

NON SYNONYMOUS SUBSTITUTIONS IN HIV-1 GAG ARE FREQUENT IN EPITOPES OUTSIDE THE MAJOR HYDROPHOBIC REGION AND ASSOCIATED WITH SUBTYPE DIFFERENCES

Babatunde A. OLUSOLA¹, David O. OLALEYE¹, Georgina N. ODAIBO^{1*}

¹Department of Virology, College of Medicine, University of Ibadan, Ibadan, Oyo State, Nigeria.

*Correspondence/Requests for Reprints

Prof Georgina N. Odaibo

Abstract

In 2018, an estimated 38 million people lived with HIV-1 infection resulting in 770,000 deaths. More than 50% of this infection and its associated deaths occurred in Sub-Saharan Africa. There is the need to develop an HIV-1 vaccine if the epidemic would be effectively controlled. Cytotoxic T Lymphocytes (CTL) epitopes within the Major Hydrophobic Region (MHR) have been shown to be highly immunogenic and immuno-dominant. These conserved epitopes have recently been the focus of vaccines studies. Despite the West African epicenter having one of the highest numbers of diverse circulating HIV-1 strains, very few longitudinal studies have checked the frequencies of CTL immune escape variants on epitopes within and without the MHR for HIV-1 strains circulating in the region.

In this study we describe non-synonymous substitutions within and without the MHR of HIV-1 GAG genes isolated from 10 infected Nigerians at the early stages of HIV-1 infection. Furthermore, we analyzed these substitutions longitudinally in five infected individuals from the early stages of infection up until when antibodies become detectable. We identified 3 non synonymous substitutions within the MHR of HIV-1 GAG genes isolated from the early HIV infected individuals. Fourteen and nineteen mutations outside the MHR were observed before and after detection of antibodies respectively while 4 mutations were found within the MHR. These substitutions include previously mapped CTL epitope immune escape mutants. CTL immune pressure likely leave different footprints on HIV-1 GAG epitopes within and outside the MHR. This information is crucial for future vaccine designs for use in the West African region.

Introduction

In 2018, an estimated 38 million people lived with HIV-1 infection resulting in 770,000 deaths. The rate of new infections also increased in the Sub-Saharan Africa region. More than 50% of HIV infection and its associated deaths also occurred in this region[1]. Combination antiretroviral therapies (cART) have been effective in suppressing viremia to undetectable levels, increasing survival, improving quality of life and decreasing infectiousness in infected individuals[2,3]. However, just about half of infected individuals receive treatment. Sub-Saharan Africa despite being the highest hit region with the virus, has the lowest access to treatment[1]. Access to treatment is unlikely to increase to optimal levels because of economic, social and pharmacologic challenges associated with cART use[2,4]. Moreover, cART is not curative as persons on treatment has to use drugs to control emergence of latent reservoir HIV (immune escape as well as drug resistant strains) for the rest of their lives[2,5]. Therefore, to control the HIV epidemic especially in African countries, there is the need to develop a safe, cost-effective, durable and accessible HIV-1 vaccine[4,6–8].

Induction of broadly neutralizing antibodies is the main stay of all protective anti-viral vaccines. However, this has been very difficult to generate for HIV-1 because of the plasticity of the virus genes due to the action of reverse transcriptase enzyme[2,4,7]. Also, broadly neutralizing antibodies develop very late in infection after establishment of the latent reservoir strains[9,10]. Recent vaccine efforts have been directed towards the stimulation of cellular immunity. Therefore, consideration of HIV-1 vaccine candidates that stimulate cellular immunity has been the focus of many recent vaccine studies. Cytotoxic T Lymphocytes (CTL) commonly referred to as CD8+ T cells have been extensively showed to control HIV-1 replication especially during the early stages of infection[4,11,12]. Previous studies have associated expansion of CTL with the control of acute infections[13–17].

Although non-synonymous substitutions are more predominant in the ENV gene of HIV-1, studies have shown that these substitutions also occur in CTL epitopes of the HIV-1 GAG gene especially in the P17 and P24 regions[18–21]. Initially, CTL control of acute HIV infections were associated with HLA protection[22–25], recent studies have however showed that magnitude and breadth of CTL responses as well as CTL epitopes presented by similar HLA alleles during the early stages of infection may better predict disease outcomes[26–32]. CTL epitopes within the major hydrophobic region (MHR) have been shown to be promiscuously presented by similar HLA alleles[33]. The MHR, found within the capsid

gene, is a conserved motif among retroviruses[22,34]. This region is important in particle assembly and viral infectivity[35]. Furthermore, CTL epitopes within the MHR have been shown to be highly immunogenic and immuno-dominant [19–21,36]. It is also hypothesized that escape mutations within the MHR are likely to be deleterious to HIV as they seem to be associated with fitness cost[37–39]. These conserved epitopes (Gag 240–249, TSTLQEQIGT; Gag 162–172, KAFSPEVIPMF and Gag 203–212, ETINEEAAEW) have recently been the focus of universal therapeutic vaccines[2,6,7,19,40].

However, there is limited information on immune escape strains within and without the MHR for HIV-1 strains circulating in the West African epicenter. Immune Escape due to CTL epitopes outside the MHR may be a crucial factor to consider in the design of therapeutic HIV vaccines[41] as they may provide opportunities for compensatory mutations on replicative fitness[42]. It is also important for HIV-1 vaccines when developed, to be effective against strains circulating in African countries[4]. Despite the West African epicenter having one of the highest numbers of diverse circulating HIV-1 strains, very few longitudinal studies have checked the frequencies of CTL immune escape variants on epitopes within and without the MHR for HIV-1 strains circulating in the region. In this study we describe non-synonymous substitutions within and without the MHR of HIV-1 GAG genes isolated from 10 infected Nigerians at the early stages of HIV-1 infection. These individuals were part of a previously described study[43] (and study 3). Furthermore, using phylogenetic tools as well as programs and databases in the Los Alamos National Laboratory HIV Sequence Database (<https://www.hiv.lanl.gov>), we analyze these substitutions longitudinally in five infected individuals from the early stages of infection up until when antibodies become detectable.

