Increasing Frequency and Transmission of HIV-1 Non-B Subtypes among Men Who Have Sex with Men in the Swiss HIV Cohort Study

Jessy J. Duran Ramirez*[1,2], Tala Ballouz*[1,2,3], Huyen Nguyen[1,2], Katharina Kusejko[1,2], Sandra E. Chaudron[1,2], Michael Huber[2], Hans H. Hirsch[4,5], Matthieu Perreau[6], Alban Ramette[7], Sabine Yerly[8], Matthias Cavassini[9], Marcel Stöckle[4], Hansjakob Furrer[10], Pietro Vernazza[11], Enos Bernasconi[12], Huldrych F. Günthard*[1,2], Roger D. Kouyos*[1,2], and the Swiss HIV Cohort Study[¶]

1 Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, 8091 Zurich, Switzerland **2** Institute of Medical Virology, University of Zurich, 8057 Zurich, Switzerland **3** Epidemiology, Biostatistics and Prevention Institute, University of Zurich, 8001 Zurich, Switzerland **4** Division of Infectious Diseases and Hospital Epidemiology, University Hospital Basel, University of Basel, 4031 Basel, Switzerland **5** Transplantation and Clinical Virology, Department of Biomedicine, University of Basel, 4009 Basel, Switzerland **6** Division of Immunology and Allergy, Lausanne University Hospital, University of Lausanne, 1011 Lausanne, Switzerland **7** Institute for Infectious Diseases, University of Bern, 3001 Bern, Switzerland **8** Laboratory of Virology and Division of Infectious Diseases, Geneva University Hospital, University of Geneva, 1205 Geneva, Switzerland **9** Division of Infectious Diseases, Lausanne University Hospital, 1011 Lausanne, Switzerland **10** Department of Infectious Diseases, Bern University Hospital, University of Bern, 3010 Bern, Switzerland **11** Division of Infectious Diseases, Cantonal Hospital St. Gallen, 9007 St. Gallen, Switzerland **12** Division of Infectious Diseases, Regional Hospital Lugano, 6900 Lugano, Switzerland

***** These authors contributed equally to the manuscript

[¶] Membership of the Swiss HIV Cohort Study is listed in the Acknowledgments

Correspondence: Prof. Dr. Roger D. Kouyos, +41 44 255 36 10, roger.kouyos@usz.ch

Short Summary: The proportion of MSM with newly diagnosed HIV-1 non-B subtypes has strongly increased since 1990, reaching 34% of all MSM diagnosed in 2019. This is partly due to ongoing local transmission in this population, as evidenced by large transmission clusters.

Running title: Non-B Subtypes among MSM in Switzerland

Word count abstract: 199

Word count article: 3496

Abstract

Background

In Switzerland, HIV-1 transmission among men who have sex with men (MSM) has been dominated by subtype B, whilst non-B subtypes are commonly attributed to infections acquired abroad among heterosexuals. Here, we evaluated the temporal trends of non-B subtypes and the characteristics of molecular transmission clusters (MTCs) among MSM.

Methods

Sociodemographic and clinical data and partial *pol* sequences were obtained from participants enrolled in the Swiss HIV Cohort Study (SHCS). For non-B subtypes, maximum likelihood trees were constructed, from which Swiss MTCs were identified and analysed by transmission group.

Results

Non-B subtypes were identified in 8.1% (416/5,116) of MSM participants. CRF01_AE was the most prevalent strain (3.5%), followed by A (1.2%), F (1.1%), CRF02_AG (1.1%), C (0.9%), and G (0.3%). Between 1990 and 2019, an increase in the proportion of newly diagnosed individuals (0/123[0%] to 11/32 [34%]) with non-B subtypes in MSM was found. Across all non-B subtypes, the majority of MSM MTCs were European. Larger MTCs were observed for MSM than heterosexuals.

Conclusions

We found a substantial increase in HIV-1 non-B subtypes among MSM in Switzerland and the occurrence of large MTCs, highlighting the importance of molecular surveillance in guiding public health strategies targeting the HIV-1 epidemic.

Keywords: HIV-1, transmission cluster, non-B subtypes, MSM, molecular epidemiology, phylogeny

Introduction

Human immunodeficiency virus (HIV) is characterized by extensive genetic diversity with high replication, recombination, and mutation rates due to its reverse transcriptase enzyme[1,2]. HIV-1 consists of four distinct strains: (M: “main”, O: “outlier”, N: “non-M/ -O”, and P: “pending”), each of which origins in distinct cross-species (zoonotic) transmission event[3]. HIV-1 group M, responsible for the majority of the global pandemic, diversifies into nine subtypes (named A-D, F-H, J, and K), six sub-subtypes (A1, A2, A3, A4, F1, and F2)[4], and at least 102 circulating recombinant forms (CRFs) and several recombinant forms[5].

Subtype C accounts for around 50% of all HIV-1 infections worldwide and is predominantly found in Sub-Saharan Africa[6,7]. In Western Europe, the HIV-1 epidemic among men who have sex with men (MSM) has been classically dominated by subtype B transmissions[8]. On the other hand, non-B subtypes are mainly transmitted among heterosexual contacts (HETs) and are prevalent in individuals of non-European origin[9]. However, in the last decade, the United Kingdom and some European countries such as Spain, Italy, the Netherlands and France have observed a rise in non-B subtypes and circulating recombinant forms (CRFs) infections in MSM[10–15]. This has been commonly attributed to an increased proportion of non-European immigrants as well as spread within the local European populations.

The epidemiology of HIV-1 in Switzerland is similar to the rest of Western Europe with subtype B constituting the most prevalent subtype, in particular among European MSM[16,17]. In 2011, an overall increase in the occurrence of non-B infections in the Swiss HIV Cohort Study (SHCS) between 1996 and 2010 was observed, particularly among HETs and to a lesser extent among MSM[18]. However, the proportion of non-B subtypes among MSM were negligible[18]. Since then, it has not been evaluated to what extent non-B subtypes are transmitted locally and whether non-B subtypes constitute a comparably larger proportion of newly diagnosed HIV-1 infections among MSM.

In the present study, we aimed to characterize the distribution and transmission of HIV-1 non-B subtypes among MSM in Switzerland based on available sequences and clinical data within the SHCS. The resulting understanding of the trends in HIV-1 diversity and transmission of non-B subtypes may help guide public health interventions in Switzerland by identifying the main sources of ongoing transmission.

Materials and Methods

The Swiss HIV Cohort Study and Viral Sequences

The SHCS is an ongoing multicenter, nationwide, prospective cohort established in 1988 with continuous enrolment and at least biannual study visits[19]. As of April 2020, more than 20,000 HIV-infected adults have been enrolled in the cohort[19]. The representativeness of the SHCS is estimated to be high, covering at least 53% of all adults ever diagnosed with HIV in Switzerland, 80% of all HIV-infected MSM[20], 69% of all registered AIDS cases and 72% of all antiretroviral treated individuals[21]. All participants have provided written informed consent and the SHCS has been approved by the participating institutions' ethics committees[19].

