JAIDS Journal of Acquired Immune Deficiency Syndromes Publish Ahead of Print
DOI: 10.1097/QAI.0000000000002949

Similar but different: Integrated phylogenetic analysis of Austrian and Swiss HIV-1 sequences reveal differences in transmission patterns of the local HIV-1 epidemics

Katharina Kusejko[1,2], Nadine Tschumi[3,4], Sandra E. Chaudron[1,2], Huyen Nguyen[1,2], Manuel Battegay[5], Enos Bernasconi[6], Jürg Böni[2], Michael Huber[2], Alexandra Calmy[7], Matthias Cavassini[8], Alexander Egle[9], Katharina Grabmeier-Pfistershammer[10], Bernhard Haas[11], Hans Hirsch[12], Thomas Klimkait[12], Angela Öllinger[13], Matthieu Perreau[14], Alban Ramette[15], Baharak Babouee Flury[15,17], Mario Sarcletti[16], Alexandra Scherrer[1], Patrick Schmid[17], Sabine Yerly[7], Robert Zangerle[16], Huldrych F. Günthard[1,2], Roger D. Kouyos[1,2], the Swiss HIV Cohort Study and the Austrian HIV Cohort Study

[1]*Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland*

[2]*Institute of Medical Virology, University of Zurich, Zurich, Switzerland*

[3]*Department of Medicine, Swiss Tropical and Public Health Institute, Basel, Switzerland.*

[4]*University of Basel, Basel, Switzerland.*

[5]*Division of Infectious Diseases and Hospital Epidemiology, University Hospital Basel, University of Basel, Basel, Switzerland*

[6]*Division of Infectious Diseases, Regional Hospital Lugano, University of Geneva and University of Southern Switzerland, Lugano, Switzerland*

[7]*Laboratory of Virology and Division of Infectious Diseases, Geneva University Hospital, University of Geneva, Geneva, Switzerland*

[8]*Division of Infectious Diseases, Lausanne University Hospital, Lausanne, Switzerland*

[9]*Department of Internal Medicine III with Haematology, Medical Oncology, Haemostaseology, Infectiology and Rheumatology, Oncologic Center, Paracelsus Medical University, Salzburg, Austria.*

[10]*Department of Dermatology, Medical University Vienna, Vienna, Austria*

[11]*Institute of Hospital Hygiene and Microbiology, Styrian Hospital Corporation – The Styrian Healthcare Company, Graz, Austria*

[12]*Molecular Virology, Department of Biomedicine–Petersplatz, University of Basel, Basel, Switzerland*

[13]*Department of Dermatology, Kepler University Hospital, Linz, Austria*

[14]*University of Lausanne, Lausanne, Switzerland*

[15]*Department of Infectious Diseases, Bern University Hospital, University of Bern, Bern, Switzerland*

[16]*Department of Dermatology, Venereology and Allergology, Medical University of Innsbruck, Innsbruck, Austria*

[17]*Division of Infectious Diseases, Cantonal Hospital St Gallen, St. Gallen, Switzerland*

Corresponding author: Katharina Kusejko, PhD Division of Infectious Diseases and Hospital Epidemiology University Hospital Zürich, Rämistrasse 100, CH-8091 Zürich +41 43 253 0188, katharina.kusejko@usz.ch

Conflicts of Interest and Source of Funding

This work was financed in the framework of the Swiss HIV Cohort Study supported by the Swiss National Science Foundation (SNF) (Grant # 177499 and 201369), by SNF grant 179571 (to H.F. Günthard), by SNF Grant BSSGI0_155851 to R. Kouyos, the Swiss HIV Research Foundation, and the Yvonne Jacob Foundation (to H.F. Günthard). The institution of EB received fees for his participation to advisory boards and as travel grants from Gilead Sciences, MSD, ViiV Healthcare, Pfizer and Abbvie.

Part of these results were presented at the 17th European AIDS Conference in Basel, Switzerland, November 6-9th 2019.

Abstract

Objectives

Phylogenetic analyses of two or more countries allow to detect differences in transmission dynamics of local HIV-1 epidemics beyond differences in demographic characteristics.

Methods

A maximum-likelihood phylogenetic tree was built using *pol*-sequences of the Swiss HIV Cohort Study (SHCS) and the Austrian HIV Cohort Study (AHIVCOS), with international background sequences. Three types of phylogenetic cherries (clusters of size 2) were analyzed further: 1) Domestic cherries, 2) International cherries and 3) SHCS/AHIVCOS-cherries. Transmission group and ethnicities observed within the cherries were compared to the respective distribution expected from a random distribution of patients on the phylogeny.

Results

The demographic characteristics of the AHIVCOS (included patients: 3’141) and the SHCS (included patients: 12’902) are very similar. In the AHIVCOS, 36.5% of the patients were in domestic cherries, 8.3% in international cherries, and 7.0% in SHCS/AHIVCOS cherries. Similarly, in the SHCS, 43.0% of the patients were in domestic cherries, 8.2% in international cherries, and 1.7% in SHCS/AHIVCOS cherries. While international cherries in the SHCS were dominated by heterosexuals (HET) with MSM being underrepresented, the opposite was the case for the AHIVCOS. In both cohorts, cherries with one patient belonging to the transmission group intravenous drug user (IDU) and the other one non-IDU were underrepresented.

Conclusion

In both cohorts, international HIV transmission plays a major role in the local epidemics, mostly driven by MSM in the AHIVOS, and by HET in the SHCS, highlighting the importance of international collaborations to understand global HIV transmission links on the way to eliminate HIV.

Keywords: Epidemiology, HIV, phylogenetics, transmission patterns

Introduction

Combination antiretroviral therapy (cART) cannot cure Human Immunodeficiency Virus (HIV) infection, but has the potential to curb the HIV epidemic, as individuals under successful cART are not infectious (1–3). In resource-rich countries, there is almost universal HIV treatment available for people living with HIV (PLWH). Nevertheless, there is still ongoing HIV transmission, driven by key populations, with around 11% decrease in the number of new HIV infections in 2010-2020 in Western and Central Europe and North America (4). Reasons why a stronger decline of new HIV infections has not been achieved yet are manifold, including delayed diagnosis of PLWH and ongoing national and international transmission. Transmission patterns differ between countries due to differences in population structure, policies, culture, and the level of influence

by global HIV epidemics, i.e., travelling and migration. It is crucial to understand the local transmission network to inform policy makers about weaknesses in the cascade of care, education, and awareness about HIV in their own country.

Phylogenetic methods have been widely used to help understanding local transmission patterns of HIV epidemics in different countries and regions (5–11). So far, there is no consensus in how to build a phylogeny or define transmission clusters, making it difficult to compare different studies (12,13). Such a comparison might however be particularly useful in the case of neighboring countries with similar HIV epidemics, e.g., to quantify the impact of different public health decisions on certain HIV transmission dynamics. In the case of an active exchange and commuting between neighboring countries, a comparison of results from phylogenetic analyses might be complicated by a potential non-negligible mutual impact.

