



# Increased Frequency of Inter-Subtype HIV-1 Recombinants Identified by Near Full-Length Virus Sequencing in Rwandan Acute Transmission Cohorts

## OPEN ACCESS

### Edited by:

Kok Keng Tee,  
University of Malaya, Malaysia

### Reviewed by:

Sudhanshu Shekhar Pandey,  
Triviron healthcare Ltd., India  
Marcel Tongo,  
Center for Research on Emerging  
and Re-Emerging  
Diseases (CREMER), Cameroon

### \*Correspondence:

Ling Yue  
lyue2@emory.edu  
Eric Hunter  
ehunte4@emory.edu

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

<sup>‡</sup>These authors have contributed  
equally to this work and share senior  
authorship

### Specialty section:

This article was submitted to  
Virology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 01 July 2021

**Accepted:** 07 September 2021

**Published:** 07 October 2021

### Citation:

Umviligihozo G, Muok E, Nyirimihigo  
Gisa E, Xu R, Dileria D, Herard K,  
Song H, Qin Q, Bizimana J, Farmer P,  
Hare J, Gilmour J, Allen S, Karita E,  
Hunter E and Yue L (2021) Increased  
Frequency of Inter-Subtype HIV-1  
Recombinants Identified by Near  
Full-Length Virus Sequencing in  
Rwandan Acute Transmission  
Cohorts.  
*Front. Microbiol.* 12:734929.  
doi: 10.3389/fmicb.2021.734929

Gisele Umviligihozo<sup>1†</sup>, Erick Muok<sup>1†</sup>, Emmanuel Nyirimihigo Gisa<sup>1</sup>, Rui Xu<sup>2</sup>, Dario Dileria<sup>2</sup>, Kimberley Herard<sup>2</sup>, Heeyah Song<sup>2</sup>, Qianhong Qin<sup>2</sup>, Jean Bizimana<sup>1</sup>, Paul Farmer<sup>2</sup>, Jonathan Hare<sup>3</sup>, Jill Gilmour<sup>4</sup>, Susan Allen<sup>5</sup>, Etienne Karita<sup>1</sup>, Eric Hunter<sup>2,5†\*</sup> and Ling Yue<sup>2†\*</sup>

<sup>1</sup>Centre for Family Health Research, Kigali, Rwanda, <sup>2</sup>Emory Vaccine Center at Yerkes National Primate Research Center, Atlanta, GA, United States, <sup>3</sup>IAVI, New York, NY, United States, <sup>4</sup>Faculty of Medicine, Imperial College London, London, United Kingdom, <sup>5</sup>Department of Pathology and Laboratory Medicine, Emory University, Atlanta, GA, United States

Most studies of HIV-1 transmission have focused on subtypes B and C. In this study, we determined the genomic sequences of the transmitted founder (TF) viruses from acutely infected individuals enrolled between 2005 and 2011 into IAVI protocol C in Rwanda and have compared these isolates to viruses from more recent (2016–2019) acute/early infections in three at risk populations – MSM, high risk women (HRW), and discordant couples (DC). For the Protocol C samples, we utilized near full-length single genome (NFLG) amplification to generate 288 HIV-1 amplicons from 26 acutely infected seroconverters (SC), while for the 21 recent seroconverter samples (13 from HRW, two from DC, and six from MSM), we PCR amplified overlapping half-genomes. Using PacBio SMRT technology combined with the MDPseq workflow, we performed multiplex sequencing to obtain high accuracy sequences for each amplicon. Phylogenetic analyses indicated that the majority of recent transmitted viruses from DC and HRW clustered within those of the earlier Protocol C cohort. However, five of six sequences from the MSM cohort branched together and were greater than 97% identical. Recombination analyses revealed a high frequency (6/26; 23%) of unique inter-subtype recombination in Protocol C with 19% AC and 4% CD recombinant viruses, which contrasted with only 6.5% of recombinants defined by sequencing of the *pol* gene previously. The frequency of recombinants was significantly higher (12/21; 57%) in the more recent isolates, although, the five related viruses from the MSM cohort had identical recombination break points. While major drug resistance mutations were absent from Protocol C viruses, 4/21 of recent isolates exhibited transmitted nevirapine resistance. These results demonstrate the ongoing evolution and increased prevalence of recombinant and drug resistant transmitted viruses in Rwanda and highlight the importance of defining NFLG sequences to fully understand the nature

of TF viruses and in particular the prevalence of unique recombinant forms (URFs) in transmission cohorts.

**Keywords:** HIV-1 transmission, genetic bottleneck, HIV subtype, inter-subtype recombination, antiretroviral drug resistance mutations, near full-length genome sequencing

## INTRODUCTION

Worldwide 37 million people are living with HIV, two-thirds of these infected individuals are found in Sub-Saharan Africa (UNAIDS, 2020). Even though more than half are receiving ART, a significant fraction of treated patients are not virally suppressed (Hamers et al., 2012; Hauser et al., 2019), and HIV prevention remains a major problem in the fight against HIV. A global effort to design and develop an effective HIV-1 vaccine has been carried out over the last 30 years, but one of its major challenges is the enormous diversity of HIV-1. This can be attributed to several factors: the high error rate of the viral reverse transcriptase, since it lacks a proof-reading function; host immune responses that constantly apply selection pressure for less susceptible virus; and the propensity for the virus to undergo recombination (Lukashov and Goudsmit, 1998; Mansky, 1998; Korber et al., 2001; Song et al., 2018). A number of investigators have studied the interplay between HIV-1 and host immunity, and have shown that viral adaptation, particularly to the cellular immune response during the course of infection, can be a major contributor to viral evolution (Moore et al., 2002; Brumme and Walker, 2009; Crawford et al., 2009; Kawashima et al., 2009; Carlson et al., 2012, 2016; Monaco et al., 2016). However, recombination between genetically distinct viruses has the greatest potential to generate diversity (McCutchan et al., 1996; Butler et al., 2007; Lau and Wong, 2013; Giovanetti et al., 2020). HIV-1 has been classified into four phylogenetic groups M, O, N, and P based on nucleic acid sequencing of the viral genomic RNA, with group M being by far the most widespread (Robertson et al., 2000; Desire et al., 2018). The latter is subdivided into nine different subtypes (A–D, F–H, J, K, and the newly identified L), with genetic variation between subtypes ranging from 20 to 35% depending on the genomic regions and the subtypes being compared (Korber et al., 2001; Desire et al., 2018; Yamaguchi et al., 2020). The process of recombination between viruses belonging to different subtypes and the ongoing spread of those recombinants is the basis for the emergence of circulating recombinant forms or CRFs. To date over 102 inter-subtype CRFs have been described (Hemelaar et al., 2020; LANL, 2021). Therefore, to develop a broadly effective prophylactic vaccine, there is a clear need to gain insight into the genotypic and phenotypic features of the viruses from various geographic locations against which a potential vaccine must act.

Rwanda is an East-Central African country bordered by the Democratic Republic of Congo, with a highly diverse HIV-1 population (Rodgers et al., 2017), Burundi, where subtype C is most prevalent (Delatorre and Bello, 2012) and Uganda where subtypes A1 and D predominate, but with a high percentage of unique recombinant forms (URFs; Lee et al., 2017;

Grant et al., 2020). Thus, defining the nature of HIV-1 diversity over time in this geographically small, land-locked country will be relevant to ongoing HIV-1 vaccine efforts. Based on a region encompassing the *pol* gene, an earlier subtype analysis of over 90 incident infections enrolled under IAVI Protocol C in Rwanda identified 80% as subtype A1 and only 6.5% as recombinant viruses (Amornkul et al., 2013), while a second study of smaller sample size, where *gag*, *pol*, and *env* genes were sequenced, reported 13.5% recombinant forms in the same cohort (Kemal et al., 2013).

In the current study, we have amplified near full-length single genomes (NFLG) of viruses from the plasma of a total of 26 acutely HIV infected individuals from the Rwandan heterosexual acute infection cohort Protocol C and 21 recently infected individuals from high-risk cohorts. This allowed us to define the sequence of the infecting viruses and compare over two time periods (2005–2012 and 2016–2019) the frequency of inter-subtype recombination across the full genome. In addition, because these two periods define very different availability of anti-retroviral therapies, we were able to compare the prevalence of antiretroviral drug resistance. These data point to an increase in genetic mixing and prevalence of transmitted drug resistance in Rwanda over the last 15 years and highlight the importance of NFLG sequencing for assessing diversity in viral populations.

## MATERIALS AND METHODS

### Ethics Statement

Subjects in this study were enrolled in human subjects protocols approved by the Rwanda National Ethics Committee and Emory University Institutional Review Board. All study subjects have provided written informed consent.

