Observational Study

Medicine

OPEN

Net charge and position 22 of the V3 loop are associated with HIV-1 tropism in recently infected female sex workers in Nairobi, Kenya

Hellen K Abisi, MS[a], Leon E Otieno, BS[b], Erastus Irungu, BS[c], Frank G Onyambu, PhD[d], Angeline Chepchirchir, PhD[e], Omu Anzala, MD, PhD[f,g], Dalton C Wamalwa, MD[i], Ruth W Nduati, MD[i], Lyle McKinnon, PhD[g,j], Joshua Kimani, MD[c], Martin M Mulinge, PhD[a,f,*]

Abstract

Human immunodeficiency virus (HIV) infection affects around 37 million people worldwide, and in Kenya, key populations especially female sex workers (FSW), are thought to play a substantial role in the wider, mostly heterosexual HIV-1 transmission structure. Notably, HIV tropism has been found to correlate with HIV-1 transmission and disease progression in HIV-infected patients. In this study, recently infected FSWs from Nairobi, Kenya, were assessed for HIV tropism and the factors related to it. We used a cross-sectional study design to analyze 76 HIV-1 positive plasma samples obtained from FSWs enrolled in sex worker outreach program clinics in Nairobi between November 2020 and April 2021. The effects of clinical, demographic, and viral genetic characteristics were determined using multivariable logistic regression. HIV-1 subtype A1 accounted for 89.5% of all cases, with a prevalence of CXCR4-tropic viruses of 26.3%. WebPSSMR5X4 and Geno2Pheno [G2P:10–15% false positive rate] showed high concordance of 88%. Subjects infected with CXCR4-tropic viruses had statistically significant lower baseline CD4$^+$T-cell counts than those infected with CCR5-tropic viruses (P = .044). Using multivariable logistic regression and adjusting for potential confounders, we found that net charge, the amino acid at position 22 of the V3 loop, and the geographic location of the subject were associated with tropism. A unit increase in V3 loop's net-charge increased the odds of a virus being CXCR4-tropic by 2.4 times (OR = 2.40, 95%CI = 1.35–5.00, P = .007). Second, amino acid threonine at position 22 of V3 loop increased the odds of a strain being X4 by 55.7 times compared to the alanine which occurred in CCR5-tropic strains (OR = 55.7, 95%CI = 4.04–84.1, P < .003). The Kawangware sex worker outreach program clinic was associated with CXCR4-tropic strains (P = .034), but there was there was no evidence of a distinct CXCR4-tropic transmission cluster. In conclusion, this study revealed a high concordance of WebPSSMR5X4 and Geno2Pheno in predicting HIV tropism. The most striking finding was that amino acid position 22 of the V3 loop is linked to tropism in HIV-1 subtype A1. Additional studies with a large dataset are warranted to confirm our findings.

Abbreviations: CCR5 = C-C motif chemokine receptor 5, CXCR4 = CXCR4 C-X-C motif chemokine receptor 4, env = envelope gene, FPR = false positive rate, G2P = geno2pheno, FSW = female sex worker; HIV-1 = human immunodeficiency virus type 1, PSSM = position specific scoring matrix, SWOP = sex work outreach program, V3 = third variable region of gp120.

Keywords: CCR5, CXCR4, female sex workers, Geno2Pheno, HIV-1 subtype A1, Kenya, WebPSSM

1. Introduction

Human immunodeficiency virus 1 (HIV-1) is a retrovirus that infects mainly CD4$^+$T lymphocytes, macrophages, and dendritic cells.[1] This tropism is determined at the level of viral entry by the use of CD4 as a primary receptor and the use of coreceptors that are strain and target specific. R5 strains of HIV use CC-chemokine receptor 5 (CCR5) as their coreceptor and can, therefore, enter macrophages, dendritic cells, and T cells, whereas X4 strains of HIV use CXCR4 as a coreceptor and can infect T cells only.[2] Genetic polymorphism in the CCR5 gene has been correlated with HIV resistance and this discovery led to development of

The research work was supported by Partnership for Health Research Training in Kenya (P-HERT) grant project number 5D43TW010141.

The authors declare no competing interests.

All data generated or analyzed during this study are included in this published article [and its supplementary information files.

Supplemental Digital Content is available for this article.

[a] Department of Biochemistry, University of Nairobi, Nairobi, Kenya, [b] Molecular Medicine and Infectious Diseases Laboratory, University of Nairobi, Nairobi, Kenya, [c] Partners for Health and Development in Africa (PHDA), Nairobi, Kenya, [d] School of Health Sciences, Meru University of Science and Technology, Meru, Kenya, [e] Department of Nursing Sciences, University of Nairobi, Nairobi, Kenya, [f] Kenya AIDS Vaccine Initiative - Institute of Clinical Research, University of Nairobi, Nairobi, Kenya, [g] Department of Medical Microbiology, University of Nairobi, Nairobi, Kenya, [i] Department of Paediatrics and Child Health, University of Nairobi, Nairobi, Kenya, [j] Department of Medical Microbiology and Infectious Diseases, Max Rady College of Medicine, University of Manitoba, Manitoba, MB, Canada.

** Correspondence: Martin M Mulinge, University of Nairobi, Chiromo Campus - Nairobi 30197-00100, Kenya (e-mail: mmulinge@uonbi.ac.ke).*

How to cite this article: Abisi HK, Otieno LE, Irungu E, Onyambu FG, Chepchirchir A, Anzala O, Wamalwa DC, Nduati RW, McKinnon L, Kimani J, Mulinge MM. Net charge and position 22 of the V3 loop are associated with HIV-1 tropism in recently infected female sex workers in Nairobi, Kenya. Medicine 2022;101:49(e32024).

Received: 19 August 2022 / Received in final form: 4 November 2022 / Accepted: 4 November 2022

http://dx.doi.org/10.1097/MD.0000000000032024

CCR5 inhibitors such as maraviroc.[3,4] Early in infection, R5 viruses are the predominant strains in HIV-infected individuals whereas X4 viruses come to predominate late with profound loss of CD4 + T cells and rapid progression to AIDS disease.[5] This is partly attributed to the effect of Interferon-induced transmembrane protein-2 (IFITM2), which preferentially suppresses replication of X4 HIV-1 strains but not R5 HIV-1 strains in the early stage of HIV-1 infection.[6] A tropism assay is conducted by sequencing the third variable region (V3) of envelope gene gp120 and prediction of tropism using algorithms.[7] The widely used algorithms include geno2pheno[coreceptor],[8] PhenoSeq,[9] WebPSSM[10], and 11/25 rule[11] which deliver high levels of sensitivity among a wide range of HIV-1 subtypes.

As HIV-1 continues to spread worldwide, its genetic diversification increases due to emergence of point mutations, and recombination between HIV-1 strains, as a result of coinfection or superinfection.[12] The consequence of this diversity has an influence on cellular tropism, kinetics of viral replication, and ultimately transmission and disease progression.[13] HIV-1 diversity is advantageous to the virus as it confers an important survival advantage against negative selection pressures such as host immune response and antiretroviral agents.[14] For instance, subtype C, which accounts for ~ 48% of all HIV-1 infections has a reduced replicative efficiency relative to subtype B in the order of 10–100-fold and is related to properties residing in gp120, in particular to a reduced avidity for binding the CD4/CCR5 receptors on host cells.[15]

HIV-1 transmission is often characterized by the presence of numerous transmission clusters that are likely to play a key role in sustaining the epidemic and thus should be prioritized for HIV prevention.[16,17] In Kenya, HIV-1 key populations including men having sex with men, people who inject drugs and female sex workers (FSWs) are thought to significantly contribute to HIV-1 transmission in the wider, mostly heterosexual HIV-1 transmission network.[18] Of concern is the profound effect on FSWs where prevalence estimates are 13.5-fold higher than non sex-worker women, including a 9.8-fold increased risk for transmission in Kenya.[19] This highlights the burden of HIV in this key population, which has demonstrated extreme vulnerability to HIV infection. It remains unclear to what extent virally encoded factors and host immune status influences HIV transmission in such a key population. It is therefore critical that we monitor viral evolution, in order to understand the pandemic's trajectory that would ultimately improve Kenya public health response. Thus, in the current study, we evaluated the performances of 8 algorithms on predicting viral tropism from recently infected FSWs recruited from 7 sex worker outreach program (SWOP) clinics in Nairobi. In addition, we identified the determinants of tropism using multivariable logistic regression.

