

Unique genetic signatures in HIV-1 subtype A1 and A1D recombinant envelope glycoprotein distinguish contemporary transmitted/founder viruses from historical strains in East Africa

Foley, Brian Thomas; Kapaata, Anne; Kato, Frank; Galiwango, Ronald; Nakyanzi, Angella; Ndekezi, Christian; Natwijuka, Fortunate; Omara, Denis; Obuku, Andrew Ekii; Kaleebu, Pontiano; Nduati, Eunice; Balinda, Sheila Nina

Provided by the author(s) and the Los Alamos National Laboratory

To be published in: Frontiers in Microbiology, Article 16, iss. 1632581, 2025-08-04

DOI to publisher's version: 10.3389/fmicb.2025.1632581

Permalink to record:

<https://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-25-29005>



Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.



OPEN ACCESS

EDITED BY

Hao Wu,
Capital Medical University, China

REVIEWED BY

Tawanda Mandizvo,
IAVI, United States
Zetao Cheng,
National Cancer Institute at Frederick (NIH),
United States

*CORRESPONDENCE

Frank Kato
✉ Frank.Kato@mrcuganda.org

[†]These authors have contributed equally to this work

RECEIVED 21 May 2025

ACCEPTED 17 July 2025

PUBLISHED 04 August 2025

CITATION

Kato F, Kapaata A, Galiwango R, Nakyanzi A, Ndekezi C, Natwijuka F, Omara D, Obuku AE, Foley B, Kaleebu P, Nduati E and Balinda SN (2025) Unique genetic signatures in HIV-1 subtype A1 and A1D recombinant envelope glycoprotein distinguish contemporary transmitted/founder viruses from historical strains in East Africa. *Front. Microbiol.* 16:1632581. doi: 10.3389/fmicb.2025.1632581

COPYRIGHT

© 2025 Kato, Kapaata, Galiwango, Nakyanzi, Ndekezi, Natwijuka, Omara, Obuku, Foley, Kaleebu, Nduati and Balinda. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Unique genetic signatures in HIV-1 subtype A1 and A1D recombinant envelope glycoprotein distinguish contemporary transmitted/founder viruses from historical strains in East Africa

Frank Kato^{1,2*}, Anne Kapaata^{1†}, Ronald Galiwango^{3,4}, Angella Nakyanzi⁵, Christian Ndekezi^{1,2}, Fortunate Natwijuka^{1,2}, Denis Omara^{1,2}, Andrew Ekii Obuku¹, Brian Foley⁶, Pontiano Kaleebu^{1,5}, Eunice Nduati⁷ and Sheila Nina Balinda^{1,5}

¹MRC/UVRI & LSHTM Uganda Research Unit, Department of Viral Pathogens, Entebbe, Uganda, ²Department of Immunology and Molecular Biology, College of Health Sciences, Makerere University, Kampala, Uganda, ³The African Centre of Excellence in Bioinformatics and Data Intensive Sciences, Kampala, Uganda, ⁴The Infectious Diseases Institute, Makerere University, Kampala, Uganda, ⁵Uganda Virus Research Institute, Entebbe, Uganda, ⁶Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM, United States, ⁷KEMRI Wellcome Trust Research Programme, Kilifi, Kenya

Introduction: The envelope glycoprotein (Env) of HIV-1 Transmitted/Founder (T/F) viruses in subtypes B and C carries distinct genetic signatures that enhance transmission fitness, augment infectivity and immune evasion. However, there is limited data on such signatures in T/F subtypes A1, D and A1D recombinants that predominate East Africa's HIV epidemic.

Methods: We used phylogenetically corrected approaches to detect distinct genetic signatures by comparing 44 contemporary HIV-1 T/F Envs with 229 historical Envs of the same subtype in East Africa.

Results and Discussion: Subtype analysis based on the full-length Env gene of contemporary T/F viruses revealed a high proportion of subtype A1, followed by A1D recombinants, and fewer subtype D. Signature analysis revealed that the contemporary subtype A1 T/Fs were more likely to select distinct amino acids, including M22 in the signal peptide, R82 in gp120, A172 in the V2 loop, E230 in the glycosite 230, K275 in the D loop, Y317 in the V3 loop, K476 and N477 in the CD4 contact site, when compared with the historical Envs (q-value < 0.2). Conversely, the contemporary subtype A1 T/F Envs were less likely to carry the amino acids Q432 in the CD4 contact site, and the L784 signature within the LLP-2 (q-value < 0.2). The A1D recombinant T/Fs were more likely to select the D620 in the C-helix, but under selected the L34 in gp120, P299 in the V3 loop and Y643 in the Heptad repeat-2, compared to the historical Envs (q-value < 0.2). The distinct signature sites reported in this study may contribute to the successful establishment of acute infection as well as the persistence of long-term infection. Therefore, effective therapeutics and vaccines may target these distinct amino acid signatures especially for the East African region as it may be necessary to employ subtype-specific vaccines according to the subtype distribution.

KEYWORDS

HIV-1, envelope, subtype A1, A1D recombinants, signature, transmitted/founder, genetic

1 Introduction

HIV-1 remains a major health concern, particularly in East and Southern Africa. Vaccine development has been hampered by viral genetic diversity and immune escape. HIV group M, responsible for the global pandemic, is divided into 10 subtypes (A-D, F-H, J, K, and L), six A subtypes (A1-A6), F subtypes (F1-F2), circulating recombinant forms (CRFs) and unique recombinant forms (URFs) (Désiré et al., 2018; Reis et al., 2019; Robertson et al., 2000; Yamaguchi et al., 2020). In the East African region, subtypes A1, D, and a recent increase in A1D recombinants drive the pandemic (Balinda et al., 2022; Bbosa et al., 2019; Grant et al., 2020; Adhiambo et al., 2021; Kemal et al., 2013; Yang et al., 2004). The increasing prevalence of A1D recombinants may reflect selective pressures against pure subtype D, which exhibits lower transmissibility (Kiwana et al., 2009), faster rates of CD4 T cell loss (Kaleebu et al., 2002) and faster disease progression (Kapaata et al., 2021; Ssemwanga et al., 2013).

The most diverse gene of HIV-1 is the Env, comprising of gp41 and gp120 which are associated with viral transmission (Herbeck et al., 2006) and host cell tropism (Yang et al., 2004). Although the Env is the sole target for neutralizing antibodies (Wibmer et al., 2015), its immense diversity regulates the functional properties of the virus and aids in rapid evolution, leading to the establishment of a viral reservoir that hinders cure and vaccine development (Ndung'u et al., 2019; Van Regenmortel, 2017; Zhou et al., 2021). The Env (gp160) encoded by the viral genome harbors the transmembrane domain and interacts with the cell surface-associated receptor (CD4) and coreceptors (CCR5 and CXCR4) by one of its non-covalently associated gp120 subunit spikes linked to gp41 (Checkley et al., 2011).

Among the broad repertoire of viral variants that circulate in an infected individual, the T/F virus that successfully establishes productive infection in a recipient is of utmost importance for prevention strategies. Understanding key genetic features of contemporary T/F viruses provides insights into mechanisms underlying transmission, which is important for both vaccine design and therapeutic interventions. Unique genetic signatures have been identified among subtype B T/F Env sequences, for example a Histidine signature in position 12 in the signal peptide (SP) and loss of an N-linked glycosylation site at positions 413–415 were associated with high Env expression levels in acute infection. This indicates that immune evasion patterns that recur in many individuals during chronic infection when antibodies are present can be selected against when the infection is being established (Gnanakaran et al., 2011). Also, an Isoleucine at position 841 instead of arginine in gp41 CT (LLP-1) was enriched in subtype B Env (Kafando et al., 2019). A K6I mutation located at the signal peptide region was found more likely among chronic viruses than the T/F viruses among subtype B Env (Kafando et al., 2019). The RV144 vaccine signature sites; lysine at position 169 in V2 and Isoleucine at position 307 in V3 loops occurred less in subtype C T/F Env sequences compared to chronic viruses (Rademeyer et al., 2016). The above studies emphasize that HIV-1 T/F viruses possess inherent properties for establishment in a new host. However, there is limited

data on the unique genetic signatures in the Env for HIV-1 T/F subtypes A1, D, and A1D recombinants circulating in the East African region.

To address this gap, we utilized Single genome amplification (SGA), Sanger sequencing, Entropy tool, GenSig tool, and analyze align tool to detect the distinct genetic signatures in the Env gene of contemporary HIV-1 T/F viruses from the East African region. These were compared to historical viruses of the same subtype. Additionally, the HIV genome browser was deployed for associating the detected genetic signatures with Env protein structural characteristics to predict their sensitivity or resistance to broadly neutralizing antibodies (bnAbs).

2 Materials and methods

2.1 Study design

This was a cross-sectional study involving laboratory data generated for the contemporary HIV-1 T/F and historical viral sequences downloaded from the Los Alamos National Laboratory HIV sequence database (LANL HIV sequence db). To generate the full-length T/F Env, we used samples from the International AIDS Vaccine Initiative's (IAVI) Virus Surveillance study working with recent infection cohorts, i.e., Protocol N in Kenya, Good Health for Women Project (GHWP) and KILGORIS at the MRC/UVRI and LSHTM Uganda Research Unit. The samples were collected between 2015 and 2021. Briefly, GHWP was a prospective cohort of women at risk who are involved in commercial sex in Uganda (Mayanja et al., 2020; Vandepitte et al., 2013). KILIGORIS cohort was a pilot study to evaluate the possibility of identifying, enrolling and following up a high-risk cohort of acute HIV infection in Uganda. Protocol N was an observation study to determine the immune and viral characteristics during acute infection in Kenya. Protocol N participants were identified from high-risk populations, e.g., Protocol B, where, despite HIV prevention care and counselling, a small proportion of HIV incidence cases occurred (Price et al., 2020). The cohort characteristics of these early infection cohorts are shown in [Supplementary Table S2](#). We purposively selected acutely infected samples from participants who had seroconverted to HIV-1, screened for early HIV infection and those between Fiebig stage 1 to V. These acute samples were included in the present study if sufficient plasma was sufficient for Single Genome Amplification analysis. The contemporary T/F Envs from early infection cohorts were supplemented with publicly available and similar T/F Env sequences from IAVI's protocol C in Rwanda. These additional T/F Envs from Rwanda were retrieved from the CATNAP webserver <http://hiv.lanl.gov/catnap> (Umviligihozo et al., 2021; Yoon et al., 2015) and Super Filtered (SFL) web alignments <https://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html#filter> at LANL HIV db. Also, the historical full length Env sequences from Uganda, Kenya, Tanzania and Rwanda were downloaded from the CATNAP webserver and SFL web alignments at LANL HIV db.

2.2 Ethical consideration

Ethics approvals were obtained for the GHWP from Uganda Virus Research Institute-Research and Ethics Committee (UVRI-REC) (GC 127). Protocol N cohort was approved by the KEMRI Scientific and Ethics Review Unit (SERU) (KEMRI/RES/7/3/1). Kiligoris cohort was approved by UVRI-REC (GC/127/714). Also, this study was approved by the School of Biomedical Sciences-Research and Ethics Committee (SBS-REC) (Ref: SBS-2023-38) at Makerere University. The permission to use the archived plasma samples (from GHWP, Protocol N and Kiligoris cohorts) was sought from the MRC/UVRI and LSHTM Uganda Research Unit.

