AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 33, Number 5, 2017

DOI: 10.1089/aid.2016.0321

# Complex Subtype Diversity of HIV-1 Among Drug Users in Major Kenyan Cities

Kamini Gounder,[1,2] Micah Oyaro,[3] Nagavelli Padayachi,[1,2] Thando Mbali Zulu,[1,2] Tulio de Oliveira,[4,5] John Wylie,[6] and Thumbi Ndung'u[1,2,7,8]

## Abstract

Drug users are increasingly recognized as a key population driving human immunodeficiency virus (HIV) spread in sub-Saharan Africa. To determine HIV-1 subtypes circulating in this population group and explore possible geographic differences, we analyzed HIV-1 sequences among drug users from Nairobi, Mombasa, and Kisumu in Kenya. We sequenced *gag* and *env* from 55 drug users. Subtype analysis from 220 *gag* clonal sequences from 54 of 55 participants (median = 4/participant) showed that 44.4% were A, 16.7% were C, 3.7% were D, and 35.2% were intersubtype recombinants. Of 156 *env* clonal sequences from 48 of 55 subjects (median = 3/participant), 45.8% were subtype A, 14.6% were C, 6.3% were D, and 33.3% were recombinants. Comparative analysis of both genes showed that 30 (63.8%) participants had concordant subtypes, while 17 (36.2%) were discordant. We identified one genetically linked transmission pair and two cases of dual infection. These data are indicative of extensive HIV-1 intersubtype recombination in Kenya and suggest decline in subtype D prevalence.

**Keywords:** drug users, HIV-1 subtypes, intersubtype recombinants, Kenya

## Introduction

It is estimated that ~60 million people have been infected with the human immunodeficiency virus type 1 (HIV-1) and 25 million people have died of HIV-1-related causes since the beginning of the epidemic. A hallmark of the HIV-1 pandemic is the genetic diversity of viruses globally. HIV-1 sequences cluster into distinct subtypes or clades, and studies have shown that subtypes are unevenly distributed around the world, but the underlying reasons for the distinct geographical distribution have not been fully resolved.[1,2] Sub-Saharan Africa bears the brunt of the epidemic, with ~70% of all infections, and the extensive HIV-1 subtype diversity within the continent may pose an impediment to vaccine development and antiretroviral drug efficacy.[1,3,4] Furthermore, there is evidence that disease progression rate and HIV-1 epidemic spread may differ within a geographic area according to subtypes,[5,6] and monitoring changing subtype dynamics in a population may therefore be important for clinical interventions among those infected and for population-level strategies to curtail spread of the epidemic.

In many countries around the world, the HIV/AIDS epidemic is centered among particular high-risk groups, particularly young women, sex workers, men who have sex with men, and injecting drug users (IDUs). Kenya is among the countries of sub-Saharan Africa severely affected by HIV/AIDS and in particular, studies have shown that HIV prevalence among IDUs, at 52% and 15% in Nairobi and Mombasa, respectively, is much higher than in the general population.[7] If countries such as Kenya are to make further progress in reducing HIV-1 incidence, it is imperative that efforts be devoted to key drivers of the epidemic such as IDUs. Genetic studies of viruses circulating in key populations may offer insights into the effectiveness of public health

[1]HIV Pathogenesis Programme, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa.
[2]Africa Health Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa.
[3]Immunology Unit, Department of Human Pathology, University of Nairobi, Nairobi, Kenya.
[4]School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa.
[5]Africa Centre for Population Health, University of KwaZulu-Natal, Durban, South Africa.
[6]Department of Medical Microbiology, University of Manitoba, Winnipeg, Canada.
[7]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Boston, Massachusetts.
[8]Max Planck Institute for Infection Biology, Berlin, Germany.

interventions and provide information on pathogenesis and design of novel strategies to treat and prevent HIV-1 transmission. The main objective of this study was to determine the HIV-1 subtype diversity among a key population group of drug users in Kenya. We generated 376 full-length HIV-1 *gag* and *env* clonal sequences from 55 drug users, identified from the three largest Kenyan cities of Kisumu, Nairobi, and Mombasa. Our data reveal a complex picture of HIV-1 diversity in this population, suggestive of an active, ongoing epidemic with multiple infections and recombination.

Between January 2011 and July 2012, drug users from Kisumu, Mombasa, and Nairobi, the three largest cities in Kenya, were recruited into a larger cross-sectional study investigating the social networks, status, and molecular epidemiology of HIV, hepatitis B virus (HBV), and hepatitis C virus (HCV) among drug users in Kenya. Snowball sampling was used to recruit a total of 673 drug users, both IDUs and noninjecting drugs users (NIDUs). Study participants were males or females, 18 years or older, who reported recreational drug use, and had engaged in sexual activity at least once monthly over the past 3 months before recruitment and willing to provide blood samples for HIV testing. Participants who met eligibility criteria were administered a study questionnaire that captured demographic information and history of drug abuse, sexual behavior, employment status, healthcare access, sexual identity, and other risk factors for HIV infection. Participants also received HIV prevention counseling. Written informed consent was obtained from participants. The study protocol was approved by the Biomedical Research Ethics Committee of the University of KwaZulu-Natal, Kenyatta National Hospital/University of Nairobi Ethics and Research Committee (KNH/UoN-ERC), and the Health Research Ethics Board of the University of Manitoba. Those who consented and agreed to provide blood were sampled at identified sample collection centers. Both plasma and serum specimens were aliquoted into two and stored frozen at −80°C until sent to laboratories for testing for sexually transmitted infections (STIs), HIV, HBV, and HCV. HIV infection status was determined using the fourth generation enzyme-linked immunosorbent assay (ELISA) kit *Vironostika HIV* Uni-Form II *Ag-Ab* (Biomerieux) according to the manufacturer's instructions.

The QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) was used to isolate viral RNA from plasma samples. For the *gag* polymerase chain reaction (PCR), RNA was used to synthesize complementary DNA (cDNA) using the Thermoscript reverse transcription kit (Invitrogen, Life Technologies, Foster City, CA) and the specific primer, GagD reverse (5′-AAT TCC TCC TAT CAT TTT TGG-3′). The Expand High Fidelity PCR System (Roche) was used to amplify the ∼1.6 kb *gag* gene using a nested PCR reaction. The first-round PCR primers were GagD forward (5′-TCT CTA GCA GTG GCG CCC G-3′) and GagD reverse. The second-round PCR primers were GagA forward (5′-CTC TCG ACG CAG GAC TCG GCT T-3′) and GagC reverse (5′-TCT TCT AAT ACT GTA TCA TCT GC-3′). The resulting PCR product was then purified using the Illustra™ GFX™ PCR DNA Gel Band Purification Kit (GE Healthcare, Pittsburgh, PA) and cloned into the PCR 2.1-TOPO vector (Invitrogen). Following transformation, plasmid DNA (GeneJet Plasmid Mini Prep Kit; Thermo Fisher Scientific) was isolated from white colonies on plates containing *Xgal* and screened for the presence of the insert by digestion with the *Eco*RI restriction enzyme (New England BioLabs, Ipswich, MA).