### Study Subjects

In this study, we have studied HIV-1 Early infection subjects from two distinct time periods during last 15 years in Kigali Rwanda. During the first period from 2005 to 2011, plasma samples were collected under IAVI Protocol C (Price et al., 2020) from 26/97 (27%) seroconvertors from a heterosexual transmission cohort in Kigali. Individuals enrolled in this cohort were from HIV-1 discordant couples who underwent couples counseling and testing and who were followed, with additional counseling and testing, every 1–3 months to reduce the incidence of transmission. HIV infection was identified by p24 ELISA antigen testing or seroconversion. The 26 seroconvertors were selected from the original 97 based on the availability of sample during the very early/acute period of HIV infection. During the

second period from 2016 to 2019, 21 individuals who were followed in government clinics every 3–6 months in virtual cohorts comprised of high-risk women, men having sex with men and discordant couples, were enrolled immediately after seroconversion.

## Viral RNA Extraction and cDNA Synthesis

Viral RNA was extracted from patient's plasma using the QIAamp RNA mini kit (Qiagen, Valencia, CA). For near-full-length genome amplification, 140  $\mu$ l plasma were used for vRNA isolation, and then the purified vRNA was converted to full-length cDNA with SuperScript III Reverse Transcriptase (Life Technologies) enzyme with a reverse HIV-1 primer that designed at the end of the R region in LTR (Yue et al., 2015). For 5' and 3' half genome amplifications, the amount of plasma sample was calculated according to VL and aliquoted based on 300 copies  $\times$  # of reaction per each extraction, and then diluted to 140  $\mu$ l with PBS if less than 140  $\mu$ l.

## Near Full-Length Single HIV-1 Genome Amplification

cDNA was serially diluted to yield approximately 30% PCR positive to ensure the majority of amplicons were derived from single virus RNA molecule (Yue et al., 2015). A 9 kb PCR fragment extending from the 5' U5 to 3' R region of the genome was generated by using Q5 Hot Start High Fidelity DNA Polymerase (NEB; Deymier et al., 2014). The amplification primers are shown in **Table 1** and conditions were as described previously (Deymier et al., 2014).

## One-Step RT-PCR Population Half-Genome Amplifications

One-step PCR was conducted using the SuperScript™ III One-Step RT-PCR System with Platinum™ Taq High Fidelity DNA Polymerase (Invitrogen). Master mix I (MMI) contained 25  $\mu$ m reverse first round primer and template vRNA (300 copies per reaction), adjusted to a total volume of 11  $\mu$ l per

reaction with H<sub>2</sub>O. The MMI was incubated at 65°C for 5 min to melt secondary structures in the RNA then temperature was decreased to 4°C to anneal the 5' or 3' first round reverse primer (**Table 1**) with the RNA template. Master mix II (MMII) contains 2XReaction buffer, 5' or 3' first round forward primer (25  $\mu$ m; **Table 1**), SuperScript III/Platinum Taq Mix in a total volume of 39  $\mu$ l per each reaction. The 39  $\mu$ l MMII was added to 11  $\mu$ l MMI at 4°C, and then the entire 50  $\mu$ l reaction was incubated at 55°C for 30 min in PCR cycler to synthesize cDNA.

After the cDNA was synthesized, the initial PCR step was 2 min at 94°C; followed by 30 cycles of 94°C 15s, 52°C 30s, 68°C 6min; and an additional one step of 68°C for 10 min for stabilization.

Second round PCR was carried out by using Q5 Hot Start High Fidelity DNA Polymerase (NEB). The sequences and positions of the 5' half second round primers are shown in **Table 1**. PCR conditions were 98°C 30s as the initial step, followed by 35 cycles of 98°C 10s, 64°C 30s and 72°C 4min; plus, an additional step of 72°C 10min prior to keeping the reaction at 4°C.

The sequences and positions of the 3' half second round forward primer and reverse primer OFM19 are shown in **Table 1**. PCR conditions were 98°C 30s as the initial step, followed by 35 cycles at 98°C 10s, 58°C 30s for annealing and 72°C 4min for extension; plus, an additional step of 72°C 10min prior to keeping the reaction at 4°C. The second round PCR resulted in a 4,456bp fragment from the 5'half, and a 4,742bp fragment from the 3'half.

## PacBio DNA Sequencing Library Preparation

Four SMRTbell™ libraries of NFLSGA and four SMRTbell™ libraries of half genome amplicons were built to gain deep sequencing data. The PacBio sequencing method was described previously (Dilernia et al., 2015). In brief, we combined 75 NFLSGA amplicons for each RSII library; 10 patients' half genome PCR products were collected for each RSII library. The final library DNA concentration was more than 20 ng/ $\mu$ l.

**TABLE 1** | cDNA and PCR Primers.

Primer name	Primer sequence (5'-3')	Position in HXB2	Application
OFM19	5'-GCACTCAAGGCAAGCTTTATTGAGGCTTA-3'	9,604–9,632	cDNA synthesis for NFLSGA
1.3'3'pICb	5'-ACTACTTAAAGCACTCAAGGCAAGCTTTATTG-3'	9,611–9,642	cDNA synthesis for NFLSGA
1U5Cc	5'-CCTTGAGTGCTCTAAGTAGTGTGTGCCCGTCTGT-3'	538–571	First round NFLSGA forward primer
1.3'3'pICb	5'-ACTACTTAAAGCACTCAAGGCAAGCTTTATTG-3'	9,611–9,642	First round NFLSGA reverse primer
2U5Cd	5'-AGTAGTGTGTGCCCGTCTGTTGTGTGACTC-3'	552–581	Second round NFLSGA forward primer
2.3'3'pICb	5'-TAGAGCACTCAAGGCAAGCTTTATTGAGGCTTA-3'	9,604–9,636	Second round NFLSGA reverse primer
1half_R1_For	5'-TTTGACTAGCGGAGGCTAGAA-3'	761–781	5'HF first round PCR forward primer
1half_R1_Rev	5'-TTCTATGGAGACYCCATGACCC-3'	5,304–5,283	5'HF cDNA synthesis and first round PCR reverse primer
2half_R1_For	5'-GGGTTTATTACAGGGACAGCAGAG-3'	4,900–4,923	3'HF first round PCR forward primer
OFM19	5'-GCACTCAAGGCAAGCTTTATTGAGGCTTA-3'	9,604–9,632	3'HF cDNA synthesis and first and second round PCR reverse primer
1half_R2_For	5'-TTTGACTAGCGGAGGCTAGAAGGA-3'	761–784	5'HF second round PCR forward primer
1half_R2_Rev	5'-TCCCCTARTGGGATGTGTACTTCTGAAC-3'	5,195–5,222	5'HF second round PCR reverse primer
2half_R2_For	5'-GCAAACTACTCTGAAAGGTGAAGGG-3'	4,944–4,970	3'HF second round PCR forward primer

purity 260/280 ratio was greater than 1.8, and 260/230 ratio was greater than 2.0; total volume was 30  $\mu$ l. SMRT sequencing was performed on a PacBio RSII at the University of Delaware DNA Sequencing & Genotyping Center.

## Sequence Analysis

Data derived from the PacBio RSII was run using the error correction algorithm MDPseq (Dilernia et al., 2015). Defining transmitted founder (TF) viral sequences and phylogenetical analysis were carried out through Geneious v9.1.8 (Biomatters Ltd). Codon-align for each HIV-1 protein was performed by Gene Cutter (LANL). Subtyping and recombinant identification were carried out by Recombinant analysis program (RIP) and jpHMM at GOBICS (LANL).

## GenBank Submission

Near full-length (NFL) sequences were submitted to GenBank. The accession numbers for the 26 Protocol C derived viruses are JX236678.1, JX236677.1, and MT942708-MT942972. Those for the 21 recent seroconverters are MZ642260-MZ642280.