2.3 Sample size

In this study, full-length contemporary HIV-1 T/F Env sequences ($n = 44$) were compared with historical Env sequences ($n = 229$) of similar subtypes from East Africa. Contemporary T/F Env sequences were generated from the laboratory at the MRC/UVRI and LSHTM Uganda Research Unit from recent cohorts as follows: GHWP ($n = 18$), Protocol N ($n = 11$), and KILIGORIS ($n = 7$). The accession numbers of the publicly available contemporary T/F Envs from IAVI's Protocol C ($n = 8$) and historical sequences retrieved from the LANL HIV db are available in the [Supplementary Tables S3–S6](#). The historical Env sequences comprised of old acute ($n = 24$), chronic ($n = 200$) and viruses of unknown infection phase ($n = 5$), and spanning the years 1986 to 2006.

2.4 Viral RNA isolation and cDNA synthesis

The Viral RNA was extracted from 140 μ L of archived EDTA plasma (from the IAVI's protocol N, GHWP and Kiligoris cohorts) using the Qiagen viral RNA extraction kit (Qiagen Inc., Valencia, CA, United States) following the manufactures' instructions. The recovered RNA was converted into complimentary DNA (cDNA) using SuperScript IV reverse transcriptase (Invitrogen, Ljubljana, Slovenia) as previously described ([Salazar-Gonzalez et al., 2009](#)). The cDNA templates for complete HIV-1 Env single genome amplification were synthesized with the reverse primer 1. R3B3R ([Supplementary Table S1](#)). The cDNA was used immediately to generate single genome amplicons.

2.5 Single genome amplification

The event of inter-subtype recombinants *in vivo* and artificial recombinants that may be generated *in vitro* because of template switching during bulk amplification of heterogeneous cDNA target sequences confound earlier findings on acute HIV-1 infection. A common strategy to tackle these challenges has been to identify participants within the acute phase of infection and using SGA, derive viral sequences from proviral DNA or plasma RNA, followed by sequencing, and phylogenetic analysis ([Salazar-Gonzalez et al., 2009](#)). Similarly, we serially diluted the cDNA and performed nested PCR amplification with HIV-1 specific primers ([Supplementary Table S1](#)) as previously described ([Salazar-Gonzalez et al., 2009](#)). All products

derived from cDNA dilutions and PCR amplifications yielding <30% positive wells and amplicon length (± 2.6 -kb) were subjected to Sanger sequencing.

2.6 Sanger sequencing

To confirm amplification from single cDNA templates and avoid *in vitro* PCR artefacts, 5–10 SGA derived amplicons per participant were sequenced using BigDye Terminator v3.1 chemistry (Applied Biosystems, Foster City, CA), and an Applied Biosystems 3500xl Genetic analyzer (Thermo Fisher Scientific, Foster City, CA, United States).

2.7 HIV-1T/F identification from recent infection cohorts

The raw sequence files were acquired from the genetic analyzer, base-called and *de novo* assembled using the Sequencer program (v5.4.6; Gene Codes, Ann Arbor, MI) ([Salazar-Gonzalez et al., 2008](#)). The assembled full-length Env sequences were aligned using MAFFT package v7.505 ([Katoh et al., 2002](#)). Maximum likelihood tree construction for the aligned SGA Env sequences was done using IQ-TREE v2.0.3 ([Nguyen et al., 2015](#); [Trifunopoulos et al., 2016](#)), with free rate of evolution, GTR model, and 1,000 bootstrap replicates, including multiple sequences from each participant to rule out contamination issues. The tree file (Newick format) was visualized using the Figtree package v1.4.4 ([Rambaut, 2010](#)). To identify and enumerate contemporary HIV-1 T/F variants, we used maximum likelihood tree reconstruction for within participant-specific phylogenetic clustering and the LANL HIV sequence visualization tool https://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter_top.html to generate highlighter plots and determine recombinant mosaic structures between cognate major and minor T/F variants ([Balinda et al., 2022](#); [Keele et al., 2008](#); [Macharia et al., 2020](#)).

2.8 Retrieval of historical Env sequences

The historical dataset ($n = 229$) comprised of old acute, old chronics and old Env sequences from HIV-1 viruses of unknown infection status. These were retrieved from the LANL HIV db using filters: HIV-1, subtype (either A1, D or A1D recombinants), Env CDS, patient health, days from seroconversion, Fiebig stage, days from infection, infection year, Sub-Saharan Africa, including sequences with less than 0.5 Percent non-ACGT to rule out problematic sequences. Next, we applied the one sequence per patient filter to rule out bias introduced by analyzing multiple sequences from one patient. Sequences were retrieved from LANL HIV-1 super filtered (SFL) alignments using filters: HIV-1/SIVcpz, DNA, and All M group with CRFS. In addition, sequences were retrieved from CATNAP Env alignment at LANL HIV db. This was followed by an extensive literature search and exclusion of sequences out of the sampling year range (1986–2006), non-subtype A1, non-subtype D, non-A1D recombinants and sequences sampled outside the East African region.

2.9 Retrieval of contemporary T/F Env sequences

The contemporary HIV-1 T/F Env sequences ($n = 8$) from acute infection were retrieved using the procedure in 2.8, followed by an extensive literature search and exclusion of sequences outside the sampling year range (2015–2021), chronic and sequences of unknown infection status, non-subtype A1, non-subtype D, and non-A1D recombinants and those sampled outside the East African region.

2.10 HIV-1 subtype analysis

The full-length HIV-1 T/F Env sequences were subtyped using the Recombinant identification Program (RIP, window size = 400) (Siepel et al., 1995) and jumping profile Hidden Markov Model (jpHMM) (Schultz et al., 2009).

2.11 Quality control

The contemporary T/F and historical Envs from East Africa were codon aligned using Gene Cutter https://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html. The codon-aligned sequence alignment was fed into ElimDupes tool <https://www.hiv.lanl.gov/content/sequence/elimdupesv2/elimdupes.html> to eliminate 100% identical sequences with consideration of sub-sequences as duplicates. The unique sequences from the ElimDupes tool were subjected to the quality control tool at LANL HIV db <https://www.hiv.lanl.gov/content/sequence/QC/index.html> to exclude sequences with stop codons and non-ACGT characters within the Env CDS. Maximum likelihood trees were reconstructed in IQ-TREE v2.0.3 using free rate of evolution, 1,000 bootstrap replicates and the GTR model to eliminate the sequence far away from the root in cases of homogeneous or clustering sequences (Trifinopoulos et al., 2016). The quality-controlled sequences were then used for entropy and unique genetic signature analysis. The detected signatures were associated with Env structural characteristics such as sensitivity or resistance to bnAbs using the HIV genome browser at LANL HIV db. These analyses were stratified by each subtype (A1, and A1D recombinants).

2.12 Entropy analysis of contemporary HIV-1T/F compared to historical Env sequences

For each alignment position, the Entropy-two tool <https://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy.html> was deployed to compare the variability in contemporary T/F Env subtype A1 (query) relative to historical subtype A1 Env (background). To guide against type 1 errors, the false discovery rate (q-value) for p -values (from Entropy tool) was determined using the source code <https://github.com/nfusi/qvalue> (Storey and Tibshirani, 2003). All significant entropy sites (q-value < 0.2) were visualized using Analyze align tool https://www.hiv.lanl.gov/content/sequence/ANALYZEALIGN/analyze_align.html. The entropy analysis procedure was repeated for A1D recombinants.

2.13 Genetic signature analysis of contemporary HIV-1T/F compared to historical Env sequences

Using the GenSig tool at LANL HIV sequence db, we performed a phylogenetically corrected analysis to detect unique genetic signatures in the codon aligned DNA alignment of contemporary T/F subtype A1 and historical subtype1 A1 Envs (Bhattacharya et al., 2007; Bricault et al., 2019). The GenSig searched for statistically significant signatures with a site depth of 1 which tests for the association between the contemporary T/F and historical Envs, and each amino acid in all independent sites in the alignment (Bhattacharya et al., 2007; Bricault et al., 2019). The Gensig tool provides p -values for Fisher's exact test and corresponding q-values in the signature output to minimize false positives due to lineage effects (Bhattacharya et al., 2007; Bricault et al., 2019). The absolute counts, frequency by position and weblogs for the significant genetic signatures (q-value < 0.2) were generated using Analyze align tool. The genetic signature analysis procedure was repeated for the A1D recombinants.

2.14 Exploration of the association between the unique genetic signatures and HIV-1T/F Env characteristics

The LANL HIV db genome browser tool was used to investigate the association between the significant genetic signatures and the HIV-1 T/F Env subtype structural characteristics (Skinner et al., 2009; Skinner and Holmes, 2010). Briefly, the Env option on the HIV genome browser was selected, dragged the tracks to the top left section and entered the HXB2 numbering position of the significant signature site in the search box. Also, the unique genetic signatures were associated with available bnAb data in the “HIV-1 Neutralizing Antibody signatures and applications to epitope targeted vaccine study” (Bricault et al., 2019).

3 Results

3.1 Identification of contemporary HIV-1T/F and historical Env sequences

A total of 344 SGA-derived viral sequences from 36 acutely infected individuals (GHWP, Protocol N and KILIGORIS Cohorts) were phylogenetically analyzed to infer 36 contemporary T/F Env sequences. Across the acute cohort, the mean, median and range numbers of SGA-derived sequences per participant are 9.56, 9 and 15, respectively, as shown in the [Supplementary Table S2](#). Maximum likelihood phylogenetic trees formed subject-specific lineages that conformed to a single transmission event [Figures 1, 2a–d](#). The identified T/F Env sequences from early infection cohorts were supplemented with available T/F Envs ($n = 8$) from LANL HIV sequence db, resulting in 44 T/F Env sequences for this study. These were compared with available historical Envs ($n = 229$) from LANL HIV db.