Methods

Study sites and patient population

Twenty-three individuals at the early stages of HIV-1 infection identified and described previously (study3) were recruited for this study (see Table 1). Ten persons out of these were studied for early HIV-1 infection. Another 5 individuals identified in 2017 were followed up until antibodies were detected. Profile of the follow up schedule is shown in Table 2. HIV-1 antibodies were detected after two years, although the exact time when HIV-1 antibodies became detectable is unknown. There was about 12 month's interval between the penultimate

visit and the one in which antibodies were detected. At every visit, these individuals were screened for HIV-1 infection using earlier described protocol[43,44].

Recruitment of participants, sample collection and processing

Participants were recruited for this study after obtaining informed consent. Experiments were conducted with the understanding and the consent of participants. Socio demographic data were also collected from these persons. Feedback on results were provided within a week of sample collection. Individuals were counselled and encouraged to continue presentation for testing. These persons were not placed on ART since they were still at the early stages of infection. In Nigeria, only persons with detectable antibodies are placed on treatment. Five milliliters of whole blood were collected in EDTA bottles from participants at every visit. Plasma was separated from the samples immediately after collection, stored at -20°C and transported in cold chain to a central laboratory for analysis. The samples were then stored at -80°C until analyzed. Blood samples were analyzed for HIV antigen/antibody, serum creatinine, HIV-1 RNA viral load (at baseline) and HIV-1 GAG DNA.

Identification of early HIV-1 infection and detection of antibodies

The updated CDC algorithm of laboratory testing for the diagnosis of early and chronic HIV-1 infection was used for this study as previously described[43]. Briefly, samples were identified as early HIV infection if HIV-1 DNA was detected after being positive with a 4th generation HIV ELISA (detection of both antigen and antibody) and negative with 3rd generation HIV ELISA (detection of HIV antibody only) while those from which HIV-1 DNA scored positive with both 4th and 3rd generation ELISA, were classified as chronic HIV infection. Genscreen™ ULTRA HIV-Ag-Ab ELISA kit (Biorad, Hercules, California) was used as the 4th generation ELISA kit while AiD™ anti-HIV 1+2 ELISA kit (Wantai, Beijing, China) was used as the 3rd generation ELISA kit. All assays were performed under strict biosafety conditions and according to the manufacturer's recommendation. Furthermore, detection of HIV antibodies were further confirmed using Alere™ Determine™ HIV-1/2 Rapid Antibody Test kit (Alere, Chiba, Japan).

Clinical chemistry assay for serum creatinine

Plasma samples were analyzed on a Roche cobas® C11 blood chemistry analyzer (Roche Diagnostics, Indianapolis). Each sample was analyzed to determine the level of serum

creatinine according to the manufacturer's instruction. Normal reference ranges for plasma creatinine is 62-133 $\mu\text{mol/L}$.

HIV-1 RNA Viral load testing

Serum samples collected at baseline were tested for Plasma HIV-1 viral load (copies/ml) using the COBAS[®] Ampliprep/COBAS TaqMan96[®] HIV-1 Test, v2:0 (Roche Molecular Diagnostics, Branchburg, NJ, USA) according to manufacturer's instruction or by an in house real-time PCR protocol.

PCR amplification and sequencing

Total DNA was extracted from whole blood samples collected at each visit using guanidium thiocyanate in house protocol. A fragment of the gag-pol region (900 base pairs) of the virus was amplified using previously published primers and cycling conditions (Gall 2012)[45] with slight modifications . Briefly, PCR was performed using a SuperScript III One-step RT-PCR system with platinum TaqDNA High fidelity polymerase (Jena Bioscience). Each 25 μl reaction mixture contained 12.5 μl reaction mix (2x), 4.5 μl RNase-free water, 1 μl each of each primer (20pmol/ μl), 1 μl Superscript III RT/Platinum Taq High Fidelity mix, and 5 μl of template DNA. Pan-HIV-1_1R (CCT CCA ATT CCY CCT ATC ATT TT) and Pan-HIV-1_2F (GGG AAG TGA YAT AGC WGG AAC) were used. Cycling conditions were 94°C for 5min; 35 cycles of 94°C for 15s, 58°C for 30s, and 68°C for 1min 30s; and finally, 68°C for 10mins. Positive HIV samples that was undetectable using the above stated primers were retested using another set of GAG primers for nested PCR as described previously[46]. Positive PCR reactions were shipped on ice to Macrogen, South Korea for Big Dye sequencing using the same amplification primers (Pan-HIV-1_1R and Pan-HIV-1_2F; or G60 and G25).

Detection of HIV-1 subtypes and phylogenetic analysis

The sequences were cleaned and edited using Chromas and Bioedit softwares. Subtyping was performed using a combination of four subtyping tools: The Rega HIV-1 Subtyping Tool, version 3.0 (<http://dbpartners.stanford.edu/RegaSubtyping/>), Comet, version 2.2 (<http://comet.retrovirology.lu>), National Center for Biotechnology Information, Bethesda, MD (<http://www.ncbi.nlm.nih.gov/Blast.cgi>) and jpHMM: Improving the reliability of recombination prediction in HIV-1 (http://jphmm.gobics.de/submission_hiv). The first three

tools were used simultaneously while jpHMM was used to resolve discordant subtypes. Phylogenetic analyses were performed using MEGA software version 10. Alignment of sequences were performed using MAFFTS online software. Genetic distances were inferred using the Tamura-Nei model and a phylogenetic tree was generated using the maximum likelihood method. The robustness of the tree was evaluated with 1000 bootstrap replicates. All consensus nucleotide sequences obtained in this study were submitted to GenBank Database and assigned accession numbers MN943617-635.

Non-synonymous substitutions in Cytotoxic T Lymphocytes (CTL) epitopes within the Major Hydrophobic Region (MHR) of HIV-1 GAG gene isolated from 10 early infected individuals

Reference GAG sequences for subtypes G, A and CRF02-AG were downloaded from the Los Alamos National Laboratory HIV Sequence Database (<https://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>). Deduced amino acid (aa) sequences were translated for both reference and sample sequences with the standard genetic code using Bioedit software. CTL epitope corresponding to the three highly conserved sites (HCS) of the MHR namely KAFSPEVIPMFSAALSEGATPQD, DTINEEAAEWDR and TSTLQEQIR[2,6,19] were used for comparison and identification of amino acid substitution. HIV-1 GAG sequences identified as subtypes A, G and CRF02_AG in this study were aligned with Reference A (GenBank accession numbers DQ676872; AB253421 and AB253429), G (GenBank accession numbers AF084936; AF061641; U88826 and AY612637) and CRF02_AG (GenBank accession numbers L39106 and DQ168578) sequences respectively.

