Virus Evolution, 2024, **10(1)**, veae092
DOI: https://doi.org/10.1093/ve/veae092
Advance Access Publication Date: 11 November 2024
Research Article

A phylogenetic assessment of HIV-1 transmission trends among people who inject drugs from Coastal and Nairobi, Kenya

Hanley Kingston[1,*], George Nduva[2], Bhavna H. Chohan[3,4], Loice Mbogo[5,6], Aliza Monroe-Wise[4], Betsy Sambai[7], Brandon L. Guthrie[4,8], Eduan Wilkinson[9,10], Jennifer Giandhari[9], Sarah Masyuko[4,11], William Sinkele[12], Tulio de Oliveria[4,9,10], David Bukusi[6], John Scott[13], Carey Farquhar[4,8,13,‡], Joshua T. Herbeck[14,‡]

[1]Institute for Public Health Genetics, University of Washington, 1410 NE Campus Parkway, Seattle, WA 98195, United States
[2]Department of Translational Medicine, Lund University, Box 117, Lund SE-221 00, Sweden
[3]Centre for Virus Research, Kenya Medical Research Institute, Mbagathi Rd, Nairobi P.O. Box 54628-00200, Kenya
[4]Department of Global Health, University of Washington, 3980 15th Avenue, Seattle, WA 98195, United States
[5]University of Washington Global Assistance Program-Kenya, Nairobi, Kenya
[6]Kenyatta National Hospital, Hospital Rd, Nairobi, Kenya
[7]Population Council-Kenya, Avenue 5, Rose Ave, Nairobi, Kenya
[8]Department of Epidemiology, University of Washington, 3980 15th Avenue, Seattle, WA 98195, United States
[9]KwaZulu-Natal Research Innovation and Sequencing Platform, Nelson R Mandel School of Medicine, University of KwaZulu-Natal, 719 Umbilo Rd, Berea, Durban, KwaZulu-Natal 4001, South Africa
[10]Centre for Epidemic Response and Innovation (CERI), Stellenbosch University, Hammanshand Rd, Stellenbosch Central, Stellenbosch 7600, South Africa
[11]Ministry of Health, Cathedral Rd, Kilimani, Nairobi, Kenya
[12]Support for Addiction Prevention and Treatment in Africa, Corner House, Nairobi, Kenya
[13]Department of Medicine, University of Washington, 1410 NE Campus Parkway, Seattle, WA 98195, United States
[14]Institute for Disease Modeling, Bill & Melinda Gates Foundation, 500 5th Ave N, Seattle, WA 98109, United States

[‡]These authors contributed equally to this work as senior authors.
*Corresponding author. Institute of Public Health Genetics, 1959 NE Pacific St, University of Washington, Seattle, WA 98195, United States.
E-mail: hkingsto@gmail.com

Abstract

Although recent modeling suggests that needle–syringe programs (NSPs) have reduced parenteral HIV transmission among people who inject drugs (PWID) in Kenya, the prevalence in this population remains high (~14–20%, compared to ~4% in the larger population). Reducing transmission or acquisition requires understanding historic and modern transmission trends, but the relationship between the PWID HIV-1 sub-epidemic and the general epidemic in Kenya is not well understood. We incorporated 303 new (2018–21) HIV-1 *pol* sequences from PWID and their sexual and injecting partners with 2666 previously published Kenyan HIV-1 sequences to quantify relative rates and direction of HIV-1 transmissions involving PWID from the coast and Nairobi regions of Kenya. We used genetic similarity cluster analysis (thresholds: patristic distance <0.045 and <0.015) and maximum likelihood and Bayesian ancestral state reconstruction to estimate transmission histories at the population group (female sex workers, men who have sex with men, PWID, or general population) and regional (coast or Nairobi) levels. Of 1081 participants living with HIV-1, 274 (25%) were not virally suppressed and 303 (28%) had sequences available. Of new sequences from PWID, 58% were in phylogenetic clusters at distance threshold <0.045. Only 21% of clusters containing sequences from PWID included a second PWID sequence. Sequences from PWID were similarly likely to cluster with sequences from female sex workers, men who have sex with men, and the general population. Ancestral state reconstruction suggested that transmission to PWID from other populations was more common than from PWID to other populations. This study expands our understanding of the HIV-1 sub-epidemic among PWID in Kenya by incorporating four times more HIV-1 sequences from this population than prior studies. Despite recruiting many PWID from local sexual and injecting networks, we found low levels of linked transmission in this population. This may suggest lower relative levels of parenteral transmission in recent years and supports maintaining NSPs among PWID, while also strengthening interventions to reduce HIV-1 sexual acquisition and transmission for this population.

Keywords: HIV; Kenya; people who inject drugs (PWID); phylogenetic; phylogeography; molecular epidemiology

© The Author(s) 2024. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site–for further information please contact journals.permissions@oup.com.

Introduction

To address harms associated with injection drug use, including risk of parenteral HIV-1 transmission, the Kenyan Government introduced needle and syringe programs (NSPs) in 2013 and methadone maintenance treatment in 2014 [National AIDS and STI Control Programme (NASCOP) 2014, Rhodes et al. 2015]. However, the prevalence of HIV-1 in people who inject drugs (PWID) in Nairobi and the coastal region of Kenya remains four-fold higher than in the larger Kenyan population, at 14–21%, [Degenhardt et al. 2017, National AIDS and STI Control Programme (NASCOP) 2020, Stone et al. 2022] and is similar to the prevalence among men who have sex with men (MSM), but likely lower than the estimated prevalence among female sex workers (FSW, 12–31%) [National AIDS Control Council, National AIDS & STI Control Program (NASCOP) 2018].

Recent modeling suggests that NSPs have been highly effective at reducing recent HIV transmission among PWID in Kenya, as was already understood to be the case in the USA and Western Europe (Asher et al. 2013, Ragonnet-Cronin et al. 2019, Stone et al. 2022). These models suggest that PWID, and particularly women who inject drugs, are increasingly likely to acquire HIV sexually, as opposed to from contaminated needles (Stone et al. 2022). However, the question of primary sources and mechanisms of HIV-1 acquisition and transmission among Kenyan PWID has not been extensively addressed through phylogenetic methods based on HIV-1 genetic data. Questions also remain about the extent to which the HIV-1 epidemic among PWID is self-sustaining, versus part of, or a reflection of, the epidemic in the larger population. Injection drug use in Kenya is heavily concentrated in urban areas (primarily in coastal cities and Nairobi), and inter-regional HIV-1 transmission dynamics involving PWID are also not well-understood [National AIDS and STI Control Programme (NASCOP) 2019, Nduva et al. 2022a]. The most recent HIV biobehavioral survey in Kenya was conducted in Kisumu and Nairobi counties in 2011, prior to the introduction of needle–syringe programs and methadone clinics (Tun et al. 2015). Epidemiological and molecular epidemiological modeling can empirically fill gaps in behavioral data for PWID and connect behavioral changes to changes in HIV prevalence and/or transmission patterns.

Phylogenetic analyses are well-suited to resolve questions about HIV transmission patterns between populations, between regions, and over time (Suchard et al. 2018, Bbosa et al. 2019, Brenner et al. 2021, Magosi et al. 2022). By overlaying geographical or epidemiological data onto a molecular phylogeny, we can also infer how the behaviors or experiences of individuals living with a disease relate to transmission trends. Rapid virus evolution, which occurs within the timescale of transmission, makes HIV-1 particularly conducive to phylogenetic analysis (Grenfell et al. 2004).

Prior phylogenetic analysis of HIV-1 in Kenya primarily focused on populations other than PWID and showed more HIV-1 transmission from the higher prevalence western regions to lower prevalence eastern regions (Bezemer et al. 2014, Nduva et al. 2022a). However, because HIV prevalence among PWID is highest in coastal Kenya (Kurth et al. 2015, Sambai et al. 2022) [opposite of regional prevalence trends in the general population (GP)] and because prior phylodynamic studies have shown differing regional transmission patterns among MSM, another key population, (Nduva et al. 2022b) it is not clear if this HIV-1 transmission trend is replicated among PWID. Prior research identified a cluster of 41 HIV-1 sequences collected from PWID in the coastal city of Mombasa in 2010 (Nduva et al. 2022a). This finding suggested substantial, isolated transmission within the PWID population, with needle-sharing likely playing a role, but also contradicted a common and potentially stigmatizing narrative that transmission from PWID "seeds" cases in other populations (Beckerleg and Hundt 2004, Nduva et al. 2020, 2022a). In light of substantial additions to harm reduction programs for PWID in Kenya in the last 10 years, this analysis seeks to evaluate more recent and geographically diverse trends in HIV transmission involving PWID. This study reports a four-fold higher number of HIV-1 sequences from PWID in Kenya and is the first to collect HIV-1 sequences from PWID after the introduction of NSPs in Kenya.

Methods

Study population and enrollment

The study of HIV, hepatitis C (HCV), APS, and Phylogenetics for PWID (SHARP) is a prospective cohort study that recruited people who had injected drugs in the last 3 years and were living with HIV (indexes) from 2018 to 2021 and used assisted partner services (APS) to identify, test, and treat their sexual and injecting partners (Monroe-Wise et al. 2021). Index participants were recruited from NSPs and methadone clinics in Nairobi (central Kenya) and Kilifi and Mombasa counties (coastal region). Eligibility criteria for indexes was: ≥18 years of age, injected at least once in the past year, tested positive for HIV, had not experienced intimate partner violence in the last month, and gave written informed consent. Indexes were enrolled in APS, through which they identified sexual (vaginal, anal, or oral intercourse) and injecting (regardless of needle sharing) partners (≥18 years of age) from the previous 3 years, who were also invited to enroll in the study.

