Molecular Epidemiological Investigations of HIV Transmission Patterns Among People Who Inject Drugs in Kenya

Hanley Kingston

A dissertation

submitted in partial fulfilment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Carey Farquhar, Chair

Josh Herbeck

Elizabeth Bukusi

Program Authorized to Offer Degree:

Public Health Genetics

University of Washington

Abstract

Molecular Epidemiological Investigations of HIV Transmission Patterns Among People Who Inject Drugs in Kenya

Hanley Kingston

Chair of the Supervisory Committee:

Carey Farquhar

Departments of Global Health, Epidemiology, and Medicine

In Kenya people who inject drugs (PWID) have a 4-fold greater prevalence of HIV and at least 10-fold greater prevalence of hepatitis C compared to the general population. While preliminary epidemiological evidence suggests needle-sharing, and associated parenteral HIV transmission, is decreasing, more evidence is needed to evaluate these conclusions and to understand factors contributing to sexual HIV acquisition and transmission in this population. Phylogenetic and molecular epidemiology approaches can reveal transmission trends within and between populations to help determine the most impactful interventions; however, such studies must also be mindful of the vulnerability of populations, like PWID, to stigma or other group harms.

To characterize the HIV epidemic among PWID in Kenya, we first used cluster analysis and ancestral state reconstruction to estimate the following relative frequencies of HIV transmission: transmission within the PWID population; between the PWID population and other key populations (female sex workers and men who have sex with men); between the PWID population and the general population; and between PWID from the coastal region and Nairobi. Second, we investigated whether PWID shared an HIV or hepatitis C transmission network with the sexual or injecting partners they identified through assisted partner services, based on genetic distances of the virus sequences.

We found substantial mixing between HIV-1 sequences from PWID with those from other populations and no excess similarity between the HIV-1 sequences from pairs of individuals identified as injecting partners. These results support prior evidence of the effectiveness of needle-syringe programs and the shift towards sexual acquisition and transmission as an important factor in this epidemic for PWID. We propose a renewed emphasis on addressing risk factors for sexual transmission and on understanding the environment within which sexual HIV acquisition and transmission occurs for PWID. Further research on hepatitis C incidence and transmission risk factors could strengthen conclusions about changes in injecting behaviors.

Finally, we explored ethical issues in molecular epidemiological research of pathogen transmission trends, noting a lack of productive change coming from the current discourse centered around privacy risks to research subjects. We recommend that by treating pathogen sequence data as primarily a community resource, we can both increase the responsiveness of researchers to community concerns and improve community acceptance of molecular epidemiology as a public health and research tool.

ACKNOWLEDGEMENTS

I have to firstly thank my patient and supportive committee, Dr. Carey Farquhar, Dr. Josh Herbeck, Dr. Elizabeth Bukusi, and Dr. Nancy Jecker. Dr. Farquhar made time to give me guidance and advice even while balancing enormous duties as interim global health chair. She also gave me the wonderful opportunity to travel to Kenya, where I was able to explore this beautiful country and appreciate the incredible work and empathy that goes into running the harm reduction sites. As the phylogenetics expert on my committee, Dr. Herbeck shouldered most of my questions concerning my first two aims and gave me thoughtful and consistent guidance throughout. Dr. Bukusi and Dr. Jecker gave me valuable guidance in my ethical analysis, and I am extremely grateful that they shared their highly-sought-after expertise with me. I also received wonderful support and mentorship from Dr. George Nduva and Dr. Bhavna Chohan. I piggy-backed off a lot of Dr. Nduva's prior work, including his compilation of previously published HIV sequences from Kenya, and would never have figured out how to use BEAST without his expertise. Dr. Chohan shared with me her many years of knowledge working in HIV as well as her professional guidance.

The SHARP study, through which the analyzed data was collected, is of course the work of many more people, including other research team members and the health advisors and peer educators working on the ground. A special thanks to our SHARP study data coordinator, Betsy Sambai, for being my first collaborator and answering all my data questions. Lastly, a very sincere thank you to the research participants who donated their time and trusted us with their samples and data.

In my program, I want to thank Dr. Bruce Weir for his guidance, teaching, and helping me find employment. My amazing thesis advisor and first PI, Dr. Liz Blue helped me develop critical research skills. Thanks also to Dr. Alison Fohner for going above and beyond as an advisor and being so invested in the PHG program; Jenny Gogarten for going to bat for students; and Annique Atwater for meticulously answering my many queries and ensuring I got paid and met graduation requirements.

While conducting this research, I received financial support from the National Institutes of Health Statistical Genetics Training Grant (T32 GM81062). I also variously received support and assistance through the University of Washington Division of Medical Genetics, Department of Epidemiology, Institute of Public Health Genetics, Department of Biostatistics, and Department of Global Health.

I have an incredible and supportive family, Mom, Dad, Zach, Emily, and Charlotte. My parents have worked tirelessly to give me incredible opportunities. My dad is a wealth of sage wisdom. My mom has many times been my test audience and sounding board and never runs out of emotional support. My brother gives excellent advice and patiently helps me with computer calamities. And my sister-in-law is an incredibly thoughtful person and almost single-handedly kept our family connected from a distance during COVID. I also have amazing grandparents, Grandma and Grandpa Hepler, Grandma Dee Dee, and Papa Jack.

Finally, I have the most incredible group of friends near and far: Natalie, Cat, Katie, Abe, Dan, Gideon, Rachel, Chris, Petr, Katie, Haley, Mackenzie, Hwei Ling, Sanne, Georgia, Jillian, Alee, and Shradha. I appreciate your moral support; patience; telling me not to do a PhD, but still being mostly sympathetic as I complained about doing a PhD; haircuts of varying quality; making delicious food; trying new things with me; welcoming me to Seattle; being hiking and climbing buddies; and much else. Diane and Alyna get a special shoutout for being not only amazing friends but also helping me so much with paper edits and troubleshooting at critical junctures.

To those I inevitably left out, please know how much I appreciate your time, expertise, guidance, and friendship.

TABLE OF CONTENTS

Abstract **3**

Acknowledgements **5**
Table of Contents **7**
List of Figures **8**
List of Tables **9**
Introduction **11**
Chapter 1: A Phylogenetic Assessment of HIV-1 Transmission Trends among People Who Inject Drugs from Coastal and Nairobi, Kenya **15**
- Abstract 15
- Introduction 17
- Methods 18
- Results 24
- Discussion 27
- Tables & Figures 33

Chapter 2: Using HIV and Hepatitis C Molecular Epidemiology to Investigate Assisted-Partner Services Recruitment among People Who Inject Drugs in Kenya **40**
- Abstract 40
- Introduction 42
- Methods 43
- Results 47
- Discussion 49
- Tables & Figures 54

Chapter 3: The Ethics of Pathogen Molecular Epidemiology Research from a Community Data Perspective **59**
- Part 1: Beyond Privacy – Conceptualizing Pathogen Sequence Data as Community Data 60
- Part 2: Community Interests in Molecular Epidemiology Data 64
- Part 3: Examples of Community Data Models 69

Conclusion **76**
References **78**
Supplemental **90**

LIST OF FIGURES

Figure 1.1. Phylogenies of HIV-1 subtype A1 sequences, with sequences from PWID highlighted 34-35

Figure 1.2. HIV-1 A1 transitions / Markov jumps between PWID and not-PWID populations from the coast and Nairobi 39

Figure 2.1. Cascades of participant and partner pair counts for HIV analysis 54

Supplementary Figure S1.1. Phylogeny of HIV-1 sequences from the SHARP study 92

Supplementary Figure S1.2. Cluster distributions for population group and region 95

Supplementary Figure S1.3. Cluster size distributions by population and cluster-definition 96

Supplemental Figure S2.1. Phylogeny of index-partner pair connections among HIV-1 sequence data 105

Supplemental Figure S2.2. Percent of index-control pairs with higher sequence molecular distance than named partner pairs 108

Supplemental Figure S2.3. TN93 distance by (A) partner type and (B&C) time-since diagnosis 109

Supplemental Figure S2.4. Cascades of participant and partner pairs counts for HCV analysis 110

Supplemental Figure S2.5. Phylogeny of index-partner pair connections among HCV 4a sequence data 111

LIST OF TABLES

Table 1.1. Participant demographics and distribution of new (from SHARP study) and previously published Kenyan HIV-1 *pol* sequences by population 33

Table 1.2. HIV-1 sequences from other populations in clusters with HIV-1 sequences from PWID 36

Table 1.3. Transitions between population groups using maximum likelihood and Bayesian tree-building and ancestral state reconstruction methods 37-38

Table 2.1. Characteristics for all participants living with HIV and for the 151 participants included in this analysis 55

Table 2.2. Distribution of HIV-1 molecular distances across index-partner pairs stratified by partner-type 56

Table 2.3. Association between partner-pair characteristics and HIV-1 TN93 distance 57

Table 2.4. Distribution of HCV molecular distances across index-partner pairs, stratified by subtype 58

Supplementary Table S1.1. Demographics, test results, and characteristics of SHARP study participants with HIV-1 sequences included in the analysis 90

Supplementary Table S1.2. Participant demographics and distribution of new (from SHARP study) and previously published Kenyan HIV-1 *pol* sequences by region 91

Supplementary Table S1.3. Distribution of population groups within clusters that contain sequences from coastal Kenya or Nairobi 93

Supplementary Table S1.4. HIV-1 sequences from other populations in clusters with HIV-1 sequences from female PWID with transactional sex risk factors or in clusters with sequences from PWID who report sharing needles 94

Supplementary Table S1.5. HIV-1 A1 subtype transitions between coastal and Nairobi regions in a tree restricted to sequences from PWID (2018-2021) using maximum likelihood tree building and ancestral state reconstruction 97

Supplementary Table S1.6. HIV-1 A1 subtype transitions between coastal and Nairobi regions using maximum likelihood tree building and ancestral state reconstruction (terminal branches only) 98

Supplementary Table S1.7. HIV-1 transitions between population groups using maximum likelihood tree-building and ancestral state reconstruction (terminal branches only) 99

Supplementary Table S1.8. HIV-1 transitions between PWID and not-PWID populations from the coast and Nairobi using both maximum likelihood and Bayesian tree-building and ancestral state reconstruction models under uniform subsampling 100-102

Supplementary Table S2.1. Distribution of HIV patristic distances across index-partner pairs stratified by partner-type 103-104

Supplementary Table S2.2. Associations between partner-pair characteristics and HIV-1 TN93 distance, using various thresholds 109

Supplementary Table S2.3. Distribution of HCV molecular distances across index-partner pairs, stratified by subtype 112

INTRODUCTION

While around 1.4 million people in Kenya live with HIV, or about 4-6% of the total population,[1,2] the country has made important strides in HIV prevention and treatment, such as through the use of assisted partner services (APS) to identify people not aware of their HIV status or not optimally engaged in care. APS involves asking people living with HIV to identify and provide contact information for their sexual and injecting partners and has already been linked to increased HIV testing, diagnosis, and enrollment in care among partners of people living with HIV in Kenya.[3–6]

Continued progress requires multi-pronged approaches that address the specific needs of different populations.[7] For example, PWID have an estimated HIV prevalence about 4-fold higher than in the general Kenyan population (14.5% in Nairobi and 20.5% in the coastal region)[8–10] and face barriers to accessing testing and care. These include community stigma, stigma in health care settings, and fear of violence.[11–13] The high prevalence of HIV among PWID informs a common narrative that transmission from PWID "seeds" cases in the larger population; however, these claims are often not based on empirical evidence, and warrant more cautious investigation.[14–16] Molecular epidemiology can be a powerful tool to understand patterns of transmission. By combining estimates of the evolutionary distance between virus sequences with geographic or epidemiological data, we can estimate the amount and history of transmission between regions or different populations. When evaluated in the context of other epidemiological evidence, these estimates can suggest behaviors or environments that may be associated with transmission.[17,18]

Chapters 1 and 2 of this work use molecular epidemiological and phylogenetic approaches, like cluster analysis and ancestral state reconstruction, to estimate the primary mechanisms and sources of HIV acquisition and transmission for PWID, with the goal of understanding the future role for needle-syringe

program scale-up and/or other interventions for epidemic control. Chapter 3 departs from the application of molecular epidemiology methods to explore ethical issues in these types of analyses.

Prior research suggests needle sharing is an important, but decreasing, factor in HIV transmission among PWID in Kenya. Injectable heroin became prevalent in Kenya in the 1990's,[19] and in 2018, an estimated 18,000 people in Kenya injected drugs, the majority living in urban areas in Nairobi, Mombasa, and Kilifi counties.[1,20] Parenteral transmission was likely a substantial historical driver of HIV prevalence in this population, evidenced by high estimates of needle-sharing prior to 2013,[21–23] epidemiological modeling,[14] and some molecular evidence.[24]

However, Kenya has taken a number of steps to reduce harms associated with injection drug use, including introducing needle and syringe programs in 2013 and methadone maintenance treatment in 2014.[21,25] An estimated 13.5% of PWID were enrolled in opioid substitution therapy programs in 2018.[11] Although estimates vary, evidence supports that needle-syringe programs have reduced needle-sharing[1,9,21–23,26–29] and parenteral transmission, and epidemiologic models estimate a substantial number of averted HIV cases among PWID as a result of these programs.[30] Despite this progress, more research is needed to clarify the current extent of parenteral HIV transmission and to understand whether needle-syringe program scale up or other interventions will be most impactful for averting future HIV cases.

Sexual transmission may be one under-addressed factor in this epidemic for PWID, who have higher estimated frequencies of transactional sex (with women particularly likely to report receiving money for sex),[13,30] lower condom use than other populations,[1,11] and like other key populations, low knowledge about PrEP. More research is also needed to understand how programs used in the general population work for PWID and whether adaptations to these programs can improve effectiveness in this population. For example, the World Health Organization recommends the use of APS to reach PWID at elevated risk of living with HIV,[31] but specific data on APS performance and acceptability in this population is limited.

The Peer Educator Program, which was started in Nairobi to train former PWID to conduct outreach and provide harm reduction services, created an opportunity to tailor APS to the specific needs of the PWID population in the parent study to this analysis.[32,33] In Chapter 2, we specifically explore whether APS for PWID identifies sexual or injecting partners from a shared HIV transmission network.

Hepatitis C (HCV), which is more readily transmitted through needle-sharing than HIV, also presents a substantial health risk to PWID. At 13-22% among PWID accessing harm reduction services in the Nairobi and coastal regions, HCV prevalence is significantly higher than among the general population (prevalence <1%).[28] The WHO has endorsed the goal of 90% reduction in HCV incidence and 65% reduction in HCV-related deaths between 2016 and 2030,[34] which will require scaling up harm reduction services for PWID. Historically, a long asymptomatic period, lack of fast-acting and effective treatments, and concerns about reinfection[35] have presented barriers to translating programs like APS for HCV testing and treatment.[36] However, highly effective direct-acting antivirals for HCV were introduced to Kenya in 2021,[10] making the WHO's goal of rapid reduction in HCV cases feasible if we're able to reach people with testing and treatment. Although the parent study to this research project implemented APS only for HIV, the inclusion of HCV testing and treatment presented opportunities to explore the potential future function of APS for HCV among PWID, which we briefly discuss in Chapter 2.

The pathogen sequence data used in molecular investigations of disease transmission can come from different sources, like primary research, public health surveillance systems (for example, when sequencing is done clinically then shared with departments of public health), or deidentified sequences uploaded to publicly accessible databases. Critiques of molecular epidemiological methods in the US have raised concerns about privacy risks to individuals, particularly those whose sequences are collected through surveillance or shared to databases. However, there is disagreement and confusion about the extent of these privacy risks and a need for an ethical model that can identify and address salient concerns about this research without impeding its important public health goals. In Chapter 3, we explore how we can

address ethical issues in this research by recognizing pathogen sequence data as a type of community data.

CHAPTER 1: A Phylogenetic Assessment of HIV-1 Transmission Trends among People Who Inject Drugs from Coastal and Nairobi, Kenya

ABSTRACT

Background

Although recent modeling suggests needle-syringe programs 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 sub-epidemic and the general epidemic in Kenya is not well understood. Incorporating 5-times more HIV sequences from PWID in Kenya than in prior studies, we quantified rates and direction of HIV-1 transmission involving PWID and other populations from the coast and Nairobi regions.

Methods

We aligned 303 new (2018-2021) HIV-1 *pol* sequences from PWID and their sexual and injecting partners with 2,666 previously published Kenyan sequences. We used genetic similarity cluster analysis (thresholds: patristic distance <0.045) and maximum likelihood and Bayesian ancestral state reconstruction to estimate transmission histories at the group (female sex workers, men who have sex with men, PWID, or not key population) and regional (coast or Nairobi) levels. Transition counts estimate how often an ancestor sequence gave rise to a descendant sequence from a different population and/or region.

Results

In this cohort, 1,081 participants lived with HIV, of whom 274 (25%) were not virally suppressed and 303 (28%) had sequences available. Of new PWID sequences, 58% were in phylogenetic clusters, primarily interspersed with sequences from other key populations and from those not in key populations. Only 21% of clusters containing PWID sequences included a second PWID sequence. Ancestral state reconstruction identified substantial transmission between the coast and Nairobi regions and more not-PWID to PWID transmission than transmission in the other direction.

Conclusion

Despite recruiting PWID from local sexual and injecting networks, we found low levels of linked transmission in this population. This suggests relatively low rates of recent parenteral transmission and supports interventions to reduce sexual transmission while maintaining needle-syringe programs. Because the epidemic among PWID and other populations 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 in PWID. However, greater understanding of how PWID and non-PWID populations interact is needed.

INTRODUCTION

In the 1990's, global drug trafficking networks from Asia to Europe brought injectable heroin to the East African coast.[19] To address harms associated with injection drug use, including risk of parenteral HIV transmission, the Kenyan Government introduced needle and syringe programs (NSP) in 2013 and methadone maintenance treatment in 2014.[21,25] However, at about 14-21%, 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[8,10,30] and is similar to the prevalence among men who have sex with men (MSM), but lower than the estimated prevalence among female sex workers (FSW, 29.3%).[11] Studies have yet to resolve several important questions about the role that injection drug use has played in the HIV epidemic in Africa.[22]

Phylogenetic analyses are well-suited to resolve questions about HIV transmission patterns between populations, between regions, and over time.[37–40] By overlaying geographical or epidemiological data onto a phylogeny, we can also infer how the behaviors or experiences of individuals living with a disease relate to transmission trends. Rapid evolution, which occurs within the timescale of transmission, makes HIV particularly conducive to phylogenetic analysis.[41]

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 US and Western Europe.[22,27,30] These models suggest sexual transmission is an increasingly common form of transmission in this population, particularly for women.[30] However, the question of primary transmission mechanisms among PWID has not been addressed through molecular data. Questions also remain about the extent to which the HIV-1 epidemic among PWID is self-contained, versus 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 – and its relationship to the generalized epidemic – is also not well-understood.

Prior phylogenetic analysis of HIV-1 in Kenya primarily focused on not-PWID populations and showed higher HIV-1 transmission from the higher prevalence western regions to lower prevalence eastern regions.[24] However, because HIV prevalence among PWID is highest in coastal Kenya[9,26] (opposite of general regional prevalence trends) and because prior phylodynamic studies have shown differing patterns among other key populations,[42] it's not clear if this HIV-1 transmission trend duplicates among PWID. Prior research identified a cluster of 41 HIV-1 sequences collected from PWID in the coastal city of Mombasa in 2010.[24] This finding suggested substantial, isolated transmission within the PWID populations, with needle-sharing likely playing a role, but also contradicted a common narrative that transmission from PWID seeds cases in other populations.[16,24,43] Our study represents a 5-fold higher number of HIV-1 sequences from PWID in Kenya and is the first to collect HIV-1 sequences from PWID in Nairobi, allowing us to reassess transmission trends post-introduction of NSPs and also to assess regional trends in HIV-1 transmission among this population.

METHODS

Study population and enrollment

The Study of HIV, 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 live with HIV (indexes) from 2018 to 2021 and used assisted partner services to identify, test, and treat their sexual and injecting partners. Index participants were recruited from needle-syringe programs 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 assisted partner services (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. Partners were contacted with the help of peer educators and invited to enroll in the study, and those fitting index eligibility criteria could additionally enroll in APS.

Socio-demographic data, HIV and hepatitis C history, and sexual and injection drug use history were obtained for all participants via survey. Rapid HIV-1 testing using fingerstick samples was performed during the interview sessions following an established Kenya national algorithm.[44] Detailed study procedures are reported in the published study protocol.[4]

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 were shipped to the Kwazulu-Natal Research Innovation and Sequencing Platform (KRISP) laboratory at the University of KwaZulu-Natal, South Africa for Sanger sequencing or next generation sequencing (NGS).

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,[45] 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) and quantified using the Qubit dsDNA High Sensitivity assay kit on a Qubit fluorometer (Life Technologies, Carslbad, CA). Sequencing was performed on an Illumina MiSeq platform (Illumina).

Sequence dataset and phylogenetic determination of clusters

Subtypes were determined using REGA HIV-1 v3.[46] A total of 4058 previously published HIV-1 *pol* sequences were available from Kenya,[24] of which we incorporated 3587 into an alignment based on: year ≥2000 and not from a person <15 years old (where known). SHARP and previously published sequences were annotated as being from: PWID (people who had (ever) injected drugs), MSM, FSW, or not-KP (assumed not in a key population). We excluded all NRTI and NNRTI mutations listed on the Stanford Drug Resistance Database[47] 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.[48–50]

A two-pronged approach (i.e., maximum likelihood (ML) and Bayesian inference) was used to construct phylogenies for cluster analysis.[51,52] For the ML approach, we used IQ-Tree[53,54] and a maximum likelihood calculator using a general time-reversable substitution model with gamma-distributed rate variation (GTR+R(9)).[55,56] Cluster summaries were combined across HIV-1 subtype-specific maximum likelihood (ML) trees (N tips: SHARP subtype A1: 196, published subtype A1: 2650; SHARP subtype C: 37, 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.[57] We further implemented a Bayesian coalescent tree model using BEAST (v1.10.4) under the assumption of constant population size with a strict clock and inferred under the GTR + Γ4 substitution model.[37,58] We calculated Markov chain Monte Carlo (MCMC) runs with a chain frequency of 250-500 million generations, logging every 50K iterations, and discarding the first 10% as chain burn-in. Convergence was determined in Tracer V1.7.2 (defined as effective sample sizes (ESS) ≥100 for most trait rates) To estimate the dates of origin (time to most recent common ancestor; tMRCA) of clusters containing sequences from a PWID, we separately generated a maximum clade credibility trees in BEAST, specifying an uncorrelated relaxed clock.[37,59]

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 non-KP) that formed a previously-identified cluster,[24] as this cluster is a substantial outlier in terms of size and we wanted to draw inference in the larger 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 not-KP. 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: 40 sequence from coast and 76 sequence from Nairobi). The second approach assessed transmission between different population groups (74 sequence each from not-KP, FSW, MSM, PWID). The third approach estimated transitions between PWID and not-PWID from the coast and Nairobi regions (110 sequence each). As a secondary analysis to test agreement with prior studies, we assessed regional transmission between the coast and Nairobi in the total population. Analyses were performed for HIV-1 subtypes A1, C, and D where sample-size permitted.

For ML trees, sequences from other regions were incorporated as references but dropped prior to ancestral state reconstruction, with further filtering of the tree 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.[60,61] We resolved ancestor states to according to their 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). We defined a state transition when the trait of the descendant node differed from the state of its immediate ancestor on the tree. 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 resamplings 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 remove address uncertainty that develops for ancestral state reconstruction deeper in the tree, where states tend to converge near the tree root.

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.[58,62] 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.[63,64] Well-supported movements and Bayes factors (BF) assessing statistical support were summarized using SPREAD v0.9.6, (BF ≥3 was considered significant).[65] A robust counting approach implemented in BEAST was used 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.[37] We averaged jump counts across 7-9 subtrees.

Unless otherwise stated, we report transition counts or Markov jumps 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. For example, for four traits, there are 16 possible transition events (counting A->B and B->A as

separate events) and each transition event has a null probability of 6.25%. Transitions or jumps are reported in the text for the A1 subtype.

Nucleotide sequence accession numbers

The HIV-1 sequence data collected as part of the parent study will be available in the GenBank repository (accession numbers: OQ299131-OQ299439) prior to publication. Prior published HIV-1 reference sequences are a the 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 (HK) on reasonable request.

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 to participate and have their data published.

Author Contributions

JTH and CF conceptualized the parent study. BS and BLG managed the study data. HK, JTH, and GN devised the analytical approach. JG, EW, and TO conducted sequencing. GN sourced and assembled metadata on HIV-1 sequences from prior studies. HK implemented analyses and wrote the manuscript, and coauthors gave approval of the manuscript.

Funding

This work was supported by U.S. National Institutes of Health (NIH), grant number R01 DA043409. The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

RESULTS

Study participants and HIV-1 sequence data

This study enrolled 1081 participants living with HIV, of whom 274 (25%) did not have viral suppression (viral load >1,000 copies/ml). We were able to sequence 313 samples; we excluded 1 sequence for missing epidemiological data, 2 for poor alignment, and 7 for >15% missing or ambiguous bases, leaving 303 HIV-1 sequences.