For the *env* PCR, cDNA was synthesized using the SuperScript III protocol according to the manufacturer's instructions (Invitrogen) with the specific primer, OFM19 (5′-GCACTCAAGGCAAGCTTTATTGAGGCTTA-3′). The first-round PCR was carried out using the High Fidelity Platinum enzyme (Invitrogen) and the external primers used were VIF1 (5′-GGGTTTATTACAGGGACAGCAGAG-3′) and OFM19 5′-GCACTCAAGGCAAGCTTTATTGAGGC TTA-3′). Phusion High Fidelity polymerase was used for the second-round PCR with the primers ENV A (5′-GCTTAG GCATCTCCTATGGCAGGAAGAA-3′) and ENV N (5′-CTGCCAATCAGGGAAGTAGCCTTGTGT-3′).[8] The final PCR product (∼3.2 kb) was ligated into the pcDNA3.1.V5-His TOPO TA vector according to the manufacturer's instructions (Invitrogen). Plasmid DNA was isolated from individual randomly picked white bacterial colonies and digested with the *Hin*dIII and *Not*I restriction enzymes (New England BioLabs) to screen for the presence of the insert.

For single genome amplification (SGA) of the full-length *gag* gene, a limiting endpoint dilution was performed on cDNA synthesized for one participant (PID 426) to confirm the possibility of dual infection suggested by the bulk clonal sequences. SGA was performed as previously described.[9] Briefly, cDNA was serially diluted in replicates of eight PCR wells and *gag* PCR was performed as described above. cDNA dilutions were tested to identify a dilution where ∼30% of wells were positive for amplification products. All products derived from this dilution were sequenced directly.

*Gag* and *env* sequencing were performed using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction kit version 3.4 (Applied Biosystems, Foster City, CA). Sequences were assembled and edited using Sequencher Program v8.1.2 (Gene Codes Corporation) and aligned using

Table 1. Characteristics of Study Participants

| *Characteristics* | *No. of subjects (%)* |
|---|---|
| Sex | |
| Male | 43 (78.2) |
| Female | 12 (21.8) |
| Median age (IQR) | 30 (26–34) |
| Self-reported drug use category | |
| NIDU | 38 (69.1) |
| IDU | 17 (30.9) |
| Serology | |
| HIV | 55 (100) |
| HBV | 6 (10.9) |
| HCV | 15 (27.3) |
| Study sites | |
| Nairobi | 10 (18.2) |
| Kisumu | 17 (30.9) |
| Mombasa | 28 (50.9) |
| Employed | |
| Nairobi | 4 (40.0) |
| Kisumu | 15 (88.2) |
| Mombasa | 19 (67.9) |

HBV, hepatitis B virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; IDU, injecting drug use; IQR, interquartile range; NIDU, noninjecting drug use.

ClustalW. Maximum likelihood phylogenetic trees were constructed as implemented by Geneious v.8.1.6 to evaluate clustering of sequences with each other as well as with subtype reference strains obtained from the Los Alamos HIV sequence database (www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html). The maximum likelihood tree topology was inferred in phyML with a general time-reversible (GTR) substitution model (gamma distribution). Branching topology was visualized in FigTree (http://tree.bio.ed.ac.uk/software/figtree). Intralaboratory plasmid contamination was assessed using different methods: first, by using neighbor-joining phylogenetic trees to confirm phylogenetic clustering within and between participants. Second, sequences were blasted against sequences present in GenBank using the BLAST tool (GenBank, National Centre for Biotechnology Information). Finally, ViroBlast was used to compare all sequences to HIV sequences derived in our laboratory (http://indra.mullins.microbiol.washington.edu/viroblast/viroblast.php). Any clones that were nearly identical against local and published databases were removed from analysis.

The REGA HIV-1 subtyping Tool Version 2.0 (http://bioafrica.net/regagenotype/html/subtypinghiv.html) and the Recombinant Identification Program, RIP version 3 (www.hiv.lanl.gov/content/sequence/RIP/RIP.html) were used to infer subtypes. Subtype breakpoints were confirmed using Simplot v3.5. Using the predicted recombination breakpoints,