Additionally, the highlighter plot analysis showed those individuals infected by a single virus ([Figures 3a–d](#)). In all single transmission events, mismatches compared to the master (consensus) in each amplicon were randomly distributed across the complete

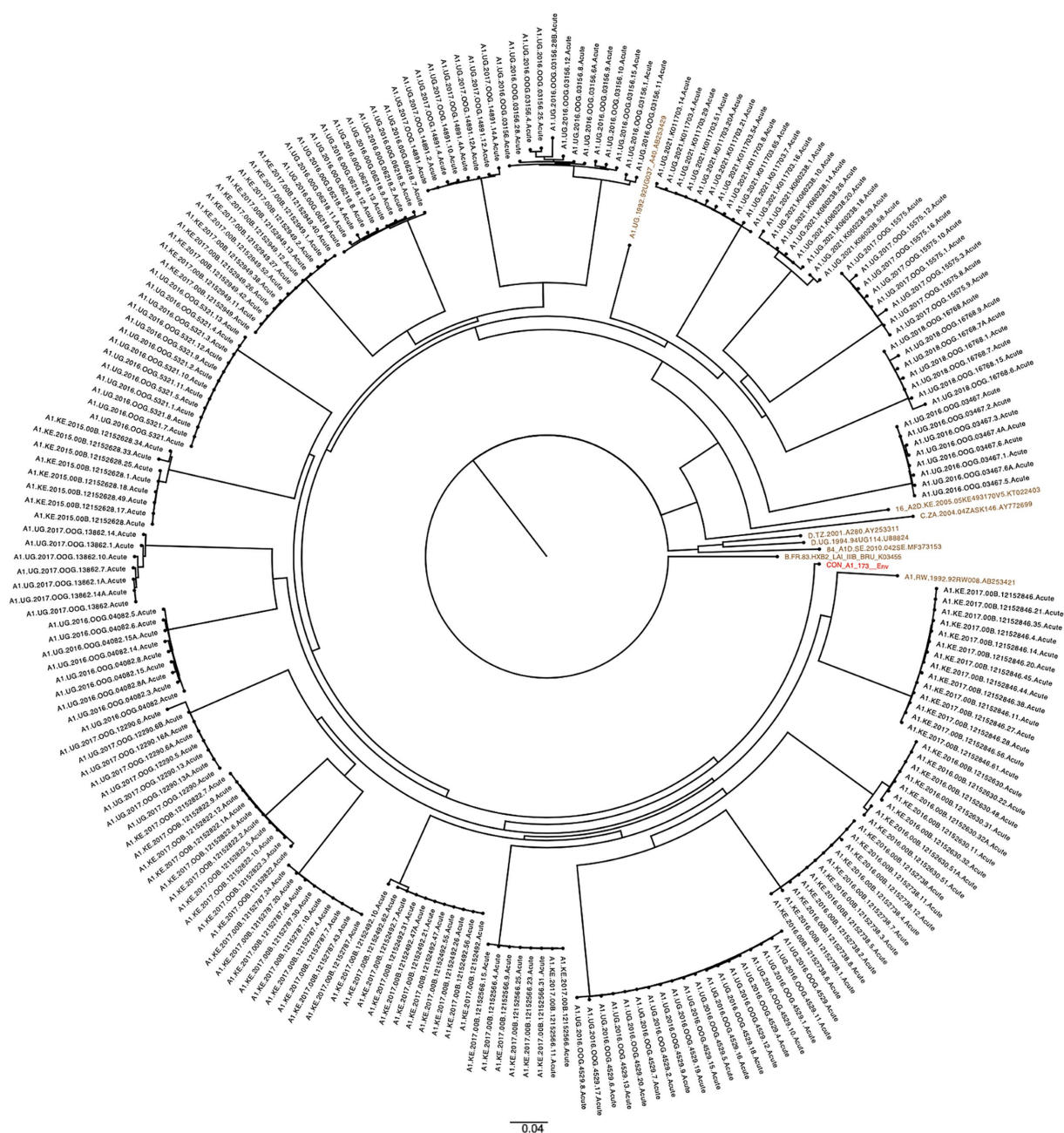


FIGURE 1

SGA-derived subtype A1 Env sequences ($n = 221$) from participants ($n = 22$) clustered into distinct lineages indicating Single transmission events. HIV-1 reference sequences (brown) and group M subtype A1 consensus sequence (red) from the LANL HIV db. The scale bar represents genetic distance.

HIV-1 Env genome. The sequence that corresponded to the consensus sequence was inferred to be the T/F sequence.

3.2 HIV-1 Env subtype A1 was more predominant than A1D recombinant and subtype D

The contemporary HIV-1 T/F Subtype A1 (68.2%, 30/44) was the most predominant, followed by A1D recombinants (20.5%, 9/44), subtype D (4.5%, 2/44), subtype C (4.5%, 2/44) and A1C recombinant

(2.3%, 1/44) (Figure 4a). Similarly, for the historical HIV-1 Env sequences, the most prevalent subtype was subtype A1(60.3%, 138/229), followed by A1D recombinants (24.5%, 56/229) and then subtype D (15.3%, 35/229) (Figure 4b).

3.3 Amino acid variation in contemporary T/F compared to historical HIV-1 Envs

In the case of subtype A1, the contemporary T/F Envs were more variable at HXB2 alignment positions (22, 82, 172, 230, 275, 317, 432,

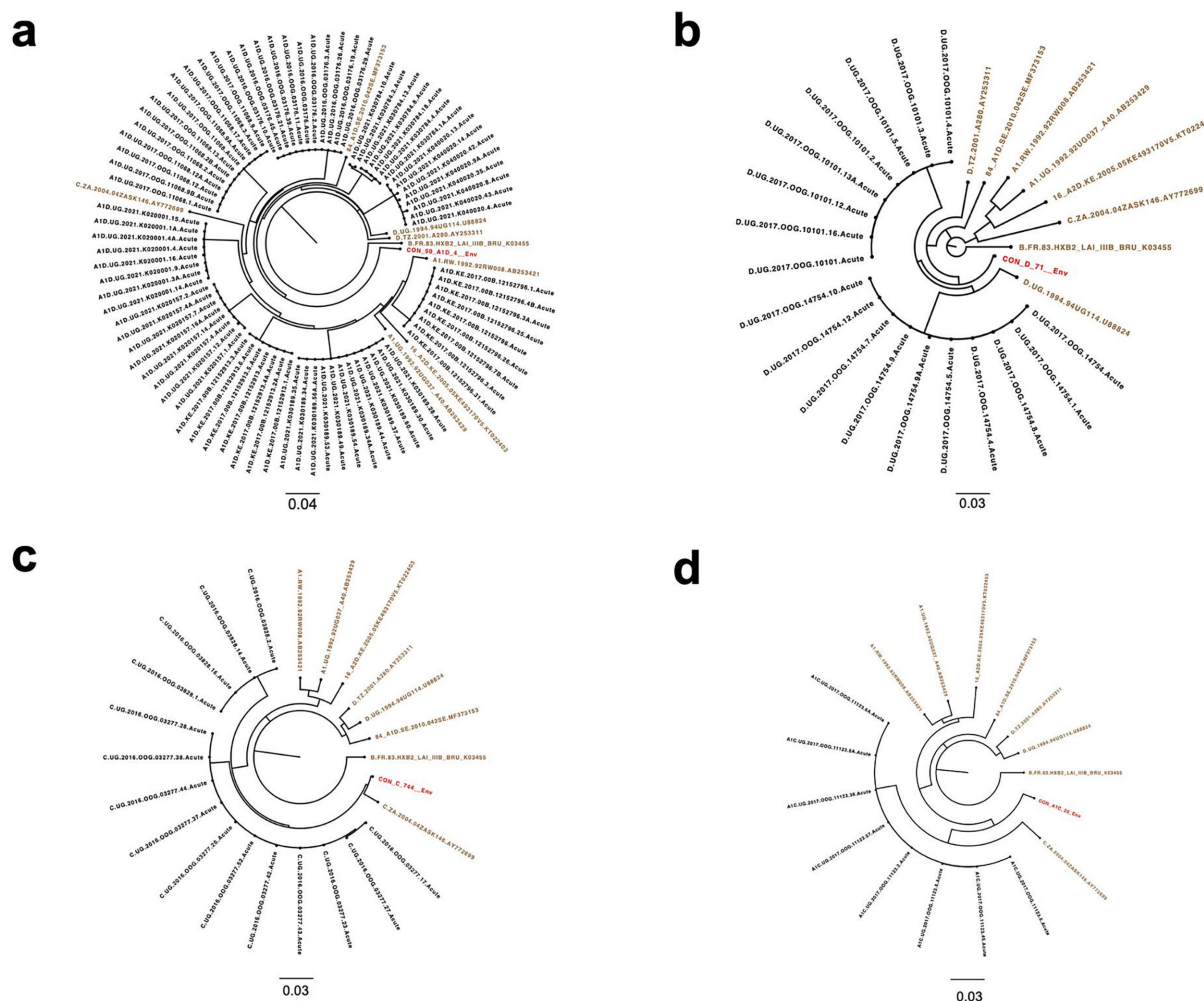


FIGURE 2

Maximum likelihood phylogenetic trees of SGA-derived Env sequences indicating single transmission events in each participant. (a) Subtype A1D recombinant Env sequences ($n = 82$) from participants ($n = 9$). (b) Subtype D Env sequences ($n = 18$) from subjects ($n = 2$). (c) Subtype C SGA sequences ($n = 15$) from participants ($n = 2$). (d) Subtype A1C Env sequences ($n = 8$) from participant ($n = 1$). HIV-1 reference sequences (brown) and group M subtypes D, C, A1D and A1C recombinant consensus sequences (red) from the LANL HIV db. The scale bar represents genetic distance.

476, 477 and 784) when compared to the historical sequences (q -value < 0.2) (Table 1). Among A1D recombinants, the T/F Envs exhibited higher diversity at HXB2 alignment positions (34, 299, and 643) when compared to the historical sequences (q -value < 0.2) (Table 1). Conversely, the A1D recombinant historical Envs were more variable at position 620 when compared to the contemporary A1D recombinant T/F (q -value < 0.2). The small sample size of subtype D sequences limited the statistical power to compare the variation of amino acids between contemporary T/F and historical subtype D Env sequences.

3.4 Unique genetic signatures in the HIV-1 Env associated with contemporary T/F compared to historical sequences

Table 2 presents the significant genetic signatures sites. We detected the robust Leucine signature at position of HXB2 numbering 22 (L22) in the hydrophobic core of the signal peptide (SP)

domain in subtype A1 (p -value = 0.000613, q -value < 0.2 , Fisher's test). The robust L22 signature site was less enriched in the SP domain of T/F subtype A1 (80.00%) compared to the historical sequences (99.28%). The L22 is a robust genetic signature site because it was supported by multiple lines of evidence, i.e., both the Entropy and GenSig tools. Instead, the T/F subtype A1 sequences (6.67%) were more likely to select a Methionine (M22) signature, compared to the historical counterparts which lacked it (p -value = 0.031, q -value < 0.2 , Fisher's test).