Non-synonymous substitutions in Cytotoxic T Lymphocytes (CTL) epitopes outside the Major Hydrophobic Region (MHR) of HIV-1 GAG gene

The Virus Epidemiology Signature Patterns Analysis (<http://www.hiv.lanl.gov/content/sequence/VESPA/vespa.html>) program was used to identify variations in other sites of the HIV-1 GAG sequence outside the HCS corresponding to CTL epitope regions. Deduced amino acid (aa) sequences were translated for both reference and sample sequences with the standard genetic code using Bioedit. HIV-1 GAG sequences identified as subtypes A, G and CRF02_AG in this study were aligned with Reference A (GenBank accession numbers DQ676872; AB253421 and AB253429), G (GenBank accession numbers AF084936; AF061641; U88826 and AY612637) and CRF02_AG (GenBank accession numbers L39106

and DQ168578) sequences respectively. Only amino acid replacements in which there was 100% non-synonymous substitution between reference sequences and sample sequences were considered. Percentage substitution rates were calculated by finding the ratio of number of substitutions to the total possible substitution sites. These mutations were compared with the Los Alamos National Laboratory HIV Immunology Database for CTL/CD8+ Epitope Variants and Escape Mutations (https://www.hiv.lanl.gov/content/immunology/variants/ctl_variant.html).

Ethical approval

This research was conducted in accordance with the declaration of Helsinki. Experiments were conducted with the understanding and the consent of each participant. Ethical approvals for this research were obtained from the University of Ibadan/ University College Hospital (UI/UCH) Research and Ethics Committee (UI/EC/15/0076) and the Oyo State Ministry of Health Committee on Human Research (AD13/479/951). All results were delinked from patient identifiers and anonymized.

Eligibility/ Exclusion criteria

Only individuals between 18 and 65 years of age were included in the study. Individuals who already knew their HIV status were excluded from the study.

Data management and statistical analysis

Statistical analyses were performed using SPSS version 20. Data are expressed as means \pm standard deviations. Statistical significance was estimated using the Kruskal-Wallis test, with SPSS package version 12.0. Statistical significance was defined as P values ≤ 0.05 .

Results

Participants' characteristics

As described previously, twenty three individuals were identified to be at the early stages of HIV-1 infection. Figure 1 shows the phylogeny of HIV-1 subtypes. Out of the 10 early infected persons studied, 5 were infected with Subtype A, 3 with subtype G and the rest CRF02-AG. Five individuals at the early stages of HIV-1 infection were followed up until detection of antibodies. These individuals were identified to be at the early stages of infection

at different periods in 2017. One individual in October, July and May each and two in August. Three of the individuals were males and were voluntary blood donors. The remaining two females were identified when referred for malaria antigen test. Three of the individuals were infected with HIV-1 subtype G while the other two were infected with subtype A (Table 1).

Table 2 shows the analysis of the samples collected from these individuals from baseline until after detection of antibodies, two years after. In four individuals, blood samples were collected at three time points, twice during the early stages of infection and once after detection of antibodies. Samples were collected four times in the fifth individual, thrice during the early stages of infection and once after detection of antibodies. HIV-1 GAG sequences of these infected individuals were determined at every time point of blood collection.

Phylogenetic analysis

Figure 1 shows the estimated phylogeny of HIV-1 subtypes with respect to reference sequences available in the HIV Los Alamos National HIV Sequence database. As shown in the Figure, Subtypes A, G and CRF02_AG were identified with green, blue and pink symbols respectively. HIV-1 subtypes A identified in this study were closely related to Ref A1 DQ676872 as well as subtypes AF457075, KF716486 and AY521631 from Kenya, Uganda and Senegal respectively. Those identified as subtypes G and CRF02-AG were closely related to the Nigerian subtypes DQ168573 and Ref.02 AG IBNG. L39106 respectively.

There is high rate of substitutions in HCS of CTL epitope regions of Subtype A HIV-1 GAG gene during early infection

In this study we compared intra and inter variations among 10 HIV-1 GAG sequences isolated from persons at the early stages of infection. These sequences were grouped by subtypes and analyzed alongside reference sequences. As shown in Table 3, variations occurred mostly in HIV-1 Subtype A at the CTL epitope region of 243-251aa. In fact, the conserved epitope of TSTLQEQIR was not found in both the reference subtypes and those from early infected individuals. HIV-1 Subtype A also had the highest variations (50%) for HCS corresponding to CTL epitope region (203-214aa). HCS for CTL epitope region 162-183aa was the most conserved among the subtypes, yet a substitution rate of 2.7% was observed for HIV-1 subtype A. CTL escape region (162-183)

KAFSPEVIPMFSALSEGATPQD had lowest frequency of mutations. However, two samples (MN943615; MN943616) had two mutations-K162R and A163G. MN943615 had these two mutations while MN943616 had A163G only. These two mutations have been previously associated with immune escape (non-susceptible forms) on the KAFSPEVIPMFSALSEGATPQD epitope[47]. D203E substitution found in a sample (MN943625) in subtype G seems to be a reversion[41].

Sequences isolated after detection of antibodies are associated with a higher rate of substitutions in HCS of CTL epitope regions

Substitutions in the HCS of subtype B HIV-1 GAG gene have been shown to come with a fitness cost. These substitutions have shown to be deleterious to HIV-1 strains eventual survival and transmission. However, very limited information exist on substitutions associated with immune epitopes during early non-subtype B HIV-1 infection. We followed five individuals from early stage of infection till detection of antibodies. Three of these individuals were infected with subtype G while the remaining two subtype A (see Table 4). There were no substitutions in CTL epitope regions for EHIV016 while EHIV022 had single aa substitutions before and after antibodies detection.