2. Methods

2.1. Research ethics

Ethical approval to collect HIV-positive plasma for the HIV prevention and care program was approved by the Kenyatta National Hospital—University of Nairobi Ethics and Research Committee (KNH-UON-ERC: P258/09/2008). The study's participants gave an informed consent for their samples to be stored and used in future related HIV research. Given that the current study focused on additional virological aspects of HIV infection on the same biological samples, a waiver of consent was sought as it was not deemed to adversely affect the rights and welfare of the subjects approval number P556/07/2019.

2.2. Study design and setting

This was a cross-sectional study that utilized 157 plasma samples from HIV-1 infected and treatment-naïve FSWs who lived in Nairobi county, between November 2020 and April 2021. The subjects were recruited from 7 clinics in Nairobi county namely: City, Donholm, Kawangware, Korogocho, Langata, Majengo and Thika road. The CD4$^+$T cells profiling was carried out at Partners for Health and Development in Africa (PHDA) laboratories while HIV tropism assays were conducted at the Molecular Medicine and Infectious Diseases Laboratory all located in University of Nairobi.

2.3. Study participants

The study population consisted of HIV-infected FSWs enrolled on either of the 7 clinics. Subjects who met the following criteria were included into the study: 18 years old; confirmed HIV-1 seropositive; treatment naïve; residing in Nairobi county during the enrollment period (November 2020–April 2021); and involved in sex work defined as an adult who consents to sexual exchange with the primary purpose of monetary benefit. Exclusion criteria included treatment-experienced subjects; subjects under the age of 18 during the enrollment period; and subjects not involved in sex-work as defined above.

2.4. CD4$^+$T cells profiling

CD4$^+$T cell count and CD4$^+$T cell percentage (%) were determined using BD FACSPrestoTM as per manufacturer instructions. Briefly, blood samples were added to the inlet ports on the BD FACSPrestoTM cartridge using pipets provided. The cartridges were incubated for 18 minutes at room temperature on the workstation and a timer was selected for the corresponding incubation slot on the workstation. After incubation, the tear strips were removed and the sample IDs corresponding to the BD FACSPrestoTM cartridges were placed in the machine and read.

2.5. HIV RNA extraction

HIV-1 RNA extraction was carried out using the PureLink™ Viral RNA/DNA Mini Kit (Thermo Fisher Scientific, San Francisco, CA) in accordance with the manufacturer's instructions. Briefly 500 µL of plasma was lysed using 500 µL viral Lysis buffer and 62.5 µL Proteinase K. The lysate was transferred to sterile spin column, washed twice with 500 µL of wash buffer, centrifuged at 6800 g for 1 minute and then 17000 g for 1 minute, and finally eluted in 40 µL of sterile, RNAase-free water. RNA was stored at −80°C.

2.6. The env gene amplification and sequencing

4 µL of extracted RNA was incubated at 65°C for 10 minutes and mixed with 21 µL OneTaq® One-Step RT-PCR Mix (New England Biolabs, Massachusetts). The RT-PCR Master Mix comprised of the following components: 12.5 µL One Taq One step reaction mix; 2 µL One Taq one step enzyme; 2.5 µL RNAse free water; 2 µL Forward Primer M5 5′-CCAATTCCCA TACATTATTGTGCCCCAGCTGG-3′; and 2 µL M10 5′-CCAATTGTCCCTCATATCTCCTCC TCCAGG-3′ corresponding to the C2-V3 region of the *env* gene (6975-7520).[20] The cycling conditions were as follows: Reverse transcription; 48°C for 30 min; PCR initial denaturation at 94°C for 1 minute and 40 cycles (94°C for 15s, 55°C for 30s, 68°C for 1 min) and final 5 minutes extension at 68°C.

For nested PCR, 4 µL of PCR product was amplified in a 20 µL reaction with GoTaq® G2 Hot Start Green Master Mix (Promega, Wisconsin). The master mix comprised of the following components: 10 µL Go Taq green master mix; 5.2 µL RNAse free water; 0.4 µL Forward primer M3 5′-GTCAGCACAGTACAATGCACACATGG-3′ and 0.4 µL Reverse primer M8

5′-TCCTTGGATGGGAGGGGCATACATTGC-3′ reverse).[21] The cycling conditions were as follows: initial denaturation 95°C for 10 minutes, 35 cycles (95°C for 30 s, 55°C for 30 s, 72°C for 1 min), and a final 10 minutes step at 72°C. Amplified PCR product (666-bp) was verified by 1% agarose gel electrophoresis stained with SYBRTM safe DNA gel stain (Thermo Fisher Scientific, San Francisco, CA) and visualized under UV light.

The amplified PCR fragment was purified using EXOSAP-ITTM (Applied Biosystems, Massachusetts) as per manufacturer instructions. The *env* gene sequencing was performed using Big-Dye Terminator v3.1 kit (Thermo Fisher Scientific, San Francisco, CA). 3 µL of purified PCR product was added to 8.5 µL of sequencing mix comprised of the following: 1.5 µl Big Dye Terminator; 2 µL 5× Buffer; 3 µL RNAse free water and 1 µL Sequencing primer (SP1) 5′-AGYRCAGTA-CAATGYACACATGG-3′ or forward sequencing primer 2 (SP2) 5′-TCAACHCAAYTRCTGTTAAATGG-3′.[22] Cycle sequencing conditions were as follows: 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes. Sanger sequencing was performed using BigDye XTerminator kit (Thermo Fisher Scientific, San Francisco, CA) on Applied Biosystems 3730 DX genetic analyzer (Applied Biosystems, Foster City). Sequencing files were obtained and aligned against the reference HIV-1 env genome using the Molecular Evolutionary Genetics Analysis software (Pennsylvania, USA) for the env gene. The *env* sequences were archived in the Data Bank of Japan Database with the following accession numbers (LC722376- LC722451).

2.7. HIV-1 subtyping

HIV-1 subtyping was carried out by submitting raw *env* sequences to 2 HIV-1 subtyping tools; COMET[23] and REGA[24] which predicted the subtype and associated bootstrap support level. The higher the bootstrap support level, the more precise the prediction.

2.8. HIV-1 tropism determination

The raw sequences were trimmed following multiple sequence alignment with MEGA11 against the reference V3 loop sequence to obtain sample V3 loop sequences. The V3 nucleotide sequences were translated to the corresponding amino acid sequences using the translate tool https://www.bioinformatics.org/sms2/translate.html Aligned, and edited sequences covering the V3 region were analyzed for coreceptor usage using various in silico genotypic tools: Geno2Pheno [5% false positive rate (FPR), 10% FPR, 15%FPR and 20%FPR]; $WebPSSM_{[sinsi]}$, $WebPSSM_{[R5X4]}$, $WebPSSM_{[sinsi\ C]}$; 11/25 rule and HIVCoR.

2.9. Phylogenetic tree construction

For the phylogenetic relationships, the V3-loop nucleotide sequences were aligned with MEGA11.[25] Following the alignment, the optimal tree model was constructed using Maximum Likelihood method with 1000 bootstrap and visualized by using FigTree v1.4.4. http://tree.bio.ed.ac.uk/software/figtree/. Clustering was assessed using Cluster Picker 1.2 with the following default settings: Initial threshold, 0.9; Main support threshold, 0.9; Genetic threshold distance threshold, 4.5; and large distance cluster threshold, 5.

