

# 1 **Increasing intra- and inter-subtype HIV diversity despite declining** 2 **HIV incidence in Uganda**

3

4 Seungwon Kim<sup>1</sup>, Godfrey Kigozi<sup>2</sup>, Michael A. Martin<sup>1</sup>, Ronald M. Galiwango<sup>2</sup>, Thomas C.  
5 Quinn<sup>2,3,4</sup>, Andrew D. Redd<sup>3,4,5</sup>, Robert Ssekubugu<sup>2</sup>, David Bonsall<sup>6,7</sup>, Deogratius Ssemwanga<sup>8,9</sup>,  
6 Andrew Rambaut<sup>10</sup>, Joshua T. Herbeck<sup>11</sup>, Steven J. Reynolds<sup>2,3,4</sup>, Brian Foley<sup>12</sup>, Lucie Abeler-  
7 Dörner<sup>6,13</sup>, Christophe Fraser<sup>6,13</sup>, Oliver Ratmann<sup>14</sup>, Joseph Kagaayi<sup>2,15</sup>, Oliver  
8 Laeyendecker<sup>3,4,16\*</sup>, M. Kate Grabowski<sup>1,2,16\*</sup>

9

10 <sup>1</sup>Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD, USA

11 <sup>2</sup>Rakai Health Sciences Program, Kalisizo, Uganda

12 <sup>3</sup>Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA

13 <sup>4</sup>National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda,  
14 MD, USA

15 <sup>5</sup>Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town,  
16 South Africa

17 <sup>6</sup>Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield  
18 Department of Medicine, University of Oxford, Oxford, UK

19 <sup>7</sup>Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

20 <sup>8</sup>Medical Research Council (MRC)/Uganda Virus Research Institute (UVRI) and London School  
21 of Hygiene and Tropical Medicine (LSHTM) Uganda Research Unit, Entebbe, Uganda

22 <sup>9</sup>Uganda Virus Research Institute, Entebbe, Uganda

23 <sup>10</sup>Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

24 <sup>11</sup>Department of Global Health, University of Washington, Seattle, WA, USA

25 <sup>12</sup>Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, New  
26 Mexico, USA

27 <sup>13</sup>Pandemic Sciences Institute, University of Oxford, Oxford, UK

28 <sup>14</sup>Department of Mathematics, Imperial College London, London, England, United Kingdom

29 <sup>15</sup>Makerere University School of Public Health, Kampala, Uganda

30 <sup>16</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore,  
31 MD, USA

32 \*These authors contributed equally.

33

34 The address of the corresponding author: Seungwon Kim, 600 N. Wolfe Street, Suite 446 D2,  
35 Baltimore, Maryland 21205 USA (skim519@jh.edu)

## 36 Abstract

37 HIV incidence has been declining in Africa with scale-up of HIV interventions. However,  
38 there is limited data on HIV evolutionary trends in African populations with waning epidemics.  
39 We evaluated changes in HIV viral diversity and genetic divergence in southern Uganda over a  
40 twenty-five-year period spanning the introduction and scale-up of HIV prevention and treatment  
41 programs using HIV sequence and survey data from the Rakai Community Cohort Study, an  
42 open longitudinal population-based HIV surveillance cohort. *Gag* (p24) and *env* (gp41) HIV data  
43 were generated from persons living with HIV (PLHIV) in 31 inland semi-urban trading and  
44 agrarian communities (1994 to 2018) and four hyperendemic Lake Victoria fishing communities  
45 (2011 to 2018) under continuous surveillance. HIV subtype was assigned using the  
46 Recombination Identification Program with phylogenetic confirmation. Inter-subtype diversity  
47 was estimated using the Shannon diversity index and intra-subtype diversity with the nucleotide  
48 diversity and pairwise TN93 genetic distance. Genetic divergence was measured using root-to-tip  
49 distance and pairwise TN93 genetic distance analyses. Evolutionary dynamics were assessed  
50 among demographic and behavioral sub-groups, including by migration status. 9,931 HIV  
51 sequences were available from 4,999 PLHIV, including 3,060 and 1,939 persons residing in  
52 inland and fishing communities, respectively. In inland communities, subtype A1 viruses  
53 proportionately increased from 14.3% in 1995 to 25.9% in 2017 ( $p < 0.001$ ), while those of  
54 subtype D declined from 73.2% in 1995 to 28.2% in 2017 ( $p < 0.001$ ). The proportion of viruses  
55 classified as recombinants significantly increased by more than four-fold. Inter-subtype HIV  
56 diversity has generally increased. While p24 intra-subtype genetic diversity and divergence  
57 leveled off after 2014, diversity and divergence of gp41 increased through 2017. Inter- and intra-  
58 subtype viral diversity increased across all population sub-groups, including among individuals

59 with no recent migration history or extra-community sexual partners. This study provides  
60 insights into population-level HIV evolutionary dynamics in declining African HIV epidemics  
61 following the scale-up of HIV prevention and treatment programs. Continued molecular  
62 surveillance may provide a better understanding of the dynamics driving population HIV  
63 evolution and yield important insights for epidemic control and vaccine development.

## 64 Introduction

65 Human immunodeficiency virus type 1 (HIV-1) remains a significant public health threat.  
66 In 2022, an estimated 1.3 million new HIV cases and 630,000 AIDS-related deaths occurred  
67 globally, of which more than one-third were in eastern and southern Africa (UNAIDS, 2023).  
68 Currently, there is no HIV vaccine to prevent HIV acquisition, partly due to extensive virus  
69 genetic diversity (Barouch & Korber, 2010; Haynes et al., 2016; Ng'uni et al., 2020). Within the  
70 predominant HIV-1 group M viruses, there exist seventeen distinct HIV-1 subtypes and sub-  
71 subtypes, and numerous inter and intra-subtype recombinant forms (LANL, 2023; Mendes Da  
72 Silva et al., 2021; B. S. Taylor et al., 2008; Yamaguchi et al., 2020). Similar to the global burden  
73 of cases and deaths, HIV genetic diversity is greatest in Africa (Hemelaar et al., 2019; Lihana et  
74 al., 2012; Peeters et al., 2003; Vidal et al., 2000). Thus, an effective preventative vaccine for  
75 African populations would mostly likely need to protect against a breadth of HIV subtypes and  
76 their derivatives (Burton et al., 2012; Burton & Hangartner, 2016; Sadanand et al., 2016;  
77 Steichen et al., 2019).

78 Notwithstanding its significance for HIV vaccine and treatment (e.g., broadly neutralizing  
79 antibodies) development, data on accurate and up-to-date longitudinal trends in HIV genetic  
80 diversity are limited, particularly in eastern and southern Africa (Hemelaar et al., 2019; Lihana et  
81 al., 2012). Prior analyses of publicly available HIV sequence data, mostly conducted before  
82 widespread expansion of HIV treatment and prevention programs, showed increasing viral  
83 diversity in Africa as the epidemic grew (Hemelaar et al., 2019, 2020; Lihana et al., 2012;  
84 Peeters et al., 2003). However, spatially and temporally uneven sampling of HIV genomes may  
85 have biased these earlier assessments. Moreover, by the end of 2022, 83% of people living with  
86 HIV (PLHIV) in eastern and southern Africa were using antiretroviral therapy (ART), resulting

87 in a decline of more than 57% in the number of new HIV infections since 2010 (UNAIDS,  
88 2023). However, it remains unclear to what extent such significant changes in the African HIV  
89 epidemic dynamics have affected viral diversity within the region.

90 Understanding patterns in HIV evolution, particularly in response to roll-out of treatment  
91 and prevention programs, may provide information critical for treatment development and  
92 insights into epidemic dynamics for virus control efforts. For example, declining diversity might  
93 indicate that less common strains of virus have disappeared, concentration of transmission within  
94 population sub-groups, or an epidemic increasingly driven by a small number of local outbreaks.  
95 In contrast, increasing diversity might highlight the growing role of virus introduction due to  
96 migration, having sex with non-local partners (Grabowski et al., 2014; Olawore et al., 2018), or  
97 indicate diversifying selection.

98 Here, we used data from more than 30 continuously surveyed communities in southern  
99 Uganda over a nearly three-decade long period to evaluate trends in intra- and inter-subtype HIV  
100 viral diversity. Data were collected prior to and during the scale-up of HIV prevention and  
101 treatment programs through the Rakai Community Cohort Study (RCCS), an open population-  
102 based household census and HIV surveillance cohort. Following scale up of antiretroviral  
103 therapy and voluntary medical male circumcision programs beginning in 2004, HIV incidence  
104 rapidly declined in RCCS communities (Grabowski et al., 2017), consistent with regional trends  
105 (Joshi et al., 2021). RCCS communities have among the highest HIV case burdens worldwide  
106 (Chang et al., 2016), with extensive HIV subtype diversity, including co-circulation of HIV-1  
107 subtypes A1, C, D, and numerous recombinant forms (Arroyo et al., 2006; Capoferri et al., 2020;  
108 Conroy et al., 2010; Lamers et al., 2020). Our analyses provide insight into population-level HIV  
109 evolutionary dynamics in a high burden region with declining incidence following HIV program

110 scale-up and may inform development and strategic deployment of HIV prevention and  
111 treatment interventions.  
112

## 113 Methods

### 114 Study design and population

115 The RCCS is conducted in inland agrarian and semi-urban trading communities (HIV  
116 prevalence 9-26%) and hyperendemic Lake Victoria fishing communities (HIV prevalence 37-  
117 43%) in southern Uganda (Chang et al., 2016). The RCCS has been ongoing since 1994 in inland  
118 agrarian and trading communities and since 2011 in fishing communities. Longitudinal  
119 surveillance is based on a household census and individual survey conducted at ~18 months  
120 intervals. The census enumerates all individuals in the household irrespective of age or presence  
121 in the community and documents residence status, including migrations into and out of the  
122 household. The survey follows the census and enrolls consenting adolescents and adults 15-49  
123 years who are resident in the community. During the survey, demographic, sexual behavior (e.g.,  
124 sex with partners outside community of residence), and HIV service utilization data (e.g., ART  
125 use and male circumcision status) are collected and blood samples are obtained for HIV testing  
126 and molecular epidemiology (Grabowski et al., 2017; Wawer et al., 1998). All RCCS participants  
127 are provided with HIV test results and offered free pre and post-test counseling with referral to  
128 prevention and treatment services. All study participants provide written informed consent at  
129 baseline and follow-up visits.

130 Here, we used HIV genomic sequence and survey data from PLHIV in 31 inland agrarian  
131 and semi-urban trading communities surveyed between November 1994 and May 2018 and four  
132 hyperendemic Lake Victoria fishing communities surveyed between November 2011 and August  
133 2017.

## 134 Sequencing

135 HIV sequence data has been generated using blood samples from RCCS participants  
136 living with HIV since the 1990s. Sequencing was attempted in some surveys for all participants  
137 with viremic HIV or not self-reporting ART use if viral load data was unavailable (Round 1:  
138 1994-1995, Round 9: 2002-2003, and Round 13-18: 2008-2018) and only for selected participant  
139 visits for others. For blood samples collected prior to 2010, Sanger sequencing was used to  
140 generate consensus sequences for p24 in the *gag* gene (HXB2: 1249-1704) and gp41 in the *env*  
141 gene (HXB2: 7857-8260) (C. Yang et al., 2000). For blood samples collected from 2010 onwards  
142 (beginning with survey round 14), near full-length genomes were generated using a next  
143 generation sequencing (NGS) approach in collaboration with the PANGEA-HIV consortium  
144 (Abeler-Dörner et al., 2019; Pillay et al., 2015) using the Gall protocol (2010-2015) and the ve-  
145 SEQ HIV protocol (2015-2018) (Bonsall et al., 2020; Gall et al., 2012). Resulting deep-  
146 sequencing reads were assembled using *shiver* to generate HIV consensus sequences (Wymant et  
147 al., 2018). Near full-length HIV genomes collected between January 2010 and May 2018 were  
148 aligned using MAFFT (v 7.52) and the HIV-1 regions of p24 and gp41 were extracted from the  
149 aligned sequences to compare consensus HIV genome sequences between earlier and later  
150 surveys (Katoh & Standley, 2013).