Another robust signature was the Glutamine detected at position 82 (Q82) in the gp120 domain of subtype A1 (p -value = 0.00979, q -value < 0.2 , Fisher's test). The gp120 of subtype A1 historical sequences (98.55%) are more likely to carry a robust Q82 signature compared to the T/F counterparts (86.67%). In contrast, gp120 domain of T/F subtype A1 (13.33%) was more likely to select an Arginine (R82) when compared to historical sequences (1.45%) (p -value = 0.00979, q -value < 0.2 , Fisher's test). Notably, a robust Valine signature detected at position 172 (V172) in the V2 loop of subtype A1 was greatly enriched in the historical sequences (92.03%),

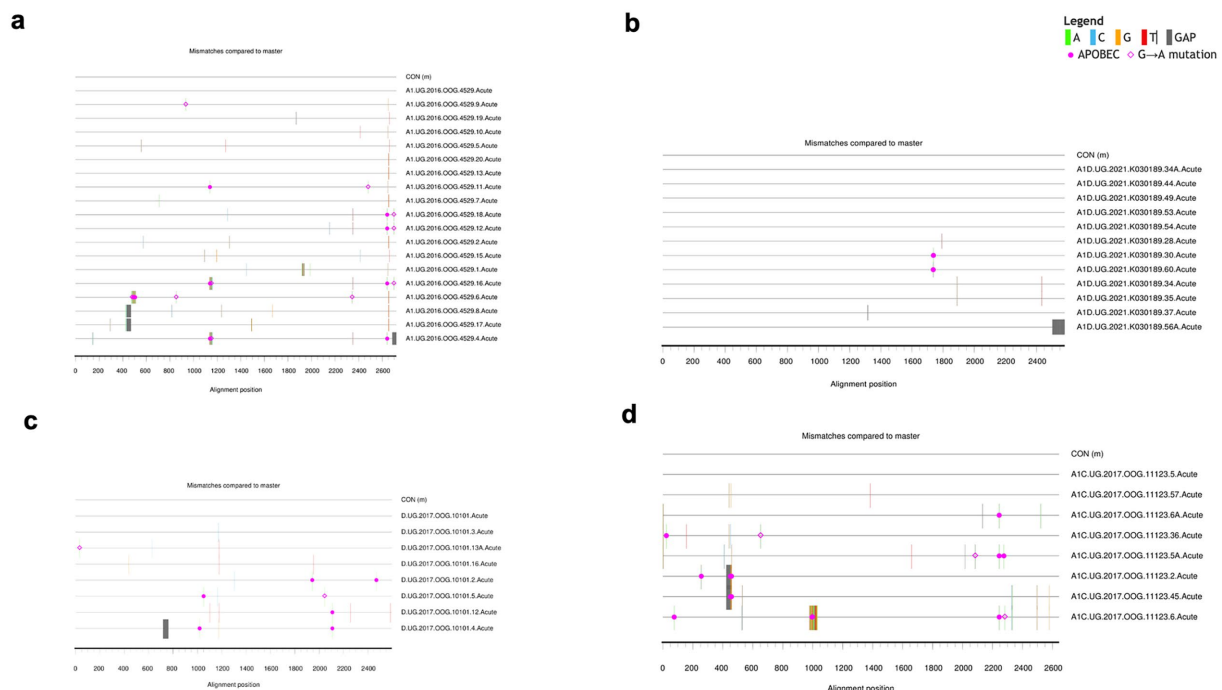


FIGURE 3
Highlighter plots. Representative subjects with single virus transmission (a–d). Tic marks represent nucleotide substitutions as compared to the top-most master (consensus) or T/F sequence in each highlighter plot.

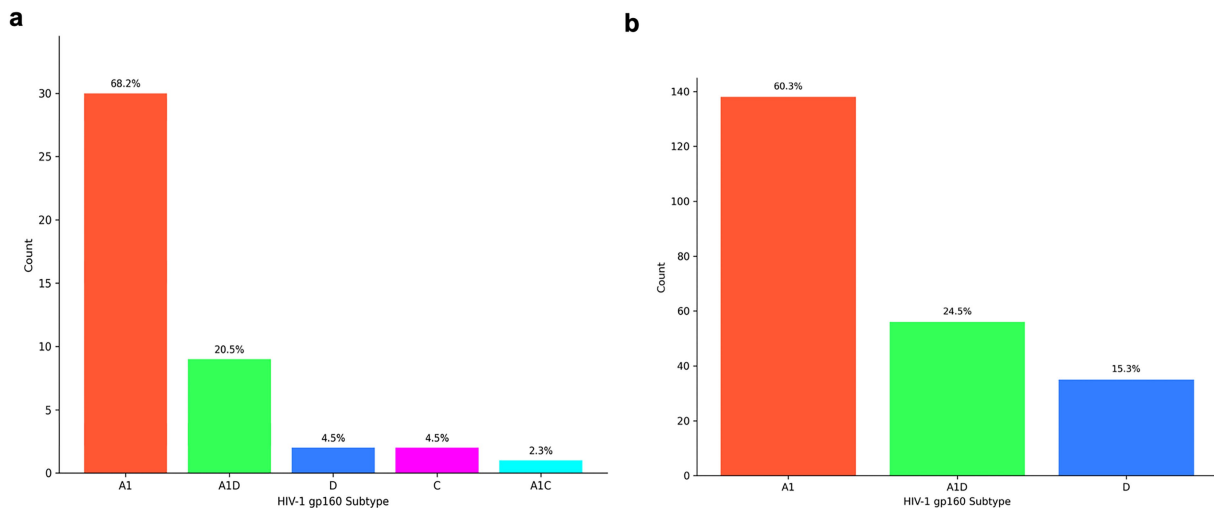


FIGURE 4
Distribution of HIV-1 envelope subtypes in East Africa according to RIP and jpHMM tools. The x-axis represents the full-length Env (gp160) subtype while the y-axis represents the number of sequences (a) HIV-1 subtype distribution of contemporary T/F Envs from 2015 to 2021. (b) HIV-1 subtype distribution of historical Envs from 1986 to 2006.

compared to the T/F counterparts (70.00%) (p -value = 0.00245, q -value < 0.2, Fisher's test). Conversely, the V1 loop of T/F subtype A1 (6.67%) exhibited a high frequency of the Alanine (A172) signature, which was absent in the historical sequences (p -value = 0.031, q -value < 0.2, Fisher's test). Also, the robust aspartate signature detected at position 230 (D230) in the glycosite 230 of the subtype A1 was less enriched in T/F (63.33%), compared to the historical sequences

(86.96%) (p -value = 0.00436, q -value < 0.2, Fisher's test). Conversely, the glycosite 230 of T/F subtype A1 (23.33%) was more likely to select a Glutamic acid (E230) signature compared to the historical sequences (9.42%) (p -value = 0.0349, q -value < 0.2, Fisher's test). What's more, a robust Glutamic Acid signature at position 275 (E275) in the D loop of subtype A1 involved its lower frequency in T/F sequences (80.00%) compared to the historical counterparts (96.38%) (p -value = 0.00486,

TABLE 1 Entropy analysis of amino acids among HIV-1 Env sequences between contemporary T/F and historical viruses.

HXB2 alignment position	HXB2 amino acid	Alignment position	Query consensus	Randomized entropy diff> = Hdiff	Highest randomized entropy diff	1st seq in query set	Background entropy (Hb)	Query Entropy (Hq)	(Hdiff = Hb-Hq)	P-value	Q-value	HIV-1 Env subtype
22	L	28	L	0	0.448	L	0.043	0.703	−0.66	0	0	A1
82	Q	92	Q	4	0.396	Q	0.076	0.393	−0.317	0	0	A1
172	E	203	V	1	0.715	V	0.36	1.068	−0.707	0	0	A1
230	N	307	D	4	0.712	E	0.464	1.036	−0.572	0	0	A1
275	V	355	E	3	0.541	Q	0.19	0.674	−0.484	0	0	A1
317	F	399	F	4	0.523	F	0.086	0.531	−0.446	0	0	A1
432	K	534	Q	4	0.706	Q	0.431	1.05	−0.62	0	0	A1
476	R	591	R	4	0.396	R	0.179	0.543	−0.364	0	0	A1
477	D	592	D	2	0.325	D	0	0.325	−0.325	0	0	A1
784	L	902	L	3	0.586	L	0.161	0.673	−0.511	0	0	A1
203	Q	277	Q	15	0.325	Q	0	0.245	−0.245	0.01	0.4	A1
282	K	362	K	7	0.346	K	0	0.291	−0.291	0.01	0.4	A1
34	L	46	L	4	1.006	L	0.268	1.149	−0.881	0	0	A1D
299	P	361	P	4	0.783	P	0	0.684	−0.684	0	0	A1D
620	T	705	D	1	1.596	D	1.857	0.637	1.221	0	0	A1D
643	L	728	Y	5	0.547	Y	0	0.53	−0.53	0	0	A1D
145	G	180	–	8	1.722	–	2.204	0.849	1.355	0.01	0.6	A1D

Query entropy represents the variation of the contemporary T/F Env while background entropy denotes that of the historical Env for a given subtype. Hdiff is the entropy difference between background and query. The HXB2 alignment positions of the statistically significant entropy sites (q-value < 0.2) are highlighted in bold.

q-value < 0.2, Fisher's test). Instead, the D loop of subtype A1 T/Fs (13.33%) exhibited a high frequency of the Lysine (K275) signature, which was absent in the historical sequences (p -value = 0.00374, q-value < 0.2, Fisher's test). Importantly, the T/F subtype A1 sequences (86.67%) were less likely to select the robust Phenylalanine (F317) signature at position 317 compared to the historical counterparts (98.55%), in the V3 loop (p -value = 0.00979, q-value < 0.2, Fisher's test). Instead, the V3 loop of T/F subtype A1 (6.67%) exhibited a higher frequency of the Tyrosine (Y317) signature, which was absent in the historical counterparts (p -value = 0.031, q-value < 0.2, Fisher's test). Furthermore, the T/F subtype A1 viruses (70.00%) are less likely to select the robust Glutamine signature at position 432 (Q432) compared to the historical counterparts (88.41%), mainly in the CD4 Contact residue (p -value = 0.00531, q-value < 0.2, Fisher's test). Similarly, the T/F subtype A1 (76.67%) exhibited a lower frequency of the robust Arginine signature at position 476 (R476) compared to the historical counterparts (95.65%), mainly in the CD4 Contact residue (p -value = 0.0214, q-value < 0.2, Fisher's test). On the contrary, the CD4 contact residue in T/F subtype A1 (23.33%) exhibited a higher frequency of the unique K476 signature compared to the historical sequences (4.35%) (p -value = 0.0214, q-value < 0.2, Fisher's test). Along the same lines, the T/F subtype A1 (90.00%) exhibited a lower frequency of the robust D477 signature compared to the historical viruses (100.00%), mainly in the CD4 Contact residue (p -value = 0.00523, q-value < 0.2, Fisher's test). Conversely, the CD4 contact residue in T/F subtype A1 (10.00%) exhibited a higher frequency of the N477 signature, which was absent in the historical counterpart (p -value = 0.00523, q-value < 0.2, Fisher's test). A robust Leucine signature (L784) detected in subtype A1 was greatly enriched in historical sequences (98.53%), compared to the T/F counterparts (86.21%), mainly in the LLP-2 lentiviral lytic peptide alpha helix (p -value = 0.00918, q-value < 0.2, Fisher's test). The genetic signature frequencies by position and absolute counts determined by the Analyze Align tool are presented in [Supplementary Table S7](#) for the contemporary T/F subtype A1 and [Supplementary Table S8](#) for historical subtype A1 Envs.

For the A1D recombinants, the C-helix in the T/F (66.67%) exhibited a high frequency of the robust D620 compared to the historical counterparts (23.21%) (p -value = 0.00171, q-value < 0.2, Fisher's test). Furthermore, the robust L34 signature detected in the gp120 of A1D recombinants occurred less in T/Fs (55.56%), compared to the historical counterparts (96.36%) (p -value = 0.00259, q-value < 0.2). The robust Proline signature detected at position 299 (P299) in A1D recombinants occurred less in the T/F sequences (77.78%) compared to the historical counterparts (100.00%), specifically in the V3 loop (p -value = 0.0173, q-value < 0.2). Additionally, the robust Y643 in the Fusion Heptad Repeat (HR) -2 was greatly enriched in A1D historical (100.00%) recombinants compared to the T/F counterparts (77.78%) (p -value = 0.0173, q-value < 0.2). The genetic signature frequencies by position and absolute counts determined by the Analyze Align tool are presented in [Supplementary Table S9](#) for T/F A1D recombinant and [Supplementary Table S10](#) for historical A1D recombinants. The weblogs show the robust genetic signature sites in contemporary T/F compared to the historical Env across subtypes A1, and A1D recombinants ([Figure 5](#)).