2.10. Statistical analyses

Tropism coded (0 = CCR5; 1 = CXCR4) was the primary outcome in this study. The predictor variables were: HIV subtype, V3 loop characteristics, $CD4^+$ T cell count and %, age, and geographical location. Continuous variables were compared using Wilcoxon rank sum test or Kruskal–Wallis test whereas Fisher's exact test was used to compare categorical variables between the 2 groups, with 2-sided *P*-values reported in both cases. Differences between proportions were tested by the χ^2 test. The consistency of the different algorithms in predicting coreceptor usage was calculated using the Kuder-Richardson Formula 20 (KR-20). Univariable and multivariable logistic regression models were used to estimate the predictors of tropism, with estimated odds ratios (ORs) and 95% confidence intervals reported. Variables with a P.25 at univariate analysis were postulated to influence the outcome and were included in the full model. A 2-sided P value < 0.05 was considered to be statistically significant. Statistical analyses were conducted with the R statistical package (R version 4.1.0).

3. Results

3.1. Patient demographics, viral load, HIV-1 subtype, and ART regimens

The demographic, virological, and clinical characteristics of the study participants are summarized in Table 1. Subjects infected with CCR5 viruses had 410 copies/mm^3 (IQR, 229-630), whereas those infected with CXCR4 viruses had 338 copies/mm^3 (IQR, 141-444). Among subjects infected with CCR5-tropic viruses, 20% (n = 11) had $CD4^+$ T cell counts < 200 copies/mm^3, 25% (n = 14) had $CD4^+$ T cell counts between 200 and 349 copies/mm^3, 18% (n = 10) had $CD4^+$ T cell counts between 350 and 500 copies/mm^3, and 38% (n = 21) had $CD4^+$ T cell counts > 500 copies/mm^3. In subjects infected with CXCR4-tropic viruses, 30% (n = 6), 25% (n = 5), 20% (n = 4), and 25% (n = 5) had < 200 copies/mm^3, 200–349 copies/mm^3, 350–500 copies/mm^3, and > 500 copies/mm^3, respectively. Viruses that were CCR5-tropic had a median $CD4^+$T cell percentage of 23 (IQR, 14–31) while CXCR4-tropic viruses had a median $CD4^+$T cell percentage of 19 (IQR, 10–26). CXCR4-tropic viruses predominated at Kawangware SWOP clinic, accounting for 53% (n = 10) of all viruses whereas CCR5-tropic viruses were predominant at Thika road clinic 76% (n = 16). Three subtypes of HIV-1 were reported, with subtype A1 accounting for 89.5% of cases (n = 68), subtype D for 6.6% of cases (n = 5), and subtype C for 3.9% of cases (n = 3). A classification of viral strains based on tropism shown that 75% (n = 51) were CCR5 tropic, while 25% (n = 17) were CXCR4 tropic.

3.2. Concordance in prediction of coreceptor usage by different algorithms

In our study, there were differences between the bioinformatics tools, with an overall concordance of 74% (95%CI 68%–79%). The Geno2Pheno algorithm had better prediction agreements: at FPRs of 5%, 10%, 15%, and 20% CXCR4-tropic strains were predicted in 11%, 24%, 26%, and 28% of the samples, respectively. The agreement between all 4 Geno2Pheno cutoffs was 93% (95%CI 90%–95%). By removing the G2P 5% FPR and calculating the concordance of G2P 10%–20%, the agreement increased to 98% (95%CI 97%–99%). When compared to Geno2pheno, WebPSSM matrixes had a poor concordance of 52% (95%CI 29%–68%). $WebPSSM_{[SINSI\_C]}$ overestimated CXCR4-tropic strains by 84%, compared to $WebPSSM_{[R5X4]}$ and $WebPSSM_{[SINSI]}$, which predicted 21% and 12% of viruses to be CXCR4-tropic, respectively. The 11/25 rule predicted 6.6% of the viral strains as X4 viruses and HIVCoR assigned all the viral strain CCR5 tropism. WebPSSM [$WebPSSM_{R5X4}$, $WebPSSM_{SINSI}$] and Geno2Pheno [G2P:10-20% FPR] were highly concordant at 88% (95%CI 79%–90%).

3.3. Association between $CD4^+$ T cell count, HIV subtype, and HIV-1 tropism

To assess the extent to which the virus's replication without treatment has harmed the $CD4^+$ T cell population, we compared

Table 1

Demographic and clinical characteristics of 76 study subjects included in the analysis.

Characteristic	CCR5, N = 56	CXCR4, N = 20
Age (years), median (IQR)	34 (29–40)	37 (30–39)
CD4 %, median (IQR)	23 (14–31)	19 (10–26)
CD4 count (copies/mm^3), median (IQR)	410 (249–630)	338 (141–444)
CD4 count strata (copies/mm^3), n (%)		
<200	11 (20)	6 (30)
200–349	14 (25)	5 (25)
350–500	10 (18)	4 (20)
>500	21 (38)	5 (25)
HIV-1 subtype, n (%)		
A1	51 (91)	17 (48.9)
C	3 (4.9)	0 (10.6)
D	2(8.6)	3 (21.3)
SWOP clinic n (%)		
Thika Road	16 (29)	5 (25)
City	12 (21)	1 (5.0)
Kawangware	9 (16)	10 (50)
Donholm	4 (7.1)	0 (0)
Korogocho	9 (16)	2 (10)
Langata	2 (3.6)	1 (5.0)
Majengo	4 (7.1)	2 (5)

Results are median (IQR) or frequency (%), as appropriate.

CCR5 = C-C motif chemokine receptor 5, CXCR4 = CXCR4 C-X-C motif chemokine receptor 4, SWOP = sex worker outreach program.

CD4$^+$ T cell counts in subjects infected with CCR5-tropic and CXCR4-tropic viruses. There were 2 outliers on the CXCR4 upper tail, Grubbs test (P = .020) which were excluded. Subjects infected with CXCR4-tropic viruses had statistically significant lower baseline CD4$^+$T-cell counts 289 ± 189 copies/mm^3 than those infected with CCR5-tropic viruses 433 ± 260 copies/mm^3, (P = .044, Fig. 1A). In patients harboring X4 strains, the baseline CD4$^+$T cell % was lower 17 ± 10% compared to 22 ± 11% in patients infected with R5 strains. The difference was statistically significant (P = .042, Fig 1B). The baseline CD4$^+$T-cell counts among different HIV-1 subtypes was not statistically significant (P = .98, Fig. 1C). Stratification of HIV subtype by tropism revealed 91% (n = 51), 5.4% (n = 3), and 3.6% (n = 2), respectively in subtype A1, C and D for subjects infected with CCR5-tropic viruses. Similarly, in CXCR4 infected subjects, 85% (n = 17), 0% (n = 0), and 15% (n = 3), respectively, in subtype A1, C, and D. The HIV-1 subtype differences between R5 and X4 strains was not statistically significant (P = .20).