151

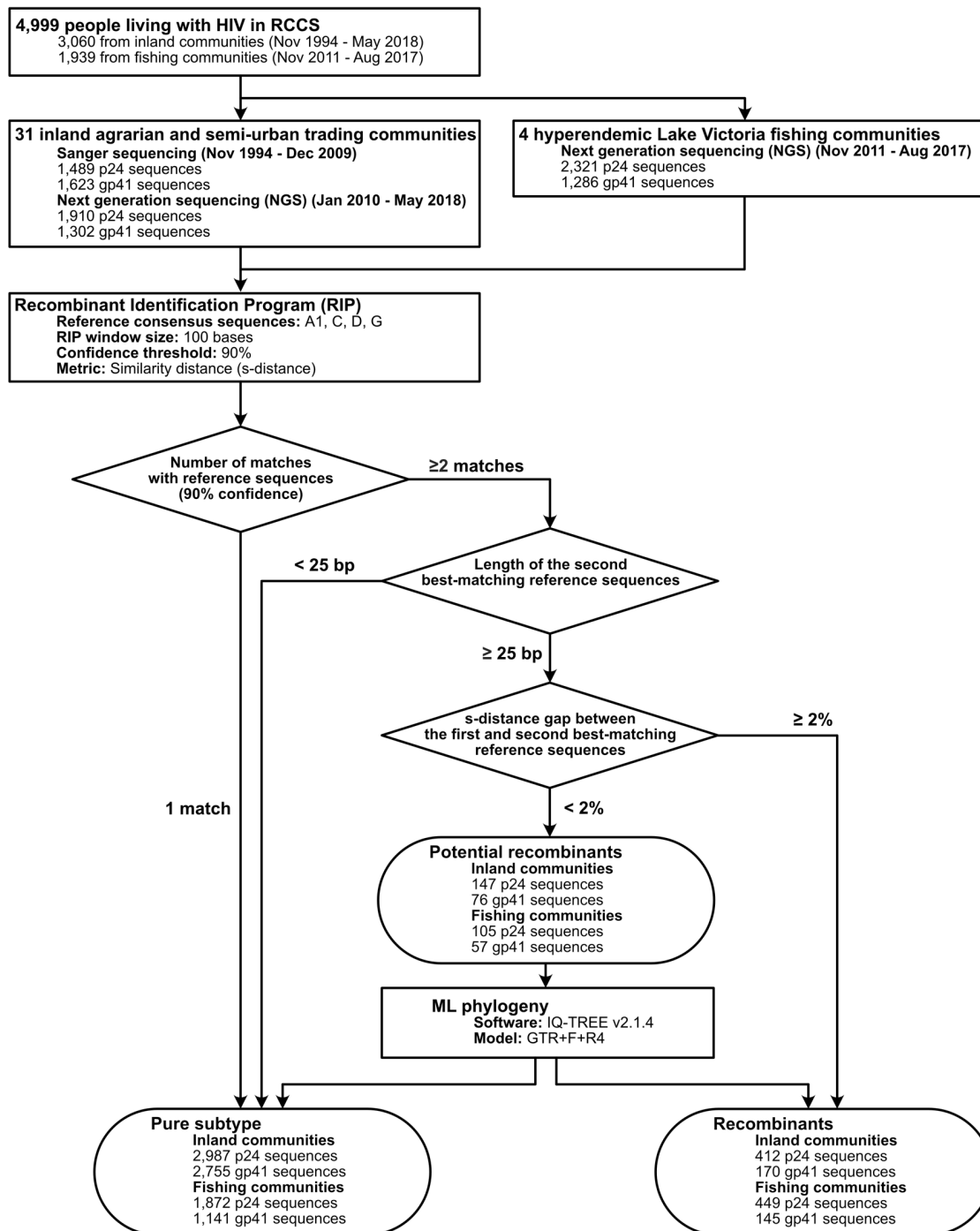
## 152 Subtyping analyses

153 The Recombinant Identification Program (RIP) version 3.0 was used to assign HIV  
154 subtypes to p24 and gp41 HIV genomes (Fig. 1) (Siepel et al., 1995). A RIP window of 100 bases  
155 and 90% confidence threshold was used to calculate similarity scores (s-distance) with the four  
156 most common circulating HIV-1 group M subtypes in Rakai (consensus reference sequences of



157 A1, C, D, and G) (Supplementary Table 1). HIV sequences were assigned as a pure subtype if  
158 they had a single best match with one of the reference sequences with 90% confidence. If there  
159 were more than two matches with the reference sequences with 90% confidence in any region,  
160 the length of the second best-matching reference sequence was measured, and if it was shorter  
161 than 25 bases, viruses were classified as pure subtypes. Otherwise, the s-distance gap between  
162 the first and second best-matching reference sequences in the potential recombinant region was  
163 used to further classify them; if the gap was greater than 2%, sequences were subtyped as  
164 recombinants, otherwise they were classified as potential recombinants. Next, pure, definitive,  
165 and potential recombinant sequences were aligned. Maximum Likelihood (ML) phylogenetic  
166 trees were generated from these alignments in IQ-TREE (v2.1.4) (Minh et al., 2020), and using  
167 resulting trees, recombinant sequences were classified as either pure subtype or recombinant  
168 viruses. We further analyzed HIV subtypes of p24 and gp41 from RCCS participants who had  
169 available sequence data from both genes in the same surveys. HIV subtypes from these two gene  
170 regions were combined if the sequence data were obtained from the same individuals in the same  
171 survey round. For example, a participant with subtype A1 virus in both genes, would be  
172 considered as having a subtype A1 infection whereas a participant with subtype A1 in one gene  
173 and subtype D in the other would be considered an A1/D recombinant infection. We also  
174 assigned HIV subtypes to p24 and gp41 sequences using COMET and REGA subtyping tools,  
175 both with a 70% of bootstrap support threshold, and results were compared to subtype  
176 assignment based on RIP and phylogenetic analyses (Supplementary Tables 2 and 3) (Pineda-  
177 Peña et al., 2013; Struck et al., 2014). The Cochran-Armitage test for trend in proportions was  
178 used to assess the statistical significance of longitudinal trends in HIV-1 subtype distributions

179 over calendar time using *stats* package (v4.3.2) (part of base R) in R (v4.3.1) (R Core Team,  
 180 2021).



181  
 182 **Figure 1.** HIV genome sequence data collected from RCCS participants living with HIV in 31 inland  
 183 agrarian and semi-urban trading communities and four hyperendemic Lake Victoria fishing communities  
 184 and subtype analyses using the Recombinant Identification Program (RIP) and phylogenetic  
 185 reconstruction

## 186 Diversity measures

### 187 Shannon diversity index

188 We used the Shannon Diversity Index (SDI) to evaluate longitudinal trends in HIV inter-  
189 subtype diversity in inland and fishing communities in Rakai. SDI quantifies biodiversity and  
190 species richness within a population (Shannon, 1948). SDI was calculated as follows:  $H' =$   
191  $-\sum_{i=1}^S (p_i \ln p_i)$ , where  $S$  is the total number of HIV subtypes in one survey round and  $p_i$  is the  
192 proportion of subtype  $i$  sequences relative to the total number of sequences in one survey round.  
193 A high  $H'$  value indicates greater inter-subtype diversity and more even distribution of sequences  
194 among subtypes, whereas low  $H'$  implies lower intra-subtype diversity and the dominance of  
195 particular subtypes. For these analyses inter-subtype recombinants of the same subtypes were  
196 classified as a single subtype irrespective of breakpoint.

197 We measured SDI in p24 and gp41 within inland and fishing communities overall and  
198 then within various population sub-groups including by sex, age group, occupation, HIV-related  
199 risk behavior (numbers of past year sexual partners, sexual partners outside the community),  
200 incident case (i.e. participants testing positive for the first time with an HIV seronegative test  
201 result at the prior visit), and history of recent migration into study communities. Occupations  
202 were grouped differently for inland (agricultural, trading/trucking, other) and fishing  
203 communities (fisherfolk/bar workers/restaurant workers, other) because their demographic  
204 composition varies significantly between inland and fishing communities (Chang et al., 2016).  
205 SDI values were calculated using *poppr* R package (v2.9.4) and 95% bootstrap confidence  
206 intervals were generated using the package *boot* (v1.3.28.1) (part of base R) in R (Kamvar et al.,  
207 2014).

208

## 209 Nucleotide diversity

210 We next estimated nucleotide diversity ( $\pi$ ) and pairwise TN93 genetic distance in p24  
211 and gp41 to measure longitudinal trends in both inter- and intra-subtype genetic diversity. The  
212 nucleotide diversity index is a measure of genetic variation that quantifies the average number of  
213 nucleotide differences per site between all possible pairs of sequences in a population.  $\pi$  was  
214 calculated as follows:  $\hat{\pi} = \frac{n}{n-1} \sum_{ij} x_i x_j \pi_{ij}$ , where  $n$  is the number of sequences,  $x_i$  and  $x_j$  are the  
215 frequencies of the  $i$ th and  $j$ th sequences, and  $\pi_{ij}$  is the number of nucleotide differences per site  
216 between  $i$ th and  $j$ th sequences. Firstly, the overall nucleotide diversity statistics, regardless of  
217 subtype, were calculated to evaluate the degree of genetic variation of HIV within the entire  
218 population in each survey round. Then, subtype-specific estimates within HIV subtypes A1 and  
219 D was measured to evaluate temporal trends in the genetic diversity of two most common HIV  
220 subtypes in our dataset. Similar to SDI, we calculated HIV nucleotide diversity in the entire  
221 population and within various population sub-groups. Nucleotide diversity measures were  
222 calculated using *pegas* package (v1.2) and 95% bootstrap confidence intervals were calculated  
223 using *boot* package in R (Paradis, 2010).

224

## 225 Genetic divergence

226 We also calculated root-to-tip distances and pairwise TN93 genetic distances to evaluate  
227 longitudinal trends of population-level HIV genetic divergence in p24 and gp41. Firstly, subtype-  
228 specific (A1, C, and D) p24 and gp41 ML phylogenetic trees were reconstructed in IQ-TREE  
229 using the General Time Reversible model with empirical base frequencies and free rate model  
230 with four categories (GTR+F+R4) (Z. Yang, 1995). ML phylogenetic trees were rooted using  
231 sequences from other subtypes as the outgroup. Patristic distances from the root to the tips of

232 phylogenetic trees (root-to-tip distances) were measured and plotted against the sampling times  
233 of sequences. The evolutionary rates and time to the most recent common ancestor were  
234 calculated using a simple linear regression. Root-to-tip distances were calculated using *adephylo*  
235 package (v1.1.16) (Jombart et al., 2010).

236         Additionally, we measured pairwise TN93 genetic distances of HIV subtype A1, C, and D  
237 between the HIV consensus sequences of Los Alamos National Laboratory (LANL) HIV  
238 Database subtype references and HIV sequence data collected in Rakai (Tamura & Nei, 1993).  
239 LANL reference sequences of subtypes A1, C, and D (Supplementary Table 1) were retrieved  
240 and the consensus sequences were generated using the Consensus Maker Tools  
241 (<https://www.hiv.lanl.gov/content/sequence/CONSENSUS/consensus.html>). Pairwise TN93  
242 genetic distances were calculated using *ape* package (v5.7.1) in R (Paradis & Schliep, 2019).

243

#### 244 HIV evolution and selective pressures

245         Lastly, we assessed if there was evidence of selective virus evolution. First, we calculated  
246 Tajima's D statistic in each survey to quantify deviations from the expected genetic diversity  
247 under neutral evolution, where most of the genetic variation arises from neutral mutations.  
248 Tajima's D is calculated as the normalized difference between the nucleotide genetic diversity  
249 and the expected number of segregating sites (Tajima, 1989). A substantial deviation of Tajima's  
250 D from 0 indicates the potential effects of natural selection, changes in population size, or other  
251 non-neutral processes. To prevent overestimation of segregating sites, p24 and gp41 sequences  
252 with gaps greater than 10 bases were excluded for the calculation of Tajima's D statistic. The p-  
253 values of Tajima's D statistic were calculated using a beta distribution with a mean of zero and a  
254 variance of one. Tajima's D statistic was calculated using *pegas* package in R.