Of the 303 participants with a sequence available, most (80%) were diagnosed prior to study enrolment, with a median time since diagnosis of 4 years (Supplementary Table S1.1). The vast majority had ever injected drugs (96%) and 81% had injected in the previous month, but recent needle-sharing was <10% in both regions. While receiving money or goods for sex was more common among women in Nairobi, participants on the Kenyan coast reported more sexual partners, on average. We did not ask about recent sex work, but 53 female participants had received money or goods for sex and reporting ≥3 sex partners

in the prior month. Participants from the coast were significantly more likely to report lack of transportation as a barrier to accessing HIV care (53% vs 31%, P<0.001).

After incorporating previously published sequences, subtype A1, C, and D Sequences were available from 148 FSW, 300 MSM, 295 PWID, and 1598 not-KP individuals from the coast and Nairobi regions (Table 1.1). Most HIV-1 sequences from FSW, MSM, and not-KP sequences were from the previously published sequence dataset, 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 S1.2).

Cluster analysis

Using a maximum patristic distance threshold of 0.045, there were 680 phylogenetic clusters (subtypes A1, C, or D) 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 1.2, and Supplementary Figure S1.2). The mean estimated year of origin 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 was relatively low: 1955 (IQR: 1939-1968).

Overall, 142 (58%) of new (2018-2021) and 190 (63%) of all sequences from PWID fell into clusters with other sequences. The previously identified cluster of 41 sequences from PWID in Mombasa represented the largest cluster in our data (and the only cluster with >5 PWID sequences); the next-largest cluster contained sequences from 13 individuals (Supplementary Figure S1.3). Sequences from PWID were not more likely to cluster with any one other population or (apart from the previously identified large cluster) with other PWID sequences. Sequences from PWID were found in 16% of all clusters containing at least 1 not-KP sequences, 12% of all clusters containing FSW sequences, and 14% of clusters containing MSM sequences (Table 1.2, Supplementary Table 1.3). Despite most PWID sequences being collected through partner referral within this single study, only 21% of clusters with PWID sequences contained another

sequence from a PWID. 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 S1.4). Population-specific clustering was, however, much more common when using the more stringent <0.015 distance threshold (Table 1.2, Supplementary Figure S1.2)

Transmission among PWID and other populations

After adjusting for population size via uniform subsampling, ancestral state reconstruction in the A1 subtype suggests while about half of transmissions occurred within populations (ML: 48.9%; Bayesian: 47.5%; null: 25%); however, 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 1.3, Supplementary Table S1.7), with ML methods suggesting most transmission to the PWID population came from the not-KP population (31, P<0.01) and Bayesian methods estimating the highest number of jumps (37, BF=15.3) to PWID coming from FSW. Transitions between PWID and MSM were rare in either direction.

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 S1.8).

Regional transmission trends among PWID

Ancestral state reconstruction restricted to newly collected PWID sequences (Supplementary Table S1.5) estimated substantial between-region transitions (31.8% of terminal branch transitions; null: 50%) and similar rates of transitions in either direction (P=0.05), similar to the pattern of regional transitions observed in the total population (Supplementary Table S1.6). To further investigate drivers of regional transmission, we conducted a combined analysis looking at the relationship between the PWID and not-

PWID populations by region (Figure 1.2, and Supplementary Table 1.S7), which suggested that most between-region transmission occurred within the not-PWID, rather than the PWID, population. In the not-PWID population, ML methods suggested similar frequency in both directions for the (7.5% coast to Nairobi vs 6.8% Nairobi to coast, P=0.20; null: 6.25% each), while Bayesian methods suggested substantially greater transmission from the coast to Nairobi (16.9% vs 1.4% Nairobi to coast, BF>100) (Figure 1.2, Supplementary Table S1.8).

Transitions were consistently (although not always significantly) more common from the not-PWID populations to PWID populations (vs from PWID to not-PWID) both within and between both regions. Specifically, within the coast, not-PWID 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 not-PWID and PWID population 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.[20] Our analysis reveals that the PWID HIV-1 sub-epidemic is highly connected to that in the not-KP and FSW populations and that these populations contribute a larger fraction of HIV-1 transmission to PWID than vice versa.

Mechanisms of transmission among PWID

Our finding of few PWID-exclusive clusters provides the first molecular epidemiological supporting limited parenteral transmission among PWID in recent years, in Kenya and in an African setting. Compared to

earlier studies, and given that we recruited from sexual and injecting partner networks in recent years (2018-2021), our study was better powered to resolve recent PWID HIV-1 transmission clusters involving Kenyan PWID.[43] Nevertheless, the only cluster we identified containing >4 PWID sequences was a previously described 41-sequence cluster in Mombasa.[43,66]

Our results show that the HIV epidemic among PWID is not self-sustained. Strategies to reduce parenteral transmission between PWID, while effective, are alone, not sufficient to address high HIV-1 prevalence in this population. Minimal 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.[21,25] Epidemiological studies support this hypothesis, with 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),[9,21,23,29] while studies conducted after 2015 estimated a much lower prevalence of 2-5% per month (2-12% at last injection) within these regions.[1,26,28] Recent modeling estimated that needle-syringe programs reduced HIV transmission among PWID in Kenya by 40-46% in 2020.[30] Nevertheless, prior research, pre- and post-introduction of needle syringe programs, shows that HIV prevalence increases with number of years injecting.[9,26] 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.[22] 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.[9] 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 most recent transmission among PWID is through sex.[29,30] That PWID likely face elevated risk of HIV acquisition even (and possibly primarily 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.[30] It has been hypothesized that PWID may have greater HIV exposure through sex because their sex partners (for example, other PWID) are more likely to live with HIV.[9,30] However, we estimate that most transmission to PWID actually comes from the general population (and to a lesser-extent, FSW), suggesting that strategies that only address transmission within the PWID population are not sufficient.

Transmission among PWID and other populations

We found substantial mixing between HIV-1 sequences from PWID and not-PWID populations and a general trend of greater transmission to (vs from) the PWID population. Prior research found 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%).[24] In contrast, we found relatively few PWID-specific clusters, but similarly calculated that transmission to PWID populations from the larger (not-PWID) population was more common than from PWID populations to the larger population.

While our results suggest greater need to address transmission between FSW and PWID, after accounting for differences in the size of the population living with HIV, we conclude that most transmission into the PWID population comes from the general population (interestingly, we found one prior epidemiological models had entirely excluded this route of transmission from its considered parameters).[14] 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 often report exchanging sex for goods, but they are not classified as sex workers in this analysis and may not be reached by most services for female sex workers.[67]

Acquisition risk and transmission risk are different.[68] Our findings suggest that the disproportionately high HIV prevalence among PWID may reflect greater risk of acquisition more so than greater risk of

transmission. Nevertheless, as many risk factors of HIV acquisition are also risk factors for transmission, our findings are somewhat surprising in-light of the high prevalence of HIV among PWID. Other phylogenetic studies have also contradicted the previously common belief that populations with high burdens of HIV are likely to be sources of cases in the larger population.[15] For example, robust evidence suggests that, despite an HIV prevalence of 40%, fishing villages are a sink, rather than sources, of HIV cases in Uganda.[27,39] 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.

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) is higher.[19] However, we estimated low rates of transmission from PWID to non-PWID 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 not-PWID and PWID populations in the other city.

Patterns of regional transmission in our data contrast a prior study that study showed most HIV transmission is from West to East (and more from Nairobi to the coast).[24] These differences in findings may be due to different model assumptions or because we restricted sequences to two regions versus 8 regions as was the case in the previous study – which was restricted to samplings from PWID from 2010. We also estimated cross-region and within-region transitions between the not-PWID and PWID population at

similar rates to within-region transitions, which is unlikely and may have resulted from low sampling density within the relevant transmission networks. Among PWID, our 2 methods yielded different estimates for the prevailing regional direction of transmission. Although HIV prevalence is estimated to be higher for 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 transitions we observed for PWID in Nairobi versus PWID from the coast under ML methods.[9,10]

Strengths and Limitations

Despite this being the largest study to analyze the PWID HIV-1 sub-epidemic in the African setting, the study has some limitations. High viral suppression rates led to lower sequencing success and lower sampling density. While those with viral suppression are unlikely to transmit HIV, their sequences would also have provided valuable information on older HIV-1 transmission networks and on HIV acquisition. Another limitation is that outside of the SHARP study, key population descriptors are primarily based on study enrollment criteria. This creates potential for misclassification, and we did not have data on whether other individuals belonged to multiple key populations.

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 not-KP incorporated primarily from prior studies. While this could create sampling bias, our results trended in the opposite direction that we'd 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 not-KP (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 model assumptions. Nevertheless, we were 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. Despite recruiting PWID through injecting partner networks, we found 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 epidemics among PWID and the general population 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 not-PWID population 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/transmission risk factors.

TABLES AND FIGURES

Demographics and sequence data summary

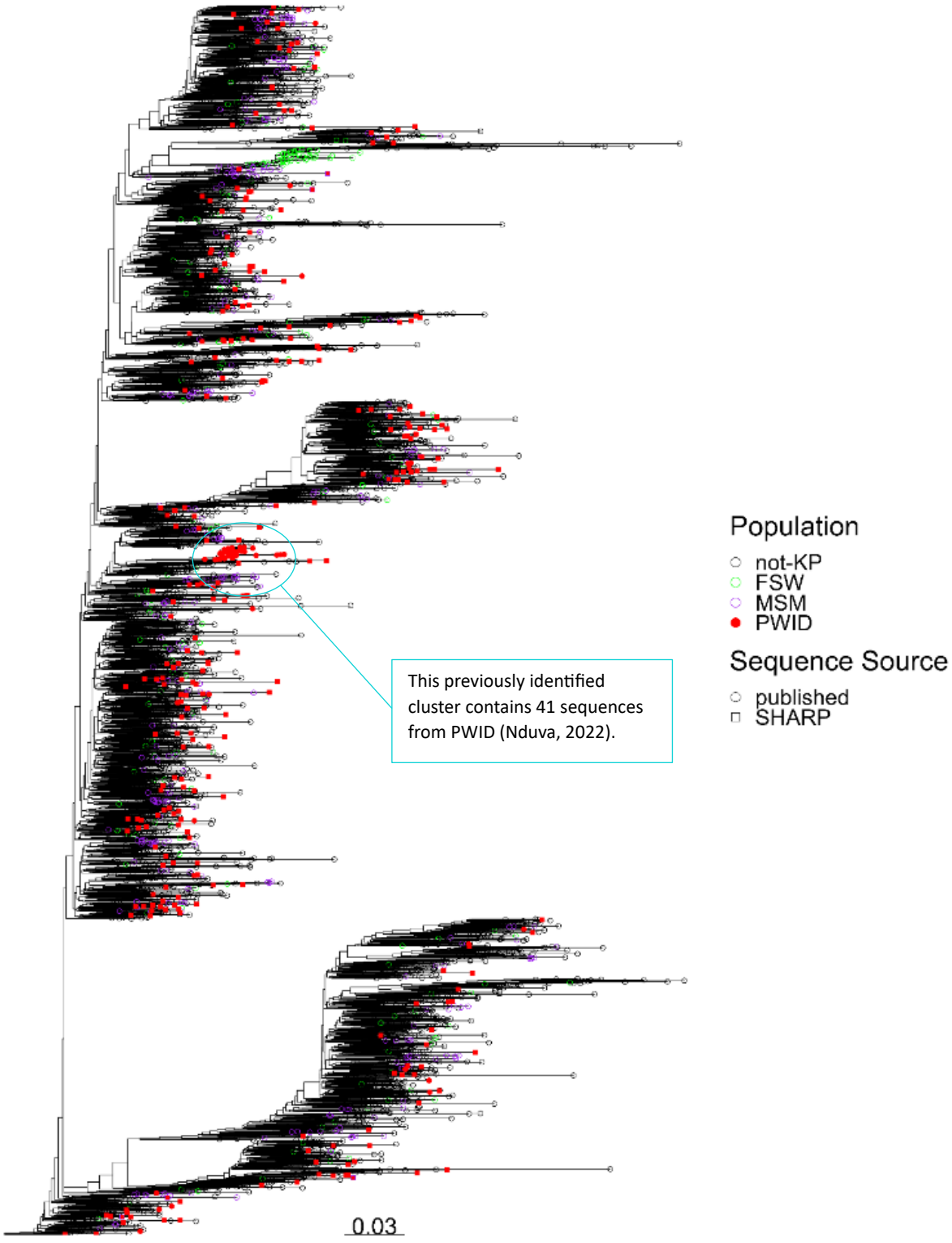
Table 1.1. Participant demographics and distribution of new (from SHARP study) and previously published Kenyan HIV-1 *pol* sequences by population. Abbreviations: FSW: female sex workers, MSM: men who have sex with men, PWID: people who injects drugs, not-KP: not key population. See Supplementary Table S1.2 for demographic distributions by region.

	FSW (N=207)	MSM (N=376)	PWID (N=341)	not-KP (N=2966)	Total (N=3890)
Source					
published	206 (99.5%)	368 (97.9%)	58 (17.0%)*	2955 (99.6%)	3587 (92.2%)
SHARP	**1 (0.5%)**	**8 (2.1%)****	**283 (83.0%)**	**11 (0.4%)**	**303 (7.8%)**
Sampling year					
2001-2015	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					
N-Miss	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	42 (20.3%)	16 (4.3%)	22 (6.7%)	328 (11.1%)	408 (10.6%)
Sex					
N-Miss	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%)

*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.

**7/8 MSM were also PWID. Sequences from these MSM individuals were counted as MSM sequences in discrete trait analysis

A)


This previously identified cluster contains 41 sequences from PWID (Nduva, 2022).
Population
not-KP
FSW
MSM
PWID
Sequence Source
published
SHARP
0.03

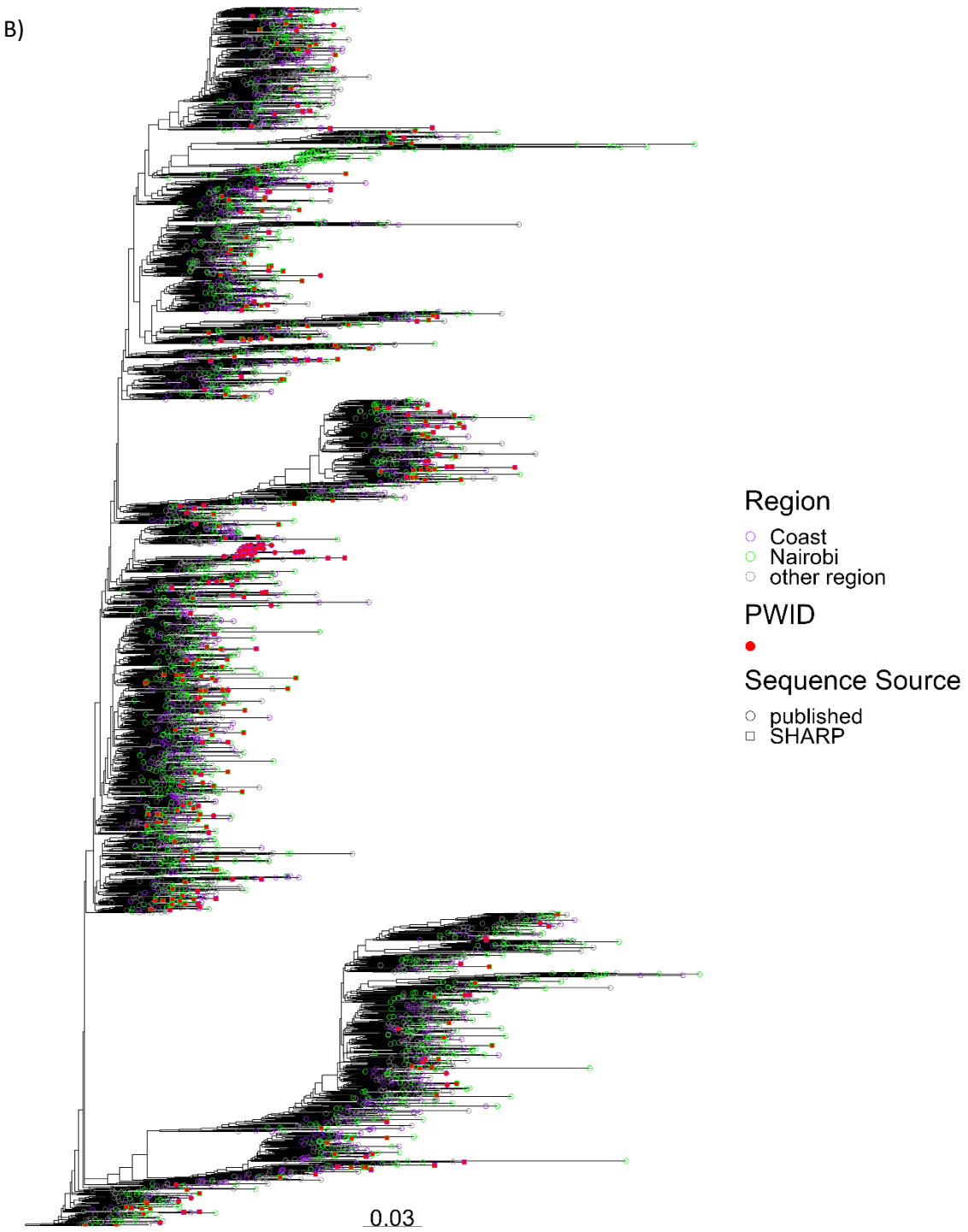


Figure 1.1. Phylogenies of HIV-1 subtype A1 sequences, with sequences from PWID highlighted. A) Sequences are colored by population. B) Sequences are colored by region. Abbreviations: FSW: female sex workers, MSM: men who have sex with men, PWID: people who injects drugs, not-KP: not key population.

Table 1.2. HIV-1 sequences from other populations in clusters with HIV-1 sequences from PWID. Clusters containing at least 1 sequence from the coast or Nairobi are included. Percents are based on the total number of clusters that contain at least 1 sequence from the given population (see Supplementary Table S1.3 for comprehensive cluster counts). Custers are based on 0.045 or 0.015 maximum genetic distance threshold on maximum likelihood trees and combined across subtypes A1, C, and D. Abbreviations: FSW: female sex workers, MSM: men who have sex with men, PWID: people who injects drugs, not-KP: not key population.

Clusters containing sequences from PWID			
Patristic Distance < 4.5 (N = 120)		**Patristic Distance < 1.5 (N = 15)**	
In cluster with	**Count***	**In cluster with**	**Count***
not-KP	95 (16.9%)	not-KP	3 (1.8%)
FSW	10 (12.2%)	FSW	0
MSM	16 (13.9%)	MSM	2 (4.5%)
PWID (ie. clusters with ≥2 PWID sequences)	25 (20.8%)	PWID (ie. clusters with ≥2 PWID sequences)	12 (80.0%)

*Clusters containing >2 population groups are represented more than once

Table 1.3. Transitions between population groups using maximum likelihood and Bayesian tree-building and ancestral state reconstruction methods. 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 Bayes factor (BF). Supplementary Table S1.8 presents a summary of transition counts limited to terminal branches.

	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		
MAXIMUM LIKELIHOOD						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%)**	**<0.01**	
C	14 (50%)	7 (25%)	7 (25%)	2 (0-4) (2.8%)	2 (1-4) (4.1%)	0.01	
D	18 (50%)	9 (25%)	9 (25%)	1 (0-2) (0.9%)	2 (1-3) (2.6%)	<0.01	
		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%)**	**0.53**	
C	14 (50%)	7 (25%)	7 (25%)	1 (1-4) (2.7%)	3 (2-4) (4.8%)	<0.01	
D	18 (50%)	9 (25%)	9 (25%)	1 (0-5) (2.0%)	4 (1-6) (5.2%)	<0.01	
		not-KP	PWID	not-KP -> PWID	PWID -> not-KP	not-KP -> 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%)**	**<0.01**	
C (53)	14 (50%)	7 (25%)	7 (25%)	4 (0-7) (7.3%)	4 (2-6) (7.1%)	0.08	
D (69)	18 (50%)	9 (25%)	9 (25%)	3 (0-6) (4.7%)	4 (2-7) (6.1%)	0.01	
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
		not-KP	PWID	not-KP -> PWID	PWID -> not-KP	not-KP-> PWID	PWID -> not-KP
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 to state 1)/(transitions from state 1 to state 0. The null distribution is generated from randomly resampling traits on the tree tips 200 times.

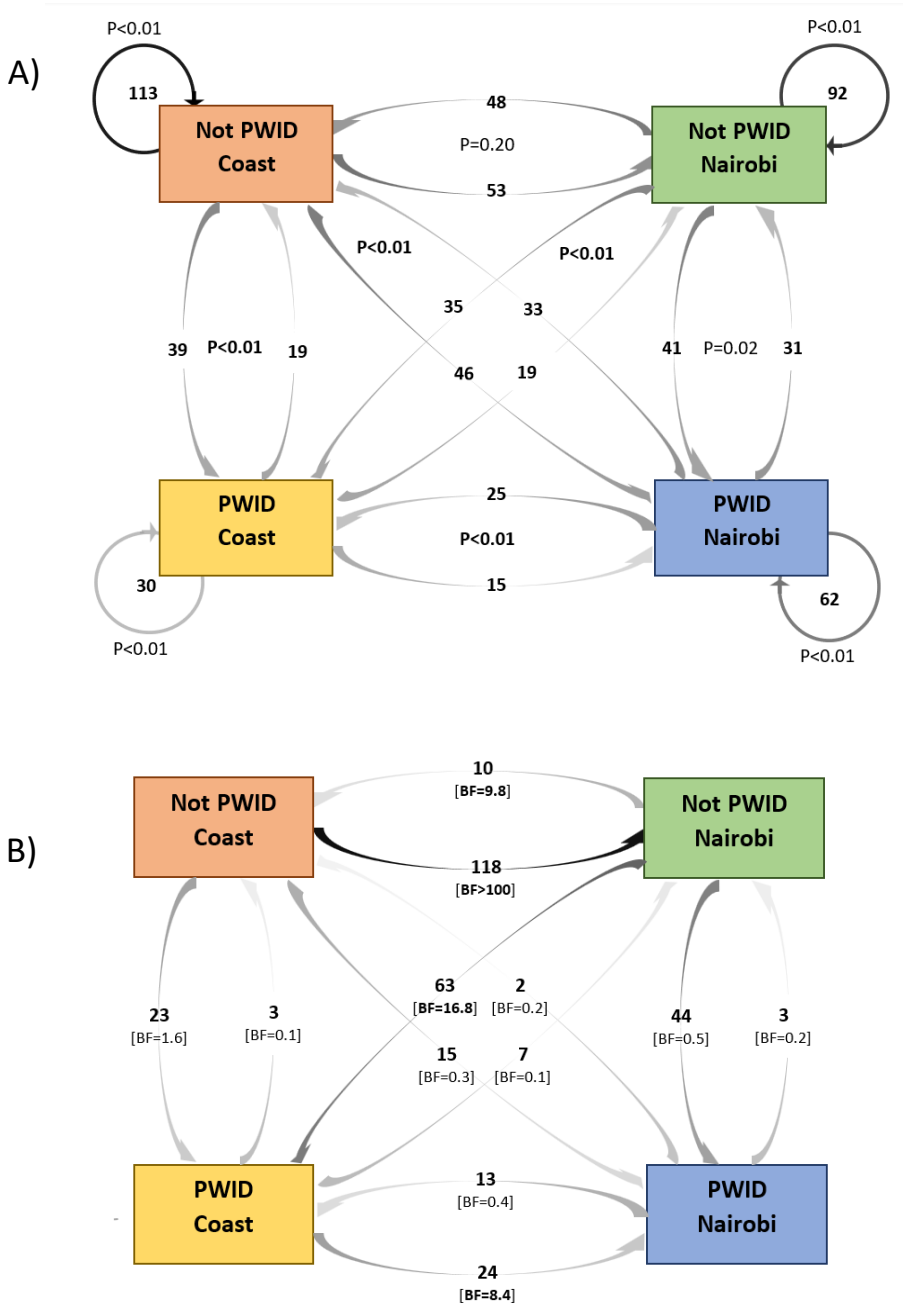


Figure 1.2. HIV-1 A1 transitions / Markov jumps between PWID and not-PWID populations from the coast and Nairobi. For the not-PWID population, not-KP, FSW, and MSM sequences are sampled proportionate to the estimated number of HIV cases in each population in the given region, with the majority of sequences coming from the not-KP population. Results are reported under maximum likelihood (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, Bayes factor (BF) >3 indicates strong support for Markov jumps in the given direction. Greyscale reflects the estimated amount of transmission in each direction. Data is presented in table format in Supplementary Table S1.7.