Sociodemographic data, HIV and HCV history, and sexual and injection drug use history were obtained for all participants by a survey. Rapid HIV-1 testing using fingerstick samples was performed during the interview sessions following an established Kenya national algorithm (Guidelines for HIV Testing 2010). Detailed study procedures are reported in the published study protocol (Monroe-Wise et al. 2021).

Laboratory procedures and sequencing

Blood samples were collected from all participants who tested positive for HIV and used to prepare dried blood spots and plasma samples for viral load testing and sequencing. Plasma samples (for 15 samples, dried blood spots were used) were shipped to the Kwazulu-Natal Research Innovation and Sequencing Platform (KRISP) laboratory at the University of KwaZulu-Natal, South Africa for Sanger sequencing ($N = 255$) or next-generation sequencing (NGS; $N = 48$).

For Sanger sequencing, PCR amplification was performed on the HIV-1 polymerase (*pol*) region using Genotyping Kit Amplification Module (ThermoFisher Scientific). The HIV-1 Genotyping Kit Cycle Sequencing Module (ThermoFisher Scientific) was used for cycle sequencing followed by purification (BigDye Xterminator kit, ThermoFisher Scientific) and capillary electrophoresis using a 3730xl DNA Analyser (Applied Biosystems).

For NGS, short overlapping amplicons spanning the full genome were generated using a tiling PCR approach, (Quick et al. 2017) and consensus sequences were later trimmed to the pol region. All sequencing libraries were prepared using the Nextera DNA Flex Library Prep kit with Nextera CD indexes (Illumina, San Diego, CA, USA) and quantified using the Qubit dsDNA High Sensitivity assay kit on a Qubit fluorometer (Life Technologies, Carslbad, CA, USA). Sequencing was performed on an Illumina MiSeq platform (Illumina).

Sequence dataset and phylogenetic determination of clusters

A total of 4058 previously published HIV-1 *pol* sequences (approximately 1020 nucleotides, HXB2 [K03455] positions 2267–3287) were available from Kenya (Nduva et al. 2022a), of which we incorporated 3587 into an alignment based on: year >2000 and not from a person <15 years old or a maternal to child transmission study (where known). SHARP and previously published sequences were annotated as being from: PWID, MSM, FSW, or GP (defined in this analysis as not known to be <15 years old and not known to be in a key population). Because participants from the SHARP study were not asked about recent sex work and because transactional sex among PWID is often highly connected with obtaining and using drugs, (Mburu et al. 2019), we conducted secondary cluster analyses on a subset of HIV-1 sequences from female PWID SHARP study participants who had transactional sex risk factors: ≥3 sexual partners in the prior month and having ever received money for sex.

For previously published sequences, subtypes were determined by maximum-likelihood phylogenetic analysis in PhyML under the general time reversable (GTR) + Γ4 + I model and the SH-aLRT algorithm for branch support (Guindon et al. 2010), and circulating recombinant forms (CRFs) were resolved by bootscan analysis in Simplot as described by Nduva and colleagues (Nduva et al. 2022a). For the new SHARP study sequences, we assessed HIV-1 subtypes using REGA HIV subtyping tool v3.6 (available at http://dbpartners.stanford.edu:8080/RegaSubtyping/stanford-hiv/typingtool/) and resolved CRFs using a combination of REGA, COMET (available at https://comet.lih.lu/), and clade placement on a phylogenetic tree with reference sequences (Pineda-Peña et al. 2013).

For alignment, we excluded all NRTI and NNRTI mutations listed on the Stanford Drug Resistance Database [Bennett et al.] and performed multiple sequence alignment using fast Fourier transform (MAFFT) (defaults: 200PAM scoring matrix, transitions-transversions ratio: 2, gap open penalty: 1.53, offset value: 0.123), implemented in Geneious Prime v11.0.11 (Katoh et al. 2002, Kearse et al. 2012, Katoh and Standley 2013).

We used a two-pronged approach [i.e. maximum likelihood (ML) and Bayesian inference] to construct phylogenies, with ML trees used for analysis of trait distribution across clusters and Bayesian trees used for coalescence (Royer-Carenzi et al. 2013, Joy et al. 2016). Both methods were used for ancestral state reconstruction. For the ML approach, we used IQ-Tree, (Nguyen et al. 2015, Minh et al. 2020) a maximum likelihood calculator, and used the built-in model selection tool (Kalyaanamoorthy et al. 2017) to choose the general time-reversable substitution model with gamma-distributed rate variation (GTR + R(9)) (Soubrier et al. 2012, Kalyaanamoorthy et al. 2017). Cluster summaries were combined across HIV-1 subtype-specific ML trees (N tips: SHARP subtype A1: 196, published subtype A1: 2650; SHARP subtype C: 38, published subtype C: 270; SHARP subtype D: 19, published subtype D: 436; total: 3432) and defined using maximum patristic distances <0.015 and ≤0.045; these are commonly used thresholds for defining recent transmission and more distantly related networks, respectively (Guindon et al. 2010, Oster et al. 2015, Junqueira et al. 2019). Only clusters containing at least one sequence from the coast or Nairobi were included in summary statistics. The primary outcome of interest in cluster analysis was the percent of clusters for each population group that contained a sequence from a PWID ([N clusters with 1+ PWID sequence and 1+ sequence from population group X]/[N clusters with 1+ sequence from population group X]).

We estimated the dates of origin (time to most recent common ancestor; tMRCA) of clusters containing subtype A1 sequences from a PWID. For coalescent analysis, sequences were limited to the 316 that fell into a cluster containing at least 1 PWID sequence based on the ML tree. We implemented a Bayesian coalescent tree model in BEAST (v1.10.4) using the SkyGrid model (Drummond et al. 2002, Gill et al. 2013) with GTR + Γ4 substitution model, and specifying an uncorrelated, relaxed clock (Drummond et al. 2006). We calculated Markov chain Monte Carlo (MCMC) runs with a chain frequency of 300 million generations, logging every 50 K iterations. After discarding 10% as burn-in, we built a maximum clade credibility tree and calculated the average and IQR for origin dates of both clusters containing PWID sequences and PWID-exclusive clusters, based on a patristic distance threshold of 0.045.

Ancestral state reconstruction

We performed ancestral state reconstruction using subsets of 2342 SHARP and previously published HIV-1 subtype A1, C, and D sequences from the coastal and Nairobi regions. We excluded from ancestral state reconstruction, 43 HIV-1 subtype A1 sequences (41 from PWID and 2 from the GP) that formed a previously identified cluster (Nduva et al. 2022a). This cluster is a substantial outlier in terms of size and, given the relatively small sample size available from PWID, would likely drive transmission pattern estimates. Having noted the importance of this cluster analysis, we wanted to draw inference relevant to the rest of the PWID population.

For geographic ancestral state reconstruction, we considered two discrete states: Coastal region or Nairobi County. For risk-group ancestral state reconstruction, we specified four discreet states: FSW, MSM, PWID, and GP. The ML and Bayesian approach were used to perform three main ancestral state reconstruction analyses, stratified by HIV-1 subtype. The first approach assessed regional transmission (coast and Nairobi) among PWID recruited from the SHARP study (ML method only; uniform subsampling: 76 sequences each; proportionate subsampling (based on estimated regional HIV-1 prevalence and PWID population sizes): 40 sequences from coast and 76 sequences from Nairobi). The second approach assessed transmission between different population groups (uniform sampling, 74 sequences each from GP, FSW, MSM, PWID). The third approach estimated transitions between PWID and not-PWID from the coast and Nairobi regions (110 sequences per group). As a secondary analysis to test agreement with prior studies, we assessed regional transmission between the coast and Nairobi overall. Analyses were performed for HIV-1 subtype A1 and for subtypes C and D where sample-size permitted.

Sequences from other regions in Kenya were incorporated as references when constructing ML trees, but were excluded from ancestral state reconstruction, as no sequences were available from PWID, the primary population of interest, in these regions. ML trees were further filtered to obtain proportionate (approach #1 only) or uniform subsampling. For Bayesian analysis, because it is computationally intensive and because ancestral-state reconstruction is conducted concurrently in the MCMC chain, alignments were filtered prior to tree reconstruction to include only coast and Nairobi sequences and to achieve uniform subsampling of the trait of interest.

Ancestral state reconstruction was performed on ML trees using a marginal ML algorithm (Phanghorn package, R), which calculates the likelihood of each state at each ancestor (Schliep 2011, R: A Language and Environment for Statistical Computing 2019). We defined a state transition when the trait of the descendant node differed from the state of its immediate ancestor on the

tree. We resolved ancestor states according to their estimated likelihood, averaging over 20 resolutions. We also averaged transition counts across 10 subtrees for HIV-1 subtype A1 (and 30 subtrees for the smaller HIV-1 subtypes C and D). *P*-values were calculated for the statistic: (transitions from state 0 to state 1)/(transitions from state 1 to state 0) for each pair of traits in order to assess the hypothesis that transitions were more common in one direction. The non-parametric null distribution is based on 200 resampling of the tree tip traits (X20 resolutions of ML ancestor states X10 subtrees). We also conducted secondary analyses of "terminal transitions" (whether a transition event occurred on the terminal branch of the tree leading to an observed sequence). We did this to investigate more recent trends and to address uncertainty that develops for ancestral state reconstruction deeper in the tree, where states tend to converge near the tree root.