255 We further calculated the dN/dS ratio ( $\omega$ ) to evaluate a population-level gene-wide  
256 selective pressure on HIV-1 subtypes A1 and D p24 and gp41 (Z. Yang & Bielawski, 2000). This  
257 metric measures the ratio of non-synonymous substitutions to synonymous substitutions in a  
258 gene to assess evolutionary forces acting on the gene within a population. A value of  $\omega < 1$   
259 suggests purifying or negative selection, indicating that non-synonymous changes are being  
260 removed or are not favored within a population. A value of  $\omega = 1$  implies neutral evolution,  
261 whereas  $\omega > 1$  indicates positive or diversifying selection, suggesting that non-synonymous  
262 changes are favored and occur more frequently than synonymous changes. To estimate, we first  
263 reconstructed ML phylogenetic trees in IQ-TREE (v2.1.4) and from these trees, we then  
264 calculated  $\omega$  in codeml in the PAML package (v4.9j) and BUSTED in the HyPhy package  
265 (v2.5.33) (Murrell et al., 2015; Z. Yang, 2007). For the dN/dS ratio analyses conducted in  
266 codeml, we subsampled down to 50 sequences from each survey round (n=400 in inland  
267 communities and n=176 in fishing communities) to maintain computational efficiency.  
268 Sensitivity analyses with five different random sequence sets were performed to assess the  
269 impact of down-sampling on the estimated parameters values.  
270

## 271 Results

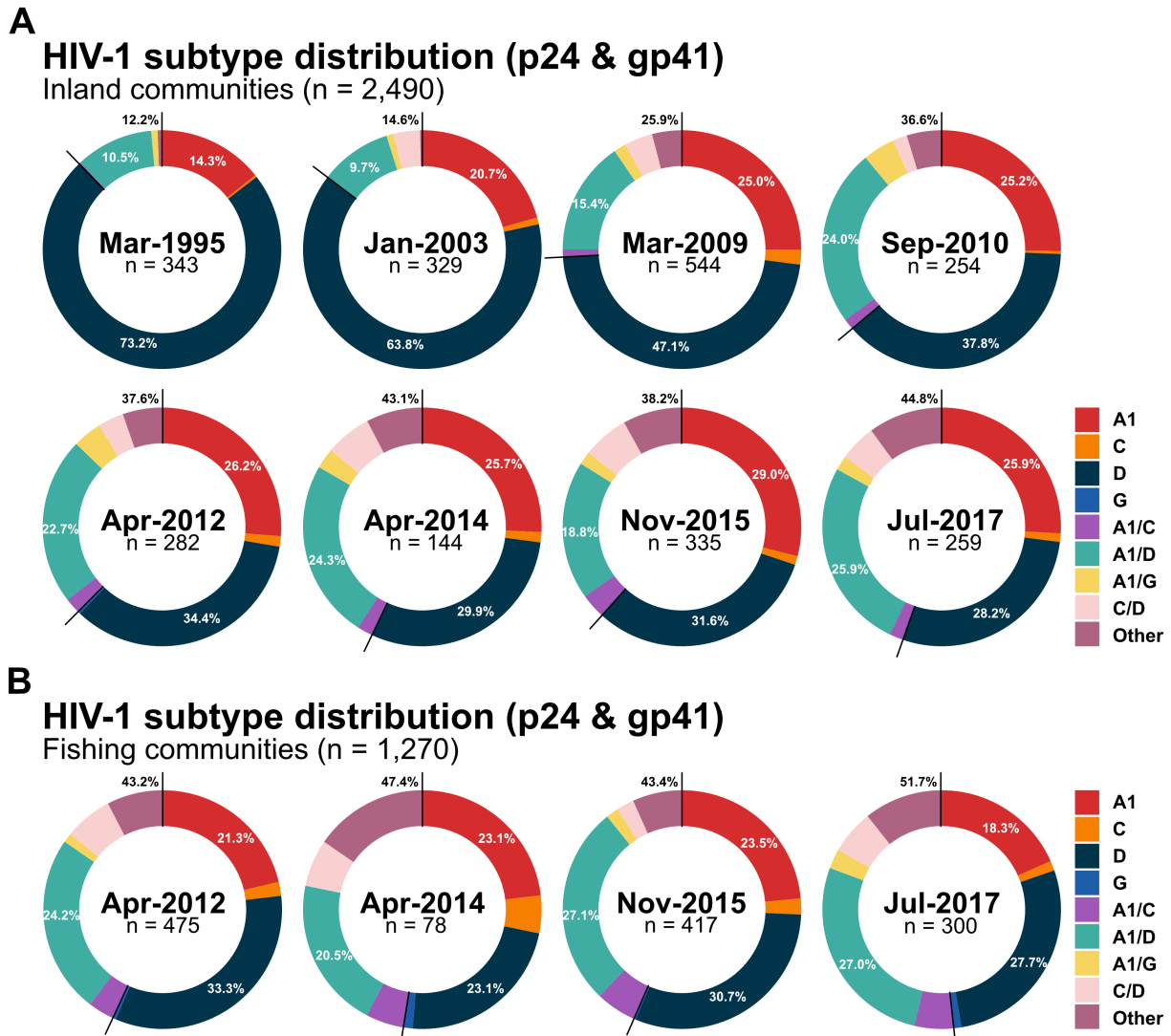
### 272 Population characteristics

273           Between 1994 and 2018, 46,565 individuals participated in the RCCS of whom 37,606  
274 lived in inland agrarian and semi-urban trading communities and 8,959 in hyperendemic Lake  
275 Victoria fishing communities. Of these participants, 5,352 (14.2%) in inland communities and  
276 3,402 (38.0%) in fishing communities were living with HIV. Among participants living with  
277 HIV, 3,060 (57.2%) in inland communities and 1,939 (57.0%) in fishing communities had HIV  
278 sequence data available in either p24 or gp41 genes from at least one survey visit. Of these  
279 participants with HIV sequence data available, 2,111 (69.0%) in inland communities and 1,121  
280 (57.8%) in fishing communities had both p24 and gp41 sequence data. Individuals with HIV  
281 sequence data were less likely to be using ART, be male, and younger than individuals without  
282 HIV sequence data (Table 1, Supplementary Table 4 and 5). Despite more frequent sampling in  
283 later survey periods, HIV sequence data was less commonly available among persons living with  
284 HIV because of the scale-up of ART and resulting increases in HIV viral load suppression over  
285 calendar time.  
286

287 **Table 1.** Demographic and HIV-related risk behavioral characteristics of RCCS participants living with  
 288 HIV, with and without HIV sequence data in 31 inland agrarian and semi-urban trading communities  
 289 between 1994 and 2018

Inland communities	R1 (Nov-1994 – Aug-1995)			R13 (June-2008 – Dec-2009)			R18 (Oct-2016 – May-2018)		
	PLHIV	PLHIV w/ seq	PLHIV wo/ seq	PLHIV	PLHIV w/ seq	PLHIV wo/ seq	PLHIV	PLHIV w/ seq	PLHIV wo/ seq
<b>N</b>	1,162	604	558	1,218	703	515	1,648	269	1,379
<b>Age</b>									
Median age (IQR)	29 (24, 35)	30 (24, 35)	28 (24, 35)	33 (28, 39)	31 (26, 37)	35 (31, 41)	36 (30, 42)	30 (25, 35)	37 (31, 42)
15 – 24	315 (27.1%)	154 (25.5%)	161 (28.9%)	170 (14%)	130 (18.5%)	40 (7.8%)	173 (10.5%)	66 (24.5%)	107 (7.8%)
25 – 34	531 (45.7%)	279 (46.2%)	252 (45.2%)	542 (44.5%)	346 (49.2%)	196 (38.1%)	559 (33.9%)	130 (48.3%)	429 (31.1%)
35 +	316 (27.2%)	171 (28.3%)	145 (26%)	506 (41.5%)	227 (32.3%)	279 (54.2%)	916 (55.6%)	73 (27.1%)	843 (61.1%)
<b>Sex</b>									
Male	440 (37.9%)	251 (41.6%)	189 (33.9%)	425 (34.9%)	263 (37.4%)	162 (31.5%)	507 (30.8%)	116 (43.1%)	391 (28.4%)
Female	722 (62.1%)	353 (58.4%)	369 (66.1%)	793 (65.1%)	440 (62.6%)	353 (68.5%)	1,141 (69.2%)	153 (56.9%)	988 (71.6%)
<b>Occupation</b>									
Agriculture	654 (56.3%)	331 (54.8%)	323 (57.9%)	637 (52.3%)	359 (51.1%)	278 (54%)	870 (52.8%)	115 (42.8%)	755 (54.7%)
Trade/Truck	150 (12.9%)	90 (14.9%)	60 (10.8%)	168 (13.8%)	94 (13.4%)	74 (14.4%)	216 (13.1%)	52 (19.3%)	164 (11.9%)
Restaurant/Bar	32 (2.8%)	18 (3%)	14 (2.5%)	58 (4.8%)	32 (4.6%)	26 (5%)	103 (6.2%)	10 (3.7%)	93 (6.7%)
Other	326 (28.1%)	165 (27.3%)	161 (28.9%)	355 (29.1%)	218 (31%)	137 (26.6%)	459 (27.9%)	92 (34.2%)	367 (26.6%)
<b>Number of sexual partners in the past year</b>									
0 – 1	959 (82.5%)	488 (80.8%)	471 (84.4%)	794 (65.2%)	478 (68%)	316 (61.4%)	1,020 (61.9%)	149 (55.4%)	871 (63.2%)
2 +	197 (17%)	114 (18.9%)	83 (14.9%)	239 (19.6%)	149 (21.2%)	90 (17.5%)	346 (21%)	82 (30.5%)	264 (19.1%)
No response	6 (0.5%)	2 (0.3%)	4 (0.7%)	185 (15.2%)	76 (10.8%)	109 (21.2%)	282 (17.1%)	38 (14.1%)	244 (17.7%)
<b>Any sexual partner outside the community in the past year</b>									
Yes	193 (16.6%)	101 (16.7%)	92 (16.5%)	333 (27.3%)	202 (28.7%)	131 (25.4%)	482 (29.2%)	101 (37.5%)	381 (27.6%)
No	968 (83.3%)	502 (83.1%)	466 (83.5%)	700 (57.5%)	425 (60.5%)	275 (53.4%)	884 (53.6%)	130 (48.3%)	754 (54.7%)
No response	1 (0.1%)	1 (0.2%)	0 (0%)	185 (15.2%)	76 (10.8%)	109 (21.2%)	282 (17.1%)	38 (14.1%)	244 (17.7%)
<b>Recent history of migration</b>									
Yes	0 (0%)	0 (0%)	0 (0%)	171 (14%)	113 (16.1%)	58 (11.3%)	405 (24.6%)	124 (46.1%)	281 (20.4%)
No	1,115 (96%)	578 (95.7%)	537 (96.2%)	1,045 (85.8%)	588 (83.6%)	457 (88.7%)	1,243 (75.4%)	145 (53.9%)	1,098 (79.6%)
Non-resident	47 (4%)	26 (4.3%)	21 (3.8%)	2 (0.2%)	2 (0.3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
<b>Self-reported ART use</b>									
Yes	-	-	-	268 (22%)	9 (1.3%)	259 (50.3%)	1,277 (77.5%)	64 (23.8%)	1,213 (88%)
No	-	-	-	731 (60%)	524 (74.5%)	207 (40.2%)	332 (20.1%)	186 (69.1%)	146 (10.6%)
No response	-	-	-	219 (18%)	170 (24.2%)	49 (9.5%)	39 (2.4%)	19 (7.1%)	20 (1.5%)
<b>Viral load</b>									
≥ 1,000	462 (39.8%)	277 (45.9%)	185 (33.2%)	690 (56.7%)	573 (81.5%)	117 (22.7%)	312 (18.9%)	269 (100%)	43 (3.1%)
50 ≤ VL < 1,000	24 (2.1%)	11 (1.8%)	13 (2.3%)	117 (9.6%)	71 (10.1%)	46 (8.9%)	56 (3.4%)	0 (0%)	56 (4.1%)
BD	21 (1.8%)	2 (0.3%)	19 (3.4%)	355 (29.1%)	29 (4.1%)	326 (63.3%)	1,264 (76.7%)	0 (0%)	1,264 (91.7%)
No data	655 (56.4%)	314 (52%)	341 (61.1%)	56 (4.6%)	30 (4.3%)	26 (5%)	16 (1%)	0 (0%)	16 (1.2%)





291  
 292 **Figure 2.** HIV subtype distribution of p24 and gp41 from RCCS participants living with HIV who had  
 293 available sequence data from both genes in the same survey in 31 inland agrarian and semi-urban trading  
 294 communities (Panel A) and four hyperendemic Lake Victoria fishing communities (Panel B) by calendar  
 295 time. HIV subtypes of p24 and gp41 were independently assigned in the RIP and HIV subtypes of p24  
 296 and gp41 from the same individuals in the same survey were combined.  
 297

298 In inland communities, the most common HIV subtypes were A1 and D subtypes,  
 299 followed by A1/D recombinant viruses, although the proportions of these subtypes varied over  
 300 calendar time (Fig. 2A, Supplementary Fig. 1A, 2A, Supplementary Table 6, 7, and 8). Among  
 301 participants who had both p24 and gp41 sequence data in the same survey, the proportion of

302 subtype A viruses increased from 14.3% in 1995 to 25.0% until 2009 ( $p < 0.001$ ), and then leveled  
303 off around 25% ( $p = 0.8515$ ). Subtype D viruses consistently declined from 73.2% in 1995 to  
304 28.2% in 2017 ( $p < 0.001$ ). Similar patterns were observed in HIV subtype distribution in p24 and  
305 gp41. Subtype A1 viruses proportionately increased from 16.6% in p24 and 21% in gp41 until  
306 2010 (p24:  $p < 0.001$ , gp41:  $p < 0.001$ ), after which stabilizing around ~31% in p24 and ~47% in  
307 gp41 between 2012 and 2017 (p24:  $p = 0.8232$ , gp41:  $p = 0.9843$ ). Meanwhile, the proportion of  
308 subtype D viruses significantly declined in p24 from 79.6% in 1995 to 44.7% in 2017 ( $p < 0.001$ )  
309 and in gp41 from 75.9% in 1995 to 38.2% in 2017 ( $p < 0.001$ ).