The small sample size of subtype D sequences limited our ability to detect robust or informative signature sites in contemporary subtype D T/F Envs compared to historical counterparts.

3.5 Exploration of the association between the unique genetic signature sites and HIV-1 Env structural characteristics

We deployed the LANL tool Genome browser to explore the relationship between the robust and unique genetic signatures in the Env and information such as HXB2 coding sites of interest, antibody epitopes, neutralizing antibody contexts such as sensitivity and resistance. Also, the signature sites were associated with other published literature such as sensitivity or resistance to bNAbs in HIV-1 subtypes, B and C ([Table 3](#)).

4 Discussion

The study aimed to detect the unique genetic signatures in the Env of the contemporary HIV-1 T/F viruses among subtypes A1, A1D recombinants and D, which are predominant in East Africa. We identified HIV-1 T/F Env sequences and subtype composition in East Africa, spanning the years 2015 to 2021 in comparison to historical sequences (1986 to 2006). The robust and unique genetic signatures in T/F Envs were analyzed in comparison to historical Envs of the same subtypes. The link between the unique genetic and the Env (gp160) structural characteristics such as sensitivity or resistance to bNAbs, and HXB2 sites of interest was also determined.

We report single variant transmission in acute infections from GHWP, Protocol B and C, and KILIGORIS cohorts. These findings are consistent with previous studies that showed that only one HIV-1 variant traverses the mucosa and establishes a productive infection in 80% of heterosexual cases ([Salazar-Gonzalez et al., 2009](#); [Keele et al., 2008](#)) further asserting the existence of a genetic bottleneck during HIV-1 transmission.

In our study of the full length Env gene, subtype A1 was the most frequently transmitted, followed by A1D recombinants and subtype D, consistent with previous reports ([Balinda et al., 2022](#); [Bbosa et al., 2019](#); [Kemal et al., 2013](#); [Billings et al., 2017](#); [Gounder et al., 2017](#)). These results suggest that subtype A1 and A1D recombinant Envs possess unique genetic signatures that enhance their transmission and persistence in East Africa. Also, A1D recombinants may benefit from a combination of advantageous genetic features from both parental subtypes.

The comparison between contemporary T/F viruses (2015–2021) and the historical viral population (1986–2006) remains informative for detecting principal genetic signatures associated with viral adaptation at the population level, given that the earliest sequence in the contemporary dataset (2015) is 10 years away from the latest sequence in the historical dataset (2006). This approach provides a meaningful temporal axis to detect unique amino acid signatures, as supported by previous studies ([Rademeyer et al., 2016](#); [Cho et al., 2019](#); [Berry et al., 2007](#); [Chaillon et al., 2012](#)).

The present study reports that while the hydrophobic core of the SP domain in subtype A1 T/Fs exhibited a unique LM22 mutation, the historical counterparts retained a conserved L22. Previous studies argue that the SP regulates Env-glycan interactions, which affect antibody binding to the V1V2 configuration, V3 apex, and gp41 epitopes ([Lambert and Upadhyay, 2021](#); [Li et al., 2000](#); [Upadhyay et al., 2018](#); [Upadhyay et al., 2020](#)). Thus, the SP is instrumental in escaping host immune responses by altering recognition by antibodies. The emerging M22 signature in

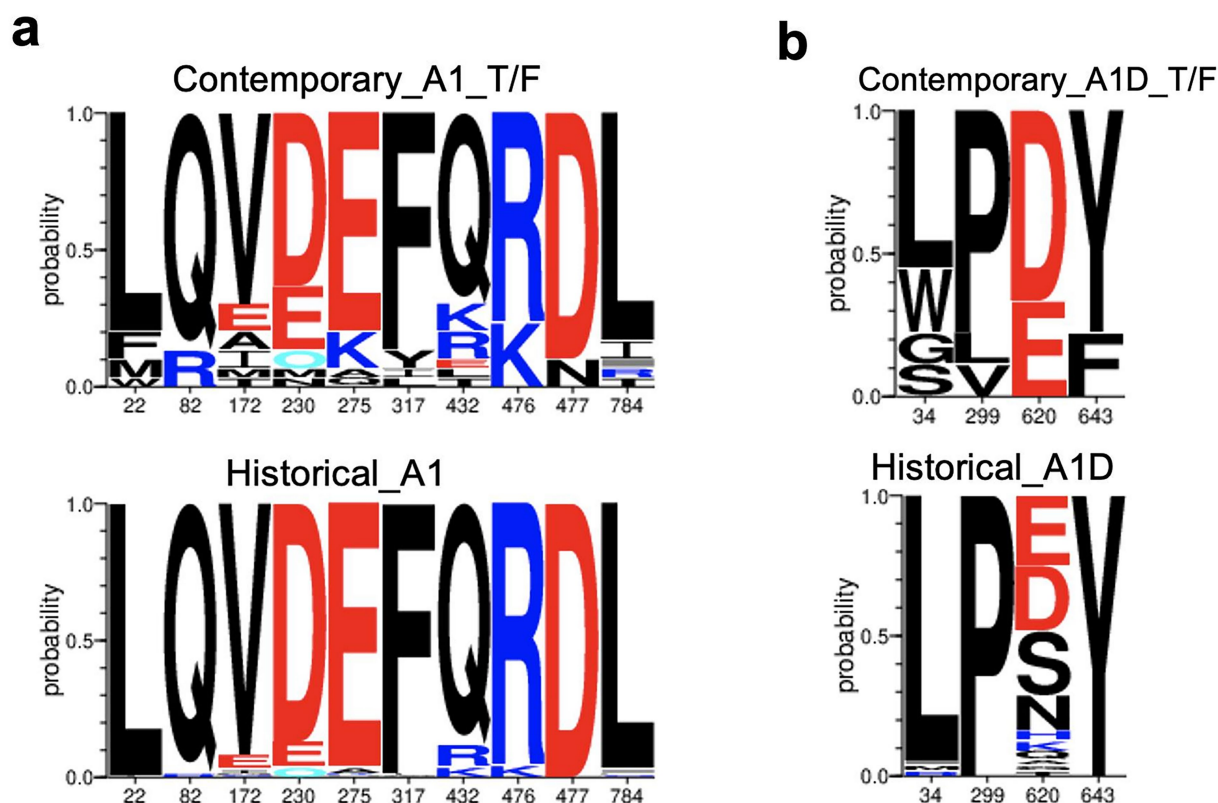


FIGURE 5

Genetic signatures identified under the HIV-1 envelope associated with contemporary T/F, compared to historical sequences. Amino acid letter probability is proportional to its relative frequency in the alignment. (a) Robust genetic signature sites, L22, M22, Q82, R82, V172, A172, D230, E230, E275, K275, F317, Y317, Q432, K476, R476, D477, N477, L784 associated with contemporary T/F subtype A1 ($n = 30$) compared to historical subtype A1 Envs ($n = 138$). (b) Robust genetic signatures, L34, P290, D620 and Y643 associated with contemporary T/F A1D recombinants ($n = 9$) compared to historical A1D recombinants ($n = 56$).

contemporary T/F subtype A1 may enhance their transmission fitness in acute infection and persistence in East Africa. The enrichment of the recurrent L22 signature in the SP of historical subtype A1 may sustain long-term infection. Moreover, the robust L22 signature has been associated with neutralisation sensitivity in subtype C viruses (Wang et al., 2011), implying that its enrichment in long-term infections could reflect a trade-off between immune evasion and viral fitness.

While the gp120 domain of T/F subtype A1 exhibited an enrichment of the Q82R mutation, the historical counterparts maintained a conserved Q82 signature site. This may facilitate viral entry into host cells during acute infection by improving receptor binding, while also contributing to viral persistence in long-term infection by influencing T cell immune responses (Santosuosso et al., 2009; Yoon et al., 2010). The robust Q82 signature site is a contact region for VRC34.01 across a geographically diverse panel of HIV strains (Chuang et al., 2019) but mediated escape from PGT15 in a subtype B strain (Dingens et al., 2019). Thus, phenotypic studies are needed to clearly define the role of the emerging Q82R mutation in subtype A1 T/F viruses.

We report a robust V172 signature, highly enriched in the V2 loop of subtype A1 historical viruses, compared to the contemporary T/Fs which exhibited a unique V172A mutation. Its presence may stabilize the Env trimer and maintain its structural integrity on the

viral surface in subtype A1 T/Fs while shielding neutralisation-sensitive domains such as the V3 loop and CD4 binding sites in long-term infection (Rao et al., 2013; Rusert et al., 2011). Despite its role in immune evasion, the V172 signature is sensitive to VRC26.08 in group M strains (Roark et al., 2021), revealing potential vulnerabilities that could be targeted for therapeutic intervention. Conversely, the unique A172 signature in subtype A1 T/F Envs and its absence in historical counterparts may alter the local structure and reaction to bnAbs. Future studies should validate the impact of the V172A mutation on the structure and function of V2 loops in subtype A1 T/F viruses.

Previous studies claim that glycosylation sites are critical for immune evasion, enhancing viral persistence in chronic infection (Ho et al., 2008; Zhang et al., 2021), while also stimulating glycan-dependent HIV-neutralizing antibodies that contribute to protective immunity (Zhang et al., 2021). The recurrent D230 signature in historical subtype A1 viruses may support glycosite 230's role in shielding the virus from immune recognition, allowing persistence. The reduced glycan content in T/F viruses may facilitate infection establishment by enhancing affinity for mucosal surfaces (Nawaz et al., 2011). The elevated E230 site in subtype A1 T/F Envs may contribute to neutralisation resistance in acute infection, thereby enhancing transmission and infectivity. While N230 confers resistance to PGT127 and sensitivity to PGT121 in subtype C Envs, and D230 is

TABLE 2 Unique genetic signatures associated with contemporary T/F compared to the historical HIV-1 Env sequences.