Ancestral state reconstruction in BEAST was conducted concurrently with the tree-building process. Trees were developed under the assumption of constant population size with a strict clock (to improve computational efficiency) and inferred under the GTR + Γ4 substitution model (Lemey et al. 2009, Suchard et al. 2018). For each subset, chains of 250–500 million iterations were run to ensure convergence to the correct posterior distribution. Convergence was assessed using Tracer, and 10% of states were removed to account for burn-in. Convergence was determined in Tracer V1.7.2 [defined as an effective sample sizes (ESS) ≥100 for most traits]. Bayesian ancestral state reconstruction was performed using an asymmetric continuous-time Markov chain (CTMC) model, as it relaxes the assumption of constant diffusion rates through time to realistically model phylogeographic processes (Lemey et al. 2009, Edwards et al. 2011). We used the PrioriTree Software to empirically derive a prior on the dispersal rate for each discrete trait—using the first quantile of the parsimony score; this decreased the dependance of the estimated dispersal rate on the prior (Gao et al. 2022, 2023). Well-supported movements and Bayes factors (BF) assessing statistical support were summarized using SPREAD v0.9.6, (BF ≥3 was considered significant) (Bielejec et al. 2016). We also used a robust counting approach implemented in BEAST to estimate the forward and reverse HIV-1 movement events (Markov jumps) between locations and population group states along the branches of time-dated phylogenetic trees (Suchard et al. 2018). We averaged jump counts across 7-9 subtrees.

Unless otherwise stated, we report transition counts (for ML methods) or Markov jumps (the corresponding measure for Bayesian methods) as a percent of all branches and indicate the null as the random probability of the transition event based on the frequency of the sampled traits. Transitions or jumps are reported in the text for the A1 subtype.

Cluster and ancestral state reconstruction summary statistics, tables, and figures were developed using the ape (Paradis et al. 2019), ggtree (Wickham et al. 2016, Yu et al. 2017) and R (R: A Language and Environment for Statistical Computing 2019) packages.

Ethical consideration

Ethical approval was provided by the Institutional Review Board at the University of Washington (STUDY00001536) and the Ethical Review Committee at Kenyatta National Hospital/University of Nairobi (P265/05/2017). All the participants in this study provided informed consent for inclusion in the study.

Results

Study participants and HIV-1 sequences

This study enrolled 1081 participants living with HIV, of whom 274 (25%) did not have viral suppression (viral load >1000 copies/ml). We were able to sequence 313 samples; we excluded 1 sequence for missing epidemiological data, 2 for failure to capture the partial pol region, and 7 for >15% missing or ambiguous bases, leaving 303 HIV-1 sequences (representing 28% of all SHARP study participants with an HIV diagnosis).

Of the 303 participants with a sequence available, most (80%) were diagnosed prior to study enrolment, primarily after 2013 (*N* = 156, 64%), when NSPs were introduced (Supplementary Table S1). The vast majority had injected drugs (96%) and 81% had injected in the previous month, but recent self-reported needle-sharing was <10% in both regions. While receiving money or goods for sex was more common among women in Nairobi (48% compared to 33% on the Kenyan coast), participants on the coast reported more sexual partners, on average. Fifty-three female participants fit our definition of having transactional sex risk factors: having ever received money or goods for sex and reporting ≥3 sex partners in the prior month.

The most common subtypes were A1 (SHARP: 67.0%, all: 71.0%), C (SHARP: 12.5%, all: 7.3%), and D (SHARP: 6.6%, all: 10.0%), with 28 (9.2%) recombinants in the SHARP data and 288 (10.8%) recombinants in the full dataset (Supplementary Tables S1 and S2, Fig. 1). After incorporating previously published sequences, subtype A1, C, or D Sequences were available from 148 FSW, 300 MSM, 295 PWID, and 1598 people in the GP from the coast and Nairobi regions (Supplementary Table S2, Supplementary Fig. S1). The estimated sampling density of people living with HIV from Nairobi was 0.7% of people in the GP, 0.7% of FSW, 7.4% of MSM, and 10.6% of PWID. Sampling density from the coast was 0.6% of people in the GP, 6.1% of FSW, 17.8% of MSM, and 12.2% of PWID (Supplementary Table S3). Most HIV-1 sequences from FSW, MSM, and the GP were from the previously published sequence dataset (Table 1), while most PWID HIV-1 sequences were newly collected through this study. HIV-1 subtype distribution was similar between the coast and Nairobi regions (Supplementary Table S2).

Cluster analysis

Using a maximum patristic distance threshold of 0.045, there were 680 phylogenetic clusters of subtypes A1 (*N* = 540), C (*N* = 59), or D (*N* = 81) (Supplementary Table S4, Fig. 2) containing at least 1 sequence from the coast or Nairobi region, and 120 (18%) of these clusters also contained at least one sequence from a PWID (Table 2, Supplementary Fig. S2). The mean estimated year of origin (for the common ancestor sequence) for clusters containing a PWID sequence was 2001 (IQR: 1995–2005) and the mean estimated year of origin for PWID-exclusive clusters was 2010 (IQR: 2004–2017); however, the estimated root age of the A1 phylogeny [1955 (HPD: 1939–68)] was relatively low, given that the first HIV case in Kenya was detected in 1984 (Simat 2022).

Overall, 142 (58%) of new (SHARP study: 2018-2021) and 190 (63%) of all sequences from PWID fell into clusters at threshold 0.045. The previously identified cluster of 41 sequences (40 of which fell within distance 0.015) from PWID in Mombasa (Supplementary Tables S4 and S5, Supplementary Fig. S1) represented the largest cluster in our data and the only cluster with >5 PWID sequences. The next-largest cluster contained sequences from 13

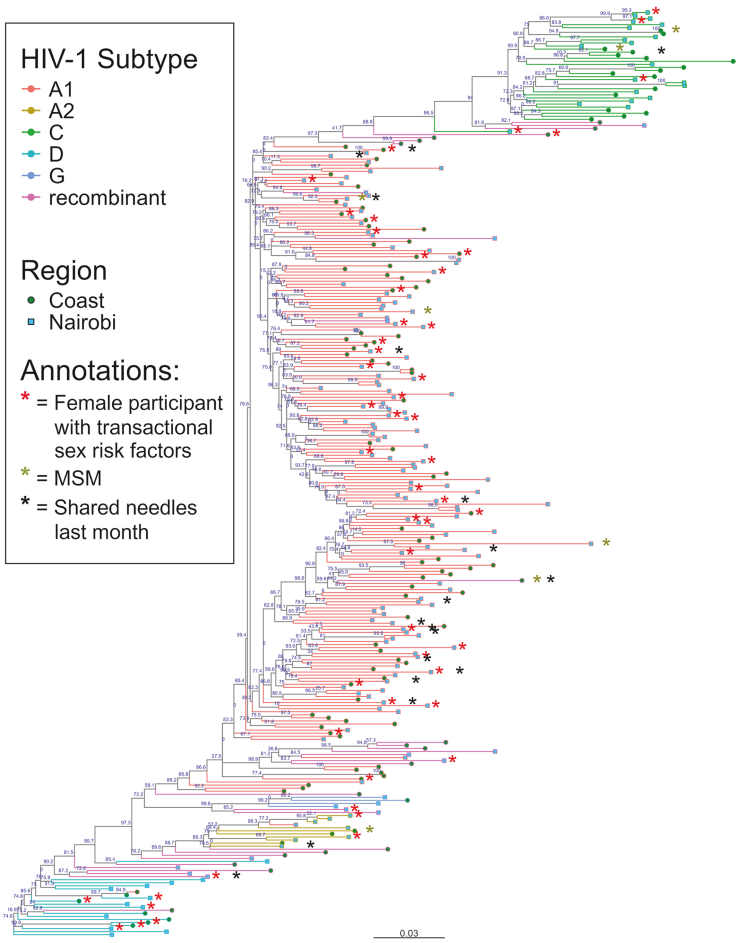

Figure 1. A molecular phylogeny of 303 HIV-1 sequences from the SHARP study. The phylogeny was assembled using a ML approach (IQ-Tree; Nguyen et al. 2015, Minh et al. 2020) using the general time-reversable substitution model with gamma-distributed rate variation (GTR + R(9)). Visualization was developed in R. (R: A Language and Environment for Statistical Computing 2019).

individuals (Supplementary Fig. S3). Sequences from PWID most commonly clustered with sequences from other PWID but were similarly likely to cluster with sequences from each of the other three populations considered: Sequences from PWID were found in 16% of all clusters containing at least 1 GP sequence, 12% of all clusters containing FSW sequences, and 14% of clusters containing MSM sequences (Table 2, Supplementary Table S6). Of new PWID sequences clustering with sequences from other populations, 50 (45%) came from men and 61 (55%) came from women. Despite most PWID sequences being collected through partner referral within the SHARP study, only 25 (21%) of clusters with PWID sequences contained another sequence from a PWID (5 with PWID sequences only from women, 5 only from men, 12 from men and women, and 3 missing data). This represents 52 (21%) of all non-recombinant SHARP sequences from PWID.