3.4. V3 loop sequence diversity

In order to establish whether specific characteristics have an impact on predicted tropism, we profiled the genetic characteristics of the translated V3 loop sequences. Detailed description of clinical characteristics and associated V3 features are provided in the online supplementary material (supplementary Table 1, Supplemental Digital Content, http://links.lww.com/MD/I8). Twenty percent of viral strains (n = 15) had a V3-loop with 34 amino acids (AA), 79% (n = 60) had a V3-loop with 35 AA, and 1.3 % (n = 1) had a V3-loop with 36 AA. The proportion of sequences that were CXCR4 tropic for the 34AA V3 loop was 33% (n = 5), compared to 23% (n = 14) for the 35AA V3 loop and 100% (n = 1) for the 36AA V3 loop. The highly conserved GPGQ motif at the crown was noticeable in both the X4 and R5 viral strains (Fig. 2). The most notable result of this study was the finding that alanine (A) was more frequent at position 22 in R5 strains than X4 strains(Fig. 2). V3 loops with alanine at position 22 accounted for 80% (n = 45) of R5-viruses and were associated with tropism (χ^2 = 15.586, df = 8, P = .048). There was no association between tropism and the amino acid at position 11 (χ^2 = 6.478, df = 2, P = .166) as well as amino acid at position 25 (χ^2 = 19.967, df = 10, P = .075). However, statistically significant association was observed with the predicted 11/25 tropism and position 11 rule usage (χ^2 = 44.96, df = 4, P < .001) and position 25 (χ^2 = 54.848, df = 10, P < .001). We observed a correlation between the increased positive charge of V3 loop and CXCR4 tropism, with the CXCR4-using viruses having higher positive charge (+6) compared to CCR5-using viruses (+4). The amino acid in position 22 was not associated with the net charge (χ^2 = 61.741, df = 56, P = .278). This is significant because it did not lead to multicollinearity in the final multivariable logistic regression model.

3.5. Phylogenetic tree analyses

The ML phylogenetic tree of the 76 recently infected subjects in this study revealed 3 statistically supported nodes (bootstrap > 90%). In the phylogenetic tree the sequences were shown to form 2 divergent clusters with an uneven distribution CXCR4-tropic viruses (highlighted in red) in the lower arm (Fig. 3). There was a tendency of subjects recruited from Kawangware SWOP clinic to be CXCR4-tropic though there is no clear evidence of an exclusive CXCR4-tropic cluster nor a SWOP clinic cluster.

3.6. Logistic regression analysis

On the basis of adjusted statistical significance, our multivariable logistic regression models identified net charge, amino acid at position 22 and subject's geographical location as the factors associated with tropism (Table 2). After adjusting for all variables included in the model, the presence of amino acid threonine(T) at position 22 increased the odds of a strain being CXCR4-tropic by 55.7 times compared to the reference (alanine), (OR = 55.7, 95%CI = 4.04–84.1, P = .003). The second most important predictor of tropism was net-charge. The odds of a virus being CXCR4-tropic increased by 2.4 times for every unit increase in net charge (OR = 2.40, 95%CI = 1.35–5.00, P = .007). Tropism was influenced by a subject's geographic location. Thus, given that all other independent variables are equal, a FSW who resides in Kawangware has nearly 7 times the odds of having the CXCR4 tropic virus as a FSW who does not (OR = 6.94; 95%CI 1.22–48.81; P = .034). However, due to a low sample size we encountered a wider confidence interval. In

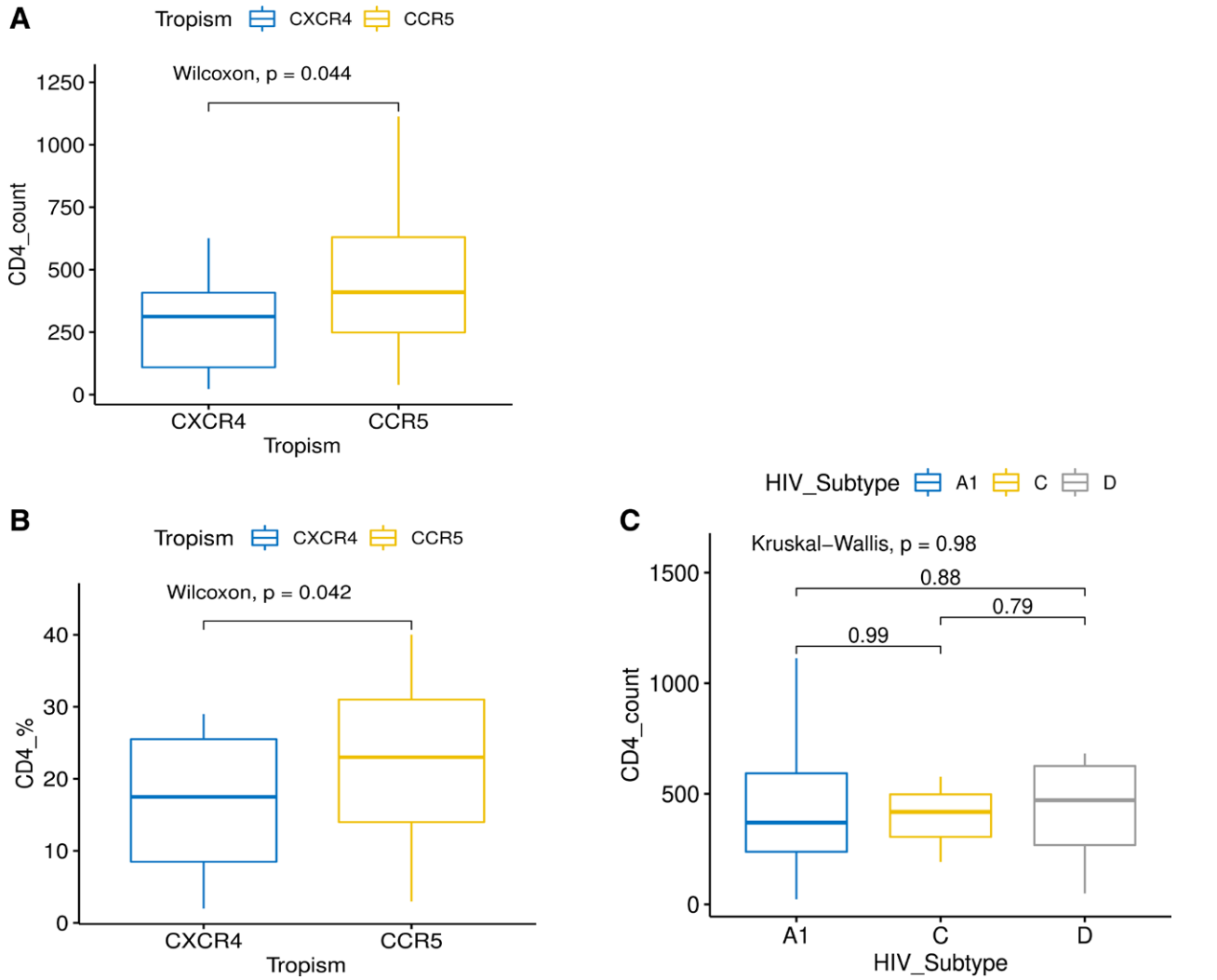

Figure 1. Comparisons of median baseline CD4⁺T cell counts and CD4⁺T cell percentages between CCR5 and CXCR4 tropisms. Subjects infected with CXCR4-tropic viruses had lower CD4⁺T-cell counts than those infected with CCR5-tropic viruses, Wilcoxon rank sum test (P = .044) (A). In graph (B) Subjects infected with CXCR4-tropic viruses had lower CD4⁺T-cell percentages than those infected with CCR5-tropic viruses, Wilcoxon rank sum test (P = .042). The levels of CD4⁺T cell count in subjects infected with HIV-1 A1, C, and subtype D was compared using Kruskal–Wallis test but there were no statistically significant differences among the different HIV-1 subtypes, (P = .98) (C). CCR5 = C-C motif chemokine receptor 5, CXCR4 = CXCR4 C-X-C motif chemokine receptor 4.

the univariable model, the subjects' SWOP clinic was marginally not associated with tropism (P = .065), emphasizing the importance of including the SWOP clinic in the multivariable model. Comparably, the CD4 T cell count was statistically significant in the CD4 T cell count only model (P = .039), but not in the multivariable model (P = .065). With caution, we can conclude that net-charge, CD4⁺T cell count, amino acid at position 22 and SWOP clinic explain 61.1% of the variance in predicted tropism, Nagelkerke (0.6117). The remaining 38.9 % of the variability is random, associated with factors not included in the study, or is associated with variables included in this study but not in the forms entered in the model. To determine how well our model would perform with the same or different data sets, we validated our model using k-fold cross-validation approach. This model's accuracy was 90%, and Cohen's kappa was 73%, which would be considered "substantial" based on acceptable thresholds. The sensitivity, specificity, and prevalence rates were 95.4%, 75%, and 26.67%, respectively.