The mutation identified after detection of antibodies –E203D- in EHIV022 have been previously associated with escape[41]. EHIV012 had single aa substitutions each after detection of antibodies in the three CTL epitope regions studied. These substitutions were not previously observed. For HIV-1 subtype A samples, EHIV023 had very high substitution rates in CTL epitope region of 203-214aa (31.2%), although there were reversions by the time antibodies were detected. Amino acid substitutions were also observed in CTL epitope region spanning 243-251aa (8.3%). Significant differences in substitution rates before and after detection of antibodies across two of the three CTL epitope regions was observed for EHIV023 sample (see Table 4). Substitutions were more associated with the HIV-1 GAG gene sequenced after detection of antibodies for EHIV013 across the three CTL epitope regions studied. The two amino acid substitutions, A166G and L216V, observed in EHIV013 occurred after antibody detection and were not previously identified.

Amino acid signature patterns in variable sites of HIV-1 GAG genes

Assessment of signature patterns in this study were limited to non-synonymous substitutions in which there were 100% aa replacements. In Table 5, these substitutions were compared

between reference sequence and sequences isolated from individuals at the early stages of infection while in Tables 6 and 7, the substitutions were compared within sequences isolated per sample spanning early HIV infection till detection of antibodies. As shown in Table 5, non-synonymous substitutions were mostly observed in subtype G. However, a substitution, E105K, observed in CRF02_AG had been previously identified as a variant that did not recognize the HXB2 epitope[48]. Subtype A had a frame shift mutation at aa85-88. There were no substitutions in sequences isolated from sample EHIV016 while EHIV 012 had the highest number of substitution within Subtype G sequences (see Table 6). EHIV022 and EHIV 012 had substitutions within sequences at aa positions of 106 and 110-113.

Lysine is the commonest aa used at position 106 (3/5) while a single substitution were observed for glutamate and asparagine. KSQ was the commonest aa usage in positions 110-113 (3/5), other aa in these positions were RRE (1/5) and KSK (1/5). In Table 7, non-synonymous substitutions associated with HIV-1 subtype A are described. EHIV013 had the highest number of substitutions spanning aa region 15 to 267. This is the largest number of substitutions found in this study. Samples EHIV023 and EHIV013 had substitutions within sequences at aa positions 215, 218 and 228. In position 215, Leucine was the highest aa used (4/7) while glutamine, asparagine and valine were present in one sequence each. Valine (4/7) and leucine (3/7) were the only aa used in position 218. Aspartate (3/7) was the highest aa used in position 228 while methionine (2/7), proline (1/7) and isoleucine (1/7) were also present in some sequences.

Non synonymous substitutions associated with immune escape variants are more within epitopes outside the MHR

Although substitutions E105K (in CRF02AG), E203D (in subtype G) as well as K162R and A163G (in subtype A) are associated with immune escape and were observed as occurring within the MHR in this study, more substitutions previously associated with immune escape were found in regions outside the MHR. Out of the 5 individuals followed up, 2 had substitutions in GAG gene sites outside the MHR associated with immune escape strains (see Figures 2 and 3). Most mutations associated with previously described immune escape strains were identified after detection of antibodies in this study. Fourteen mutations in 16 HIV-1 GAG sites outside the MHR were identified before antibodies detection while 21 mutations in 23 HIV-1 GAG sites outside the MHR were identified after detection of antibodies. As shown in Figures 2 and 3, EHIV012 had three types of mutations before (L31M, L101V and

S172T) and after (V82I, Y86W and F172S) detection of antibodies. H28K, M30R, A224P and A248G mutations were identified before and after detection of antibodies while V82I and Y86W mutations were identified only after detection of antibodies.

Serum creatinine concentration during longitudinal follow up

As shown in Figure 4, there was a significant difference ($P < 0.0044$) in serum creatinine concentrations between early HIV-1 infection and detection of antibodies for the three individuals studied among the five persons followed up. EHIV 023 had lowest serum creatinine concentration at baseline (0.9mg/dl) and after antibody detection (0.8mg/dl) while EHIV 022 had the highest serum concentration of creatinine at baseline (1.1mg/dl) and after antibodies detection (1.0mg/dl). Sample EHIV016 had no changes in serum creatinine concentration from baseline until antibodies detection (1.0mg/dl).

Discussion

The results of this study show that diverse HIV-1 subtypes circulate in Nigeria, a West African country as subtypes A, G and CRF02_AG were identified. In our study, we identified 3 non synonymous substitutions within the MHR of HIV-1 GAG genes isolated from 10 early infected Nigerians. One substitution was however observed outside the MHR epitopes. Three (E105K, K162R and A163G) of these substitutions have been previously associated with immune escape[47,48]. These substitutions were associated with subtypes A and CRF02_AG. Most substitutions in subtype G within the MHR were associated with detection of antibodies while those of subtype A were associated with early HIV infection. Recent studies have showed that apart from immune escape mutations, other factors such as subtype differences and NEF mediated evasion of antibodies may account for replication and increased viremia during HIV-1 infection[49,50]. Fourteen mutations outside the MHR previously associated with immune escape were observed before detection of antibodies while 19 were observed after detection of antibodies in this study. This is a far cry from the four observed within the MHR. This finding is in consonance with previous reports that observed very limited non-synonymous substitutions in CTL epitopes within the MHR[7,28,51].

Gounder *et al.*, 2015[52] however suggested that CTL escape immune variants observed during the early stages of HIV infection may be due to the transmitted virus as well as stochastic processes and not necessarily CD8 immune pressures while Roberts *et al.*, 2015

reported that the average time before targeted epitope evolved an escape mutation was longer than two years[53]. We cannot fully corroborate this claim in our study because of our low sample size. However, since latent reservoir strains are established during the early stages of infection and are known to compose majorly of CTL immune escape strains[54,55], immune escape mutations may be incorporated more into the latent reservoirs irrespective of the mechanism with which they were generated. These latent reservoirs may require a broad CTL response for clearance as previously alluded[54]. Also, previous studies have showed that although escape mutants come with a fitness cost, this may be compensated for by HLA independent polymorphisms as well as other host-virus interaction factors[42]. Observations of CTL escape mutants even in the presence of antibodies in this study suggests that immune pressures by other immune cells other than CTL may aid the generation of CTL immune escape mutants. Mapping HIV immune epitopes in other regions of the genome will further clarify this hypothesis[27,56].