Switzerland and Austria are two resource-rich neighboring countries of similar population size and culture. The characteristics of the respective HIV epidemics are comparable: The numbers of new HIV diagnoses were 425 and 421 in Switzerland, and 323 and 336 (plus 74 and 94 anonymous) in Austria, in 2018 and 2019, respectively (14,15). The densely sampled drug resistance sequence data base of the Swiss HIV Cohort Study (SHCS) was used in several previous projects to analyze key aspects of the HIV epidemics in Switzerland (5,6,16–19). So far, the drug resistance sequence data base of the Austrian HIV Cohort Study (AHIVCOS) in Austria was not used for a nation-wide phylogenetic analysis to study local transmission patterns, but subsets were used for collaborations (20,21).

Our main aim is to compare transmission dynamics of the local epidemics of Austria and Switzerland, including a quantification of infections occurring outside the country and between these two countries.

Methods

The Cohorts

The Swiss HIV Cohort Study (SHCS), launched in 1988, is a prospective, multi-center cohort study enrolling adult PLWH in Switzerland. The SHCS is a nation-wide cohort with seven centers and represents over 70% of people on cART in Switzerland (22,23). The SHCS drug resistance database contains HIV partial polymerase (*pol)* sequences from around 65% of patients across all centers. The AHIVCOS was initiated in 2001 and represents about 74% of people currently receiving cART in nine centers in Austria (24), with HIV sequences available from around one third of the patients. All patients gave informed consent for participation in the SHCS or AHIVCOS. Detailed information about patient characteristics in these two cohorts can be found here: SHCS (25) and AHIVCOS (26).

Definitions

HIV transmission group was defined as the most likely source of HIV infection self-reported by the patient: men who have sex with men (MSM), heterosexual contacts (HET), intravenous drug use (IDU), and other or unknown transmission route. HIV subtypes for descriptive purposes were determined using the Context-based Modeling for Expeditious Typing (COMET) tool for classification of HIV-1 subtypes (27).

Construction of the phylogeny

We compared SHCS and AHIVCOS sequences, without prior subtyping, to all non-Swiss and non-Austrian sequences from the international Los Alamos (LA) data base using *Basic Local Alignment Search Tool* (BLAST, https://blast.ncbi.nlm.nih.gov/Blast.cgi), (Los Alamos download March 2019). Non-Swiss and non-Austrian LA sequences with at least 90% identity with one of the cohort sequences were selected. All SHCS, AHIVCOS and LA sequences were aligned against the reference genome HXB2 (accession number: K03455.1) using *Multiple Sequence Comparison by Log-Expectation* (MUSCLE). Nucleotide positions for the most common cART drug resistance mutations, based on the Stanford and International Antiviral Society USA drug resistance mutations list (28,29), were deleted to avoid a bias introduced by cART-driven convergent evolution. We built a maximum-likelihood phylogenetic tree using the generalized time-reversible model of nucleotide evolution and the CAT approximation for rate variation across sites of *FastTree* (30–32). This approach of building a phylogeny was used and validated in other SHCS projects (18,33).

Extraction cherries

We extracted all monophyletic pairs (henceforth called “cherries”) with at least one patient being in the SHCS or AHIVCOS, using the tree package *Analyses of Phylogenetics and Evolution* (APE) (34). Only cherries with a maximal cophenetic distance of 0.045 were considered

(17,35,36). We concentrated on three types of cherries: 1) *Domestic cherries* with both patients being in the same cohort, i.e., AHIVCOS or SHCS, termed AHIVCOS/AHIVCOS-cherries or SHCS/SHCS-cherries, 2) *International cherries* with one patient in the AHIVCOS or SHCS and the other patient from the Los Alamos (LA) data base, termed AHIVCOS/LA-cherries or SHCS/LA-cherries, and 3) *SHCS/AHIVCOS-cherries* with one patient being from the SHCS and the other patient being from the AHIVCOS**.**

Sensitivity Analysis

We repeated all analyses by stepwise narrowing the cophenetic distance threshold from 0.045 to 0.015 (37). Moreover, since the SHCS is more densely sampled as compared to the AHIVCOS, we performed simulation analyses by stepwise down-sampling the SHCS sequence data set: we trimmed the original phylogenetic tree by randomly cutting up to 80% of the SHCS tree tips and extracting new sets of cherries from the trimmed trees. For each given SHCS sample proportion, the procedure was repeated 100 times and the results were averaged (38).

Statistical analysis

HIV subtype distribution: We counted the number of different subtype cherries for the different cherry types (domestic, international and SHCS/AHIVCOS-cherries).

Age: For domestic cherries and SHCS/AHIVCOS-cherries, we determined the age difference of the patients based on the birth year of the patient and compared the two cohorts using the Wilcoxon test. For LA sequences, no age information was available.

Transmission group and ethnicity: We first determined the frequencies of the traits among SHCS and AHIVCOS patients in the tree. We then analyzed the three types of cherries: 1) *Domestic cherries*: Based on the occurrence of traits in the tree, we calculated the frequencies of traits one would expect by randomly pairing patients of the same cohort. We call the ratio of the expected and observed pairings of traits in the SHCS/SHCS-cherries and AHIVOCS/AHIVCOS-cherries *assortativity factor (AF)*. 2) *International cherries*: We compared the frequency of traits in the tree with the frequency of traits in international cherries. The ratio of these frequencies was then used to assess whether a trait is more or less common in SHCS/LA-cherries or AHIVOCS/LA-cherries than expected from the frequency of traits on the tree 3) *AHIVCOS/SHCS-cherries*: Similarly, we used the ratio of observed and expected distributions (based on the trait distribution in the whole tree) to assess whether traits are more or less common as compared to randomly pairing SHCS and AHIVCOS-patients. See **Supplementary material section S1** for the detailed description of the formulas. We used *MultinomCI* of the R package *DescTool (39)* for calculating confidence intervals of categorical variables, i.e., the distribution of the traits in the cherries, and with that derived confidence intervals for the ratios and AFs.

Results

Patient characteristics and number of cherries

We included 3141 AHIVCOS and 12902 SHCS patients in the phylogenetic tree. Of the 188917 background sequences downloaded from the Los Alamos data base, 7970 sequences were included in the phylogenetic tree. The majority of SHCS and AHIVCOS patients was male, of

white ethnicity and the transmission group MSM (**Table 1**). See **Table S1 and S2** for the characteristics of all patients in the cohorts as compared to patients in the phylogeny. We obtained 1148 (36.5%) AHIVCOS patients in domestic cherries and 260 (8.3%) in international cherries. Similarly, SHCS patients were predominantly in domestic cherries (5544, 43.0%) as compared to international cherries (1061, 8.2%). We obtained 220 SHCS/AHIVCOS-cherries, reflecting 1.7% of all SHCS and 7.0% of all AHIVCOS patients. We re-calculated the number of different cherry types for different distance thresholds and sampling densities of the SHCS: With an SHCS sample density of 0.5, i.e., randomly selecting around 50% of the SHCS sequences, we obtain a similar percentage of domestic SHCS cherries as compared to domestic AHIVCOS cherries (37.4% vs 36.4%) (**see Table S3 and S4**). At the same time, after down-sampling the SHCS to 50% of the sequences, the percentage of international SHCS cherries increases to 11.1%, compared to 9% international cherries in the AHIVCOS, pointing towards a similar fraction of international cherries in the SHCS and the AHIVCOS (**see Table S5 and S6**). Moreover, the fraction of AHIVCOS patients in AHIVCOS/SHCS-cherries is higher as compared to SHCS patients in AHIVCOS/SHCS-cherries, even after down-sampling 50% of the SHCS sequences (**see Table S7 and S8**). Of note, the total number of AHIVCOS/SHCS-cherries, i.e., 220, is higher than the number of SHCS/LA-cherries with the LA sequence from the United States, the country with most links to the SHCS (189 SHCS/LA-cherries with the LA sequence from the United States). **See Table S9** for the countries of origin of the LA sequences in all AHIVCOS/LA- and SHCS/LA-cherries.