310 In the analyses of inland community participants with subtype data in both genes, the  
311 proportion of recombinants increased from 12.2% in 1995 to 44.8% in 2017 ( $p < 0.001$ ) (Fig 2A  
312 and Supplementary Table 6). The proportion of A1/D recombinant viruses substantially increased  
313 during the earlier surveys from 10.5% in 1995 to 24.0% in 2010 ( $p < 0.001$ ), after which leveled  
314 off around 23% ( $p = 0.2046$ ). Recombinant viruses of more than three different subtypes (e.g.,  
315 A1/C/D, A/D/G) or rare recombinant forms in our sequence data (e.g., D/G, C/G) also  
316 significantly increased over calendar time, from 0.5% in 1995 to 10.0% in 2017 ( $p < 0.001$ ). By  
317 2017, 44.8% of participants with sequence data in both genes were classified as having an inter-  
318 subtype recombinant infection. p24 sequences were significantly more likely to be classified as  
319 recombinants over the analysis period. The proportion of recombinants in p24 significantly  
320 increased from 2.7% in 1995 to 21.4% in 2017 ( $p < 0.001$ ) (Supplementary Figure 1A and  
321 Supplementary Table 7), in gp41 from 2.4% in 1995 to 9.5% in 2017 ( $p < 0.001$ ) (Supplementary  
322 Fig. 2A and Supplementary Table 8). In both genes, the most common recombinant infection was  
323 A1/D recombinants.

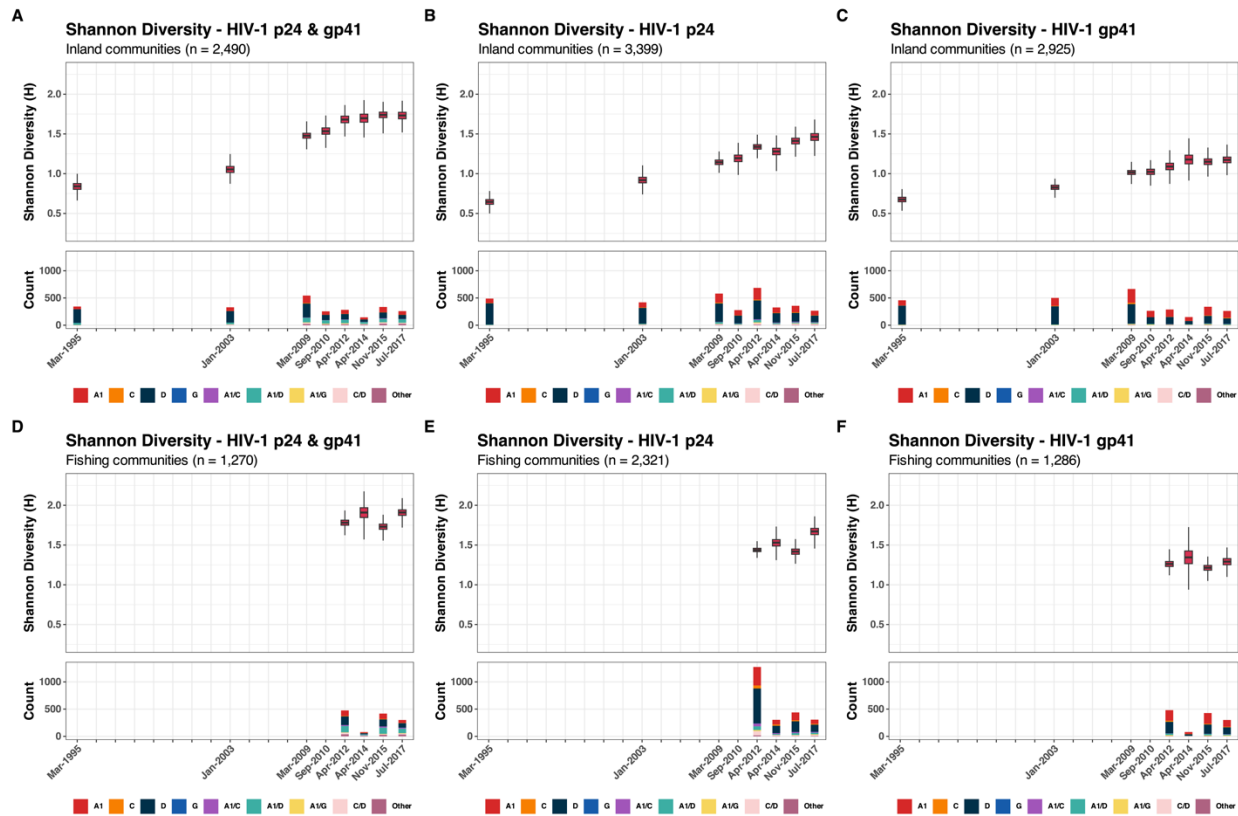
324 In Lake Victoria fishing communities, the proportion of viruses classified as either  
325 subtype A1 or D did not substantially change between 2012 and 2017 (Fig. 2B, Supplementary  
326 Fig. 1B, 2B, Supplementary Table 9, 10, and 11). Overall, recombinant viruses were slightly  
327 more common in fishing communities than inland communities, but these differences were not  
328 statistically significant in either p24 or gp41. Among individuals with subtype data in both genes,  
329 the proportions of subtypes A1, D, and A1/D recombinant viruses were generally similar to those  
330 in inland communities over the analysis period. Similar to inland communities, recombinants  
331 were more frequently detected in p24 than in gp41, and were usually A1/D, A1/C, or C/D  
332 recombinants.

333 We further stratified HIV subtypes by participant demographic and HIV-related risk  
334 behavioral characteristics across three calendar epochs for inland and fishing communities: pre-  
335 ART (1994-2004), ART roll-out and expansion (2008-2013), and Universal Test and Treatment  
336 era (2013-2018) (Supplementary Fig. 3-8). Overall, the proportion of subtype A1 and  
337 recombinant viruses increased, whereas subtype D viruses proportionally decreased regardless of  
338 population sub-groups over calendar time.

339

340

341 Inter-subtype diversity as measured by Shannon Diversity Index (SDI)

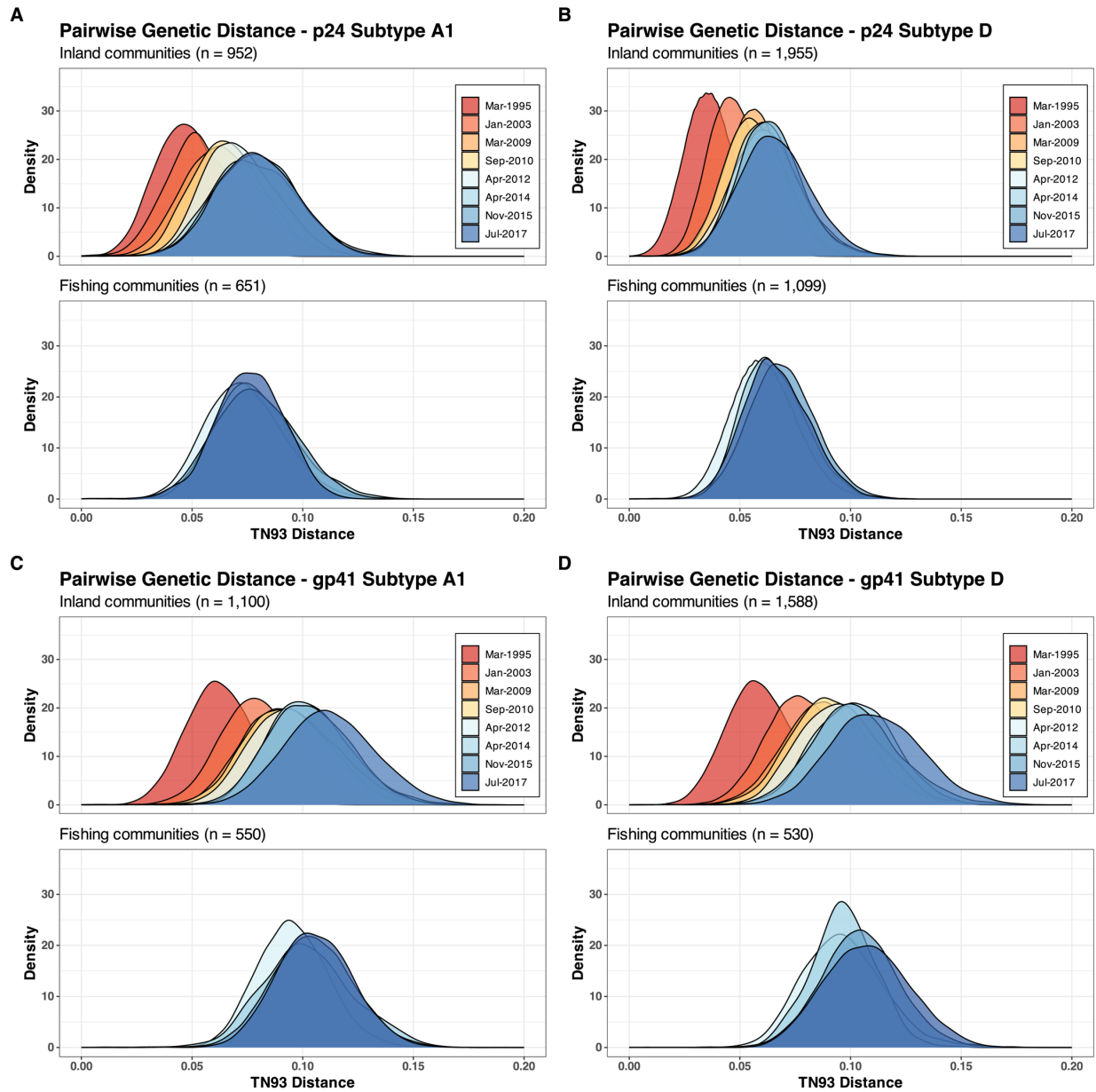


342  
 343 **Figure 3.** The Shannon diversity index of p24, gp41, and both p24 and gp41 in 31 inland agrarian and  
 344 semi-urban trading communities and four Lake Victoria fishing communities between 1995 and 2017  
 345

346 Next, we evaluated trends in inter-subtype HIV diversity by calculating the SDI in p24  
 347 and gp41 genes over calendar time for both inland and fishing communities (Fig. 3). Overall, the  
 348 SDI increased in p24 through the study period regardless of community type. The SDI also  
 349 increased in gp41, although no significant changes from 2014 onwards were identified in either  
 350 inland or fishing communities. SDI trends were generally similar irrespective of age, sex,  
 351 occupation, number of sexual partners in the past year, extra-community partners in the past year,  
 352 or among incident/non-incident cases (Supplementary Fig. 9-15). Notable exceptions included  
 353 adolescents and young adults (15-24 years) and persons with a recent migration history in inland  
 354 communities, with these groups exhibiting slightly higher levels of p24 inter-subtype HIV

355 diversity than the rest of the population in the most recent survey period (Supplementary Fig. 10  
356 and 14).