CHAPTER 2: Using HIV and Hepatitis C Molecular Epidemiology to Investigate Assisted-Partner Services Recruitment Among People Who Inject Drugs in Kenya

ABSTRACT

Background

Sexual and/or injecting partners of people who inject drugs (PWID) may have elevated risk of HIV infection either from sharing a transmission network or an epidemiological environment. Because viruses from individuals within a transmission network usually have greater similarity, we compared HIV and hepatitis C (HCV) sequences from PWID and their partners to assess whether partner-based recruitment identifies sexual or injecting partners within transmission networks.

Methods

We used assisted partner services (APS) to recruit sexual and injecting partners of PWID living with HIV in Kenya and evaluated trends in the TN93 distances (an adjusted measure of sequence similarity) of the HIV-1 and HCV sequences from partner pairs.

Results

Of 135 unique pairs identified with APS, we found 2 sexual, 2 injecting, and 3 sexual and injecting partner pairs with similar HIV sequences (TN93 distance<0.045), and 4 unique partner pairs had HCV sequences with distances <0.015. Sexual, but not injecting, partner pairs had HIV sequences with significantly smaller molecular distances than non-partners; injecting partner pairs did have significantly smaller HCV-4a molecular distances than non-partners.

Conclusion

APS recruitment partly reflects the HIV transmission network among sexual, but not injecting, partners of PWID. The relationship between the injecting partner recruitment and molecular networks is stronger for HCV than HIV and may reflect recent parenteral HCV transmission. Our results show the importance of continued focus on reducing sexual transmission among PWID and on education and services to address HCV transmission through needle and/or equipment-sharing.

INTRODUCTION

The prevalence of HIV among people who inject drugs (PWID) in Kenya is estimated at 15-18%, similar to global estimates among PWID, and ~4-times higher than among the general Kenyan population.[8,10,19] Estimates of hepatitis C (HCV) prevalence in this population are even higher, at 13-61%, compared to ~0.9% in the general population).[19,69,28,70] HIV testing and antiretroviral therapy (ART) have become more accessible for PWID in Kenya, and highly effective direct-acting antivirals for HCV were introduced in 2021.[10] Coupled with methadone treatment and community-based interventions, these advancements could drastically reduce disease prevalences; however, this requires reaching individuals who face a number of barriers to accessing testing and care.[17,18]

The World Health Organization recognizes assisted partner services (APS), in which people living with HIV provide contact information for their sexual and/or injecting partners, as a safe and effective tool to identify individuals who may not be aware of their HIV status or optimally engaged in care, including among hard-to-reach populations like PWID.[31] APS has previously been linked to increased HIV testing, diagnosis, and enrollment in care among partners of people living with HIV in Kenya.[3–6]

There are several motivations for identifying sexual and injecting partners of people living with HIV. Partners may have been directly exposed to HIV, or other infectious diseases, through sexual contact or needle- or equipment-sharing with the index who named them (or, inversely, have transmitted to that index). They may share a transmission network with the index who named them (such as having sexual or injecting partners in common). Alternatively, partners may share demographic or socio-behavioral factors associated with higher likelihood of living with HIV or HCV, even if they were not exposed through a close network.

Viruses that share a more recent ancestral infection usually have greater sequence similarity. By comparing a network based on pathogen sequence similarity to the APS recruitment network, we can gage the relative closeness of partner pairs within an HIV transmission network and investigate transmission dynamics in this

population. Specifically, the extent to which APS recruitment networks overlap with the underlying transmission network can suggest drivers of transmission, clarify when and how APS will be most effective, and asses the utility of molecular epidemiology approaches to supplement APS-focused activities. This study expands on existing literature comparing social networks with pathogen molecular networks by assessing differences between sexual vs injecting partners in an under-studied population (PWID) and setting (sub-Saharan Africa).

METHODS

Study population and enrollment

In this prospective cohort study, we recruited PWID living with HIV from 2018 to 2021 and used APS to identify, test, and link-to-care their sexual and injecting partners. Index participants were recruited from needle-syringe programs and methadone clinics in Nairobi (central Kenya) and Kilifi and Mombasa counties (coastal region). Index participants were eligible if ≥18 years old, injected at least once in the past year, tested positive for HIV, not at high risk of intimate partner violence, and provided locator information for at least one sexual (vaginal, anal, or oral intercourse) or injecting (regardless of needle sharing) partner from the last 3 years.

Partners (≥18 years old) were contacted via peer educators embedded in the community under the guidance of study health advisors. The identity of naming indexes was not disclosed to partners, although some indexes voluntarily notified injecting partners who were difficult to contact. Individuals could enroll multiple times as partners of different indexes, and partners fitting index eligibility criteria were also invited to enroll as indexes.

Socio-demographic data, HIV and HCV history, and sexual and drug use history were obtained for all participants via questionnaire. Rapid HIV testing using fingerstick samples was conducted following the Kenya national algorithm,[44] and HCV antibody testing was performed using the Abbott SD Bioline rapid one-step HCV testing kit (Abbott Pharmaceuticals).[71] Detailed study procedures are reported in the study protocol.[4]

Laboratory procedures and sequencing

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

For HIV sequencing using Sanger, PCR amplification was performed on the HIV-1 *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 HIV sequencing using NGS, short overlapping amplicons spanning the genome were generated using a tiling PCR approach,[45] and consensus sequences were later trimmed to the *pol* region. For HCV NGS, the NS5B region was targeted. All sequencing libraries were prepared using the Nextera DNA Flex Library Prep kit with Nextera CD indexes (Illumina, San Diego) and quantified using the Qubit dsDNA High Sensitivity assay kit on a Qubit fluorometer (Life Technologies, Carslbad, CA). Sequencing was performed on an Illumina MiSeq platform (Illumina).

Distances

We used Geneious Prime V11.0.18[48] for HIV and HCV sequence quality check and alignment via the MAFFT algorithm (default parameters).[49,50] For HIV, we trimmed sequences to a 966 bp region within *pol*, with near-complete sequence coverage, and excluded all codons associated with resistance to nucleoside and non-nucleoside reverse transcriptase inhibitors (NRTI's and NNRTI's).[47] For HCV, we separately aligned HCV-1a and HCV-4a subtypes and trimmed to the ~1,690 bp NS5B region. Average sequence coverage was 63% for the 4a alignment and 80% for 1a. Pairwise distances were calculated using the pairwise Tamura-Nei 93 model (TN93), which is a closed-form measure of sequence similarity that allows for 3 substitution rates (2 different transition rates and a transversion rate) and unequal base frequencies.[72,73] Due to incomplete coverage for some HCV sequences, we excluded all pairwise comparisons (partners and controls) between sequences with <300 overlapping bases. For topographical inference and secondary molecular distance-based analyses, we generated maximum likelihood trees (GTR+R+R9; IQ-Tree V1.6.12;).[53–55] For HIV, we also incorporated 952 previously published regional reference *pol* sequences from 2010-2018.

Data Analysis

For all participants living with HIV and for the subset of participants included within partner pairs in this analysis, we summarized baseline demographic, clinical, and socio-behavioral characteristics with hypothesized relevancy to transmission trends. Participants were included in the primary partner-pair analysis for HIV or HCV if a sequence was available for them and at least one of their sexual or injecting partners. For illustration, say participant A was named as a partner of indexes B and C then also enrolled as an index, naming B and D as partners. If A, B, and D had available HIV sequences, participant A's sequence would be represented in three total partner pairs (A_I-B_P, B_I-A_P, and A_I-D_P) and two "unique partner pairs" (A-B and A-D).

First, we assessed the number of unique partner pairs with HIV sequences in shared clusters, based on complete linkage and a TN93 distance threshold of 0.045 (commonly used to define transmission networks)

or 0.015 (a threshold consistent with recent outbreaks or possible direct transmission).[38,72,74] To assess the statistical strength of index-partner sequence similarity, we generated a control set for each index – all possible pairwise comparisons between the index sequence and participant sequences from the same region, excluding other partners of that index. We described the number and percent of index-partner pairs below various distance thresholds for 8 groups: all named partner pairs, only sexual partner pairs, only injecting partner pairs, and the subset of injecting pairs who reported sharing needles or injecting equipment with each other, as well as for the control set for each group. We also calculated the percent of control pairs with higher distances (greater dissimilarity) than true pairs and generated a *p* value from the distribution of this metric using a 2-tailed non-parametric test with 2,000 resamples of the named-partner in the pair. We performed a similar analysis for partner pairs sharing an HCV sequence but did not stratify by partner type due to low sample size. Because the HCV phylogeny was so substantially differentiated by subtype, we stratified by subtype in HCV analyses.

Finally, to understand how various index-partner characteristics relate to the molecular and possible transmission network, we tested the association between index-partner-pair characteristics and binary measures of HIV molecular distance using logistic regression, controlling for region, to get an adjusted odds ratio (aOR).

Ethics Approval and Participant Consent

This study followed ethical guidelines of the University of Washington and Kenyatta National Hospital. 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 to participate and have their data published.

RESULTS

Participant Characteristics – HIV analysis

Sequences were available from 28.0% of participants living with HIV. Of 5536 index-partner pairs identified in the data, 1180 consisted of an index and partner both living with HIV, and we were able to obtain HIV sequences from both participants in 150 (12.6%) of such pairs (135 unique pairs) (Figure 2.1, Supplementary Figure S2.1). As some individuals are represented in multiple pairs, this data represents 151 unique participants. The majority were only injecting partners (N=116), of which 13 reported having shared needles with each other; all 18 sexual partner pairs were male female.

Most participants in this analysis had a previous HIV diagnosis (76.8%), with a 3.5-year median time-since-diagnosis. Because HIV sequencing success depends on viral load, viral non-suppression was common (82.5%), and ART use was lower (55.0%) than among the larger study population. Almost all had previously injected drugs (99.4%), with a median start time 4.5 years before enrollment; 5.3% shared needles and 11.3% shared equipment in the last month. Participants reported a median of 1 sexual partner in the last 3 months, but several participants had substantially higher numbers of sex partners; 85% of female participants and 38% of male participants had been paid for sex (Table 2.1).

Comparing recruitment and HIV molecular networks

Analysis of HIV sequence distances by partner type showed the recruitment network for sexual partners is related to the molecular HIV network. Among 303 total HIV sequences, we identified 39 clusters using a maximum TN93 distance of 0.045 (mean cluster size: 2.23, range: 2-4) and 15 within 0.015 (all size 2). Of the 135 unique named partner pairs in our data, 7 (5.2%; 2 sexual, 2 injecting, and 3 both sexual and injecting) fell within clusters at threshold 0.045, representing 16.7%, 1.8%, and 21.4% percent of sexual, injecting, and both sexual and injecting partner pairs, respectively. Four unique pairs (1 sexual and 3 both sexual and injecting) fell within clusters at threshold 0.015. Although the majority of named partner pairs

did not fall within clusters, they were over-represented (compared to controls) among the closest distances across all partner types (Table 2.2). Except for one injecting pair, pairs in clusters (threshold: 0.045) were also nearest neighbors on a phylogenetic tree that included 952 regional reference sequences (Supplementary Figure S2.1).

Low clustering is not surprising in this dataset because of low sampling density and because we did not restrict analyses to recent diagnoses. For this reason, and to avoid confusion from non-standardized cluster definitions, we also evaluated trends across a distribution of sequence distances, with a control set for comparison. Only distances for sexual partner pairs were significantly closer (p=0.016), on average, than control pairs (Table 2.2, Supplementary Figure S2.2; analysis reproduced using molecular distances in Supplementary Table S2.1).

Associations between index-partner pair characteristics and HIV TN93 distance

Although sexual partners had lower molecular distances than injecting partners, this difference was not significant at TN93 cut-off of 0.10 (OR: 2.63; 95% CI: 0.91-8.77) (Table 2.3, Supplementary Figure 2.3A). Low distances were significantly more common in partner pairs where neither partner was virally suppressed (OR: 2.46; 95% CI: 1.21-5.13) and those with more recent diagnoses, a proxy for time-since infection ($OR_{\text{avg. years since diagnosis}}$: 0.90; 95% CI: 0.81-1.00) (Table 3, Supplementary Figure S2.3B&C). As a secondary analysis to explore the possibility of confounding by time-since-infection, we determined that partner pair type is not associated with average time-since diagnosis (avg. years) (OR: 1.02, 95% CI: 0.87-1.17) or difference in diagnosis times (years) (OR: 0.98, 95% CI: 0.84-1.12).

Comparing recruitment network and HCV molecular networks

Sequences were available for 35.5% of participants living with HCV (64 HCV-1a and 84 HCV-4a sequences), and there were 42 partner pairs where an HCV sequence was available from both partners. Due to high divergence by HCV subtype, we stratified our analysis by subtype, which left 22 (20 unique)

subtype-matched partner pairs (18 4a and 4 1a), representing 28 unique individuals (Supplementary Figure S2.4&S2.5). All 22 partner pairs were identified from the coastal region, 20 were only injecting pairs and 2 (1 unique) were a sexual and injecting pair. None of the 28 participants reported HCV diagnoses before enrollment.

Despite a small sample size, there was evidence of excess similarity among the HCV subtype 4a sequences for the partner pairs (mean distance=0.021, mean distance for injecting-only pairs=0.023), compared to controls (mean distance: 0.029; P=0.008). This trend did not reproduce among the small (N=4) HCV 1a subtype dataset (Table 2.4; analysis reproduced using molecular distances in Supplementary Table S2.3). For partner pairs where at least one partner reported sharing injecting needles in the last month (N=6), the mean HCV sequence distance (0.024) was slightly smaller than for partners where neither shared injecting needles (0.027) (N=16). The only pair where both reported sharing needles in the last month had the most similar HCV sequences (TN93=0.004), but the only pair that reported sharing needles with each other had the least similar sequences (TN93=0.07). Only 2 pairs were nearest neighbors on the HCV phylogeny (Supplementary Figure S2.5).

DISCUSSION

This study investigated the relationship between index-partner pairs and the HIV and HCV molecular network to test whether APS can find individuals in shared networks relating to sexual and/or injecting modes of transmission. Among 135 unique partner pairs in Kenya recruited using APS, we found 7 partner pairs where HIV molecular evidence (based on TN93 sequence distance and phylogenetic topological adjacency) supports a shared semi-recent transmission network. Overall, we found excess similarity among HIV sequences from sexual but not injecting partner pairs. However, there was slight excess similarity among some HCV subtype 4a sequences from injecting partners.

Sequence similarity among HIV sequences from sexual partners is consistent with prior research and shows that molecular HIV analyses can reveal information about sexual transmission networks even when most infections occurred multiple years before enrollment.[75–77] More than half of participants in this analysis reported being paid for sex, which has been associated with HIV prevalence in other studies.[78] Although transactional sex was not associated with HIV sequence distance in this analysis, we believe it's important that HIV prevention and harm-reduction services for PWID also consider those who may exchange sex for drug but who are not reached by traditional services available to sex workers.

Our results suggest that few participants who injected together fell within shared HIV transmission networks. This finding corroborates a prior study in Canada that found several HIV clusters indicative of recent transmissions among PWID, but showed recruitment networks had little to no link with these molecular networks.[79] Two prior studies that did find some molecular link among HIV sequences from PWID and their social contacts, recruited only indexes with recent infections, whereas most individuals in our study were previously diagnosed.[72,80]

Lack of sequence similarity among HIV sequences from injecting partners could suggest reduced recent parenteral HIV transmission in Nairobi and coastal Kenya; however, our findings cannot decipher the extent to which parenteral transmission among PWID was a historic driver of the high HIV prevalence in this population. Prior research in this dataset showed an age-independent association between a longer history of injecting drugs and both HIV and HCV infection.[26] It is possible that a participant's injecting partners at the time they acquired HIV do belong to a shared transmission network but differ from the partners they named at enrollment, preventing us identifying this link. Our analysis may also be under-powered to detect excess sequence similarity associated with older HIV transmission events.[26,74,81,82] Lower sequence similarity among sequences from partners who shared needles or equipment was surprising and not easily explained,

but supports that APS in this setting did not readily identify individuals who contracted HIV parenterally through a partner.

We found evidence that injecting partner pairs from coastal Kenya have significantly more similar HCV sequences than control pairs in the HCV-4a subtype strata, although the result is in a small sample, and about 41% of partner pairs had HCV sequences with different subtypes, indicating no close molecular link. This result may indicate that APS identified some pairs connected through parenteral HCV transmission networks (although the possibility of sexual transmission through unsampled partners cannot be excluded). HCV is more readily transmitted through needle-sharing than HIV, so the stronger connection between the HCV network and injecting network is not surprising and emphasizes the importance of needle-syringe programs for reducing incidence of diseases beyond HIV.[28,70] The distribution of distances we observed for HCV are quite small; however, subtype differences, low sample size, and lack of subtype-matched references limits comparability to distance thresholds used in other analyses to define outbreaks or close transmission networks.[83]

Prior research comparing HCV molecular networks with recruitment networks is minimal and draws mixed conclusions. Our result agree with a study in Australia that found PWID who injected together were more likely to have HCV sequences within the same cluster;[84] however, two other prior studies found minimal overlap between HCV molecular networks and social recruitment networks among PWID.[79,85]

Our findings should be placed in the broader context of what we know about needle-sharing behaviors in this population. Prior research suggests that needle sharing rates dropped among the PWID after the 2013 introduction of needle-syringe programs.[9,14,21,23,28,29] Participants in this study reported lower rates of needle-sharing than in prior studies in similar populations, but 11% reported sharing other injecting equipment, which has been previously associated with HIV and HCV infection.[86,87] Data on the longer-

term history of needle-sharing in this population could clarify how needle-syringe programs may have impacted HIV and HCV parenteral transmission patterns over time.

In addition to inference about modes of transmission, this analysis has implications for how we understand the function of APS in this population. Previous research shows APS is an effective tool to identify individuals living with HIV and connect them to care,[3,5] and many partners of PWID identified through APS also live with untreated HCV.[67] These results suggest that APS can identify individuals who belong to shared sexual transmission networks with PWID living with HIV (with direct transmission being plausible in a few cases). But APS for injecting partners may be better placed to identify individuals with shared environmental risk factors, such as a history of injection drug use, than as a method of contact tracing for recent transmission. It is possible that APS functions more like contact tracing (finding individuals exposed through a partner) when used among newly diagnosed individuals. Although APS is less commonly used with HCV, we found that APS for HIV was also effective at finding some partners within shared HCV transmission networks. Our results have implications for communicating information about risk and exposure to partners identified through APS and tailoring future recruitment designs to efficiently reach individuals with the highest need.

Our study has limitations. Sequencing success is biased towards higher viral loads, which reduces sample size and limits generalizability; historical transmission patterns may be quite different for PWID and their partners who were virally suppressed at baseline. We were unable to consider sequences relevant to this network from individuals outside the target population. We were unable to distinguish between differences in sequence distances driven by distance within a transmission network vs driven by time-since infection, although we confirmed that average time-since diagnosis or difference in partners' diagnosis times is unlikely to explain the molecular distance differences that we observed between sexual vs injecting partners. Finally, inference about modes of transmission is also sensitive to other sources of confounding linked to sampling or to the likelihood of viral suppression.

CONCLUSION

Many APS-identified sexual pairs belong to a broad HIV transmission network and some partners belong to a more recent transmission network. HIV transmission among non-sexual partners who inject together is either sufficiently rare or occurred too long ago to detect in the molecular HIV data; however, evidence supports some parenteral HCV transmission in this population. We suggest that future research works to quantify the contribution of needle sharing to current HIV and HCV transmission and that future recruitment and intervention approaches are cognizant of the role that sexual transmission continues to play in this population.

TABLES AND FIGURES

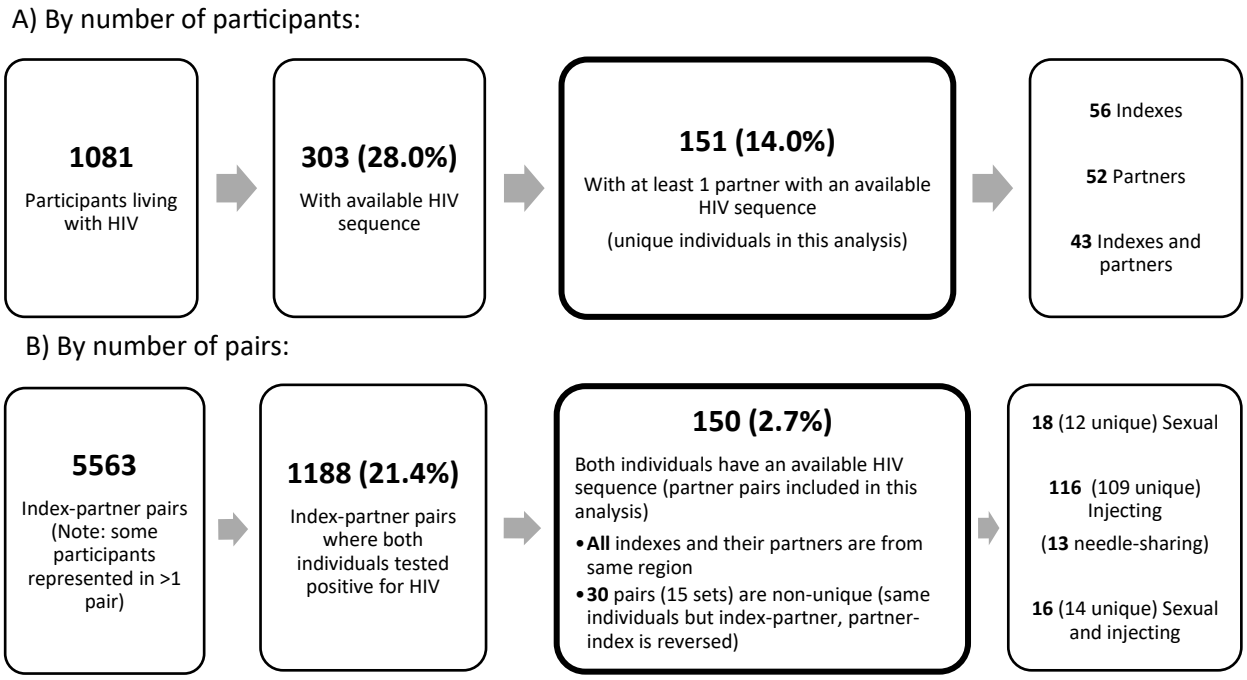


Figure 2.1 Cascades of participant and partner pair counts for HIV analysis A) Counts of participants, culminating in the number of unique participants included in the HIV analysis and stratified by enrollment type. B) Counts of named partner pairs, culminating in the number of pairs included in the HIV analysis and stratified by pair type. Some individuals are represented in multiple partner pairs, but counts of unique partner pairs are also provided.

Table 1.1. Characteristics for all participants living with HIV and for the 151 participants included in this analysis: those with an available HIV sequence and at least one partner with an available HIV sequence**.**

	Included in partner-pair phylogenetic analysis (N=151)	All participants living with HIV (N=1081)
Enrolment[a]		
Only index	56 (37.1%)	487 (45.1%)
Index and partner	43 (28.5%)	502 (46.4%)
Only partner	52 (34.4%)	92 (8.5%)
Times included (this analysis) or named (total dataset) as a partner (excludes indexes who were never enrolled as partners)		
Mean (SD)	1.6 (1.3)	2.0 (1.7)
Median (Q1, Q3)	1 (1,2)	2 (1,2)
Partner type (partners only)		
Injecting	71 (74.7%)	-
Sexual	18 (18.9%)	-
Sexual & Injecting	6 (6.3%)	-
Female	85 (56.3%)	546 (50.5%)
Previously tested positive for HIV	116 (76.8%)	934 (86.4%)
Years since first positive HIV test[b]		
Mean (SD)	4.7 (4.5)	6.3 (5.5)
Median (Q1, Q3)	3.5 (1,8)	5 (2,10)
On ART at enrolment	82 (54.3%)	852 (78.8%)
HIV viral load		
<1,000 (virally suppressed)	25 (16.9%)	583 (67.9%)
1,000-10,000	42 (28.4%)	108 (12.6%)
>10,000	80 (54.1%)	166 (19.3%)
Ever injected drugs	148 (99.3%)	1031 (96.7%)
Years injecting		
Mean (SD)	4.5 (5.1)	5.5 (5.2)
Median (Q1, Q3)	3 (1 – 6)	3 (2-8)
Shared needles in the last month (among PWID)	8 (5.4%)	69 (6.5%)
Shared injecting equipment (ex. cookers, cottons, rinses) in the last month (among PWID)	17 (11.3%)	109 (10.1%)
Number of sex partners in the previous 3 months		
Mean (SD)	5.8 (17.4%)	5.8 (23.9%)
Median (Q1, Q3)	1 (0,2)	1 (0,2)
Ever received money or goods for sex (Females)	72 (84.7%)	431 (78.9%)
Ever received money or goods for sex (Males)	25 (37.9%)	135 (25.3%)

[a]Whether the participant was enrolled only as an index, included as both an index and a partner (in different partner pairs), or only a partner.
[b]Diagnosis times prior to enrollment are self-reported. Diagnosis at enrollment = time 0.

Table 2.2. Distribution of HIV-1 molecular distances across index-partner pairs stratified by partner-type. The *p* value comes from the null distribution of the test statistic (Avg(% controls further than named pairs)) under 2,000 resamplings of the partner for each index.