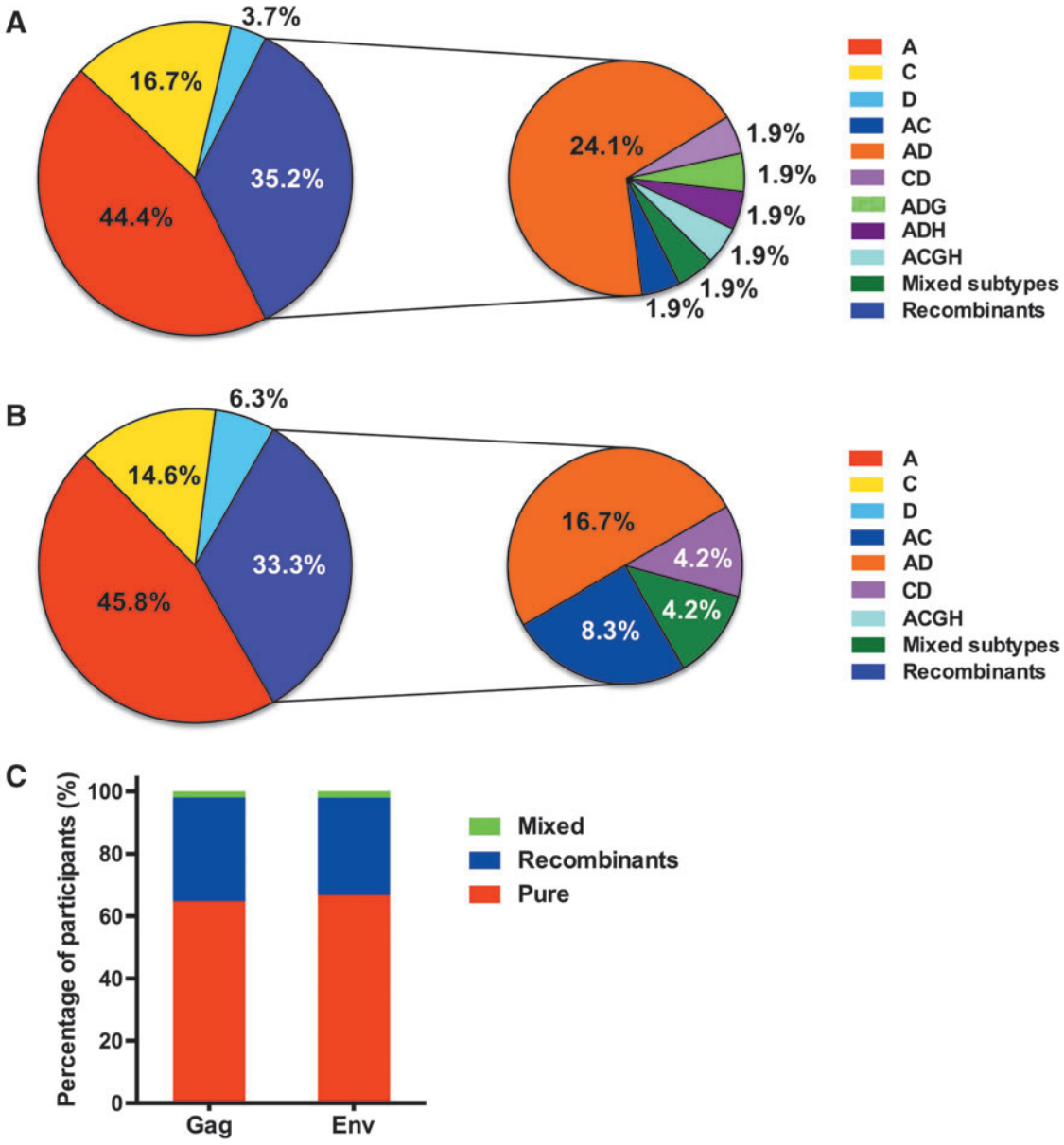


**FIG. 1.** Relative proportions of pure subtypes and intersubtype recombinants among drug users in Kenya. **(A)** Proportion of participants infected with pure and recombinant subtypes together with the breakdown of the relative recombinant percentages based on HIV-1 *gag* sequences. **(B)** Proportion of participants infected with pure and recombinant subtypes together with the breakdown of the relative recombinant percentages based on HIV-1 *env* sequences and **(C)** relative proportions of participants containing pure subtypes, recombinants, or a mixture of both pure and recombinant subtypes for either the *gag* or *env* region of the HIV-1 genome. HIV-1, human immunodeficiency virus type 1.

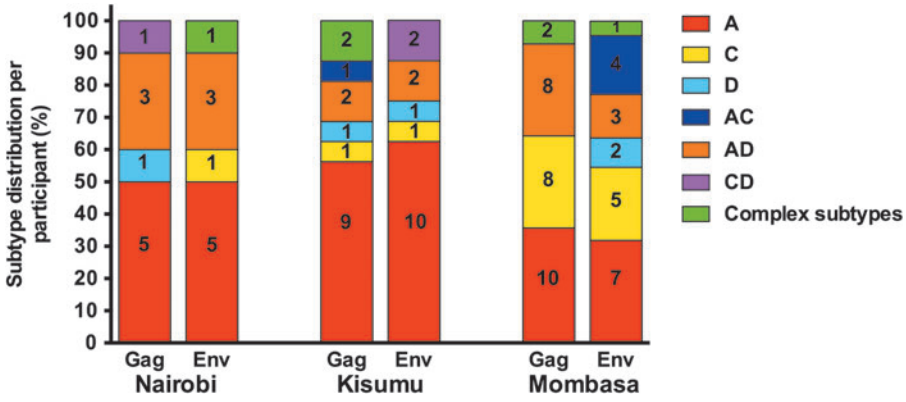


**FIG. 2.** Distribution (%) of HIV-1 subtypes and intersubtype recombinants among drug users in Nairobi, Kisumu, and Mombasa based on the number of *gag* and *env* clonal sequences obtained for each region. The number of participants screened for each city in Kenya is shown within the individual subtype bars.

