HIV-1 pol Diversity among Female Bar and Hotel Workers in Northern Tanzania

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.1371/journal.pone.0102258</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:12717417">http://nrs.harvard.edu/urn-3:HUL.InstRepos:12717417</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
HIV-1 pol Diversity among Female Bar and Hotel Workers in Northern Tanzania

Ireen E. Kiwelu1,2, Vladimir Novitsky2, Elimsaada Kituma1,4, Lauren Margolin2, Jeannie Baca2, Rachel Manongi1,4, Noel Sam1,4, John Shao1, Mary F. McLane2, Saidi H. Kapiga3,4, M. Essex2*

1 Kilimanjaro Christian Medical Centre and College, Tumaini University, Moshi, Tanzania, 2 Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts, United States of America, 3 London School of Hygiene and Tropical Medicine, London, United Kingdom, 4 Kilimanjaro Reproductive Health Program, Moshi, Tanzania

Abstract

A national ART program was launched in Tanzania in October 2004. Due to the existence of multiple HIV-1 subtypes and recombinant viruses co-circulating in Tanzania, it is important to monitor rates of drug resistance. The present study determined the prevalence of HIV-1 drug resistance mutations among ART-naive female bar and hotel workers, a high-risk population for HIV-1 infection in Moshi, Tanzania. A partial HIV-1 pol gene was analyzed by single-genome amplification and sequencing in 45 subjects (622 pol sequences total; median number of sequences per subject, 13; IQR 5–20) in samples collected in 2005. The prevalence of HIV-1 subtypes A1, C, and D, and inter-subtype recombinant viruses, was 36%, 29%, 9% and 27%, respectively. Thirteen different recombination patterns included D/A1/D, A1/C, A1/C/A1, A1/C/A1/A1/C, A2/C/A2, CRF10 CD/C/CRF10_CD and CRF35 AD/A1/CRF10 CD, CRF35 AD. CRF10 CD was identified in Tanzania for the first time. All recombinant viruses in this study were unique, suggesting ongoing recombination processes among circulating HIV-1 variants. The prevalence of multiple infections in this population was 16% (n = 7). Primary HIV-1 drug resistance mutations to RT inhibitors were identified in three (7%) subjects (K65R plus Y181C; N60D; and V106M). In some subjects, polymorphisms were observed at the RT positions 41, 69, 75, 98, 101, 179, 190, and 215. Secondary mutations associated with NNRTIs were observed at the RT positions 90 (7%) and 138 (6%). In the protease gene, three subjects (7%) had M46I/L mutations. All subjects in this study had HIV-1 subtype-specific natural polymorphisms at positions 36, 69, 89 and 93 that are associated with drug resistance in HIV-1 subtype B. These results suggested that HIV-1 drug resistance mutations and natural polymorphisms existed in this population before the initiation of the national ART program. With increasing use of ARV, these results highlight the importance of drug resistance monitoring in Tanzania.

Introduction

Antiretroviral therapy (ART) has resulted in dramatic reduction of morbidity and mortality among HIV-1 infected individuals [1–3]. However, the emergence of drug-resistant viral variants and their potential spread remains a legitimate concern with serious implications for the course of the epidemic [4–7].

A virologic failure during the course of ART regimen is frequently related to HIV drug resistance, which arises from mutations in the genes that encode the molecular targets for the drugs, i.e., the HIV-1 protease (PR) and reverse transcriptase (RT) pol gene products. The HIV-1 RT is highly error-prone due to a lack of proofreading capacity, which often results in numerous polymorphisms. If viral mutations are associated with HIV drug resistance, these viral variants can have selective advantage and avoid drug pressure [8–10].

HIV-1 mutations associated with drug resistance are classified as either primary (major) or secondary (minor). Primary mutations are selected under drug pressure, may lead to a several-fold decrease in sensitivity to one or more antiretroviral drugs, and are extremely rare in the absence of treatment [11]. Secondary mutations are defined as having little or no effect on drug susceptibility, but may lead to increased resistance or increased replication capacity in the presence of major mutations [11,12]. Thus the appearance of a primary mutation in a genome already containing secondary mutations could influence the speed with which highly resistant viruses are selected during ART [13].

As access to ART rapidly increases in resource-limited countries, the prevalence of circulating HIV-1 drug resistant strains is also expected to increase. Acquired HIV-1 drug resistance developed during the course of treatment can spread upon viral transmission to newly infected individuals. The transmitted HIV-1 drug resistance may pose a challenge for therapeutic control of infection, by reducing the efficacy of first-line antiretroviral (ARV) treatment, and impact clinical outcome.