The SHCS has a linked drug resistance database (DRDB) which includes all genotypic resistance tests performed on enrolled participants. As of April 2020, the DRDB contains 29,837 sequences belonging to 13,278 of participants. Since 2002, these tests are routinely done at the time of enrolment or when viral rebound is observed while on ART. Furthermore, more than 12,000 sequences have been generated retrospectively from the biobank, from when drug resistance testing was not routinely performed[21,22]. For approximately 60% of the SHCS participants, at least one partial polymerase (*pol*) sequence (comprising the full protease and a minimum of codons 28-225 of the reverse transcriptase) are available[23].

Sequence Selection and Study Population

We considered all partial *pol* sequences of the SHCS DRDB, which were linked to a SHCS participant. The subtyping was conducted on the sequence using a combination of two subtyping tools: REGA v3.0[24] and COMET[25].

Multiple sequences and subtype classification variations were identified in 300 individuals. In these cases, the following sequence selection and subtype classification algorithm was applied to assign each individual to one subtype. Sequences were prioritized if they were retrospectively linked to a participant and if the sequence length fell above the 870bp threshold (corresponding to the first quartile of all sequence lengths, Figure S1). If there was still more than one sequence per participant available, we selected the sequence with the earliest sample date and also the earliest sequence date. Duplicated sequences were removed. In cases where the subtype/CRF was not unambiguously assigned in the selected sequence, this sequence was classified as a "potential recombinant". For such potential recombinants, we tried to identify the most likely subtype by the following rule: If the selected sequence was clearly classified by COMET[25] as subtype X with at least one additional sequence from that patient with the same subtype classification available by REGA[24], then the subtype was assigned to the subtype called by COMET[25]. For instance, if COMET[25] had classified the selected sequence of a patient as subtype A and REGA[24] classified at least one sequence as subtype A, then the individual infection was assigned to subtype A. Additionally, if REGA[24] had classified the sequence as recombinant of B, D or B, F1, and COMET[25] as subtype B, the sequences were assigned to subtype B. As these sequences clustered within subtype B, we considered that the discrepancy in classification could thus be an artefact from the subtyping procedure. After applying the sequence selection and classification algorithm, 12,852 sequences were considered for the subsequent analyses (Figure 1).

Definitions

Only the subtypes and recombinants A, C, F, G, CRF01_AE, and CRF02_AG were considered as separate groups to ensure sufficient sample size for statistical analysis. Other pure subtypes, CRFs,

and sequences with undefined subtype (N=706) were pooled and excluded from the primary analysis (Table S1), but considered in a sensitivity analysis (Figure S6).

Transmission groups were categorized into MSM and HETs based on the physician-reported source of infection. Other transmission modes (e.g., intravenous drug user, mother to child, transfusion risks) were not considered in these analyses. Ethnicity was self-reported by participants according to the following categories: white, black, Asian, Hispanic, or other. The SHCS centers are located in Zurich, Basel, Bern, Geneva, Lausanne, Lugano and St. Gallen. As the Zurich center contributes to approximately 50% of all study participants, we grouped the centers into Zurich and non-Zurich. Participants' countries of origin were grouped using the UNAIDS[26] classification with some modification into the following geographic regions of origin: "Western, Northern and Southern Europe", "Eastern Europe", "Latin America and the Caribbean", "North America and Oceania", "Western and Central Africa", "Eastern and Southern Africa", "Western Asia and North Africa", "Eastern and Southern Asia" (Figure S2). We used the geographic region of origin as a proxy assessment of whether the subtypes were introduced from abroad since information on the likely location of transmission has been collected only since 2007 in the SHCS. Furthermore, only 35.1% (1,793/5,114) of MSM and 27.9% (1,136/4,076) of HETs included in the analysis reported whether they were infected most likely in Switzerland or abroad. Therefore, the likely location of transmission was not considered in the analysis. We additionally checked for association with the geographic region of origin after 2007 (Table S6-7, Figure S5). Clinical variables comprised the CD4 cell counts and viral load at diagnosis. To estimate the HIV diagnosis date as accurate as possible, we used the earliest date of the following: first positive HIV-1 screening test, documented positive HIV-1 test result, and study enrolment. Finally, the yearly fraction of newly diagnosed people living with HIV-1 (PLHIV) B/non-B subtypes was calculated.

Phylogenetic Tree Construction

A separate phylogenetic tree was built for each of the six most frequent non-B subtypes (Figure S3). The participants' partial *pol* sequences were blasted against non-Swiss background sequences from the Los Alamos National Laboratory HIV database[27]. For each SHCS sequence, the ten closest hits, Los Alamos sequences with at least 90% identity to the participant's sequence, were included. The sequences were aligned to the reference genome HXB2 using MUSCLE[27,28]. Insertions related to HXB2 and known resistance mutation positions[29,30] were removed, and gaps were trimmed by trimAl[31]. Maximum likelihood phylogeny was reconstructed with FastTree using a generalized time-reversible (GTR) and CAT model[32,33]. For each subtype and CRF, one hundred bootstrap trees were constructed using FSEQBOOT[32].

Identification and statistical analysis of transmission clusters

For each phylogenetic tree, Swiss molecular transmission clusters (MTCs), were defined as those consisting of at least 2 sequences with ≥ 80% Swiss sequences, ≥ 80% bootstrap and a maximum pairwise genetic distance ≤ 4.5%[34]. Information about the most likely route of HIV-infection, ethnicity, geographical region of origin and the year of HIV diagnosis was mapped to these Swiss clusters. As we aimed to study MSM MTCs, more than 50% of the cluster members needed to correspond to MSM [34]. With this definition, we also captured MTCs not containing exclusively MSM, which is important to characterize potential MSM not disclosed as MSM, as has been previously shown[17]. Additionally, to compare MSM with HETs, HETs MTCs were extracted following the same principles. MTCs with ≥10 individuals were considered as large MTCs.

To distinguish between imported and domestic MTCs, the percentage of participants from a certain geographical region of origin was assessed by calculating the number of participants from Europe, Asia, America, or Africa to the total number of participants clustering or not-clustering by transmission group and subtype.

Statistical Analysis

Continuous variables were reported as median and interquartile range (IQR) and categorical variables as frequency (N) and percentage (%). Differences between groups were assessed using Wilcoxon or Kruskal-Wallis tests for continuous variables and by $\chi 2$ or Fisher exact tests where appropriate for categorical variables. Statistical significance was defined as p-value < 0.05.

To evaluate the trend in occurrence of non-B subtypes over time, logistic regression was performed using B-spline regression with four degrees of freedom for the year of HIV diagnosis. Results were reported as the odds ratio (OR) with 95% confidence interval (CI).