		Sexual only		**Injecting only**		**Injecting partners who shared needles or equipment**		**All (Sexual, Injecting, and Sexual & Injecting)**	
		named pairs	controls	named pairs	controls	named pairs	controls	named pairs	controls
total	distance threshold	**18**	**2636**	**116**	**18577**	**13**	**2130**	**150**	**23630**
Pairs < distance threshold N (%)	0.015	3 (16.7%)	1 (<0.01%)	0 (0%)	8 (<0.01%)	0 (0%)	0 (0%)	6 (4%)	6 (<0.01%)
	0.045	4 (22.2%)	13 (0.5%)	2 (1.7%)	95 (0.5%)	0 (0%)	6 (0.3%)	9 (6%)	117 (0.5%)
	0.1	13 (72.2%)	1441 (54.7%)	57 (49.1%)	10427 (56.1%)	4 (30.8%)	679 (31.9%)	77 (51.3%)	12944 (54.8%)
	0.15	17 (94.4%)	2619 (99.4%)	114 (98.3%)	18457 (99.4%)	13 (100%)	2124 (99.7%)	147 (98%)	23486 (99.4%)
mean distance		0.075	0.098	0.099	0.096	0.111	0.107	0.095	0.097
Avg (% control pairs further than named pairs)[a]		0.67	-	0.461	-	0.46	-	0.492	-
***p* value**		**0.016**	-	0.133	-	0.632	-	0.747	-

[a]Percent of index-control pairs with sequence molecular distance > the named index-partner pair, averaged across all pairs. Percents >50% indicate that indexes' sequences are more similar, on average, to those of their identified partner than to controls (sequences from the same region).

Table 2.3. Association between partner-pair characteristics and HIV-1 TN93 distance. A cut-off of <0.10 was chosen to define more phylogenetically similar molecular distances. Supplementary Table 2 shows regression results using various molecular distance thresholds.

Pair characteristic	**values**	**mean** (continuous) **or proportion** (categorical)	**odds** (molecular distance <0.10 vs. >0.10)	**aOR (95% CI)** (controlling for region)
Partner type (N=134) (excludes those identified as both sexual and injecting partners)	Ref: Injecting only	116 (86.6%)	57/59 = 0.97	Ref
	Sexual only	18 (13.4%)	13/5 = 2.6	2.63 (0.91-8.77)
Years since diagnosis (N=150) averaged between index and partner (diagnoses prior to enrollment are self-reported)	-	3.66	-	**0.90 (0.81-1.00)**
Number of years separating index-partner diagnoses[a]	-	3.98	-	0.96 (0.88-1.05)
Positive HIV diagnosis prior to enrollment (N=150)	Ref: both previously tested positive	89 (59.3%)	45/44 = 1.02	Ref
	one partner previously tested positive	54 (36%)	29/25 = 1.16	1.14 (0.58-2.25)
	neither partner previously tested positive	7 (4.7%)	3/4 = 0.75	0.72 (0.14-3.46)
ART use at enrollment (N=150)	Ref: Both partners on ART at enrollment	37 (24.7%)	19/18 = 1.06	Ref
	One partner on ART at enrollment	65 (43.3%)	38/27 = 1.41	1.35 (0.59-3.12)
	Neither partner on ART at enrollment	48 (32.0%)	20/28 = 0.71	0.69 (0.28-1.73)
Shared needles (with anyone) last month (N=148)	Ref: Neither partner shared needles	133 (89.9%)	66/67 = 0.99	Ref
	One or both partners shared needles	15 (10.1%)	10/5 = 2.00	2.12 (0.71-7.15)
Shared needle or injecting equipment with named partner[b] (injecting partners only; N = 132)	Ref: Never shared with partner	119 (90.2%)	60/59 = 1.02	Ref
	Have shared with partner	13 (9.8%)	4/9 = 0.44	0.44 (0.11-1.44)
Ever received money or goods for sex (N=150)	Ref: Neither partner received money or goods for sex	18 (12%)	10/8 = 1.25	Ref
	One partner received money or goods for sex	66 (44%)	39/27 = 1.44	1.17 (0.40-3.36)
	Both partners received money or goods for sex	66 (44%)	28/38 = 0.74	0.59 (0.20-1.70)
Viral suppression (N=145) (<1,000 copies/ml at enrollment)	Ref: one or both partners virally suppressed	47 (32.4%)	17/30 = 0.57	Ref
	Neither partner virally suppressed	98 (67.6%)	57/41 = 1.39	**2.46 (1.21-5.13**)
HIV subtype (N=91) low confidence subtypes and recombinants are excluded	same HIV subtype	60 (65.9%)	58/1 = 58	-
	different HIV subtypes	31 (34.12%)	0/31 = 0	-

[a]Diagnosis times prior to enrollment are self-reported

[b]Based on reporting by index

Table 2.4. Distribution of HCV molecular distances across index-partner pairs, stratified by subtype. The *p* value comes from the null distribution of the test statistic (Avg(% controls further than named pairs)) under 2,000 resamplings of the partner for each index.

		4a		1a	
		named pairs	controls	named pairs	controls
total	**distance threshold**	**18**	**1087**	**4**	**136**
Pairs < distance threshold N (%)	0.005	1 (5.6%)	1 (0.1%)	0 (0%)	0 (0%)
	0.015	5 (27.8%)	61 (5.6%)	0 (0%)	1 (0.7%)
	0.025	10 (55.6%)	326 (30.0%)	0 (0%)	13 (9.6%)
	0.045	18 (100.0%)	1041 (95.8%)	3 (75.0%)	102 (75.0%)
mean distance		0.021	0.029	0.048	0.042
Avg (% control pairs further than named pairs)[a]		0.70	-	0.35	
***p* value**		**0.008**	-	0.34	

[a]Percent of index-control pairs with sequence molecular distance > the named index-partner pair, averaged across all pairs. Percents >50% indicate that indexes' sequences are more similar, on average, to those of their identified partner than to controls (sequences from the same region).

CHAPTER 3: The Ethics of Pathogen Molecular Epidemiology Research from a Community Data Perspective

Throughout human history, infectious diseases have evolved alongside us, substantially shaping and being shaped by our patterns of migration, behaviors, and societal and political dynamics. To prepare for the emergence of new pathogens and end existing epidemics, we must constantly update our understanding of pathogen transmission at the community and global levels.[88] Molecular/genomic epidemiological research combines pathogen gene sequence or phylogenetic data with epidemiological data from the population to address questions about diseases' origins, spread, and mechanisms of transmission. Findings from these studies can inform rapid public health responses to new outbreaks, identify emerging resistance, guide resource allocation, and prepare future responses.[40,89–91]

However, molecular epidemiological studies have raised unique ethical considerations, which can be roughly divided into concerns about risks to individuals and concerns about potential group harms. While principles of ethical research, particularly within the West, focus heavily on individual autonomy and the protection of individually sensitive data, this paper argues that pathogen data is largely a community resource and that more focus should be paid to its potential community impacts. People living with HIV and HIV advocacy groups have already spearheaded efforts for greater community collaboration, promoting research that benefits and does not stigmatize people living with HIV. However, molecular epidemiological research on transmission dynamics can impact groups beyond those living with the studied disease, and there is less guidance on how to balance the needs of these different types of impacted communities. There is also a need for guidance that can be generalized to other types of pathogens. Thinking about pathogen data as community data has the potential to promote research practices that strengthen community relationships and work within unique community dynamics.

This paper is split into three sections. The first explores existing critiques about the privacy risks of molecular epidemiological research to individual research participants and shows that many of these critiques could be best addressed at the community level. Section 2 explores the ways that molecular epidemiology research could impact different types of communities. We suggest that a thorough understanding of community interests includes understanding the relationship between communities living with and without the studied disease and understanding how the research impacts people within geographically defined or identity-based communities. The last section provides examples of existing concepts of community data and explores the value they could bring to pathogen molecular epidemiology research.

1. BEYOND PRIVACY: CONCEPTUALIZING PATHOGEN SEQUENCE DATA AS COMMUNITY DATA

Ethics principles, particularly within the West, are heavily concerned with the sensitivity of individual health data, and concerns about privacy are frequently mentioned in molecular epidemiology literature.[92–96] However, there is disagreement and confusion about the extent of privacy risks to research participants,[97–99] and calls to restructure approaches to pathogen sample collection[100,101] out of privacy concerns have been largely ignored. Some confusion may stem from the fact that concerns about group impacts of molecular epidemiology research or collective moral objections to research are sometimes framed as concerns about individual privacy. But molecular epidemiology analyses often reveal more sensitive information about a community than they do about specific individuals. As such, many of these concerns may be addressed, not by restricting sequence collection, but by ensuring research approaches have community acceptance.

Pathogen sequence data used in research may come from primary study data, public health surveillance data, or sequences from either source that are deidentified and shared to publicly accessible databases.

Privacy concerns with molecular epidemiology research tend to be heavily intertwined with critiques of infectious disease surveillance (primarily for HIV), and to a lesser extent, data sharing.[97,100–103] Most pathogen molecular surveillance is done in high resource settings.[89] For example, HIV sequencing is done clinically in the US to inform treatment regimens, and these sequences are often shared with departments of health. As sequencing capacity and the availability of genotype-informed treatments increases, however, surveillance for different diseases will likely become more common worldwide,[89] and more regions will have to think about how to engage with communities whose data is collected outside of traditional research settings. One advantage of conceptualizing pathogen sequence data as a community resource is that the source of the sequence data becomes less important than the way the sequences are used, meaning community consultation can provide a path-forward regardless of whether the sequence data was collected through research or other means.

The extent of privacy concerns among molecular epidemiology research stakeholders is unclear.[92,95] Some US-based advocates have raised privacy as a critical issue in HIV molecular epidemiology research.[97,104] On the other hand, two US-based studies found a more complicated picture. One study found that only 50% of people seen at an infectious disease clinic agreed to participate in phylogenetic research, but few cited privacy as the reason for refusal.[105] A different study found that stakeholders – people living with HIV, those deemed at risk for HIV, and HIV medical professionals – grew significantly more concerned about privacy in molecular epidemiology research only after it was brought up as part of a guided interview.[92] Studies have also shown persistent misconceptions, even among medical professionals, about the capabilities of molecular epidemiological methods. Misconceptions that over-estimate the power or intent of these studies include interpreting lines in a diagram meant to highlight similar viruses as proof of transmission between individuals or assuming molecular epidemiology could be used to find someone living with undiagnosed HIV.[92,106] Of course, disagreement and/or confusion about privacy does not mean this issue is unimportant or unresolvable; however, it does caution against

the utility of asking individual research participants to weigh in-depth critiques about privacy risks. Additionally, individuals agreeing to participate in research should not be taken as proof of the broader acceptability of that research.[92,107]

Molecular epidemiological methods, and their descriptions of transmission-pathways, -pairs, and -networks, understandably invoke notions of privacy. Poorly explained or taken-out-of-context studies may obscure the fact that the research goal is usually to uncover community-level trends. Consider the challenge of explaining a method like source attribution analysis, which models the probability of a transmission event between pairs of individuals, based largely on the relationship of their pathogen sequences on a phylogenetic tree.[108] Source attribution is one of several approaches to understand how certain behaviors relate to transmission risk. While the use of these pairwise transmission probabilities are a tool, and rarely the end goal of this method, it's easy to see how descriptions of source attribution could be perceived as trying to find "who transmitted a pathogen to whom?" Research, even if it does not create tangible risks to research participants, must respect commonly held values within a community. Through better community engagement, we can resolve common misunderstandings about study methods, goals, or capabilities while still being open to the possibility that some research questions or approaches may feel inherently sensitive or problematic to the wider community.

Many critiques about data privacy are actually critiques about perceived unethical uses of community data or about research questions seen as insensitive or invasive. For example, criminal repercussions are a frequently noted concern in HIV phylogenetic research. But the situations in which these concerns are raised are usually ones in which research subjects' data have fairly substantial protections, but the findings of the research could have negative implications for a wider communtiy.[97,100,101,107] McClelland cautions, for example, that Canada has used sequences from the publicly-accessible database, Los Alamos National Laboratory, for criminal investigations.[100] In another example, advocates criticized a study of HIV among transgender women in Los Angeles[109] for "position[ing] transgender women at the

centre of high-risk sexual networks; a particular concern in the USA, where there is heightened criminalization of sex work, migration, drug use, and HIV…."[97] Finally, a Washington state study chose not to publish a phylogenetic analysis about HIV transmission among men who have sex with men in response to community concerns, particularly about criminalization.[107] While acknowledging that fields evolving as rapidly as molecular epidemiology and phylogenetics must diligently reassess the protections in place for individuals' data, the above examples largely illustrate a conflation of individual and group risks. Each set of data in the above examples was deidentified (i.e., the researchers could not have chosen to reveal individuals' identities). In the US, laws and institutional policies also provide substantial protections against outside requests to disclose the identities of molecular epidemiology research participants or individuals whose data is collected through public health surveillance.[105,110] Nevertheless, policies are less clear in most other countries,[111] and it is impossible to definitively assert that molecular epidemiology research data or results could never be compelled by law enforcement.[110]

Much more likely, and as was the case with the use of the Los Alamos database, is that deidentified database sequences can be incorporated as reference sequences (basically background information about community transmission patterns) in investigations involving others. Community engagement can and should address concerns about the use of such data to benefit the prosecution of HIV transmission, but it is important not to conflate this group-level harm with the low risk that an individual's privacy might be breached. Most research on stakeholder perceptions of molecular epidemiology, as well as most advocacy work on the issue, is focused within the US.[92,97,101,105,107] There is substantial need to understand more diverse stakeholder perspectives, particularly in other regions, where different expectations of privacy, informed consent, trust in researchers, levels of scientific literacy, and varying HIV criminalization laws could contribute to different overall sentiments.

In conclusion, while we must certainly continue to ask questions about how molecular epidemiology research will protect individuals' privacy, restricting public health surveillance efforts or relying too

heavily on individual informed consent processes to address privacy concerns is often ineffective and usually fails to address the bigger picture about the group impacts of the research. Molecular epidemiology research is inherently about communities and community dynamics, as we see from the fact that critiques ostensibly about privacy often focus on larger-scale community harms. Therefore, it is worthwhile to consider what types of communities and community dynamics are implicated by this type of research.

2. COMMUNITY INTERESTS IN MOLECULAR EPIDEMIOLOGY RESEARCH

When we collect pathogen data, we collect data about a community, such as travel patterns and how the behaviors and diseases of one sub-population can impact other sub-populations. Prior ethical analyses of phylogenetic or molecular epidemiological research already recognize the potential for group harms and the need for community consultation.[99,100] But the goal often appears to be to find spokespersons or representatives who can advocate specifically for the individuals' whose sequences are used. By contrast, we have proposed starting from the assumption that pathogen data is a form of community data and that a diverse group of individuals within a community may have a vested interest in the research done with that data. By engaging with the different communities potentially impacted by the research, researchers can devise studies that are relevant to diverse community needs and sensitive to community concerns. Community engagement efforts must also account for the ways molecular epidemiological studies of transmission dynamics can impact groups beyond those enrolled in the study.

Determining who is impacted by molecular epidemiology research is not always straight-forward. As Mutenherwa points out, "communities are often not stable, static entities with clearly defined boundaries. Rather, they are fluid and may consist of sub-communities whose members may not share similar research needs and priorities."[99] We suggest that molecular epidemiology often impacts

communities connected by living with the same disease, communities based around shared aspects of identity, and communities within a shared geographic region. Researchers should consider these aspects of community when planning their research and eliciting feedback.

Those who live with the studied disease and those who do not

It is important to understand the relationship between people living with the studied disease and the larger population; in societies where the disease is highly stigmatized and people with the disease are treated as outsiders, we must be particularly cautious that our research does not perpetuate this stigma and isolation.[112] At the same time, we should not assume that people living with a disease have interests unconnected to those of the rest of the community. A more community-focused ethics considers that individuals living with a disease can benefit from being able to help a wider community that they feel a part of and/or from contributing to the overall health of that community.[113,114]

Community by identity and identities within a community

Molecular epidemiological research often seeks to understand how traits, behaviors, or experiences inform disease transmission risk. Sometimes traits associated with the risk of contracting a contagious disease are also related to important aspects of people's identities. For example, gay men share a greater burden from HIV and monkeypox;[7,115] older people, people with larger bodies, and those with certain chronic health conditions face greater risk from contracting COVID-19;[116] and pregnant women were recommended to take special precautions to avoid the Zika virus.[117] MacQueen et al. interviewed stakeholders involved in HIV participatory research to define the most salient aspects of community in that research setting and concluded communities are diverse sets of individuals linked by geographical location, social ties, shared perspectives, and/or engagement in joint action.[118] With the exception of geography, these aspects cannot be discerned from basic demographic information alone, again highlighting that local knowledge and broader community input are essential to understand community

dynamics. Such input can add critical context to how the data collected about study participants – for example, their age, sexual behaviors, or religious practices – may relate to more saliant aspects of identity in the study setting.

Results of some types of molecular epidemiology research could further stigmatize identity-based communities as sources of an outbreak or for having higher rates of transmission. Researchers should present results in a way that does not depict specific identity groups as *vectors* and spreaders of that disease.[119] Ways to reduce stigma in the presentation of study results include: cautious use of arrows to depict possible transmission events;[112] avoiding blaming language like "cause" or "super-spreader;"[120] and clearly highlighting uncertainty in findings. In some cases, molecular epidemiology investigations may reduce stigma by reassessing long-held assumptions that certain populations disproportionately contribute to an epidemic.[24,39] Prior to the study, community collaboration efforts should include discussion of how different possible findings might impact the community and ensure that there is community approval to report and act on all possible findings.

Understanding the identities represented among people impacted by molecular epidemiology research, and how their identities influence their power and vulnerability in the larger society, is important for anticipating study impacts. Consider the previously-mentioned example from Tordoff *et al.*, who decided not to publish study results based on feedback elicited from various community members and other stakeholders concerning a study about HIV transmission among men who have sex with men in Washington, US.[107] They concluded that earlier community engagement could have helped assuage concerns, especially about criminalization, and led to the collective development of more community-accepted questions.

Community engagement can also be a chance to demonstrate benefits of research for different groups. People may be motivated to participate because of a sense of solidarity with others who share similar

backgrounds or identities, particularly if they also face an elevated risk of contracting the studied disease.[121,122] Community outreach efforts must be cognizant of how societal norms and stigma could impact this sense of solidarity. Stigma and/or persecution may motivate the development of advocacy or representation groups that researchers can work with, but those who experience the most stigma may not be able to connect with such groups and may, therefore, be poorly represented.[121] Finally, any messaging about study participation benefits should avoid implying that an individual is obligated to participate in research in order to show solidarity to their community.[121]

Traditional epidemiologic studies can only draw inference about behaviors participants willingly disclose, but molecular epidemiology studies can draw inference about undisclosed behaviors (sometimes called cryptic transmission), and this presents an additional challenge for obtaining input from *all* impacted populations. For example, some studies suggest men who don't disclose male sexual partners may be a *bridge* population for transmission between heterosexual and gay male populations.[27] These men who have sex with men but don't identify as gay may not be well represented by advocacy from gay communities. To better understand the potential impacts of research on these individuals, researchers might consider eliciting feedback from openly gay men who report a history of not disclosing male sexual encounters in prior clinical visits.

Lastly, it may be tempting to assume that behaviors or experiences connected to advocacy groups in one setting, such as a researcher's home country, are important parts of people's identities in the research setting. Such assumptions could lead to misidentifying impacted communities and/or misunderstanding power dynamics between groups. Incorporating local knowledge in community outreach efforts can ensure representation of identity groups that are relevant in the local setting.

Geographic and political communities

Concerns about injustice are especially acute when studies work with communities that are largely isolated from the research beneficiaries. For example, immigrants and migrants may be socially, culturally, or politically isolated from the geographic community and be especially vulnerable to a narrative that depicts them as outsiders, bringing disease into that community.[123,124] People who share a geographic community can also be collectively vulnerable to outside stigma, such as narratives that depict the region as a source of an outbreak or blame the region for failing to control disease spread. For example, South Africa faced travel restrictions after being the first country to identify the SARS-CoV-2 Omicron variant,[125] despite the fact that the variant likely did not arise in South Africa and increasing evidence that targeted travel restrictions were ineffective.[126] While that research was conducted locally (then misinterpreted on the world stage), this problem could worsen when the narrative is driven by outside researchers, who may lack a vested interest or understanding of the economic and social vulnerabilities of geographic communities.

Studies that identify high rates of stigmatized behaviors can also be harmful to a region, both because they may impact how outside groups view that region and because they could insight internal conflict or punitive measures.[27] Examples of unreported behaviors that molecular epidemiology might infer include high numbers of age-disparate relationships or high numbers of sexual partners. Local leadership may provide valuable guidance about local community interests; however, existing systems may not be set up to equally protect all members of a geographic community. For instance, Ogunrin *et al.* notes that while older African participants in human genetics research tended to defer to local leadership, young participants tended to express more skepticism of local leadership and felt more solidarity with their social network.[122] This is why it's important to understand drivers of community from multiple levels.

Recommendations

Researchers can better collaborate with communities if they: 1) Define the community being studied in terms of identity and/or geographic or political boundaries. 2) Understand community dynamics, such as the relationship between people living with and without the studied disease and how people's identities relate to the behaviors studied. 3) Try to represent each impacted group in community engagement efforts and develop research questions that are sensitive to these community relationships. It will be important to ensure that all relevant community voices are heard, and not merely the most empowered voices.

It's also important to note that not every source of pathogen data or application of molecular epidemiology will be sensitive or necessitate engaging all the above concepts of community. For example, not all diseases have strong disproportionate impacts by identity groups. The passing of time, particularly for infections with high rates of recovery, should also lessen the sensitivity of the research question to the community. More guidance is needed to specify paradigms for when pathogen data is likely to be sensitive to a community so that resources can be invested on productive consultations or collaborations without hindering the use of many large and valuable pathogen datasets.

3. EXAMPLES OF COMMUNITY DATA MODELS

Several societies already have codified concepts of community data, for example governing biobanks and genetic information.[121,122] This section explores how communities with an established precedent for some types of collective community data have been able to impart community values into research using this data. Their ethical reasons for conceptualizing data in this way carry relevance for pathogen data as well.

American Indian tribal data sovereignty as a model of community data

Several American Indian tribes, like the Havasupai and Navaho, regulate use of genetic data and some biological samples from tribe members on the grounds that this is a form of cultural property.[127] Some tribes were motivated to assert data sovereignty in response to unauthorized research by universities using Havasupai genetic samples to study schizophrenia, alcohol dependency, "inbreeding," and ancient human migration.[127–130] Taking control over how its tribal members' genetic data is used allows the Havasupai to ensure research considers cultural beliefs and sensitivities, such as respecting the tribal origin story and protecting the sacredness of blood.

Laws and regulations do not always reflect community sentiment, but many stakeholders among American Indian, Alaska Native, and Native Hawaiian communities have also expressed a preference for community ownership and control over genetic data (as opposed to control by researchers or institutions).[131] While pathogen sequence data is unlikely to hold the same cultural significance as human genetic data, and its applications are less broad – SARS-CoV-2 sequences could not, for example, be used to study HIV transmission – we've already shown that transmission dynamic research can have other important political and cultural significances. Given that discrimination and crimes against American Indian tribes were often justified by assertions about the inferiority of their behaviors or cultural practices, it's not difficult to imagine potential harms from ill-conceived investigations into how certain behaviors contribute to the spread of disease in these communities.

Data protection acts from sub-Saharan African countries as models of community data

Data protection Acts and guidelines in many sub-Saharan African countries[132,133] reflect a similar, though less explicit, concept of community data. A history of, and fears about, colonialist research were the likely impetus behind the laws of several countries that variously: restrict sample shipments out of the country, prohibit indefinite sample storage, require local investigator involvement, build local capacity, and/or require that research by outside *vendors* must benefit the local community.[132–135] These laws

formalize the view that local communities are best placed to handle local samples and data. Community leadership or governance can add important context about community dynamics, weigh the value of the research question against its social and political ramifications, and identify and manage non-obvious risks.[136] These examples illustrate ways in which communities have effectively exerted their collective interests in the samples collected from individuals.

Data protection laws may serve to build local capacity and promote more research equity – that is, more research developed and implemented by local researchers using local infrastructure. Research equity is still lacking in pathogen research, particularly for novel pathogens where rapid analyses benefit from existing infrastructure and funds. For example, only 0.6% of COVID-19 related publications in the first three months of the outbreak were from African researchers or institutions. Despite the high burden of HIV in many African countries, Africa and the Middle East contribute only 21.61% of HIV research and were further under-represented as primary authors.[137,138] Data sharing and molecular pathogen surveillance – areas of pathogen research sometimes critiqued for failing to respect the autonomy and privacy of the individual – may be valuable tools for promoting more local research because they enable the utilization of sequence data from local health departments or available from prior studies. But while local control over research should increase knowledge about communities and community dynamics, the geographic connection is only one aspect of community. Healthcare workers, researchers, and other staff may have different primary languages or different cultural customs from study participants and others impacted by the research, particularly those from marginalized populations.[139] Locally-led research, therefore, does not negate the need for community outreach, but will likely make outreach more effective and impactful.