Subtype distribution and age difference

The fraction of HIV subtype B cherries was highest in SHCS/SHCS-cherries (2230, 82.5%), followed by AHIVCOS/SHCS-cherries (155, 71.1%) and AHIVCOS/AHIVCOS-cherries (388, 69.7%). In the case of international cherries, the subtype distribution between SHCS and AHIVCOS cherries was similar, except for subtype 02_AG and F (see **Figure 1A**). The fraction of subtype B cherries was lower for both cohorts (AHIVCOS/LA-cherries: 143 (55%), SHCS/LA-cherries: 570 (53.7%)) as compared to the fraction of subtype B participants in the cohorts. The median age difference in SHCS/SHCS-cherries as well as AHIVCOS/AHIVCOS-cherries was 7 years (QR = 3-13), with no significant difference in the age difference (p = 0.51). The median age difference in AHIVCOS/SHCS-cherries was 8 years (IQR = 3-14), see **Figure 1B.**

Transmission group and ethnicity of domestic cherries

In both cohorts, MSM/MSM, IDU/IDU and male HET/female HET cherries were overrepresented, i.e., they were more frequent than expected by randomly pairing patients in the cohorts. This corresponds to an AF greater than one for these pairs (see **Methods**). IDU/IDU-cherries were most assortative (AHIVCOS AF = 4.24, SHCS AF = 3.76), followed by female HET/male HET-cherries (AHIVCOS AF = 2.71, SHCS AF = 2.27) and MSM/MSM-cherries (AHVICOS AF = 2.00, SHCS AF = 2.01) (see **Figure 2** for all traits). Of note, IDU/non-IDU cherries were more common in the SHCS (AF = 0.52) as compared to the AHIVCOS (AF = 0.38). The assortativity regarding white ethnicity was similar in both cohorts (AHIVCOS AF = 1.14, SHCS AF = 1.21). All differences between SHCS and AHIVCOS domestic cherries were stable with respect to down-sampling and varying the cophenetic distance threshold: In the case of IDU/IDU-cherries, the AF ranged between 3.5 and 5.2, indicating a clear over-representation of this cherry-type, and still higher as compared to MSM/MSM-cherries (range 2.0-2.6) (**Figures S3 and S4)**. The

situation was less clear in the case of female HET/female HET-cherries with range 0.7 to 1.4, but the AF was clearly higher in the SHCS regardless of the distance threshold and SHCS sample density (**Figure S5**). The AF concerning ethnicity was very stable for both cohorts (**Figure S6**).

Characteristics of international cherries

In the SHCS, international cherries were dominated by HET (male HET: 19.5%, ratio = 1.18, female HET: 25.9%, ratio = 1.34), while MSM were not overrepresented (40.3%, ratio = 1.01). In contrast, in the AHIVCOS, international cherries were dominated by MSM (48.5%, ratio = 1.12), while HET were only slightly overrepresented (male HET: 18.8%, ratio = 1.05, female HET: 18.5%, ratio = 1.04). In both cohorts, IDU were underrepresented in international cherries, in the SHCS even more as compared to the AHIVCOS (AHIVCOS ratio: 0.56, SHCS ratio: 0.46). Similarly, in both cohorts patients of white ethnicity were less present in international cherries as would be expected from the cohort distribution (AHIVCOS ratio: 0.90, SHCS ratio: 0.83). Contrariwise, patients of black and Asian ethnicity were more frequent in international cherries as compared to the frequency in the cohort (see **Figure 3**). These results were stable in our sensitivity analysis, with the ratio for IDU being below 1 (under-represented) and for MSM above 1 (over-represented), see **Figure S7 and S8**. In the case of ethnicity, however, the ratio was approaching 1 for a very low distance, most likely due to the low sample size for non-white patients in international cherries of low distance (see **Figure S9**).
In the AHIVCOS, the most frequent link in international cherries were sequences from Germany (43, 11.7%), which was only the case in 84 (6.1%) of the SHCS sequences, with the USA having most international links to the SHCS (189, 13.7%). See Table S7 for all countries in SHCS/LA- and AHIVCOS/LA-cherries (see **Table S9**).

SHCS/AHIVCOS-cherries

Of the 220 SHCS/AHIVCOS-cherries, 57 (25.9%) were MSM/MSM-cherries, followed by 21 (9.5%) male HET/female HET-cherries and 18 (8.1%) male HET/male HET-cherries. The only transmission group combination which was significantly overrepresented, as compared to the distribution expected by randomly pairing SHCS and AHIVCOS patients, were MSM/MSM-cherries (ratio = 1.49). See **Table 2** for all transmission group combinations and ethnicities in SHCS/AHIVCOS-cherries.

Discussion

Building a phylogenetic tree including Austrian, Swiss and international HIV-1 sequences revealed interesting insights into national and international transmission patterns. In both cohorts, the SHCS and AHIVCOS, around 50% of all patients were in a phylogenetic cherry. In the AHIVCOS, 30% of the patients in a cherry had a link to a non-Austrian sequence (16% international Los Alamos data base and 13.5% Switzerland). Similarly, a significant amount of SHCS patients had links to non-Swiss sequences (15.5% international Los Alamos sequences and 3.2% Austria). Given that the Los Alamos background sequence data base is less representative of the global HIV-1 epidemic as compared to the two local cohorts, we can assume that the fraction of international, i.e., non-Austrian and non-Swiss, sequences is under-represented. This means that in both countries, international links have a major impact on the local HIV-1 epidemics. This highlights the importance of transnational collaboration to understand the dynamics of the on-going HIV-1 epidemics. By combining the sequence data bases of the Austrian and Swiss cohorts, we were able

to compare transmission in the two local epidemics. Regarding links to the Los Alamos database, we could identify differences in transmission groups in the two countries: While in Austria international links are dominated by MSM, in Switzerland international links are over-represented by HET. This indicates differences in international HIV-1 transmission sources between Austria and Switzerland. The interpretation of our results is that in Austria, the HIV-1 epidemic among MSM is more influenced by international transmission, i.e., MSM being infected by their partner abroad, as compared to Switzerland. Of note, since we do not distinguish between nationalities in this project, our findings only reflect the amount of transmission events between patients registered in the local cohorts (SHCS or AHIVCOS) and patients somewhere else, but does not tell us anything about the role of immigrants in the respective countries, as immigrants are be part of the local cohorts too and hence count as domestic transmission. In Switzerland, it was shown before that the HIV-1 epidemic among HET is not self-sustained, indicating the major impact of international transmission and domestic transmission of other transmission groups in the case of HET (18). In both cohorts, few international links were found among IDU, indicating that HIV-1 transmission among IDU predominantly occurred within local transmission networks, in Switzerland even more than in Austria. This mostly domestic transmission dynamics of HIV-1 among IDU might have helped the successful prevention and virtual eradication of HIV transmission among IDU in Switzerland, and presumably also in Austria (19).