4. Discussion

In Kenya, the HIV prevalence among FSWs is 29.3% compared to 6.6% among women in the general population.[26] This extra risk is attributed to nature of sex work which involves likelihood of engagement in unprotected sex and having multiple partners thus making FSWs one of the most-at-risk population for HIV acquisition and transmission.[27] The involvement of multiple partners is responsible for genetic diversification of viral strains due to recombination between HIV-1 strains from multiple partners, as a result of coinfection or superinfection.[12] This underscores the need to focus our attention on such a key population in order to understand the pandemic's evolution trajectory and thus inform public health mitigation strategies. As a result, study focused on the third hypervariable region (V3) of the gp120 protein which is recognized as the major determinant for tropism.[7]

Several algorithms have been developed for genotypic prediction of tropism based on the V3 loop, and there is no consensus about which is the best. The estimated prevalence of CXCR4 tropism among 5 independent algorithms was 26% using the Kuder-Richardson 20 formula.[28] This formula calculates internal consistency reliability, or how similar coreceptor usage prediction scores (CCR5 or CXCR4) are to one another, and a cutoff of 70% is considered acceptable. Additionally, we assessed how well different tropism prediction algorithms performed when predicting tropism of V3 loop sequences derived

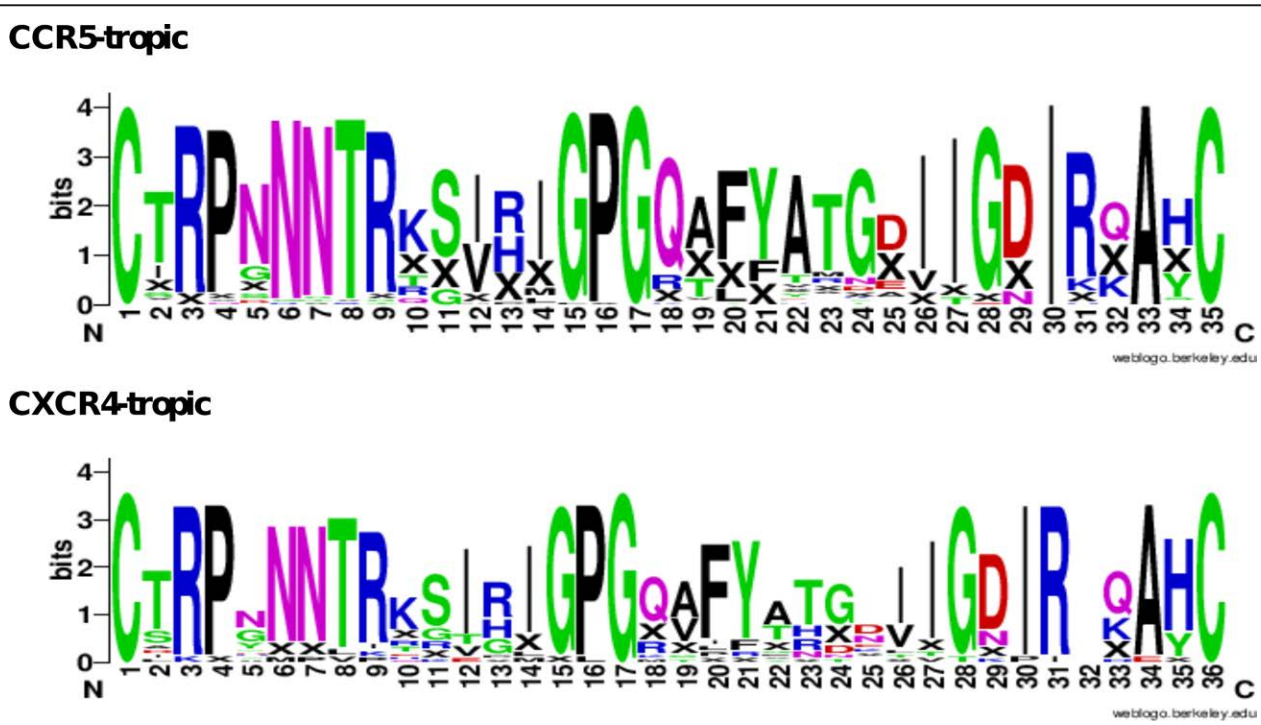

Figure 2. Sequence variation in the V3 region of CCR5-tropic viruses (top), and CXCR4-tropic viruses (bottom) generated by WebLogo 3 software. The X axis indicates the amino acid position, whereas the symbol height (Y axis) indicates the relative frequency of each amino acid at that position. The amino acid alanine (A) dominates the CCR5-tropic graph and one CXCR4-tropic virus was 36 amino acid long. Different colors indicate the physiochemical characteristics of the amino acid (black: nonpolar, green: polar, red: aromatic, blue: positively charged, purple: negatively charged). CCR5 = C-C motif chemokine receptor 5, CXCR4 = CXCR4 C-X-C motif chemokine receptor 4.

from a cohort that was predominantly subtype A1 with a minority of C and D strains. The first algorithm used was the 11/25 rule, which determines tropism based on the presence of a charge at amino acid positions 11 and/or 25 of the 35 amino acids in the V3-loop.[29] The 11/25 rule predicted 6.6% as CXCR4 against a score of > 20% of the highly concordant Geno2Pheno[8] and WebPSSM algoritms.[10] This is not surprising given that the 11/25 algorithm was trained on subtype B viruses with 35 amino acids, while the V3 loops in our dataset ranged in length from 34 to 36 amino acids.[30] The HIVCoR[31] webserver developed for CRF01_AE exclusively predicted all viral strains as CCR5 emphasizing the need for subtype specific algorithms or algorithms that have high sensitivity across a wide range of subtypes. Geno2Pheno[coreceptor FPR:10-20%],[8] and $WebPSSM_{X4R5}$[10] were highly concordant 88% (95%CI 79%–90%), in line with previous studies which demonstrated the best concordance (>85%) between these 2 algorithms.[32–34] There was poor agreement among the WebPSSM matrixes: X4R5, SINSI, and SINSI C at 52% due to overestimation of CXCR4 coreceptor usage by $WebPSSM_{SINSI\_C}$ which was trained on subtype C which were under-represented in this study.[35]

The global distribution and genetic complexity of HIV-1 subtypes and circulating recombinant forms (CRFs) have continued to evolve, exhibiting altered transmissibility, cellular tropism, kinetics of viral replication, and disease progression.[36] This study's findings that subtype A1 accounted for 89.5% of infections and subtypes D and C for 6.6% and 3.9%, respectively, are consistent with earlier research done in Kenya.[37–42] Despite earlier research indicating that subtype D of HIV-1 is linked to CXCR4 tropism, this study found no conclusive evidence of such an association.[36] We cautiously hypothesize that this is due to the dominance of subtype A1, which obscured any differences from the minority subtypes, thereby reducing the model's ability to detect a difference. A distinctive feature of HIV-1 infection is a profound loss of $CD4^+$ T cells, the main home of HIV, and a rise in plasma HIV RNA, which eventually leads to opportunistic infections, development of AIDS and AIDS-related deaths if left untreated.[43] As a result, the CD4 + T cell count is an immunological marker for HIV disease diagnosis and was previously used to determine the priority for initiating antiretroviral therapy.[44] This study found a statistically significant lower levels of baseline CD4 + T cell count among subjects infected with CXCR4-tropic viruses, strongly suggesting that CXCR4-tropic viruses are more pathogenic and might play a significant role in early $CD4^+$ T cells depletion and possible fast disease progression.[45,46] This establishes that when HIV infection is left untreated, as was the case with the study's participants, there is a profound depletion of $CD4^+$ T cells that is comparable to the chronic phase of HIV disease.[5] This is of importance to governments, since early detection of HIV infection, timely monitoring of disease progression, and early linkage to antiretroviral therapy are critical steps to curbing the spread of HIV.