Contingency table	HXB2 number	HXB2 amino acid	Genetic signature region	Local aa	Test aa	<i>p</i> -value	T/F		BG		<i>q</i> -value	Odds ratio	HIV-1 Env subtype
							r1c1	r1c2	r2c1	r2c2			
t1	5	E	SP	6	I	0.000128	21	9	43	95	0.0112	5.1	A1
t2*	22	L	SP	28	M	0.031	2	28	0	138	0.166	Inf	A1
t3*	22	L	SP	28	L	0.000613	5	24	1	137	0.00739	27.6	A1
t2*	82	Q	gp120	92	R	0.00979	4	26	2	136	0.085	10.2	A1
t3*	82	Q	gp120	92	Q	0.00979	4	26	2	136	0.0455	10.2	A1
t2*	172	E	V2 loop	203	A	0.031	2	28	0	138	0.166	Inf	A1
t3*	172	E	V2 loop	203	V	0.00245	9	21	11	127	0.0184	4.88	A1
t2*	230	N	glycosite 230	307	E	0.0349	6	22	10	125	0.173	3.38	A1
t3*	230	N	glycosite 230	307	D	0.00436	9	19	13	120	0.0285	4.32	A1
t2*	275	V	D loop	355	K	0.00374	4	26	1	137	0.0537	20.5	A1
t3*	275	V	D loop	355	E	0.00486	6	24	5	133	0.0318	6.54	A1
t2*	317	F	V3 loop	399	Y	0.031	2	28	0	138	0.166	Inf	A1
t3*	317	F	V3 loop	399	F	0.00979	4	26	2	136	0.0455	10.2	A1
t3*	432	K	CD4 Contact Residue	534	Q	0.00531	8	21	10	122	0.0362	4.59	A1
t2*	476	R	CD4 Contact Residue	591	K	0.0214	5	23	6	132	0.123	4.72	A1
t3*	476	R	CD4 Contact Residue	591	R	0.0214	5	23	6	132	0.0691	4.72	A1
t2*	477	D	CD4 Contact Residue	592	N	0.00523	3	27	0	138	0.0686	Inf	A1
t3*	477	D	CD4 Contact Residue	592	D	0.00523	3	27	0	138	0.0362	Inf	A1
t3*	784	L	LLP-2	902	L	0.00918	4	25	2	134	0.0455	10.5	A1
t2*	620	E	c-helix	705	D	0.00171	3	1	1	34	0.121	66.2	A1D
t3*	34	L	gp120	46	L	0.00259	4	5	2	53	0.126	19.2	A1D
t3*	299	P	V3 loop	361	P	0.0173	2	7	0	56	0.174	Inf	A1D
t3*	643	H	Fusion HR-2	728	Y	0.0173	2	7	0	56	0.174	Inf	A1D

HXB2 number: HXB2 alignment position of the signature. **Local aa:** Local alignment amino acid position of the signature. **Test aa:** principal amino acid signatures tested or associated with T/F compared to historical Envs. **p-value:** probability calculated for the respective test. **q-value:** False discovery rate for the *p*-value. **T/F:** Transmitted/Founder. **BG:** Historical. **t:** table. **r1 = row 1 = feature = T/F, r2 = row 2 =!feature = historical.** **t1:** results for Fisher's exact tests without phylogenetic correction. **t3*:** results for Fisher's exact tests with phylogenetic correction, i.e., if the amino acid being considered is "A," **t3*** shows the events in which the ancestral sequence had A, with a significant difference. For **t3***, c1 = column 1 = A- >A, c2 = column 2 = A- > A. **t2*:** results for Fisher's tests with phylogenetic correction, i.e., if the amino acid being considered is "A," **t2*** displays the cases in which the ancestral sequence lacked "A." for **t2***, c1 = column 1 =!A- > A, c2 = column 2 =! A- >!A. An exclamation mark (!) means "not."

TABLE 3 Genetic signature relationship to bnAb context and HXB2 sites of interest.

HXB2 number	HXB2 aa	Signature region	Test aa	Feature
22	L	SP	L and M	L confers neutralisation sensitivity in clade C.
82	Q	gp120	Q and R	VRC34.01 and PGT151 contacts in subtype A and B
172	E	V2 loop	V and A	V confers neutralisation sensitivity to VRC26.08. PG16 contact with clade C.
230	N	glycosite 230	D and E	N confers neutralisation resistance to PGT127 in clade C. D confers neutralisation resistance to 10.1074. N confers neutralisation sensitivity to PGT121
275	V	D loop	E and K	E confers neutralisation sensitivity to 12A12. K confers neutralisation resistance to VRC13 in subtype C.
317	F	V3 loop	F and Y	Coreceptor binding site inside the V3 loop. Coreceptor-specific (R5/X4) site. Mutations affect neutralisation
432	K	CD4 Contact residue	Q	VRC27, VRC16 and N60P23 contacts. Amino acid Q confers resistance to 2F5. Amino acid K is sensitivity to 2F5.
476	R	CD4 Contact residue	R and K	R confers neutralisation sensitivity to CH31, VRC06b, and 8ANC131 in subtype C. K confers neutralisation resistance to CH31 VRC06b 8ANC131 R476K mutation confers loss of apoptosis induction.
477	D	CD4 Contact residue	D and N	D477 is 2411a contacts. N49P7 contact. VRC13 contacts
784	L	LLP-2	L	L784 confers sensitivity 12A12 VRC07 VRC07.523.LS VRC01 in clade C. L784I mutation confers sensitivity to 3BNC117 in clade C and B. L784I mutation Confers resistance to PGDM1400 in SHIV.CAP256SU.375S clade C clone.
620	E	c-helix	D	A confers resistance in subtype C. E associated with sensitivity to 2F5 in subtype C
34	L	gp120	L	Antibody epitope in subtype C
299	P	V3 loop	P	vFP20.01 functional escape mapping.
643	H	Fusion HR-2	Y	PGT151 contacts. Significant site of viral escape

linked to resistance to 10.1074 (Bricault et al., 2019), the functional implications of the D230E mutation remain unclear. Future studies must characterize bnAbs that target the E230 signature site in the understudied T/F subtype A1 viruses.

In the D loop, the robust E275 signature site was enriched in the historical subtype A1 viruses, whereas the T/F counterparts exhibited the unique K275 signature, which was absent in historical viruses. Mutations in the D loop have been associated with viral resistance to VRC01-like antibodies in subtype B infection (Zhang et al., 2022). Specifically, E275 was linked to sensitivity to 12A12, whereas K275 conferred resistance to VRC13 in subtype C viruses (Bricault et al., 2019). This suggests that the emerging E275K mutation in subtype A1 T/F Envs may enhance viral entry and immune evasion during acute infection, while E275 in historical Envs may contribute to long-term viral persistence.

While the V3 loop of T/F subtype A1 exhibited an enrichment of the unique V317 signature, the historical sequences retained a conserved F317 signature at this site. This contradicts findings from subtype C T/F viruses, where I307 in the V3 loop was less common in acute than chronic infections (Rademeyer et al., 2016). The V317 site in T/F subtype A1 may enhance viral infectivity, influence coreceptor usage and host cell tropism during acute infection. Similarly, the enriched F317 site in the V3 loop of historical sequences may contribute to tropism changes, and sustenance of long-term infection. Mutations in the V3 loop influence neutralisation sensitivity; for example, the F317A mutation in a subtype B strain reduced PGV04 and VRC01 neutralisation but enhanced CD4-IgG and b12

neutralisation (Falkowska et al., 2012). However, the impact of the F317V mutation on bnAbs remains unclear in subtype A1 T/F viruses.

This study highlights that the CD4 contact residues in subtype A1 are under evolutionary pressure in both contemporary T/F and historical sequences. The CD4 contact residues are integral to viral entry. The recurring Q432 signature may aid in sustaining both long-term infection and enhancement of acute subtype T/F A1 transmission. A previous study argues that the R432 residue in the CD4 binding site in subtype C infection confers resistance to 8ANC131 while Q432 and K432 were associated with sensitivity to 2F5 (Bricault et al., 2019). This suggests that the principal Q432 signature in subtype A1 T/Fs could be explored and targeted by therapeutics. Additionally, the emerging R476K mutation in the CD4 contact residue in T/F subtype A1 may enhance viral entry in acute infection while the recurrent R476 in historical counterparts may sustain long-term infection. Structural biology studies claim that R476 is a CD4 contact with a buried surface area (BSA) of 16.5 Å² on BG505 SOSIP.664 (PDB 6CM3) (Stanfield et al., 2020). Functionally, R476 was associated with sensitivity to CH31, VRC06b, and 8ANC131 while K476 confers resistance to CH31 VRC06b 8ANC131 in subtype C (Bricault et al., 2019). A study in subtypes B, C, G, CRF13_cpx, AG/A1, A/E, and A/G recombinants claimed that N425R and R476K mutations were strongly linked to a loss of apoptosis induction (Joshi et al., 2013). This implies that the enriched K476 site in T/F subtype A1 may reduce the fitness of the virus to trigger the death of uninfected CD4 + T cells in acute infection. Phenotypic studies are warranted to understand the

function of the R476K mutation in T/F subtype A1 viruses. The great enrichment of the robust D477 signature in the CD477 contact residue of historical sequences may sustain long-term infection. The D477 residue interacts with antibodies such as VRC13, 2411a and N49P7, suggesting that it could be targeted by therapeutics (Stanfield et al., 2020; Chen et al., 2021; Sajadi et al., 2018). Inversely, the emerging N477 signature in T/F subtype A1 may enhance viral entry by interacting with the CD4 receptor on host cells in acute infection. A previous study reported that D477A mutation significantly decreases b12 binding to less than 50% of wildtype in strain JRCSF (Pantophlet et al., 2003). Therefore, future studies should investigate the impact of the D477N mutation on binding monoclonal antibodies in the context of subtype A T/F viruses. Elucidating the principal amino acid patterns in the CD4bs region across the infection timeline could inform effective therapeutic design.

In the LLP-2 domain, the robust L784 signature was less associated with T/F subtype A1 compared to the historical sequences. The LLPs disrupt membrane permeability, leading to host cell death in subtype D and B infection (Costin et al., 2007). The L784 signature site might enhance the role of LLP-2 domain in perforating the host cell in acute T/F viral infection as well as sustaining long-term subtype A1 infection. Therefore, novel therapies that can target the L784 in LLP-2 could deter formation of viroporin. The recurrent L784 signature site is susceptible to several bnAbs, including 12A12, VRC07, VRC07.523, LS and VRC01 while L784I mutation is sensitive to 3BNC117 in subtype C and B viruses (Bricault et al., 2019). Contrarywise, L784I mutation is associated with resistance to PGDM1400 (Roark et al., 2021). Overall, the evidence suggests that the recurrent L784 signature is a promising target for HIV-1 vaccine development.

Among the A1D recombinants, the robust D620 signature site occurred more in the C-helix of T/F compared to the historical viruses. The D620 in the C-helix within the gp41 domain may enhance membrane fusion during acute T/F infection as well as sustain long-term infection. It is noteworthy that the A620 signature confers resistance to 4E10 while E620 is sensitive to 2F5 in subtype C (Bricault et al., 2019). While these bnAbs target the MPER domain, there is no evidence regarding the impact of D620 site on antibody recognition within the A1D recombinants. Subsequent studies should design bnAbs that target the D620 site in the neglected A1D recombinants. Furthermore, the robust L34 signature site in the gp120 domain was highly selected by A1D recombinant historical Envs compared to the T/F counterparts. The gp120 domain facilitates viral entry into host cells in the acute phase infection by binding to target cell receptors as well as mediating viral persistence by influencing the T cell immune response in the chronics (Santoso et al., 2009; Yoon et al., 2010). The recurrent L34 signature may initiate acute infection and sustain long-term infection. The lower frequency of the robust P299 signature in the V3 loop of A1D recombinant T/Fs compared to the historical counterparts may enhance coreceptor usage and host cell tropism throughout infection. A previous study showed that the P299 in the V3 loop of BG505. T332N is a site of viral escape, where P299A have a strong effect and P299H have a moderate effect on immune evasion (Dingens et al., 2019; Zhang et al., 2021). The frequency of the Y643 in the HR2 region occurred less in the A1D recombinant T/Fs compared to historical viruses. The HR2 contributes to transmission and replication by bringing the viral and host cell membranes into proximity. On that note, Y643 could

potentially enhance transmission fitness to A1D recombinant T/Fs in the acute phase as well as immune escape in long-term infection. The H643 site in the strain JR-FL interacts with the broadly neutralizing antibody PGT151, which in turn makes it an important target for HIV antibody-based therapeutics (Dingens et al., 2019). Future studies should characterize bnAbs that target the principal Y643 signature site in A1D recombinants.

The low prevalence of subtype D Env sequences in this study limited the detection of informative subtype D specific genetic signatures.