Sensitivity analyses were performed by (1) including the other non-B subtypes (Figure S6) to the analysis; (2) varying cluster definitions (combinations of bootstrap support values (no or ≥80%) and inter-cluster genetic distances (≤1.5% and ≤4.5%)) (Figures S7-9) and (3) minimizing the bias in transmission cluster detection by extracting MTCs with RAxML[35,36] and comparing them with the previously detected MTCs by FastTree[32] (Figures S7-8). All analyses were done with R, version 3.6.1.

Results

Characteristics of the study population

We included 12,146 PLHIV enrolled in the SHCS for whom HIV genotype information was available. Of these individuals 5,114 (42.1%) had acquired HIV through MSM contact and are considered the primary study population (Table 1). A total of 4,076 (33.5%) were infected by HET contact and are considered for descriptive comparison.

Overall, the median age of MSM participants at diagnosis was 35.0 years (IQR, 29.0-43.0) and the majority of individuals were of white ethnicity (4,541/5,114[88.8%]). HIV-1 subtype B was identified for 91.8% (4,698/5,114) of the MSM cases and non-B subtypes for 8.2% (416/5,114). Among those harbouring a non-B subtype, the proportion of non-white individuals was higher than those with a B subtype. No other differences in demographics or clinical characteristics (e.g., cohort center, CD4 cell count or viral load) between individuals harbouring B or non-B subtypes were found (Table 1).

Similarly, the median age of HETs was 34.0 years (IQR, 27.0-43.0) and most were of white ethnicity (2,503/4,076[61.4%|), albeit, the proportion of white individuals was lower than in MSM. Approximately 48% (1,953/4,076) of HET participants were infected with a non-B subtype. A difference in ethnicities and geographical region of origin was observed between B and non-B subtypes. Among those harbouring a non-B subtype, higher proportions of HET participants were of non-white ethnicity and non- European origin (Table S2).

Among MSM participants, CRF01_AE (180/5,114[3.5%]), A (60/5,114[1.2%]), F (58/5,114[1.1%]), CRF02_AG (57/5,114[1.1%]), C (44/5,114[0.9%]) and G (17/5,114[0.3%]) were the most frequent non-B subtypes. As shown in Table 2, a higher proportion of participants with CRF02_AG, C, and G subtypes were enrolled in non-Zurich cohort centers (71.9%, 65.9%, and 64.7%, respectively). Differences in self-reported ethnicity were observed mainly among participants with CRF01_AE and CRF02_AG, with higher proportions of Asian and Black individuals in the two subtypes, respectively. By contrast, among HET participants, CRF02_AG (493/4,076[12.1%]), C (457/4,076[11.2%]), A

(447/4,076[11.0%]), CRF01_AE (326/4,076[8.0%]), G (147/4,076[3.6%]), and F (83/4,076[2.0%]) were the six most frequent non-B subtypes. Among those with CRF01_AE, we found higher proportions of participants enrolled in the Zurich cohort center (45.1%), originating from Europe (63.8%), and of white ethnicity (58.6%). No other differences were observed (Table S3).

Increasing non-B subtype proportions between 1990 and 2019

Among newly diagnosed MSM, the proportion of non-B subtypes increased from 0% (0/123) in 1990 to approximately 1% (9/113) in 2000, 10% (19/187) in 2010 and 34% (11/32) in 2019 (Figure 2AB). These represent in recent years a substantial fraction of new diagnoses. Accordingly, we observed an increase of non-B subtypes among MSM (univariable logistic regression, OR [95% CI] = 1.13[1.12-1.15]/year; Figure 2C), with the highest increase for subtype A (1.16 [1.11-1.21]/year) and CRF02_AG (1.15 [1.10-1.20]/year; Figure 2D). After including the year of diagnosis with B-splines, the model significantly improved (Likelihood Ratio (LR) = 14.37, p-value = 0.002). However, even the splines did not exhibit a saturation of the increase of non-B subtypes, indicating a continuous, almost-linear increase of the probability of harbouring a non-B subtype over time (Figure 2C).

In comparison, among newly diagnosed HETs, the proportion of non-B subtypes saturated after 2000 (Figure S4AB). In the HETs model as well, when including the diagnosis year with B-splines, a clear saturation in harbouring non-B subtype over time was exhibited (Figure S4C). Comparable results for both MSM and HETs were obtained after including the other non-B subtypes in a sensitivity analysis (Figure S6).

Detection and description of MSM clusters

Overall, in non-B subtypes, we found less MSM than HETs (both among clustering and non-clustering individuals). Within the phylogeny and across the transmission groups, the majority of individuals were not in MTCs for MSM and HETs. However, the fraction of clustering individuals was larger for

MSM compared to HETs (41% vs. 23%, p-value < 0.001). The comparison between MSM/HETs in MTCs and MSM/HETs on the phylogeny, which did not cluster, is shown in Figure 3A.

A total of 33 MTCs among non-B subtypes in MSM (8 A, 1 C, 5 F, 1 G, 12 CRF01_AE, and 6 CRF02_AG) were identified (Table S4). Across all the non-B subtypes, members of possible MSM MTCs identified were mainly from Europe (Figure 3B). The highest number of MSM MTCs was found for CRF01_AE (Figure S7A, C) and the highest proportion of MSM MTCs was for subtype F (Figure S7BD).

The MTCs sizes ranged from 2 to 19 and larger MTCs were found for MSM compared to HETs (Figure 4). Among the larger ones, an MTC with 11-13 members was found for each of CRF01_AE, CRF02_AG, and subtype F. Only one CRF01_AE MTC reached values above 17 members (Table S4). The characteristics of the four largest MTCs are described in Table S5. Only the largest CRF01_AE MTC had 100% of MSM cluster members (19/19). The highest proportion of white ethnicity (11/11[100%]) and Europeans (10/11[90.9%]) were found for the large subtype F MTC. The size of MTC for other subtypes and CRFs ranged mostly between 2 to 4 (Table S4).

Several sensitivity analyses varying the cluster definition (Figures S7-9) and approaches (i.e., RAxML[35,36]) (Figures S7-8) were performed, which led to similar results for MSM and HETs.

Discussion

Our study is the first to thoroughly investigate the frequent occurrence and transmission of HIV-1 non-subtype B genetic forms among 5,114 MSM in the SHCS using phylogenetic methods combined with epidemiological and clinical data. We found that the proportion of newly diagnosed non-B subtypes among MSM increased from 0% (0/123) in 1990 to 34% (11/32) in 2019. Furthermore, larger Swiss MTCs of non-subtype B forms were found in MSM compared to HETs, indicating higher transmission among the MSM networks in Switzerland.