HIV-1 sequences from female PWID with transactional sex risk factors also did not cluster substantially more with sequences from any one population nor did sequences from PWID who reported sharing needles (Supplementary Table S7). When using the more stringent <0.015 distance threshold, population-specific clustering was more common (8 out of 15 clusters containing a PWID sequence, 53%); nevertheless, as this represents only 8% ($N = 20$) of all non-recombinant SHARP PWID sequences clustering with a second PWID sequence (Table 2, Supplementary Fig. S2).

Transmission between PWID and other populations

After adjusting for population size via uniform subsampling, ancestral state reconstruction in the A1 subtype suggests that about half of transmissions occurred within the defined population groups (ML: 48.9%; Bayesian: 47.5%; null: 25%); there was no excess transmission within the PWID population (ML: 5.8%; null: 6.3%). Transitions to the PWID population were 1.5-times (ML) to 2.8-times (Bayesian) more common than transitions from the PWID population (Table 3). Both methods found that transitions/jumps between PWID and MSM were rare in either direction. ML methods estimated most transitions to the PWID population came from the GP (31, $P < .01$), followed by FSW (23, $P < .01$) and Bayesian methods estimated the highest number of jumps to PWID came from FSW (37, BF = 15.3), followed by the GP (24, BF = 10.6). In an analysis stratified by region, however, excess transmission from the GP to PWID populations was significant only for the coast and only with the ML approach.

Secondary analysis restricted transition counts to terminal branches (the branches leading to the observed sequences on the tree using ML methods only) to assess more recent transmission trends. This revealed similar patterns for population groups, with 2.20-times more transitions to PWID than from PWID to other populations (Supplementary Table S8).

Table 1. Demographics and distribution of new (from SHARP study) and previously published Kenyan HIV-1 *pol* sequences by population. New sequences from the SHARP study are indicated in bold, as are sequences with A1, C, or D subtype (the primary sequences incorporated in these analyses).

	FSW (*N* = 207), *n* (%)	MSM (*N* = 376), *n* (%)	PWID (*N* = 341), *n* (%)	GP (*N* = 2966), *n* (%)	Total (*N* = 3890), *n* (%)
Source					
Published	206 (99.5)	368 (97.9)	58 (17.0)[a]	2955 (99.6)	3587 (92.2)
SHARP (new)	**1 (0.5)**	**8 (2.1)**[b]	**283 (83.0)**	**11 (0.4)**	**303 (7.8)**
Sampling year					
2001–15	178 (86.0)	144 (38.3)	58 (17.0)	2709 (91.3)	3089 (79.4)
2015+	29 (14.0)	232 (61.7)	283 (83.0)	257 (8.7)	801 (20.6)
Region					
Central	0 (0.0)	0 (0.0)	0 (0.0)	44 (1.5)	44 (1.1)
Coast	110 (53.1)	178 (47.3)	171 (50.1)	700 (23.6)	1159 (29.8)
Eastern	0 (0.0)	0 (0.0)	0 (0.0)	6 (0.2)	6 (0.2)
Nairobi	82 (39.6)	141 (37.5)	170 (49.9)	1114 (37.6)	1507 (38.7)
Nyanza	14 (6.8)	57 (15.2)	0 (0.0)	501 (16.9)	572 (14.7)
Rift Valley	1 (0.5)	0 (0.0)	0 (0.0)	497 (16.8)	498 (12.8)
Western	0 (0.0)	0 (0.0)	0 (0.0)	104 (3.5)	104 (2.7)
HIV subtype					
Missing	0	2	11	1	14
G	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.0)	1 (0.0)
A1	**122 (58.9)**	**276 (73.8)**	**243 (73.6)**	**2040 (68.8)**	**2681 (69.2)**
A2	0 (0.0)	0 (0.0)	9 (2.7)	2 (0.1)	11 (0.3)
B	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.0	1 (0.0)
C	**20 (9.7)**	**33 (8.8)**	**36 (10.9)**	**213 (7.2)**	**302 (7.9)**
D	**21 (10.1)**	**48 (12.8)**	**17 (5.2)**	**363 (12.2)**	**449 (11.6)**
G	2 (1.0)	1 (0.3)	3 (0.9)	17 (0.6)	23 (0.6)
Recombinant and other	42 (20.3)	16 (4.3)	22 (6.7)	328 (11.1)	408 (10.6)
Sex or gender[c]					
Missing	0	2	58	2253	2313
Female	207 (100.0)	0 (0.0)	152 (53.7)	491 (68.9)	850 (53.9)
Male	0 (0.0)	374 (100.0)	131 (46.3)	222 (31.1)	727 (46.1)

[a]41/58 published sequences were part of a large, previously identified cluster and these were included in cluster analysis but excluded from discrete trait analysis.
[b]7/8 MSM were also PWID. Sequences from MSM who were also PWID were counted as MSM sequences in discrete trait analysis.
[c]Most data did not distinguish between sex and gender.
See Supplementary Table S1 for demographic distributions by region and a breakdown of recombinant subtypes.

Regional transmission trends among PWID

Ancestral state reconstruction restricted to newly collected PWID sequences (Supplementary Table S9) estimated substantial between-region transitions (31.8% of terminal branch transitions; null: 50%) and similar rates of transitions in either direction [$P = .05$, similar to the pattern of regional transitions observed in the total population (Supplementary Table S10)]. To further investigate drivers of regional transmission, we conducted a combined analysis looking at the relationship between the GP and PWID population by region (Fig. 3, Supplementary Tables S11 and S12), which estimated the most between-region transitions within the GP, rather than the PWID population (after sampling equal numbers of sequences from both populations). In the GP, ML methods estimated similar transition frequency in both regional directions (7.5% coast to Nairobi versus 6.8% Nairobi to coast, $P = .20$; null: 6.25% each), while Bayesian methods suggested substantially greater jumps from the coast to Nairobi (16.9% versus 1.4% Nairobi to coast, BF > 100; Fig. 2, Supplementary Table S11).

Transitions were consistently (although not always significantly) more common from the GP populations to PWID populations (versus from PWID to the GP) within and between both regions. Specifically, within the coast, GP to PWID transitions were 2- (ML, $P < 0.01$) to 10-times (Bayesian, BF = 1.6) more common and for Nairobi they were 1.3- (ML, $P = 0.02$) to 12-times (Bayesian, BF = 0.5) more common. Surprisingly, cross-region transitions between the GP and PWID populations were estimated at similar rates as within-region transitions in both ML and Bayesian models.

Discussion

We leveraged APS recruitment to collect sequences from 303 PWID living with HIV (and their sexual and injecting partners) from coastal and Nairobi, Kenya, the regions with the highest estimated levels of injection drug use in the country [National AIDS and STI Control Programme (NASCOP) 2019]. Our analysis reveals that the PWID HIV-1 sub-epidemic is connected to that in the GP and FSW populations.

Mechanisms of transmission among PWID

Our finding of few PWID exclusive clusters provides the first molecular epidemiological evidence supporting more limited transmission among PWID in recent years, in Kenya and in an African setting. Given that we recruited from sexual and injecting partner networks in recent years (2018–21), our study was likely better powered than previous studies to resolve recent PWID HIV-1 transmission clusters involving Kenyan PWID (Nduva et al. 2020). Nevertheless, the only cluster we identified containing >4 PWID sequences was a previously described 41-sequence cluster in Mombasa (Osman et al. 2013, Nduva et al. 2020). While transmission within PWID may be more common than between PWID and other populations (although this may also be explained by population-specific sampling), the results of the cluster analysis do not suggest widespread, isolated transmission among PWID within the last 10 years.

Our results suggest that the HIV epidemic among PWID is not self-sustained. Strategies to reduce parenteral transmission