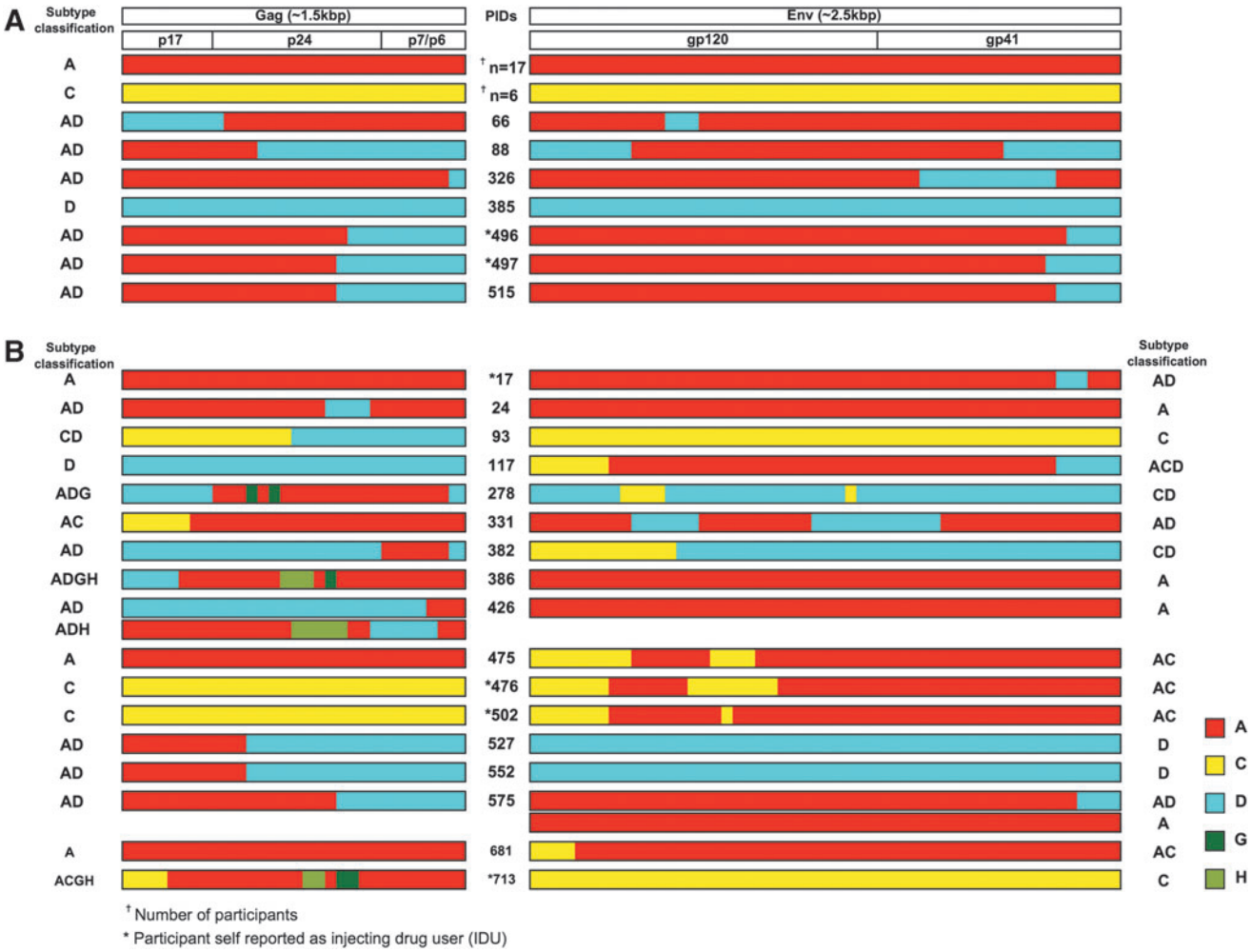


**FIG. 3.** Schematic representation of concordant and discordant subtypes based on full-length clonal *gag* and *env* genes analyzed from drug users in Kenya. **(A)** *gag* and *env* gene structure of the 30 concordant subtypes. **(B)** Gene structures of full-length *gag* and *env* showing discordant subtype distribution in 17 drug user participants. Clonal sequencing revealed mixed subtypes (PID 426 had mixed subtypes in the *gag* region, while PID 575 had mixed subtypes in the *env* region). *Represents known injecting drug users. Subtype classification of each gene is indicated. Subtype distribution is shown in different colors (A, *red*; C, *yellow*; D, *light blue*; G, *dark green*, and H, *light green*). Note that the gene lengths are not drawn to scale.

sequence fragments were trimmed, realigned to subtype reference sequences, and phylogenetic trees were constructed to verify subtype classification for each sequence fragment between identified subtype breakpoints.

Of the 673 study participants enrolled into the parent drug user study, 80 (11.9%) were HIV infected. Out of 80 samples subjected to amplification protocols, only 55 (69%) yielded an amplicon. The demographic and epidemiological characteristics of these participants are shown in Table 1. The majority of participants were male (43 of 55, 78.2%) and the median age of the participants was 30, with an interquartile range (IQR) of 26–34. Drug users were recruited from three sites within Kenya, 10 participants (18.2%) from Nairobi, 17 (30.9%) from Kisumu, and 28 (50.9%) from Mombasa. Only 17 (30.9%) of the participants self-reported as IDUs. Six of 55 (10.9%) were positive for HBV and 15 of 55 (27.3%) tested positive for HCV.

From the 55 HIV-positive drug users with HIV PCR products that could be evaluated, we successfully generated a total of 220 *gag* clonal sequences from 54 of the 55 participants (98.2%) and 156 *env* clonal sequences from 47 of 55 (85.5%) participants. The median number of sequences analyzed for *gag* was four per participant and three sequences for *env*. The remaining samples either had negative PCR results or ambiguous sequences, which were discarded.