Similar to other Western European countries, the most prevalent HIV-1 genetic form in Switzerland is subtype B, mainly associated with ongoing transmissions among MSM [8]. Yet, non-B subtypes have been predominantly found in HETs[9,12]. In our study, even though HETs still constitute the largest proportion of non-B subtype infections (1,953/2,369[82.4%]), MSM accounted for a considerable fraction of these infections (416/2,353[17.6%]). CRF01_AE was the most frequently identified non-B subtype among MSM participants, followed by A, F, CRF02_AG, C and G. CRF01_AE, which is most prevalent in Thailand and neighbouring countries in Southeast Asia, has been circulating for some time in Switzerland. This has been reported to be most likely associated with sex tourism and emigration from Southeast Asia[37], the latter of which is supported by the high proportion of members of Asian ethnicity in the MTCs for CRF01_AE. Generally, differences in participant ethnicities and region of origin followed the known geographical distribution of the subtypes. In line with other studies, we did not observe any other demographic or clinical differences such as age, viral load or CD4 cell counts among the non-B subtypes[9]. However, this must be carefully interpreted as these were based on a single measurement and additional possible confounders were not adjusted for.

Several European countries have reported increased transmission of non-B subtypes in MSM. This increase has been linked to imported infections from countries where these non-B subtypes are predominant, followed by transmission among the local MSM population[9,10,12–14,38–41].

Similarly, we found that five pure subtypes, 13 CRFs and several unassigned recombinant forms have been increasingly introduced into the MSM population in the SHCS, mainly from 2008 onwards. Overall, the highest temporal increase in non-B forms among MSM was observed for subtype A and CRF02_AG, both of which are commonly found among HETs and are among the most predominant forms (together with subtype C) in Western Europe[7]. In general, non-B subtype MSM transmission events were more common among Europeans, as the majority of MTCs members across all non-B subtypes were of European origin. This reflects probable domestic transmission and increasing circulation within the local population, in accordance with findings in other European countries [10–15,42]. Even though the majority of individuals in both MSM and HETs were not found in an MTC, larger MTCs were found for MSM. This increased clustering suggests a more probable ongoing transmission among MSM in Switzerland. CRF01_AE accounted for both the highest number of MSM MTCs and the largest number of members per MTC. Only six MTCs of CRF02_AG comprising up to 12 individuals and eight MTCs for subtype A containing even fewer individuals (i.e., up to three) were found. This contrasts with other publications reporting larger MTCs for CRF02_AG and subtype A in Europe. For example, in a recently published analysis of HIV-1 non-B subtype clusters in Spain, CRF02_AG was one of the four largest MSM clusters comprising 115 individuals[11]. Interestingly, we found that the highest proportion of MSM MTCs among all MTCs was for subtype F. Similar results have been described in Belgium[43], France[15], Italy[14], and Spain[44] with rapid expansion of subtype F MTCs suggesting a higher potential for transmissibility of the virus.

Taken together, these findings show a recent increase of non-B subtype transmission events among MSM, as previously reported in Germany[40] and the United Kingdom[10]. These non-B subtypes' import into Swiss transmission networks may be traceable to migrants from areas in which these forms are predominant (or via Europeans associated with individuals from such areas). Multiple non-B subtype transmission events between HETs and MSM and within the same transmission group may have occurred, and these subtypes and CRFs are now circulating in the local population. We have previously shown that approximately 10% of HET sequences were associated with MSM MTCs,

most likely due to a non-disclosure as MSM and women infected by bisexual men [17,34]. Timely implementation of prevention strategies, including behavioural interventions and pre-exposure prophylaxis, among this high-risk population is thus crucial to curb further transmission.

Like all studies relying on phylogenetic inference, our study is subject to several caveats such as biased and incomplete sampling. The small size of most MTCs may reflect imperfect sampling or phylogenetic uncertainty. To minimize the bias in MTCs detection, we confirmed our findings by using another phylogenetic reconstruction method (i.e., RAxML[35,36]**)** to extract MTCs. A further limitation arises from subtype misclassifications by subtyping tools. However, adequate subtyping is essential when studying different non-B subtypes. In this study, we established a subtype classification algorithm to define rules for 300 (2.3%) individuals having multiple *pol* sequences from different time points with subtype variations. In previous studies, the subtyping tools REGA[24] and COMET[25] demonstrated concordant classification for 96.5% [38] or 85.6% [9], respectively, mitigating this problem. Despite these limitations, the densely sampled SHCS remains highly representative and provides important insights on the ongoing transmission of non-subtype B genetic forms in Switzerland[19][19].

Conclusion

Our data provide evidence for an increasing long-term trend of non-subtypes B infections and ongoing transmission among MSM in a well-defined, representative Swiss study population. This highlights the need to implement additional, more specific public health measurements aiming at preventing HIV-1 transmission among MSM, and to closely monitor the expansion of non-B subtype transmission clusters.

Acknowledgements

We thank the participants of the Swiss HIV Cohort Study; the physicians, and study nurses for excellent patient care; the resistance laboratories for high-quality genotyping drug resistance testing; the SHCS data center (A. Scherrer, K. Kusejko, J. Meier, Y. Schäfer and O. Follonier) for excellent data management; and D. Perraudin and M. Amstad for administrative assistance.

Members of the Swiss HIV Cohort Study are as follows:

K. Aebi-Popp, A. Anagnostopoulos, M. Battegay, E. Bernasconi, J. Böni, D. L. Braun, H. C. Bucher, A. Calmy, M. Cavassini, A. Ciuffi, G. Dollenmaier, M. Egger, L. Elzi, J. Fehr, J. Fellay, H. Furrer, C. A. Fux, H. F. Günthard (President of the SHCS), D. Haerry (deputy of "Positive Council"), B. Hasse, H. H. Hirsch, M. Hoffmann, I. Hösli, M. Huber, C. R. Kahlert (Chairman of the Mother & Child Substudy), L. Kaiser, O. Keiser, T. Klimkait, R. D. Kouyos, H. Kovari, K. Kusejko (Head of Data Centre), B. Ledergerber, G. Martinetti, B. Martinez de Tejada, C. Marzolini, K. J. Metzner, N. Müller, D. Nicca, P. Paioni, G. Pantaleo, M. Perreau, A. Rauch (Chairman of the Scientific Board), C. Rudin, , P. Schmid, R. Speck, M. Stöckle (Chairman of the Clinical and Laboratory Committee), P. Tarr, A. Trkola, P. Vernazza, G. Wandeler, R. Weber, S. Yerly.

Footnote

Funding

This work was supported by the Swiss National Science Foundation [SNF grant number 33CS30_177499 to H. F. G.] in the framework of the Swiss HIV Cohort Study. The data were gathered by the five Swiss university hospitals, two cantonal hospitals, 15 affiliated hospitals, and 36 private physicians (listed at http://www.shcs.ch/180-health-care-providers). The work was further supported by the SNF [grant numbers 324730B_179571 and 310030_141067 to H. F. G. and PZ00P3-

142411 and BSSGI0_155851 to R. D. K.]; the Yvonne-Jacob Foundation (to H. F. G.); the University of Zurich's Clinical Research Priority Program for viral infectious disease; the Zurich Primary HIV Infection Cohort Study (to H. F. G.); and an unrestricted research grant from Gilead Sciences to the SHCS research foundation.