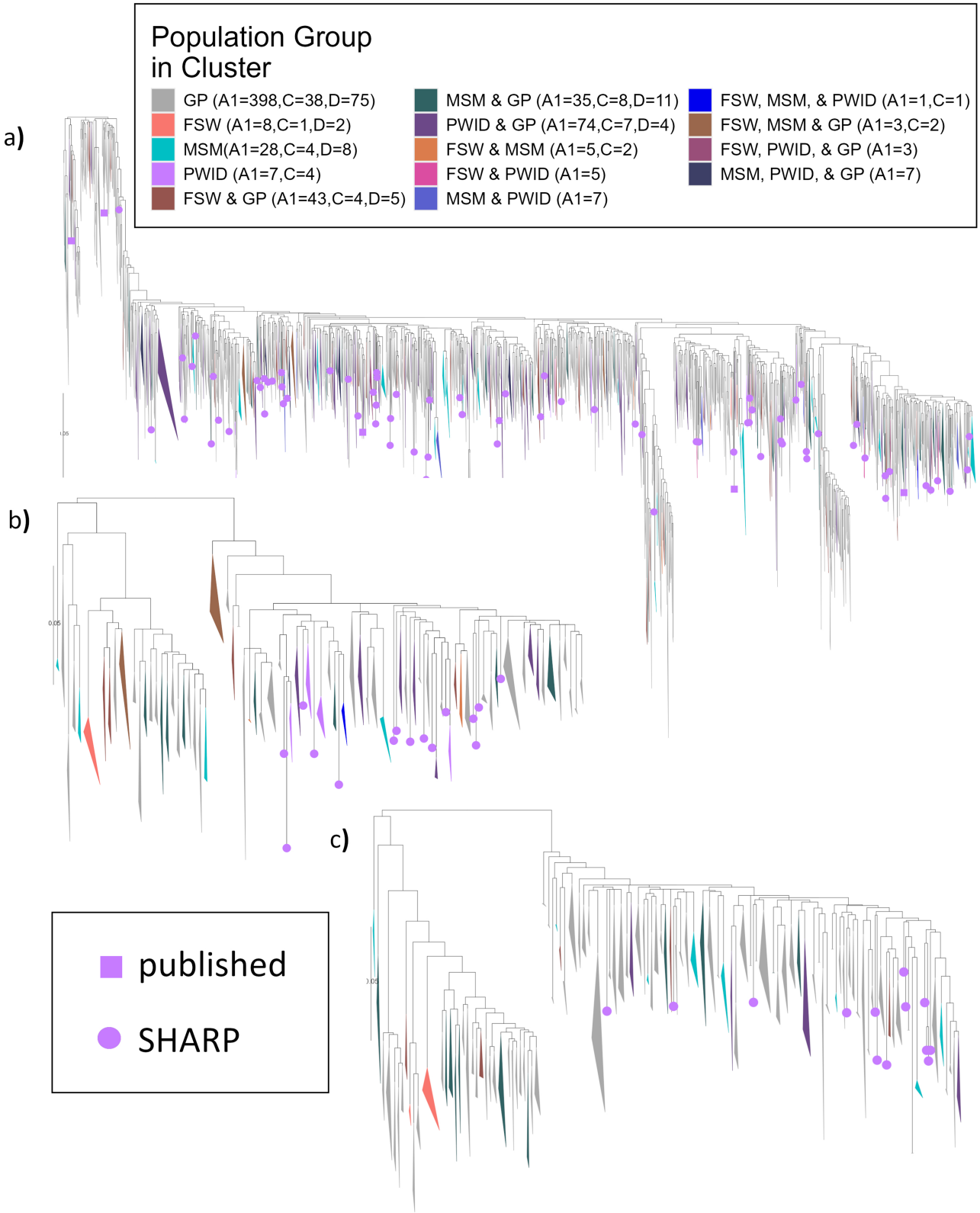

Figure 2. Molecular phylogenies of HIV-1 sequence clusters for (a) 203 subtype A1, (b) 38 subtype C, and (c) 19 subtype D. Clusters are depicted as triangles with nodes at the most-recent common ancestor and the closest and furthest tip. PWID sequences not in clusters are shown in purple and all other sequences not in clusters are excluded. Phylogenies were assembled using a ML approach (IQ-Tree) (Nguyen et al. 2015, Minh et al. 2020) using the general time-reversable substitution model with gamma-distributed rate variation (GTR + R(9)). Visualization was developed in R. (R: A Language and Environment for Statistical Computing 2019) Abbreviations: FSW: female sex workers, MSM: men who have sex with men, PWID: people who injects drugs, GP: general population.

Table 2. Sequences from other populations in clusters with HIV sequences from PWID.

Clusters containing sequences from PWID			
Patristic distance < 4.5 ($N = 120$)		Patristic distance < 1.5 ($N = 15$)	
In cluster with	Count, *n* (%)[a]	In cluster with	Count, *n* (%)[a]
GP	95 (16.9)	GP	3 (1.8)
FSW	10 (12.2)	FSW	0 (0.0)
MSM	16 (13.9)	MSM	2 (4.5)
PWID (i.e. clusters with ≥2 PWID sequences)	25 (20.8)	PWID (ie. clusters with ≥2 PWID sequences)	12 (80.0)

[a]Clusters containing >2 population groups are represented more than once. Clusters containing at least one sequence from the coast or Nairobi are included. Percents are based on the total number of clusters that contain at least one sequence from the given population (see Supplementary Tables S4–S6 for comprehensive cluster counts). Clusters are based on 0.045 or 0.015 maximum genetic distance threshold on maximum likelihood trees and combined across subtypes A1, C, and D.

between PWID, while effective, are alone not sufficient to address high HIV-1 prevalence in this population. Low levels of excess clustering among sequences from PWID, the majority of whom were diagnosed after the advent of NSPs, may reflect the effectiveness of these programs at preventing parenteral HIV-1 transmission [National AIDS and STI Control Programme (NASCOP) 2014,Rhodes et al. 2015]. Recent modeling estimated that NSPs reduced HIV transmission among PWID in Kenya by 40–46% in 2020 (Stone et al. 2022). Modeling studies are supported by empirical evidence, with between-study comparison suggesting decreasing rates of sharing injection equipment among PWID: studies conducted in the coastal and/or Nairobi regions between 2010 and 2012 estimating that 28–55% of PWID shared needles in the prior month [4–48% at last injection; National AIDS and STI Control Programme (NASCOP) 2014), Kurth et al. 2015, Brodish et al. 2011, Oguya et al. 2021] while studies conducted after 2015 estimated a much lower prevalence for needles sharing of 2–5% per month (2–12% at last injection) within these regions (Akiyama et al. 2019, Joint United Nations Programme on HIV/AIDS 2022, Sambai et al. 2022). Nevertheless, prior research, pre- and post-introduction of NSPs, shows that HIV prevalence increases with number of years injecting (Kurth et al. 2015, Sambai et al. 2022). Discerning the cause of this increasing risk is critical to provide appropriate resources to this population.

To this end, PWID may also have greater risk factors for HIV acquisition through non-parenteral routes (Asher et al. 2013). A 2012 retrospective analysis showed, for example, that the prevalence of HIV among people who *later* started injecting drugs (7% in Nairobi and 9% in coastal Kenya) was higher than among the general population (Kurth et al. 2015). Epidemiological and network analyses support that PWID, particularly young PWID, are more likely to engage in sexual behaviors associated with HIV acquisition and transmission and estimate that for female PWID, sexual acquisition may now be more common than acquisition through contaminated needles (Oguya et al. 2021, Stone et al. 2022). We were not able to draw conclusions about differences in acquisition or transmission patterns for male and female PWID because we did not observe substantial differences in the frequency with which HIV-1 sequences from male or female PWID clustered together or with sequences from other populations. That PWID may face elevated risk of HIV acquisition from non-parenteral routes also raises the possibility that any elevated acquisition risk may be shared by people who use drugs (without injecting), a population that is often overlooked (Stone et al. 2022). It has been hypothesized that PWID may have greater HIV exposure through sex because their sex partners are more likely to be other PWID living with HIV (Kurth et al. 2015, Stone et al. 2022). However, these results suggest that transmission involving the GP and FSW are important avenues of HIV acquisition for PWID, suggesting that strategies that only address transmission between PWID are not sufficient. Key populations are groups that face an elevated risk of living with HIV and/or barriers to accessing services or care (World Health Organization 2022). While tailored approaches are often needed to address the unique needs of different key populations, which should not automatically lead us to assume that the epidemics among these populations are isolated (Smith et al. 2009).

Transmission among PWID and other populations

We found mixing between HIV-1 sequences from the GP and PWID population and (Nduva et al. 2022a) a general trend of greater transmission to (versus from) the PWID population, although this trend was largely not significant in a region-stratified analysis. Similarly to a prior study, we calculated that transmission to PWID populations from the larger (not-PWID) population was more common than from PWID populations to the larger population (although we caution that differences in sampling times may bias results). This prior research found that most sequence clusters (88.5%) were population-specific (using similar categories to ours) and that the majority of between-population transmissions were from the general population to key populations (82.9%) (Nduva et al. 2022a). In contrast, and despite a sampling density higher than for most other population groups, we found relatively few PWID-specific clusters.

Our results suggest the need to better understand and address transmission involving FSW and PWID and, particularly, the GP and PWID (interestingly, we found one prior epidemiological model had entirely excluded this route of transmission from its considered parameters) (Strathdee et al. 2010). More research is needed to uncover primary modes and mechanisms of transmission between populations and to understand transmission patterns for people who fit into multiple key populations. For example, many women who inject drugs report exchanging sex for money or goods, but they are not classified as sex workers in this analysis and may not be reached by most services for female sex workers (Monroe-Wise 2023).

Acquisition risk and transmission risk are different, (Patel et al. 2014) and our findings may suggest that PWID were historically more likely to acquire than transmit HIV. Nevertheless, as many risk factors of HIV acquisition are also risk factors for transmission, such findings are somewhat surprising in light of the high prevalence of HIV among PWID and may reflect possible bias from the sequence sampling timeframe for different population groups. Other phylogenetic studies have contradicted the previously common belief that populations with high burdens of HIV are likely to be sources of cases in the larger population (Gogia et al. 2019). For example, evidence suggests that, despite an HIV prevalence of 40%, fishing villages are a sink, rather than sources, of HIV cases in Uganda (Bbosa et al. 2019, Ragonnet-Cronin et al. 2019). It's important to note that estimates we present are relative to population size; as PWID are a small minority (0.2–0.3% of the population of the coast and Nairobi, respectively), the estimated absolute amount of transmission from PWID would be even lower than what we report here.

Table 3. Transitions between population groups using maximum likelihood and Bayesian tree-building and ancestral state reconstruction methods.