Universal HIV-1 vaccines are supposed to be broadly effective against all HIV-1 clades. However, immune escape strains encoding CTL epitopes outside the MHR may reduce the sensitivities of these vaccines. Also therapeutic HIV-1 vaccines are intended to be used after cART stoppage particularly against latent reservoir strains. However, as shown in this study, immune escape strains generated during the early stages of infection which are likely major constituents of the latent reservoir may lead to therapeutic vaccine failures[2,6,7,40]. Since immune escape occurs during the early stages of infection, it is imperative that cART commence early in order to reduce the reservoir size as well as the incorporation of immune escape variants[57,58]. Other studies have alluded to the fact that post treatment control may be possible if treatment begins early since blips observed after cART stoppage are largely due to immune escape variants incorporated into reservoir cells[59]. However, rapid and high magnitude CTL responses observed during the early stages of infection[11] may be affected by early treatment[58,60,61] which can in turn impair subsequent responses during cART stoppage in post treatment control trials. We have showed in this study that immune escape variants may be from those arising from CTL epitopes outside the MHR. However, functional and molecular studies on the nature and characteristics of HIV-1 strains in latent reservoirs need to be carried out to further ascertain our claims.

As observed previously in other studies, CTL epitope KAFSPEVIPMF was the most conserved[62,63]. This epitope has been associated with very low frequencies of CTL

response selective pressures and has been a choice candidate for many T cell based HIV vaccines[2]. However, two substitutions previously associated with immune escape, K162R and A163G were observed in this epitope for Subtype A during the early stages of infection. On the other hand, CTL epitope DTINEEAAEWDR was associated with more substitutions, although most of these mutations reverted to wild type by the time antibodies were detectable. This phenomenon was observable in both subtypes A and G. It seems that although more mutations were observed during the early stages of HIV infection, reversions of these mutations occurred later on in infection. Previous studies have also associated reversions of CTL epitopes with the early stages of HIV infection[48]. A non-synonymous substitution, E203D, was observed in CTL epitope DTINEEAAEWDR. Selection of this epitope for immune escape strains have been previously described[41]. CTL epitope TSTLQEQIR was conserved for subtype G. However, this epitope was not prominent for subtype A sequences. Previous studies have documented that TSTLQEQIR may not be presented early and is associated with elite controllers[64,65]. Inclusion of this epitope in vaccines should induce strong CTL responses associated with viral load control and NEF gene down regulation. However, as observed in this study, the epitope is expressed by rare HLAs.

Several previously recognized immune escape substitutions were observed in this study. Majority of these substitutions emanated from epitopes outside the MHR. Although functional assays were not carried out to associate these epitopes with HLA polymorphisms, to the best of our knowledge, this is the first longitudinal study from West Africa on the natural history of HIV-1 especially as it relates to the kinetics of HIV-1 GAG CTL epitopes. We have also showed that numerous non-synonymous substitutions associated with CTL epitopes outside the MHR occur during the early stages of HIV-1 infection among HIV-1 subtypes and recombinant forms circulating in West Africa. It is important to state that these substitutions were identified from HIV-1 DNA sequences as against plasma RNA used in a previous similar study[52]. Proviral sequences have previously been associated with rare mutations on CTL epitopes[60,62]. These substitutions may have to be taken into consideration in future universal and therapeutic vaccines designs for HIV-1 strains circulating in West African countries[4,66]. More so that recent studies have showed that the major role of poorly recognized CTL epitopes is viral escape[67].

While some non-synonymous substitutions were observed before and after antibodies detection (H28K, M30R, A224P and A248G), some, such as V82I and Y86W were consistently identified after antibodies detection. V82I has been previously identified in studies among infected individuals[68,69]. This substitution among other things were associated with emergence of higher viral loads while Y86W was previously associated with HIV-1 clade B and E[70]. We as well as others have reported the high levels of creatinine among HIV infected Africans and African-Americans during the early stages of infection[43,44,71,72], findings from this present study show that these high levels may be linked to activation of HLA receptors on CD8 T cells during this stage of infection. Down regulation of this activation by the time antibodies appear and peak viremia subsides may correlate with reduction in creatinine levels.

In summary, we have showed that there is high genetic diversity of HIV-1 strains in Nigeria. Also, very high frequencies of non-synonymous substitutions occur in the HIV-1 GAG gene during the early stages of infection up until when antibodies become detectable. These substitutions include previous mapped CTL epitope immune escape mutants. We also observed different rates of HIV-1 GAG substitutions in the different subtypes especially during the early stages of infection. CTL immune pressure likely leave different footprints and signature patterns on HIV-1 GAG epitopes within and outside the MHR. This information is crucial for future vaccine designs for use in the West African region be it preventive or therapeutic.

Acknowledgements

BO, OD and GO conceptualized and designed the study. BO performed experiments, analysed and interpreted the data as well as wrote the first draft of manuscript. OD and GO supervised the work and reviewed the manuscript.

References

- 1 UNAIDS. Global HIV & AIDS statistics — 2019 fact sheet | UNAIDS. 2019.<https://www.unaids.org/en/resources/fact-sheet> (accessed 1 Mar2020).
- 2 Hanke T. **Aiming for protective T-cell responses: a focus on the first generation conserved-region HIV_{consv} vaccines in preventive and therapeutic clinical trials.**