ART was introduced to Tanzania in 1995 with mono and dual regimens available to only a small number of patients due to the high cost of the drugs [14,15]. Access to ART has increased since the Tanzanian government launched its public-sector ART program free of charge in October 2004 [14,15].
The current standard first-line ART for HIV-1 infection in Tanzania consists of two nucleoside reverse transcriptase inhibitors (NRTIs), zidovudine (ZDV) or stavudine plus lamivudine (3TC), and one non-nucleoside reverse transcriptase inhibitor (NNRTI), nevirapine (NVP) or efavirenz (EFV). If the patient fails to respond to the first-line regimens, the second-line regimens include abacavir/didanosine (ABC/ddI) in combination with lopinavir or saquinavir boosted with ritonavir (LPV/r or SQV/r) [14–16]. Protease inhibitors (PIs) have been used rarely in Tanzania, and were not available in the public sector at the time the specimens for this study were collected.

Tanzania is one of the African countries severely affected by the HIV/AIDS epidemic with 5.7% of its 40 million people infected with HIV [17]. The HIV-1 subtypes A1, C, and D, as well as CRF01_CD and unique inter- and intra-subtype recombinant viruses, have been reported in Tanzania [18–26]. Recently we found that HIV-1 subtypes A1, C, and D, and inter- and intra-subtype recombinant viruses, were prevalent among female bar and hotel workers in Northern Tanzania [21,26]. HIV-1 subtypes and recombinants may be associated with various phenotypes such as disease progression [27], transmission patterns [28], as well as different pathways of drug resistant evolution [29–32].

HIV-1 subtypes may respond differently to ARV regimens [33–35]. Within the HIV-1 group M, it has been reported that isolates of subtype D tend to be less susceptible to ZDV, 3TC, ddI, NVP, and ritonavir [35]. Similarly, it has been reported that some subtype G strains have decreased susceptibility to PIs [36,37]. In HIV-1 CRF01_AE infection, the RT mutations T69N and V75M were seen more frequently than in HIV-1 subtype B [38].

The evolution of drug-resistant mutations in the non-B HIV-1 epidemic may not necessarily follow the patterns observed in HIV-1B infection [32]. However, limited information is available on non-B HIV-1 subtypes, particularly in regions like Tanzania where multiple HIV-1 subtypes A1, C, and D, as well as a high number of unique inter- and intra-subtype recombinant viruses, co-circulate. It is important to estimate the baseline prevalence of viral polymorphisms that might be associated with HIV-1 drug resistance in regions with multiple HIV-1 subtypes.

The present study estimated the prevalence of HIV-1 drug resistance mutations in pol within a high-risk population of HIV-1-infected ART-naive female bar and hotel workers, by single-genome amplification and sequencing (SGA/S) of specimens collected in 2005.

**Methodology**

**Ethics statement**

This study was conducted according to the principles expressed in the Declaration of Helsinki, and was approved by the research ethics committees at the Kilimanjaro Christian Medical Centre (KCMC), Tanzania National Institute for Medical Research, and Harvard School of Public Health (HSPH). All study subjects provided written informed consent for participation in the study.

**Study population**

The samples for this study were collected from treatment-naive female bar and hotel workers who were enrolled in a prospective cohort study between December 2004 and March 2007. Descriptions of assessment of HIV-1 status, recruitment of study subjects, characteristics of the cohort, and sampling procedures have been provided elsewhere [21,39,40]. All subjects enrolled in this study had similar sexual risk behaviors and are considered one of the high-risk populations for HIV-1 infection in Tanzania [21]. Subjects were followed-up quarterly over one year. At each study visit women were examined, consented, and interviewed about their sexual behavior and HIV-related risk factors, and blood samples were collected for further analysis.

Among 900 subjects enrolled in the study, 139 (17%) were HIV-1 positive by serological testing [21,39,40]. A subset of 50 out of 139 HIV-1 positive subjects with at least two samples collected one year apart has been recently characterized [21].

In this study we estimated the prevalence of HIV-1 drug resistance mutations and pol diversity. Thus a subset of 50 samples collected at enrollment was genotyped. The median age of subjects at study entry was 30 years (IQR 26–37). None of the study subjects reported previous exposure to ART. Viral load in plasma was quantified [21]. The viral load results are shown in Table S1.

**Single-genome amplification and sequencing (SGA/S)**

The isolation of peripheral blood mononuclear cells (PBMCs) from whole blood and genomic DNA have been described previously [21]. A fragment of the HIV-1 pol gene of about 1,660 bp encoding the entire PR and part of RT (position 2085–3763; HXB2 numbering) was amplified using a modified SGA/S technique [41,42] based on the limiting dilutions method [43]. The first-round PCR was conducted with primers BFI1 (5'-AAA TGA TGA CAG CAT GTC AGG GAG -3'); nucleotides 1826–1847; HXB2 numbering) and 3891L (5'-TCC TCT GTC AGT AAC ATA CCC TG-3'; nucleotides 3913–3932; HXB2 numbering). PCR amplification was performed in 20 μl and contained 1 μl of proviral DNA, 1.8 mM FastStart High Fidelity Buffer (Roche), 10 mM deoxynucleotide triphosphate (dNTPs (dATP, dCTP, dGTP and dTTP) (Roche), 10 pmol of each primer (Integrated DNA Technologies) and 5U FastStart High Fidelity Enzyme (Roche). The second-round PCR reaction was done with primers 2018U (5’-TTG GAA ATG TGG AAA GGA AGG AC-3'); nucleotides 2031–2050; HXB2 numbering) and 3815; HXB2 numbering). The reaction was carried out in a final volume of 25 μl and contained 1 μl of the first-round PCR product diluted 1:50 and 24 μl of master mix containing 1.8 mM FastStart High Fidelity buffer, 10 mM dNTPs, 10 pmol of each primer and 5U FastStart High Fidelity Enzyme. Thermal cycling conditions for both PCR rounds were as follows: 95°C for 2 minutes, followed by 35 cycles at 95°C for 20 sec, 54°C for 20 sec and 72°C for 2 sec with a final extension step at 72°C for 7 min. Reaction mixtures were stored at 4°C until use.