## RESULTS

### Study Volunteers

This project was conducted in partnership with Projet San Francisco/Centre for Family Health Research which was established in Kigali, Rwanda in 1986. Two distinct groups of HIV-1 infected volunteers were studied: Group 1 represented acutely and very-early infected individuals from 2005 to 2011 (IAVI's Protocol C cohort), while Group 2 represented infected individuals with early HIV infection from 2016 to 2019. For Group 1, HIV discordant couples enrolled in a longitudinal prospective prevention study were provided with counseling, condoms and HIV testing of the seronegative partner during the study (Allen et al., 1992). Couples voluntary counseling and testing (CVCT) in high prevalence areas has been shown to reduce transmission incidence of HIV in cohabiting couples by more than two-thirds (Wall et al., 2019). When infection of the seronegative partner was identified as described in methods, they were enrolled in IAVI Protocol C, an acute infection, long-term follow-up study, and samples were obtained from both partners (Price et al., 2020). Originally, 94 volunteers with incident HIV infection were enrolled into Protocol C from Kigali; in the current study, we analyzed plasma viruses from the 26 seroconverters with the shortest estimated time from the date of infection (median time from EDI=23 days) calculated as described in methods; with 16 out of 26 plasma collected less than 30 days post-EDI (Table 2). For Group 2, individuals were identified in collaboration with government clinics generally within 3 months of their last seronegative visit (median time from EDI=91 days; Table 3), so that we could compare the phylogenetics and subtypes of contemporaneous viruses circulating in 2016 through 2019 with those from 2005 to 2011. We analyzed plasma viruses from 21 of these newly infected individuals who included partners of HIV-1 discordant

couples (2), female sex workers (FSW; 13), and young MSM (6; Table 3).

### Severity of Genetic Bottleneck During HIV Transmission in the Protocol C Heterosexual Acute Transmission Cohort

In sub-Saharan Africa, heterosexual transmission remains the predominant mode of infection by HIV-1, and accounts for approximately 75% of newly infected cases worldwide (Monaco et al., 2017). We performed near full-length, single genome amplification (NFLSGA ~9,000 bp) on HIV-1 from the 26 acute plasma samples using a high-fidelity nested PCR approach described previously (Rousseau et al., 2006; Deymier et al., 2014; Yue et al., 2015; Kinloch et al., 2019). A total of 288 PCR amplicons from viral RNA were sequenced with an average of 11 NFLSGAs per individual using a multiplexed, highly accurate, DNA sequencing approach based on the PacBio Sequencing platform combined with the MDPseq work flow we have described previously (Dilernia et al., 2015). Phylogenetic analysis of the near full-length genome sequences showed that in 20 out of 26 Rwandan seroconverters a single virus variant established infection, with sequences for each individual clustering in a very homogeneous rake (Figure 1). In contrast, in six of 26 individuals, infection was initiated by more than one virus variant (Figure 1); as an example, for the acutely infected individual 175,071 three clearly distinct sequence branches were observed in the phylogenetic tree and three different populations were shown by the highlighter analysis (Figure 2).

### Inter-Subtype Recombinants Recognized by Near Full-Length Sequencing

Worldwide HIV-1 Group M is the major source of the global pandemic. In this group there are 9 subtypes, over 100 CRF, and many URF (Hemelaar et al., 2020; LANL, 2021). Globally, the proportions of recombinants has increased over time, reaching almost 23% of all infections in the period 2010 to 2015 (Hemelaar et al., 2020), and URF infections occur frequently in the regions and countries where more than one subtype circulate (Tebit and Arts, 2011). Based on previous subtyping of HIV-1 infection for Protocol C volunteers which employed *pol* gene sequences (Fabiani et al., 1998; Amornkul et al., 2013), we expected that the 26 viruses would comprise approximately 21 Subtype A1 (81%); three subtype C (11%); and two Recombinant (8%, 1 A1/C and 1 C/D). In contrast, utilizing the 9 kb near full-length genome sequences and the programs RIP and jpHMM<sup>1</sup> to detect recombinants, we observed a 3-fold higher percentage of recombinant viruses, with six (5A1/C and 1C/D) recombinants (23%) in addition to 18 subtype A1 (69%), and two subtype C (8%) infections (Figures 1, 3A and Table 2). Each of the recombinant genomes exhibited unique recombination breakpoints (URFs), with two viruses 175,011 and 175,017 resulting from multiple crossovers between their subtype A1 and C progenitors (Figure 3A). A majority of the A1/C recombinants retained a portion of the Env gene and/or

<sup>1</sup>www.hiv.lanl.gov

**TABLE 2** | Transmission variants and subtypes in protocol C individuals.

PCID	Coded ID	Sample date	EDI	VL	Transmission variants	Sequence type	Subtype by NFLG	Subtype by pol
175,019	R49M	30-Mar-05	10	3,000,000	Single	TF	A1	A1
175,042	R463F	9-Mar-07	10	152,000,000	Single	TF	A1	A1
175,079	R3103M	19-Sep-08	13	750,001	Single	TF	A1	A1
175,089	R3584M	15-Dec-09	13	223,440	Single	TF	A1C	A1
175,005	R53F	16-Jun-05	14	2,090,040	Single	TF	A1	A1
175,062	R1135M	11-Feb-08	14	1,256,940	Single	TF	A1	A1
175,065	R269M	24-Mar-08	14	31,700,000	Single	Con	A1	A1
175,072	R3137M	28-May-08	14	9,061,540	Single	Con	C	C
175,053	R254F	8-Aug-07	15	1,876,000	Single	TF	A1C	C
175,090	R3469F	15-Apr-10	15	2,920,000	>2	TF(P1)	A1	A1
175,093	R2302M	9-Jul-10	15	3,760,000	Single	TF	A1	A1
175,094	R3843F	14-Oct-10	16	4,400,000	Single	TF	A1	A1
175,059	R977F	14-Dec-07	17	7,290,000	Single	TF	A1	A1
175,038	R880F	12-Jan-07	21	730,000	Single	TF	A1	A1
175,097	R3894M	18-Mar-11	23	6,934,997	Single	TF	CD	CD
175,092	R3671F	8-Jun-10	25	3,940,000	Single	NA	A1	A1
175,074	R873F	24-Jul-08	37	219,920	2	TF(P1)	A1	A1
175,071	R1077F	12-Jun-08	41	4,702,444	3	TF(P2)	A1	A1
175,008	R50M	25-May-05	45	117,608	>2	NA	C	C
175,020	R57F	12-Oct-05	46	134,472	>2	NA	A1	A1
175,014	R59M	21-Oct-05	46	806,290	Single	TF	A1	A1
175,017	R44M	3-May-05	49	123,560	Single	TF	A1C	A1
175,011	R63M	14-Dec-05	50	184,270	Single	Con	A1C	A1C
175,012	R40F	31-Mar-05	53	1,398,004	Single	TF	A1C	A1
175,027	R72M	10-Aug-06	67	425,000	>2	NA	A1	A1
175,010	R65M	27-Jan-06	73	148,220	Single	TF	A1	A1

PCID, protocol C identification number of the participant; Coded ID, coded identification number used in two previous publications (Haaland et al., 2009; Yue et al., 2015); Sample date, date of sample collection; EDI, time since estimated date of infection; VL, viral load of sample; Transmission variants, number of genetic variants transmitted from partner; Sequence type, sequence defined as TF: transmitted founder virus sequence amplicon identical to consensus identified in NFLG amplicons; (P1), (P2): TF sequence identified in subpopulation 1 or 2 of multiple variants, Con: consensus sequence of NFLG amplicons; Subtype by NFLG: subtype as deduced from the NFLG sequence, Subtype by pol: Subtype defined previously from pol gene sequencing.

**TABLE 3** | Summary of recent infection samples and derived near full-length sequences.

Coded ID	Sample date	EDI	VL	Subtype	Risk group*
BUS71F	10-July-19	91.5	328,000	A1	FSW
GIT84F	1-August-19	131	102,000	A1	FSW
GWE47F	10-October-18	111.5	150,000	A1	FSW
GWE68F	25-June-19	216	120,000	A1	FSW
KAG34F	10-November-17	88	191,000	A1/C/D	FSW
KIN18F	17-March-17	46	113,000	A1	FSW
MAS1F	18-August-16	44.5	553,000	A1	FSW
MAS21F	30-March-17	164.5	856,000	A1	DC
MAS6M	22-November-16	174.5	202,000	A1/C	DC
MAS7F	23-November-16	46	47,200	A1/C	FSW
MAT81F	24-July-19	181	292,000	A1/C/D	FSW
NGA76F	17-July-19	90	141,000	A1/C	FSW
NGA77F	17-July-19	90	473,000	A1/C	FSW
PSF24M	2-June-17	96	306,000	A1/C	MSM
PSF33M	20-October-17	194	160,000	A1/C	MSM
PSF36F	4-January-18	94	702,000	A1/C	FSW
PSF38M	17-September-18	174	145,000	A1/C	MSM
PSF39M	17-February-18	72	100,000	A1/C	MSM
PSF3M	21-October-16	51	86,500	A1/C	MSM
PSF80M	19-July-19	17	683,000	C	MSM
REM29F	24-August-17	91	148,000	A1	FSW

Coded ID, coded identification number of the participant; sample date, date of sample collection; EDI, time since estimated date of infection; VL, viral load of sample; subtype, subtype defined by near full-length genomic sequence; risk group, risk group of the individual, FSW: female sex worker; DC, HIV discordant couple; and MSM, men having sex with men.