357



359  
 360 **Figure 4.** Pairwise TN93 genetic distances of p24 and gp41 subtypes A1 and D in 31 inland  
 361 agrarian and semi-urban trading communities and four hyperendemic Lake Victoria  
 362 fishing communities between 1995 and 2017  
 363

364 Within inland communities, overall nucleotide diversity, irrespective of viral subtype,  
 365 increased in both genes over the analysis period (Supplementary Fig. 16). However, intra-  
 366 subtype diversity in A1 and D p24 subtypes leveled off after 2014 (Fig. 4). In contrast, A1 and D

367 gp41 nucleotide diversity increased continuously. In fishing communities, p24 nucleotide  
368 diversity in subtype A1 increased through 2014 then decreased, while p24 nucleotide diversity of  
369 subtype D increased through 2015 then decreased (Fig. 4 and Supplementary Fig. 17). Similar to  
370 inland communities, overall and intra-subtype specific nucleotide gp41 diversity increased  
371 throughout the analysis periods. Nucleotide diversity trends were generally similar across  
372 population-subgroups (Supplementary Fig. 18-30), although slightly higher overall p24  
373 nucleotide diversity, but not intra-subtype diversity, was observed among younger individuals  
374 15-24 years in inland communities (Supplementary Fig. 20).

375 We further evaluated population-level HIV genetic divergence by using root-to-tip  
376 divergence analyses (Supplementary Fig. 31) and pairwise TN93 distances between subtype  
377 reference consensus sequences (A1, C, and D) and HIV sequences in Rakai (Supplementary Fig.  
378 32). Similar to nucleotide diversity, p24 diverged through 2014 and then leveled off, while gp41  
379 diverged over the analysis period. In the analyses of root-to-tip divergence, higher degrees of  
380 genetic variability were observed in gp41 compared to p24 in the same surveys.

381

## 382 HIV selection pressure measures

383 Overall, Tajima's D values were less than 0, lower in p24 than in gp41 over the analysis  
384 period, indicating stronger purifying selection in p24 than in gp41 (Supplementary Fig. 33 and  
385 34). In both inland and fishing communities,  $\omega$  values were less than 1, indicating a gene-wide  
386 purifying selection acting on p24 and gp41 subtype A1 and D viruses (Supplementary Table 12).  
387 Consistent with Tajima's D analyses,  $\omega$  values in p24 were also lower than those in gp41. In p24,  
388 subtype A1 viruses had slightly higher  $\omega$  than subtype D viruses, suggesting relatively weaker  
389 purifying selection pressure acting on subtype A1. In contrast,  $\omega$  values were slightly higher

390 among gp41 subtype D than among gp41 subtype A1, indicating relatively stronger purifying  
391 selection acting on subtype A1 gp41.

392



## 393 Discussion

394 In this study, we examined longitudinal trends in population-level HIV diversity within  
395 31 inland agrarian and semi-urban trading communities and four hyperendemic Lake Victoria  
396 fishing communities in Rakai, Uganda over a 25-year period including the introduction and the  
397 expansion of HIV prevention and treatment programs. While HIV incidence in eastern and  
398 southern Africa has declined by more than 50% since 2010 (Grabowski et al., 2017; Joshi et al.,  
399 2021; UNAIDS, 2023), we found that population HIV genetic diversity and divergence increased  
400 over calendar time. Increases in HIV genetic diversity were primarily driven by an increasing  
401 proportion of inter-subtype recombinant forms and rising intra-subtype gp41 genetic diversity.  
402 While Lake Victoria fishing communities are recognized as key populations with a high HIV  
403 burden and population mobility (Chang et al., 2016; Kagaayi et al., 2019), neither inter- nor  
404 intra-subtype genetic diversity was significantly higher compared to inland communities.

405 Overall, our HIV-1 genomic sequence data showed that trends in HIV genetic diversity  
406 were generally consistent with those from previous analyses conducted in Uganda and Eastern  
407 Africa. Notably, the proportions of inter-subtype recombinant viruses and their increasing trends,  
408 and a decline in subtype D viruses were remarkably similar to those in previous systematic  
409 reviews and molecular studies of HIV sequence data from Uganda (Grant et al., 2020; Hemelaar  
410 et al., 2020) and within Africa more broadly (Hemelaar et al., 2019; Lee et al., 2017). Here, we  
411 further show that changes in HIV subtype distribution persisted following the extensive scale-up  
412 of HIV programs; subtypes A1 and D remained predominant, while the proportion of inter-  
413 subtype recombinants increased. A previous molecular epidemiologic study from the RCCS  
414 reported a decline in HIV intra-subtype genetic diversity following the ART roll-out (Lamers et  
415 al., 2020). However, our findings from both nucleotide diversity and pairwise TN93 distance

416 analyses indicate a consistent increase in genetic diversity in gp41, whereas diversity of subtypes  
417 A1 and D in p24 leveled off after the scale-up of HIV programs.

418         Between 1994 and 2018, we observed a substantial decrease in the proportion of subtype  
419 D viruses and an increasing proportion of subtype A1 viruses among individuals living in inland  
420 communities. These trends are consistent with findings in previous studies from Uganda (Conroy  
421 et al., 2010; Hemelaar et al., 2019; Lamers et al., 2020), and is likely attributed to pathogenic and  
422 clinical characteristics of subtypes A1 and D viruses. For example, infection with HIV-1 subtype  
423 D is associated with a rapid decrease in CD4<sup>+</sup> cell count compared to subtype A1 and other non-  
424 D subtypes, resulting in faster progression to acquired immunodeficiency syndrome (AIDS) and  
425 earlier initiation of ART. Thus, there is a relatively shorter amount of time for subtype D viruses  
426 to be transmitted than subtype A1 or other subtypes (Baeten et al., 2007; Kaleebu et al., 2002;  
427 Kiwanuka et al., 2008; Ssemwanga et al., 2013). Additionally, in a previous study in Uganda,  
428 individuals in HIV serodiscordant couples whose partners were infected with subtype D viruses  
429 were found to be at a lower risk of acquiring HIV compared to those with subtype A1 (Kiwanuka  
430 et al., 2009). The higher rate of HIV disease progression and lower transmissibility of subtype D  
431 compared to subtype A1 or other subtypes may have led to the observed patterns in the  
432 proportion of subtype A1 and D viruses in this study.

433         We further showed that increases in HIV genetic diversity were largely driven by an  
434 increase in HIV inter-subtype recombinant infections. In both inland and fishing communities,  
435 the most common recombinants were A1/D recombinants, consistent with previous molecular  
436 and surveillance studies in Uganda (Arroyo et al., 2006; Capoferri et al., 2020; Conroy et al.,  
437 2010; Eshleman et al., 2002; Grant et al., 2020; Harris et al., 2002; Lamers et al., 2020; Yirrell et  
438 al., 2002). Interestingly, despite the low prevalence of pure subtype C viruses in our HIV

439 sequence data, we observed a substantial increase in inter-subtype C recombinants (A1/C and  
440 C/D) in p24. This might be attributed to introduction of subtype C, which accounted for 15% of  
441 HIV infections in Eastern Africa by 2015 (Hemelaar et al., 2019). A previous study in Lake  
442 Victoria fishing communities in southern Uganda indicated a higher HIV prevalence among in-  
443 migrating populations from other geographic locations, with the highest rates observed among  
444 migrants from Tanzania (Grabowski et al., 2020). Another molecular study, conducted prior to  
445 ART roll-out in Rakai, revealed that 19% of inter-subtype recombinants were A1/C and C/D  
446 viruses in trading centers along Masaka road connecting Kampala to the border with Tanzania  
447 (Arroyo et al., 2006). These studies suggest that there might be a continuous influx of subtype C  
448 viruses or C recombinants into Uganda (Bbosa et al., 2019; Grabowski et al., 2020), resulting in  
449 high levels of inter-subtype viral mixing and increasing HIV genetic diversity.

450         Increasing HIV subtype diversity might be partly driven by individuals acquiring  
451 multiple infections of different HIV subtypes. Previous analyses from this same population  
452 revealed that individuals who remain viremic tended to have substantially higher levels of HIV-  
453 related risk behaviors compared to those who were virally suppressed (Brophy et al., 2021;  
454 Kennedy et al., 2023), potentially increasing the risk of coinfections and superinfections.  
455 Interestingly, however, our analyses of inter-subtype genetic diversity revealed no significant  
456 differences in levels of recombinant infection across various population sub-groups, including  
457 those with and without recent history of migration or those having extra-community sexual  
458 partners. In particular, despite high levels of HIV-related risk behaviors, population mobility in  
459 fishing communities, inter-subtype diversity was not significantly higher compared to inland  
460 communities. This might be due to widespread viral population mixing and pervasive inter-  
461 subtype recombination in the generalized and diverse HIV epidemic in Uganda (Grant et al.,

462 2020; Lee et al., 2017). In analyses of HIV subtype distribution from individuals with HIV  
463 sequence data in both genes, the proportion of recombinant viruses in 2017 were 44.8% in inland  
464 communities and 51.7% in fishing communities. These estimates are consistent with findings in  
465 a previous study of near full-length genomes of HIV in Uganda, indicating that nearly 50% of  
466 viruses were unique recombinants forms (URFs) (Grant et al., 2020). Inter-subtype  
467 recombination via coinfections and superinfections is common in Rakai (Redd et al., 2012, 2013;  
468 J. E. Taylor & Korber, 2005), and potentially driving HIV diversity in declining epidemic in this  
469 region.

470 Our data showed increasing trends in intra-subtype diversity and genetic divergence in  
471 gp41 irrespective of population sub-groups. We also found higher Tajima's D and  $\omega$  values in  
472 gp41, consistent with the previous analyses at both individual and population-level in Rakai and  
473 elsewhere (Alizon & Fraser, 2013; Raghwani et al., 2018). These findings suggest that  
474 diversifying selection is acting more intensively on gp41 (or *env*), possibly due to its critical role  
475 in immune evasion, while p24 (or *gag*) is likely under stronger functional constraints (Lin et al.,  
476 2019). However, it is unclear why subtype-specific genetic diversity in p24 leveled off following  
477 the widespread expansion of HIV treatment and prevention programs. One possibility is that the  
478 success of HIV intervention programs might limit the opportunities for mutation and  
479 recombination events. Continued molecular surveillance with full-length HIV genome sequence  
480 data from diverse geographic locations may be needed to understand the changing dynamics of  
481 HIV and the long-term efficacy of these HIV programs.

482 High levels of genetic diversity in gp41 have significant implications for vaccine  
483 development. Various candidate HIV vaccines have either failed or demonstrated limited efficacy  
484 due to the HIV genetic diversity, which led to calls for developing new vaccines that elicit

485 broadly neutralizing antibodies (bNAbs) responses (Burton, 2019; Burton & Hangartner, 2016;  
486 Sadanand et al., 2016; Steichen et al., 2019). bNAbs are capable of recognizing and neutralizing  
487 a breadth of HIV subtypes by binding to relatively conserved regions on the *env* protein. The  
488 primary challenge in developing vaccines is to reliably induce these bNAbs responses, which  
489 could be complicated by the high mutation rate and variability on the *env* gene. For example, 2F5  
490 and 4E10 are two extensively studied bNAbs that target the membrane-proximal external region  
491 (MPER) of the *env* protein, a part of the region in our gp41 sequence analyses (Alam et al., 2007;  
492 Brunel et al., 2006; Cardoso et al., 2005; Ofek et al., 2004). Increasing genetic diversity in gp41,  
493 especially around the MPER or other potential target regions in the whole *env* protein, may pose  
494 significant challenges for vaccine design, as genetic changes through mutations or inter-subtype  
495 recombination in these regions may allow the viruses to escape bNAbs responses, potentially  
496 reducing the efficacy of vaccines. Our findings highlight the need for a multifaceted approach in  
497 HIV vaccine development that not only focuses on the robust induction of bNAbs but also  
498 considers the rapid genetic shifts in gp41 through enhanced molecular surveillance and  
499 monitoring.