Combining Austrian and Swiss sequences into one phylogenetic tree allows to study and compare characteristics of the local epidemics. In both cohorts, cherries of the expected HIV-1 transmission group combinations, i.e., MSM/MSM, IDU/IDU and male HET/female HET were most assortative, in the AHIVCOS even more as compared to the SHCS. In the SHCS, IDU/non-IDU, i.e., IDU/MSM, IDU/male HET and IDU/female HET, were more frequent as compared to the AHIVCOS. This suggests that the overspill of the HIV epidemic among IDU to other transmission

groups was larger in Switzerland as compared to Austria. Interestingly, the assortativity factor of female HET/female HET pairs, the transmission group combination with a very small HIV transmission probability, was above 1 in both cohorts. Although not statistically significant, this is an indication of unsampled male HET in both cohorts, in the SHCS even more than in the AHIVCOS. In Switzerland as well as Austria, HIV testing is done routinely in pregnant women and hence female HET have a higher chance of being diagnosed. In addition, there might be more reluctance towards HIV testing in male HET as compared to female HET in general, as was observed for other countries (40).

Building a phylogenetic tree, extracting clusters from the tree as well as analyzing properties of clusters involve numerous modelling and parameter choices. To date, there is no consensus regarding the ideal way to construct an HIV phylogeny, neither for extraction of HIV transmission clusters (12,13). Comparing results from different publications, such as the fraction of international transmission links or properties of local transmission clusters, is hence rarely possible. To target HIV-1 prevention, it is important to understand where previous prevention campaigns are lagging behind, also in comparison to other countries. A comparison is only possible if the same methods were used to quantify the respective problem. Using the "HIV estimates accuracy tool" provided by the European Center for Disease Control (ECDC) (41,42), estimates concerning local HIV epidemics can be made, such as estimates about the percentage of undiagnosed HIV-infected people. With this tool, it is hence possible to compare the WHO 90-90-90 goals between different countries, using the same method, differences in the sample density and missing data are taken into account. To our knowledge, no such tool is available including phylogenetic analyses. Hence, our paper showcases how combined phylogenies could be used to understand, compare, and quantify transmission patterns of local epidemics.

Ragonnet, Shilaih et al. (36) performed a phylogenetic study to compare transmission patterns of the HIV epidemics in Switzerland and the UK by applying the same methods on the drug resistance data base of the SHCS and the UK HIV resistance data base. They found similar characteristics of the Swiss and UK local HIV epidemics after correcting for differences in sample size. In our comparison of Switzerland and Austria, in addition to analyzing the local epidemics, we investigate properties of international links with the Los Alamos data base, as well as properties of potential overlaps between the Swiss and Austrian HIV epidemic. By contrast to Ragonnet, Shilaih et al, our approach however necessitates the transfer of sequences from one cohort to the other cohort.

Our study has several strengths and limitations. One strength is that for showcasing the use of a combined phylogeny, we were able to include Switzerland and Austria, two neighboring countries of comparable size, culture, and similar basic HIV epidemics. One major limitation is inevitable to all phylogenetic analyses: throughout the construction of the tree and extraction of phylogenetic clusters, a multitude of parameters need to be chosen. In this project, we concentrated on clusters of size 2, “cherries”, based on the tree topology and an additional genetic distance criterion. This simplification on the one hand disregards a lot of sequences potentially closely clustered with these cherries, on the other hand provides an intuitive interpretation of the transmission patterns through the proposed assortative factor. However, previous work has shown that transmission characteristics of cherries are a good proxy of characteristics of larger clusters (43).

Extensive sensitivity analyses were performed to understand the impact of sampling density and distance threshold for the cherry definition on our results and performed numerous simulations (**Section S6**). In addition, we performed a sensitivity analysis by re-building the phylogenetic tree with subtype B sequences only. The distribution of characteristics is very similar, indicating robustness of our results (see **Section S7.1**). Similar, we

rebuilt the phylogenies of Austrian and Swiss sequences separately, again with robust results (see **Section S7.2**). Further, the SHCS has sequenced a large number of samples retrospectively for research purposes. As a sensitivity analysis, we re-built the phylogenetic tree with retrospectively sequenced samples removed, and obtain similar results as presented in the main analysis (see **Section S8**). Another limitation is that sequences included in the Los Alamos data base might be biased and do not reflect the trait distribution of the respective countries. Given the higher median genetic distance in international cherries as compared to domestic cherries, we conclude that the international cherries observed in our project most likely do not reflect direct transmission events, but rather two sequences on a transmission chain with few intermediate transmission events. Hence, the country of origin of the Los Alamos sequences might not reflect the countries of the direct transmission events. One of the main characteristics we study is the HIV transmission route, which is self-reported by the patients, and potentially underestimates the assortativeness of patients from the same transmission group.

In summary, the local epidemics of Austria and Switzerland are of remarkable similarity, with only minor differences observed in transmission patterns. In both cohorts, international transmission links play a major role, mainly driven by MSM in Austria and HET in Switzerland. This underlines the importance of international collaborations to understand the links between HIV epidemics in different areas on the way to eliminate HIV. The overrepresentation of female HET cherries indicates missing HIV diagnoses of male HET in both cohorts, calling for tailored HIV testing strategies among male HET. Moreover, the underrepresentation of IDU in international cherries in both cohorts highlight the success of the virtual elimination of HIV transmission among IDU.

Word count: 3669

Ethical statement

AHIVCOS: Approval for this study was obtained from the local ethical committees of all participating centres: ethics committee of the Vienna Medical University (No. 898/ 2010), of the Salzburg Federal Government (No. 1159/ 2010), of the Graz Medical University (No. 21-431/ 2010), of the Innsbruck Medical University (No 283/4.4/ 2009), of the Upper Austria Federal Government (No C-3-10/ 2010) and of the Carynthian Federal State (No A-13-11/2011). Written informed consent was given by the patients for their information to be stored in the hospital database and used for research.

SHCS: The SHCS was approved by the local ethical committees of the participating centres: Ethikkommission beider **Basel** ("Die Ethikkommission beider Basel hat die Dokumente zur Studie zustimmend zur Kenntnis genommen und genehmigt."); Kantonale Ethikkommission **Bern** (21/88); Comité départemental d'éthique des spécialités médicales et de médecine communautaire et de premier recours, Hôpitaux Universitaires de **Genève** (01–142); Commission cantonale d'éthique de la recherche sur l'être humain, Canton de **Vaud** (131/01); Comitato etico cantonale, Repubblica e Cantone **Ticino** (CE 813); Ethikkommission des Kantons **St. Gallen** (EKSG 12/003); Kantonale Ethikkommission **Zürich** (KEK-ZH-NR: EK-793), and written informed consent was obtained from all participants.