	Sequence counts *N* (% of total sequences)			Transitions/Jumps: full tree *N* (% of total branches)*		Significance/Support	
Subtype	**Total**	**Group 0**	**Group 1**	**0 -> 1**	**1 -> 0**		
ML						Directional *P*-value (2-tailed)*	
All between group transitions							
All		296 (100)		293 of 584 total branches (50.2)			
A1		**232** (100)		**236 of 462 total branches (51.1)**			
C		28 (100)		28 of 53 total branches (52.8)			
D		36 (100)		29 of 69 total branches (42.0)			
		FSW	PWID	FSW -> PWID	PWID -> FSW	FSW -> PWID	
All	148 (50)	74 (25)	74 (25)	26 (10–43) (5.4)	19 (12–21) (3.3)		
A1	**116 (50)**	**58 (25)**	**58 (25)**	**23 (14–37) (5.0)**	**15 (10–22) (3.2)**	**<.01**	
C	14 (50)	7 (25)	7 (25)	2 (0–4) (2.8)	2 (1–4) (4.1)	.01	
D	18 (50)	9 (25)	9 (25)	1 (0–2) (0.9)	2 (1–3) (2.6)	<.01	
		MSM	PWID	MSM -> PWID	PWID -> MSM	PWID -> MSM	
All	148 (50)	74 (25)	74 (25)	16 (5–26) (2.7)	18 (9–26) (3.1)		
A1	**116 (50)**	**58 (25)**	**58 (25)**	**11 (5–14) (2.3)**	**10 (6–14) (2.1)**	**.53**	
C	14 (50)	7 (25)	7 (25)	1 (1–4) (2.7)	3 (2–4) (4.8)	<.01	
D	18 (50)	9 (25)	9 (25)	1 (0–5) (2.0)	4 (1–6) (5.2)	<.01	
		GP	PWID	GP -> PWID	PWID -> GP	GP -> PWID	
All (584)	148 (50)	74 (25)	74 (25)	35 (22–54) (6.0)	24 (16–33) (4.1)		
A1 (462)	**116 (50)**	**58 (25)**	**58 (25)**	**31 (21–44) (6.8)**	**17 (12–22) (3.1)**	**<.01**	
C (53)	14 (50)	7 (25)	7 (25)	4 (0–7) (7.3)	4 (2–6) (7.1)	.08	
D (69)	18 (50)	9 (25)	9 (25)	3 (0–6) (4.7)	4 (2–7) (6.1)	.01	
Total		Not-PWID	PWID	Not-PWID -> PWID	PWID -> Not-PWID		
All	295 (100)	222 (75)	74 (25)	77 (37–114) (13.2)	61 (37–80) (10.4)		
A1	**232 (100)**	**174 (75)**	**58 (25)**	**65 (40–95) (14.1)**	**42 (28–48) (8.4)**		
C	27 (100)	21 (75)	7 (25)	7 (1–15) (12.8)	9 (5–14) (16.0)		
D	36 (100)	27 (75)	9 (25)	5 (0–13) (7.6)	10 (4–16) (13.9)		
Bayesian						Bayes factor	
All between-group transitions							
A1	**232** (100)			**243 of 462 total branches (52.5)**			
		FSW	PWID	FSW -> PWID	PWID -> FSW	FSW -> PWID	PWID -> FSW
A1	**116 (50)**	**58 (25)**	**58 (25)**	**37 (27-47) (8.1)**	**9 (5-17) (1.9)**	**15.3**	**1.5**
		MSM	PWID	MSM -> PWID	PWID -> MSM	MSM -> PWID	PWID -> MSM
A1	**116 (50)**	**58 (25)**	**58 (25)**	**3 (2–6) (0.6)**	**8 (5–12) (1.6)**	**2.9**	**1.3**
		GP	PWID	GP -> PWID	PWID -> GP	GP-> PWID	PWID -> GP
A1	**11**6 (50)	**58 (25)**	**58 (25)**	**24 (12–35) (5.2)**	**7 (3–17) (1.5)**	**10.6**	**1.2**
Total		Not-PWID	PWID	Not-PWID -> PWID	PWID -> Not-PWID	Not-PWID -> PWID	PWID -> Not-PWID
A1	**232 (100)**	**174 (75)**	**58 (25)**	**65 (58–78) (14.1**	**23 (17–40) (5.0)**		

*P-values test for disproportionate transitions in either direction based on the statistic: transitions from state 0 > state 1)/(transitions from state 1 > state 0. The null distribution is generated from randomly resampling traits on the tree tips 200 times.
ML trees are down-sampled to have equal numbers of sequences from each group, and counts are averaged across subtrees (10 for subtype A1 and 30 for subtypes C and D) and 20 resolutions of ancestor states. For Bayesian analyses, counts are averaged across the 1000 highest posterior probability trees for 7 subtrees. *P*-values test for disproportionate transitions in either direction, resampling traits on the tree tips 200 times to derive the null distribution. Support for Markov jumps is assessed via BF. Supplementary Table S11 presents a summary of transition counts limited to terminal branches.

Regional transmission trends among PWID

We estimated that transmission between the coast and Nairobi regions was primarily driven by transmission among the larger (not PWID) population and that regional transmission trends among PWID likely reflect trends in the larger population, rather than the impacts of behaviors specific to PWID. We were particularly interested in trends in coastal Kenya, where HIV prevalence in the general population is lower than in Nairobi, but where rates of injection drug use (and the prevalence of HIV among PWID) are higher (Beckerleg et al. 2005). However, we estimated low rates of transmission from PWID to the GP even in the coastal region. Between-region transmission is likely to continue to increase as the world becomes increasingly more mobile, and changes in HIV prevalence in the coast or in Nairobi will impact the GP and PWID populations in the other region.

Patterns of regional transmission in our data contrast findings in a prior study that showed most HIV transmission is from West to East (and more from Nairobi to the coast; Nduva et al. 2022a). These differences in findings may be due to different model assumptions or because we restricted sequences to two regions compared to eight regions in the previous study—which only had sequences from 58 PWID enrolled in Mombasa in 2010. We also estimated transmissions between the GP and PWID population to occur at similar rates for between versus within-regions. This result is likely due to low sampling density of the transmission network between the GP and PWID population, rather than truly equal frequencies of cross-region and between-region transmission for these populations. Among PWID, our two methods yielded conflicting estimates for the prevailing regional direction of transmission. Although HIV prevalence is estimated to be higher for

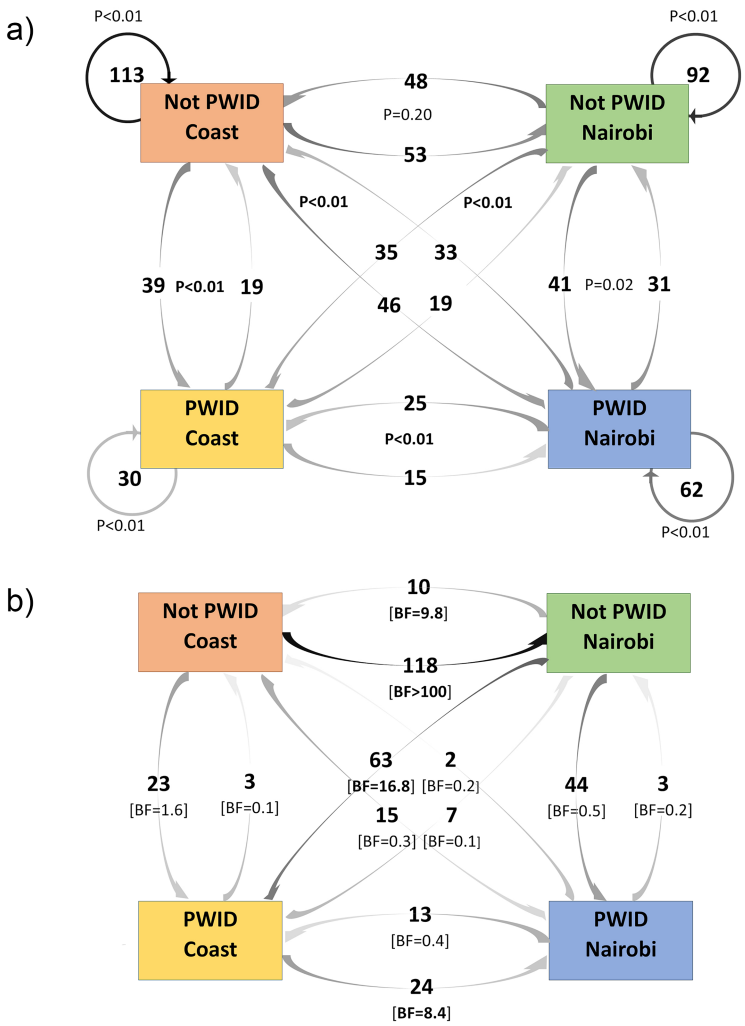

Figure 3. HIV-1 A1 transitions/Markov jumps between PWID and not-PWID populations from the coast and Nairobi. For the not-PWID population, GP, FSW, and MSM sequences are sampled proportionate to the estimated number of HIV cases in each population in the given region, with the majority of sequences coming from the GP population. Results are reported under ML (a) and Bayesian (b) models. *P*-values test for disproportionate transitions in either direction, resampling traits on the tree tips 200 times to derive the null distribution. For Bayesian models, BF >3 indicates strong support for Markov jumps in the given direction. Grayscale reflects the estimated amount of transmission in each direction. Data are presented in table format in Supplementary Table S10.

PWID on the coast, Kurth et al. estimated a higher annual HIV incidence in Nairobi (2.5%) compared to the coast (1.6%, in the year 2012), which is consistent with the more frequent of within-group transmissions we observed for PWID in Nairobi versus PWID from the coast under ML methods [National AIDS and STI Control Programme (NASCOP) 2020, Kurth et al. 2015].