500         There are several limitations to this study. We used two short regions of HIV for our  
501 analyses; thus, it is possible that our assessment of the HIV subtype distribution might be biased  
502 and the extent of HIV genetic diversity is likely underestimated. Previous analyses of near full-  
503 length genomes from Uganda estimated that 30-50% of viruses were inter-subtype recombinants  
504 (Grant et al., 2020; Harris et al., 2002; Lee et al., 2017). In particular, the frequency of  
505 recombination breakpoints tends to be lower in the essential regions (e.g., *gag*, *pol*, and *env*),  
506 with the *env* gene exhibiting the lowest frequency (Grant et al., 2020; Lee et al., 2017). This  
507 could explain the relatively lower proportion of inter-subtype recombinant viruses in gp41

508 compared to p24 in our data. Although we analyzed HIV diversity across both genes, this  
509 underestimates the proportion of recombinant viruses. Thus, the analyses of short regions do not  
510 fully capture the overall subtype distribution, and genome-wide studies would be necessary to  
511 provide a more accurate assessment of the HIV genetic diversity. Additionally, intra-subtype  
512 recombination, which may occur just as frequently as inter-subtype recombination, could  
513 significantly contribute to genetic variation within subtypes, but we did not evaluate it here. We  
514 utilized HIV sequence data generated from different sequencing methods and protocols. These  
515 changes in sequencing methods may have influenced our assessment of intra-subtype HIV  
516 genetic diversity. Young individuals and men in Rakai generally had lower participation rates in  
517 the RCCS (Grabowski et al., 2020). However, these population sub-groups were more likely to  
518 be sequenced, particularly in the later survey rounds, thus these individuals living with HIV were  
519 overrepresented in our sequence data. Lastly, we used self-reported data to evaluate the HIV-1  
520 genetic diversity among individuals who had HIV-related high-risk behaviors, which may be  
521 influenced by reporting bias.

522         In conclusion, we observed continued HIV evolution, with a growing number of  
523 recombinant viruses and increasing intra-subtype evolution of gp41, despite declining HIV  
524 incidence. Overall, HIV genetic diversity increased in both inland agrarian and trading  
525 communities and hyperendemic Lake Victoria fishing communities, irrespective of population  
526 sub-groups, potentially reflecting the generalized nature of the HIV epidemic in Uganda.  
527 Comprehensive and wide-ranging molecular surveillance is essential for a better understanding  
528 of HIV dynamics driving population HIV evolution and the development of new HIV vaccines  
529 and treatments for epidemic control.

530

531 **Acknowledgements**

532           This study is supported by the Division of Intramural Research, National Institute of  
533 Allergy and Infectious Diseases, NIH (R01AI110324, R01AI143333-01, R01AI155080-01A1,  
534 R01MH115799-02, R01HL152813-01A1), the HIV Prevention Trials Network Laboratory  
535 Center (UM1-AI068613), and the Bill and Melinda Gates Foundation (OPP1175094,  
536 OPP1084362).

537 Reference

538

539 Abeler-Dörner, L., Grabowski, M. K., Rambaut, A., Pillay, D., & Fraser, C. (2019). PANGEA-HIV 2:  
540 Phylogenetics and Networks for Generalised Epidemics in Africa. In *Current Opinion in HIV*  
541 *and AIDS* (Vol. 14, Issue 3, pp. 173–180). Lippincott Williams and Wilkins.

542 <https://doi.org/10.1097/COH.0000000000000542>

543 Alam, S. M., McAdams, M., Boren, D., Rak, M., Scarce, R. M., Gao, F., Camacho, Z. T., Gewirth,  
544 D., Kelsoe, G., Chen, P., & Haynes, B. F. (2007). The Role of Antibody Polyspecificity and  
545 Lipid Reactivity in Binding of Broadly Neutralizing Anti-HIV-1 Envelope Human Monoclonal  
546 Antibodies 2F5 and 4E10 to Glycoprotein 41 Membrane Proximal Envelope Epitopes. *The*  
547 *Journal of Immunology*, 178(7), 4424–4435. <https://doi.org/10.4049/jimmunol.178.7.4424>

548 Alizon, S., & Fraser, C. (2013). Within-host and between-host evolutionary rates across the HIV-1  
549 genome. *Retrovirology*, 10(1), 49. <https://doi.org/10.1186/1742-4690-10-49>

550 Arroyo, M. A., Saterén, W. B., Serwadda, D., Gray, R. H., Wawer, M. J., Sewankambo, N. K.,  
551 Kiwanuka, N., Kigozi, G., Wabwire-Mangen, F., Eller, M., Eller, L. A., Birx, D. L., Robb, M. L., &  
552 McCutchan, F. E. (2006). Higher HIV-1 Incidence and Genetic Complexity Along Main Roads  
553 in Rakai District, Uganda. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 43(4),  
554 440–445. <https://doi.org/10.1097/01.qai.0000243053.80945.f0>

555 Baeten, J. M., Chohan, B., Lavreys, L., Chohan, V., McClelland, R. S., Certain, L., Mandaliya, K.,  
556 Jaoko, W., & Overbaugh, J. (2007). HIV-1 Subtype D Infection Is Associated with Faster  
557 Disease Progression than Subtype A in Spite of Similar Plasma HIV-1 Loads. *The Journal of*  
558 *Infectious Diseases*, 195(8), 1177–1180. <https://doi.org/10.1086/512682>

559 Barouch, D. H., & Korber, B. (2010). HIV-1 vaccine development after STEP. *Annual Review of*  
560 *Medicine*, 61, 153–167. <https://doi.org/10.1146/annurev.med.042508.093728>

561 Bbosa, N., Ssemwanga, D., Nsubuga, R. N., Salazar-Gonzalez, J. F., Salazar, M. G., Nanyonjo, M.,  
562 Kuteesa, M., Seeley, J., Kiwanuka, N., Bagaya, B. S., Yebra, G., Leigh-Brown, A., & Kaleebu, P.  
563 (2019). Phylogeography of HIV-1 suggests that Ugandan fishing communities are a sink for,  
564 not a source of, virus from general populations. *Scientific Reports*, 9(1), 1051.  
565 <https://doi.org/10.1038/s41598-018-37458-x>

566 Bonsall, D., Golubchik, T., De Cesare, M., Limbada, M., Kosloff, B., Macintyre-Cockett, G., Hall,  
567 M., Wymant, C., Azim Ansari, M., Abeler-Dörner, L., Schaap, A., Brown, A., Barnes, E.,  
568 Piwowar-Manning, E., Eshleman, S., Wilson, E., Emel, L., Hayes, R., Fidler, S., ... Team, P.  
569 (2020). A Comprehensive Genomics Solution for HIV Surveillance and Clinical Monitoring in  
570 Low-Income Settings. In *jcm.asm.org 1 Journal of Clinical Microbiology* (Vol. 58).  
571 <http://www.hiv.lanl.gov/>

572 Brophy, J. E., Lessler, J., Ssekubugu, R., Kennedy, C. E., Chang, L. W., Kigozi, G., Ndyanabo, A.,  
573 Kigozi, G., Nalugoda, F., Gray, R. H., Wawer, M. J., Serwadda, D. M., Reynolds, S. J., Kagaayi,  
574 J., & Grabowski, M. K. (2021). Prevalence of Untreated HIV and Associated Risk Behaviors  
575 Among the Sexual Partners of Recent Migrants and Long-term Residents in Rakai, Uganda.  
576 *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 88(3), 243–251.  
577 <https://doi.org/10.1097/QAI.0000000000002764>

578 Brunel, F. M., Zwick, M. B., Cardoso, R. M. F., Nelson, J. D., Wilson, I. A., Burton, D. R., & Dawson,  
579 P. E. (2006). Structure-Function Analysis of the Epitope for 4E10, a Broadly Neutralizing



580 Human Immunodeficiency Virus Type 1 Antibody. *Journal of Virology*, 80(4), 1680–1687.  
581 <https://doi.org/10.1128/jvi.80.4.1680-1687.2006>

582 Burton, D. R. (2019). Advancing an HIV vaccine; advancing vaccinology. *Nature Reviews*  
583 *Immunology*, 19(2), 77–78. <https://doi.org/10.1038/s41577-018-0103-6>

584 Burton, D. R., & Hangartner, L. (2016). Broadly Neutralizing Antibodies to HIV and Their Role in  
585 Vaccine Design. *Annual Review of Immunology*, 34, 635–659.  
586 <https://doi.org/10.1146/annurev-immunol-041015-055515>

587 Burton, D. R., Poignard, P., Stanfield, R. L., & Wilson, I. A. (2012). Broadly neutralizing antibodies  
588 present new prospects to counter highly antigenically diverse viruses. *Science*, 337(6091),  
589 183–186. <https://doi.org/10.1126/science.1225416>

590 Capoferri, A. A., Lamers, S. L., Grabowski, M. K., Rose, R., Wawer, M. J., Serwadda, D., Gray, R.  
591 H., Quinn, T. C., Kigozi, G., Kagaayi, J., Laeyendecker, O., Abeler-Dörner, L., Ayles, H.,  
592 Bonsall, D., Bowden, R., Calvez, V., Cohen, M., Denis, A., Frampton, D., ... Tanser, F. (2020).  
593 Recombination Analysis of Near Full-Length HIV-1 Sequences and the Identification of a  
594 Potential New Circulating Recombinant Form from Rakai, Uganda. *AIDS Research and*  
595 *Human Retroviruses*, 36(6), 467–474. <https://doi.org/10.1089/aid.2019.0150>

596 Cardoso, R. M. F., Zwick, M. B., Stanfield, R. L., Kunert, R., Binley, J. M., Katinger, H., Burton, D.  
597 R., & Wilson, I. A. (2005). Broadly neutralizing anti-HIV antibody 4E10 recognizes a helical  
598 conformation of a highly conserved fusion-associated motif in gp41. *Immunity*, 22(2), 163–  
599 173. <https://doi.org/10.1016/j.immuni.2004.12.011>

600 Chang, L. W., Grabowski, M. K., Ssekubugu, R., Nalugoda, F., Kigozi, G., Nantume, B., Lessler, J.,  
601 Moore, S. M., Quinn, T. C., Reynolds, S. J., Gray, R. H., Serwadda, D., & Wawer, M. J. (2016).  
602 Heterogeneity of the HIV epidemic in agrarian, trading, and fishing communities in Rakai,  
603 Uganda: an observational epidemiological study. *The Lancet HIV*, 3(8), e388–e396.  
604 [https://doi.org/10.1016/S2352-3018\(16\)30034-0](https://doi.org/10.1016/S2352-3018(16)30034-0)

605 Conroy, S. A., Laeyendecker, O., Redd, A. D., Collinson-Streng, A., Kong, X., Makumbi, F., Lutalo,  
606 T., Sewankambo, N., Kiwanuka, N., Gray, R. H., Wawer, M. J., Serwadda, D., & Quinn, T. C.  
607 (2010). Changes in the Distribution of HIV Type 1 Subtypes D and A in Rakai District,  
608 Uganda Between 1994 and 2002. *AIDS Research and Human Retroviruses*, 26(10), 1087–  
609 1091. <https://doi.org/10.1089/aid.2010.0054>

610 Dennis, A. M., Volz, E., Frost, A. S. Md. S. D. W., Hossain, M., Poon, A. F. Y., Rebeiro, P. F.,  
611 Vermund, S. H., Sterling, T. R., & Kalish, M. L. (2018). HIV-1 Transmission Clustering and  
612 Phylodynamics Highlight the Important Role of Young Men Who Have Sex with Men. *AIDS*  
613 *Research and Human Retroviruses*, 34(10), 879–888.  
614 <https://doi.org/10.1089/aid.2018.0039>

615 Eshleman, S. H., Gonzales, M. J., Becker-Pergola, G., Cunningham, S. C., Guay, L. A., Jackson, J. B.,  
616 & Shafer, R. W. (2002). Sequence Note Identification of Ugandan HIV Type 1 Variants with  
617 Unique Patterns of Recombination in pol Involving Subtypes A and D. In *AIDS RESEARCH*  
618 *AND HUMAN RETROVIRUSES* (Vol. 18, Issue 7). [www.liebertpub.com](http://www.liebertpub.com)

619 Gall, A., Ferns, B., Morris, C., Watson, S., Cotten, M., Robinson, M., Berry, N., Pillay, D., & Kellam,  
620 P. (2012). Universal amplification, next-generation sequencing, and assembly of HIV-1  
621 genomes. *Journal of Clinical Microbiology*, 50(12), 3838–3844.  
622 <https://doi.org/10.1128/JCM.01516-12>