The concept of community data has not been explicitly tied to pathogen sequence research, and not all societies have a formal conception of community data. The US, for example, does not have an equivalent data protection act concerning ownership, storage, or reuse of biological specimens or a legal

acknowledgement that certain data may be sensitive to a wider community.[140,141] US data laws, like the *Genetic Information Nondiscrimination Act* (2008) and *HIPAA,* tend to emphasize individual privacy protections.[142,143] Jecker et al. note this contrast, stating that many African governments are more likely to see their role as "promoting moral good" in their citizens, while Western governments are geared more toward protecting the rights of individuals to be self-deterministic.[144] Community data models may sometimes present a trade-off with individualism-focused ethics, such as by denying individuals the benefits of some research participation or preventing them contributing to a project they personally feel has value.[145,146] However, as previously discussed, the processes for collecting and sharing pathogen sequence data are often not conducive to an individual-level weighing of risks and benefits. A community data model, therefore, may sometimes be the most feasible way to consider diverse community viewpoints.

Ubuntu

The ethical and philosophical principle of *ubuntu*, while not a legally binding doctrine, illustrates how a communal view of pathogen sequence data might be accepted at the societal level. The term *ubuntu* likely originated with the Nguni people of South Africa, but the concept is referenced in political, social, and philosophic contexts throughout much of sub-Saharan Africa.[147] A critical component of this philosophy is the inter-connectedness of human beings and, as such, it has been proposed as an ethical justification or psychological motivation for adherence to public health measures and for participation in research, particularly during a pandemic.[144,148–150] We propose that the *ubuntu* philosophy would require us to consider the positive and negative impacts of molecular epidemiology transmission dynamic research for a widely-defined community. This is in contrast to the emphasis Molldrem and Smith put on weighing study impacts for the research participant when they worry that "expert stakeholders who conducted molecular HIV research studies (Mutenherwa et al. 2019, 68) could not categorically state whether these studies offered a favorable risk-benefit ratio for participants, an important condition of

ethical research (Buchanan and Miller 2006)."[112] Under the *ubuntu* philosophy, we both recognize that a wider community than the research participants is impacted by the research and recognize (as discussed in Section 2) that benefits to the larger community can also benefit research participants.

Implementation and limitations of community data models

Researchers have a duty to understand and respect prevailing views about individual rights or community obligations where the research is conducted. While no law or ethical philosophy can capture the full diversity of viewpoints held by different communities within a region, tailoring research questions to the values of a community may require giving more weight to individual autonomy in some societies and to community benefit in others. Nevertheless, there are several ways that we believe the community data models we propose should be factored into *all* molecular epidemiology research, even in societies with more individualistic focuses.

First, we should recognize that effective community engagement can benefit the individuals whose data is used in these analyses by identifying and advancing values that many research participants share with a larger community. We discussed in Section 1 that most molecular epidemiology research does not present individual privacy risks, but some uses may run afoul of community norms. Respecting these norms is one way that community input can benefit individual study participants.

Second, talking about data as a community resource can help avoid "othering" people living with infectious diseases. HIV-related advocacy has demonstrated the importance of emphasizing the shared humanity of people living with and without diseases like HIV, with efforts like the promotion of person-first language and campaigns like "Everyone has a status."[118] When proposing studies or presenting results, researchers can focus on a pathogen as an environmental danger within a community and potential research participants as community members who can provide valuable information about the history of the disease in that community.[95,104] By talking about people living with these diseases as part

of a larger community, we both address our own potential biases and increase the likelihood that people will accept pathogen sequence data as a community resource.

Nevertheless, some studies will invariably run into non-aligned interests between different communities or sub-communities. For example, people living with a disease may fear the research will be stigmatizing, while others may approve of its goals. In rare instances, it may be necessary to weigh the interests of different sub-communities. One example of a study that was justified despite the dissent of an impacted group is a phylogenetic investigation of the origins of the 2010 cholera epidemic in Haiti. The study provided supporting evidence that peacekeeping troops from Nepal introduced cholera to Haiti, resulting in >7,000 deaths within 14 months.[151] Because the Nepalese troops refused to provide samples, researchers conducted analyses comparing cholera samples from Haiti to reference samples from Nepal and India. This study had substantial public health importance, highlighting the testing, sanitation, and disclosure failures that contributed to the devastating outbreak. Molecular epidemiology investigations must be sensitive to the political background and social sensitivities, but where there is disapproval from impacted communities, researchers must weigh this against public health benefits and risks.

In weighing differing views from impacted communities, we should consider factors like the power dynamics between these communities, being especially cautious if objectors are from an already marginalized group. We should carefully evaluate whether the community input we've elicited is representative. For populations that are heavily marginalized and difficult to reach, this may mean instead eliciting feedback from those who work directly with people in these populations. These evaluations must be specific to the local social or cultural context. We should consider not only potential harm but also potential benefits for different communities. Finally, we can look for ways to reduce the disparate interests of groups, such as by working to formulate research questions, goals, and presentations in a way that is least stigmatizing or to find intersecting interests between groups.

CONCLUSION

Conceptualizing pathogen data as a community resource, with the potential for benefit and harm, helps us consider community sensitivities that might be overlooked if the focus is exclusively on individual sensitivities. A communal data view shows that many critiques about data privacy in molecular epidemiology research can be addressed through community consultation, without stymieing important research efforts, particularly in disease surveillance. While most settings don't have legal doctrines and/or leadership specifically established to protect community data, molecular epidemiology research may be improved through careful evaluation of how it implicates different community groups, beyond just those whose data is used in research. Efforts to improve community engagement must acknowledge that the public health benefits of molecular epidemiological research depend heavily on efficient collection of large datasets. In most cases, these public health benefits inherently reflect community interests, but in some instances they may also present community risks, particularly to minority and/or marginalized sub-communities. While certainly not a comprehensive list, ways to reduce potential groups harms might include: more communication about study methods, goals, and benefits; more collaborating on how to present results in non-stigmatizing ways; or further discussing the process of data anonymization and sharing.

CONCLUSION

In Chapter 1, cluster and discrete trait analyses showed mixing between HIV-1 sequences from PWID and other populations, suggesting overlapping transmission networks. In Chapter 2, we found weak evidence of HIV sequence similarity among sexual partners identified through APS and possible HCV sequence similarity among injecting partners identified through APS. Finally, in Chapter 3, we suggested that a community-data model could add much needed clarity to the existing ethical discourse around molecular epidemiology research and promote productive community engagement efforts.

The molecular epidemiological evidence presented here corroborates prior research,[30] suggesting decreasing levels of parenteral HIV transmission among PWID in Kenya. While prior epidemiological analyses were not able to resolve sources of sexual HIV transmission,[30] our results suggest that PWID may acquire HIV sexually from people not in key populations and from other PWID and to a lesser extent, from female sex workers. Our conclusions about the relative commonality of sexual vs parenteral HIV transmission are cautious in light of low sampling density and possible sampling bias. Finally, although limited by sample size, we found suggestive evidence that APS helped reach some injecting partners within a shared HCV transmission network and some sexual partners within shared HIV transmission networks.

The best tools to address gaps in HIV and HCV diagnoses and reduce transmission today may not be the same tools that were effective in the past. Our findings suggest the need for more services to address sexual transmission in this population and a need to make existing harm reduction tools more accessible to people who use drugs. For example, prior research shows low knowledge about PrEP among PWID but suggests high demand for PrEP delivered at harm reduction drop-in centers.[152] Future research could also investigate whether PWID could benefit from longer-lasting PrEP delivery mechanisms, discrete packaging, or community education to reduce stigma and misconceptions. Exchanging money or goods

for sex is common among PWID and transactional sex – including among people who are not traditionally included in services for sex workers – may be an under-addressed contributor to transmission.[30] Future research could investigate the experiences and specific transmission and acquisition risk factors for PWID who engage in transactional sex.[13]

While better understanding the HIV and HCV epidemics among PWID can guide the development of more tailored HIV harm-reduction programs, addressing the stigma associated with drug use is crucial to making these services accessible. In the attempt to characterize transmission patterns, we must be careful not to promote further stigma. We have suggested that because pathogen data is inherently about communities and community dynamics, our primary focus in evaluating the ethics of this research should be understanding and addressing potential group harms and benefits. Voices from lower-resourced regions and from disempowered populations, like PWID, are lacking in this space. Through community-level collaboration and education – that is, educating communities about molecular epidemiology and educating researchers about community needs and dynamics – we can elicit a more complete picture of how this research impacts different communities.

REFERENCES

1. 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. Accessed August 29, 2022.

2. National AIDS and STI Control Program, National AIDS Control Council. *Kenya HIV Estimates Report 2018*. Ministry of Health, Kenya; 2018. https://nsdcc.go.ke/download/kenya-hiv-estimates-report-2018/. Accessed November 27, 2023.

3. Cherutich P, Golden MR, Wamuti B, et al. Assisted partner services for HIV in Kenya: a cluster randomised controlled trial. *Lancet HIV*. 2017;4(2):e74-e82. doi:10.1016/S2352-3018(16)30214-4

4. 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. April 2021. doi:10.1136/bmjopen-2020-041083

5. Sharma M, Kariithi E, Kemunto E, et al. High Acceptability of Assisted Partner Notification Services Among HIV-Positive Females in Kenya: Results From an Ongoing Implementation Study. *Journal of acquired immune deficiency syndromes (1999)*. 2021;86(1). doi:10.1097/QAI.0000000000002527

6. Dijkstra M, Mohamed K, Kigoro A, et al. Peer Mobilization and Human Immunodeficiency Virus (HIV) Partner Notification Services Among Gay, Bisexual, and Other Men Who Have Sex With Men and Transgender Women in Coastal Kenya Identified a High Number of Undiagnosed HIV Infections. *Open Forum Infect Dis*. 2021;8(6):ofab219. doi:10.1093/ofid/ofab219

7. Barr D, P Garnett G, Mayer KH, Morrison M. Key populations are the future of the African HIV/AIDS pandemic. *J Int AIDS Soc*. 2021;24(Suppl 3):e25750. doi:10.1002/jia2.25750

8. 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(12):e1192-e1207. doi:10.1016/S2214-109X(17)30375-3

9. 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(4):420-427. doi:10.1097/QAI.0000000000000769

10. 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/. Accessed August 29, 2022.

11. National AIDS Control Council, National AIDS & STI Control Program (NASCOP). *Kenya AIDS Response Progress Report 2018*. Nairobi, Kenya: Kenya Ministry of Health; 2018. https://nsdcc.go.ke/download/kenya-aids-response-progress-report-2018/.

12. Joint United Nations Programme on HIV/AIDS. *The Path That Ends AIDS: UNAIDS Global AIDS Update 2023*. Geneva, Switzerland: UNAIDS; 2023.

https://www.unaids.org/en/resources/documents/2023/global-aids-update-2023. Accessed November 27, 2023.

13. Aung SWKH, Kingston H, Mbogo LW, et al. Prevalence and correlates of violence among sexual and injecting partners of people who inject drugs living with HIV in Kenya: a cross-sectional study. *Harm Reduct J*. 2023;20(1):164. doi:10.1186/s12954-023-00895-7

14. Strathdee SA, Hallett TB, Bobrova N, et al. HIV and risk environment for injecting drug users: the past, present, and future. *Lancet*. 2010;376(9737):268-284. doi:10.1016/S0140-6736(10)60743-X

15. Gogia M, Lawlor C, Shengelia N, Stvilia K, Raymond HF. 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):1-12. doi:10.1186/s12954-019-0287-5

16. Beckerleg SE, Hundt GL. The characteristics and recent growth of heroin injecting in a Kenyan coastal town. *Addiction Research & Theory*. 2004;12(1):41-53. doi:10.1080/16066350410001646605

17. Ludwig-Barron NT, Guthrie BL, Mbogo L, et al. Barriers and facilitators of HIV and hepatitis C care among people who inject drugs in Nairobi, Kenya: a qualitative study with peer educators. *Harm Reduction Journal*. 2021;18(1):133. doi:10.1186/s12954-021-00580-7

18. World Health Organization. *Guidelines for the Care and Treatment of Persons Diagnosed with Chronic Hepatitis C Virus Infection*. Geneva: World Health Organization; 2018. https://apps.who.int/iris/handle/10665/273174. Accessed March 13, 2023.

19. Beckerleg S, Telfer M, Hundt GL. The rise of injecting drug use in east Africa: a case study from Kenya. *Harm Reduction Journal*. 2005;2(1):1-9. doi:10.1186/1477-7517-2-12

20. 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. https://hivpreventioncoalition.unaids.org/country-action/key-population-mapping-and-size-estimation-in-selected-counties-in-kenya-phase-one-key-findings-april-2019/.

21. 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. Published 2014.

22. Asher AK, Hahn JA, Couture MC, Maher K, Page K. People who Inject Drugs, HIV Risk, and HIV Testing Uptake in Sub-Saharan Africa. *J Assoc Nurses AIDS Care*. 2013;24(6):e35-e44. doi:10.1016/j.jana.2012.09.003

23. Brodish P, Singh K, Rinyuri A, et al. Evidence of high-risk sexual behaviors among injection drug users in the Kenya PLACE Study. *Drug and alcohol dependence*. 2011;119(1-2):138-141. doi:10.1016/j.drugalcdep.2011.05.030

24. 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 Evolution*. 2022;8(1):veac016. doi:10.1093/ve/veac016

25. 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(3):e007198. doi:10.1136/bmjopen-2014-007198

26. 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. doi:10.1186/s12879-022-07036-8

27. 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(3):205-212. doi:10.1097/COH.0000000000000537

28. Akiyama MJ, Cleland CM, Lizcano JA, Cherutich P, Kurth AE. 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. *The Lancet Infectious diseases*. 2019;19(11):1255-1263. doi:10.1016/S1473-3099(19)30264-6

29. 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(1):1549. doi:10.1186/s12889-021-11373-9

30. 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(15):2191. doi:10.1097/QAD.0000000000003382

31. World Health Organization. *Guidelines on HIV Self-Testing and Partner Notification: Supplement to Consolidated Guidelines on HIV Testing Services*. Geneva: World Health Organization; 2016. https://www.ncbi.nlm.nih.gov/books/NBK401684/. Accessed August 29, 2022.

32. Ludwig-Barron NT, Guthrie BL, Mbogo L, et al. Barriers and facilitators of HIV and hepatitis C care among people who inject drugs in Nairobi, Kenya: a qualitative study with peer educators. *Harm Reduct J*. 2021;18(1):133. doi:10.1186/s12954-021-00580-7

33. Masese LN, Ludwig-Barron NT, Mbogo L, et al. Occupational roles and risks of community-embedded peer educators providing HIV, hepatitis C and harm reduction services to persons who inject drugs in Nairobi, Kenya. *PLOS ONE*. 2022;17(12):e0278210. doi:10.1371/journal.pone.0278210

34. World Health Organization. Hepatitis. World Health Organization. https://www.who.int/health-topics/hepatitis/elimination-of-hepatitis-by-2030. Accessed November 24, 2023.

35. Harris M, Rhodes T. Hepatitis C treatment access and uptake for people who inject drugs: a review mapping the role of social factors. *Harm Reduction Journal*. 2013;10(1):7. doi:10.1186/1477-7517-10-7

36. Katzman C, Mateu-Gelabert P, Kapadia SN, Eckhardt BJ. Contact tracing for hepatitis C: The case for novel screening strategies as we strive for viral elimination. *Int J Drug Policy*. 2019;72:33-39. doi:10.1016/j.drugpo.2019.04.003

37. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*. 2018;4(1):vey016. doi:10.1093/ve/vey016

38. 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. doi:10.7554/eLife.72657

39. 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. *Scientific Reports*. 2019;9(1):1-8. doi:10.1038/s41598-018-37458-x

40. 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(8):1643. doi:10.3390/v13081643

41. Grenfell BT, Pybus OG, Gog JR, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*. 2004;303(5656):327-332. doi:10.1126/science.1090727

42. 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*. 2022;13:843330. doi:10.3389/fmicb.2022.843330

43. Nduva GM, Hassan AS, Nazziwa J, Graham SM, Esbjörnsson J, Sanders EJ. HIV-1 Transmission Patterns Within and Between Risk Groups in Coastal Kenya. *Sci Rep*. 2020;10(1):6775. doi:10.1038/s41598-020-63731-z

44. Guidelines for HIV Testing and Counselling in Kenya: 2nd edition. 2010. http://guidelines.health.go.ke/. Accessed June 10, 2022.

45. 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(6):1261-1276. doi:10.1038/nprot.2017.066

46. 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;19:337-348. doi:10.1016/j.meegid.2013.04.032

47. WHO SDRM List - HIV Drug Resistance Database. https://hivdb.stanford.edu/page/who-sdrm-list/. Accessed June 8, 2022.

48. 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 (Oxford, England)*. 2012;28(12). doi:10.1093/bioinformatics/bts199

49. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*. 2013;30(4). doi:10.1093/molbev/mst010

50. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*. 2002;30(14). doi:10.1093/nar/gkf436

51. Joy JB, Liang RH, McCloskey RM, Nguyen T, Poon AFY. Ancestral Reconstruction. *PLoS Comput Biol*. 2016;12(7):e1004763. doi:10.1371/journal.pcbi.1004763

52. Royer-Carenzi M, Pontarotti P, Didier G. Choosing the best ancestral character state reconstruction method. *Mathematical Biosciences*. 2013;242(1):95-109. doi:10.1016/j.mbs.2012.12.003

53. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*. 2015;32(1):268-274. doi:10.1093/molbev/msu300

54. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*. 2020;37(5):1530-1534. doi:10.1093/molbev/msaa015

55. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587-589. doi:10.1038/nmeth.4285

56. Soubrier J, Steel M, Lee MSY, et al. The Influence of Rate Heterogeneity among Sites on the Time Dependence of Molecular Rates. *Molecular Biology and Evolution*. 2012;29(11):3345-3358. doi:10.1093/molbev/mss140

57. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*. 2010;59(3):307-321. doi:10.1093/sysbio/syq010

58. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian Phylogeography Finds Its Roots. *PLOS Computational Biology*. 2009;5(9):e1000520. doi:10.1371/journal.pcbi.1000520

59. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Molecular Biology and Evolution*. 2005;22(5):1185-1192. doi:10.1093/molbev/msi103

60. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics*. 2011;27(4):592-593. doi:10.1093/bioinformatics/btq706

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

62. Edwards CJ, Suchard MA, Lemey P, et al. Ancient Hybridization and an Irish Origin for the Modern Polar Bear Matriline. *Curr Biol*. 2011;21(15):1251-1258. doi:10.1016/j.cub.2011.05.058

63. Gao J, May MR, Rannala B, Moore BR. Model misspecification misleads inference of the spatial dynamics of disease outbreaks. *Proceedings of the National Academy of Sciences*. 2023;120(11):e2213913120. doi:10.1073/pnas.2213913120

64. Gao J, May MR, Rannala B, Moore BR. PrioriTree: a utility for improving phylodynamic analyses in BEAST. August 2022:2022.08.24.505196. doi:10.1101/2022.08.24.505196

65. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. *Molecular Biology and Evolution*. 2016;33(8):2167-2169. doi:10.1093/molbev/msw082

66. 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 Retroviruses*. 2013;29(1):187-190. doi:10.1089/AID.2012.0182

67. Monroe-Wise A. Assisted Partner Services for PWID to Identify Partners Living with HIV and HCV. 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.

68. Patel P, Borkowf CB, Brooks JT, Lasry A, Lansky A, Mermin J. Estimating per-act HIV transmission risk: a systematic review. *AIDS*. 2014;28(10):1509-1519. doi:10.1097/QAD.0000000000000298

69. Mwatelah RS, Lwembe RM, Osman S, et al. Co-Infection Burden of Hepatitis C Virus and Human Immunodeficiency Virus among Injecting Heroin Users at the Kenyan Coast. *PLoS One*. 2015;10(7):e0132287. doi:10.1371/journal.pone.0132287

70. Karoney MJ, Siika AM. Hepatitis C virus (HCV) infection in Africa: a review. *The Pan African Medical Journal*. 2013;14. doi:10.11604/pamj.2013.14.44.2199

71. Prequalified In Vitro Diagnostics. WHO - Prequalification of Medical Products (IVDs, Medicines, Vaccines and Immunization Devices, Vector Control). https://extranet.who.int/pqweb/vitro-diagnostics/vitro-diagnostics-lists. Published January 27, 2020. Accessed May 15, 2023.

72. Wertheim JO, Pond SLK, Forgione LA, et al. Social and Genetic Networks of HIV-1 Transmission in New York City. *PLOS Pathogens*. 2017;13(1):e1006000. doi:10.1371/journal.ppat.1006000

73. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*. 1993;10(3):512-526. doi:10.1093/oxfordjournals.molbev.a040023

74. Junqueira DM, Sibisi Z, Wilkinson E, de Oliveira T. Factors influencing HIV-1 phylogenetic clustering. *Curr Opin HIV AIDS*. 2019;14(3):161-172. doi:10.1097/COH.0000000000000540

75. Gore DJ, Schueler K, Ramani S, et al. HIV Response Interventions that Integrate HIV Molecular Cluster and Social Network Analysis: A Systematic Review. *AIDS Behav*. November 2021. doi:10.1007/s10461-021-03525-0

76. Dennis AM, Murillo W, de Maria Hernandez F, et al. Social network based recruitment successfully reveals HIV-1 transmission networks among high risk individuals in El Salvador. *J Acquir Immune Defic Syndr*. 2013;63(1):135-141. doi:10.1097/QAI.0b013e318288b246

77. Dennis AM, Cohen MS, Rucinski KB, et al. Human Immunodeficiency Virus (HIV)-1 Transmission Among Persons With Acute HIV-1 Infection in Malawi: Demographic, Behavioral, and Phylogenetic Relationships. *Clin Infect Dis*. 2019;69(5):853-860. doi:10.1093/cid/ciy1006

78. Nerlander LM, Hess KL, Rose CE, et al. Exchange Sex and HIV Infection Among Women Who Inject Drugs-20 US Cities, 2009. *J Acquir Immune Defic Syndr*. 2017;75 Suppl 3(Suppl 3):S333-S340. doi:10.1097/QAI.0000000000001408

79. Pilon R, Leonard L, Kim J, et al. Transmission Patterns of HIV and Hepatitis C Virus among Networks of People Who Inject Drugs. *PLOS ONE*. 2011;6(7):e22245. doi:10.1371/journal.pone.0022245

80. Kostaki EG, Nikolopoulos GK, Pavlitina E, et al. Molecular Analysis of Human Immunodeficiency Virus Type 1 (HIV-1)–Infected Individuals in a Network-Based Intervention (Transmission Reduction Intervention Project): Phylogenetics Identify HIV-1–Infected Individuals With Social Links. *J Infect Dis*. 2018;218(5):707-715. doi:10.1093/infdis/jiy239

81. Resik S, Lemey P, Ping LH, et al. Limitations to contact tracing and phylogenetic analysis in establishing HIV type 1 transmission networks in Cuba. *AIDS Res Hum Retroviruses*. 2007;23(3):347-356. doi:10.1089/aid.2006.0158

82. Poon AFY. Impacts and shortcomings of genetic clustering methods for infectious disease outbreaks. *Virus Evol*. 2016;2(2):vew031. doi:10.1093/ve/vew031

83. Campo D, Xia G, Dimitrova Z, et al. Accurate genetic detection of hepatitis C virus transmissions in outbreak settings. *J Infect Dis*. 2016;213(6):957-965. doi:10.1093/infdis/jiv542

84. Sacks-Davis R, Daraganova G, Aitken C, et al. Hepatitis C Virus Phylogenetic Clustering Is Associated with the Social-Injecting Network in a Cohort of People Who Inject Drugs. *PLoS One*. 2012;7(10):e47335. doi:10.1371/journal.pone.0047335

85. Aitken CK, McCaw RF, Bowden DS, et al. Molecular epidemiology of hepatitis C virus in a social network of injection drug users. *J Infect Dis*. 2004;190(9):1586-1595. doi:10.1086/424678

86. Ball LJ, Puka K, Speechley M, et al. Sharing of Injection Drug Preparation Equipment Is Associated With HIV Infection: A Cross-sectional Study. *J Acquir Immune Defic Syndr*. 2019;81(4):e99-e103. doi:10.1097/QAI.0000000000002062

87. Hagan H, Thiede H, Weiss NS, Hopkins SG, Duchin JS, Alexander ER. Sharing of drug preparation equipment as a risk factor for hepatitis C. *Am J Public Health*. 2001;91(1):42-46.