Conflicts of interests

R. D. K. has received grants from the SNF and personal fees from Gilead Sciences, outside the submitted work. H. F. G. has received unrestricted research grants from Gilead Sciences and Roche; fees for data and safety monitoring board membership from Merck; and consulting/advisory board membership fees from Gilead Sciences, ViiV, Sandoz, and Mepha. The institution of H.F. has received unrestricted educational grants from Gilead Sciences, MSD, ViiV, Sandoz, and Abbvie.

Corresponding author:

Prof. Dr. Roger Kouyos

Division of Infectious Diseases and Hospital Epidemiology

University Hospital Zurich, Rämistrasse 100, CH-8091 Zürich

+41 44 255 36 10, roger.kouyos@usz.ch

References

1. Hemelaar J. Implications of HIV diversity for the HIV-1 pandemic. Journal of Infection. W.B. Saunders; 2013. p. 391–400.

2. Bbosa N, Kaleebu P, Ssemwanga D. HIV subtype diversity worldwide. Current Opinion in HIV and AIDS. **2019**; 14(3):153–160.

3. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. Cold Spring Harbor Perspectives in Medicine. **2011**; 1(1).

4. Hemelaar J. The origin and diversity of the HIV-1 pandemic. Trends in Molecular Medicine. Elsevier Current Trends; 2012. p. 182–192.

5. LOS ALAMOS. HIV Circulating Recombinant Forms (CRFs) [Internet]. HIV Circulating Recombinant Forms (CRFs). 2020 [cited 2020 Oct 5]. p. https://www.hiv.lanl.gov/content/sequence/HIV/CRFs. Available from: https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html

6. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS. NIH Public Access; **2011**; 25(5):679–689.

7. Hemelaar J, Elangovan R, Yun J, et al. Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. The Lancet Infectious Diseases. Lancet Publishing Group; **2019**; 19(2):143–155.

8. Abecasis AB, Wensing AMMJ, Paraskevis D, et al. HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. Retrovirology. BioMed Central; **2013**; 10(1):7.

9. Dauwe K, Mortier V, Schauvliege M, et al. Characteristics and spread to the native population of HIV-1 non-B subtypes in two European countries with high migration rate. BMC Infectious Diseases. **2015**; 15(1):524.

10. Ragonnet-Cronin M, Lycett SJ, Hodcroft EB, et al. Transmission of Non-B HIV Subtypes in the United Kingdom Is Increasingly Driven by Large Non-Heterosexual Transmission Clusters Manon. The Journal of Infectious Diseases. **2016**; 213:1410–8.

11. Delgado E, Benito S, Montero V, et al. Diverse large HIV-1 non-subtype B clusters are spreading among men who have sex with men in Spain. Frontiers in Microbiology. **2019**; 10(APR).

12. Snoeck J, Laethem K Van, Hermans P, et al. Rising Prevalence of HIV-1 Non-B Subtypes in Belgium: 1983-2001. Journal of Acquired Immune Deficiency Syndromes. **2004**; 35(3):279–285.

13. Chaix ML, Seng R, Frange P, et al. Increasing HIV-1 non-B subtype primary infections in patients in France and effect of HIV subtypes on virological and immunological responses to combined antiretroviral therapy. Clinical Infectious Diseases. Oxford Academic; **2013**; 56(6):880–887.

14. Lai A, Riva C, Marconi A, et al. Changing patterns in HIV-1 non-B clade prevalence and diversity in Italy over three decades. HIV Medicine. John Wiley & Sons, Ltd; **2010**; 11(9):593–602.

15. Frange P, Meyer L, Deveau C, et al. Recent HIV-1 infection contributes to the viral diffusion over the French territory with a recent increasing frequency. Vartanian J-P, editor. PLoS ONE. Public Library of Science; **2012**; 7(2):e31695.

16. Turk T, Bachmann N, Kadelka C, et al. Assessing the danger of self-sustained HIV epidemics in heterosexuals by population based phylogenetic cluster analysis. eLife. **2017**; 6:28721.

17. Kouyos RD, Wyl V Von, Yerly S, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. Journal of Infectious Diseases. **2010**; 201(10):1488–1497.

18. Wyl V Von, Kouyos RD, Yerly S, et al. The role of migration and domestic transmission in the spread of HIV-1 non-B subtypes in Switzerland. Journal of Infectious Diseases. The Journal of infectious diseases.; 2011. p. 1095–1103.

19. Schoeni-Affolter F, Ledergerber B, Rickenbach M, et al. Cohort profile: The Swiss HIV cohort study. International Journal of Epidemiology. Oxford Academic; **2010**; 39(5):1179–1189.

20. Sighem A van, Vidondo B, Glass TR, et al. Resurgence of HIV Infection among Men Who Have Sex with Men in Switzerland: Mathematical Modelling Study. PLoS ONE. **2012**; 7(9):e44819.

21. Yang WL, Kouyos R, Scherrer AU, et al. Assessing the paradox between transmitted and acquired HIV type 1 drug resistance mutations in the Swiss HIV Cohort Study from 1998 to 2012. Journal of Infectious Diseases. **2015**; 212(1):28–38.

22. Yang WL, Kouyos RD, Böni J, et al. Persistence of Transmitted HIV-1 Drug Resistance Mutations Associated with Fitness Costs and Viral Genetic Backgrounds. PLoS Pathogens. **2015**; 11(3):1–13.

23. Shilaih M, Marzel A, Yang WL, et al. Genotypic Resistance Tests Sequences Reveal the Role of Marginalized Populations in HIV-1 Transmission in Switzerland. Scientific Reports. Nature Publishing Group; **2016**; 6:27580.

24. Pineda-Peña AC, Faria NR, Imbrechts S, et al. 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. Elsevier; **2013**; 19:337–348.

25. Struck D, Lawyer G, Ternes AM, Schmit JC, Bercoff DP. COMET: Adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic Acids Research. Oxford University Press; **2014**; 42(18):e144.

26. UNAIDS. UNAIDS - Countries [Internet]. 2016 [cited 2020 Dec 10]. Available from: http://www.unaids.org/en/regionscountries/countries

27. Los Alamos National Laboratory. HIV sequence database [Internet]. 2016 [cited 2020 Mar 19]. Available from: https://www.hiv.lanl.gov/content/index

28. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. **2004**; 32(5):1792–1797.

29. Shafer R. HIV Drug Resistance Database [Internet]. Journal of Infectious Diseases. 2006 [cited 2020 Jun 9]. p. S51-8. Available from: https://hivdb.stanford.edu/

30. Gilles P, Vancutsem LM. International Antiviral [Internet]. 2015 [cited 2020 Dec 9]. p. 2–4. Available from: https://www.iasusa.org/

31. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. Bioinformatics; **2009**; 25(15):1972–1973.

32. Price MN, Dehal PS, Arkin AP. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. Molecular Biology and Evolution. **2009**; 26(7):1641–1650.

33. Stamatakis A. Phylogenetic models of rate heterogeneity: A high performance computing perspective. 20th International Parallel and Distributed Processing Symposium, IPDPS 2006. 2006.