Nef from their subtype A parent, but no single region was conserved in all five of these A1/C recombinants. Of these the five A1/C URFs and 1 C/D URF were single virus transmission cases: while one of two subtype C and five of 18 subtype A1 involved multi-variant transmission (Figure 1).

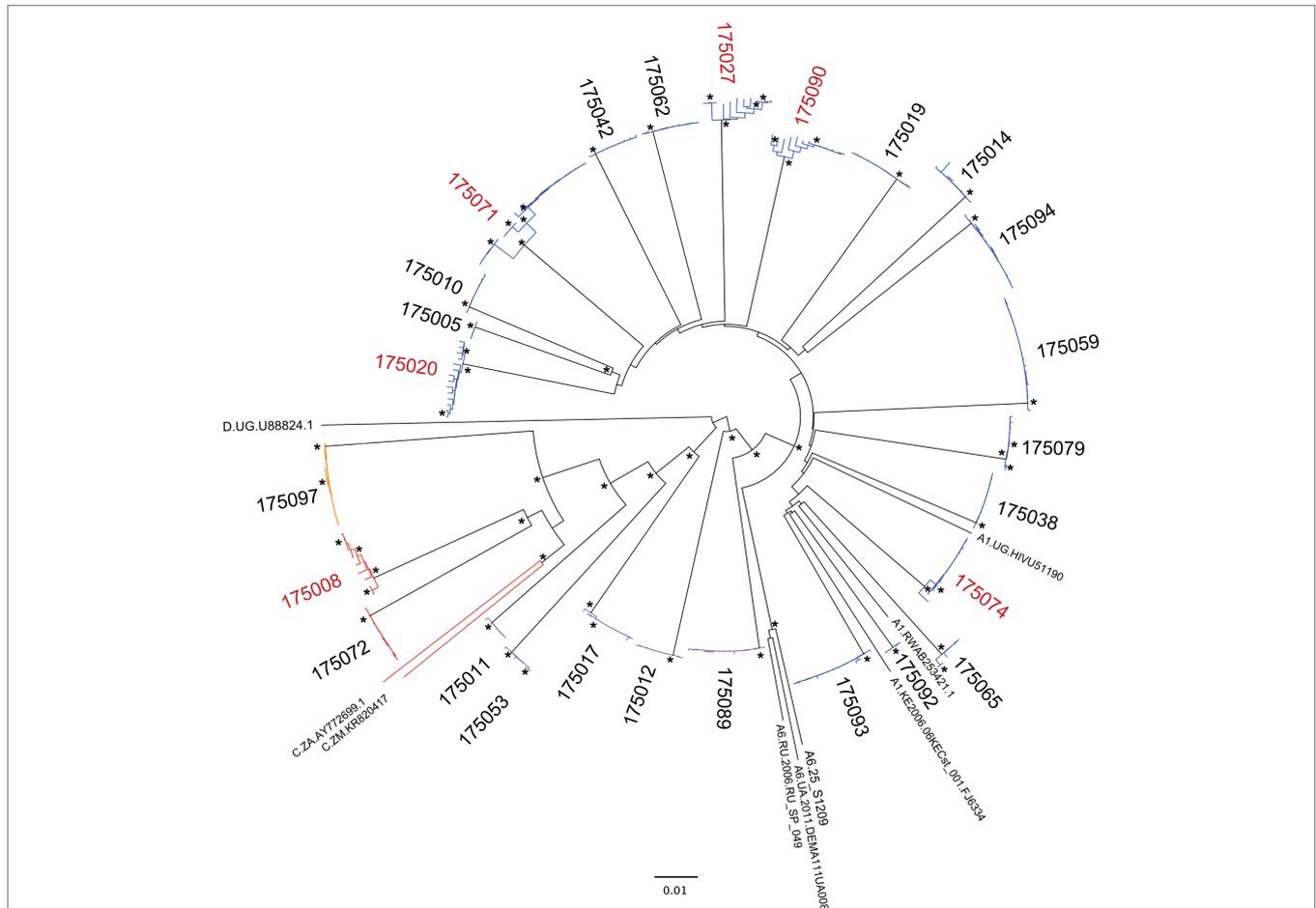
### Near Full-Length Sequencing of Plasma Virus in Recent Seroconverters

Near full-length (9kb) single genome amplification is very inefficient, expensive and time consuming, therefore, for recent samples, we opted to utilize the more efficient population PCR amplification of 5' and 3' half-genome regions, with amplicons that overlapped by ~250 nucleotides. The population amplicons were then sequenced using next-generation PacBio single molecule real time sequencing and individual reads were analyzed using the MDPseq workflow. This allowed us to determine the consensus sequence for the early virus population and breakpoints in those determined to be recombinants. A phylogenetic analysis of the 21 recent seroconversion viruses

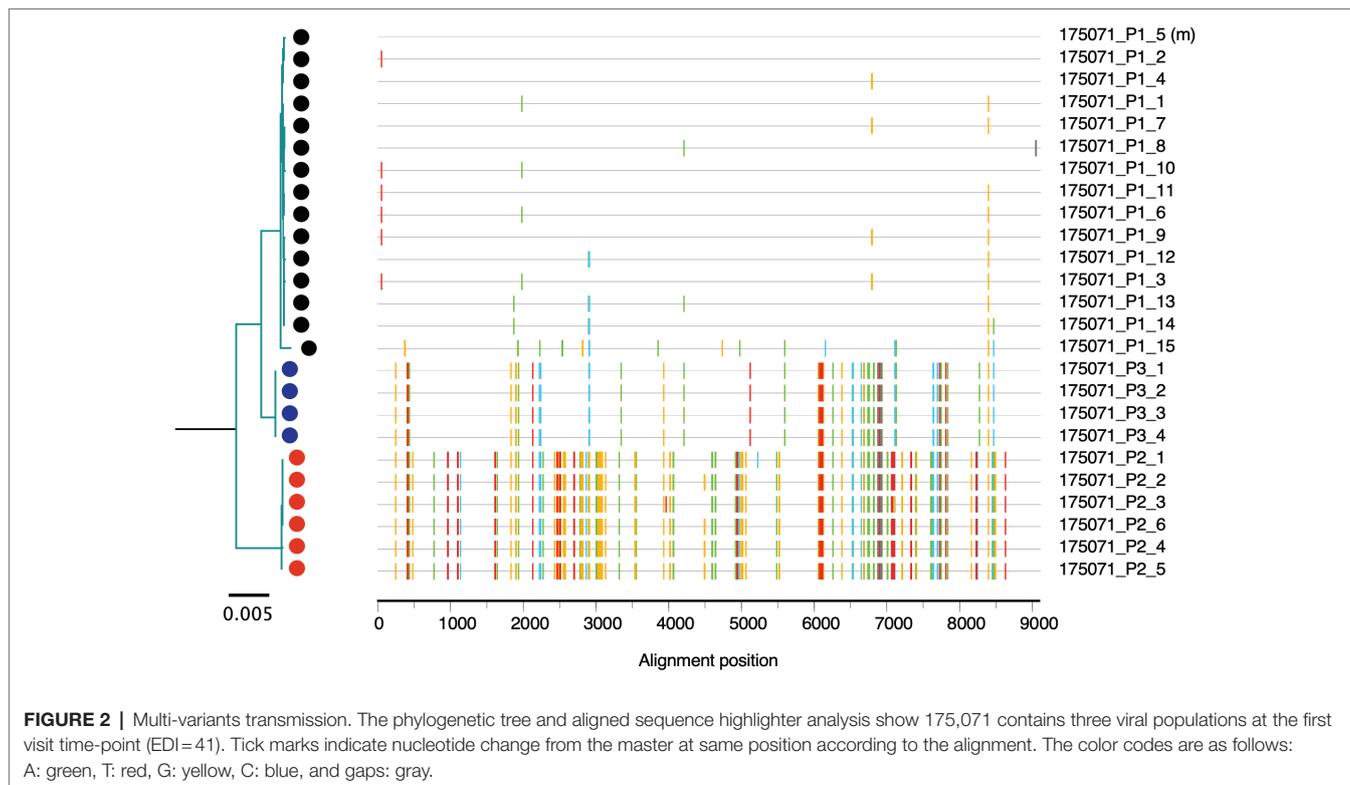
in the context of the 26 viruses from Protocol C (Red; Figure 4) shows that overall, the two viruses from discordant couple transmissions (Blue) and a majority of the viruses from newly infected FSWs (Magenta) clustered within the diversity of the older viral isolates. In contrast five of the six newly infecting viruses from the cohort of young MSM (Cyan) clustered together on the phylogenetic tree with the sequences exhibiting very limited diversity (median 97.5% identity). This would be consistent with a recent transmission network within a risk group that otherwise has demonstrated low sero-incidence (Karita et al., in preparation).