Data availability statement

The individual level datasets generated or analyzed during the current study do not fulfill the requirement for open data access: 1) The SHCS informed consent states that sharing data outside the SHCS network is only permitted for specific studies on HIV infection and its complications, and to researchers who have signed an agreement detailing the use of the data and biological samples; and 2) the data is too dense and comprehensive to preserve patient privacy in persons living with HIV. According to the Swiss law, data cannot be shared if data subjects have not agreed or data is too sensitive to share. Investigators with a request for selected data should send a proposal to the respective SHCS address (**www.shcs.ch/contact**). The provision of data will be considered by the Scientific Board of the SHCS and the study team and is subject to Swiss legal and ethical regulations, and is outlined in a material and data transfer agreement.

References

1. Rodger AJ, Cambiano V, Bruun T, Vernazza P, Collins S, Degen O, et al. Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study. Lancet Lond Engl. 2019 15;393(10189):2428–38.

2. Rodger AJ, Cambiano V, Bruun T, Vernazza P, Collins S, van Lunzen J, et al. Sexual Activity Without Condoms and Risk of HIV

Transmission in Serodifferent Couples When the HIV-Positive Partner Is Using Suppressive Antiretroviral Therapy. JAMA. 2016 Jul 12;316(2):171–81.

3. Bavinton BR, Pinto AN, Phanuphak N, Grinsztejn B, Prestage GP, Zablotska-Manos IB, et al. Viral suppression and HIV transmission in serodiscordant male couples: an international, prospective, observational, cohort study. Lancet HIV. 2018;5(8):e438–47.

4. UNAIDS. UNAIDS: GLOBAL COMMITMENTS, LOCAL ACTION: After 40 years of AIDS, charting a course to end the pandemic [Internet]. Available from: https://www.unaids.org/sites/default/files/media_asset/global-commitments-local-action_en.pdf

5. von Wyl V, Kouyos RD, Yerly S, Böni J, Shah C, Bürgisser P, et al. The role of migration and domestic transmission in the spread of HIV-1 non-B subtypes in Switzerland. J Infect Dis. 2011 Oct 1;204(7):1095–103.

6. Kouyos RD, von Wyl V, Yerly S, Böni J, Taffé P, Shah C, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. J Infect Dis. 2010 May 15;201(10):1488–97.

7. Hughes GJ, Fearnhill E, Dunn D, Lycett SJ, Rambaut A, Leigh Brown AJ, et al. Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. PLoS Pathog. 2009 Sep;5(9):e1000590.

8. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. PLoS Med. 2008 Mar 18;5(3):e50.

9. van de Vijver DAMC, Boucher CAB. Insights on transmission of HIV from phylogenetic analysis to locally optimize HIV prevention strategies. Curr Opin HIV AIDS. 2018 Mar;13(2):95–101.

10. Bezemer D, Cori A, Ratmann O, van Sighem A, Hermanides HS, Dutilh BE, et al. Dispersion of the HIV-1 Epidemic in Men Who Have Sex with Men in the Netherlands: A Combined Mathematical Model and Phylogenetic Analysis. PLoS Med. 2015 Nov;12(11):e1001898; discussion e1001898.

11. Castor D, Low A, Evering T, Karmon S, Davis B, Figueroa A, et al. Transmitted drug resistance and phylogenetic relationships among acute and early HIV-1-infected individuals in New York City. J Acquir Immune Defic Syndr 1999. 2012 Sep 1;61(1):1–8.

12. Junqueira DM, Sibisi Z, Wilkinson E, de Oliveira T. Factors influencing HIV-1 phylogenetic clustering. Curr Opin HIV AIDS. 2019 May;14(3):161–72.

13. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjörnsson J. Defining HIV-1 transmission clusters based on sequence data. AIDS Lond Engl. 2017 Jun 1;31(9):1211–22.

14. Zahlen zu HIV/Aids [Internet]. Aids-Hilfe Schweiz. [cited 2020 May 16]. Available from: https://www.aids.ch/de/haeufigste-fragen/hiv-aids/zahlen/

15. ”VIRUSEPIDEMIOLOGISCHE INFORMATION” NR. 03/20 [Internet]. [cited 2021 Jun 15]. Available from: https://www.virologie.meduniwien.ac.at/fileadmin/virologie/files/Epidemiologie/2020/0320.pdf

16. Duran Ramirez JJ, Ballouz T, Nguyen H, Kusejko K, Chaudron SE, Huber M, et al. Increasing Frequency and Transmission of HIV-1 Non-B Subtypes among Men Who Have Sex with Men in the Swiss HIV Cohort Study. J Infect Dis. 2021 Jul 14;jiab360.

17. Marzel A, Shilaih M, Yang W-L, Böni J, Yerly S, Klimkait T, et al. HIV-1 Transmission During Recent Infection and During Treatment

Interruptions as Major Drivers of New Infections in the Swiss HIV Cohort Study. Clin Infect Dis Off Publ Infect Dis Soc Am. 2016 Jan 1;62(1):115–22.

18. Turk T, Bachmann N, Kadelka C, Böni J, Yerly S, Aubert V, et al. Assessing the danger of self-sustained HIV epidemics in heterosexuals by population based phylogenetic cluster analysis. eLife. 2017 Sep 12;6.

19. Marzel A, Kusejko K, Weber R, Bruggmann P, Rauch A, Roth JA, et al. The Cumulative Impact of Harm Reduction on the Swiss HIV Epidemic: Cohort Study, Mathematical Model, and Phylogenetic Analysis. Open Forum Infect Dis. 2018 May;5(5):ofy078.

20. Paraskevis D, Beloukas A, Stasinos K, Pantazis N, de Mendoza C, Bannert N, et al. HIV-1 molecular transmission clusters in nine European countries and Canada: association with demographic and clinical factors. BMC Med. 2019 Jan 8;17(1):4.

21. Hoenigl M, Chaillon A, Kessler HH, Haas B, Stelzl E, Weninger K, et al. Characterization of HIV Transmission in South-East Austria. PLoS ONE [Internet]. 2016 Mar 11 [cited 2019 Feb 20];11(3). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4788428/

22. Swiss HIV Cohort Study, Schoeni-Affolter F, Ledergerber B, Rickenbach M, Rudin C, Günthard HF, et al. Cohort profile: the Swiss HIV Cohort study. Int J Epidemiol. 2010 Oct;39(5):1179–89.

23. Scherrer AU, Traytel A, Braun DL, Calmy A, Battegay M, Cavassini M, et al. Cohort Profile Update: The Swiss HIV Cohort Study (SHCS). Int J Epidemiol. 2021 Aug 7;dyab141.

24. 37th Report of the Austrian HIV Cohort Study [Internet]. AGES - Österreichische Agentur für Gesundheit und Ernährungssicherheit. [cited 2020 May 16]. Available from: https://www.ages.at/service/service-oeffentliche-gesundheit/berichte-folder-und-formulare/37th-report-of-the-

austrian-hiv-cohort-study/

25. about SHCS - Swiss HIV Cohort Study [Internet]. [cited 2022 Jan 31]. Available from: https://www.shcs.ch/157-about-shcs

26. Berichte & Folder der Öffentlichen Gesundheit [Internet]. AGES. [cited 2022 Feb 18]. Available from: https://www.ages.at/mensch/krankheit/oeffentliche-gesundheit-services/berichte-folder

27. Struck D, Lawyer G, Ternes A-M, Schmit J-C, Bercoff DP. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic Acids Res. 2014 Oct;42(18):e144.