Due to the global intensification of ART with the goal of achieving the UNAIDS 95-95-95 target, there has been a trend toward fewer new HIV infections. However, the declining HIV-1 incidence rates have not been consistent across all regions and risk groups, with FSWs having a 13.5-fold higher risk compared to nonsex-worker women,[19] highlighting the role that various HIV-1 transmission pathways and risky behavior play in this process. To address this vulnerable population, SWOP clinics were established in 2008 to scale-up accessible, acceptable, and friendly HIV prevention and treatment services. This program has 7 clinics throughout Nairobi County, all in close proximity to areas with higher concentrations of hot spots in which sex workers hangout to find clients.[47] In this study, there was statistically significant convergence of CXCR4-tropic strains at the Kawangware SWOP clinic. Given that CXCR4 strains are linked to aggressive HIV disease, attention should be paid to this site in an attempt to understand HIV epidemiology in this SWOP clinic.

To scientifically understand how various factors influenced tropism, we used a multivariable logistic regression model that

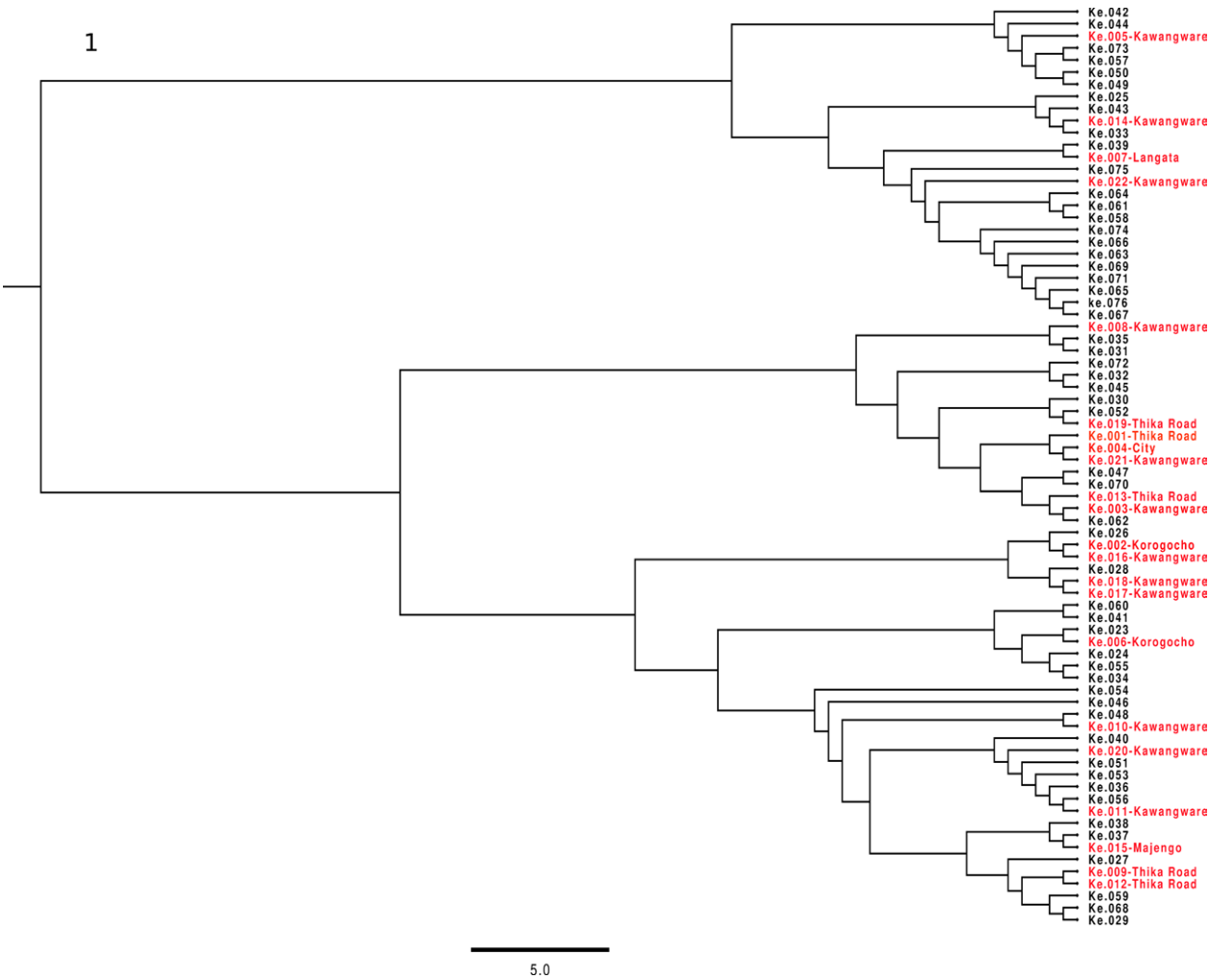

Figure 3. Phylogenetic relationship of the 76 V3 loop sequences from treatment naïve female sex workers from 7-SWOP clinics in Kenya. There is no evidence of excusive CXCR4-tropic clusters. The CXCR4-tropic viruses are highlighted in red including the source SWOP clinic. CXCR4 = CXCR4 C-X-C motif chemokine receptor 4; SWOP = sex worker outreach program.

predicted tropism (dependent variable) as a function of multiple independent variables. Part of the scientific goal was a basic understanding of the contributions of each risk factor, so as to aid public-health efforts and help prioritize and help future research in this key population. Our model was validated using a k-fold cross validation approach, and it was found to be 90% accurate. We found a strong association between net-charge and coreceptor usage which is consistent with previous studies.[48,49] A notable finding was the association between tropism and the amino acid at position 22 in the HIV-1 subtype A1 V3 loop ($P = .003$). Amino acid alanine predominated in CCR5-tropic viruses while amino acid threonine was predominant in CXCR4 strains. Even though the widely-studied subtype B does not frequently have this 22-AA genotype, when it does, the virus's infectivity is increased.[50,51] In laboratory growth competition experiments with primary HIV-1 isolates infecting peripheral blood mononuclear cells, the order of relative fitness is subtypes A, B, D, CRF01_AE > subtype C > HIV-2 > group O.[13] The reason why there would be differences in tropism following an amino acid switch from 2 uncharged amino acids (alanine to threonine) is unexplained. In attempt to identify the role of this switch, Marozsan et al, used the Garnier–Osguthorpe–Robson method to investigate the role of this switch. They did not find any differences linked to the switch, pointing to a predominance of alanine in R5-tropic strains as opposed to X4-tropic strains in subtype B strains.[52] This warrants further investigation to ascertain whether the increased fitness HIV-1 subtype A is associated with this amino acid switch. Contrary to expectations, neither a significant association with tropism nor a difference in amino acid composition between CCR5-tropic and CXCR4-tropic viruses were found at positions 11 and 25.[29,53] The confounding effects of net charge, amino acid at position 22 and the Kawangware SWOP clinic rendered $CD4^+$ T cell count effect insignificant in our multivariable model despite it being a significant factor in univariable model as reported in other.[45,46,54,55]