- Expert Rev Vaccines* 2019; **18**:1029–1041.
- 3 Herout S, Mandorfer M, Breitenecker F, Reiberger T, Grabmeier-Pfistershammer K, Rieger A, *et al.* **Impact of Early Initiation of Antiretroviral Therapy in Patients with Acute HIV Infection in Vienna, Austria.** *PLoS One* 2016; **11**:e0152910.
 - 4 Ndung'u T, McCune JM, Deeks SG. **Why and where an HIV cure is needed and how it might be achieved.** *Nature* 2019; **576**:397–405.
 - 5 Papuchon J, Pinson P, Lazaro E, Reigadas S, Guidicelli G, Taupin JL, *et al.* **Resistance Mutations and CTL Epitopes in Archived HIV-1 DNA of Patients on Antiviral Treatment: Toward a New Concept of Vaccine.** *PLoS One* 2013; **8**.
doi:10.1371/journal.pone.0069029
 - 6 Lé tourneau S, Im E-J, Mashishi T, Brereton C, Bridgeman A, Yang H, *et al.* **Design and Pre-Clinical Evaluation of a Universal HIV-1 Vaccine.**
doi:10.1371/journal.pone.0000984
 - 7 Ondondo B, Murakoshi H, Clutton G, Abdul-Jawad S, Wee EGT, Gatanaga H, *et al.* **Novel conserved-region T-cell mosaic vaccine with high global HIV-1 coverage is recognized by protective responses in untreated infection.** *Mol Ther* 2016; **24**:832–842.
 - 8 Bekker LG, Tatoud R, Dabis F, Feinberg M, Kaleebu P, Marovich M, *et al.* The complex challenges of HIV vaccine development require renewed and expanded global commitment. *Lancet.* 2020; **395**:384–388.
 - 9 Burton DR, Desrosiers RC, Doms RW, Koff WC, Kwong PD, Moore JP, *et al.* HIV vaccine design and the neutralizing antibody problem. *Nat. Immunol.* 2004; **5**:233–236.
 - 10 Muema DM, Macharia GN, Olusola BA, Hassan AS, Fegan GW, Berkley JA, *et al.* **Proportions of circulating follicular helper T cells are reduced and correlate with memory B cells in HIV-infected children.** *PLoS One* 2017; **12**.
doi:10.1371/journal.pone.0175570
 - 11 Ndhlovu ZM, Kanya P, Mewalal N, Kløverpris HN, Nkosi T, Pretorius K, *et al.* **Magnitude and Kinetics of CD8+ T Cell Activation during Hyperacute HIV Infection Impact Viral Set Point.** *Immunity* 2015; **43**:591–604.

- 12 Leitman EM, Thobakgale CF, Adland E, Ansari MA, Raghwani J, Prendergast AJ, *et al.* **Role of HIV-specific CD8 + T cells in pediatric HIV cure strategies after widespread early viral escape.** 2016; :3239–3262.
- 13 Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. **Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection.** *J Virol* 1994; **68**:6103–6110.
- 14 Goonetilleke N, Liu MKP, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Ghanusov V V., *et al.* **The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection.** *J Exp Med* 2009; **206**:1253–1272.
- 15 Goulder PJR, Watkins DI. HIV and SIV CTL escape: Implications for vaccine design. *Nat. Rev. Immunol.* 2004; **4**:630–640.
- 16 Ogg GS, Jin X, Bonhoeffer S, Dunbar PR, Nowak MA, Monard S, *et al.* **Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA.** *Science (80-)* 1998; **279**:2103–2106.
- 17 Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, *et al.* **Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome.** *J Virol* 1994; **68**:4650–4655.
- 18 Kosakovsky Pond SL, Poon AFY, Zárate S, Smith DM, Little SJ, Pillai SK, *et al.* **Estimating selection pressures on HIV-1 using phylogenetic likelihood models.** *Stat Med* 2008; **27**:4779–4789.
- 19 Yang OO, Ali A, Kasahara N, Faure-Kumar E, Bae JY, Picker LJ, *et al.* **Short Conserved Sequences of HIV-1 Are Highly Immunogenic and Shift Immunodominance.** *J Virol* 2015; **89**:1195–1204.
- 20 Tenzer S, Wee E, Burgevin A, Stewart-Jones G, Friis L, Lamberth K, *et al.* **Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance.** *Nat Immunol* 2009; **10**:636–646.
- 21 Bennett MS, Joseph A, Ng HL, Goldstein H, Yang OO. **Fine-tuning of T-cell receptor avidity to increase HIV epitope variant recognition by cytotoxic T lymphocytes.** *AIDS* 2010; **24**:2619–2628.

- 22 Honeyborne I, Codoñer FM, Leslie A, Tudor-Williams G, Luzzi G, Ndung'u T, *et al.* **HLA-Cw*03-restricted CD8+ T-cell responses targeting the HIV-1 gag major homology region drive virus immune escape and fitness constraints compensated for by intracodon variation.** *J Virol* 2010; **84**:11279–88.
- 23 Mori M, Wichukchinda N, Miyahara R, Rojanawiwat A, Pathipvanich P, Maekawa T, *et al.* **HLA-B*35: 05 is a protective allele with a unique structure among HIV-1 CRF01-AE-infected Thais, in whom the B*57 frequency is low.** *AIDS* 2014; **28**:959–967.
- 24 Brumme ZL, Brumme CJ, Carlson J, Streeck H, John M, Eichbaum Q, *et al.* **Marked Epitope-and Allele-Specific Differences in Rates of Mutation in Human Immunodeficiency Type 1 (HIV-1) Gag, Pol, and Nef Cytotoxic T-Lymphocyte Epitopes in Acute/Early HIV-1 Infection.** *J Virol* 2008; **82**:9216–9227.
- 25 Gillespie GMA, Kaul R, Dong T, Yang HB, Rostron T, Bwayo J, *et al.* **Cross-reactive cytotoxic T lymphocytes against a HIV-1 p24 epitope in slow progressors with B*57.** *AIDS* 2002; **16**:961–972.
- 26 Balla-Jhagjhoorsingh SS, Koopman G, Mooij P, Haaksma TG, Teeuwssen VJ, Bontrop RE, *et al.* **Conserved CTL epitopes shared between HIV-infected human long-term survivors and chimpanzees.** *J Immunol* 1999; **162**:2308–14.
- 27 Matthews PC, Koyanagi M, Kløverpris HN, Harndahl M, Stryhn A, Akahoshi T, *et al.* **Differential clade-specific HLA-B*3501 association with HIV-1 disease outcome is linked to immunogenicity of a single Gag epitope.** *J Virol* 2012; **86**:12643–54.
- 28 Lazaro E, Kadie C, Stamegna P, Zhang SC, Gourdain P, Lai NY, *et al.* **Variable HIV peptide stability in human cytosol is critical to epitope presentation and immune escape.** *J Clin Invest* 2011; **121**:2480–2492.
- 29 Kløverpris HN, Adland E, Koyanagi M, Stryhn A, Harndahl M, Matthews PC, *et al.* **HIV subtype influences HLA-B*07:02-associated HIV disease outcome.** *AIDS Res Hum Retroviruses* 2014; **30**:468–75.
- 30 Ndhlovu ZM, Kanya P, Mewalal N, Kløverpris HN, Nkosi T, Pretorius K, *et al.* **Magnitude and Kinetics of CD8+ T Cell Activation during Hyperacute HIV Infection Impact Viral Set Point.** *Immunity* 2015; **43**:591–604.