88. Baker RE, Mahmud AS, Miller IF, et al. Infectious disease in an era of global change. *Nat Rev Microbiol*. 2022;20(4):193-205. doi:10.1038/s41579-021-00639-z

89. Inzaule S, Tessema S, Kebede Y, Ogwell Ouma A, Nkengasong J. Genomic-informed pathogen surveillance in Africa: opportunities and challenges - The Lancet Infectious Diseases. 2021. doi:10.1016/S1473-3099(20)30939-7

90. Nduva GM, Nazziwa J, Hassan AS, Sanders EJ, Esbjörnsson J. The Role of Phylogenetics in Discerning HIV-1 Mixing among Vulnerable Populations and Geographic Regions in Sub-Saharan Africa: A Systematic Review. *Viruses*. 2021;13(6):1174. doi:10.3390/v13061174

91. German D, Grabowski MK, Beyrer C. Enhanced use of phylogenetic data to inform public health approaches to HIV among men who have sex with men. *Sex Health*. 2017;14(1):89-96. doi:10.1071/SH16056

92. Schairer C, Mehta SR, Vinterbo SA, Hoenigl M, Kalichman M, Little S. Perceptions of molecular epidemiology studies of HIV among stakeholders. *J Public Health Res*. 2017;6(3):992. doi:10.4081/jphr.2017.992

93. Martinez-Martin N, Wieten S, Magnus D, Cho MK. Digital Contact Tracing, Privacy, and Public Health. *Hastings Center Report*. 2020;50(3):43-46. doi:10.1002/hast.1131

94. Mehta SR, Schairer C, Little S. Ethical Issues in HIV Phylogenetics and Molecular Epidemiology. *Curr Opin HIV AIDS*. 2019;14(3):221-226. doi:10.1097/COH.0000000000000538

95. Mutenherwa F, Wassenaar DR, Oliveira T de. Experts' Perspectives on Key Ethical Issues Associated With HIV Phylogenetics as Applied in HIV Transmission Dynamics Research: *https://doi.org/101177/1556264618809608*. November 2018. https://journals.sagepub.com/doi/10.1177/1556264618809608.

96. Rothstein MA. Public Health and Privacy in the Pandemic. *Am J Public Health*. 2020;110(9):1374-1375. doi:10.2105/AJPH.2020.305849

97. Chung C, Khanna N, Cardell B, et al. Consent and criminalisation concerns over phylogenetic analysis of surveillance data. *The Lancet HIV*. 2019;6(7):e420. doi:10.1016/S2352-3018(19)30138-9

98. Wertheim JO, Morris S, Ragonnet-Cronin M. Consent and criminalisation concerns over phylogenetic analysis of surveillance data – Authors' reply. *The Lancet HIV*. 2019;6(7):e420-e421. doi:10.1016/S2352-3018(19)30142-0

99. Mutenherwa F, Wassenaar DR, de Oliveira T. Ethical issues associated with HIV phylogenetics in HIV transmission dynamics research: A review of the literature using the Emanuel Framework. *Developing World Bioethics*. 2019;19(1):25-35. doi:10.1111/dewb.12191

100. McClelland A, Guta A, Gagnon M. The rise of molecular HIV surveillance: implications on consent and criminalization. *Critical Public Health*. 2020;30(4):487-493. doi:10.1080/09581596.2019.1582755

101. Bernard EJ, McClelland A, Cardell B, et al. We Are People, Not Clusters! *The American Journal of Bioethics*. 2020;20(10). https://www.tandfonline.com/doi/full/10.1080/15265161.2020.1809550.

102. Benbow ND, Evans D. *Ethical Considerations for Molecular HIV Surveillance in the United States*. Galter Health Science Library & Learning Center; 2018. doi:10.18131/G3MT7B

103. Lee LM. Health Information in the Background: Justifying Public Health Surveillance Without Patient Consent. *Emerging Pervasive Information and Communication Technologies (PICT)*. 2013;11:39-53. doi:10.1007/978-94-007-6833-8_3

104. US PLHIV Caucus. Open Letter to the CDC: Networks of People Living with HIV Demand a Moratorium on Molecular HIV Surveillance. 2021. https://actionnetwork.org/forms/open-letter-to-the-cdc-we-demand-a-moratorium-on-molecular-hiv-surveillance. Accessed July 11, 2022.

105. Rudolph AE, Martinez O, Davison R, Amuchi CB. Informed consent for HIV phylogenetic research: A case study of urban individuals living with HIV approached for enrollment in an HIV study. *EHQUIDAD*. 2020;2020(14):129-143.

106. Mutenherwa F, Wassenaar DR, de Oliveira T. Ethical issues associated with HIV molecular epidemiology: a qualitative exploratory study using inductive analytic approaches. *BMC Medical Ethics*. 2019;20(1):67. doi:10.1186/s12910-019-0403-9

107. Tordoff D. Lessons Learned from Conducting Community Engagement on Phylodynamic Research with Molecular HIV Surveillance Data [preprint].

108. Le Vu S, Ratmann O, Delpech V, et al. Comparison of cluster-based and source-attribution methods for estimating transmission risk using large HIV sequence databases. *Epidemics*. 2018;23:1-10. doi:10.1016/j.epidem.2017.10.001

109. Ragonnet-Cronin M, Hu YW, Morris SR, Sheng Z, Poortinga K, Wertheim JO. HIV transmission networks among transgender women in Los Angeles County, CA, USA: a phylogenetic analysis of surveillance data. *Lancet HIV*. 2019;6(3):e164-e172. doi:10.1016/S2352-3018(18)30359-X

110. Dawson L, Benbow N, Fletcher FE, et al. Addressing Ethical Challenges in US-Based HIV Phylogenetic Research. *The Journal of Infectious Diseases*. 2020;222(12):1997-2006. doi:10.1093/infdis/jiaa107

111. Coltart CEM, Hoppe A, Parker M, et al. Ethical considerations in global HIV phylogenetic research. *Lancet HIV*. 2018;5(11):e656-e666. doi:10.1016/S2352-3018(18)30134-6

112. Molldrem S, Smith AKJ. Reassessing the Ethics of Molecular HIV Surveillance in the Era of Cluster Detection and Response: Toward HIV Data Justice. *https://doi.org/101080/1526516120201806373*. September 2020. https://www.tandfonline.com/doi/full/10.1080/15265161.2020.1806373.

113. Maria Vang Johansen, Jens Aagaard-Hansen, Povl Riis. Benefit--a neglected aspect of health research ethics. *Danish medical bulletin*. 2008;55(4). https://www.ncbi.nlm.nih.gov/pubmed/19232161.

114. Barrow JM, Brannan GD, Khandhar PB. Research Ethics. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; 2023. http://www.ncbi.nlm.nih.gov/books/NBK459281/. Accessed October 18, 2023.

115. Bragazzi NL, Khamisy-Farah R, Tsigalou C, Mahroum N, Converti M. Attaching a stigma to the LGBTQI+ community should be avoided during the monkeypox epidemic. *Journal of Medical Virology*. n/a(n/a). doi:10.1002/jmv.27913

116. Centers for Disease Control and Prevention (CDC). COVID-19: Medical Conditions. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html. Published May 11, 2023. Accessed December 11, 2023.

117. Centers for Disease Control and Prevention (CDC), National Center for Emerging and Zoonotic Infectious Diseases (NCEZID), Division of Vector-Borne Diseases (DVBD). Zika and Pregnancy. Centers for Disease Control and Prevention: Zika Virus. https://www.cdc.gov/zika/pregnancy/index.html. Published May 9, 2019. Accessed December 14, 2023.

118. MacQueen KM, McLellan E, Metzger DS, et al. What Is Community? An Evidence-Based Definition for Participatory Public Health. *Am J Public Health*. 2001;91(12):1929-1938.

119. Liu J. Disease outbreak: Finish the fight against Ebola. *Nature*. 2015;524(7563):27-29. doi:10.1038/524027a

120. SciELO - Brazil - Impact of super-spreaders on COVID-19: systematic review Impact of super-spreaders on COVID-19: systematic review. https://www.scielo.br/j/spmj/a/fMJjf9mKW3d8Rg9bmX4hHjR/?lang=en&format=html. Accessed November 18, 2023.

121. Neuhaus CP. Does Solidarity Require "All of Us" to Participate in Genomics Research? *Hastings Center Report*. 2020;50(S1):S62-S69. doi:10.1002/hast.1157

122. Ogunrin O, Woolfall K, Gabbay M, Frith L. Relative solidarity: Conceptualising communal participation in genomic research among potential research participants in a developing Sub-Saharan African setting. *PLOS ONE*. 2018;13(4):e0195171. doi:10.1371/journal.pone.0195171

123. Sandset T. The ethical and epistemological pitfalls of translating phylogenetic HIV testing: from patient-centered care to surveillance. *Humanities and Social Sciences Communications*. 2020;7(1):1-10. doi:10.1057/s41599-020-0522-4

124. Cella E, Foley BT, Riva E, et al. HIV-2 Infection in a Migrant from Gambia: The History of the Disease Combined with Phylogenetic Analysis Revealed the Real Source of Infection. *AIDS Research and Human Retroviruses*. 2018;34(12):1090-1094. doi:10.1089/aid.2018.0112

125. Tegally H, Moir M, Everatt J, et al. Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. *Nat Med*. 2022;28(9):1785-1790. doi:10.1038/s41591-022-01911-2

126. Zhang Q, Ye Y, Zeng DD. Travel restrictions cannot prevent the introduction of new COVID variants. *Journal of Travel Medicine*. 2023;30(5):taad066. doi:10.1093/jtm/taad066

127. Harry D, Kanehe LM. Asserting Tribal Sovereignty over Cultural Property: Moving Towards Protection of Genetic Material and Indigenous Knowledge.

128. Garrison NA. Genomic Justice for Native Americans: Impact of the Havasupai Case on Genetic Research. *Sci Technol Human Values*. 2013;38(2):201-223. doi:10.1177/0162243912470009

129. Sterling RL. Genetic Research among the Havasupai: A Cautionary Tale. *AMA Journal of Ethics*. 2011;13(2):113-117. doi:10.1001/virtualmentor.2011.13.2.hlaw1-1102

130. Chadwick JQ, Copeland KC, Branam DE, et al. Genomic Research and American Indian Tribal Communities in Oklahoma: Learning From Past Research Misconduct and Building Future Trusting Partnerships. *Am J Epidemiol*. 2019;188(7):1206-1212. doi:10.1093/aje/kwz062

131. Garrison NA, Barton KS, Porter KM, Mai T, Burke W, Carroll SR. Access and Management: Indigenous Perspectives on Genomic Data Sharing. *Ethnicity & Disease*. 2019;29(Suppl 3):659. doi:10.18865/ed.29.S3.659

132. Nienaber A. Consent to and authorisation of the export and use of human biological specimens for future research - perspectives from three African countries. 2011:30.

133. de Vries J, Munung SN, Matimba A, et al. Regulation of genomic and biobanking research in Africa: a content analysis of ethics guidelines, policies and procedures from 22 African countries. *BMC Medical Ethics*. 2017;18(1):8. doi:10.1186/s12910-016-0165-6

134. Data Protection Act – Office of the Data Protection Commissioner of Kenya. https://www.odpc.go.ke/dpa-act/. Accessed March 3, 2023.

135. Staunton C, de Vries J. The governance of genomic biobank research in Africa: reframing the regulatory tilt. *J Law Biosci*. 2020;7(1):lsz018. doi:10.1093/jlb/lsz018

136. Fregonese F. Community involvement in biomedical research conducted in the global health context; what can be done to make it really matter? *BMC Med Ethics*. 2018;19(Suppl 1):44. doi:10.1186/s12910-018-0283-4

137. González-Alcaide G, Menchi-Elanzi M, Nacarapa E, Ramos-Rincón JM. HIV/AIDS research in Africa and the Middle East: participation and equity in North-South collaborations and relationships. *Globalization and Health*. 2020;16(1):1-18. doi:10.1186/s12992-020-00609-9

138. Gwenzi W, Rzymski P. When silence goes viral, Africa sneezes! A perspective on Africa's subdued research response to COVID-19 and a call for local scientific evidence. *Environmental research*. 2021;194. doi:10.1016/j.envres.2020.110637

139. Tindana P, Depuur C, de Vries J, Seeley J, Parker M. Informed consent in genomic research and biobanking: taking feedback of findings seriously. *Glob Bioeth*. 31(1):200-215. doi:10.1080/11287462.2020.1717896

140. Langat SK. Reuse of samples: ethical issues encountered by two institutional ethics review committees in Kenya. *Bioethics*. 2005;19(5-6):537-549. doi:10.1111/j.1467-8519.2005.00463.x

141. Biobanking in health care: evolution and future directions - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6532145/. Accessed November 17, 2022.

142. The Genetic Information Nondiscrimination Act of 2008. 2008;(Bills/Resolutions).

143. *The Health Insurance Portability and Accountability Act of 1996*.; 1996.

144. Jecker NS, Atuire CA, Kenworthy N. Realizing Ubuntu in Global Health: An African Approach to Global Health Justice. *Public Health Ethics*. September 2022:phac022. doi:10.1093/phe/phac022

145. Mungwira RG, Nyangulu W, Misiri J, et al. Is it ethical to prevent secondary use of stored biological samples and data derived from consenting research participants? The case of Malawi. *BMC Med Ethics*. 2015;16(1):83. doi:10.1186/s12910-015-0077-x

146. Reardon S. Navajo Nation reconsiders ban on genetic research. *Nature*. 2017;550(7675):165-166. doi:10.1038/nature.2017.22780

147. Gade CB. The Historical Development of the Written Discourses on \emphUbuntu. *South African Journal of Philosophy*. 2011;30(3):303-329. doi:10.4314/sajpem.v30i3.69578

148. Sambala EZ, Cooper S, Manderson L. Ubuntu as a Framework for Ethical Decision Making in Africa: Responding to Epidemics. *Ethics & Behavior*. 2020;30(1):1-13. doi:10.1080/10508422.2019.1583565

149. Chigangaidze RK, Matanga AA, Katsuro TR. Ubuntu Philosophy as a Humanistic–Existential Framework for the Fight Against the COVID-19 Pandemic. *Journal of Humanistic Psychology*. 2022;62(3):319-333. doi:10.1177/00221678211044554

150. Menkiti I. African Philosophy, an Introduction Person and Community in African Traditional Thought. In: Wright R, ed. *African Philosophy: An Introduction*. 3rd ed. Lanham, MD: University Press of America.

151. Frerichs RR, Keim PS, Barrais R, Piarroux R. Nepalese origin of cholera epidemic in Haiti. *Clin Microbiol Infect*. 2012;18(6):E158-163. doi:10.1111/j.1469-0691.2012.03841.x

152. Cosmas M, Loice M, William S, et al. Oral Pre-Exposure Prophylaxis (PrEP) Awareness and Acceptability Among Persons who Inject Drugs (PWID) in Kenya: A Qualitative Investigation. *Res Sq*. February 2023:rs.3.rs-2512731. doi:10.21203/rs.3.rs-2512731/v1

SUPPLEMENTAL

Supplementary Table S1.1. Demographics, test results, and characteristics of SHARP study participants with HIV-1 sequences included in the analysis.

	Coast (N=124)	Nairobi (N=179)	Total (N=303)	p value*
Demographics				
Female	51 (41.1%)	108 (60.3%)	159 (52.5%)	< 0.001
Key population**				-
PWID	98 (79.0%)	133 (74.3%)	231 (76.2%)	
Female participant with transactional sex risk factors***	1 (0.8%)	0 (0.0%)	1 (0.3%)	
PWID & Female participant with transactional sex risk factors	15 (12.1%)	37 (20.7%)	52 (17.2%)	
MSM	0 (0.0%)	1 (0.6%)	1 (0.3%)	
MSM&PWID	4 (3.2%)	3 (1.7%)	7 (2.3%)	
Not-KP	6 (4.8%)	5 (2.8%)	11 (3.6%)	
HIV history				
Diagnosed before enrollment	102 (82.3%)	141 (78.8%)	243 (80.2%)	0.52
Years since diagnosis (previously diagnosed only, N=243) (median, inter-quartile range)	4 (1.1-8)	3.5 (1-9)	4 (1-8.5)	0.91
Behaviors				
Active Injection Drug Use (in last month)	97 (78.2%)	154 (86.0%)	251 (82.8%)	0.076
Shared needles in the last month				0.070
No	93 (75.0%)	140 (78.2%)	233 (76.9%)	
Yes	4 (3.2%)	14 (7.8%)	18 (5.9%)	
Don't inject or didn't inject in last month	27 (21.8%)	25 (14.0%)	52 (17.2%)	
number of sex partners in the previous 3 months (1 missing)				< 0.001
0	37 (29.8%)	99 (55.6%)	136 (45.0%)	
1to2	56 (45.2%)	41 (23.0%)	97 (32.1%)	
more_than_2	31 (25.0%)	38 (21.3%)	69 (22.8%)	
Ever received money or goods for sex (Females, N=159)	41 (33.1%)	85 (47.8%)	126 (41.7%)	0.80
Ever received money or goods for sex (Males, N=143, 1 missing)	21 (16.9%)	16 (9.0%)	37 (12.3%)	0.42
Housing and Mobility				
Report transportation as a barrier to receiving care	66 (53.2%)	56 (31.3%)	122 (40.3%)	< 0.001
Have stable housing	100 (80.6%)	151 (84.4%)	251 (82.8%)	0.40
Spent ≥1 week outside city in the last 6 months	37 (29.8%)	20 (11.2%)	57 (18.8%)	< 0.001

*For categorical variables: chi-square test, for continuous variables: t-test with equal variance

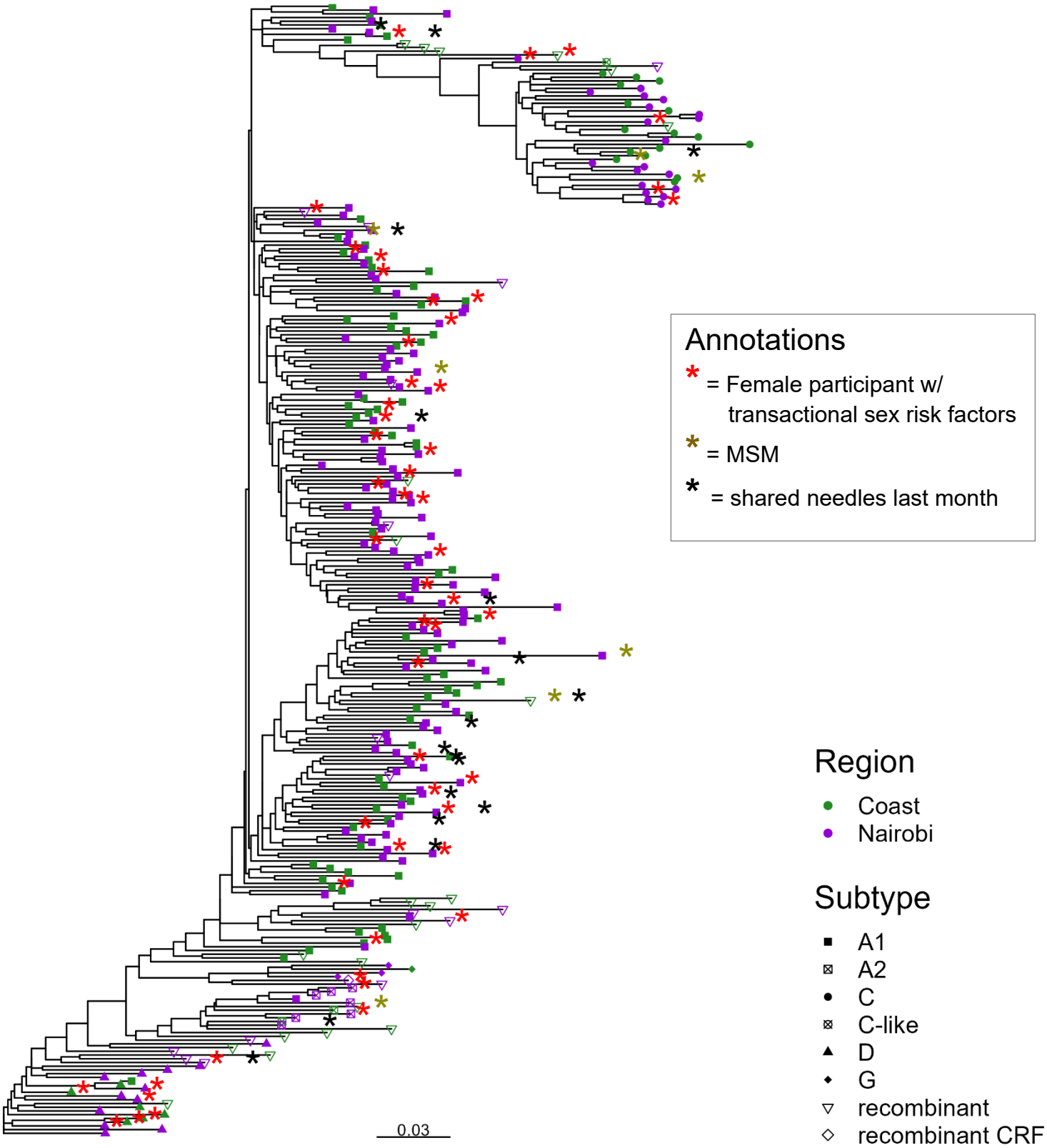
**For ancestral state reconstruction, participants are classified into only 1 population (see Table 1). MSM who are also PWID are called MSM (this was done to address recruitment bias, as most of the sequences from PWID and very few sequences from MSM are from the SHARP study).

***Have ever received money for sex and reported ≥3 sex partners in the prior month. This classification was not used to define key populations for ancestral state reconstruction.

Supplementary Table S1.2. Participant demographics and distribution of new (from SHARP study) and previously published Kenyan HIV-1 *pol* sequences by region. Only sequences from Coast and Nairobi are included. Abbreviations: FSW: female sex workers, MSM: men who have sex with men, PWID: people who injects drugs, not-KP: not key population.

	Coast (N=1159)	Nairobi (N=1507)	Total (N=2666)
Source			
published	1035 (89.3%)	1328 (88.1%)	2363 (88.6%)
SHARP	124 (10.7%)	179 (11.9%)	303 (11.4%)
Sampling year			
2001-2015	960 (82.8%)	1119 (74.3%)	2079 (78.0%)
2015+	199 (17.2%)	388 (25.7%)	587 (22.0%)
Key population			
FSW	110 (9.5%)	82 (5.4%)	192 (7.2%)
Not-KP	700 (60.4%)	1114 (73.9%)	1814 (68.0%)
MSM	178 (15.4%)	141 (9.4%)	319 (12.0%)
PWID	171 (14.8%)	170 (11.3%)	341 (12.8%)
HIV subtype			
N-Miss	9	5	14
A1	817 (71.0%)	1068 (71.1%)	1885 (71.1%)
A2	2 (0.2%)	8 (0.5%)	10 (0.4%)
B	0 (0.0%)	1 (0.1%)	1 (0.0%)
C	91 (7.9%)	101 (6.7%)	192 (7.3%)
D	91 (7.9%)	174 (11.6%)	265 (10.0%)
G	5 (0.4%)	10 (0.7%)	15 (0.6%)
recombinant	144 (12.5%)	140 (9.4%)	284 (10.7%)
Sex			
N-Miss	633	989	1622
Female	237 (45.1%)	267 (51.5%)	504 (48.3%)
Male	289 (54.9%)	251 (48.5%)	540 (51.7%)

Supplementary Figure S1.1. Phylogeny of HIV-1 sequences from the SHARP study.