HIV-1 transmission is frequently characterized by the existence of numerous transmission clusters, thought to be critical for sustaining the epidemic.[16,17] Given that FSWs are one of the population groups most at risk for contracting HIV, this is even more crucial. The Kawangware SWOP clinic was statistically associated with CXCR4-tropic viruses, despite the fact that there was not a cluster in this study that was predominantly CXCR4-specific. It is not unexpected that we did not find a cluster of CXCR4-tropic strains in one SWOP clinic at one time, which could be attributed to the high variability of the V3 loop or a situation in which a subject visits a clinic away from their work location due to stigma associated with sex work.[56] Contrary to what was found in the multivariable analysis, the SWOP clinic was not associated with CXCR4 usage in the univariable model. This is a classic example of Simpson's paradox whereby results from analyses in overall data contradict the findings from subgroups of the same data.[57]

Table 2

Factors associated with coreceptor usage in 76 HIV-1 treatment naïve subjects

	Univariable			Multivariable		
Characteristic	**OR**	**95% CI**	***P* value**	**OR**	**95% CI**	***P* value**
Net charge	1.75	1.20, 2.70	**.006**	2.40	1.35, 5.00	**.007**
$CD4^+$ T cell count	0.88	0.76, 0.97	**.039**	1.00	0.93, 1.81	.093
Amino acid at position 22						
A	—	—		—	—	
Other (E,G,P,R,V)	2.41	0.45, 10.9	.27	1.21	0.17, 7.37	.84
T	12.7	3.32, 57.0	**<.001**	55.7	4.04, 84.1	**.003**
SWOP clinic						
Thika Road	—	—		—	—	
Kawangware	3.56	0.96, 14.7	.065	6.94	1.22, 48.81	**.034**
City	0.27	0.01, 1.95	.25	2.33	1.04, 14.6	.89
Donholm	0.00	0.0, ∞	>.99	0.00	0, ∞	>.99
Korogocho	0.71	0.09, 4.10	.72	0.22	0.01, 2.48	.26
Langata	1.60	0.07, 20.6	.27	4.89	0.16, 116	.30
Majengo	0.80	0.04, 7.21	.86	4.00	0.14, 71.0	.34
HIV-1 subtype						
A1	—	—				
C	0.00	0.00, ∞	>.99			
D	4.50	0.69, 36.4	.12			

OR = unadjusted odds ratio, CI = confidence interval.

P values from χ^2 test for categorical variables or Wilcoxon rank sum test for continuous variables. A = alanine, E = glutamic acid, G = glycine, P = proline, R = arginine, SWOP = sex work outreach program, T = threonine, V = valine.

Bold value indicates $P < .05$.

Our study has some limitations. First, there was no data on viral load, and in recently infected subjects, viral load, alongside clinical features, may be a key player in coreceptor usage. Second, due to a poor amplification rate, the study's sample size was small. On the surface, 76 samples may appear small in comparison and the study underpowered, but this was not the case. The dependent variable in logistic regression is the occurrence of an event, and the rule of the thump refers to the number of subjects to whom the event occurs, rather than the total number of subjects. In our case, we predicted the occurrence of CXCR4-tropic viruses by assuming that approximately 10% of the subjects had CXCR4, and we used a model with 5 independent variables. As a result, the minimum sample size would need to be 10–15 events per independent variable, which adds up to 50–75. Despite the justification, we cannot completely rule out the possibility of low study power in circumstances where an uneven distribution of cases in a specific variable led to contradictory findings.

In conclusion, WebPSSM and Geno2Pheno were highly concordant at predicting tropism. In this cohort of FSWs, we discovered that CXCR4 coreceptor is used by subtype A1 viral strains more frequently than previously believed in HIV-1 infected individuals. The most striking finding was the significant association between tropism and the amino acid at position 22 of the V3 loop. It is necessary to conduct more research using large datasets to confirm our findings.

Acknowledgments

The authors are deeply grateful to the study subjects who participated in this study and Festus Muriuki from PHDA, for provision of de-identified subject metadata.

Author contributions

Conceptualization: Martin M Mulinge and Omu Anzala.
Funding acquisition: Martin M. Mulinge, Dalton C. Wamalwa, and Ruth W. Nduati.
Investigation: Hellen K Abisi, Leon E Otieno and Erastus Irungu.
Methodology: Frank Onyambu and Martin Mulinge.
Resources: Joshua Kimani, Lyle McKinnon and Omu Anzala.
Project administration and supervision: Angeline Chepchirchir and Martin Mulinge
Data analysis and software: Martin M Mulinge.
Writing original draft: Martin Mulinge, Hellen K Abisi, and Leon E Otieno.
Review and editing: Hellen K Abisi, Martin M Mulinge, and Lyle McKinnon.

References

[1] Hladik F, Sakchalathorn P, Ballweber L, et al. Initial events in establishing vaginal entry and infection by human immunodeficiency virus type-1. Immunity. 2007;26:257–70.
[2] Doms RW, Trono D. The plasma membrane as a combat zone in the HIV battlefield. Genes Dev. 2000;14:2677–88.
[3] Samson M, Libert F, Doranz BJ, et al. Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature. 1996;382:722–6.
[4] Liu R, Paxton WA, Choe S, et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell. 1996;86:367–77.
[5] Douek DC, Picker LJ, Koup RA. T cell dynamics in HIV-1 infection. Annu Rev Immunol. 2003;21:265–304.
[6] Wu W-L, Grotefend CR, Tsai M-T, et al. Δ20 IFITM2 differentially restricts X4 and R5 HIV-1. Proc Natl Acad Sci. 2017;114:7112–7.
[7] Hwang SS, Boyle TJ, Lyerly HK, et al. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. Science. 1991;253:71–4.
[8] Beerenwinkel N, Däumer M, Oette M, et al. Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes. Nucleic Acids Res. 2003;31:3850–5.
[9] Cashin K, Gray LR, Harvey KL, et al. Reliable genotypic tropism tests for the major HIV-1 subtypes. Sci Rep. 2015;5:21–3.
[10] Jensen MA, Li F-S, van 't Wout AB, et al. Improved coreceptor usage prediction and genotypicmonitoring of R5-to-X4 transition by motif analysis of humanimmunodeficiency virus type 1 env V3 LoopSequences. J Virol. 2003;77:13376–88.
[11] Shioda T, Levy JA, Cheng-Mayer C. Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell-line and macrophage tropism of human immunodeficiency virus type 1. Proc Natl Acad Sci USA. 1992;89:9434–8.
[12] Hu DJ, Subbarao S, Vanichseni S, et al. Frequency of HIV-1 dual subtype infections, including intersubtype superinfections, among injection drug users in Bangkok, Thailand. AIDS. 2005;19:303–8.