- 31 Koofhethile CK, Ndhlovu ZM, Thobakgale-Tshabalala C, Prado JG, Ismail N, Mncube Z, *et al.* **CD8+ T Cell Breadth and Ex Vivo Virus Inhibition Capacity Distinguish between Viremic Controllers with and without Protective HLA Class I Alleles.** *J Virol* 2016; **90**:6818–6831.
- 32 Radebe M, Gounder K, Mokgoro M, Ndhlovu ZM, Mncube Z, Mkhize L, *et al.* **Broad and persistent Gag-specific CD8+ T-cell responses are associated with viral control but rarely drive viral escape during primary HIV-1 infection.** *AIDS* 2015; **29**:23–33.
- 33 Carlson JM, Listgarten J, Pfeifer N, Tan V, Kadie C, Walker BD, *et al.* **Widespread impact of HLA restriction on immune control and escape pathways of HIV-1.** *J Virol* 2012; **86**:5230–43.
- 34 Tanaka M, Robinson BA, Chutiraka K, Geary CD, Reed JC, Lingappa JR. **Mutations of Conserved Residues in the Major Homology Region Arrest Assembling HIV-1 Gag as a Membrane-Targeted Intermediate Containing Genomic RNA and Cellular Proteins.** *J Virol* 2016; **90**:1944–1963.
- 35 Reicin AS, Paik S, Berkowitz RD, Luban J, Lowy I, Goff SP. **Linker insertion mutations in the human immunodeficiency virus type 1 gag gene: effects on virion particle assembly, release, and infectivity.** *J Virol* 1995; **69**:642–650.
- 36 Kløverpris HN, McGregor R, McLaren JE, Ladell K, Harndahl M, Stryhn A, *et al.* **CD8 + TCR Bias and Immunodominance in HIV-1 Infection .** *J Immunol* 2015; **194**:5329–5345.
- 37 Liu Y, Rao U, McClure J, Konopa P, Manochewa S, Kim M, *et al.* **Impact of Mutations in Highly Conserved Amino Acids of the HIV-1 Gag-p24 and Env-gp120 Proteins on Viral Replication in Different Genetic Backgrounds.** *PLoS One* 2014; **9**:e94240.
- 38 Troyer RM, McNevin J, Liu Y, Zhang SC, Krizan RW, Abraha A, *et al.* **Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response.** *PLoS Pathog* 2009; **5**:e1000365.
- 39 Martinez-Picado J, Prado JG, Fry EE, Pfafferott K, Leslie A, Chetty S, *et al.* **Fitness cost of escape mutations in p24 Gag in association with control of human**

- immunodeficiency virus type 1.** *J Virol* 2006; **80**:3617–23.
- 40 Daucher M, Price DA, Brenchley JM, Lamoreaux L, Metcalf JA, Rehm C, *et al.*
Virological outcome after structured interruption of antiretroviral therapy for human immunodeficiency virus infection is associated with the functional profile of virus-specific CD8+ T cells. *J Virol* 2008; **82**:4102–14.
- 41 Murakoshi H, Kuse N, Akahoshi T, Zhang Y, Chikata T, Borghan MA, *et al.* **Broad Recognition of Circulating HIV-1 by HIV-1-Specific Cytotoxic T-Lymphocytes with Strong Ability to Suppress HIV-1 Replication.** *J Virol* 2019; **93**.
doi:10.1128/JVI.01480-18
- 42 Crawford H, Matthews PC, Schaefer M, Carlson JM, Leslie A, Kilembe W, *et al.* **The Hypervariable HIV-1 Capsid Protein Residues Comprise HLA-Driven CD8+ T-Cell Escape Mutations and Covarying HLA-Independent Polymorphisms.** *J Virol* 2011; **85**:1384–1390.
- 43 Olusola BA, Olaleye DO, Odaibo GN. **Early HIV infection among persons referred for malaria parasite testing in Nigeria.** *Arch Virol* 2017; **163**:439–445.
- 44 Olusola BA, Kabelitz D, Olaleye DO, Odaibo GN. **Early HIV infection is associated with reduced proportions of gamma delta T subsets as well as high creatinine and urea levels.** *Scand J Immunol* 2020; :1–13.
- 45 Gall A, Ferns B, Morris C, Watson S, Cotten M, Robinson M, *et al.* **Universal Amplification, Next-Generation Sequencing, and Assembly of HIV-1 Genomes.** Published Online First: 2012. doi:10.1128/JCM.01516-12
- 46 Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, Ilunga W, *et al.*
Unprecedented Degree of Human Immunodeficiency Virus Type 1 (HIV-1) Group M Genetic Diversity in the Democratic Republic of Congo Suggests that the HIV-1 Pandemic Originated in Central Africa. *J Virol* 2000; **74**:10498–10507.
- 47 Currier JR, Harris ME, Cox JH, McCutchan FE, Birx DL, Maayan S, *et al.*
Immunodominance and cross-reactivity of B5703-restricted CD8 T lymphocytes from HIV type 1 subtype C-infected Ethiopians. *AIDS Res Hum Retroviruses* 2005; **21**:239–245.
- 48 Li B, Gladden AD, Altfeld M, Kaldor JM, Cooper DA, Kelleher AD, *et al.* **Rapid**