623 Grabowski, M. K., Lessler, J., Bazaale, J., Nabukalu, D., Nankinga, J., Nantume, B., Ssekasanvu, J.,  
624 Reynolds, S. J., Ssekubugu, R., Nalugoda, F., Kigozi, G., Kagaayi, J., Santelli, J. S., Kennedy, C.,  
625 Wawer, M. J., Serwadda, D., Chang, L. W., & Gray, R. H. (2020). Migration, hotspots, and  
626 dispersal of HIV infection in Rakai, Uganda. *Nature Communications*, *11*(1), 976.  
627 <https://doi.org/10.1038/s41467-020-14636-y>

628 Grabowski, M. K., Lessler, J., Redd, A. D., Kagaayi, J., Laeyendecker, O., Ndyanabo, A., Nelson, M.  
629 I., Cummings, D. A. T., Bwanika, J. B., Mueller, A. C., Reynolds, S. J., Munshaw, S., Ray, S. C.,  
630 Lutalo, T., Manucci, J., Tobian, A. A. R., Chang, L. W., Beyrer, C., Jennings, J. M., ... Gray, R. H.  
631 (2014). The Role of Viral Introductions in Sustaining Community-Based HIV Epidemics in  
632 Rural Uganda: Evidence from Spatial Clustering, Phylogenetics, and Egocentric Transmission  
633 Models. *PLoS Medicine*, *11*(3), e1001610. <https://doi.org/10.1371/journal.pmed.1001610>

634 Grabowski, M. K., Serwadda, D. M., Gray, R. H., Nakigozi, G., Kigozi, G., Kagaayi, J., Ssekubugu,  
635 R., Nalugoda, F., Lessler, J., Lutalo, T., Galiwango, R. M., Makumbi, F., Kong, X., Kabatesi, D.,  
636 Alamo, S. T., Wiersma, S., Sewankambo, N. K., Tobian, A. A. R., Laeyendecker, O., ... Chang,  
637 L. W. (2017). HIV Prevention Efforts and Incidence of HIV in Uganda. *New England Journal*  
638 *of Medicine*, *377*(22), 2154–2166. <https://doi.org/10.1056/NEJMoa1702150>

639 Grant, H. E., Hodcroft, E. B., Ssemwanga, D., Kitayimbwa, J. M., Yebra, G., Esquivel Gomez, L. R.,  
640 Frampton, D., Gall, A., Kellam, P., de Oliveira, T., Bbosa, N., Nsubuga, R. N., Kibengo, F.,  
641 Kwan, T. H., Lycett, S., Kao, R., Robertson, D. L., Ratmann, O., Fraser, C., ... Leigh Brown, A. J.  
642 (2020). Pervasive and non-random recombination in near full-length HIV genomes from  
643 Uganda. *Virus Evolution*, *6*(1). <https://doi.org/10.1093/ve/veaa004>

644 Harris, M. E., Serwadda, D., Sewankambo, N., Kim, B., Kigozi, G., Kiwanuka, N., Phillips, J. B.,  
645 Wabwire, F., Meehen, M., Lutalo, T., Lane, J. R., Merling, R., Gray, R., Wawer, M., Birx, D. L.,  
646 Robb, M. L., & McCutchan, F. E. (2002). Among 46 Near Full Length HIV Type 1 Genome  
647 Sequences from Rakai District, Uganda, Subtype D and AD Recombinants Predominate. In  
648 *AIDS RESEARCH AND HUMAN RETROVIRUSES* (Vol. 18, Issue 17). [www.liebertpub.com](http://www.liebertpub.com)

649 Haynes, B. F., Shaw, G. M., Korber, B., Kelsoe, G., Sodroski, J., Hahn, B. H., Borrow, P., &  
650 McMichael, A. J. (2016). HIV-Host Interactions: Implications for Vaccine Design. *Cell Host &*  
651 *Microbe*, *19*(3), 292–303. <https://doi.org/10.1016/j.chom.2016.02.002>

652 Hemelaar, J., Elangovan, R., Yun, J., Dickson-Tetteh, L., Fleminger, I., Kirtley, S., Williams, B.,  
653 Gouws-Williams, E., Ghys, P. D., Abimiku, A. G., Agwale, S., Archibald, C., Avidor, B., Barbás,  
654 M. G., Barre-Sinoussi, F., Barugahare, B., Belabbes, E. H., Bertagnolio, S., Birx, D., ... Zhang,  
655 R. (2019). Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic  
656 review, global survey, and trend analysis. *The Lancet Infectious Diseases*, *19*(2), 143–155.  
657 [https://doi.org/10.1016/S1473-3099\(18\)30647-9](https://doi.org/10.1016/S1473-3099(18)30647-9)

658 Hemelaar, J., Loganathan, S., Elangovan, R., Yun, J., Dickson-Tetteh, L., & Kirtley, S. (2020).  
659 Country Level Diversity of the HIV-1 Pandemic between 1990 and 2015. *Journal of Virology*,  
660 *95*(2). <https://doi.org/10.1128/JVI.01580-20>

661 Jombart, T., Balloux, F., & Dray, S. (2010). adephylo: new tools for investigating the phylogenetic  
662 signal in biological traits. *Bioinformatics*, *26*(15), 1907–1909.  
663 <https://doi.org/10.1093/bioinformatics/btq292>

664 Joshi, K., Lessler, J., Olawore, O., Loevinsohn, G., Bushey, S., Tobian, A. A. R., & Grabowski, M. K.  
665 (2021). Declining HIV incidence in sub-Saharan Africa: a systematic review and meta-

666 analysis of empiric data. *Journal of the International AIDS Society*, 24(10), 25818.  
667 <https://doi.org/10.1002/jia2.25818>

668 Kagaayi, J., Chang, L. W., Ssempijja, V., Grabowski, M. K., Ssekubugu, R., Nakigozi, G., Kigozi, G.,  
669 Serwadda, D. M., Gray, R. H., Nalugoda, F., Sewankambo, N. K., Nelson, L., Mills, L. A.,  
670 Kabatesi, D., Alamo, S., Kennedy, C. E., Tobian, A. A. R., Santelli, J. S., Ekström, A. M., ...  
671 Reynolds, S. J. (2019). Impact of combination HIV interventions on HIV incidence in  
672 hyperendemic fishing communities in Uganda: a prospective cohort study. *The Lancet HIV*,  
673 6(10), e680–e687. [https://doi.org/10.1016/S2352-3018\(19\)30190-0](https://doi.org/10.1016/S2352-3018(19)30190-0)

674 Kaleebu, P., French, N., Mahe, C., Yirell, D., Watera, C., Lyagoba, F., Nakiyingi, J.,  
675 Rutebemberwa, A., Morgan, D., Weber, J., Gilks, C., & Whitworth, J. (2002). Effect of  
676 Human Immunodeficiency Virus (HIV) Type 1 Envelope Subtypes A and D on Disease  
677 Progression in a Large Cohort of HIV-1-Positive Persons in Uganda. In *The Journal of*  
678 *Infectious Diseases* (Vol. 185).

679 Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of  
680 populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2014(1), 1–14.  
681 <https://doi.org/10.7717/peerj.281>

682 Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7:  
683 Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–  
684 780. <https://doi.org/10.1093/molbev/mst010>

685 Kennedy, C. E., Feng, X., Ssekubugu, R., Rosen, J. G., Ssekasanvu, J., Kigozi, G., Serwadda, D., Yeh,  
686 P. T., Kagaayi, J., Quinn, T. C., Tobian, A. A. R., Wawer, M. J., Reynolds, S. J., Chang, L. W.,  
687 Grabowski, M. K., & Nalugoda, F. (2023). Population prevalence of antiretroviral therapy  
688 sharing and its association with HIV viremia in rural Uganda: a cross-sectional population-  
689 based study. *Journal of the International AIDS Society*, 26, 26135.  
690 <https://doi.org/10.1002/jia2.26135/full>

691 Kiwanuka, N., Laeyendecker, O., Quinn, T. C., Wawer, M. J., Shepherd, J., Robb, M., Kigozi, G.,  
692 Kagaayi, J., Serwadda, D., Makumbi, F. E., Reynolds, S. J., & Gray, R. H. (2009). HIV-1  
693 subtypes and differences in heterosexual HIV transmission among HIV-discordant couples  
694 in Rakai, Uganda. *AIDS*, 23(18), 2479–2484.  
695 <https://doi.org/10.1097/QAD.0b013e328330cc08>

696 Kiwanuka, N., Laeyendecker, O., Robb, M., Kigozi, G., Arroyo, M., McCutchan, F., Eller, L. A., Eller,  
697 M., Makumbi, F., Birx, D., Wabwire-Mangen, F., Serwadda, D., Sewankambo, N. K., Quinn, T.  
698 C., Wawer, M., & Gray, R. (2008). Effect of Human Immunodeficiency Virus Type 1 (HIV-1)  
699 Subtype on Disease Progression in Persons from Rakai, Uganda, with Incident HIV-1  
700 Infection. *The Journal of Infectious Diseases*, 197(5), 707–713.  
701 <https://doi.org/10.1086/527416>

702 Lamers, S. L., Rose, R., Cross, S., Rodriguez, C. W., Redd, A. D., Quinn, T. C., Serwadda, D.,  
703 Kagaayi, J., Kigozi, G., Galiwango, R., Gray, R. H., Grabowski, M. K., & Laeyendecker, O.  
704 (2020). HIV-1 Subtype Distribution and Diversity Over 18 Years in Rakai, Uganda. *AIDS*  
705 *Research and Human Retroviruses*, 36(6), 522–526. <https://doi.org/10.1089/aid.2020.0062>

706 LANL. (2023, October 5). *Los Alamos Database*.

707 Lee, G. Q., Bangsberg, D. R., Mo, T., Lachowski, C., Brumme, C. J., Zhang, W., Lima, V. D., Boum,  
708 Y., Mwebesa, B. B., Muzoora, C., Andia, I., Mbalibulha, Y., Kembabazi, A., Carroll, R.,  
709 Siedner, M. J., Haberer, J. E., Mocello, A. R., Kigozi, S. H., Hunt, P. W., ... Harrigan, P. R.

710 (2017). Prevalence and clinical impacts of HIV-1 intersubtype recombinants in Uganda  
 711 revealed by near-full-genome population and deep sequencing approaches. *AIDS*, *31*(17),  
 712 2345–2354. <https://doi.org/10.1097/QAD.0000000000001619>  
 713 Lewis, F., Hughes, G. J., Rambaut, A., Pozniak, A., & Leigh Brown, A. J. (2008). Episodic Sexual  
 714 Transmission of HIV Revealed by Molecular Phylodynamics. *PLoS Medicine*, *5*(3), e50.  
 715 <https://doi.org/10.1371/journal.pmed.0050050>  
 716 Lihana, R. W., Ssemwanga, D., Abimiku, A., & Ndembi, N. (2012). Update on HIV-1 diversity in  
 717 Africa: A decade in review. *AIDS Reviews*, *14*(2), 83–100.  
 718 Lin, J. J., Bhattacharjee, M. J., Yu, C. P., Tseng, Y. Y., & Li, W. H. (2019). Many human RNA viruses  
 719 show extraordinarily stringent selective constraints on protein evolution. *Proceedings of*  
 720 *the National Academy of Sciences of the United States of America*, *116*(38), 19009–19018.  
 721 <https://doi.org/10.1073/pnas.1907626116>  
 722 Mendes Da Silva, R. K., Monteiro de Pina Araujo, I. I., Venegas Maciera, K., Gonçalves Morgado,  
 723 M., & Lindenmeyer Guimarães, M. (2021). Genetic Characterization of a New HIV-1 Sub-  
 724 Subtype A in Cabo Verde, Denominated A8. *Viruses*, *13*(6), 1093.  
 725 <https://doi.org/10.3390/v13061093>  
 726 Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., &  
 727 Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic  
 728 Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534.  
 729 <https://doi.org/10.1093/molbev/msaa015>  
 730 Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., Eren, K., Pollner,  
 731 T., Martin, D. P., Smith, D. M., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Gene-wide  
 732 identification of episodic selection. *Molecular Biology and Evolution*, *32*(5), 1365–1371.  
 733 <https://doi.org/10.1093/molbev/msv035>  
 734 Ng'uni, T., Chasara, C., & Ndhlovu, Z. M. (2020). Major Scientific Hurdles in HIV Vaccine  
 735 Development: Historical Perspective and Future Directions. *Frontiers in Immunology*,  
 736 *11*(October), 1–17. <https://doi.org/10.3389/fimmu.2020.590780>  
 737 Ofek, G., Tang, M., Sambor, A., Katinger, H., Mascola, J. R., Wyatt, R., & Kwong, P. D. (2004).  
 738 Structure and Mechanistic Analysis of the Anti-Human Immunodeficiency Virus Type 1  
 739 Antibody 2F5 in Complex with Its gp41 Epitope. *Journal of Virology*, *78*(19), 10724–10737.  
 740 <https://doi.org/10.1128/JVI.78.19.10724-10737.2004>  
 741 Olawore, O., Tobian, A. A. R., Kagaayi, J., Bazaale, J. M., Nantume, B., Kigozi, G., Nankinga, J.,  
 742 Nalugoda, F., Nakigozi, G., Kigozi, G., Gray, R. H., Wawer, M. J., Ssekubugu, R., Santelli, J. S.,  
 743 Reynolds, S. J., Chang, L. W., Serwadda, D., & Grabowski, M. K. (2018). Migration and risk of  
 744 HIV acquisition in Rakai, Uganda: a population-based cohort study. *The Lancet HIV*, *5*(4),  
 745 e181–e189. [https://doi.org/10.1016/S2352-3018\(18\)30009-2](https://doi.org/10.1016/S2352-3018(18)30009-2)  
 746 Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular  
 747 approach. In *Bioinformatics* (Vol. 26, Issue 3, pp. 419–420).  
 748 <https://doi.org/10.1093/bioinformatics/btp696>  
 749 Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and  
 750 evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528.  
 751 <https://doi.org/10.1093/bioinformatics/bty633>