Amplified products were electrophoretically analyzed by applying 5 μl of second PCR amplification product to 1% agarose gel containing ethidium bromide, and visualized under ultraviolet light. Amplicons were purified by Exo-Sap [44] and directly sequenced on both strands on the ABI 3730 DNA analyzer using BigDye technology.

**Phylogenetic analysis and subtype determination**

Generated proviral DNA sequences were assembled and edited using SeqScape V 2.7. The pol sequences were aligned together with the HIV-1 subtype reference sequences retrieved from the Los Alamos HIV-1 sequence database [45] using the MUSCLE algorithm [46] in MEGA 5.0 [47]. Minor manual adjustment was done by Bioedit version 7.0 [48]. Maximum likelihood (ML) phylogenetic trees were constructed by PhyML version 3.0.1 [49] and visualized by FigTree v1.3.1 [50]. The approximate likelihood ratio test (aLRT) was used as a statistical test for support of splits [51]. aLRT values ≥0.95 were considered significant and are displayed at the tree nodes. The neighbor-joining (NJ) trees were constructed by MEGA 5.0 using the Kimura-two parameter.
model with 1000 bootstrap replicates [32]. Bootstrap values ≥80% were considered significant [33]. HIV-1 subtypes were determined based on branching topology, clustering and splits support of the analyzed sequences and their phylogenetic relationships with HIV-1 reference subtype sequences from the Los Alamos HIV-1 sequence database, as described elsewhere [21].

The proviral DNA sequences were analyzed for evidence of APOBEC3G-induced hypermutations using Hypermut tool V2.0 [34]. Thirteen quasispecies from five subjects with a p value of ≤0.05 were considered enriched for mutations consistent with APOBEC3G signatures and were excluded from analysis. The final set included 622 pol sequences.

Screening for inter-subtype recombination and breakpoints identification

All sequences generated in this study were screened for evidence of inter-subtype recombination by Recombination Identification Program (RIP 3.0) [45] and REGA HIV-1 Subtyping Tool-Version 2.0 [35]. The identified recombinant sequences were further analyzed for breakpoints identification using bootscan by SimPlot software v3.5.1 [56] as previously described [21]. The HIV-1 subtype reference sequences were retrieved from the Los Alamos HIV Sequence Database [45]. Identified breakpoints were visually inspected in BioEdit. To confirm the HIV-1 subtypes in the inter-subtype recombinant viruses, nucleotide sequences on both sides of the breakpoint were analyzed independently by re-constructing phylogenetic trees using the splits at the putatively identified breakpoints, as described previously [21].

For the sequences with both recombinants and pure subtypes (multiple infections), we established whether or not the recombinant viruses originated within the infected individuals. The recombinant sequences were split at the putatively identified breakpoints, realigned with the pure subtypes which originated from the same subjects together with the reference sequences, including CRFs if required and examined by neighbor-joining phylogenetic tree analysis.

Multiplicity of HIV-1 infection

To determine HIV-1 infections with multiple viral variants, the HIV-1 pol gene analysis was performed as described recently for HIV-1 env gene analysis [21].

Drug resistance mutation analyses

The HIV-1 pol quasispecies were evaluated for HIV-1 drug resistance mutations and for naturally occurring polymorphisms in the PR and RT using the International AIDS Society–USA (IAS-USA) major mutation list [57] and the Stanford University HIV Drug Resistance Database [58].

Control for cross-contamination

Control of laboratory cross-contamination during specimen collection, processing, amplification, and/or sequencing was performed routinely, as described previously [21].

Statistical analysis

Descriptive statistics were performed using Sigma Stat v.3.5. The bootstrap and aLTR support values for splits in the inferred phylogenetic trees were computed by MEGA 5.0 and PhyML respectively.

Accession numbers

Sequences have been assigned GenBank database accession numbers KF530900–KF531521.

Results

HIV-1 pol subtyping

In this study we targeted the same subjects described in our recent study on diversity of the V1-C5 region of HIV-1 env gp 120 [21]. A total of 622 pol sequences were generated from 45 subjects [21]. The median number of pol sequences per subject was 13 (IQR 5–20). Samples from five subjects (codes 86, 181, 321, 404, 945) could not be amplified.