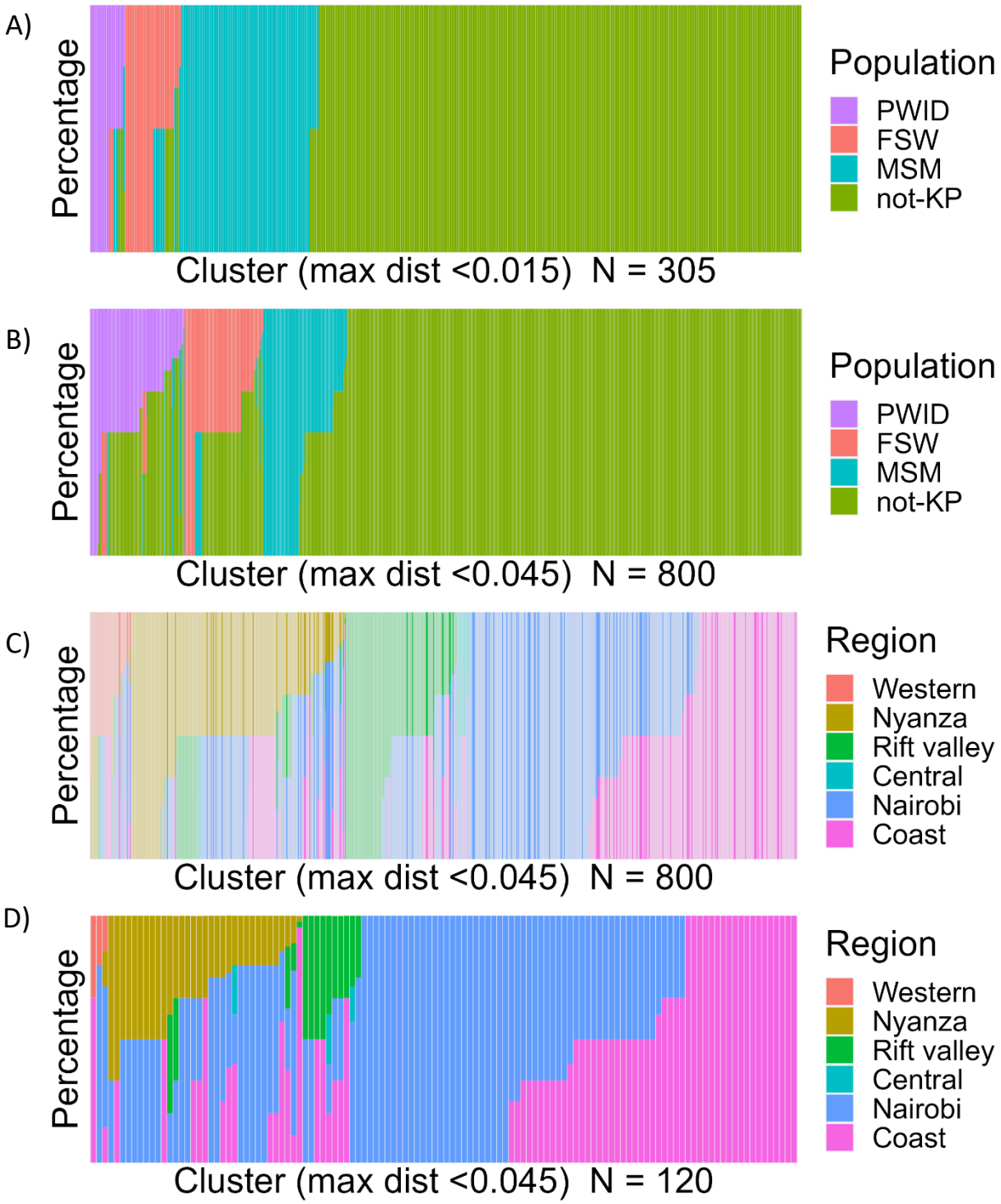
Supplementary Table S1.3. Distribution of population groups within clusters that contain sequences from coastal Kenya or Nairobi. Only clusters with at least 1 sequence from the coast or Nairobi are included. Custers are based on a 0.045 maximum patristic distance threshold on maximum likelihood trees for subtypes A1, C, and D. Clusters containing sequences from PWID are in bold. Abbreviations: FSW: female sex workers, MSM: men who have sex with men, not-KP: not key population, PWID: people who injects drugs.

Cluster membership	**dyad**	**Network (>2 sequences)**	**total**
Single-population clusters	263	197	460
not-KP only	230	170	400
FSW only	9	2	11
MSM only	16	22	38
PWID only	**8**	**3**	**11**
PWID & other populations	45	63	109
PWID, not-KP	**37**	**48**	**85**
PWID, FSW	**5**	**0**	**5**
PWID, MSM	**3**	**4**	**7**
PWID, not-KP, FSW	**-**	**3**	**3**
PWID, not-KP, MSM	**-**	**7**	**7**
PWID, FSW, MSM	**-**	**2**	**2**
PWID, not-KP, FSW, MSM	**-**	**0**	**0**
Other	67	44	111
not-KP, FSW	35	15	50
not-KP, MSM	25	24	49
FSW, MSM	7	0	7
not-KP, FSW, MSM	-	5	5
Total	375	305	680

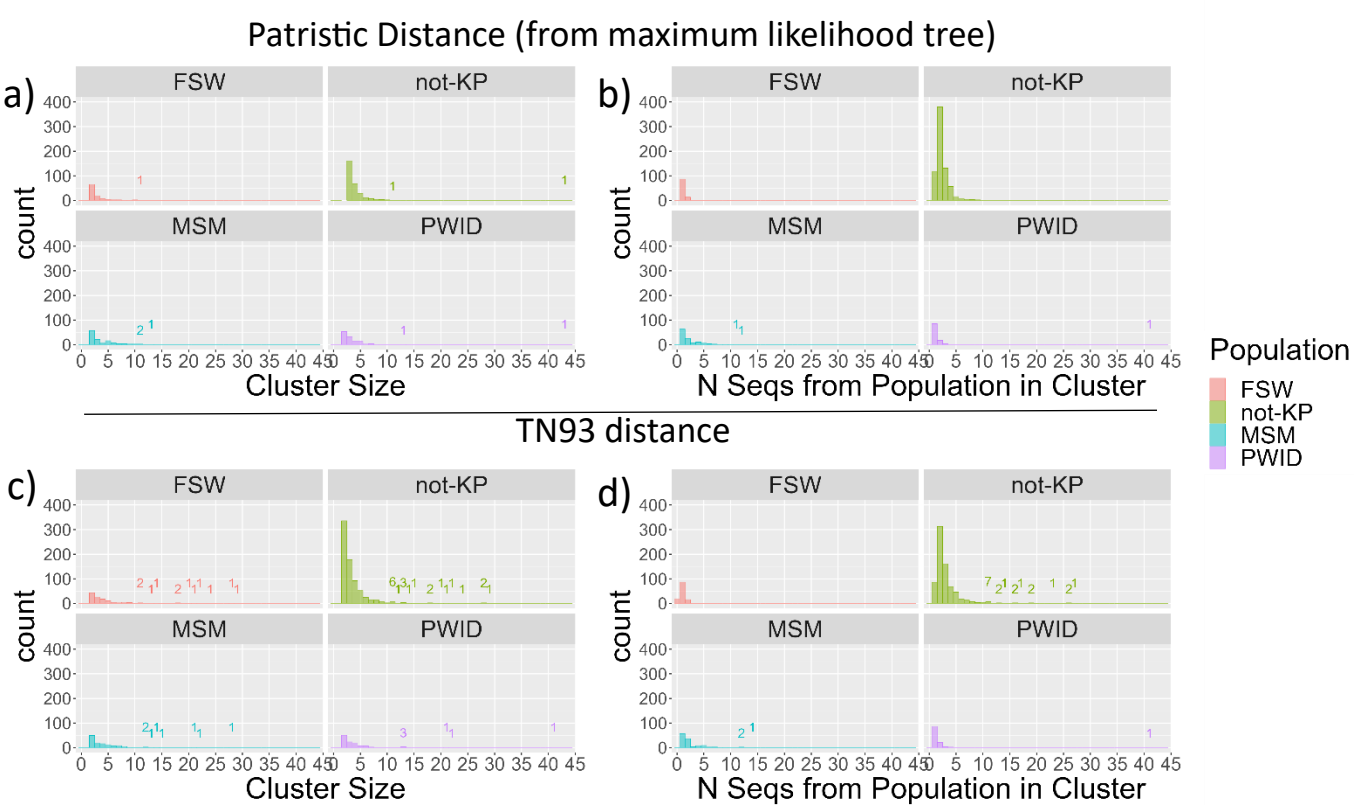
Supplementary Table S1.4. HIV-1 sequences from other populations in clusters with HIV-1 sequences from female PWID with transactional sex risk factors or in clusters with sequences from PWID who report sharing needles. Only clusters with at least 1 sequence from the coast or Nairobi are included. Custers are based on 0.045 maximum patristic distance threshold on maximum likelihood trees using subtypes A1, C, and D. Percents are based on the total number of clusters that contain at least 1 sequence from that population. Note: Clusters containing >2 population groups are represented more than once. Abbreviations: FSW: female sex workers, MSM: men who have sex with men, not-KP: not key population, PWID: people who injects drugs. See Supplementary Table S1.3 for comprehensive cluster counts.

Clusters containing sequences from female PWID with transactional sex risk factors* (N = 18)		Clusters containing sequences from PWID who shared needles last month (N=7)	
In cluster with	**Count**	**In cluster with**	
not-KP	14 (2.3%)	not-KP	5 (0.8%)
FSW	1 (1.2%)	FSW	0
MSM	3 (2.6%)	MSM	1 (0.9%)
PWID	4 (3.3%)	PWID	3 (2.5%)
PWID who shared needles	3 (42.9%)	PWID who shared needles	0

*Have ever received payment for sex and had ≥3 sex partners in the prior month



Supplementary Figure S1.2. Cluster distributions for population group and region. Each cluster is depicted as a vertical bar colored by the percent makeup of sequences from either population group (A&B) or region (C&D). Clusters are defined using a patristic distance threshold of 0.015 (A) or 0.045 (B-D). D is restricted to the 120 clusters that contain at least 1 sequence from a PWID (the darker-colored bars in C). Abbreviations: FSW: female sex workers, MSM: men who have sex with men, PWID: people who injects drugs, not-KP: not key population



Supplementary Figure S1.3. Cluster size distributions by population and cluster-definition. A&C) The size of clusters containing sequences from each population (clusters are represented in each plot for which they contain ≥1 sequence from that population group). B&D) The number of sequences from each population group in each cluster (counts sum to total number of all sequences in clusters). Clusters are defined using maximum genetic distance threshold 0.045 based on patristic distance (A,B) or TN93 distance (C,D). Counts are indicated for all clusters of size >10.

Supplementary Table S1.5. HIV-1 A1 subtype transitions between coastal and Nairobi regions in a tree restricted to sequences from PWID (2018-2021) using maximum likelihood tree building and ancestral state reconstruction. Discrete trait analysis is conducted under uniform subsampling (trees are down-sampled to have equal counts of coast and Nairobi sequences) and proportionate subsampling (trees are down sampled so counts are proportionate to the estimated number of HIV cases in each population). Counts are averaged across 10 subtrees and 20 resolutions of ancestor states. P-values test for disproportionate transitions in either direction, resampling traits on the tree tips 200 times to derive the null distribution.

	Sequence counts N (%)			TRANSITIONS N (% of ALL BRANCHES)		P
Subtype	**Total**	**Nairobi**	**Coast**	**Coast → Nairobi**	**Nairobi → Coast**	**(Coast->Nairobi vs. Nairobi -> coast) (2-tailed)**
Uniform subsampling						
A1	152 (100%)	76 (50%)	76 (50%)	50 (37-69) (16.6%)	46 (32-61) (15.2%)	0.05
Proportionate subsampling						
A1	116 (100%)	40 (34.5%)	76 (65.5%)	37 (31-40) (16.1%)	9 (6-18) (4.0%)	0.61

Supplementary Table S1.6. HIV-1 A1 subtype transitions between coastal and Nairobi regions using maximum likelihood tree building and ancestral state reconstruction (terminal branches only). Discrete trait analysis is conducted under uniform subsampling (trees are down-sampled to have equal counts of coast and Nairobi sequences). Counts are averaged across 10 subtrees and 20 resolutions of ancestor states for each tree. P-values test for disproportionate transitions in either direction, resampling traits on the tree tips 200 times to derive the null distribution.

	N sequences			TRANSITIONS: TERMINAL BRANCHES N (% OF ALL SAMPLED SEQUENCES)		P
Uniform Subsampling						
Subtype	**Total**	**Nairobi**	**Coast**	**Coast → Nairobi**	**Nairobi → Coast**	**Coast->Nairobi vs. Nairobi -> coast transitions (2-tailed)**
A1	1004 (100%)	502 (50%)	502 (50%)	176 (152-212) (17.6%)	147 (128-158) (14.6%)	<0.01

Supplementary Table S1.7. HIV-1 Transitions between population groups using maximum likelihood tree-building and ancestral state reconstruction (terminal branches only). 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.

	SEQUENCE COUNTS N (%)			TRANSITIONS / JUMPS: TERMINAL BRANCHES N (% OF ALL SAMPLED SEQUENCES)	
SUBTYPE	Total	Group 0	Group 1	0 -> 1	1 -> 0
		FSW	PWID	FSW-> PWID	PWID -> FSW
all	148 (50%)	74 (25%)	74 (25%)	15 (10-28) (5.0%)	9 (5-14) (6.1%)
A1	**116 (50%)**	**58 (25%)**	**58 (25%)**	**14 (10-25) (6.1%)**	**7 (5-10) (2.9%)**
C	14 (50%)	7 (25%)	7 (25%)	1 (0-2) (2.1%)	1 (0-2) (3.9%)
D	18 (50%)	9 (25%)	9 (25%)	0 (0-1) (0.6%)	1 (0-2) (3.1%)
		MSM	PWID	MSM -> PWID	PWID -> MSM
all	148 (50%)	74 (25%)	74 (25%)	9 (3-12) (6.1%)	8 (3-13) (5.4%)
A1	**116 (50%)**	**58 (25%)**	**58 (25%)**	**7 (3-9) (3.0%)**	**5 (3-6) (2.0%)**
C	14 (50%)	7 (25%)	7 (25%)	1 (0-2) (2.1%)	1 (0-3) (4.7%)
D	18 (50%)	9 (25%)	9 (25%)	1 (0-1) (1.4%)	2 (0-4) (5.3%)
		not-KP	PWID	not-KP -> PWID	PWID -> not-KP
all	148 (50%)	74 (25%)	74 (25%)	23 (0-8) (15.5%)	11 (5-16) (7.4%)
A1	**116 (50%)**	**58 (25%)**	**58 (25%)**	**20 (12-32) (8.6%)**	**7 (5-10) (3.0%)**
C	14 (50%)	7 (25%)	7 (25%)	2 (0-5) (6.6%)	2 (0-3) (5.8%)
D	18 (50%)	9 (25%)	9 (25%)	1 (0-3) (4.0%)	2 (0-3) (4.9%)
TOTAL		not-PWID	PWID	not-PWID -> PWID	PWID -> not-PWID
all	296 (100%)	222 (75%)	74 (25%)	47 (25-80) (15.9%)	28 (13-53) (9.5%)
A1	**232** (100%)	**174 (75%)**	**58 (25%)**	**41 (25-66) (17.7%)**	**19 (13-36) (7.9%)**
C	28 (100%)	21 (75%)	7 (25%)	4 (0-9) (14.3%)	4 (0-8) (14.3%)
D	36 (100%)	27 (75%)	9 (25%)	2 (0-5) (5.6%)	5 (0-9) (13.9%)

Supplementary Table S1.8. HIV-1 transitions between PWID and not-PWID populations from the coast and Nairobi using both maximum likelihood and Bayesian tree-building and ancestral state reconstruction models under uniform subsampling. 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 9 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 median Bayes factor (BF). This data is also presented in Figure 3. Supplementary Table S9 presents counts limited to terminal branches.

	SEQUENCE COUNTS N (% OF ALL SEQUENCES)			TRANSITIONS / JUMPS: FULL TREE N (% OF ALL BRANCHES)		
SUBTYPE	Total	Group 0	Group 1	0 -> 1	1 -> 0	
MAXIMUM LIKELIHOOD						P-value
All between group transitions						
All		440		489 of 873 total branches (56.0%)		
A1		**352**		**404 of 702 total branches (57.6%)**		
C		64		55 of 126 total branches (43.7%)		
D		24		30 of 45 total branches (66.7%)		
Between regions (not-PWID)						
		Coast not-PWID*	Nairobi not-PWID	Coast not-PWID -> Nairobi not-PWID	Nairobi not-PWID -> Coast not-PWID	Coast not-PWID -> Nairobi not-PWID vs. Nairobi not-PWID -> Coast not-PWID
All	220	110 (25%)	110 (25%)	64 (39-89) (7.3%)	57 (46-87) (6.5%)	
A1	**176**	**88 (25%)**	**88 (25%)**	**53 (33-71) (7.5%)**	**48 (22-71) (6.8%)**	**0.20**
C	32	16 (25%)	16 (25%)	8 (5-14) (6.4%)	6 (3-12) (4.8%)	0.04
D	12	6 (25%)	6 (25%)	3 (1-4) (6.2%)	3 (1-4) (5.7%)	0.93
Between regions (PWID)						
		Coast PWID	Nairobi PWID	Coast PWID -> Nairobi PWID	Nairobi PWID -> Coast PWID	Coast PWID -> Nairobi PWID vs. Nairobi PWID -> Coast PWID
All	220	110 (25%)	110 (25%)	22 (16-34) (2.5%)	33 (20-59) (4.1%)	
A1	**176**	**88 (25%)**	**88 (25%)**	**15 (12-20) (2.1%)**	**25 (16-46) (3.6%)**	**<0.01**
C	32	16 (25%)	16 (25%)	4 (2-9) (3.4%)	6 (3-10) (4.5%)	0.02
D	12	6 (25%)	6 (25%)	3 (2-5) (7.4%)	2 (1-3) (5.3%)	0.04
Between not-PWID and PWID populations (same region)						
		Coast not-PWID	Coast-PWID	Coast not-PWID -> Coast PWID	Coast PWID -> Coast not-PWID	Coast not-PWID -> Coast PWID vs. Coast PWID -> Coast not-PWID
All	220	110 (25%)	110 (25%)	45 (25-75) (5.6%)	25 (16-35) (2.9%)	
A1	**176**	**88 (25%)**	**88 (25%)**	**39 (24-60) (5.6%)**	**19 (14-24) (2.8%)**	**<0.01**
C	32	16 (25%)	16 (25%)	4 (1-11) (3.5%)	3 (2-7) (2.6%)	0.44
D	12	6 (25%)	6 (25%)	2 (0-4) (4.8%)	3 (0-4) (5.8%)	0.10

		Nairobi not-PWID	Nairobi-PWID	Nairobi not-PWID -> Nairobi PWID	Nairobi PWID -> Nairobi not-PWID	Nairobi not-PWID -> Nairobi PWID vs. Nairobi PWID -> Nairobi not-PWID
All	220	110 (25%)	110 (25%)	47 (20-81) (5.3%)	37 (23-59) (4.2%)	
A1	**176**	**88 (25%)**	**88 (25%)**	**41 (18-68) (5.8%)**	**31 (20-48) (4.5%)**	**0.02**
C	32	16 (25%)	16 (25%)	4 (1-8) (3.0%)	4 (2-8) (3.4%)	0.03
D	12	6 (25%)	6 (25%)	2 (1-5) (4.8%)	2 (1-3) (4.6%)	0.69
Between not-PWID and PWID populations (different region)						
		Coast not-PWID	Nairobi-PWID	Coast not-PWID -> Nairobi PWID	Nairobi PWID -> Coast not-PWID	Coast not-PWID -> Nairobi PWID vs Nairobi PWID -> Coast not-PWID
All	220	110 (25%)	110 (25%)	54 (26-90) (6.1%)	40 (18-59) (4.6%)	
A1	**176**	**88 (25%)**	**88 (25%)**	**46 (23-74) (6.6%)**	**33 (15-49) (4.7%)**	**<0.01**
C	32	16 (25%)	16 (25%)	5 (1-11) (4.2%)	5 (2-7) (3.6%)	0.41
D	12	6 (25%)	6 (25%)	3 (2-5) (7.4%)	2 (1-3) (4.5%)	0.04
		Nairobi not-PWID	Coast-PWID	Nairobi not-PWID -> Coast PWID	Coast PWID -> Nairobi not-PWID	Nairobi not-PWID -> Coast PWID vs. Coast PWID -> Nairobi not-PWID
All	220	110 (25%)	110 (25%)	40 (21-64) (4.6%)	25 (16-36) (2.9%)	
A1	**176**	**88 (25%)**	**88 (25%)**	**35 (19-54) (4.9%)**	**19 (14-24) (2.7%)**	**<0.01**
C	32	16 (25%)	16 (25%)	3 (1-7) (2.8%)	3 (1-4) (2.1%)	0.50
D	12	6 (25%)	6 (25%)	2 (1-3) (4.1%)	3 (1-4) (5.8%)	0.06
BAYESIAN						Bayes Factor
All between group transitions						
A1		702		315 of 702 total branches (44.9%)		
Between regions (not-PWID)						
		Coast not-PWID	Nairobi not-PWID	Coast not-PWID -> Nairobi not-PWID	Nairobi not-PWID -> Coast not-PWID	Coast not-PWID -> Nairobi not-PWID	Nairobi not-PWID -> Coast not-PWID
A1	176	88 (25%)	88 (25%)	118 (52-139) (16.9%)	10 (6-28) (1.4%)	**>100**	**9.8**
Between regions (PWID)							
		Coast PWID	Nairobi PWID	Coast PWID -> Nairobi PWID	Nairobi PWID -> Coast PWID	Coast PWID -> Nairobi PWID	Nairobi PWID -> Coast PWID
A1	176	88 (25%)	88 (25%)	24 (7-40) (3.4%)	13 (5-29) (1.8%)	**8.4**	0.4
Between not-PWID and PWID populations (same region)							
		Coast not-PWID	Coast-PWID	Coast not-PWID -> Coast PWID	Coast PWID -> Coast not-PWID	Coast not-PWID -> Coast PWID	Coast PWID -> Coast not-PWID
A1	176	88 (25%)	88 (25%)	23 (9-56) (3.2%)	3 (2-5) (0.5%)	1.6	0.1
		Nairobi not-PWID	Nairobi-PWID	Nairobi not-PWID -> Nairobi PWID	Nairobi PWID -> Nairobi not-PWID	Nairobi not-PWID -> Nairobi PWID	Nairobi PWID -> Nairobi not-PWID
A1	176	88 (25%)	88 (25%)	44 (3-68) (6.2%)	3 (2-5) (0.5%)	0.5	0.2
Between not-PWID and PWID populations (different region)							
		Coast not-PWID	Nairobi-PWID	Coast not-PWID -> Nairobi PWID	Nairobi PWID -> Coast not-PWID	Coast not-PWID -> Nairobi PWID	Nairobi PWID -> Coast not-PWID
A1	176	88 (25%)	88 (25%)	15 (6-63) (2.1%)	2 (1-2) (0.3%)	0.3	0.2

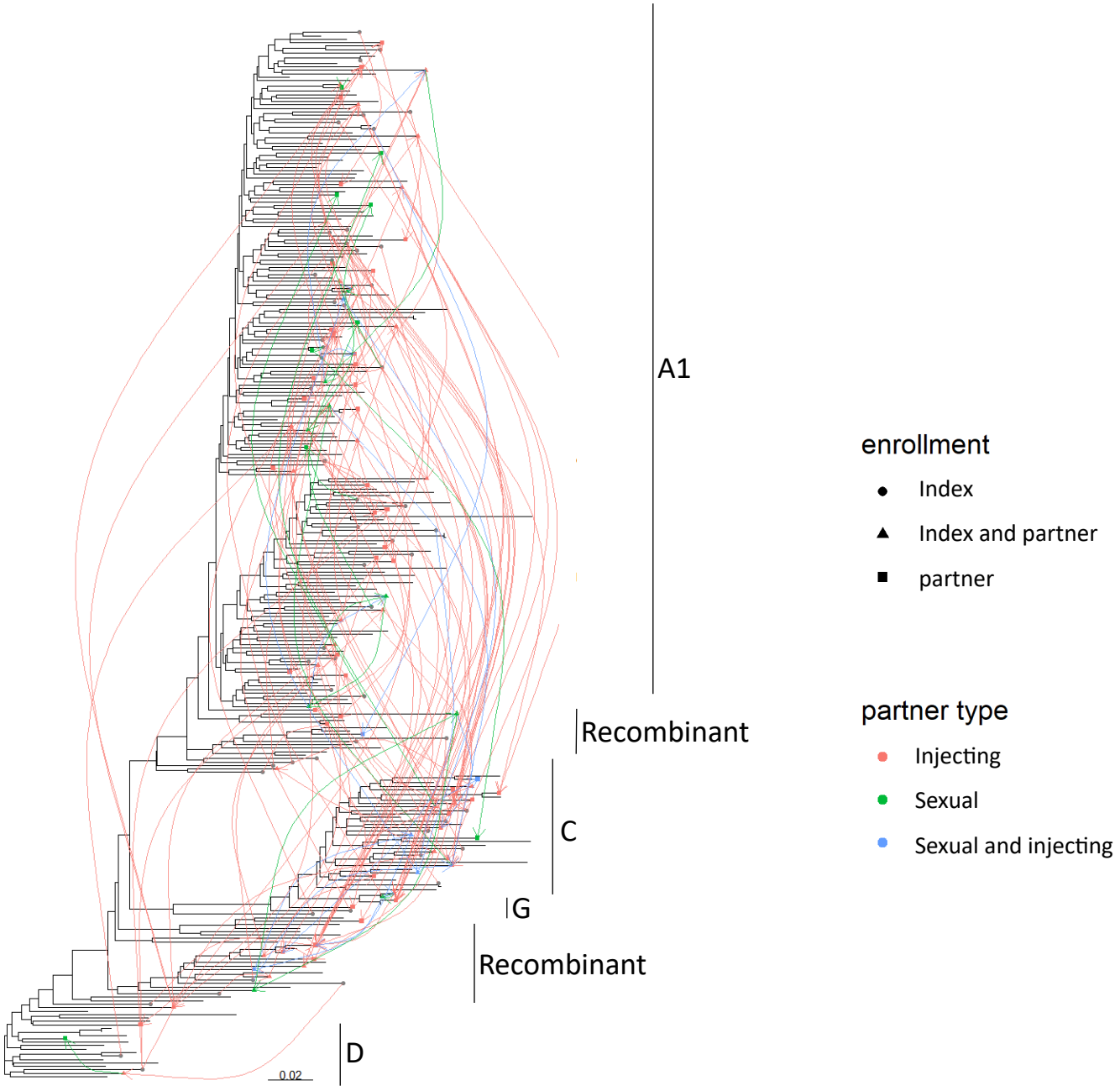
		Nairobi not-PWID	Coast-PWID	Nairobi not-PWID -> Coast PWID	Coast PWID -> Nairobi not-PWID	Nairobi not-PWID -> Coast PWID	Coast PWID -> Nairobi not-PWID
A1	176	88 (25%)	88 (25%)	63 (23-82) (9.0%)	7 (3-10) (0.9%)	**16.8**	0.1

*The not-PWID population is sampled from not-KP, FSW, and MSM sequences proportionate to their population sizes, with the majority of sequences coming from the not-KP population.