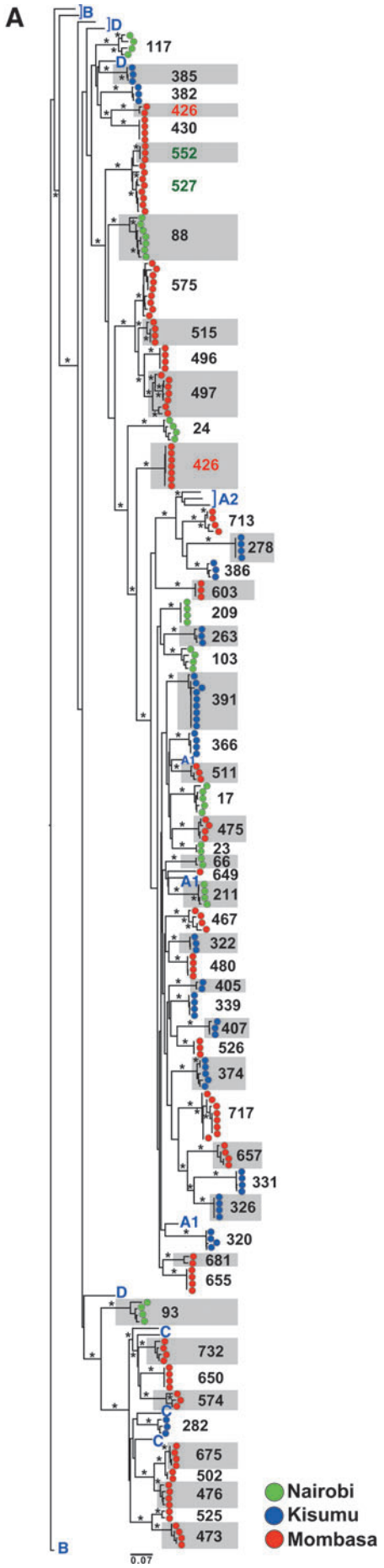


**FIG. 4.** Phylogenetic analysis of drug user participants from three major cities in Kenya. **(A)** Maximum likelihood trees were constructed using 220 HIV-1 full-length *gag* clonal sequences collected from drug user participants recruited from Nairobi (*green circles*), Kisumu (*blue circles*) and Mombasa (*red circles*). Possible linked participants (PIDs 552 and 527) are highlighted in *green text* and possible dual infected participant is highlighted in *red text* (PID 426). **(B)** Detailed phylogenetic analysis of possible dual infected drug user participant 426 (*red*) and epidemiologically-linked drug users (*green*) are shown. Both clonal sequences (*solid red circles*) together with single genome amplification (SGA) derived *gag* sequences (*open red circles*) for PID 426 show infection with subtypes A, D, AD and ADH. Phylogenetic analysis shows strong bootstrap support of epidemiological linkage between participants PID 552 and 527 (*green circles*) after sequences were downloaded and included in phylogenetic analysis after BLAST analysis of individual participant sequences. **(C)** Phylogenetic analysis of 156 HIV-1 full-length *env* sequences from drug user participants from Nairobi (*green*), Kisumu (*blue*) and Mombasa (*red*). Possible dual infected drug user participant 426 (*red text*) and epidemiologically-linked drug users PID 552 and 527 (*green text*) are shown. **(D)** Detailed phylogenetic analysis of the epidemiologically-linked participants PID 552 and 527 (*green*) after adding downloaded BLAST analysis sequences suggest strong linkage between the two participants. All HIV-1 reference sequences representing genetic subtypes were obtained from the Los Alamos National Laboratory HIV Sequence Database included: Ref.A1.RW.92, Ref.A1.UG.92, Ref.A1.AU.03, Ref.A2.CD.97, Ref.A2.CY.94, Ref.A2.CM.01, Ref.B.NL.00, Ref.B.US.98, Ref.B.FR.83, Ref.C.ET.86, Ref.C.BR.92, Ref.C.ZA.04, Ref.C.IN.95, Ref.D.CD.83, Ref.D.TZ.01, Ref.D.UG.94, and Ref.D.CM.01 for both *gag* and *env* (*blue text*). Bootstrap support ≥70 is indicated with an *asterisk* (*). The *gray boxes* highlight alternate individual participant sequence clusters.