Analysis of phylogenetic relationships between the generated pol sequences revealed that A1 was the most common HIV-1 subtype (35.6%), followed by subtype C (20.9%) and HIV-1 inter-subtype recombinant viruses (26.7%). HIV-1 subtype D was less prevalent (8.9%). Similar results were observed in our previous study on the V1-C5 env gene. However, the pol-based prevalence of HIV-1 inter-subtype recombinants (26.7%) was higher than the env-based prevalence (8.6%) [21], although this finding did not reach statistical significance (p = 0.0513, Fisher exact test). Using the combined env/pol data, the overall HIV-1 subtype distribution was A1/A1 (35.6%), C/C (24.4%), D/D (4.4%) and inter-subtype recombinants (35.6%), highlighting a higher rate of HIV-1 inter-subtype recombinants in the combined analysis (Table 1). Figure 1 shows the phylogenetic relationships among a subset of 458 non-recombinant HIV-1 pol sequences. The 134 HIV-1 inter-subtype recombinant pol DNA sequences were analyzed separately, as their topology in the phylogenetic tree was not informative.

HIV-1 inter-subtype recombinant viruses

The distribution of HIV-1 inter-subtype recombinant viruses in two regions (env and pol) is shown in Table 2. HIV-1 inter-subtype recombinant viruses were found in 12 (26.7%) of the 45 subjects (Table 1). In seven subjects (codes 33, 87, 355, 558, 733, 838 and 909), all of the quasispecies for the pol gene were represented by inter-subtype recombinant viruses, while five subjects (codes 177, 209, 322, 491 and 603) had multiple HIV-1 subtype infections, suggesting possible recombination and/or dual infections in this population. To determine the relationship between the non-recombinant subtypes and the putative recombinant regions based on the pol gene, phylogenetic analysis was performed. Results for the pol gene showed that in four of the five subjects (codes 177, 322, 491 and 603) with dual infections, the pure subtypes were found to be parental strains of the recombinant viruses, while in the remaining subject (code 209), the pure subtype was not a parental strain of the recombinant virus (data not shown).

We also identified two complex circulating recombinant forms (CRFs), CRF10_CD/C/CRF10_CD and CRF35_AD/A1/CRF35_AD in this population (Fig. 2). CRF10_CD has been previously reported in Tanzania [20,22,23], while this is the first time that CRF35_AD has been reported in this population as well as in Tanzania. The CRF35_AD/A1/CRF35_AD recombinant was further analyzed to confirm the recombination patterns of the 40 generated viral quasispecies of subject 733 (number of viral quasispecies per subject ranged from 1 to 45 quasispecies). ML trees were generated separately for the three regions, 2,080–2,536, 2,537–2,987, 2988–3746 (HXB2 numbering). Results showed that the first analyzed fragment clustered with HIV-1 CRF35_AD reference (Fig. S1B; aLTR support of 0.88). The second fragment clustered with HIV-1 subtype A1 reference (Fig. S1C; aLTR support of 0.76). The third fragment clustered with HIV-1 CRF35_AD reference (Fig. S1D; aLTR support of 0.93). The low aLTR support value for all the three fragments could possibly be due to the short length and limited number of informative sites. The relationship of these strains to other published HIV-1 pol sequences was investigated with the BLAST subtyping tool [59].
The closest available sequence was HIV-1 isolate TV725 from Canada [60]. Similar analyses were performed for the CRF10_CD/C/CRF10_CD recombinant virus to confirm the HIV-1 sub-genomic regions (data not shown).

Phylogenetic analysis of both env and pol genes indicated 16 (35.6%) of 45 subjects were infected with HIV-1 inter-subtype recombinant viruses. Among these recombinant viruses: two subjects (codes 33 and 322) had recombination breakpoints in both env and pol regions; ten subjects (codes 87, 177, 209, 355, 491, 558, 603, 733, 838 and 909) had a virus with breakpoints in the pol gene only; two subjects (codes 471 and 510) had recombination breakpoints in env gene only; and two subjects (codes 697 and 740) had discordant env and pol subtypes, A1/C and A1/D, respectively.

Recombinant strains were analyzed in detail based on the location of recombination breakpoints. Putative recombinant regions were split according to the breakpoints and analyzed by neighbor-joining trees and HIV-1 reference subtypes. Results for this analysis are shown in Figure 2. Thirteen different recombination patterns were observed in the pol gene: 11 recombination patterns were observed in 11 of the 12 subjects with recombinant viruses, while two different recombination patterns were observed in the remaining subject (Fig. 2; code 209). Of note, in the env analysis of the same subjects, we observed only five different patterns [26], suggesting that the pol region has a high recombination rate in this population.