28. Johnson VA, Calvez V, Gunthard HF, Paredes R, Pillay D, Shafer RW, et al. Update of the drug resistance mutations in HIV-1: March 2013. Top Antivir Med. 2013 Mar;21(1):6–14.

29. Bennett DE, Camacho RJ, Otelea D, Kuritzkes DR, Fleury H, Kiuchi M, et al. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. PloS One. 2009;4(3):e4724.

30. Lartillot N, Philippe H. A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process. Mol Biol Evol. 2004 Jun 1;21(6):1095–109.

31. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009 Jul;26(7):1641–50.

32. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS One. 2010 Mar 10;5(3):e9490.

33. Bachmann N, Turk T, Kadelka C, Marzel A, Shilaih M, Böni J, et al. Parent-offspring regression to estimate the heritability of an HIV-1 trait in a realistic setup. Retrovirology. 2017 May 23;14(1):33.

34. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004 Jan 22;20(2):289–90.

35. Brenner B, Wainberg MA, Roger M. Phylogenetic inferences on HIV-1 transmission: implications for the design of prevention and treatment interventions. AIDS Lond Engl. 2013 Apr 24;27(7):1045–57.

36. Ragonnet-Cronin ML, Shilaih M, Günthard HF, Hodcroft EB, Böni J, Fearnhill E, et al. A Direct Comparison of Two Densely Sampled HIV Epidemics: The UK and Switzerland. Sci Rep. 2016 19;6:32251.

37. Ragonnet-Cronin M, Hodcroft E, Hué S, Fearnhill E, Delpech V, Brown AJL, et al. Automated analysis of phylogenetic clusters. BMC Bioinformatics. 2013 Nov 6;14:317.

38. Kusejko K, Kadelka C, Marzel A, Battegay M, Bernasconi E, Calmy A, et al. Inferring the age difference in HIV transmission pairs by applying phylogenetic methods on the HIV transmission network of the Swiss HIV Cohort Study. Virus Evol [Internet]. 2018 Jul 1 [cited 2018 Sep 26];4(2). Available from: https://academic.oup.com/ve/article/4/2/vey024/5101427

39. DescTool, R package [Internet]. [cited 2021 Jul 20]. Available from: https://cran.r-project.org/web/packages/DescTools/DescTools.pdf

40. Camlin CS, Ssemmondo E, Chamie G, El Ayadi AM, Kwarisiima D, Sang N, et al. Men “missing” from population-based HIV testing: insights from qualitative research. AIDS Care. 2016;28 Suppl 3:67–73.

41. Rosinska M, Pantazis N, Janiec J, Pharris A, Amato-Gauci AJ, Quinten C, et al. Potential adjustment methodology for missing data and

reporting delay in the HIV Surveillance System, European Union/European Economic Area, 2015. Eurosurveillance. 2018 Jun 7;23(23):1700359.

42. European Centre for Disease Prevention and Control. HIV estimates accuracy tool [Internet]. Stockholm: ECDC; 2018. Available from: https://ecdc.europa.eu/en/publications-data/hiv-estimates-accuracy-tool

43. Kusejko K, Bachmann N, Chaudron SE, Nguyen H, Braun DL, Hampel B, et al. A Systematic Phylogenetic Approach to Study the Interaction of HIV-1 With Coinfections, Noncommunicable Diseases, and Opportunistic Diseases. J Infect Dis. 2019 Jun 19;220(2):244–53.

Figures

Figure 1A: Distribution of subtype in the different types of cherries: SHCS/SHCS, AHIVCOS/AHIVCOS, SHCS/LA, AHIVCOS/LA, AHIVCOS/SHCS

Figure 1B: The distribution of the age difference in the different cherry types: Domestic cherries (SHCS/SHCS, AHIVCOS/AHIVCOS) as well as AHIVCOS/SHCS cherries

Figure 2: Characteristics of domestic cherries: Frequency of cherries with different patient characteristics and the assortativity factor (AF), i.e., the observed frequency divided by the expected frequency when randomly pairing cohort patients.

Figure 3: Characteristics of international cherries: Frequency of cherries with different patient characteristics and the factor: observed frequency divided by the frequency in the cohort (expected frequency).

FIGURES AND TABLES

Table 1

		AHIVCOS	SHCS
Total		**3141**	**12902**
Cohort Center		Vienna: 1837 (58.5%)	Zürich: 4869 (37.7%)
		Linz: 487 (15.5%)	Lausanne: 1911 (14.8%)
		Graz: 354 (11.3%)	Geneva: 1762 (13.7%)
		Innsbruck: 235 (7.5%)	Bern: 1712 (13.3%)
		Salzburg: 228 (7.3%)	Basel: 1427 (11.1%)
			St Gallen: 833 (6.5%)
			Lugano: 388 (3.0%)
Sex	male	2375 (75.6%)	9272 (71.9%)
	female	766 (24.4%)	3630 (28.1%)
Birth year	median (IQR)	1972 [1964, 1981]	1965 [1959, 1972]
Registration year	median (IQR)	2009 [2003, 2013]	2001 [1996, 2009]
Sequence year	Median (IQR)	2010 [2007, 2014]	2003 [1998, 2009]
ART naïve at sequence date		2177 (69.3%)	8879 (68.8%)
Transmission group	MSM	1361 (43.3%)	5168 (40.1%)
	male HET	562 (17.9%)	2133 (16.5%)
	female HET	559 (17.8%)	2491 (19.3%)
	male IDU	335 (10.7%)	1643 (12.7%)
	female IDU	158 (5.0%)	887 (6.9%)
	male other	117 (3.7%)	357 (2.8%)
	female other	49 (1.6%)	223 (1.7%)
Ethnicity	white	2541 (80.9%)	9881 (76.6%)
	black	303 (9.6%)	1638 (12.7%)
	Hispanic	24 (0.8%)	397 (3.1%)
	Asian	80 (2.5%)	425 (3.3%)
HIV subtype	B	2031 (64.7%)	9574 (74.2%)
	CRF01_AE	189 (6.0%)	539 (4.2%)
	CRF02_AG	128 (4.1%)	508 (3.9%)
	A	268 (8.5%)	610 (4.7%)
	C	181 (5.8%)	558 (4.3%)
	F	106 (3.4%)	158 (1.2%)
	G	78 (2.5%)	300 (2.3%)
	D	31 (1.0%)	157 (1.2%)
	other or unknown	446 (14.2%)	1545 (12.0%)

Table 1: Basic characteristic of the study population; AHIVCOS: Austrian HIV Cohort Study, SHCS: Swiss HIV Cohort Study, IQR: interquartile range, MSM: men who have sex with men, HET: heterosexual, IDU: intravenous drug user