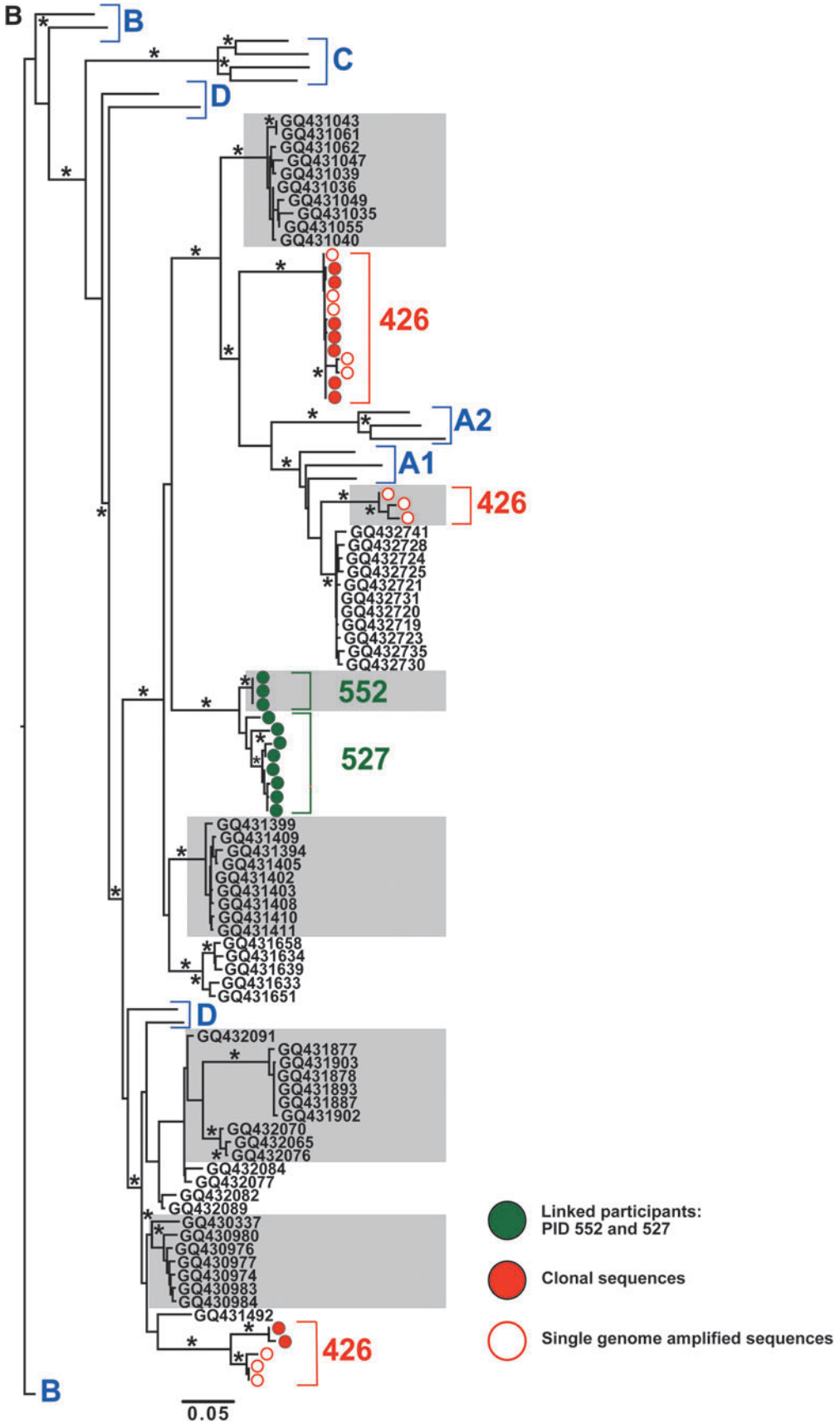


**FIG. 4.** (*Continued*).

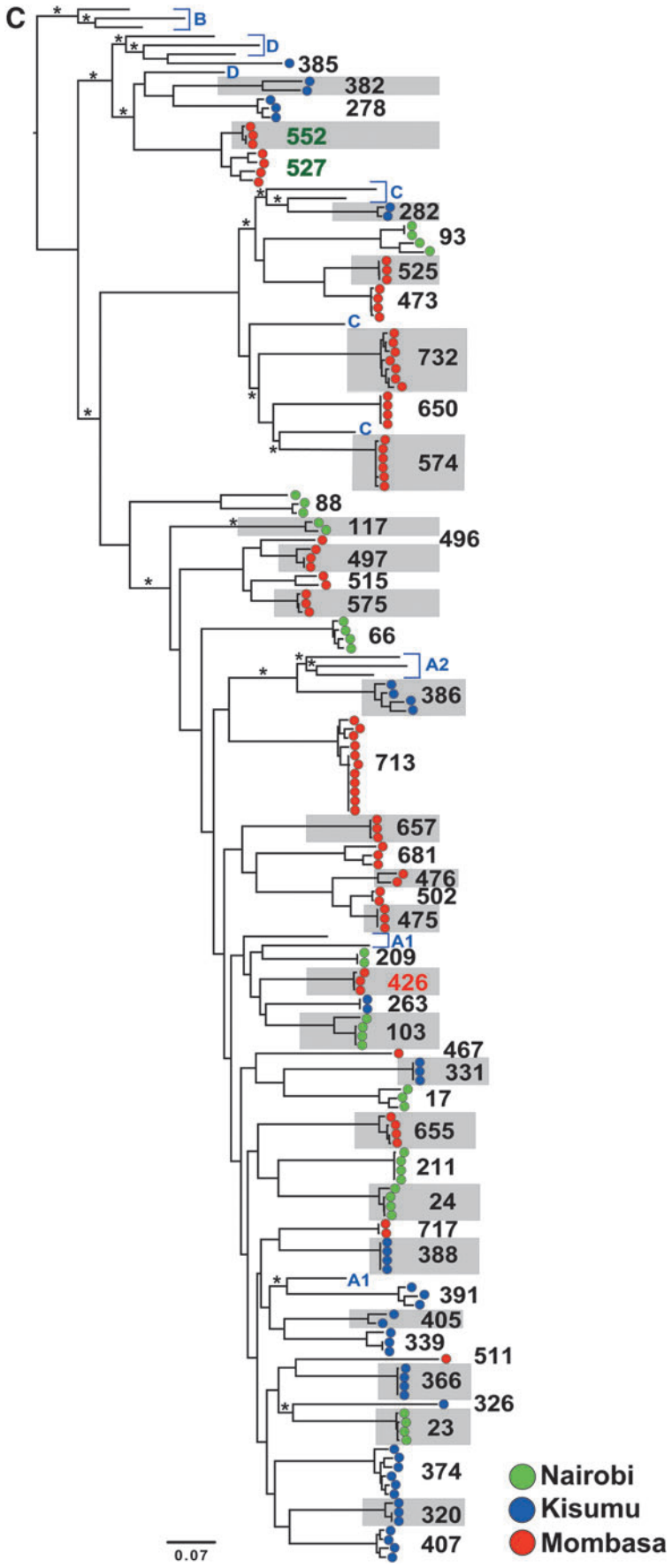


**FIG. 4.** (*Continued*).

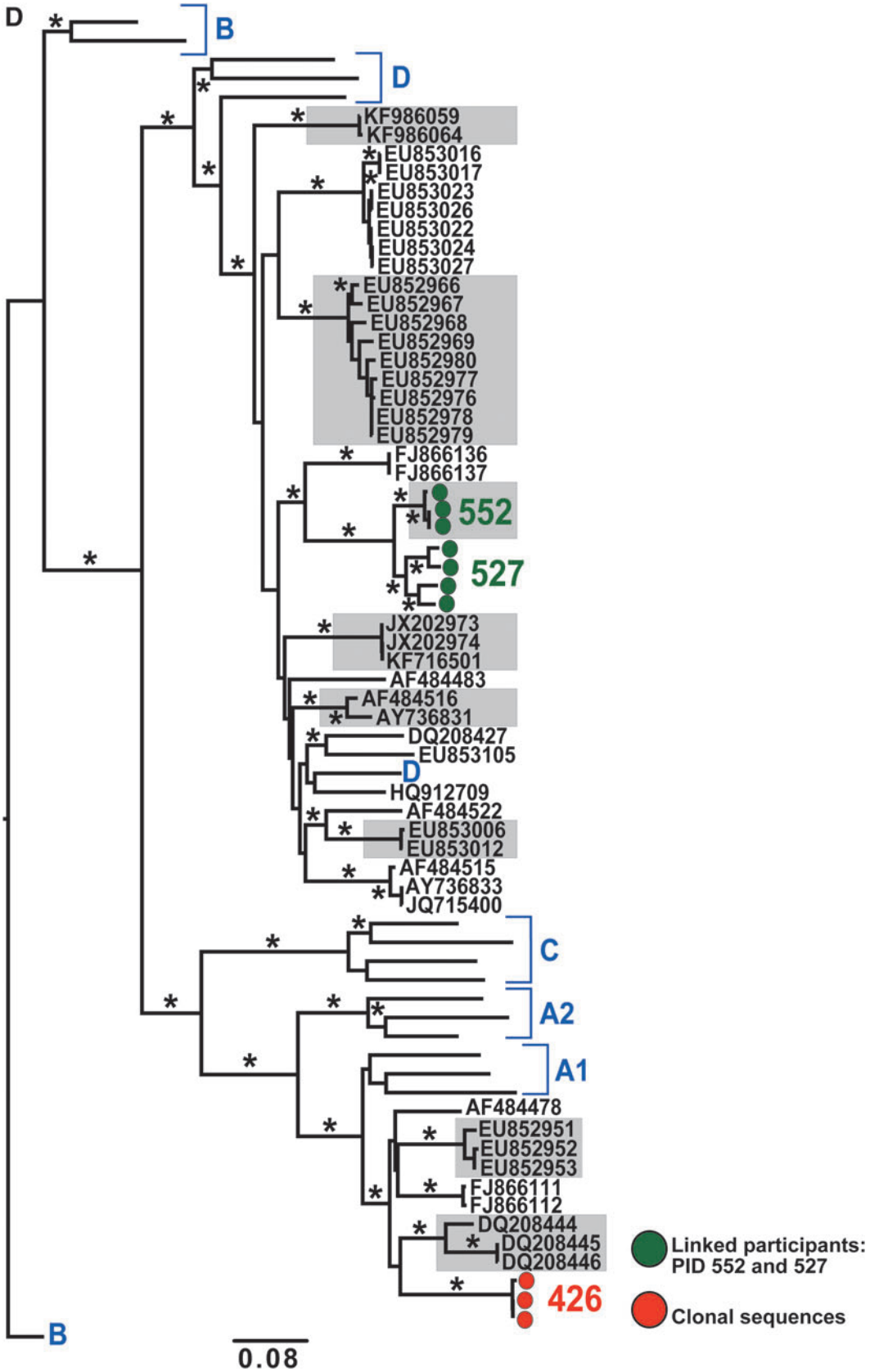


**FIG. 4.** (*Continued*).

*Gag* sequences indicated that 24 (44.4%) drug users were infected with subtype A, 9 (16.7%) were infected with subtype C, 2 (3.7%) were infected with subtype D, and the remaining 19 (35.2%) were infected with intersubtype recombinants (Fig. 1A). Similarly, *env* sequence analyses revealed that 22 (45.8%) participants were infected with subtype A, 7 (14.6%) were infected with subtype C, 3 (6.3%) were infected with subtype D, and the remaining 16 (33.3%) were infected with recombinants (Fig. 1B). We next determined the number of participants infected with a pure subtype, recombinant, or a mixture of viruses (sequences containing both a pure subtype and a recombinant virus) (Fig. 1C). Both *gag* and *env* sequence data showed that majority of individuals were infected with a pure subtype ($N=35$, 64.8% and $N=32$, 66.7%, for *gag* and *env*, respectively), compared to ($N=18$) 33.3% and ($N=15$) 31.3% of individuals infected with recombinant viruses and a relatively low number of individuals infected with a mix of both pure and recombinant viruses [$N=1$ (1.9%) and $N=1$ (2%), respectively].