An analysis for recombination using the RIP and jpHMM tools<sup>2</sup> revealed that 12 of the 21 recent seroconversions were initiated by unique A1/C and A1/C/D recombinant forms (Figure 4, denoted by blue circles). These represent, therefore, 57% of the samples analyzed, a significantly higher frequency than we observed in the Protocol C samples (23%;  $p=0.033$ ). Even if we define the five closely related viruses from MSM

<sup>2</sup><https://www.hiv.lanl.gov/content/sequence/HIV/HIVTools.html>



**FIGURE 1 |** Phylogenetic analysis of virus sequences from Protocol C acute infections. Neighbor-joining tree representing the near full-length genome (NFLG) sequences of viruses from 26 acutely infected recipients with reference subtypes A1, A6, C, and D. Subtype and recombinants are indicated with specific colors (Subtype A, Blue; Subtype C, Red; Recombinant AC, Purple; Recombinant CD, Orange). The IDs of subjects infected with more than one viral variant from their partner are highlighted in red. Nodes with bootstrap values >0.9 are denoted with an asterisk.



as a single recombinant, the frequency (47.1%) while no longer significant ( $p=0.182$ ) remains double that of the earlier samples.

Interestingly, while each of the recent recombinants exhibited unique numbers and positions of crossovers (**Figure 3B**), a region of subtype A1 Env, extending from just before an 18-residue amphipathic alpha-helical region located on the outer domain of gp120, known as the alpha 2 helix, to the membrane-spanning domain of gp41, was conserved throughout. A similar, albeit shorter region that terminated in the C-terminal heptad repeat of gp41, was present in four of the five unique A1/C recombinants identified in the Protocol C samples (**Figure 3A**).

### Identification of Antiretroviral Drug Resistance Mutations

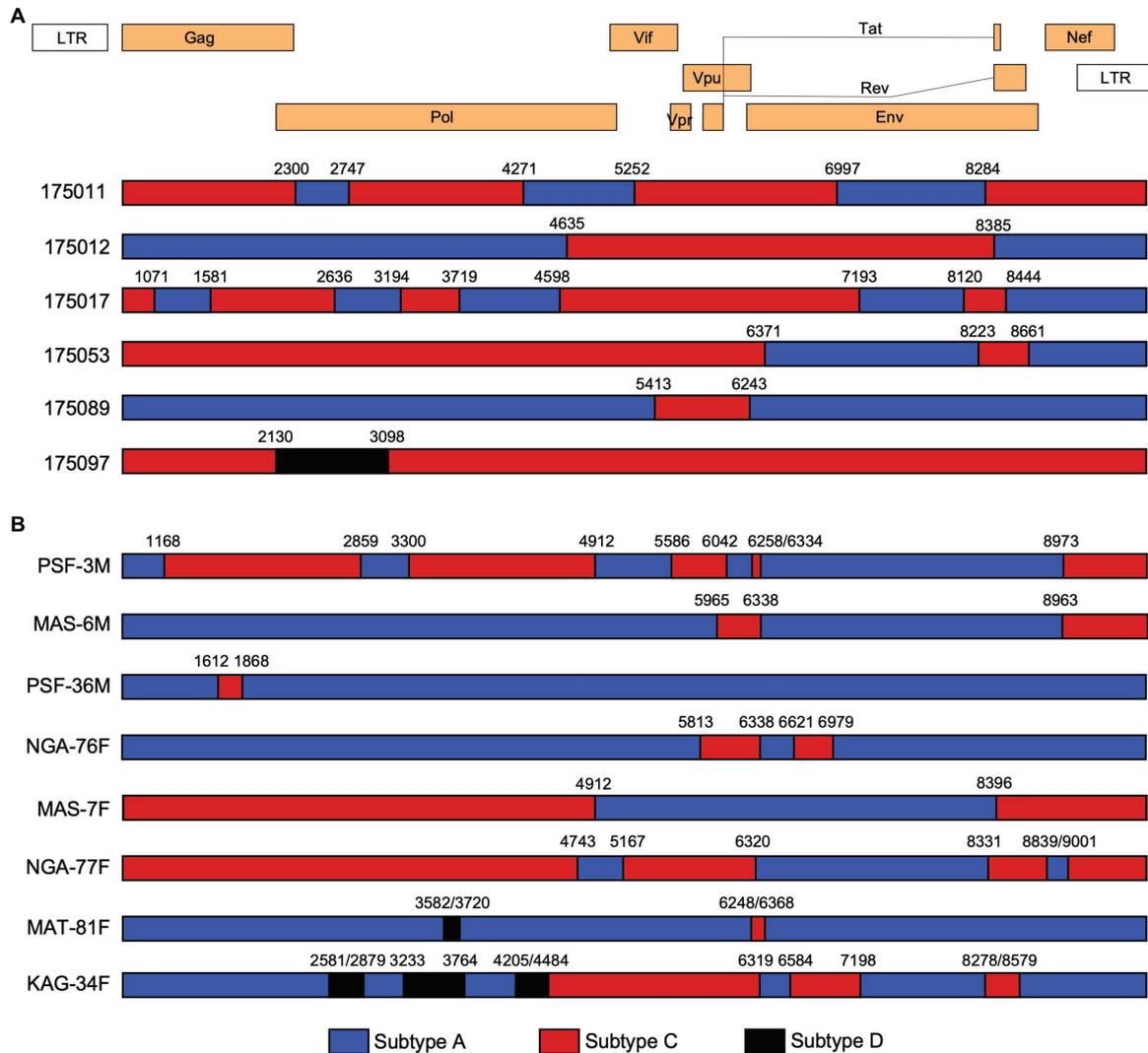
Finally, in order to determine whether, in the context of increased availability of ART, transmitted drug resistance was increasing, sequences encompassing the protease, reverse transcriptase, and integrase were extracted from all 47 near full-length genomes and submitted to the Stanford HIVdb analysis program, and any major drug resistance mutations (DRMs) documented. None of the Protocol C viral sequences analyzed for this study encoded any major DRMs. In contrast, four of the 21 recent isolates encoded DRMs in the reverse transcriptase. MAS-21F, MAS-1F, and PSF-80M encoded the K103N mutation that confers high level resistance to the non-nucleoside inhibitors nevirapine (NVP) and efavirenz (EFV). NGA-77F encoded K103S, which confers high level resistance to NVP and intermediate resistance to EFV, and

E138A, which confers low level resistance to etravirine and rilpivirine.

## DISCUSSION

Here, we report on the NFLG sequences of HIV from 47 acute and early infected Rwandans. This more than quadruples the number of near full-length sequences previously reported for this East African country (Lamers et al., 2016). For the IAVI Protocol C samples, we amplified more than 10 NFLG single-genome amplicons per individual, a total of 288, from acute infections that allowed us to define a TF virus sequence in a majority of cases. This was facilitated by the use of PacBio single molecule real-time (SMRT) long-read sequencing combined with the MDPseq workflow, to define accurate sequences (Dilernia et al., 2015). We utilized the same workflow for the more efficient, overlapping half-genome population amplicons generated from recent infections. The long-read technology allowed the population PCR to be deconvoluted to yield the individual sequences of the corresponding amplicons in order to generate a consensus sequence of the infecting virus.