34. Bachmann N, Kusejko K, Nguyen H, et al. Phylogenetic Cluster Analysis Identifies Virological and Behavioral Drivers of Human Immunodeficiency Virus Transmission in Men Who Have Sex With Men. Clinical Infectious Diseases. **2020**; :1–9.

35. Stamatakis A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. Oxford Academic; **2006**; 22(21):2688–2690.

36. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. Oxford University Press; **2014**; 30(9):1312–1313.

37. Angelis K, Albert J, Mamais I, et al. Global dispersal pattern of HIV type 1 subtype CRF01-AE: A genetic trace of human mobility related to heterosexual sexual activities centralized in southeast Asia. Journal of Infectious Diseases. Oxford Academic; **2015**; 211(11):1735–1744.

38. Lorenzin G, Gargiulo F, Caruso A, et al. Prevalence of non-B HIV-1 subtypes in North Italy and analysis of transmission clusters based on sequence data analysis. Microorganisms. **2020**; 8(1):1–20.

39. Neogi U, Häggblom A, Santacatterina M, et al. Temporal trends in the Swedish HIV-1 epidemic: Increase in non-B subtypes and recombinant forms over three decades. PLoS ONE. Public Library of Science; **2014**; 9(6):e99390.

40. Hanke K, Faria NR, Kühnert D, et al. Reconstruction of the Genetic History and the Current Spread of HIV-1 Subtype A in Germany. Journal of Virology. **2019**; 93(12):2238–2256.

41. Dunn D. The increasing genetic diversity of HIV-1 in the UK, 2002-2010. AIDS. Wolters Kluwer Health; **2014**; 28(5):773–780.

42. Beloukas A, Psarris A, Giannelou P, Kostaki E, Hatzakis A, Paraskevis D. Molecular epidemiology of HIV-1 infection in Europe: An overview. Infection, Genetics and Evolution. Elsevier; **2016**; 46:180–189.

43. Verhofstede C, Dauwe K, Fransen K, et al. 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. Elsevier; **2018**; 61:36–44.

44. Delgado E, Cuevas MT, Domínguez F, et al. Phylogeny and phylogeography of a recent HIV-1 subtype F outbreak among men who have sex with men in Spain deriving from a cluster with a wide geographic circulation in Western Europe. PLoS ONE. Public Library of Science; **2015**; 10(11):e0143325.

Figure Legends

Figure 1. Flowchart to illustrate the sequence selection and subtype classification algorithm. Abbreviation: SHCS, Swiss HIV Cohort Study; X, any subtype or circulating recombinant form

Figure 2. (A, B) Yearly proportion of B/non-B subtype among newly diagnosed MSM in absolute (A) and relative (B) numbers. The bar charts are colored by subtype B (blue) and non-B subtypes (red). X-axis: year of diagnosis **(C, D) Increasing frequency of non-B subtypes among MSM over time using logistic regression.** (**C**) The overall time trend of the non-B subtypes (**D**) and the time trend for the six most frequently observed non-B subtypes. Y-axis: Log-odds for harbouring a non-B subtype; x-axis: year of diagnosis. Abbreviations: LR, likelihood ratio; MSM, men who have sex with men

Figure 3. Higher proportion of clustered individuals and European origin among MSM. (**A**) Mosaic plot representing the association between transmission group and whether individuals were clustered or not-clustered (with p-values from two-tailed Fisher's exact test). (**B**) Mosaic plot representing the association between transmission group, geographical region of origin and clustered or not-clustered individuals. The two bottom right boxes for MSM in each subtype group are taller than the left boxes for HET. That is, the proportion of participants of European descent is higher in MSM for all subtypes. The number in the rectangle corresponds to the number of individuals. The clusters presented herein were extracted by using a bootstrap value of ≥80% and maximum pairwise genetic distance of ≤4.5%. Abbreviations: AF: Africa, blue; AM: America, brown; AS: Asia, green; CL: cluster, dark shades; EU; Europe, red; HET: heterosexual contact; NCL: non-cluster, light shades; MTC: molecular transmission cluster, MSM: men who have sex with men

Figure 4. Cluster size distribution depends on transmission group. Cluster sizes of the six most frequent non-B subtypes were pulled together. The colour represents: MSM (green) and HETs (blue) molecular transmission clusters. y-axis: frequency; x-axis: cluster size. Abbreviations: 4.5% D 80% BS, ≤4.5 % genetic distance and ≥80% bootstrap; HET, heterosexual contact; MSM, men who have sex with men.

Table 1. MSM characteristics and associations with subtype

Variables		B (N=4,698)	Non-B* (N=416)	Total (N=5,114)
Age at diagnosis (years)	Median [IQR]	35.0 [28.0,42.0]	36.0 [29.0,46.0]	35.0 [29.0,43.0]
Cohort center	Zurich	2,287 (48.7%)	201 (48.3%)	2,488 (48.7%)
	Non-Zurich[1]	2,411 (51.3%)	215 (51.7%)	2,626 (51.3%)
Ethnicity	White	4,216 (89.7%)	325 (78.1%)	4,541 (88.8%)
	Black	53 (1.1%)	18 (4.3%)	71 (1.4%)
	Hispanic	203 (4.3%)	17 (4.1%)	220 (4.3%)
	Asian	102 (2.2%)	40 (13.2%)	157 (3.1%)
Geographic region of origin	Eastern and Southern Africa	10 (0.2%)	45(1.2%)	15 (0.3%)
	Eastern and Southern Asia	81 (1.7%)	53 (12.7%)	134 (2.6%)
	Eastern Europe	64 (1.4%)	13 (3.1%)	77 (1.5%)
	Latin America	261 (5.6%)	21 (5.0%)	282 (5.5%)
	North America and Oceania	62 (1.3%)	1 (0.2%)	63 (1.2%)
	Unknown	4 (0.1%)	0 (0.0%)	4 (0.1%)
	Western and Central Africa	9 (0.2%)	10 (2.4%)	19 (0.4%)
	Western Asia and North Africa	36 (0.8%)	8 (1.9%)	44 (0.9%)
	Western, Northern and Southern Europe	4,171 (88.8%)	305 (73.3%)	4,476 (87.5%)
CD4 count at diagnosis (cells/µL)	Median [IQR]	377 [212,561]	406 [237,560]	380 [216,561]
Viral load (Log10 copies/ml)	Median [IQR]	3.93 [3.42,4.41]	4.02 [3.42,4.60]	3.94 [3.42,4.42]

NOTE all numbers are no. (%) unless otherwise stated, MSM only (N=5,114).

* Non-B: A, C, F, G, CRF01_AE, and CRF02_AG.

[1] Non-Zurich: Basel, Bern, Geneva, Lausanne, Lugano, and St. Gallen.