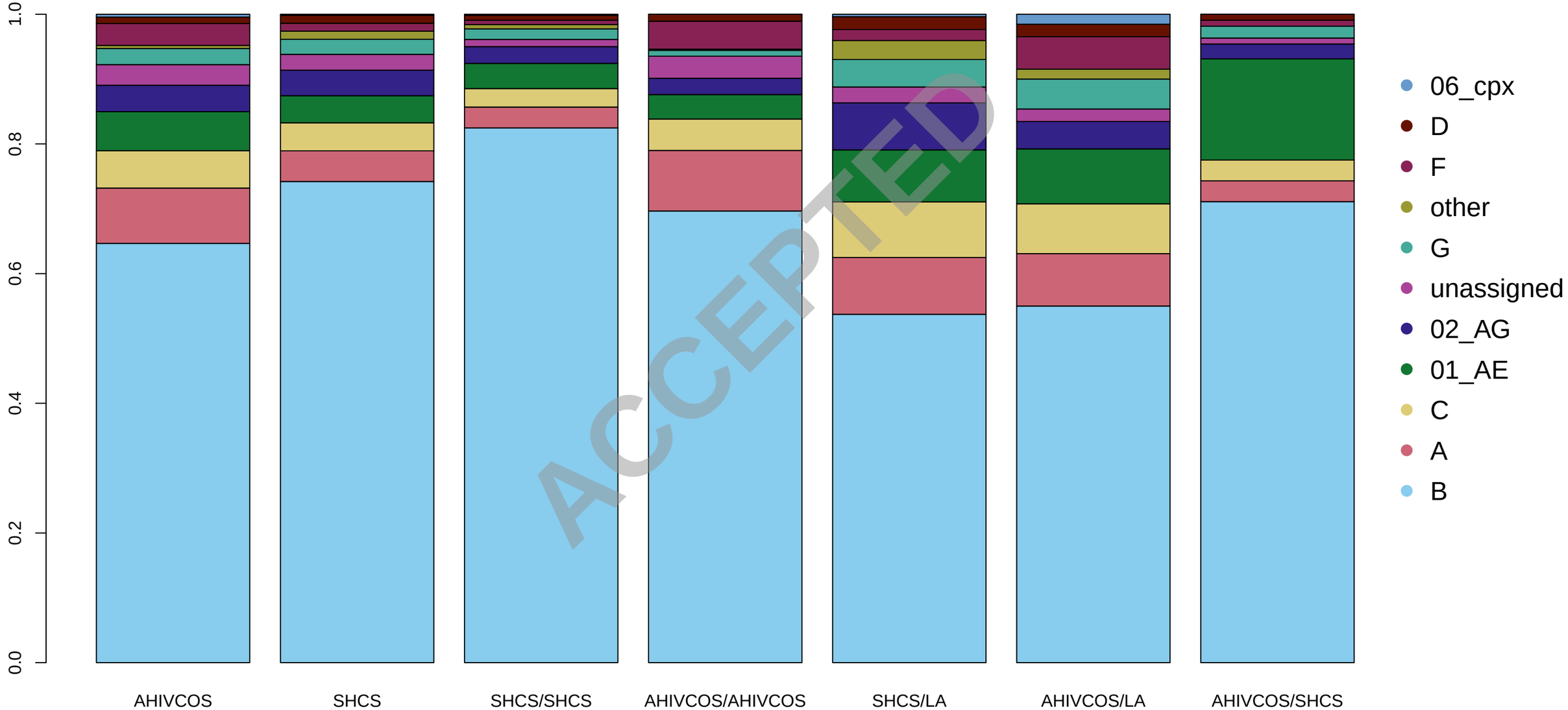
Table 2

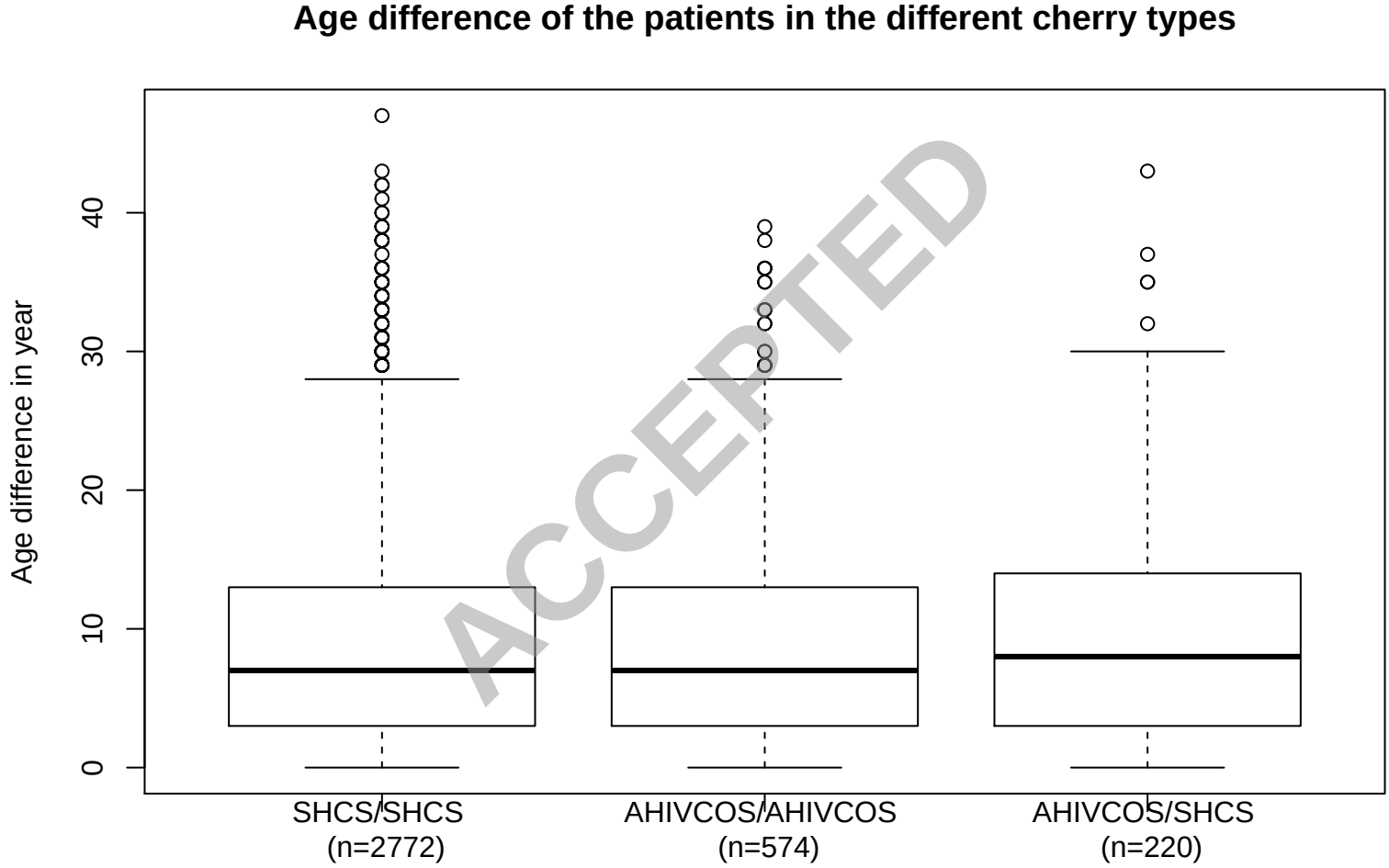
SHCS/ AHIVCOS	MSM	Male HET	Female HET	Male IDU	Female IDU	Male Other	Female Other
MSM	1.49 [1.18, 1.83]	0.32 [0, 1.12]	0.33 [0, 1.02]	0.49 [0, 1.54]	0.15 [0, 2.09]	1.52 [0, 6.33]	
Male HET	0.7 [0, 1.5]	2.77 [0.92, 4.72]	1.18 [0, 2.86]	0.8 [0, 3.33]	1.11 [0, 5.8]		2.94 [0, 21.6]
Female HET	0.45 [0, 1.26]	1.85 [0, 3.82]	1.46 [0, 3.13]	1.2 [0, 3.75]	0.37 [0, 5.09]	0.92 [0, 12.64]	1.48 [0, 20.24]
Male IDU	0.64 [0, 1.99]	1.29 [0, 4.56]	1.99 [0, 4.79]	1.67 [0, 5.92]	1.24 [0, 9.11]	1.54 [0, 21.1]	
Female IDU	0.45 [0, 3.32]		0.94 [0, 6.88]	4.26 [0, 13.27]	3.94 [0, 20.63]		
Male Other	1.22 [0, 5.09]	0.74 [0, 10.11]	1.26 [0, 9.29]	0.96 [0, 13.13]			
Female Other		3.52 [0, 25.9]	3.02 [0, 22.18]	2.29 [0, 31.34]			16.86 [0, 230.92]