Strengths and limitations

Despite this being the largest phylogenetic study of the PWID HIV-1 sub-epidemic in an African setting, the study has some limitations. Although the use of APS allowed us to reach PWID beyond those already accessing harm reduction services (a common limitation in studies involving PWID), high viral suppression rates limited sequencing success (28%) and, therefore, sampling density; this is also a source of selection bias. While those with viral suppression are unlikely to transmit HIV, their sequences would also have provided valuable information on older HIV-1 transmission networks. Another limitation is that, while the *pol* region is commonly used in HIV alignments and is sufficient for phylogenetic inference (Drescher et al. 2014, Hassan et al. 2017, Ragonnet-Cronin et al. 2018), sequencing of a larger region and the availability of replicates for all sequences would have provided more robustness to conclusions (Yebra et al. 2016). High variation in existing estimates for key population sizes and HIV prevalence also limits our ability to interpret our results in terms of sampling density or to estimate absolute contributions to transmission. Another limitation is that outside of the SHARP study, key population descriptors are primarily based on study enrollment criteria and include limited additional behavioral data. This creates potential for misclassification, as does the possibility of non-reporting because sex between men, sex work, and drug use are all criminalized in Kenya.

The sampling scheme used in this study has both advantages and disadvantages. The use of assisted partner services in collaboration with community-embedded peer-educators allowed us to recruit participants who might otherwise be difficult to reach. This approach may have helped address a common problem in molecular epidemiology of over-sampling individuals with high healthcare engagement. However, sequences from our study were almost entirely restricted to PWID and were collected more

recently than the sequences from FSW, MSM, and the GP incorporated primarily from prior studies. While this could create sampling bias, our results trended in the opposite direction that we would expect from selective sampling, with substantial mixing of PWID sequences with sequences collected from other populations. It is, nevertheless, possible that differences in sampling time-biased estimates of the direction of transmission and could specifically explain why we estimated more transmission from FSW and the GP (the populations with the oldest sequences) to the other populations.

By assessing transmission dynamics with three approaches—cluster analysis and both maximum likelihood and Bayesian ancestral state reconstruction—we were able to assess the robustness of our findings to different models, but not to resolve the reason for model-based discrepancies. ML and Bayesian models each have their own strengths for ancestral state reconstruction—Bayesian methods may be better placed to make source inference when sampling is incomplete (de Silva et al. 2012, Volz 2012) as is usually the case, but can be sensitive to the choice of dispersal rate prior and does not resolve all instances of sampling bias (Gao et al. 2023). We were also not able to resolve the difference in some regional transmission trends we observed compared to prior research and were unable to include sequences from other regions in our regional trends analysis, as no PWID sequences were available from these regions.

Conclusion

In East Africa, the availability of injection drugs has increased drastically in the last 30 years, but despite the disproportionately high prevalence of HIV among PWID, studies of HIV transmission rarely include this population. We recruited PWID through injecting partner networks; however, we found only low levels of linked transmission in this population. This suggests relatively low rates of recent parenteral transmission and supports the value of needle-syringe programs Because the epidemic among PWID and the GP are inter-related, interventions within the larger population, where we also observed the most transmission between regions, may have carry-over benefits for reducing HIV prevalence among PWID. HIV harm reduction services for this population must address risk factors for acquisition and transmission beyond injection drug use. There is also a need to better understand the environment within which transmission from the GP to the PWID population occurs. Lastly, feedback from people belonging to multiple key populations (particularly FSW who also inject drugs) is needed to understand if they experience unique acquisition or transmission risk factors.

Author contributions

J.T.H., C.F., and D.B. conceptualized the parent study. B.S. and B.L.G. managed the study data. H.K., J.T.H., and G.N. devised the analytical approach. J.G., E.W., and T.O. conducted sequencing. G.N. sourced and assembled metadata on HIV-1 sequences from prior studies. H.K. implemented analyses and wrote the manuscript, and coauthors gave approval to the manuscript.

Supplementary data

Supplementary data is available at *VEVOLU* online.

Conflict of interest: J.S. has received advisory fees from Gilead Sciences. All other authors state that they have no competing interests.

Funding

This work was supported by the National Institutes of Health (grant number R01 DA043409 and T32 GM81062 to H.K.) and Fogarty International Center (grant number D43 TW009580 to L.M. and S.M.). Funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Data availability

The HIV-1 sequence data collected as part of the parent study are available in the GenBank repository (accession numbers: OQ299131-OQ299439). Prior published HIV-1 reference sequences are a subset of the following: AF457085, FJ865384, HQ993685, HQ993707, HQ993955, HQ993975, JN011936-JN011994, JN628466-JN630893, JQ410388-JQ410431, JQ616926-JQ698429, JQ914101-JQ914103, JX123572-JX123678, KC018519-KC018954, KC517014, KC517047, KC568501-KC568530, KC900525-KC900816, KF544155, KF544276-KF544282, KF716468-KF716477, KF781839-KF781850, KJ395348, KJ502114-KJ502170, KM016220-KM016223, KM391677-KM391723, KM853096-KM853149, KP071681-KP071727, KR086420, KR138541-KR138543, KR872428-KR872543, KT213607-KT213653, KU749431, KU753728-KU753792, KX505365, KY062096-KY062142, KY364286-KY364337, MK192577-MK192628, MT084914-MT085067, OM109696-OM110282. Study materials, code, and data that support the findings of this study are available from the corresponding author (H.K.) on request.

References

Akiyama MJ, Cleland CM, Lizcano JA *et al*. Prevalence, estimated incidence, risk behaviours, and genotypic distribution of hepatitis C virus among people who inject drugs accessing harm-reduction services in Kenya: a retrospective cohort study. *Lancet Infect Dis* 2019;**19**:1255–63.

Asher AK, Hahn JA, Couture MC *et al*. People who inject drugs, HIV risk, and HIV testing uptake in Sub-Saharan Africa. *J Assoc Nurses AIDS Care* 2013;**24**:e35–e44.

Bbosa N, Ssemwanga D, Nsubuga RN *et al*. Phylogeography of HIV-1 suggests that Ugandan fishing communities are a sink for, not a source of, virus from general populations. *Sci Rep* 2019;**9**:1–8.

Beckerleg SE, Hundt GL. The characteristics and recent growth of heroin injecting in a Kenyan coastal town. *Addict Res Theory* 2004;**12**:41–53.

Beckerleg S, Telfer M, Hundt GL. The rise of injecting drug use in east Africa: a case study from Kenya. *Harm Reduct J* 2005;**2**:1–9.

Bezemer D, Faria NR, Hassan A *et al*. HIV type 1 transmission networks among men having sex with men and heterosexuals in Kenya. *AIDS Res Hum Retrovir* 2014;**30**:118–26.

Bennett DE, Camacho RJ, Otelea D. Stanford University. *HIV Drug Resistance Database: World Health Organization 2009 List of Mutations*. https://hivdb.stanford.edu/page/who-sdrm-list/.

Bielejec F, Baele G, Vrancken B *et al*. SpreaD3: interactive visualization of spatiotemporal history and trait evolutionary processes. *Mol Biol Evol* 2016;**33**:2167–69.

Brenner BG, Ibanescu RI, Osman N *et al*. The role of phylogenetics in unravelling patterns of HIV transmission towards epidemic control: the Quebec experience (2002–2020). *Viruses* 2021;**13**:1643.

Brodish P, Singh K, Rinyuri A *et al*. Evidence of high-risk sexual behaviors among injection drug users in the Kenya PLACE study. *Drug Alcohol Depend* 2011;**119**:138–41.

Degenhardt L, Peacock A, Colledge S *et al*. Global prevalence of injecting drug use and sociodemographic characteristics

and prevalence of HIV, HBV, and HCV in people who inject drugs: a multistage systematic review. *Lancet Glob Health* 2017;**5**: e1192–e1207.

de Silva E, Ferguson NM, Fraser C. Inferring pandemic growth rates from sequence data. *J R Soc Interface* 2012;**9**:1797–808.

Drescher SM, von Wyl V, Yang W-L *et al*. Treatment-naive individuals are the major source of transmitted HIV-1 drug resistance in men who have sex with men in the swiss HIV cohort study. *Clin Infect Dis* 2014;**58**:285–94.

Drummond AJ, Ho SY, Phillips MJ *et al*. Relaxed phylogenetics and dating with confidence. *PLoS Biol* 2006;**4**:e88.

Drummond AJ, Nicholls GK, Rodrigo AG *et al*. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 2002;**161**:1307–20.

Edwards CJ, Suchard MA, Lemey P *et al*. Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr Biol* 2011;**21**:1251–58.

Gao J, May MR, Rannala B *et al*. PrioriTree: a utility for improving phylodynamic analyses in BEAST. *Bioinformatics* 2022;**39**:btac849.

Gao J, May MR, Rannala B *et al*. Model misspecification misleads inference of the spatial dynamics of disease outbreaks. *Proc Natl Acad Sci* 2023;**120**:e2213913120.

Gill MS, Lemey P, Faria NR *et al*. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol Biol Evol* 2013;**30**:713–24.

Gogia M, Lawlor C, Shengelia N *et al*. Hidden populations: discovering the differences between the known and the unknown drug using populations in the Republic of Georgia. *Harm Reduct J* 2019;**16**:1–12.

Grenfell BT, Pybus OG, Gog JR *et al*. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 2004;**303**:327–32.