We have previously reported, based on sequencing of the V1-V4 region of Env that in a majority of transmission events in both a Zambian discordant couple cohort and the Rwandan Protocol C cohort infection was established by a single genetic variant from the transmitting partner (Derdeyn et al., 2004; Haaland et al., 2009). In general, multiple variant infections represented only 10–15% of the transmission pairs examined

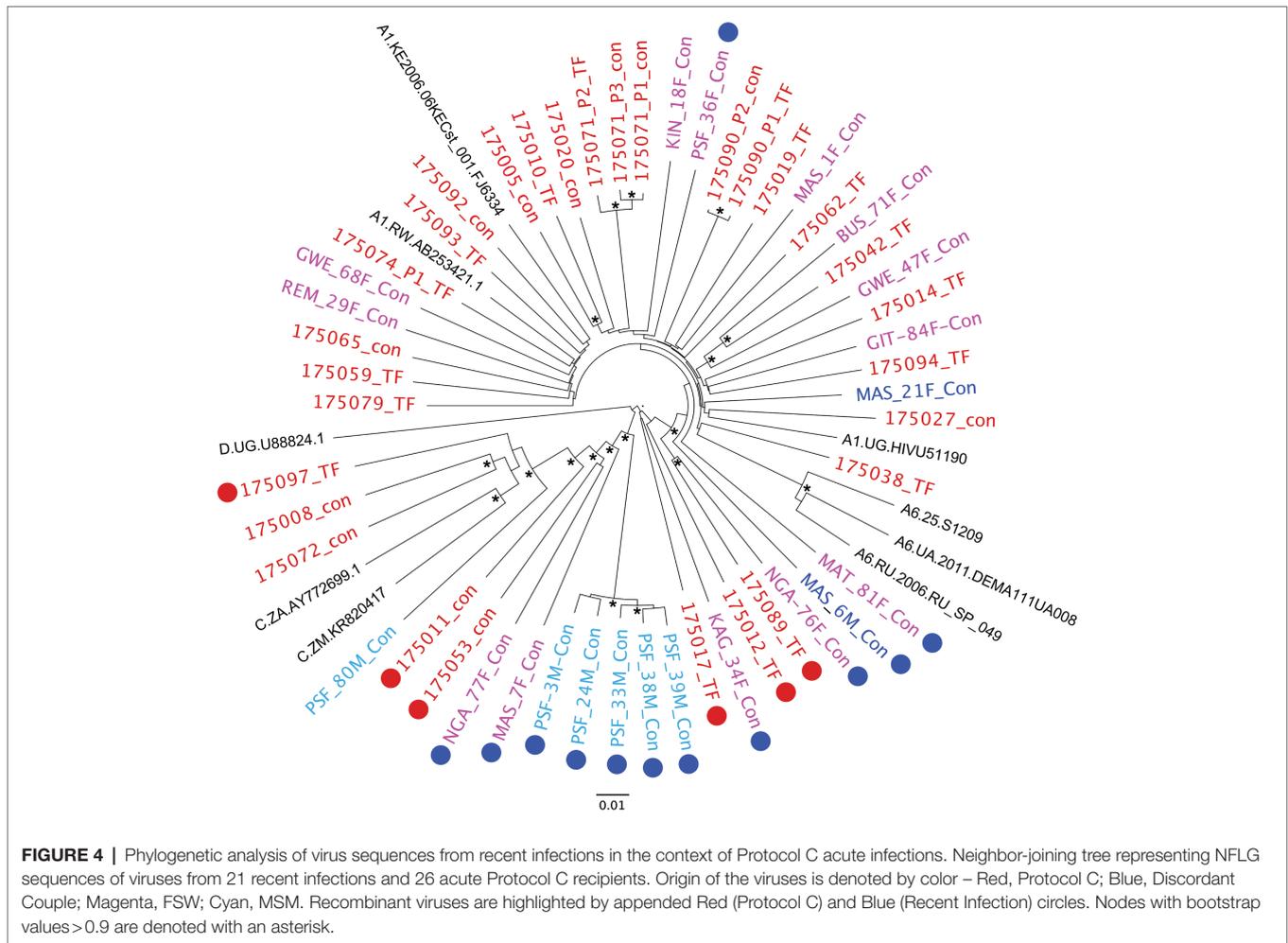


**FIGURE 3 |** Recombination analysis of Protocol C and recent infections. The positions of recombination cross-overs and the subtype origin of genomic regions are shown based on the jphMM analytical tool. Nucleotide positions refer to the equivalent positions on the HXB-2 HIV-1 genome. **(A)** The six unique recombinant forms (URFs) from the Protocol C acute infection cohort. **(B)** The recombination structure of recent Rwandan seroconverter viruses. PSF-3M illustrates the identical recombinant structure found in all five early isolates from MSM (PSF-3M, -24M, -33M, -38M, and -39M). Subtype sequences are denoted by the same colors as in **Figure 1**.

(Haaland et al., 2009). In the current study, we observed a somewhat higher frequency (23%) of infections initiated by more than one partner-derived virus variant, similar to that reported in a South African women cohort (Abrahams et al., 2009) and a predominantly MSM cohort (Keele et al., 2008). We have reported previously that evidence of genital inflammation and ulcers can lower the barrier to transmission and increase multi-variant transmission (Haaland et al., 2009; Carlson et al., 2014), and a recent study in a Kenyan MSM cohort, where sexually transmitted infections were common reported 39% (15 out of 38) of the participants were infected with multiple founder viruses (Macharia et al., 2020). In the heterosexual protocol C cohort studied here we observed evidence of genital inflammation or ulcers in just six individuals in the 6 months

prior to HIV-1 infection and these individuals were equally distributed between the single and multiple variant infections groups ( $p=0.67$ ; Chi-square with Yates correction).

Inter-subtype recombination represents a significant challenge to HIV-1 vaccine design since it provides the virus with a mechanism to rapidly diversify following co-infection of an individual with two different subtype viruses (Robertson et al., 1995). This gives rise to unique recombinant forms that can be prevalent in regions where more than one subtype circulates. An example of this is Uganda, Rwanda's neighbor, where subtypes A and D cocirculate, and recent studies have identified a frequency of recombinants between the two subtypes as high as 49% (Lee et al., 2017; Grant et al., 2020). In contrast, a previous



analysis of viral subtypes from 30 women in Rwanda in 2013 reported 80% subtype A1, 3% subtype C and D, and 13% AC or AD recombinant forms based on sequencing of *gag*, *pol*, and *env* (Kemal et al., 2013). Similarly, an earlier subtype analysis of 92 IAVI Protocol C incident infections in Rwanda, which was based on a region encompassing the *pol* gene, defined 80% as subtype A1 and only 6.5% as recombinant viruses (Amornkul et al., 2013). By contrast, using NFLG sequences, we identified 23% (6/26) recombinants, a much higher frequency than even the 8% (2/26) previously defined for these same individuals through *pol* sequencing. This highlights the importance of NFLG sequencing to fully understand the complexity of virus populations circulating in multi-subtype countries. Indeed, sequencing of 21 more recent (2016–2019) incident infection viruses from different high-risk groups suggests that recombinant viruses are increasing in frequency, since we observed that 57% of these recent infections were A1/C, or A1/C/D recombinants. With the exception of the five viruses that appeared to represent a recent viral transmission network, the recombinants in these recent infections and those in Protocol C resulted from a series of unique recombination events. In contrast,

five of the six viruses from the MSM cohort were highly related and had identical recombination patterns. Of interest, one common recombination breakpoint (4912) was shared by this MSM group (represented by PSF-3M, **Figure 3B**) and its nearest neighbor MAS-7F (**Figure 4**) raising the possibility that the former evolved from the latter following further recombination events. Although, the risk groups in Protocol C and the recent infection cohorts are different, the majority of infections in the latter represented heterosexual transmission in high-risk women and discordant couple partners (15/21), where recombinant viruses remained prevalent (47%). Nevertheless, this difference in risk groups should be considered a potential weakness of the comparison.

Although, a majority of the recombinant viruses from the recent seroconverters exhibited unique recombination breakpoints, they all retained a common region of Env that was derived from a subtype A1 parent. This region extended minimally in KAG-34F from the last few residues of the third variable (V3) loop of gp120, just before the hydrophobic alpha 2 helix, to the beginning of the membrane spanning domain of gp41. We have shown previously, that, in subtype C viruses, the alpha 2 helix is under positive selection pressure and that

variations in this region are in part linked to early neutralizing antibody escape (Rong et al., 2007). The conserved region also spans the fifth conserved domain (C5) of gp120 and the ectodomain of gp41, both critical for gp120 and gp41 interactions and trimer stability (Binley et al., 2000; Julien et al., 2013). Thus it is possible that this region from subtype A1 provides a fitness advantage to the recombinants.

We have recently reported on the presence of DRMs in the transmitted virus of newly infected partners of Rwandan couples, where the transmitting partner was on ART but carried drug resistant virus (Woodson et al., 2018). It was of interest, therefore, to compare the prevalence and nature of DRMs in the Protocol C cohort, from a time when antiretroviral drugs were not widely available, to those in more recently infected individuals, when treatment with the standard first-line combination of Tenofovir (TDF), Lamivudine (3TC), and EFV following diagnosis was routine. Although none of the 26 acutely infected individuals from Protocol C encoded DRMs, a previous study of 78 Rwandans from the same cohort did identify five individuals with NNRTI mutations (three with K103N, two with L100I) and one with the protease inhibitor mutation, M46L (Price et al., 2011). Our finding that four of the 21 recent infection viruses encoded K103N/S, which confers resistance to NVP and EFV, indicates that this mutation is becoming more prevalent ( $p=0.035$ ) within the population. Moreover, since all of these samples were collected prior to the initiation of ART, for the three females whose virus encoded this mutation, the resistance mutations must have been present in the virus of their male partners.