In conclusion, the presence of key genetic signature sites in the SP, gp120, V2, glycosite 230, D loop, V3 loop, CD4 contact residues, and LLP-2 in subtype A1 T/Fs, as well as in the C-helix, gp120, V3 loop, and Fusion HR-2 regions of A1D recombinant T/F Envs, may play a crucial role in the successful establishment of acute and the maintenance of long-term infection. However, phenotypic studies are needed to gain a deeper understanding of how these variations influence viral fitness and immune recognition in the less-studied subtypes A1 and A1D recombinants, which are predominant in East Africa. Also, recognizing that viral persistence also relies on other replication steps such as reverse transcription, integration and assembly, future research should extend the search for unique genetic signatures to other viral proteins mediating these processes in subtypes A1, and A1D recombinants. We acknowledge the limitation of our approach, as comparing contemporary T/F Env sequences from acutely infected individual to unrelated, cross-sectional historical sequences may not account for host specific confounders. While the envelope unique genetic signatures reported in the present study may inform therapeutic interventions, longitudinal cohorts tracking within host evolution from acute to chronic infection are needed to validate these signatures and their functional relevance.

Data availability statement

Publicly available datasets analyzed in this study can be found at the CATNAP webserver <http://hiv.lanl.gov/catnap> and SFL web alignments <https://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html#filter> at LANL HIV db. The accession numbers can be found in the [Supplementary Tables S3–S6](#). Also, original datasets analysed in this study, from acute cohorts (GHWP, protocol N and Kiligoris) were submitted to GenBank under submission ID (2982317).

Ethics statement

The studies involving humans were approved by ethics approvals obtained for the GHWP from Uganda Virus Research Institute-Research and Ethics Committee (UVRI-REC) (GC 127). Protocol N cohort was approved by the KEMRI Scientific and Ethics Review Unit (SERU) (KEMRI/RES/7/3/1). Kiligoris cohort was approved by UVRI-REC (GC/127/714). Additionally, this study was approved by the School of Biomedical Sciences-Research and Ethics Committee (SBS-REC) (Ref: SBS-2023-38) at Makerere University. The permission to use the archived plasma samples (from GHWP, Protocol N and Kiligoris cohorts) was sought from the MRC/UVRI and LSHTM Uganda Research Unit. The studies were conducted in accordance

with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

FK: Validation, Data curation, Methodology, Formal analysis, Visualization, Investigation, Writing – review & editing, Conceptualization, Software, Writing – original draft. AK: Writing – review & editing, Formal analysis, Investigation, Conceptualization, Methodology. RG: Investigation, Writing – review & editing, Formal analysis. AN: Writing – review & editing, Methodology. CN: Writing – review & editing, Methodology. FN: Methodology, Writing – review & editing. DO: Writing – review & editing, Methodology. AO: Writing – review & editing, Resources. BF: Writing – review & editing, Formal analysis, Resources. PK: Project administration, Supervision, Resources, Investigation, Writing – review & editing. EN: Project administration, Funding acquisition, Resources, Investigation, Supervision, Writing – review & editing. SB: Funding acquisition, Project administration, Formal analysis, Resources, Supervision, Writing – review & editing, Conceptualization, Methodology, Investigation.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the International AIDS Vaccine Initiative through an IAVI-Investigator-Initiated Research (IAVI-IIR) grant to SB and EN, grant number AID-0AA-A-16-000-32.

Acknowledgments

We are grateful for the supervisors Sheila Nina Balinda, Pontiano Kaleebu, Eunice Nduati, Anne Kapaata, and Ronald Galiwango, for their invaluable expertise and guidance throughout

References

- Adhiambo, M., Makwaga, O., Adungo, F., Kimani, H., Mulama, D. H., Korir, J. C., et al. (2021). Human immunodeficiency virus (HIV) type 1 genetic diversity in HIV positive individuals on antiretroviral therapy in a cross-sectional study conducted in Teso, Western Kenya. *Pan Afr. Med. J.* 38:335. doi: 10.11604/pamj.2021.38.335.26357
- Balinda, S. N., Kapaata, A., Xu, R., Salazar, M. G., Mezzell, A. T., Qin, Q., et al. (2022). Characterization of near full-length transmitted/founder HIV-1 subtype D and A/D recombinant genomes in a heterosexual Ugandan population (2006–2011). *Viruses* 14:334. doi: 10.3390/v14020334
- Bbosa, N., Kaleebu, P., and Ssemwanga, D. (2019). HIV subtype diversity worldwide. *Curr. Opin. HIV AIDS* 14:153. doi: 10.1097/COH.0000000000000534
- Berry, I. M., Ribeiro, R., Kothari, M., Athreya, G., Daniels, M., Lee, H. Y., et al. (2007). Unequal evolutionary rates in the assessment of HIV polymorphisms and HLA allele associations. *Science* 315, 1583–1586. doi: 10.1126/science.1131528
- Billings, E., Sanders-Buell, E., Bose, M., Kijak, G. H., Bradfield, A., and Crossler, J. (2017). HIV-1 Genetic Diversity Among Incident Infections in Mbeya, Tanzania. *AIDS Res. Hum. Retroviruses* 33, 373–381. doi: 10.1089/AID.2016.0111
- Bricault, C. A., Yusim, K., Seaman, M. S., Yoon, H., Theiler, J., Giorgi, E. E., et al. (2019). HIV-1 neutralizing antibody signatures and application to epitope-targeted vaccine design. *Cell Host Microbe* 25, 59–72.e8. doi: 10.1016/j.chom.2018.12.001
- Chaillon, A., Braibant, M., Hué, S., Bencharif, S., Enard, D., Moreau, A., et al. (2012). Human immunodeficiency virus type-1 (HIV-1) continues to evolve in presence of broadly neutralizing antibodies more than ten years after infection. *PLoS One* 7:e44163. doi: 10.1371/journal.pone.0044163
- Checkley, M. A., Luttge, B. G., and Freed, E. O. (2011). HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *J. Mol. Biol.* 410, 582–608. doi: 10.1016/j.jmb.2011.04.042
- Chen, X., Zhou, T., Schmidt, S. D., Duan, H., Cheng, C., Chuang, G. Y., et al. (2021). Vaccination induces maturation in a mouse model of diverse Unmutated VRC01-class precursors to HIV-neutralizing antibodies with >50% breadth. *Immunity* 54, 324–339.e8. doi: 10.1016/j.immuni.2020.12.014
- Cho, Y. K., Kim, J. E., and Foley, B. T. (2019). Genetic analysis of the full-length gag gene from the earliest Korean subclade B of HIV-1: an outbreak among Korean hemophiliacs. *Viruses* 11:545. doi: 10.3390/v11060545
- Chuang, G. Y., Zhou, J., Acharya, P., Rawi, R., Shen, C. H., Sheng, Z., et al. (2019). Structural survey of broadly neutralizing antibodies targeting the HIV-1 Env trimer

this project, from its inception to execution. We acknowledge all the participants in IAVI'S Protocol N, GHWP and Kilgoris cohorts whose archived plasma samples were used in this study. We also extend our appreciation to Kshitij Wagh and Brian Foley for training us in genetic signature analysis using advanced online data tools at the LANL HIV sequence db. The authors gratefully acknowledge the staff at the MRC/UVRI and LSHTM Uganda Research Unit, UVRI, Makerere University, College of Health Sciences, and Kenya Medical Research Institute, who made this study feasible.

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.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

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.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2025.1632581/full#supplementary-material>

delineates epitope categories and characteristics of recognition. *Structure* 27:e6, 196–206. doi: 10.1016/j.str.2018.10.007

Costin, J. M., Rausch, J. M., Garry, R. F., and Wimley, W. C. (2007). Viroporin potential of the lentivirus lytic peptide (LLP) domains of the HIV-1 gp41 protein. *Virology* 4:123. doi: 10.1186/1743-422X-4-123

Désiré, 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

Dings, A. S., Arenz, D., Weight, H., Overbaugh, J., and Bloom, J. D. (2019). An antigenic atlas of HIV-1 escape from broadly neutralizing antibodies distinguishes functional and structural epitopes. *Immunity* 50, 520–532.e3. doi: 10.1016/j.immuni.2018.12.017

Falkowska, E., Ramos, A., Feng, Y., Zhou, T., Moquin, S., Walker, L. M., et al. (2012). PGV04, an HIV-1 gp120 CD4 binding site antibody, is broad and potent in neutralization but does not induce conformational changes characteristic of CD4. *J. Virol.* 86, 4394–4403. doi: 10.1128/JVI.06973-11

Gnanakaran, S., Bhattacharya, T., Daniels, M., Keele, B. F., Hraber, P. T., Lapedes, A. S., et al. (2011). Recurrent signature patterns in HIV-1 B clade envelope glycoproteins associated with either early or chronic infections. *PLoS Pathog.* 7:e1002209. doi: 10.1371/journal.ppat.1002209

Gounder, K., Oyar, M., Padayachi, N., Zulu, T. M., de Oliveira, T., Wylie, J., et al. (2017). Complex subtype diversity of HIV-1 among drug users in major Kenyan cities. *AIDS Res. Hum. Retrovir.* 33, 500–510. doi: 10.1089/aid.2016.0321

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

Herbeck, J. T., Nickle, D. C., Learn, G. H., Gottlieb, G. S., Curlin, M. E., Heath, L., et al. (2006). Human immunodeficiency virus type 1 env evolves toward ancestral states upon transmission to a new host. *J. Virol.* 80, 1637–1644. doi: 10.1128/JVI.80.4.1637-1644.2006

Ho, Y. S., Abecasis, A. B., Theys, K., Deforche, K., Dwyer, D. E., Charleston, M., et al. (2008). HIV-1 gp120 N-linked glycosylation differs between plasma and leukocyte compartments. *Virology* 475, 514. doi: 10.1016/j.virol.2008.05.014

Joshi, A., Lee, R. T. C., Mohl, J., Sedano, M., Khong, W. X., Ng, O. T., et al. (2013). Genetic signatures of HIV-1 envelope-mediated bystander apoptosis. *J. Biol. Chem.* 289:2497. doi: 10.1074/jbc.M113.514018

Kafando, A., Martineau, C., El-Far, M., Fournier, E., Doualla-Bell, F., Serhir, B., et al. (2019). HIV-1 envelope glycoprotein amino acids signatures associated with clade B transmitted/founder and recent viruses. *Viruses* 11:1012. doi: 10.3390/v11111012

Kaleebu, P., French, N., Mahe, C., Yirrell, D., Wattera, C., Lyagoba, F., et al. (2002). Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1—positive persons in Uganda. *J. Infect. Dis.* 185, 1244–1250. doi: 10.1086/340130

Kapaata, A., Balinda, S. N., Xu, R., Salazar, M. G., Herard, K., Brooks, K., et al. (2021). HIV-1 gag-pol sequences from Ugandan early infections reveal sequence variants associated with elevated replication capacity. *Viruses* 13:171. doi: 10.3390/v13020171

Katoh, K., Misawa, K., Kuma, K. I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

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. USA* 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. Retroviruses* 29, 957–962. doi: 10.1089/AID.2012.0095