Our data provide potentially important insights into the changing HIV-1 subtype dynamics of the Kenyan epidemic, particularly among the key population of drug users. Our data are consistent with previous studies that subtype A is the predominant subtype in Kenya; however, we found that in this population group, subtype C (16.7% in *gag* and 14.6% in *env*), was the second most common subtype, with a far less contribution of subtype D (3.7% in *gag* and 6.3% in *env*) than the anticipated $\sim$9% to 20% based on majority of previous studies.[3,10–12] Remarkably, despite higher percentage of pure subtype C infections compared to D in both *gag* and *env* in this cohort, AD were the majority intersubtype recombinants consisting of 68.4% in *gag* and 50% in *env* (Fig. 1A, B), suggesting that recombination may not be random and subtype D is more likely to generate recombinants compared to subtype C. Alternatively, subtype D may have existed at higher prevalence in the past and may be decreasing, with its remnants still evident in the form of recombinants. Overall, our data are consistent with predominance of HIV-1 subtype A in Kenya, but relatively higher percentage of subtype C, and lower percentage of subtype D than has been reported previously.

We next assessed whether there are differences in subtype distribution in the geographically spread out Kenyan towns of Kisumu (Western Kenya), Mombasa (Coast region), and Nairobi (Central Kenya) where there may be differences in socio-demographic factors and risk behaviors influencing HIV spread. HIV-1 subtype A was the predominant subtype in all three Kenyan cities irrespective of the virus gene analyzed (Fig. 2). The percentage of participants whose sequences clustered with subtype A references in Nairobi accounted for 50% in both *gag* and *env*, respectively, compared to 56.3% and 62.5% in Kisumu. For the Mombasa dataset, the percentages of individuals infected with subtypes A was lower at 35.7% and 28.6% for *gag* and *env*, respectively, just slightly higher than the percentages observed for subtype C at 31.8% and 22.7% for *gag* and *env*, respectively.

In contrast, for all three cities, the percentage of participants infected with subtype D was generally below 10%. In Nairobi, 10% of participants could be classified as subtype D, but none of the *env* sequences were subtype D, while 10% of participant *env* sequences were classified as subtype C with no *gag* subtype C detected. Among sequences generated from participants in Kisumu, the same percentage of participants was infected with subtype D (6.3%) and C (6.3%) in both *gag* and *env*, respectively. In Mombasa, 9.1% of participants were subtype D based on *env* sequences, with no *gag* subtype D sequences detected. Of the intersubtype recombinants, only subtype AD was found in significant percentages. In Nairobi and Kisumu, 30% and 12.5% of participants were subtype AD for both *gag* and *env*, respectively, compared to 28.6% in *gag* and 13.6% in *env* in Mombasa. All other recombinants were below 18% prevalence at all sites irrespective of gene region analyzed.

It is therefore noteworthy that in this study, we observed a trend whereby subtype C mostly surpassed or was almost equal to subtype D in percentage of sequences analyzed, a trend that was most prominent in Mombasa. Mombasa is a coastal town located near the border and with strong cultural links to Tanzania, Kenya's southern neighbor, which itself has a sizable HIV-1 subtype C epidemic. It is therefore possible that high prevalence of C represents recent transmissions from movement of people across the Kenya–Tanzania border. HIV-1 subtype A was generally highest in prevalence, followed by AD intersubtype recombinants, despite the lower prevalence of pure D compared to C sequences. Our data call for further studies to explore biological differences that may explain the changing subtype landscape. Although one study suggested that women infected with HIV-1 subtype C viruses in Kenya had significantly higher viral loads and lower CD4 cell counts compared to clade A or D infection,[10] other studies from Kenya have demonstrated that subtype D infection is associated with a faster rate of $CD4^+$ T decline and time to AIDS compared to subtypes A or C.[11,13] It is plausible that faster progression to AIDS for subtype D may result in reduced lifetime odds of transmitting the virus, which in turn may result in reduced prevalence of this subtype in a heterogeneous epidemic. Alternatively, subtype D may have specific biological characteristics that provide a disadvantage for its transmission efficiency, or the vice-versa may be true for subtype C. In support of these hypotheses are studies showing subtype-specific differences in Pol-mediated replicative fitness[14] and in Nef protein functions, with subtype D suggested to be more virulent compared to either subtype A or C.[15]

We next assessed subtype concordance within individuals for the two genes analyzed. Of the 47 participants with both *gag* and *env* sequences, 30 (63.8%) showed subtype concordance of the *gag* and *env* sequences (17 subtype A, 6 subtype C, 1 subtype D, and 6 had AD intersubtype recombinants, with recombination breakpoints more common at the 3′ end of *gag* and *env*, although the dataset was too small to perform statistical associations) (Fig. 3A). Seventeen participants (36.2%) had discordance in subtype between *gag* and *env* with a variety of different subtype combinations and recombinant breakpoints (Fig. 3B) (the remaining 8 of 55 participants could not be included in this analysis because they had a negative PCR for *env*). Among the discordant participants, the subtype combinations included two each of A/AC, C/AC, and AD/D and one each of A/AD, AD/A, CD/C, D/ACD, ADG/CD, AC/AD, AD/CD, ADGH/A, and ACGH/C between the *gag* and *env* genes, respectively. Two participants (PID 426 and 575) showed evidence of dual infection. PID 426 had a combination of AD and ADH *gag* sequences with

corresponding pure subtype A *env* clonal sequences. Participant 575 had a recombinant AD *gag* sequence with a mixture of AD and pure A *env* sequences. In this cross-sectional study, it was not possible to determine whether these sequences represent mixed infections or infection with pure subtype viruses that then undergo recombination in the infected patient. However, the presence of discordant subtypes in *gag* versus *env* suggests either multivariant transmitted/founder viruses or dual infections with subsequent generation of viral recombinants among drug user populations in Kenya.