Abbreviations: IQR, interquartile range; MSM, men who have sex with men

Table 2. Characteristics of MSM Infected With the 6 Most Frequently Observed Non-B subtype

Variables		CRF01_AE (N=180)	CRF02_AG (N=57)	A (N=60)	C (N=44)	F (N=58)	G (N=17)	Total (N=416)
Age at diagnosis (years)	Median [IQR]	36.0 [29.0,46.0]	37.0 [31.0,44.0]	35.5 [27.8,43.0]	34.5 [30.0,45.3]	34.5 [28.5,42.0]	38.0 [34.0,42.0]	36.0 [29.0,46.0]
Cohort center	Zurich	108 (60.0%)	16 (28.1%)	24 (40.0%)	15 (34.1%)	32 (55.2%)	6 (35.3%)	201 (48.3%)
	Non-Zurich[1]	72 (40.0%)	41 (71.9%)	36 (60.0%)	29 (65.9%)	26 (44.8%)	11 (64.7%)	215 (51.7%)
Ethnicity	White	128 (71.1%)	46 (80.7%)	53 (88.3%)	35 (79.5%)	47 (81.0%)	16 (94.1%)	325 (78.1%)
	Black	2 (1.1%)	6 (10.5%)	1 (1.7%)	4 (9.1%)	4 (6.9%)	1 (5.9%)	18 (4.3%)
	Hispanic	3 (2.7%)	4 (7.0%)	3 (5.0%)	2 (4.5%)	5 (8.6%)	0 (0.0%)	17 (4.1%)
	Asian	46 (25.6%)	1 (1.8%)	3 (5.0%)	3 (6.8%)	2 (3.4%)	0 (0.0%)	55 (13.2%)
Geographic region of origin	Eastern and Southern Africa	0 (0.0%)	0 (0.0%)	1 (1.7%)	4 (9.1%)	0 (0.0%)	0 (0.0%)	5 (1.2%)
	Eastern and Southern Asia	35 (25.0%)	1 (1.8%)	3 (5.0%)	3 (6.8%)	1 (1.7%)	0 (0.0%)	53 (12.7%)
	Eastern Europe	3 (1.7%)	3 (5.3%)	2 (3.3%)	1 (2.3%)	3 (5.2%)	1 (5.9%)	13 (3.1%)
	Latin America	4 (2.2%)	2 (3.5%)	2 (3.3%)	5 (11.4%)	8 (13.8%	0 (0.0%)	21 (5.0%)
	North America and Oceania	1 (0.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.2%)
	Unknown	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	Western and Central Africa	1 (0.6%)	5 (8.8%)	0 (0.0%)	0 (0.0%)	3 (5.2%)	1 (5.9%)	10 (2.4%)
	Western Asia and North Africa	1 (0.6%)	2 (3.5%)	3 (5.0%)	1 (2.3%)	1 (1.7%)	0 (0.0%)	8 (1.9%)
	Western, Northern and Southern Europe	125 (69.4%)	44 (77.2%)	49 (81.7%)	30 (68.2%)	42 (72.4%)	15 (88.2%)	305 (73.3%)

CD4 count at diagnosis (cells/µL)	Median [IQR]	381 [171,552]	441 [291,623]	398 [284,536]	429 [270,560]	408 [265,580]	504 [265,741]	406 [237,560]
Viral load (Log10 copies/ml)	Median [IQR]	3.92 [3.41,4.47]	4.18 [3.58,4.82]	4.25 [3.60,4.80]	4.02 [3.21,4.38]	3.95 [3.18,4.54]	4.21 [3.75,4.58]	4.02 [3.42,4.60]

NOTE all numbers are no. (%) unless otherwise stated, MSM only (N=5,114)

[1] Non-Zurich: Basel, Bern, Geneva, Lausanne, Lugano, and St. Gallen.

Abbreviations: IQR, interquartile range; MSM, men who have sex with men.