Overall, we demonstrate here, through NFLG sequence analysis that recombinant viruses are more prevalent in Rwanda than previously reported, and that the frequency of both recombinants and NNRTI DRMs appear to have increased between two sampling periods. It will be critical therefore to continue to monitor the nature of circulating viruses to ensure the validity of ongoing vaccine development efforts.

## DATA AVAILABILITY STATEMENT

NFL sequences were submitted to GenBank. The accession numbers for the 26 Protocol C derived viruses are JX236678.1, JX236677.1, MT942708-MT942972. Those for the 21 recent seroconverters are MZ642260-MZ642280.

## REFERENCES

- Abrahams, M. R., Anderson, J. A., Giorgi, E. E., Seoighe, C., Mlisana, K., Ping, L. H., et al. (2009). Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J. Virol.* 83, 3556–3567. doi: 10.1128/JVI.02132-08
- Allen, S., Tice, J., Van De Perre, P., Serufilira, A., Hudes, E., Nsengumuremyi, E., et al. (1992). Effect of serotesting with counselling on condom use and seroconversion among HIV discordant couples in Africa. *BMJ* 304, 1605–1609. doi: 10.1136/bmj.304.6842.1605
- Amornkul, P. N., Karita, E., Kamali, A., Rida, W. N., Sanders, E. J., Lakhi, S., et al. (2013). Disease progression by infecting HIV-1 subtype in a seroconverter cohort in sub-Saharan Africa. *AIDS* 27, 2775–2786. doi: 10.1097/QAD.0000000000000012
- Binley, J. M., Sanders, R. W., Clas, B., Schuelke, N., Master, A., Guo, Y., et al. (2000). A recombinant human immunodeficiency virus type 1 envelope glycoprotein complex stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits is an antigenic mimic of the trimeric virion-associated structure. *J. Virol.* 74, 627–643. doi: 10.1128/JVI.74.2.627-643.2000
- Brumme, Z. L., and Walker, B. D. (2009). Tracking the culprit: HIV-1 evolution and immune selection revealed by single-genome amplification. *J. Exp. Med.* 206, 1215–1218. doi: 10.1084/jem.20091094

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Rwanda National Ethics Committee and Emory University Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

LY, EH, JG, SA, and EK conceived and designed the experiments. GU, EM, EN, RX, DD, KH, HS, and QQ performed the experiments. LY, GU, EM, and EN analyzed the data. LY, EH, PF, JH, JB, EK, JG, and SA contributed reagents, materials, and analysis. EH and LY wrote the paper. All authors contributed to the article and approved the submitted version.

## FUNDING

This work is made possible in part by NIH grant R01AI051231 and by IAVI contract # 06-646-9933 through the support of the American People through the United States President's Emergency Plan for AIDS Relief (PEPFAR) through the United States Agency for International Development (USAID). The contents of this manuscript are the sole responsibility of the authors and do not necessarily reflect the views of PEPFAR, USAID, or the United States Government. This work was also supported, in part, by the Virology Core at the Emory Center for AIDS Research by performing viral load determinations (grant P30 AI050409); the Yerkes National Primate Research Center base grant through the Office of Research Infrastructure Programs /OD P51OD11132. EH is a Georgia Eminent Scholar.

## ACKNOWLEDGMENTS

The authors thank all the volunteers in Rwanda who participated in this study and all the staff in the Center for Family Health Research who made this study possible. The authors would like to thank Jon Allen and Paul Farmer for technical assistance, sample management, and database management, and Pat Fast and Matt Price for thoughtful comments on the manuscript.

