommon HIV-1 non-B subtypes and circulating recombinant forms and trends in transmission of antiretroviral drug resistance in patients with primary infection during the 2013-2015 period in Marseille, Southeastern France. Journal of medical virology 90, 1559-1567. http://doi.org/10.1002/jmv.25228

Van Beckhoven, D., et al., 2015. Good continuum of HIV care in Belgium despite weaknesses in retention and linkage to care among migrants. BMC infectious diseases 15, 496. http://doi.org/10.1186/s12879-015-1230-3

Verhofstede, C., et al., 2018. Phylogenetic analysis of the Belgian HIV-1 epidemic reveals that local transmission is almost exclusively driven by men having sex with men despite presence of large African migrant communities. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 61, 36-44. http://doi.org/10.1016/j.meegid.2018.03.002

Verhofstede, C., et al., 2019. Exploring HIV-1 Transmission Dynamics by Combining Phylogenetic Analysis and Infection Timing. Viruses 11. http://doi.org/10.3390/v11121096

Vinken, L., et al., 2019. Earlier Initiation of Antiretroviral Treatment Coincides With an Initial Control of the HIV-1 Sub-Subtype F1 Outbreak Among Men-Having-Sex-With-Men in Flanders, Belgium. Front Microbiol 10, 613. http://doi.org/10.3389/fmicb.2019.00613

Wertheim, J.O., et al., 2019. Natural selection favoring more transmissible HIV detected in United States molecular transmission network. Nature communications 10, 5788. http://doi.org/10.1038/s41467-019-13723-z

Figure captions

Figure 1. Recombination pattern of the 10 new recombinant forms as predicted using the jumping profile Hidden Markov Model (jpHMM).

Figure 2. Maximum likelihood phylogenetic tree constructed with the 101 full-length genome sequences supplemented with reference sequences from the Los Alamos subtype reference panel (Ref.J.CM.04.04CMU11421.GU237072, Ref.B.NL.00.671_00T36.AY423387, Ref.B.TH.90.BK132.AY173951, Ref.B.US.98.1058_11.AY331295, Ref.F1.BE.93.VI850.AF077336, Ref.F2.CM.02.02CM_0016BBY.AY371158, Ref.01_AE.AF.07.569M.GQ477441, Ref.02_AG.CM.99.pBD6_15.AY271690, Ref.A1.AU.03.PS1044_Day0.DQ676872, Ref.A1.RW.92.92RW008.AB253421, Ref.A1.UG.92.92UG037.AB253429, Ref.A2.CD.97.97CDKTB48.AF286238, Ref.A2.CM.01.01CM_1445MV.GU201516, Ref.A2.CY.94.94CY017_41.AF286237).

The tree is rooted on the subtype J reference sequence. Patient IDs include the cluster number, subtype and a serial number. Reference sequences are marked in blue. Patients from the 10 clusters with a new recombinant are marked in red.