Figure 1

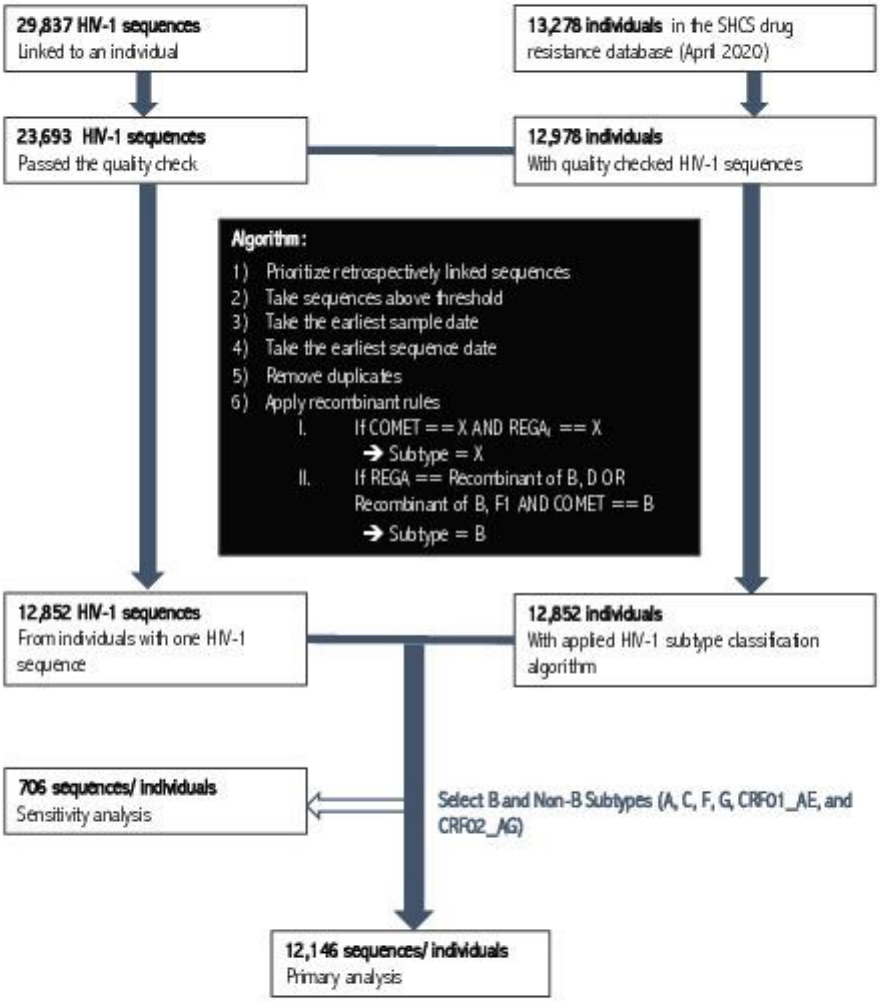

Figure 2

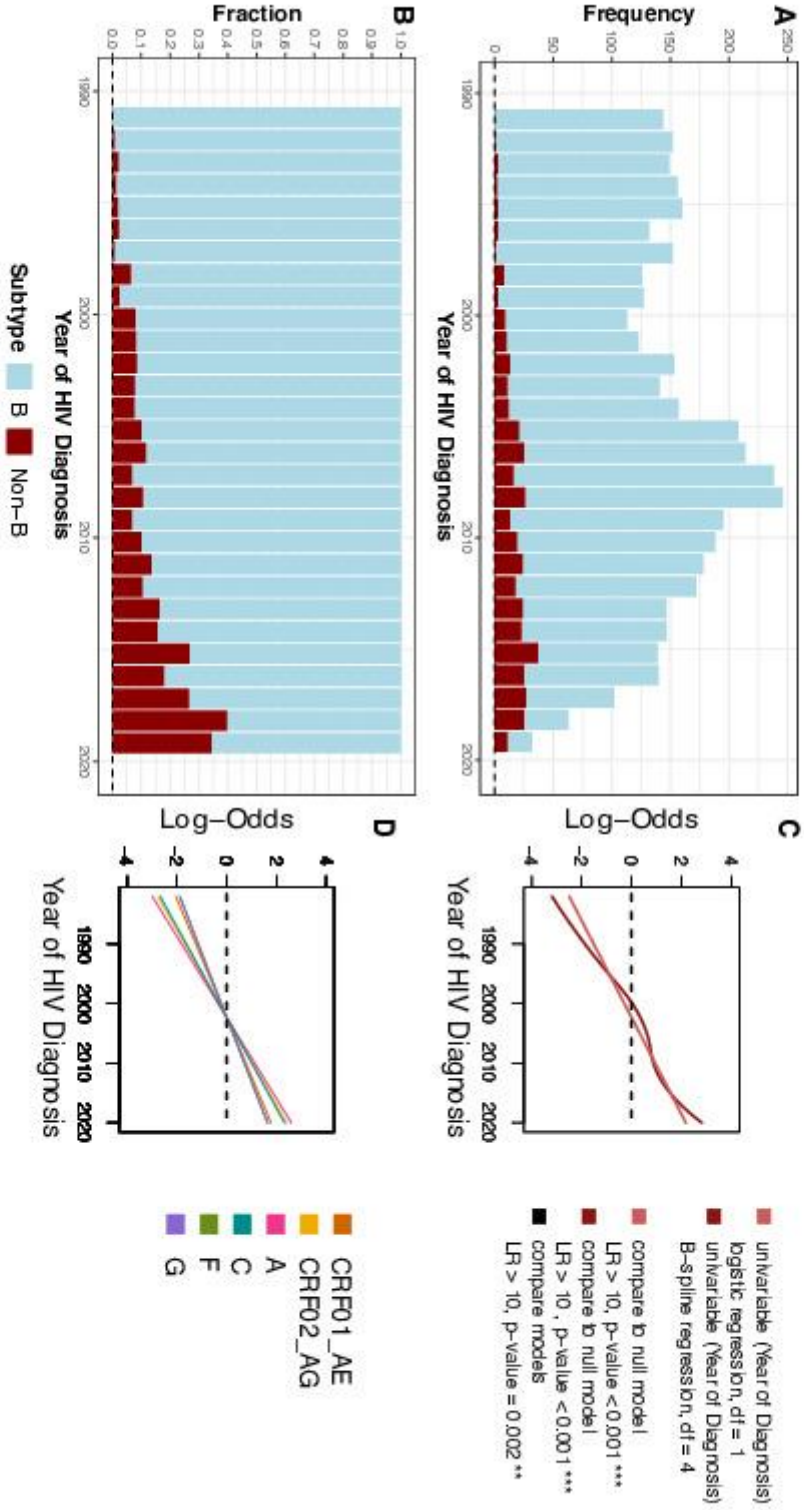

Figure 3

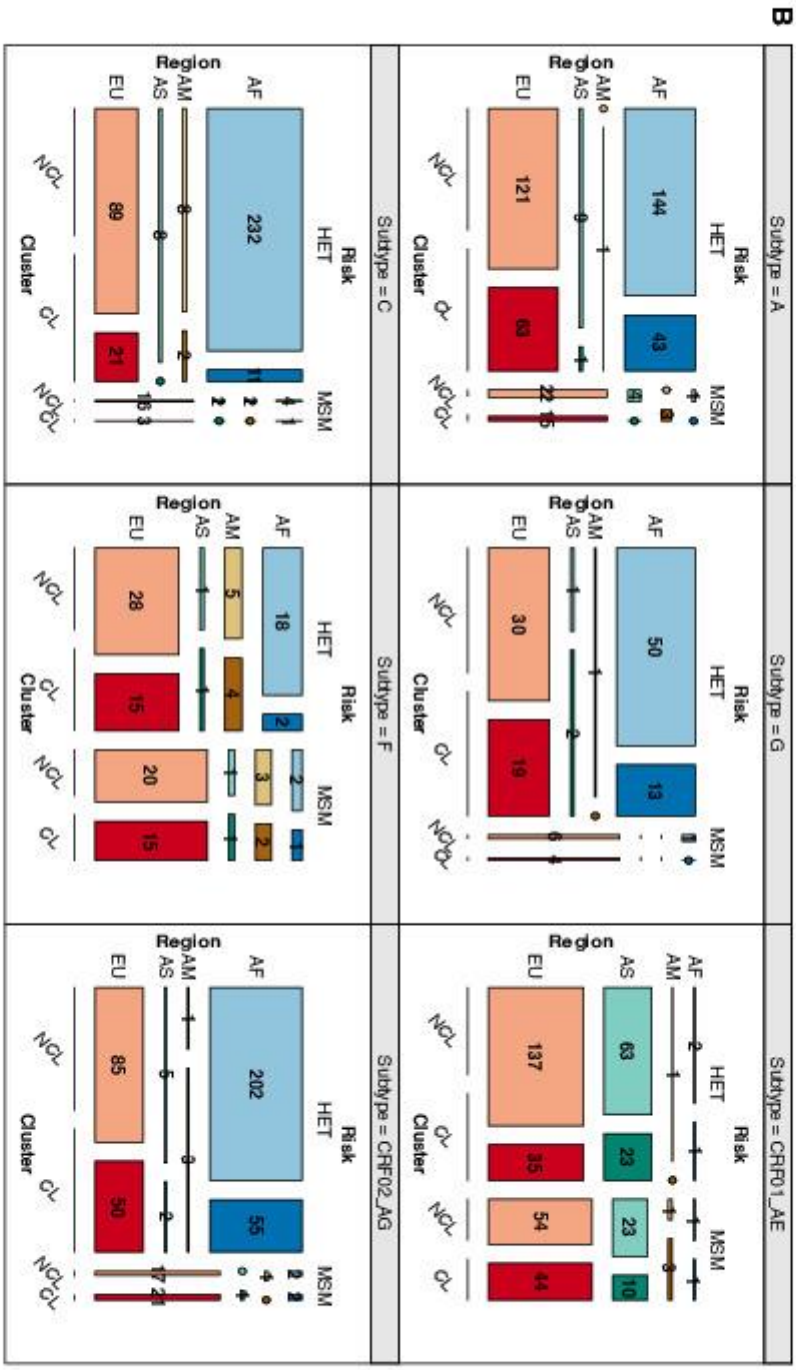

Figure 4