- Butler, I. F., Pandrea, I., Marx, P. A., and Apetrei, C. (2007). HIV genetic diversity: biological and public health consequences. *Curr. HIV Res.* 5, 23–45. doi: 10.2174/15701620779316297
- Carlson, J. M., Brumme, C. J., Martin, E., Listgarten, J., Brockman, M. A., Le, A. Q., et al. (2012). Correlates of protective cellular immunity revealed by analysis of population-level immune escape pathways in HIV-1. *J. Virol.* 86, 13202–13216. doi: 10.1128/JVI.01998-12
- Carlson, J. M., Du, V. Y., Pfeifer, N., Bansal, A., Tan, V. Y., Power, K., et al. (2016). Impact of pre-adapted HIV transmission. *Nat. Med.* 22, 606–613. doi: 10.1038/nm.4100
- Carlson, J. M., Schaefer, M., Monaco, D. C., Batorsky, R., Claiborne, D. T., Prince, J., et al. (2014). HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science* 345:1254031. doi: 10.1126/science.1254031
- Crawford, H., Lumm, W., Leslie, A., Schaefer, M., Boeras, D., Prado, J. G., et al. (2009). Evolution of HLA-B\*5703 HIV-1 escape mutations in HLA-B\*5703-positive individuals and their transmission recipients. *J. Exp. Med.* 206, 909–921. doi: 10.1084/jem.20081984
- Delatorre, E. O., and Bello, G. (2012). Phylodynamics of HIV-1 subtype C epidemic in East Africa. *PLoS One* 7:e41904. doi: 10.1371/journal.pone.0041904
- Derdeyn, C. A., Decker, J. M., Bibollet-Ruche, F., Mokili, J. L., Muldoon, M., Denham, S. A., et al. (2004). Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* 303, 2019–2022. doi: 10.1126/science.1093137
- Desire, N., Cerutti, L., Le Hingrat, Q., Perrier, M., Emler, S., Calvez, V., et al. (2018). Characterization update of HIV-1 M subtypes diversity and proposal for subtypes A and D sub-subtypes reclassification. *Retrovirology* 15:80. doi: 10.1186/s12977-018-0461-y
- Deymier, M. J., Claiborne, D. T., Ende, Z., Ratner, H. K., Kilembe, W., Allen, S., et al. (2014). Particle infectivity of HIV-1 full-length genome infectious molecular clones in a subtype C heterosexual transmission pair following high fidelity amplification and unbiased cloning. *Virology* 468–470, 454–461. doi: 10.1016/j.virol.2014.08.018
- Dilernia, D. A., Chien, J. T., Monaco, D. C., Brown, M. P., Ende, Z., Deymier, M. J., et al. (2015). Multiplexed highly-accurate DNA sequencing of closely-related HIV-1 variants using continuous long reads from single molecule, real-time sequencing. *Nucleic Acids Res.* 43:e129. doi: 10.1093/nar/gkv630
- Fabiani, M., Blè, C., Grivel, P., Lukwiya, M., and DeClích, S. (1998). 1989–1996 HIV-1 prevalence trends among different risk groups in Gulu District, North Uganda. *J. Acquir. Immune Defic. Syndr. Hum. Retrovir.* 18:514. doi: 10.1097/00042560-199808150-00015
- Giovanetti, M., Ciccozzi, M., Parolin, C., and Borsetti, A. (2020). Molecular epidemiology of HIV-1 in African countries: a comprehensive overview. *Pathogens* 9:1072. doi: 10.3390/pathogens9121072
- Grant, H. E., Hodcroft, E. B., Ssemwanga, D., Kitayimbwa, J. M., Yebra, G., Esquivel Gomez, L. R., et al. (2020). Pervasive and non-random recombination in near full-length HIV genomes from Uganda. *Virus Evol.* 6:veaa004. doi: 10.1093/ve/veaa004
- Haaland, R. E., Hawkins, P. A., Salazar-Gonzalez, J., Johnson, A., Tichacek, A., Karita, E., et al. (2009). Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS Pathog.* 5:e1000274. doi: 10.1371/journal.ppat.1000274
- Hamers, R. L., Sigaloff, K. C., Wensing, A. M., Wallis, C. L., Kityo, C., Siwale, M., et al. (2012). Patterns of HIV-1 drug resistance after first-line antiretroviral therapy (ART) failure in 6 sub-Saharan African countries: implications for second-line ART strategies. *Clin. Infect. Dis.* 54, 1660–1669. doi: 10.1093/cid/cis254
- Hauser, A., Kusejko, K., Johnson, L. F., Wandeler, G., Riou, J., Goldstein, F., et al. (2019). Bridging the gap between HIV epidemiology and antiretroviral resistance evolution: modelling the spread of resistance in South Africa. *PLoS Comput. Biol.* 15:e1007083. doi: 10.1371/journal.pcbi.1007083
- Hemelaar, J., Elangovan, R., Yun, J., Dickson-Tetteh, L., Kirtley, S., Gouws-Williams, E., et al. (2020). Global and regional epidemiology of HIV-1 recombinants in 1990–2015: a systematic review and global survey. *Lancet HIV* 7, e772–e781. doi: 10.1016/S2352-3018(20)30252-6
- Julien, J. P., Cupo, A., Sok, D., Stanfield, R. L., Lyumkis, D., Deller, M. C., et al. (2013). Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* 342, 1477–1483. doi: 10.1126/science.1245625
- Kawashima, Y., Pfafferott, K., Frater, J., Matthews, P., Payne, R., Addo, M., et al. (2009). Adaptation of HIV-1 to human leukocyte antigen class I. *Nature* 458, 641–645. doi: 10.1038/nature07746
- Keele, B. F., Giorgi, E. E., Salazar-Gonzalez, J. F., Decker, J. M., Pham, K. T., Salazar, M. G., et al. (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* 105, 7552–7557. doi: 10.1073/pnas.0802203105
- Kemal, K. S., Anastos, K., Weiser, B., Ramirez, C. M., Shi, Q., and Burger, H. (2013). Molecular epidemiology of HIV type 1 subtypes in Rwanda. *AIDS Res. Hum. Retrovir.* 29, 957–962. doi: 10.1089/aid.2012.0095
- Kinloch, N. N., Lee, G. Q., Carlson, J. M., Jin, S. W., Brumme, C. J., Byakwaga, H., et al. (2019). Genotypic and mechanistic characterization of subtype-specific HIV adaptation to host cellular immunity. *J. Virol.* 93, e01502–e01518. doi: 10.1128/JVI.01502-18
- Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., and Detours, V. (2001). Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull.* 58, 19–42. doi: 10.1093/bmb/58.1.19
- Lamers, S. L., Barbier, A. E., Ratmann, O., Fraser, C., Rose, R., Laeyendecker, O., et al. (2016). HIV-1 sequence data coverage in Central East Africa from 1959 to 2013. *AIDS Res. Hum. Retrovir.* 32, 904–908. doi: 10.1089/aid.2016.0079
- Lanl (2021). HIV Circulating Recombinant Forms (CRFs) [Online]. Los Alamos National Laboratory. Available at: <https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html> (Accessed Aug 24, 2021).
- Lau, K. A., and Wong, J. J. (2013). Current trends of HIV recombination worldwide. *Infect. Dis. Rep.* 5:e4. doi: 10.4081/idr.2013.s1.e4
- Lee, G. Q., Bangsberg, D. R., Mo, T., Lachowski, C., Brumme, C. J., Zhang, W., et al. (2017). Prevalence and clinical impacts of HIV-1 intersubtype recombinants in Uganda revealed by near-full-genome population and deep sequencing approaches. *AIDS* 31, 2345–2354. doi: 10.1097/QAD.0000000000001619
- Lukashov, V. V., and Goudsmit, J. (1998). HIV heterogeneity and disease progression in AIDS: a model of continuous virus adaptation. *AIDS* 12(Suppl. A), S43–S52.
- Macharia, G. N., Yue, L., Staller, E., Dilernia, D., Wilkins, D., Song, H., et al. (2020). Infection with multiple HIV-1 founder variants is associated with lower viral replicative capacity, faster CD4+ T cell decline and increased immune activation during acute infection. *PLoS Pathog.* 16:e1008853. doi: 10.1371/journal.ppat.1008853
- Mansky, L. M. (1998). Retrovirus mutation rates and their role in genetic variation. *J. Gen. Virol.* 79, 1337–1345. doi: 10.1099/0022-1317-79-6-1337
- Mccutchan, F. E., Salminen, M. O., Carr, J. K., and Burke, D. S. (1996). HIV-1 genetic diversity. *AIDS* 10(Suppl. 3), S13–S20.
- Monaco, D. C., Dilernia, D. A., Fiore-Gartland, A., Yu, T., Prince, J. L., Dennis, K. K., et al. (2016). Balance between transmitted HLA preadapted and nonassociated polymorphisms is a major determinant of HIV-1 disease progression. *J. Exp. Med.* 213, 2049–2063. doi: 10.1084/jem.20151984
- Monaco, D. C., Ende, Z., and Hunter, E. (2017). Virus-host gene interactions define HIV-1 disease progression. *Curr. Top. Microbiol. Immunol.* 407, 31–63. doi: 10.1007/82\_2017\_33
- Moore, C. B., John, M., James, I. R., Christiansen, F. T., Witt, C. S., and Mallal, S. A. (2002). Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* 296, 1439–1443. doi: 10.1126/science.1069660
- Price, M. A., Kilembe, W., Ruzagira, E., Karita, E., Inambao, M., Sanders, E. J., et al. (2020). Cohort profile: IAVI's HIV epidemiology and early infection cohort studies in Africa to support vaccine discovery. *Int. J. Epidemiol.* 50, 29–30. doi: 10.1093/ije/dyaa100
- Price, M. A., Wallis, C. L., Lakhi, S., Karita, E., Kamali, A., Anzala, O., et al. (2011). Transmitted HIV type 1 drug resistance among individuals with recent HIV infection in East and Southern Africa. *AIDS Res. Hum. Retrovir.* 27, 5–12. doi: 10.1089/aid.2010.0030
- Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., et al. (2000). HIV-1 nomenclature proposal. *Science* 288, 55–56. doi: 10.1126/science.288.5463.55d
- Robertson, D. L., Sharp, P. M., Mccutchan, F. E., and Hahn, B. H. (1995). Recombination in HIV-1. *Nature* 374, 124–126. doi: 10.1038/374124b0
- Rodgers, M. A., Wilkinson, E., Vallari, A., McArthur, C., Sthreshley, L., Brennan, C. A., et al. (2017). Sensitive next-generation sequencing method

- reveals deep genetic diversity of HIV-1 in the Democratic Republic of the Congo. *J. Virol.* 91, e01841–e01816. doi: 10.1128/JVI.01841-16
- Rong, R., Gnanakaran, S., Decker, J. M., Bibollet-Ruche, F., Taylor, J., Sfakianos, J. N., et al. (2007). Unique mutational patterns in the envelope alpha 2 amphipathic helix and acquisition of length in gp120 hypervariable domains are associated with resistance to autologous neutralization of subtype C human immunodeficiency virus type 1. *J. Virol.* 81, 5658–5668. doi: 10.1128/JVI.00257-07
- Rousseau, C. M., Birditt, B. A., McKay, A. R., Stoddard, J. N., Lee, T. C., McLaughlin, S., et al. (2006). Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. *J. Virol. Methods* 136, 118–125. doi: 10.1016/j.jviromet.2006.04.009
- Song, H., Giorgi, E. E., Ganasov, V. V., Cai, F., Athreya, G., Yoon, H., et al. (2018). Tracking HIV-1 recombination to resolve its contribution to HIV-1 evolution in natural infection. *Nat. Commun.* 9:1928. doi: 10.1038/s41467-018-04217-5
- Tebit, D. M., and Arts, E. J. (2011). Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect. Dis.* 11, 45–56. doi: 10.1016/S1473-3099(10)70186-9
- UNAIDS (2020). Global HIV & AIDS statistics — 2020 Fact Sheet [Online]. <https://www.unaids.org/en/resources/fact-sheet> (Accessed April 05, 2021).
- Wall, K. M., Inambao, M., Kilembe, W., Karita, E., Vwalika, B., Mulenga, J., et al. (2019). HIV testing and counselling couples together for affordable HIV prevention in Africa. *Int. J. Epidemiol.* 48, 217–227. doi: 10.1093/ije/dyy203
- Woodson, E., Goldberg, A., Michelo, C., Basu, D., Tao, S., Schinazi, R., et al. (2018). HIV transmission in discordant couples in Africa in the context of antiretroviral therapy availability. *AIDS* 32, 1613–1623. doi: 10.1097/QAD.0000000000001871
- Yamaguchi, J., Vallari, A., McArthur, C., Streshley, L., Cloherty, G. A., Berg, M. G., et al. (2020). Brief report: complete genome sequence of CG-0018a-01 establishes HIV-1 subtype L. *J. Acquir. Immune Defic. Syndr.* 83, 319–322. doi: 10.1097/QAI.0000000000002246
- Yue, L., Pfafferoth, K. J., Baalwa, J., Conrod, K., Dong, C. C., Chui, C., et al. (2015). Transmitted virus fitness and host T cell responses collectively define divergent infection outcomes in two HIV-1 recipients. *PLoS Pathog.* 11:e1004565. doi: 10.1371/journal.ppat.1004565

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Umviligihozo, Muok, Nyirimihigo Gisa, Xu, Dilernia, Herard, Song, Qin, Bizimana, Farmer, Hare, Gilmour, Allen, Karita, Hunter and Yue. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.