[13] Ariën KK, Abraha A, Quiñones-Mateu ME, et al. The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. J Virol. 2005;79:8979–90.
[14] Sherpa C, Rausch JW, Le Grice SFJ. HIV genetic diversity – superpower of a formidable virus. Curr HIV Res. 2020;18:69–73.
[15] Marozsan AJ, Moore DM, Lobritz MA, et al. Differences in the fitness of two diverse wild-type human immunodeficiency virus type 1 isolates are related to the efficiency of cell binding and entry. J Virol. 2005;79:7121–34.
[16] Petersen A, Cowan SA, Nielsen J, et al. Characterisation of HIV-1 transmission clusters and drug-resistant mutations in Denmark, 2004 to 2016. Euro Surveill. 2018;23:1–9.
[17] Kouyos RD, Von Wyl V, Yerly S, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. J Infect Dis. 2010;201:1488–97.
[18] Nduva GM, Otieno F, Kimani J, et al. Quantifying rates of HIV-1 flow between risk groups and geographic locations in Kenya: a country-wide phylogenetic study. Virus Evol. 2022;8:1–14.
[19] Baral S, Beyrer C, Muessig K, et al. Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Infect Dis. 2012;12:538–49.
[20] Lwembe R, Lihana RW, Ochieng' W, et al. Changes in the HIV type 1 envelope gene from non-subtype B HIV type 1-infected children in Kenya. AIDS Res Hum Retroviruses. 2009;25:141–7.
[21] Lihana RW, Khamadi SA, Lwembe RM, et al. HIV-1 subtype and viral tropism determination for evaluating antiretroviral therapy options: an analysis of archived Kenyan blood samples. BMC Infect Dis. 2009;9:1–8.
[22] Mulinge M, Lemaire M, Servais JY, et al. HIV-1 tropism determination using a phenotypic env recombinant viral assay highlights overestimation of CXCR4-usage by genotypic prediction algorithms for CRRF01_AE and CRF02_AG. PLoS One. 2013;8:e60566.
[23] Struck D, Lawyer G, Ternes AM, et al. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic Acids Res. 2014;42:e144.
[24] de Oliveira T, Deforche K, Cassol S, et al. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics. 2005;21:3797–800.
[25] Stecher G, Tamura K, Kumar S. Molecular evolutionary genetics analysis (MEGA) for macOS. Mol Biol Evol. 2020;37:1237–9.
[26] National AIDS and STI Control Programme (NASCOP). Kenya Population-based HIV Impact Assessment (KENPHIA) 2018 Preliminary Report. 2020. Available from https://www.health.go.ke/wp-content/uploads/2020/02/KENPHIA-2018-PREL-REP-2020-HR3-final.pdf.
[27] Tago A, McKinnon LR, Wanjiru T, et al. Declines in HIV prevalence in female sex workers accessing an HIV treatment and prevention programme in Nairobi, Kenya over a 10-year period. AIDS. 2021;35:317–24.
[28] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–334.
[29] Fouchier RA, Groenink M, Kootstra NA, et al. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gpl20 molecule. J Virol. 1992;66:3183–7.
[30] Resch W, Hoffman N, Swanstrom R. Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks. Virology. 2001;288:51–62.
[31] Hongjaisee S, Nantasenamat C, Carraway TS, et al. HIVCoR: a sequence-based tool for predicting HIV-1 CRF01_AE coreceptor usage. Comput Biol Chem. 2019;80:419–32.
[32] Cabral GB, Ferreira JL, Coelho LPO, et al. Concordance of HIV type 1 tropism phenotype to predictions using web-based analysis of V3 sequences: composite algorithms may be needed to properly assess viral tropism. AIDS Res Hum Retroviruses. 2012;28:734–8.
[33] Verhofstede C, Brudney D, Reynaerts J, et al. Concordance between HIV-1 genotypic coreceptor tropism predictions based on plasma RNA and proviral DNA. HIV Med. 2011;12:544–52.
[34] Seclén E, Soriano V, González MM, et al. High concordance between the position-specific scoring matrix and geno2pheno algorithms for genotypic interpretation of HIV-1 tropism: V3 length as the major cause of disagreement. J Clin Microbiol. 2011;49:3380–2.
[35] Ceresola ER, Nozza S, Sampaolo M, et al. Performance of commonly used genotypic assays and comparison with phenotypic assays of HIV-1 coreceptor tropism in acutely HIV-1-infected patients. J Antimicrob Chemother. 2014;70:1391–5.
[36] Taylor BS, Hammer SM. The challenge of HIV-1 subtype diversity. N Engl J Med. 2008;359(18):1965–6.
[37] Scriven YA, Mulinge MM, Saleri N, et al. Prevalence and factors associated with HIV-1 drug resistance mutations in treatment-experienced patients in Nairobi, Kenya: a cross-sectional study. Med (United States). 2021;100:27460.
[38] Kageha S, Lihana RW, Okoth V, et al. HIV type 1 subtype surveillance in central Kenya. AIDS Res Hum Retroviruses. 2012;28:228–31.
[39] Khamadi SA, Lihana RW, Osman S, et al. Genetic diversity of HIV type 1 along the coastal strip of Kenya. AIDS Res Hum Retroviruses. 2009;25:919–23.
[40] Lihana RW, Khamadi SA, Lubano K, et al. HIV type 1 subtype diversity and drug resistance among HIV Type 1-infected kenyan patients initiating antiretroviral therapy. AIDS Res Hum Retroviruses. 2009;25:1211–7.
[41] Neilson JR, John GC, Carr JK, et al. Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. J Virol. 1999;73:4393–403.
[42] Kantor R, DeLong A, Balamane M, et al. HIV diversity and drug resistance from plasma and non-plasma analytes in a large treatment programme in western Kenya. J Int AIDS Soc. 2014;17:19262.
[43] Dalgleish AG, Beverley PC, Clapham PR, et al. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature. 1984;312:763–7.
[44] Zhang F, Dou Z, Ma Y, et al. Effect of earlier initiation of antiretroviral treatment and increased treatment coverage on HIV-related mortality in China: a national observational cohort study. Lancet Infect Dis. 2011;11:516–24.
[45] Ghosn J, Bayan T, Meixenberger K, et al. CASCADE Collaboration in EuroCoord. CD4 T cell decline following HIV seroconversion in individuals with and without CXCR4-tropic virus. J Antimicrob Chemother. 2017;72:2862–8.
[46] Sechet M, Roussel C, Schmit J-L, et al. X4 tropic virus prediction is associated with a nadir CD4 T-cell count below 100 Cells/mm3. Intervirology. 2015;58:155–9.
[47] Kimani J, McKinnon LR, Wachihi C, et al. Enumeration of sex workers in the central business district of Nairobi, Kenya. PLoS One. 2013;8:e543541–5.
[48] Montagna C, De Crignis E, Bon I, et al. V3 net charge: additional tool in HIV-1 tropism prediction. AIDS Res Hum Retroviruses. 2014;30:1203–12.
[49] Naganawa S, Yokoyama M, Shiino T, et al. Net positive charge of HIV-1 CRF01Æ V3 sequence regulates viral sensitivity to humoral immunity. PLoS One. 2008;3:e32061–9.
[50] Wei XM, Xu HF, Di CX, et al. Position 22 of the V3 loop is associated with HIV infectivity. Arch Virol. 2017;162:637–43.
[51] Zhou H-Z, Xu H-F, Xin X-M, et al. Position 22 of the V3 loop is associated with co-receptor usage and disease progression in HIV-1 subtype B isolates. Curr HIV Res. 2012;9:636–41.
[52] Cormier EG, Dragic T. The crown and stem of the V3 loop play distinct roles in human immunodeficiency virus type 1 envelope glycoprotein interactions with the CCR5 coreceptor. J Virol. 2002;76:8953–7.
[53] Soulié C, Fofana DB, Boukli N, et al. Performance of genotypic algorithms for predicting tropism of HIV-1CRF02_AG subtype. J Clin Virol. 2016;76:51–4.
[54] Lee GQ, Lachowski C, Cai E, et al. Non-R5-tropic HIV-1 in subtype A1 and D infections were associated with lower pretherapy CD4+ cell count but not with PI/(N)NRTI therapy outcomes in Mbarara, Uganda. AIDS. 2016;30:1781–8.
[55] Santoro MM, Armenia D, Fabeni L, et al. The lowest X4 Geno2Pheno false-positive rate is associated with greater CD4 depletion in HIV-1 infected patients. Clin Microbiol Infect. 2012;18:E289–98.
[56] Wanjiru R, Nyariki E, Babu H, et al. Beaten but not down! Exploring resilience among female sex workers (FSWs) in Nairobi, Kenya. BMC Public Health. 2022;22:1–12.
[57] Rojanaworarit C. Misleading epidemiological and statistical evidence in the presence of Simpson's paradox: an illustrative study using simulated scenarios of observational study designs. J Med Life. 2020;13:37–44.