To determine whether the discordant mixed infections were representative of dual infection, we subjected the gene sequences to further phylogenetic analysis. We defined dual/mixed infection as an infection with two phylogenetically distinct viruses. Viruses from one drug user (PID 426) from Mombasa grouped separately on the *gag* phylogenetic tree demonstrating that this individual was infected with two different viruses (Fig. 4A). *Gag* viral populations grouped as subtype AD ($n=2$ sequences), with a second branch of recombinant subtype ADH ($n=7$) sequences. To confirm these cases of dual infection in *gag*, we performed SGA to ameliorate potential bias from PCR-generated recombination or resampling (Fig. 4B). The resulting SGA sequences confirmed the phylogenetically distinct sequences present in participant 426 (shown in orange), including pure subtype A and D sequences that were not picked up by clonal sequencing. However, phylogenetic analysis did not reveal any possible dual infection within the *env* region, with all sequences forming a single cluster (Fig. 4C). In addition, among these drug users appeared a case of epidemiologically linked viruses in participants PID 552 and PID 527, with sequences that clustered together for both *gag* and *env* with high bootstrap upon stringent phylogenetic analysis, with closely related sequences downloaded after a BLAST search as shown in Figure 4A–D. Both participants are females from Mombasa who may have shared a sexual partner or were part of the same drug user social network.

Our study has some limitations worth highlighting. First, we lacked clinical history and immune profile data such as CD4 counts and viral load, which would have been useful for more detailed interpretation of data from the study. In addition, this was a small sample size, cross-sectional drug user cohort study, restricted only to major urban sites in Kenya and may therefore not be representative of other key population groups that drive the epidemic in the region. Moreover, of 80 HIV-positive samples, we could amplify the virus from only 55 (68.8%), indicating either poor quality or extreme genetic diversity in the remaining samples, and our data should therefore be interpreted with caution. Last, as underlined by our analysis of *gag* and *env* regions, recombination patterns in a heterogeneous epidemic are complex and will require generation of full-length single genome level data to fully understand the changing dynamics of viral genotypes and infer the possible implications for medical and public health interventions.

In conclusion, our study highlights the high prevalence of HIV-1 subtype A in a key population group within the epidemic in Kenya. Our data suggest that HIV-1 subtype C is an important component of the epidemic within this group, with a hint of geographical differences across the Kenya. Interestingly, we found lower prevalence of subtype D than has been reported in most previous studies, and despite this low prevalence, subtype D comprised the majority of intersubtype recombinants. Further studies focusing on longitudinal studies and full-length genotypic characterization of viral variants in larger key population groups in Kenya are recommended to optimize public health strategies for interventions.

### Sequence Data

Full-length *gag* and *env* nucleotide sequences obtained in this study have been submitted to the GenBank database under accession numbers: KU921705-KU922091.

### Acknowledgments

We thank and acknowledge the drug user study participants. We acknowledge Dr. Johannes Viljoen and the Africa Center laboratory, Durban, South Africa, for providing access to the sequencing facility. We thank Keshni Hiramen for technical assistance. The research was supported by the South African Research Chairs Initiative, the Victor Daitz Foundation, and an International Early Career Scientist award from the Howard Hughes Medical Institute to T.N. This work was also partly supported through the DELTAS Africa Initiative (grant No. 107752/Z/15/Z). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (grant No. 107752/Z/15/Z) and the UK Government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust, or the UK Government. K.G. was funded by the South African Medical Research Council (MRC). M.O. was funded by the IAS/NIDA.

### Authors' Contributions

Conceived and designed the overall study: M.O. and J.W. Conceived and designed the molecular analyses: T.N., K.G., and M.O. Performed the molecular analyses: K.G., N.P., and T.M.Z. Analyzed the data: K.G. and T.d.O. Wrote the article: K.G. and T.N. All authors read and approved the final draft.

### Author Disclosure Statement

No competing financial interests exist.

### References

1. Virgin HW, Walker BD: Immunology and the elusive AIDS vaccine. Nature 2010;464:224–231.
2. Ndung'u T, Weiss RA: On HIV diversity. AIDS 2012;26:1255–1260.
3. 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;29:187–190.
4. Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM: The challenge of HIV-1 subtype diversity. N Engl J Med 2008;358:1590–1602.
5. Kiwanuka N, Laeyendecker O, Quinn TC, *et al.*: HIV-1 subtypes and differences in heterosexual HIV transmission

among HIV-discordant couples in Rakai, Uganda. AIDS 2009;23:2479–2484.
6. Palm AA, Esbjornsson J, Mansson F, *et al.*: Faster progression to AIDS and AIDS-related death among seroincident individuals infected with recombinant HIV-1 A3/CRF02_AG compared with sub-subtype A3. J Infect Dis 2014;209:721–728.
7. Beckerleg S, Telfer M, Hundt GL: The rise of injecting drug use in East Africa: A case study from Kenya. Harm Reduct J 2005;2:12.
8. Nofemela A, Bandawe G, Thebus R, *et al.*: Defining the human immunodeficiency virus type 1 transmission genetic bottleneck in a region with multiple circulating subtypes and recombinant forms. Virology 2011;415:107–113.
9. Salazar-Gonzalez JF, Bailes E, Pham KT, *et al.*: Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J Virol 2008;82:3952–3970.
10. Neilson JR, John GC, Carr JK, *et al.*: Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. J Virol 1999;73:4393–4403.
11. Baeten JM, Chohan B, Lavreys L, *et al.*: HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. J Infect Dis 2007;195:1177–1180.
12. Yang C, Li M, Shi YP, *et al.*: Genetic diversity and high proportion of intersubtype recombinants among HIV type 1-infected pregnant women in Kisumu, western Kenya. AIDS Res Hum Retroviruses 2004;20:565–574.
13. McKinnon LR, Nagelkerke NJ, Kaul R, *et al.*: HIV-1 clade D is associated with increased rates of CD4 decline in a Kenyan cohort. PLoS One 2012;7:e49797.
14. Ng OT, Laeyendecker O, Redd AD, *et al.*: HIV type 1 polymerase gene polymorphisms are associated with phenotypic differences in replication capacity and disease progression. J Infect Dis 2014;209:66–73.
15. Mann JK, Byakwaga H, Kuang XT, *et al.*: Ability of HIV-1 Nef to downregulate CD4 and HLA class I differs among viral subtypes. Retrovirology 2013;10:100.

Address correspondence to:
*Thumbi Ndung'u*
*HIV Pathogenesis Programme*
*Doris Duke Medical Research Institute*
*Nelson R. Mandela School of Medicine*
*University of KwaZulu-Natal*
*719 Umbilo Road*
*Durban 4013*
*South Africa*

*E-mail:* ndungu@ukzn.ac.za