The HIV-1 inter-subtype recombinant viruses in this study were unique, shared no recombination breakpoints, and demonstrated

---

**Table 1.** Distribution of HIV-1 subtypes among female bar and hotel workers in Moshi, Kilimanjaro, Tanzania in 2005.

<table>
<thead>
<tr>
<th>Subtype</th>
<th>V1-CS env gene*</th>
<th>pol gene (PR and RT)</th>
<th>Env*pol genes combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>24 (53.3%)</td>
<td>16 (35.6%)</td>
<td>16 (35.6%)</td>
</tr>
<tr>
<td>C</td>
<td>14 (31.1%)</td>
<td>13 (28.9%)</td>
<td>11 (24.4%)</td>
</tr>
<tr>
<td>D</td>
<td>3 (6.7%)</td>
<td>4 (8.9%)</td>
<td>2 (4.4%)</td>
</tr>
<tr>
<td>Recombinant</td>
<td>4 (8.9%)</td>
<td>12 (26.7%)</td>
<td>16 (35.6%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>45</strong></td>
<td><strong>45</strong></td>
<td><strong>45</strong></td>
</tr>
</tbody>
</table>

*Kiwelu et al., 2013.
doi:10.1371/journal.pone.0102258.t001

---

**Figure 1.** ML phylogenetic tree of HIV-1 protease and reverse transcriptase sequences of the pol gene from 45 subjects. The ML was constructed by PhyML 3.0.1 and visualized in FigTree. The tree is rooted with the HIV-1 group N consensus sequence as an outgroup. 488 non-recombinant pol sequences generated from the 45 subjects were analyzed with HIV-1 reference subtypes from the Los Alamos HIV-1 Sequence Database. HIV-1 inter-subtype recombinant pol DNA sequences were excluded in this figure. HIV-1 reference subtypes A1, C, and D are shown in red, pink and blue, respectively. The other references are shown in orange. Approximate likelihood ratio test (aLRT) values of >0.95 were considered significant and are shown by an asterisk (*). The scale at the bottom of the figure corresponds to 0.01 nucleotide substitutions per site.
doi:10.1371/journal.pone.0102258.g001
complex subtype structure. For example, a portion of HIV-1 in subject 209 clustered with none of the group M HIV-1 subtypes. However, when we compared the 18 viral quasispecies of subject 209 with the HIV-1 sequences in GenBank using BLAST [59], the closest available HIV-1 sequences to the 18 viral quasispecies of subject 209 were HIV-1 pol sequences from Spain with 94% identity [61].

HIV-1 multiple infections

The prevalence of multiple HIV-1 infections in this study was 16% (n = 7). Two subjects (codes 66 and 291) were infected with HIV-1 variants of HIV-1 subtype C, while the remaining five subjects (codes 177, 209, 322, 491, and 603) were infected with both pure subtypes and recombinant viruses. Recently we reported that 12 (27%) of 45 subjects had multiple HIV-1 infections based on analysis of the HIV-1 env gene [21]. However, congruence between two structural viral genes, env and pol, in identification of multiplicity of HIV-1 infection was poor, at least in this population. Thus, only one of 12 subjects with multiple env infections (code 291) was infected with multiple variants of HIV-1 subtype C based on the pol gene analysis. Multiple HIV-1 infection was not confirmed in the other subjects due to non-significant bootstrap support values (6 subjects), low number of quasispecies (2 subjects), or no evidence for multiple distinct variants (3 subjects). At the same time, one subject (code 66) with homogeneous env quasispecies indicating HIV-1 infection with a single variant, was classified as infected with multiple HIV-1 variants based on the pol gene analysis. A summary of HIV-1 infection with single and multiple viral variants is shown in Table S2.

Reverse transcriptase inhibitor (RTI) resistance mutations

Table 3 summarizes the mutations and polymorphisms associated with PR and RT inhibitors. Primary HIV-1 drug resistance mutations to RT inhibitors were identified in three (7%; codes 201, 245 and 291) of the 45 subjects. The identified NRTI mutations included D67N and K65R, while the NNRTI mutations were V106M and Y181C. The NRTI-associated polymorphisms were observed at positions 41, 69, 75 and 215. The prevalence of the secondary mutations associated with NNRTI at positions 90 and 138 was 11% (n = 5). Single polymorphisms associated with NNRTIs were detected at positions 98, 101, and 190. A subtype-specific polymorphism at position 179 (V179I) was observed among all 16 (100%) subjects infected with HIV-1 subtype A1. Some subjects harbored multiple secondary mutations (e.g., subject 905 with V90I and E138K) and/or polymorphisms (e.g., subject 237 with A98S and L101Q). The significance of the observed polymorphisms in HIV-1 non-B subtypes is unknown.
Protease polymorphisms associated with protease inhibitor resistance mutations

In the protease gene, three subjects (7%) had major mutations at position 46 (M46I/L) associated with PR drug resistance mutations (Table 3). Viral polymorphisms were present at multiple positions across protease, e.g., at positions 16, 20, 34, 36, 60, 62, 63, 64, 69, 71, 74, 77, 89, 90 and 93, that are associated with PI-resistance in HIV-1 subtype B. The most frequent polymorphisms were seen at positions M36I (81%), H69K (86%), L89M (74%), and I93L (62%). Given the abundance of viral polymorphisms potentially associated with PI-drug resistance in subtype B, their frequencies were compared with the HIV-1 subtypes A1 and C frequencies in treatment-naive individuals available from the Stanford HIV Drug Resistance Database (Fig. 3). No statistically significant difference was found (p = 0.161 and p = 0.104, respectively), suggesting HIV-1 subtype-specific polymorphisms are not associated with PI drug resistance.