SHCS/AHIVCOS	white	black	Asian	Hispanic	other
white	1.09 [1.01, 1.19]	0.35 [0, 0.94]	1.54 [0, 3.79]	0.73 [0, 3.15]	1.42 [0, 3.13]
black	0.31 [0, 1.12]	2.23 [0, 7.14]			1.08 [0, 15.43]
Asian	1.86 [0, 4.95]		21.67 [0, 93.37]		
Hispanic	0.78 [0, 11.06]				
other	1.06 [0, 2.34]		6.74 [0, 36.46]		

Table 2: Transmission group and ethnicities in AHIVCOS/SHCS-cherries: The factor of the expected and observed frequency of the different cherries, and confidence interval.

Distribution of subtypes in the cohorts and the cherries

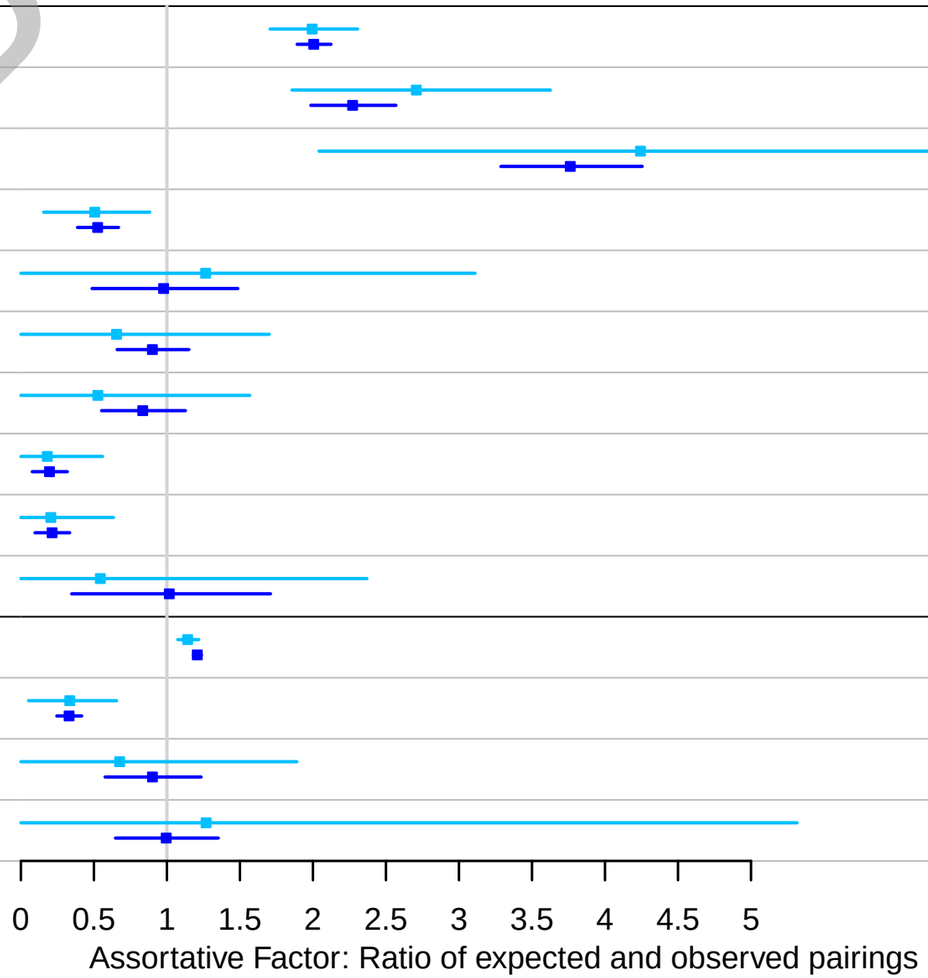




Characteristics of domestic cherries

AHIVCOS AF SHCS AF

		AHIVCOS observed	AHIVCOS expected	SHCS observed	SHCS expected
Transmission group	MSM/MSM	37.5%	18.8%	32.2%	16%
	femaleHET/maleHET	17.2%	6.4%	14.5%	6.4%
	IDU/IDU	10.5%	2.5%	14.5%	3.8%
	maleHET/MSM	7.8%	15.5%	7%	13.2%
	femaleHET/femaleHET	4%	3.2%	3.6%	3.7%
	femaleHET/IDU	3.7%	5.6%	6.8%	7.6%
	IDU/maleHET	3%	5.6%	5.4%	6.5%
	femaleHET/MSM	2.8%	15.4%	3%	15.5%
	IDU/MSM	2.8%	13.6%	3.4%	15.7%
	maleHET/maleHET	1.7%	3.2%	2.8%	2.7%
Ethnicity	white/white	74.7%	65.4%	70.9%	58.7%
	black/white	5.2%	15.6%	6.4%	19.4%
	Asian/white	2.8%	4.1%	4.5%	5%
	Hispanic/white	1.6%	1.2%	4.7%	4.7%



Characteristics of international cherries

AHIVCOS factor SHCS factor

		AHIVCOS observed	AHIVCOS expected	SHCS observed	SHCS expected
Transmission group	MSM	48.5%	43.3%	40.3%	40.1%
	female HET	18.5%	17.8%	25.9%	19.3%
	male HET	18.8%	17.9%	19.5%	16.5%
	IDU	8.8%	15.7%	9%	19.6%
Ethnicity	white	73.1%	80.9%	63.9%	76.6%
	black	13.5%	9.6%	23.1%	12.7%
	Asian	5.4%	2.5%	5.5%	3.3%
	Hispanic	1.5%	0.8%	4.2%	3.1%