Kiwanuka, N., Laeyendecker, O., Quinn, T. C. J., Wawer, M., Shepherd, J., Robb, M., et al. (2009). HIV-1 subtypes and differences in heterosexual HIV transmission among HIV-discordant couples in Rakai, Uganda. *AIDS* 23, 2479–2484. doi: 10.1097/QAD.0b013e328330cc08

Lambert, G. S., and Upadhyay, C. (2021). HIV-1 envelope glycosylation and the signal peptide. *Vaccine* 9:176. doi: 10.3390/vaccines9020176

Li, Y., Luo, L., Thomas, D. Y., and Kang, C. Y. (2000). The HIV-1 Env protein signal sequence retards its cleavage and down-regulates the glycoprotein folding. *Virology* 272, 417–428. doi: 10.1006/viro.2000.0357

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

Mayanja, Y., Abaasa, A., Namale, G., Price, M. A., and Kamali, A. (2020). Willingness of female sex workers in Kampala, Uganda to participate in future HIV vaccine trials: a case control study. *BMC Public Health* 20:1789. doi: 10.1186/s12889-020-09932-7

Nawaz, F., Cicala, C., Van Ryk, D., Block, K. E., Jelicic, K., McNally, J. P., et al. (2011). The genotype of early-transmitting HIV gp120s promotes α 4 β 7-reactivity, revealing α 4 β 7 +CD4+ T cells as key targets in mucosal transmission. *PLoS Pathog.* 7:e1001301. doi: 10.1371/journal.ppat.1001301

Ndung'u, T., McCune, J. M., and Deeks, S. G. (2019). Why and where an HIV cure is needed and how it might be achieved. *Nature* 576, 397–405. doi: 10.1038/s41586-019-1841-8

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A Fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Pantophlet, R., Saphire, E. O., Poignard, P., Parren, P. W. H. I., Wilson, I. A., and Burton, D. R. (2003). Fine mapping of the interaction of neutralizing and nonneutralizing monoclonal antibodies with the CD4 binding site of human immunodeficiency virus type 1 gp120. *J. Virol.* 77:642. doi: 10.1128/jvi.77.1.642-658.2003

Price, M. A., Kilembe, W., Ruzagira, E., Karita, E., Inambao, M., Sanders, E. J., et al. (2020). Cohort profile: IAVTS HIV epidemiology and early infection cohort studies in Africa to support vaccine discovery. *Int. J. Epidemiol.* 50, 29–30. doi: 10.1093/ije/dyaa100

Rademeyer, C., Korber, B., Seaman, M. S., Giorgi, E. E., Thebus, R., Robles, A., et al. (2016). Features of recently transmitted HIV-1 clade C viruses that impact antibody recognition: implications for active and passive immunization. *PLoS Pathog.* 12:e1005742. doi: 10.1371/journal.ppat.1005742

Rambaut, A. FigTree. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. (2010). Available online at: <http://tree.bio.ed.ac.uk/software/figtree/> (accessed Nov 25, 2022)

Rao, M., Peachman, K. K., Kim, J., Gao, G., Alving, C. R., Michael, N. L., et al. (2013). HIV-1 variable loop 2 and its importance in HIV-1 infection and vaccine development. *Curr. HIV Res.* 11, 427–438. doi: 10.2174/1570162X113116660064

Reis, M. N. G., Guimarães, M. L., Bello, G., and Stefani, M. M. A. (2019). Identification of new HIV-1 circulating recombinant forms CRF81_cpx and CRF99_BF1 in Central Western Brazil and of unique BF1 recombinant forms. *Front. Microbiol.* 10:10. doi: 10.3389/fmicb.2019.00097

Roark, R. S., Li, H., Williams, W. B., Mason, R. D., Gorman, J., et al. (2021). Recapitulation of HIV-1 Env-antibody coevolution in macaques leading to neutralization breadth. *Science* 371:eabd2638. doi: 10.1126/science.abd2638

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

Rusert, P., Krarup, A., Magnus, C., Brandenberg, O. F., Weber, J., Ehler, A. K., et al. (2011). Interaction of the gp120 V1V2 loop with a neighboring gp120 unit shields the HIV envelope trimer against cross-neutralizing antibodies. *J. Exp. Med.* 208, 1419–1433. doi: 10.1084/jem.20110196

Sajadi, M. M., Dashti, A., Tehrani, Z. R., Tolbert, W. D., Seaman, M. S., Ouyang, X., et al. (2018). Identification of near-pan-neutralizing antibodies against HIV-1 by deconvolution of plasma humoral responses. *Cell* 173, 1783–1795.e14. doi: 10.1016/j.cell.2018.03.061

Salazar-Gonzalez, J. F., Bailes, E., Pham, K. T., Salazar, M. G., Guffey, M. B., Keele, B. F., et al. (2008). Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J. Virol.* 82, 3952–3970. doi: 10.1128/JVI.02660-07

Salazar-Gonzalez, J. F., Salazar, M. G., Keele, B. F., Learn, G. H., Giorgi, E. E., Li, H., et al. (2009). Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J. Exp. Med.* 206, 1273–1289. doi: 10.1084/jem.20090378

Santoso, M., Righi, E., Lindstrom, V., Leblanc, P. R., and Poznansky, M. C. (2009). HIV-1 envelope protein gp120 is present at high concentrations in secondary lymphoid organs of individuals with chronic HIV-1 infection. *J. Infect. Dis.* 200, 1050–1053. doi: 10.1086/605695

Schultz, A. K., Zhang, M., Bulla, I., Leitner, T., Korber, B., Morgenstern, B., et al. (2009). JpHMM: improving the reliability of recombination prediction in HIV-1. *Nucleic Acids Res.* 37, W647–W651. doi: 10.1093/nar/gkp371

Siepel, A. C., Halpern, A. L., Macken, C., and Korber, B. T. (1995). A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Res. Hum. Retrovir.* 11, 1413–1416. doi: 10.1089/aid.1995.11.1413

Skinner, M. E., and Holmes, I. H. (2010). Setting up the JBrowse genome browser. *Curr. Protoc. Bioinformatics* 32, 9–13. doi: 10.1002/0471250953.bi0913s32

Skinner, M. E., Uzilov, A. V., Stein, L. D., Mungall, C. J., and Holmes, I. H. (2009). JBrowse: A next-generation genome browser. *Genome Res.* 19, 1630–1638. doi: 10.1101/gr.094607.109

Ssemwanga, D., Nsubuga, R. N., Mayanja, B. N., Lyagoba, F., Magambo, B., Yirrell, D., et al. (2013). Effect of HIV-1 subtypes on disease progression in rural Uganda: a prospective clinical cohort study. *PLoS One* 8:e71768. doi: 10.1371/journal.pone.0071768

Stanfield, R. L., Berndsen, Z. T., Huang, R., Sok, D., Warner, G., Torres, J. L., et al. (2020). Structural basis of broad HIV neutralization by a vaccine-induced cow antibody. *Sci. Adv.* 6:eaba0468. doi: 10.1126/sciadv.aba0468

Storey, J. D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* 100, 9440–9445. doi: 10.1073/pnas.1530509100

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., and Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44, W232–W235. doi: 10.1093/nar/gkw256

Umviligho, G., Muok, E., Nyirimiho Gisa, E., Xu, R., Dilernia, D., Herard, K., et al. (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

- Upadhyay, C., Feyznehzad, R., Cao, L., Chan, K. W., Liu, K., Yang, W., et al. (2020). Signal peptide of HIV-1 envelope modulates glycosylation impacting exposure of V1V2 and other epitopes. *PLoS Pathog.* 16:e1009185. doi: 10.1371/journal.ppat.1009185
- Upadhyay, C., Feyznehzad, R., Yang, W., Zhang, H., Zolla-Pazner, S., and Hioe, C. E. (2018). Alterations of HIV-1 envelope phenotype and antibody-mediated neutralization by signal peptide mutations. *PLoS Pathog.* 14:e1006812. doi: 10.1371/journal.ppat.1006812
- Van Regenmortel, M. H. V. (2017). Development of a preventive HIV vaccine requires solving inverse problems which is unattainable by rational vaccine design. *Front. Immunol.* 8:2009. doi: 10.3389/fimmu.2017.02009
- Vandepitte, J., Weiss, H. A., Bukenya, J., Nakubulwa, S., Mayanja, Y., Matovu, G., et al. (2013). Alcohol use, *mycoplasma genitalium*, and other STIs associated with HIV incidence among women at high risk in Kampala, Uganda. *J. Acquir. Immune Defic. Syndr.* 62:119. doi: 10.1097/QAI.0b013e3182777167
- Wang, S., Nie, J., and Wang, Y. (2011). Comparisons of the genetic and neutralization properties of HIV-1 subtype C and CRF07_08_BC env molecular clones isolated from infections in China. *Virus Res.* 155, 137–146. doi: 10.1016/j.virusres.2010.09.012
- Wibmer, C. K., Moore, P. L., and Morris, L. (2015). HIV broadly neutralizing antibody targets. *Curr. Opin. HIV AIDS* 10, 135–143. doi: 10.1097/COH.0000000000000153
- Yamaguchi, J., Vallari, A., McArthur, C., Sthresley, 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.00000000000002246
- Yang, Z. y., Chakrabarti, B. K., Xu, L., Welcher, B., Kong, W. p., Leung, K., et al. (2004). Selective modification of variable loops alters tropism and enhances immunogenicity of human immunodeficiency virus type 1 envelope. *J. Virol.* 78, 4029–4036. doi: 10.1128/jvi.78.8.4029-4036.2004
- Yang, C., Li, M., Shi, Y. P., Winter, J., van Eijk, A. M., Ayisi, J., et al. (2004). Genetic diversity and high proportion of intersubtype recombinants among HIV type 1-infected pregnant women in Kisumu, Western Kenya. *AIDS Res. Hum. Retrovir.* 20, 565–574. doi: 10.1089/088922204323087822
- Yoon, V., Fridkis-Hareli, M., Munisamy, S., Lee, J., Anastasiades, D., and Stevceva, L. (2010). The GP120 molecule of HIV-1 and its interaction with T cells. *Curr. Med. Chem.* 17, 741–749. doi: 10.2174/092986710790514499
- Yoon, H., Macke, J., West, A. P., Foley, B., Bjorkman, P. J., Korber, B., et al. (2015). CATNAP: a tool to compile, analyze and tally neutralizing antibody panels. *Nucleic Acids Res.* 43, W213–W219. doi: 10.1093/nar/gkv404
- Zhang, D., Liu, Z., Wang, W., Chen, M. X., Hou, J. L., Zhang, Z., et al. (2022). Viral resistance to VRC01-like antibodies with mutations in loop D and V5 from an HIV-1 B' subtype infected individual with broadly neutralization activity. *Mol. Immunol.* 145, 50–58. doi: 10.1016/j.molimm.2022.02.021
- Zhang, Y., Zheng, S., Zhao, W., Mao, Y., Cao, W., Zeng, W., et al. (2021). Sequential analysis of the N/O-glycosylation of heavily glycosylated HIV-1 gp120 using EThcD-sceHCD-MS/MS. *Front. Immunol.* 12:755568. doi: 10.3389/fimmu.2021.755568
- Zhou, Y., Maldini, C. R., Jadowsky, J., and Riley, J. L. (2021). Challenges and opportunities of using adoptive T-cell therapy as part of an HIV cure strategy. *J. Infect. Dis.* 223, S38–S45. doi: 10.1093/infdis/jiaa223