Discussion

This study determined the prevalence of HIV-1 subtypes and HIV-1 drug resistance mutations among treatment-naive female bar and hotel workers, a high-risk population for HIV-1 infection in Moshi, Tanzania. The most prevalent subtype was HIV-1 subtype A1, followed by HIV-1 subtype C, HIV-1 inter-subtype recombinant viruses and HIV-1 subtype D. Similar results were reported in our previous HIV-1 env-based study [21]. However, the frequency of HIV-1 inter-subtype recombinant viruses in the HIV-1 pol gene (26.7%) showed a trend to be higher (p = 0.051; Fisher exact test) than the frequency observed in the HIV-1 env gene (8.6%) in the same population [21]. Similarly, a high prevalence of HIV-1 inter-subtype recombinant viruses was reported in the previous studies using the pol gene [62–65]. The combined HIV-1 env [21] and pol prevalence of HIV-1 inter-subtype recombinant viruses was 33.6%. It is possible that near full-length HIV-1 genome analysis could show even higher prevalence of recombinant viruses. Our study supported the previous findings that examining multiple regions of the HIV-1
genome may allow detection of more subjects infected with multiple infections and recombinant viruses [66–68]. A high prevalence of HIV-1 inter-subtype recombinant viruses in the HIV-1 pol gene suggests that recombination occurs in the pol region. These results are consistent with the previous studies demonstrating that pol appears to be a hot spot for recombination [62,63,66,69].

The recombination patterns and breakpoints in the HIV-1 pol gene were unique in all 12 (26.7%) subjects infected with HIV-1 inter-subtype recombinant viruses. In contrast, in the HIV-1 env gene we observed only five recombination patterns in the same population [26]. Additionally, five of the 12 subjects with recombinant viruses had dual infections of pure HIV-1 subtypes and recombinant viruses. The pure HIV-1 subtypes in four subjects were parental strains of the recombinants, suggesting that dual infections were responsible for the generation of these recombinants. Six complex recombinant viruses including circulating recombinant forms (CRFs) were reported in this study: A1/U/A1, C/U/A1, U/D/U, A2/C/A2, CRF10_CD/C/ CRF10_CD, and CRF35_AD/A1/CRF35_AD. The complex recombinant virus CRF35_AD/A1/CRF35_AD was not previously reported in Tanzania. The CRF35_AD had been previously described among injecting drug users in Kabul, Afghanistan [60,70,71]. The HIV-1 sub-subtype A2 was reported for the first time in Moshi among female bar and hotel workers [72] and was later reported among pregnant women in the Kilimanjaro region [23]. The CRF10_CD recombinant has been previously described in Tanzania [20,22,23]. Our results suggest that the HIV-1 sub-subtype A2 and CRF10_CD are present at a low prevalence in this population. The three recombinant variants A1/U/A1, C/U/A1 and U/D/U include regions that did not cluster with any HIV-1 group M subtype, and were therefore considered unclassified regions (U). The source of unclassified regions remains unknown although we cannot exclude a complex recombination between recombinants of unknown degree.

All recombinant viruses identified in this study were unique, and contained the co-circulating HIV-1 subtypes A, C and D in Tanzania. Similar results were reported in the HIV-1 env gene in the same population [21] and in previously published studies in Tanzania [23,73,74].

The high prevalence of HIV-1 inter-subtype recombinant viruses in this population may be associated with multiple factors. First, there is the high-risk behavior of women working in hotels and bars in Moshi, Tanzania, who have a high rate of sexual

### Table 3. Mutations and polymorphisms at positions associated with drug resistance to PIs, NRTIs, and NNRTIs among female bar and hotel workers in Moshi, Kilimanjaro, Tanzania, in 2005.