752 Peeters, M., Toure-Kane, C., & Nkengasong, J. N. (2003). Genetic diversity of HIV in Africa:  
753 Impact on diagnosis, treatment, vaccine development and trials. *Aids*, *17*(18), 2547–2560.  
754 <https://doi.org/10.1097/00002030-200312050-00002>

755 Peters, P. J., Pontones, P., Hoover, K. W., Patel, M. R., Galang, R. R., Shields, J., Blosser, S. J.,  
756 Spiller, M. W., Combs, B., Switzer, W. M., Conrad, C., Gentry, J., Khudyakov, Y., Waterhouse,  
757 D., Owen, S. M., Chapman, E., Roseberry, J. C., McCants, V., Weidle, P. J., ... Duwve, J. M.  
758 (2016). HIV Infection Linked to Injection Use of Oxymorphone in Indiana, 2014–2015. *New*  
759 *England Journal of Medicine*, *375*(3), 229–239. <https://doi.org/10.1056/nejmoa1515195>

760 Pillay, D., Herbeck, J., Cohen, M. S., Oliveira, T. de, Fraser, C., Ratmann, O., Brown, A. L., &  
761 Kellam, P. (2015). PANGEA-HIV: Phylogenetics for generalised epidemics in Africa. In *The*  
762 *Lancet Infectious Diseases* (Vol. 15, Issue 3, pp. 259–261). Lancet Publishing Group.  
763 [https://doi.org/10.1016/S1473-3099\(15\)70036-8](https://doi.org/10.1016/S1473-3099(15)70036-8)

764 Pineda-Peña, A. C., Faria, N. R., Imbrechts, S., Libin, P., Abecasis, A. B., Deforche, K., Gómez-  
765 López, A., Camacho, R. J., De Oliveira, T., & Vandamme, A. M. (2013). Automated subtyping  
766 of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation  
767 of the new REGA version 3 and seven other tools. *Infection, Genetics and Evolution*, *19*,  
768 337–348. <https://doi.org/10.1016/j.meegid.2013.04.032>

769 R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for  
770 Statistical Computing. <https://www.r-project.org/>

771 Raghwani, J., Redd, A. D., Longosz, A. F., Wu, C. H., Serwadda, D., Martens, C., Kagaayi, J.,  
772 Sewankambo, N., Porcella, S. F., Grabowski, M. K., Quinn, T. C., Eller, M. A., Eller, L. A.,  
773 Wabwire-Mangen, F., Robb, M. L., Fraser, C., & Lythgoe, K. A. (2018). Evolution of HIV-1  
774 within untreated individuals and at the population scale in Uganda. *PLoS Pathogens*, *14*(7).  
775 <https://doi.org/10.1371/journal.ppat.1007167>

776 Redd, A. D., Mullis, C. E., Serwadda, D., Kong, X., Martens, C., Ricklefs, S. M., Tobian, A. A. R.,  
777 Xiao, C., Grabowski, M. K., Nalugoda, F., Kigozi, G., Laeyendecker, O., Kagaayi, J.,  
778 Sewankambo, N., Gray, R. H., Porcella, S. F., Wawer, M. J., & Quinn, T. C. (2012). The Rates  
779 of HIV Superinfection and Primary HIV Incidence in a General Population in Rakai, Uganda.  
780 *The Journal of Infectious Diseases*, *206*(2), 267–274. <https://doi.org/10.1093/infdis/jis325>

781 Redd, A. D., Quinn, T. C., & Tobian, A. A. (2013). Frequency and implications of HIV  
782 superinfection. *The Lancet Infectious Diseases*, *13*(7), 622–628.  
783 [https://doi.org/10.1016/S1473-3099\(13\)70066-5](https://doi.org/10.1016/S1473-3099(13)70066-5)

784 Sadanand, S., Suscovich, T. J., & Alter, G. (2016). Broadly neutralizing antibodies against HIV:  
785 New insights to inform vaccine design. *Annual Review of Medicine*, *67*, 185–200.  
786 <https://doi.org/10.1146/annurev-med-091014-090749>

787 Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*,  
788 *27*(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>

789 Siepel, A. C., Halpern, A. L., Macken, C., & Korber, B. T. M. (1995). A Computer Program  
790 Designed to Screen Rapidly for HIV Type 1 Intersubtype Recombinant Sequences. *AIDS*  
791 *Research and Human Retroviruses*, *11*(11), 1413–1416.  
792 <https://doi.org/10.1089/aid.1995.11.1413>

793 Ssemwanga, D., Nsubuga, R. N., Mayanja, B. N., Lyagoba, F., Magambo, B., Yirrell, D., van der  
794 Paal, L., Grosskurth, H., & Kaleebu, P. (2013). Effect of HIV-1 Subtypes on Disease

795 Progression in Rural Uganda: A Prospective Clinical Cohort Study. *PLoS ONE*, 8(8).  
796 <https://doi.org/10.1371/journal.pone.0071768>

797 Steichen, J. M., Lin, Y. C., Havenar-Daughton, C., Pecetta, S., Ozorowski, G., Willis, J. R., Toy, L.,  
798 Sok, D., Liguori, A., Kratochvil, S., Torres, J. L., Kalyuzhniy, O., Melzi, E., Kulp, D. W.,  
799 Raemisch, S., Hu, X., Bernard, S. M., Georgeson, E., Phelps, N., ... Schief, W. R. (2019). A  
800 generalized HIV vaccine design strategy for priming of broadly neutralizing antibody  
801 responses. *Science*, 366(6470). <https://doi.org/10.1126/science.aax4380>

802 Struck, D., Lawyer, G., Ternes, A.-M., Schmit, J.-C., & Bercoff, D. P. (2014). COMET: adaptive  
803 context-based modeling for ultrafast HIV-1 subtype identification. *Nucleic Acids Research*,  
804 42(18), e144. <https://doi.org/10.1093/nar/gku739>

805 Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA  
806 polymorphism. *Genetics*, 123(3), 585–595. <https://doi.org/10.1093/genetics/123.3.585>

807 Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the  
808 control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and*  
809 *Evolution*. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>

810 Taylor, B. S., Sobieszczyk, M. E., McCutchan, F. E., & Hammer, S. M. (2008). The Challenge of HIV-  
811 1 Subtype Diversity. *New England Journal of Medicine*, 358(15), 1590–1602.  
812 <https://doi.org/10.1056/nejmra0706737>

813 Taylor, J. E., & Korber, B. T. (2005). HIV-1 intra-subtype superinfection rates: estimates using a  
814 structured coalescent with recombination. *Infection, Genetics and Evolution : Journal of*  
815 *Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 5(1), 85–95.  
816 <https://doi.org/10.1016/j.meegid.2004.07.001>

817 UNAIDS. (2023). *The path that ends AIDS: UNAIDS Global AIDS Update 2023*.  
818 <http://www.wipo.int/amc/en/mediation/rules>

819 Vidal, N., Peeters, M., Mulanga-Kabeya, C., Nzilambi, N., Robertson, D., Ilunga, W., Sema, H.,  
820 Tshimanga, K., Bongo, B., & Delaporte, E. (2000). Unprecedented Degree of Human  
821 Immunodeficiency Virus Type 1 (HIV-1) Group M Genetic Diversity in the Democratic  
822 Republic of Congo Suggests that the HIV-1 Pandemic Originated in Central Africa. *Journal of*  
823 *Virology*, 74(22), 10498–10507. <https://doi.org/10.1128/jvi.74.22.10498-10507.2000>

824 Wawer, M. J., Gray, R. H., Sewankambo, N. K., Serwadda, D., Paxton, L., Berkley, S., McNairn, D.,  
825 Wabwire-Mangen, F., Li, C., Nalugoda, F., Kiwanuka, N., Lutalo, T., Brookmeyer, R., Kelly, R.,  
826 & Quinn, T. C. (1998). A randomized, community trial of intensive sexually transmitted  
827 disease control for AIDS prevention, Rakai, Uganda. *Aids*, 12(10), 1211–1225.  
828 <https://doi.org/10.1097/00002030-199810000-00014>

829 Wymant, C., Blanquart, F., Golubchik, T., Gall, A., Bakker, M., Bezemer, D., Croucher, N. J., Hall,  
830 M., Hillebregt, M., Ong, S. H., Ratmann, O., Albert, J., Bannert, N., Fellay, J., Fransen, K.,  
831 Gourlay, A., Grabowski, M. K., Günsenheimer-Bartmeyer, B., Günthard, H. F., ... Fraser, C.  
832 (2018). Easy and accurate reconstruction of whole HIV genomes from short-read sequence  
833 data with shiver. *Virus Evolution*, 4(1). <https://doi.org/10.1093/ve/vey007>

834 Yamaguchi, J., Vallari, A., McArthur, C., Sthreshley, L., Cloherty, G. A., Berg, M. G., & Rodgers, M.  
835 A. (2020). Brief Report: Complete Genome Sequence of CG-0018a-01 Establishes HIV-1  
836 Subtype L. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 83(3), 319–322.  
837 <https://doi.org/10.1097/QAI.0000000000002246>

838 Yang, C., Dash, B. C., Simon, F., van der Groen, G., Pieniazek, D., Gao, F., Hahn, B. H., & Lal, R. B.  
839 (2000). Detection of Diverse Variants of Human Immunodeficiency Virus–1 Groups M, N,  
840 and O and Simian Immunodeficiency Viruses from Chimpanzees by Using Generic pol and  
841 env Primer Pairs. *The Journal of Infectious Diseases*, *181*(5), 1791–1795.  
842 <https://doi.org/10.1086/315439>

843 Yang, Z. (1995). A space-time process model for the evolution of DNA sequences. *Genetics*,  
844 *139*(2), 993–1005. <https://doi.org/10.1093/genetics/139.2.993>

845 Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and*  
846 *Evolution*, *24*(8), 1586–1591. <https://doi.org/10.1093/molbev/msm088>

847 Yang, Z., & Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends*  
848 *in Ecology & Evolution*, *15*(12), 496–503. [https://doi.org/10.1016/S0169-5347\(00\)01994-7](https://doi.org/10.1016/S0169-5347(00)01994-7)

849 Yirrell, D. L., Kaleebu, P., Morgan, D., Watera, C., Magambo, B., Lyagoba, F., & Whitworth, J.  
850 (2002). Inter- and intra-genic intersubtype HIV-1 recombination in rural and semi-urban  
851 Uganda. *AIDS*, *16*(2), 279–286. <https://doi.org/10.1097/00002030-200201250-00018>  
852  
853  
  
854  
  
855  
  
856