Supplementary Table S1.9. Transitions between PWID and not-PWID populations from coastal Kenya and Nairobi using maximum likelihood tree-building and ancestral state reconstruction methods. Transition counts are reported for terminal branches only (ie. branches leading to the observed sequences). 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.

	SEQUENCE COUNTS N (% OF ALL SEQUENCES)			TRANSITIONS / JUMPS: TERMINAL BRANCHES N (% OF SAMPLED SEQEUCNES)	
Subtype	Total	Group 0	Group 1	0 -> 1	1 -> 0
Maximum Likelihood					
Between regions (not-PWID)					
		Coast not-PWID	Nairobi not-PWID	Coast not-PWID -> Nairobi not-PWID	Nairobi not-PWID -> Coast not-PWID
all	232 (50%)	110 (25%)	110 (25%)	32 (19-51) (7.2%)	24 (12-41) (5.4%))
A1	**176 (50%)**	**88 (25%)**	**88 (25%)**	**25 (15-39) (7.1%)**	**20 (11-32) (5.8%)**
C	32 (50%)	16 (25%)	16 (25%)	5 (3-9) (7.4%)	3 (1-6) (4.5%)
D	24 (50%)	6 (25%)	6 (25%)	2 (1-3) (6.8%)	1 (0-3) (5.8%)
Between regions (PWID)					
		Coast PWID	Nairobi PWID	Coast PWID -> Nairobi PWID	Nairobi PWID -> Coast PWID
all	232 (50%)	110 (25%)	110 (25%)	12 (7-18) (2.8%)	21 (12-38) (4.8%)
A1	**176 (50%)**	**88 (25%)**	**88 (25%)**	**9 (6-11) (2.5%)**	**18 (11-33) (5.0%)**
C	32 (50%)	16 (25%)	16 (25%)	2 (1-4) (2.9%)	2 (1-3) (3.0%)
D	24 (50%)	6 (25%)	6 (25%)	1 (0-3) (5.1%)	1 (0-2) (4.8%)
Between not-PWID and PWID populations (same region)					
		Coast not-PWID	Coast-PWID	Coast not-PWID -> Coast PWID	Coast PWID -> Coast not-PWID
all	232 (50%)	110 (25%)	110 (25%)	28 (14-43) (6.3%)	11 (6-16) (2.5%)
A1	**176 (50%)**	**88 (25%)**	**88 (25%)**	**25 (14-37) (7.0%)**	**9 (5-12) (2.5%)**
C	32 (50%)	16 (25%)	16 (25%)	2 (0-5) (3.0%)	1 (1-2) (2.2%)
D	24 (50%)	6 (25%)	6 (25%)	1 (0-1) (3.4%)	1 (0-2) (4.4%)
		Nairobi not-PWID	Nairobi-PWID	Nairobi not-PWID -> Nairobi PWID	Nairobi PWID -> Nairobi not-PWID
all	232 (50%)	110 (25%)	110 (25%)	27 (10-52) (6.1%)	18 (9-34) (4.1%)
A1	**176 (50%)**	**88 (25%)**	**88 (25%)**	**23 (9-44) (6.5%)**	**15 (8-28) (4.2%)**
C	32 (50%)	16 (25%)	16 (25%)	2 (0-5) (2.9%)	2 (1-4) (2.8%)
D	24 (50%)	6 (25%)	6 (25%)	2 (1-3) (8.1%)	1 (0-2) (3.9%)

Between not-PWID and PWID populations (different region)					
		Coast not-PWID	Nairobi-PWID	Coast not-PWID -> Nairobi PWID	Nairobi PWID -> Coast not-PWID
all	232 (50%)	110 (25%)	110 (25%)	29 (12-50) (6.6%)	17 (9-30) (3.9%)
A1	**176 (50%)**	**88 (25%)**	**88 (25%)**	**25 (11-41) (7.0%)**	**14 (8-26) (3.9%)**
C	32 (50%)	16 (25%)	16 (25%)	2 (0-6) (3.7%)	2 (1-3) (2.5%)
D	24 (50%)	6 (25%)	6 (25%)	2 (1-3) (8.1%)	1 (0-2) (3.9%)
		Nairobi not-PWID	Coast-PWID	Nairobi not-PWID -> Coast PWID	Coast PWID -> Nairobi not-PWID
all	232 (50%)	110 (25%)	110 (25%)	25 (13-41) (5.7%)	12 (7-18) (2.8%)
A1	**176 (50%)**	**88 (25%)**	**88 (25%)**	**22 (13-36) (6.2%)**	**10 (7-13) (2.8%)**
C	32 (50%)	16 (25%)	16 (25%)	2 (0-4) (3.0%)	1 (0-3) (1.8%)
D	24 (50%)	6 (25%)	6 (25%)	1 (0-1) (3.1%)	1 (0-2) (5.2%)

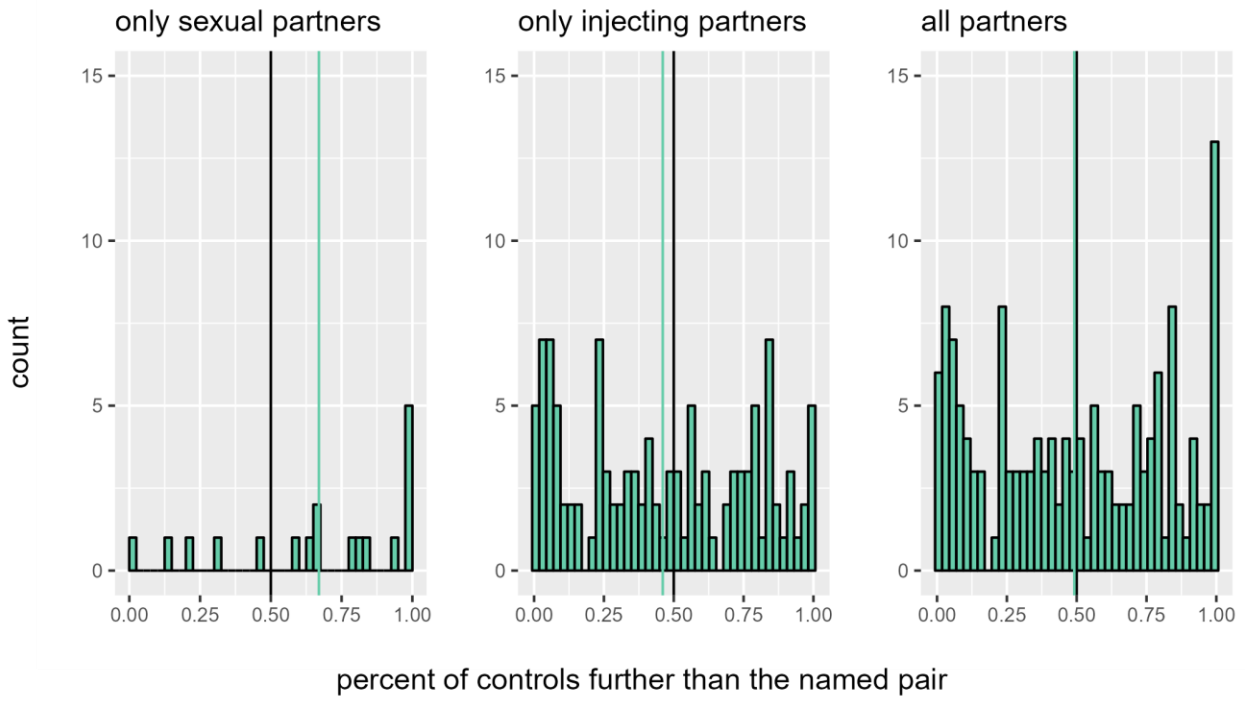


Supplementary Figure S2.1. Phylogeny of index-partner pair connections among HIV-1 sequence data. Arrows are drawn from the index to the partner they identified and colored by the partner-type. Additional Kenyan reference sequences are incorporated in phylogeny-based analyses (excluded from this image for clarity).

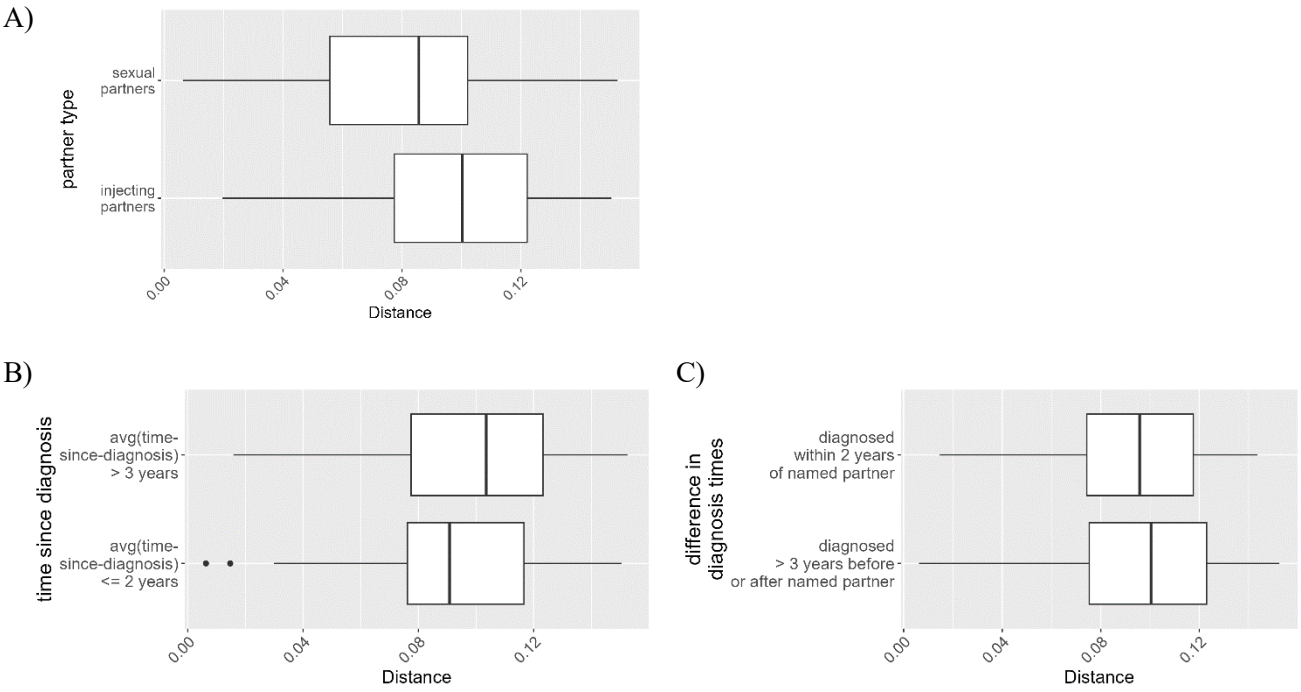
Supplementary Table S2.1. Distribution of HIV-1 patristic distances across index-partner pairs stratified by partner-type. The analysis shown in Table 2 is reproduced using molecular distances extracted from the HIV phylogeny instead of TN93 distances. The *p* value comes from the null distribution of the test statistic (Avg(% controls further than named pairs)) under 2,000 resamplings of the partner for each index.

		Sexual only		**Injecting only**		**Injecting and shared needles or equipment**		**All (Sexual, Injecting, and Sexual & Injecting)**	
		named pairs	controls	named pairs	controls	named pairs	controls	named pairs	controls
total	**distance threshold**	**18**	**2636**	**116**	**18577**	**13**	**2130**	**150**	**23630**
Pairs < distance threshold N (%)	0.015	2 (11.1%)	1 (0.0004%)	0 (0%)	6 (0.0003%)	0 (0%)	0 (0%)	4 (2.7%)	3 (0.0001%)
	0.045	4 (22.2%)	3 (0.1%)	2 (1.7%)	23 (0.1%)	0 (0%)	3 (0.1%)	9 (6%)	26 (0.1%)
	0.1	6 (33.3%)	304 (11.5%)	16 (13.8%)	2920 (16.7%)	0 (0%)	83 (3.9%)	25 (16.7%)	3488 (14.8%)
	0.15	13 (72.2%)	1189 (45.1%)	48 (41.4%)	8961 (48.2%)	2 (15.4%)	473 (22.2%)	67 (44.7%)	11037 (46.7%)
	0.2	13 (72.2%)	1605 (60.9%)	67 (57.8%)	12012 (64.7%)	3 (23.1%)	814 (58.0%)	87 (58%)	14934 (63.2%)
mean distance		0.13	0.18	0.19	0.17	0.24	0.21	0.18	0.18
Avg (% control pairs further than named pairs)[a]		0.71	-	0.45	-	0.33		0.48	-
***p* value**		**0.001**	-	0.06	-	0.025		0.34	-

[a]Percent of index-control pairs with HIV-1 patristic distances > the named index-partner pair, averaged across all pairs. Percents >50% indicate that HIV sequences from indexes are more similar, on average, to their identified partner than to sequences from the same region.



Supplementary Figure S2.2. Percent of index-control pairs with higher sequence molecular distance than named partner pairs. The percentage of index-control pairs that have less similar HIV sequences (based on TN93 distance) than the named index-partner pair. Controls consist of SHARP-study participant HIV sequences from the same region. Black lines indicate the null (50%), teal lines indicate the observed means (see Avg(% named pairs closer than control pairs) from Table 2). If named partner pairs have excess sequence similarity, the plots will be skewed towards 1.



Supplementary Figure S2.3. TN93 distance by A) partner type and B&C) time-since diagnosis. B) Partners with mean(time-since-diagnosis) >3 years (N = 64) vs. partners with mean(time-since-diagnosis) ≤2 years (N = 68). C) Partners with >3 years separating their diagnosis date (N = 72) to partners with ≤2 years separating their diagnosis time (N = 57). Boxes cover the 1st and 3rd quartiles & whiskers cover the range of values. Outliers are values >1.5X inter-quartile range.

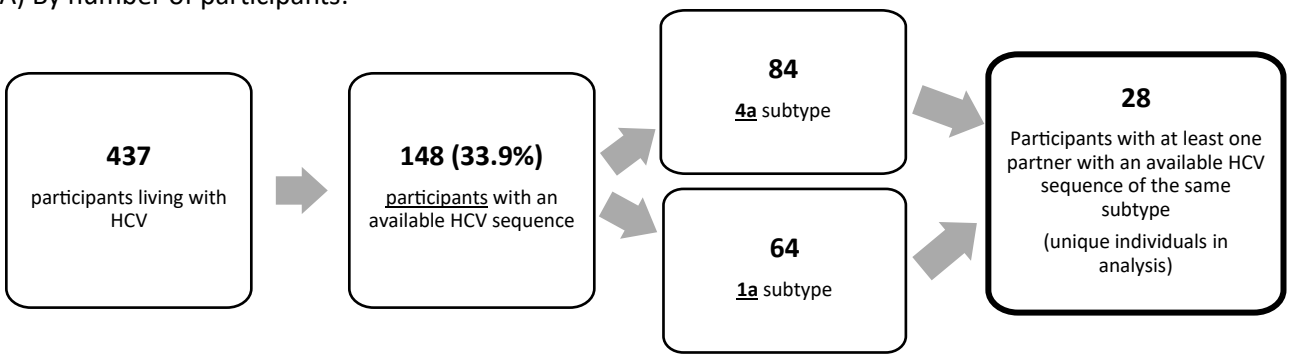
Supplementary Table S2.2. Associations between partner-pair characteristics and HIV-1 TN93 distance, using various thresholds.

Pair characteristic	values	mean (continuous) or proportion (categorical)	aOR (95% CI) (controlling for region)		
Sequence distance:			0.045	0.10	0.15
Partner type (N = 134) (excludes those identified as both sexual and injecting partners)	Ref: Injecting only	116 (86.6%)	Ref	Ref	Ref
	Sexual only	18 (13.4%)	**27.47 (4.35-239.36)**	2.63 (0.91-8.77)	0.42 (0.04-9.75)
Years since diagnosis (N =150) averaged between index and partner (diagnoses prior to enrollment are self-reported)	-	3.66	0.85 (0.61-1.07)	**0.9 (0.81-1.00)**	0.94 (0.69-1.40)
Number of years separating index-partner diagnoses[a]	-	3.98	0.87 (0.65-1.07)	0.96 (0.88-1.05)	0.75 (0.53-1.01)
Positive HIV diagnosis prior to enrollment (N =150)	Ref: both previously tested positive	89 (59.3%)	Ref	Ref	Ref
	one partner previously tested positive	54 (36%)	3.56 (0.89-17.49)	1.14 (0.58-2.25)	0.28 (0.01-3.04)
	neither partner previously tested positive	7 (4.7%)	-	0.72 (0.14-3.46)	-
ART use at enrollment (N = 150)	Ref: Both partners on ART at enrollment	37 (24.7%)	Ref	Ref	Ref
	One partner on ART at enrollment	65 (43.3%)	2.19 (0.3-44.34)	1.35 (0.59-3.12)	0.68 (0.03-7.78)
	Neither partner on ART at enrollment	48 (32.0%)	2.85 (0.36-60.01)	0.69 (0.28-1.73)	-
Shared needles last month (N = 148)	Ref: Neither partner shared needles	133 (89.9%)	Ref	Ref	Ref
	One or both partners shared needles	15 (10.1%)	-	2.12 (0.71-7.15)	-
Shared needle or injecting equipment with named partner[b] (injecting partners only; N = 132)	Ref: Never with partner	119 (90.2%)	Ref	Ref	Ref
	Have shared with partner	13 (9.8%)	-	0.44 (0.11-1.44)	-
Ever received money or goods for sex (N = 150)	Ref: Neither partner received money or goods for sex	18 (12%)	Ref	Ref	Ref
	One partner received money or goods for sex	66 (44%)	0.77 (0.16-5.61)	1.17 (0.4-3.36)	-
	Both partners received money or goods for sex	66 (44%)	0.12 (0.01-1.32)	0.59 (0.20-1.7)	-
Viral suppression (N = 145) (<1,000 copies/ml at enrollment)	Ref: one or both partners virally suppressed	47 (32.4%)	Ref	Ref	Ref
	neither partner virally suppressed	98 (67.6%)	0.79 (0.18-3.98)	**2.46 (1.21-5.13)**	1.04 (0.05-11.28)

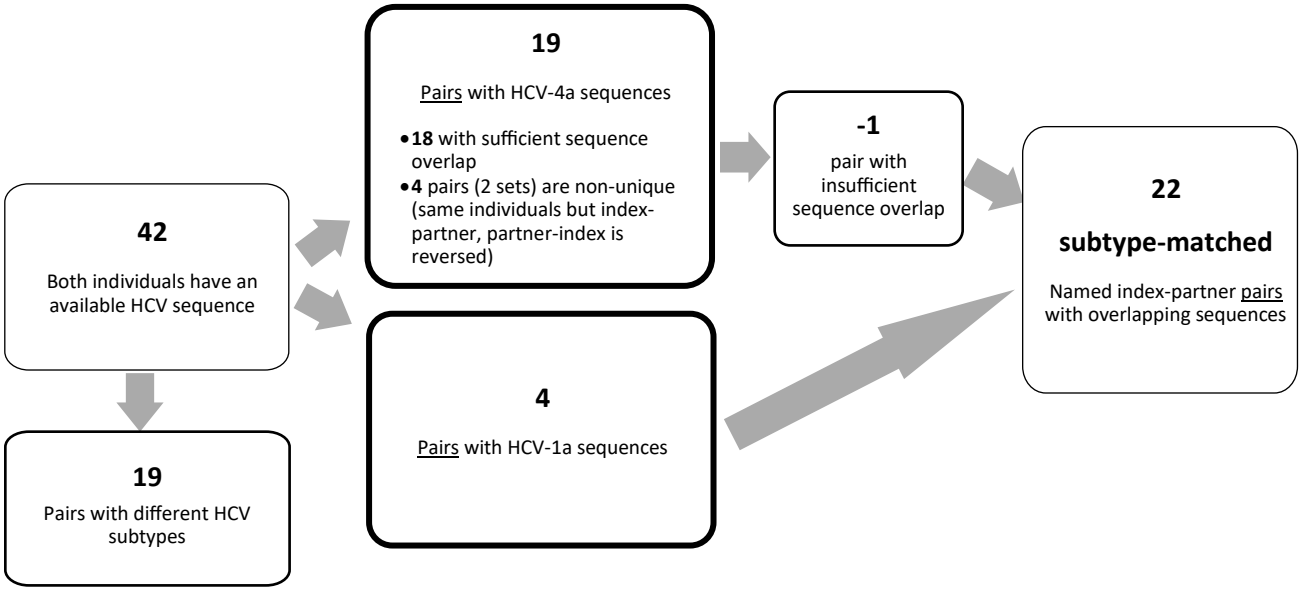
[a]Diagnosis times prior to enrollment are self-reported

[b]Based on reporting by index

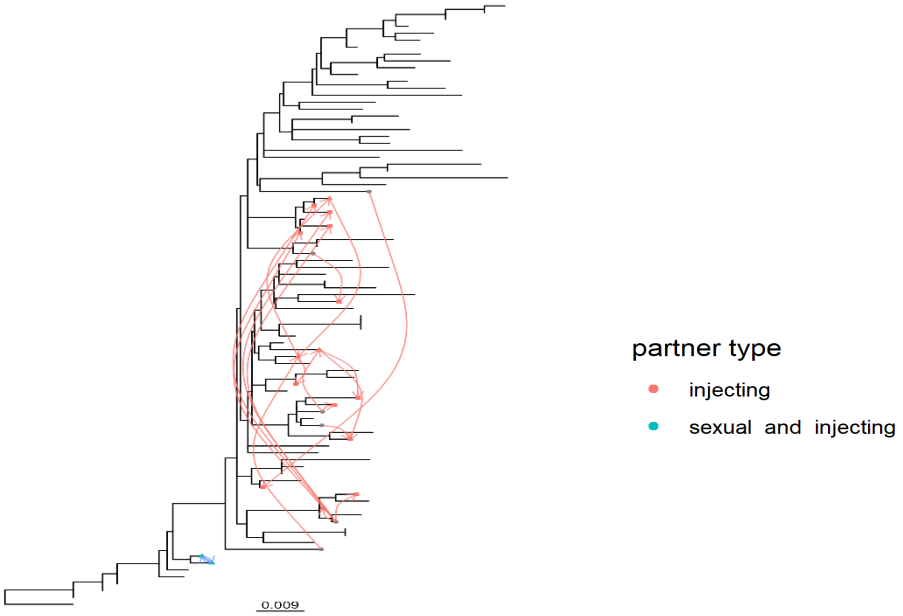
A) By number of participants:



B) By number of pairs:



Supplementary Figure S2.4. Cascades of participant and partner pairs counts for HCV analysis A) Counts of participants, culminating in the number of unique participants included in the HCV analysis and stratified by enrollment type. B) Counts of named partner pairs, culminating in the number of pairs included in the HCV analysis and stratified by pair type. Some individuals are represented in multiple partner pairs.



Supplementary Figure S2.5. Phylogeny of index-partner pair connections among HCV 4a sequence data. Arrows are drawn from the index to the partner they identified and colored by the partner-type.

Supplementary Table S2.3. Distribution of HCV molecular distances across index-partner pairs, stratified by subtype. The analysis in Table 4 is reproduced using molecular distances from an HCV phylogeny. The *p* value comes from the null distribution of the test statistic (Avg(% controls further than named pairs)) under 2,000 resamplings of the partner for each index.

		4a		**1a**	
		named pairs	controls	named pairs	controls
total	**distance threshold**	**19**	**1218**	**4**	**136**
Pairs < distance threshold N (%)	0.005	1 (5.3%)	1 (0.1%)	0	0
	0.01	3 (15.8%)	17 (1.4%)	0	1 (0.7%)
	0.025	7 (36.8%)	137 (11.2%)	0	7 (5.1%)
	0.04	19 (100%)	1107 (90.9%)	1 (25%)	82 (60.3%)
	0.10	19 (100%)	1218 (100%)	3 (75%)	121 (89%)
mean distance		0.023	0.037	0.062	0.055
Avg (% control pairs further than named pairs)[a]		0.81	-	0.43	
***p* value**		**<0.001**	-	0.68	

[a]Percent of index-control pairs with HCV molecular distance > the named index-partner pair, averaged across all pairs. Percents >50% indicate that HCV sequences from indexes are more similar, on average, to their identified partner than to other individuals from the same region.