<table>
<thead>
<tr>
<th>Subject code</th>
<th>HIV-1 Subtype</th>
<th>Total no. of quasispecies</th>
<th>Mutation position</th>
<th>No. of quasispecies with mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>276</td>
<td>C</td>
<td>10</td>
<td>M46I</td>
<td>1</td>
</tr>
<tr>
<td>480</td>
<td>A1</td>
<td>32</td>
<td>M46I</td>
<td>1</td>
</tr>
<tr>
<td>733</td>
<td>CRF35_AD/A1/CRF35_AD</td>
<td>40</td>
<td>M46I</td>
<td>1</td>
</tr>
<tr>
<td>245</td>
<td>A1</td>
<td>22</td>
<td>D67N</td>
<td>1</td>
</tr>
<tr>
<td>201</td>
<td>C</td>
<td>13</td>
<td>K65R</td>
<td>1</td>
</tr>
<tr>
<td>740</td>
<td>D</td>
<td>22</td>
<td>M41I</td>
<td>1</td>
</tr>
<tr>
<td>66</td>
<td>C</td>
<td>13</td>
<td>T69A</td>
<td>1</td>
</tr>
<tr>
<td>245</td>
<td>A1</td>
<td>22</td>
<td>T69P</td>
<td>1</td>
</tr>
<tr>
<td>209</td>
<td>A1</td>
<td>24</td>
<td>V75A</td>
<td>1</td>
</tr>
<tr>
<td>909</td>
<td>A2/C/A2</td>
<td>21</td>
<td>V75A</td>
<td>1</td>
</tr>
<tr>
<td>46</td>
<td>A1</td>
<td>24</td>
<td>T215A</td>
<td>1</td>
</tr>
<tr>
<td>201</td>
<td>C</td>
<td>13</td>
<td>Y181C</td>
<td>1</td>
</tr>
<tr>
<td>491</td>
<td>C</td>
<td>45</td>
<td>V106M</td>
<td>4</td>
</tr>
<tr>
<td>168</td>
<td>A1</td>
<td>19</td>
<td>V90I</td>
<td>1</td>
</tr>
<tr>
<td>838</td>
<td>CRF10_CD/C/CRF10_DC</td>
<td>16</td>
<td>V90I</td>
<td>2</td>
</tr>
<tr>
<td>905</td>
<td>A1</td>
<td>16</td>
<td>V90I</td>
<td>9</td>
</tr>
<tr>
<td>237</td>
<td>A1</td>
<td>13</td>
<td>E138A</td>
<td>5</td>
</tr>
<tr>
<td>291</td>
<td>C</td>
<td>32</td>
<td>E138A</td>
<td>2</td>
</tr>
<tr>
<td>237</td>
<td>A1</td>
<td>13</td>
<td>A98S</td>
<td>6</td>
</tr>
<tr>
<td>480</td>
<td>A1</td>
<td>32</td>
<td>G190E</td>
<td>1</td>
</tr>
</tbody>
</table>
partner change which may facilitate multiple HIV infections and recombinations [73,75]. Second, co-circulation of HIV-1 subtypes A1, C, D and some other HIV-1 subtypes in this population contributes to the generation of inter-subtype recombinant viruses [21]. Third, analysis of multiple regions of the HIV-1 genome including env and pol genes allows the detection of more recombinant viruses.

In this study the prevalence of HIV-1 multiple infections was 16% (n = 7 of 45). Two of the seven subjects were infected with multiple HIV-1 variants of the same subtype, while the remaining five subjects were infected with a mixture of pure HIV-1 subtypes and recombinant viruses. However, based on the HIV-1 env gene, only 12 (27%) of the 45 subjects were infected with multiple HIV-1 variants of the same subtype [21]. Based on the HIV-1 pol gene five more subjects were infected with HIV-1 multiple infections, suggesting that analysis of a single region of the HIV-1 genome may underestimate the true proportion of HIV-1 multiple infections.

Analysis of HIV-1 drug resistance mutations and polymorphisms among female bar and hotel workers revealed that three (7%) of the 45 subjects harbored HIV-1 drug resistance mutations to RT inhibitors, NRTIs and NNRTIs. This is higher than in some previous studies in Tanzania among HIV-1 treatment-naive individuals [15,64] but is in line with other studies in Tanzania [14,76]. Our results suggest that HIV-1 strains with drug-resistant mutations to RT inhibitors existed in this population due to suboptimal regimens and adherence during the earlier phase of the HIV/AIDS epidemic in Tanzania, i.e., before the implementation of the national ART program.

Three (7%) subjects had a major mutation at the protease amino acid position 46 (M46I/L) that confers high resistance to protease inhibitors (PIs) only in combination with other mutations, and can occur among untreated persons as natural polymorphisms [77–80], as was reported previously in Tanzania among treatment-naive individuals [14,15]. Since PIs were not used in Tanzania at the time that the samples were collected, the mutations M46I and M46L most likely represent natural polymorphisms rather than transmitted drug-resistant strains. However, unreported exposure to PI or HIV transmission from individuals receiving PIs cannot be excluded.

All subjects in this study harbored three or more polymorphisms at amino acid positions associated with PIs in HIV-1 subtype B. H69K (86%), M36I (81%), I93L (74%), and F68K (48%) were considered to be subtype-specific natural polymorphisms since they occur at high frequency in HIV-1 subtypes A1, C or D [81]. Data from the Stanford HIV Drug Resistance Database for HIV-1 subtypes A1 and C confirmed that the observed polymorphisms are common among HIV-1 treatment-naive individuals [58]. Polymorphisms were defined as mutations that occurred in more than 1% of sequences from untreated persons. Subtype-specific polymorphisms were defined as mutations that were significantly more prevalent in each non-B subtype than in subtype B viruses from untreated persons [30]. Subjects with and without HIV-1 drug resistance mutations had similar sexual risk behaviors.