Guidelines for HIV Testing. *Guidelines for HIV Testing and Counselling in Kenya: 2nd edition*. 2010. http://guidelines.health.go.ke/ (10 June 2022, date last accessed).

Guindon S, Dufayard JF, Lefort V *et al*. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;**59**:307–21.

Hassan AS, Pybus OG, Sanders EJ *et al*. Defining HIV-1 transmission clusters based on sequence data. *AIDS* 2017;**31**:1211–22.

Joint United Nations Programme on HIV/AIDS. *UNAIDS 2022 Data*. Geneva, Switzerland: UNAIDS; 2022. https://www.unaids.org/en/resources/documents/2023/2022_unaids_data 29 August 2022,date last accessed.

Joy JB, Liang RH, McCloskey RM *et al*. Ancestral reconstruction. *PLoS Comput Biol* 2016;**12**:e1004763.

Junqueira DM, Sibisi Z, Wilkinson E *et al*. Factors influencing HIV-1 phylogenetic clustering. *Curr Opin HIV AIDS* 2019;**14**:161–72.

Kalyaanamoorthy S, Minh BQ, Wong TKF *et al*. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;**14**:587–89.

Katoh K, Misawa K, Kuma K *et al*. MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res* 2002;**30**:3059–66.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–80.

Kearse M, Moir R, Wilson A *et al*. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012;**28**:1647–49.

Kurth AE, Cleland CM, Des Jarlais DC *et al*. HIV prevalence, estimated incidence, and risk behaviors among people who inject drugs in Kenya. *J Acquir Immune Defic Syndr* 2015;**70**:420–27.

Lemey P, Rambaut A, Drummond AJ *et al*. Bayesian phylogeography finds its roots. *PLoS Comput Biol* 2009;**5**:e1000520.

Magosi LE, Zhang Y, Golubchik T *et al*. Deep-sequence phylogenetics to quantify patterns of HIV transmission in the context of a universal testing and treatment trial – BCPP/Ya Tsie trial. *eLife* 2022;**11**:e72657.

Mburu G, Limmer M, Holland P. HIV risk behaviours among women who inject drugs in coastal Kenya: findings from secondary analysis of qualitative data. *Harm Reduct J* 2019;**16**:1–15.

Minh BQ, Schmidt HA, Chernomor O *et al*. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;**37**:1530–34.

Monroe-Wise A. Assisted partner services for PWID to identify partners living with HIV and HCV. In: *Oral Presentation at Conference on Retroviruses and Opportunistic Infections (CROI)*. *Conference Presentation Presented at the: Conference on Retroviruses and Opportunistic Infections (CROI)*. February 21, 2023, Seattle, WA.

Monroe-Wise A, Mbogo L, Guthrie B *et al*. Peer-mediated HIV assisted partner services to identify and link to care HIV-positive and HCV-positive people who inject drugs: a cohort study protocol. *BMJ Open* 2021;**11**:e041083.

National AIDS and STI Control Programme (NASCOP). *Key Population Mapping and Size Estimation in Selected Counties in Kenya: Phase 1 - Key Findings*. Nairobi, Kenya: Republic of Kenya, Ministry of Health, 2019.

National AIDS and STI Control Programme (NASCOP). *Ministry of Health, Kenya. Preliminary KENPHIA 2018 Report*. Nairobi: NASCOP, 2020. https://phia.icap.columbia.edu/kenya-preliminary-report/ (29 August 2022, date last accessed).

National AIDS and STI Control Programme (NASCOP). *National Guide for HIV/STI Programming with Key Populations*. https://www.icop.or.ke/wp-content/uploads/2016/10/KP-National-Guidelines-2014-NASCOP.pdf (Accessed 11 July 2022) Published 2014.

National AIDS Control Council, National AIDS & STI Control Program (NASCOP). *Kenya AIDS Response Progress Report 2018*. Nairobi, Kenya: Kenya Ministry of Health, 2018.

Nduva GM, Hassan AS, Nazziwa J *et al*. HIV-1 transmission patterns within and between risk groups in Coastal Kenya. *Sci Rep* 2020;**10**:6775.

Nduva GM, Otieno F, Kimani J *et al*. Quantifying rates of HIV-1 flow between risk groups and geographic locations in Kenya: a country-wide phylogenetic study. *Virus Evol* 2022a;**8**:veac016.

Nduva GM, Otieno F, Kimani J *et al*. Phylogeographic assessment reveals geographic sources of HIV-1 dissemination among men who have sex with men in Kenya. *Front Microbiol* 2022b;**13**:843330.

Nguyen LT, Schmidt HA, von Haeseler A *et al*. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;**32**: 268–74.

Oguya FO, Kenya PR, Ongecha F *et al*. Rapid situational assessment of people who inject drugs (PWID) in Nairobi and coastal regions of Kenya: a respondent driven sampling survey. *BMC Public Health* 2021;**21**:1549.

Osman S, Lihana RW, Kibaya RM *et al*. Diversity of HIV type 1 and drug resistance mutations among injecting drug users in Kenya. *AIDS Res Hum Retrovir* 2013;**29**:187–90.

Oster AM, Wertheim JO, Hernandez AL *et al*. Using molecular HIV surveillance data to understand transmission between subpopulations in the United States. *J Acquir Immune Defic Syndr* 2015;**70**:444–51.

Paradis E, Schliep K, Schwartz R. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019;**35**:526–28.

Patel P, Borkowf CB, Brooks JT *et al*. Estimating per-act HIV transmission risk: a systematic review. *AIDS* 2014;**28**:1509–19.

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. *Infect Genet Evol* 2013;**19**:337–48.

Quick J, Grubaugh ND, Pullan ST *et al*. Multiplex PCR method for MinION and illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 2017;**12**:1261–76.

Ragonnet-Cronin M, Hodcroft EB, Wertheim JO. Understanding disclosed and cryptic HIV transmission risk via genetic analysis: what are we missing and when does it matter? *Curr Opin HIV AIDS* 2019;**14**:205–12.

Ragonnet-Cronin M, Jackson C, Bradley-Stewart A *et al*. Recent and rapid transmission of HIV among people who inject drugs in Scotland revealed through phylogenetic analysis. *J Infect Dis* 2018;**217**:1875–82.

R: A Language and Environment for Statistical Computing. 2019. 3.6.1(Computer Program).

Rhodes T, Guise A, Ndimbii J *et al*. Is the promise of methadone Kenya's solution to managing HIV and addiction? A mixed-method mathematical modelling and qualitative study. *BMJ Open* 2015;**5**:e007198.

Royer-Carenzi M, Pontarotti P, Didier G. Choosing the best ancestral character state reconstruction method. *Math Biosci* 2013;**242**:95–109.

Sambai BC, Kingston H, Monroe-Wise A *et al*. Characteristics associated with HIV and hepatitis C seroprevalence among sexual and injecting partners of HIV positive persons who inject drugs in Nairobi and coastal Kenya. *BMC Infect Dis* 2022;**22**:73.

Schliep KP. Phangorn: phylogenetic analysis in R. *Bioinformatics* 2011;**27**:592–93.

National Syndemic Disease Control Council, Simat P. *HIV Situation in Kenya. National Syndemic Disease Control Council*. 2022 https://nsdcc.go.ke/hiv-situation-in-kenya/ 8 August 2022, Published. (10 January 2024, date last accessed).

Smith AD, Tapsoba P, Peshu N *et al*. Men who have sex with men and HIV/AIDS in sub-Saharan Africa. *Lancet* 2009;**374**:416–22.

Soubrier J, Steel M, Lee MSY *et al*. The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol Biol Evol* 2012;**29**:3345–58.

Stone J, Fraser H, Walker JG *et al*. Modelling the impact of HIV and hepatitis C virus prevention and treatment interventions among people who inject drugs in Kenya. *AIDS* 2022;**36**:2191.

Strathdee SA, Hallett TB, Bobrova N *et al*. HIV and risk environment for injecting drug users: the past, present, and future. *Lancet* 2010;**376**:268–84.

Suchard MA, Lemey P, Baele G *et al*. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 2018;**4**:vey016.

Tun W, Sheehy M, Broz D *et al*. HIV and STI prevalence and injection behaviors among people who inject drugs in nairobi: results from a 2011 bio-behavioral study using respondent-driven sampling. *AIDS Behav* 2015;**19**:24–35.

Volz EM. Complex population dynamics and the coalescent under neutrality. *Genetics* 2012;**190**:187–201.

Wickham H, Navarro D, Pedersen TL. *Ggplot2: Elegant Graphics for Data Analysis*, 3rd edn. New York: Springer, 2016.

World Health Organization. *Consolidated Guidelines on HIV, Viral Hepatitis and STI Prevention, Diagnosis, Treatment and Care for Key Populations*. Geneva: World Health Organization, 2022.

Yebra G, Hodcroft EB, Ragonnet-Cronin ML *et al*. Using nearly full-genome HIV sequence data improves phylogeny reconstruction in a simulated epidemic. *Sci Rep* 2016;**6**:39489.

Yu G, Smith DK, Zhu H *et al*. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Meth Ecol Evolut* 2017;**8**:28–36.

Virus Evolution, 2024, 10(1), veae092 , DOI: https://doi.org/10.1093/ve/veae092, Advance Access Publication Date: 11 November 2024, Research Article
© The Author(s) 2024. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site–for further information please contact journals.permissions@oup.com.