The undisclosed use of ART can be a hidden problem in sub-Saharan Africa. Recently Kahle et al. examined drug levels among subjects with low HIV-1 RNA loads and reported a higher than expected prevalence of unreported ARV drugs use [82]. In this study five subjects were found with primary drug resistant mutations associated with NRTIs, or protease inhibitors. Only one of these subjects, code 201, had plasma viral load below 2.7 log_{10} copies/ml. Due to a shortage of plasma specimens we were not able to measure levels of ART in these subjects, which is a clear study limitation. It would be important to address levels of ART in individuals with drug-resistant mutations and/or low HIV-1 RNA load.

The presence of a high number of substitutions at positions associated with drug resistance mutations in non-B viruses might influence the risk of treatment failure through lowering the genetic barriers to the development of drug resistance [83,84]. Further

![Figure 3. Polymorphisms in protease gene in HIV-1 subtypes A1 and C, compared to ARV-naïve individuals from the Stanford HIV-1 Drug Resistance Database.](http://example.com/figure3)
likely to be associated with low efficiency of PCR amplification. In the subjects had undetectable plasma HIV-1 viral RNA, which is a relatively low number of quasispecies available. Fourth, some of quasispecies is needed; however, some of the subjects had a super-infection, or both. Thirdly, in order to detect HIV-1 unknown, and the study had no power to determine whether the HIV-1 inter-subtype recombinant viruses, recombinants and multiplicity of infection. Secondly, the duration and stage of HIV-1 infection were unknown, and the study had no power to determine whether the HIV-1 inter-subtype recombination was due to co-infection, super-infection, or both. Thirdly, in order to detect HIV-1 multiple infections of the same subtype, analysis of multiple viral quasispecies is needed; however, some of the subjects had a relatively low number of quasispecies available. Fourth, some of the subjects had undetectable plasma HIV-1 viral RNA, which is likely to be associated with low efficiency of PCR amplification. In conclusion, our study demonstrated that the HIV-1 epidemic in Tanzania is highly diverse, with multiple HIV-1 infections and unique HIV-1 inter-subtype recombinants, as well as complex circulating recombinant forms. HIV-1 subtypes A1 and C are still prevalent in this population, including large proportions of unique HIV-1 inter-subtype recombinant viruses. CRF35_AD was reported for the first time in this population, in Moshi as well as in Tanzania. We have further reported the baseline prevalence of HIV-1 drug resistance mutations and natural polymorphisms at amino acid positions associated with HIV-1 drug resistance to NRTIs, NNRTIs and PIs before ARV drugs were widely used in Tanzania. The results of this study will help to better understand the pathogenesis of HIV-1 infection and the emergence of drug resistance, and should aid in the development of therapeutic strategies in Tanzania.

Supporting Information

Figure S1 Maximum likelihood (ML) phylogenetic tress of three segments of the 40 viral quasispecies of subject 733. Fig. S1A: The bootstrap plot generated by SimPlot analysis using a consensus DNA sequence of subject 733 with distinct recombination pattern CRF35_AD/A1/CRF35_AD. HIV-1 subtype A1 is shown in the red bar and CRF35_AD is shown in the blue bar. Fig. S1B is a ML tree of the fragment classified as HIV-1 CRF35_AD. Fig. S1C is a ML tree of the fragment classified as HIV-1 subtype A1, and Fig. S1D is a ML tree of the fragment classified as HIV-1 CRF35_AD. The viral quasispecies of subject 733 are shown in green and the legend at the right of the bootstrap plot indicates reference HIV-1 subtypes. aLRT values ≥0.95 were considered significant and are shown by asterisk (*). Selected aLRT values are shown at the branch node of the tree. Scale at the bottom of the figure corresponds to 0.1 nucleotide substitution per site.

Table S1 Plasma HIV-1 RNA viral load analyzed from the 45 subjects at the early time point (baseline).

Table S2 HIV-1 single and multiple variants of the same subtype among female bar and hotel workers in Moshi, Kilimanjaro region, Tanzania.

Acknowledgments

The authors would like to thank the study participants, research team, and Kilimanjaro Christian Medical Centre and Kilimanjaro Christian Medical College for their contributions and support of this work. We would like to thank Beth Chaplin for her assistance with the HIV-1 viral load testing. We are also grateful to Lendney Melton for his excellent editorial and administrative assistance. This study was supported by a grant from the International Partnership for Microbicides (SHK), the NIH Fogarty International Center AIDS International Training and Research Program grant D45 TW000004 (IEK), and the Fogarty International Center Scholars and Fellows Program (IEK).

Author Contributions

Conceived and designed the experiments: IEK VN RM NS JS MMF SHK ME. Performed the experiments: IEK VN EK LM JB MMF. Analyzed the data: IEK VN EK LM JB MMF ME. Contributed reagents/materials/analysis tools: IEK VN RM NS JS MMF SHK ME. Wrote the paper: IEK VN EK LM JB MMF NS JS MMF SHK ME.

References