- Reversion of Sequence Polymorphisms Dominates Early Human Immunodeficiency Virus Type 1 Evolution.** *J Virol* 2007; **81**:193–201.
- 49 Buckheit RW, Allen TG, Alme A, Salgado M, O’Connell KA, Huculak S, *et al.* **Host factors dictate control of viral replication in two HIV-1 controller/chronic progressor transmission pairs.** *Nat Commun* 2012; **3**:716.
- 50 Omondi FH, Chandrarathna S, Mujib S, Brumme CJ, Jin SW, Sudderuddin H, *et al.* **HIV Subtype and Nef-Mediated Immune Evasion Function Correlate with Viral Reservoir Size in Early-Treated Individuals Downloaded from.** *jvi.asm.org 1 J Virol* 2019; **93**:1832–1850.
- 51 Amicosante M, Gioia C, Montesano C, Casetti R, Topino S, D’Offizi G, *et al.* **Computer-based design of an HLA-haplotype and HIV-clade independent cytotoxic T-lymphocyte assay for monitoring HIV-specific immunity.** *Mol Med* 2002; **8**:798–807.
- 52 Gounder K, Padayachi N, Mann JK, Radebe M, Mokgoro M, Van Der Stok M, *et al.* **High frequency of transmitted HIV-1 Gag HLA class I-driven immune escape variants but minimal immune selection over the first year of clade C infection.** *PLoS One* 2015; **10**:e0119886.
- 53 Roberts HE, Hurst J, Robinson N, Brown H, Flanagan P, Vass L, *et al.* **Structured Observations Reveal Slow HIV-1 CTL Escape.** *PLoS Genet* 2015; **11**:1–22.
- 54 Deng K, Perteau M, Rongvaux A, Wang L, Durand CM, Ghiaur G, *et al.* **Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations.** *Nature* 2014; **517**:381–385.
- 55 Leitman EM, Thobakgale CF, Adland E, Ansari MA, Raghwani J, Prendergast AJ, *et al.* **Role of HIV-specific CD8+ T cells in pediatric HIV cure strategies after widespread early viral escape.** *J Exp Med* 2017; **214**:3239–3261.
- 56 Adland E, Carlson JM, Paioni P, Kløverpris H, Shapiro R, Ogburn A, *et al.* **Nef-specific CD8+ T cell responses contribute to HIV-1 immune control.** *PLoS One* 2013; **8**:e73117.
- 57 Brockman MA, Jones RB, Brumme ZL. **Challenges and opportunities for T-cell-mediated strategies to eliminate HIV reservoirs.** *Front Immunol* 2015; **6**:506.

- 58 Takata H, Buranapraditkun S, Kessing C, Fletcher JLK, Muir R, Tardif V, *et al.* **Delayed differentiation of potent effector CD8+ T cells reducing viremia and reservoir seeding in acute HIV infection.** *Sci Transl Med* 2017; **9**:eaag1809–eaag1809.
- 59 Conway JM, Perelson AS. **Post-treatment control of HIV infection.** *PNAS* 2014; **112**:5467–5472.
- 60 Lee GQ, Reddy K, Einkauf KB, Gounder K, Chevalier JM, Dong KL, *et al.* **HIV-1 DNA sequence diversity and evolution during acute subtype C infection.** *Nat Commun* 2019; **10**:2737.
- 61 Ndhlovu ZM, Kazer SW, Nkosi T, Ogunshola F, Muema DM, Anmole G, *et al.* **Augmentation of HIV-specific T cell function by immediate treatment of hyperacute HIV-1 infection.** *Sci Transl Med* 2019; **11**.
doi:10.1126/scitranslmed.aau0528
- 62 Gama Caetano D, Heloise Côrtes F, Bello G, Lopes Maia Teixeira S, Hoagland B, Grinsztejn B, *et al.* **Next-generation sequencing analyses of the emergence and maintenance of mutations in CTL epitopes in HIV controllers with differential viremia control.** 2018; **15**:62.
- 63 Garcia-knight MA, Slyker J, Payne BL, Pond SLK, Silva TI De, Chohan B, *et al.* **Viral Evolution and Cytotoxic T Cell Restricted Selection in Acute Infant HIV-1 Infection.** *Nat Publ Gr* 2016; :1–8.
- 64 Balamurugan A, Ali A, Boucau J, Le Gall S, Ng HL, Yang OO. **HIV-1 Gag Cytotoxic T Lymphocyte Epitopes Vary in Presentation Kinetics Relative to HLA Class I Downregulation.** Published Online First: 2013. doi:10.1128/JVI.01040-13
- 65 O’Connell KA, Hegarty RW, Siliciano RF, Blankson JN. **Viral suppression of multiple escape mutants by de novo CD8+T cell responses in a human immunodeficiency virus-1 Infected elite suppressor.** *Retrovirology* 2011; **8**:63.
- 66 Shu J, Shen W, Liu H, Zhou Y, Li J, Zhuang Y, *et al.* **The immunologic dominance of an epitope within a rationally designed poly-epitope vaccine is influenced by multiple factors.** *Vaccine* Published Online First: 29 February 2020.
doi:10.1016/j.vaccine.2020.02.069

- 67 Grossman MA, Hofmann C, Ng HL, Yang OO. **Suboptimal stimulation by weak agonist epitope variants does not drive dysfunction of HIV-1-specific cytotoxic T lymphocyte clones.** *AIDS* 2019; **33**:1565–1574.
- 68 Karlsson AC, Iversen AKN, Chapman JM, De Oliveira T, Spotts G, Mcmichael AJ, *et al.* **Sequential Broadening of CTL Responses in Early HIV-1 Infection Is Associated with Viral Escape.** doi:10.1371/journal.pone.0000225
- 69 Arcia D, Ochoa R, Hernández JC, Cristiam M, Díaz FJ, Velilla PA, *et al.* **Potential immune escape mutations under inferred selection pressure in HIV-1 strains circulating in Medellin, Colombia.** *Infect Genet Evol* 2018; **18**:#pagerange#.
- 70 Fukada K, Tomiyama H, Wasi C, Matsuda T, Kusagawa S, Sato H, *et al.* **Cytotoxic T-cell recognition of HIV-1 cross-clade and clade-specific epitopes in HIV-1-infected Thai and Japanese patients.** *AIDS* 2002; **16**:701–711.
- 71 Marras D, Bruggeman LA, Gao F, Tanji N, Mansukhani MM, Cara A, *et al.* **Replication and compartmentalization of HIV-1 in kidney epithelium of patients with HIV-associated nephropathy.** *Nat Med* 2002; **8**:522–526.
- 72 Bruggeman LA, Ross MD, Tanji N, Cara A, Dikman S, Gordon RE, *et al.* **Renal epithelium is a previously unrecognized site of HIV-1 infection.** *J Am Soc Nephrol* 2000; **11**:2079–2087.

Table 1: Summary on 23 early HIV infected Nigerians

Sample ID	Gender	Age(years)	High Risk group	Location	Date of HIV detection	HIV-1 RNA Viral Load	Subtype	GenBank accession number
EHIV001	Female	21	RM	Saki	26 Jan 2016	Undetectable	Unassigned	KY786266
EHIV002	Female	45	RM	Ibadan	26 Jan 2016	Undetectable	K	KY786267
EHIV003	Female	31	RM	Saki	26 Jan 2016	4999281	CRF 02-AG	KY786268
EHIV004	Female	32	RM	Saki	26 Jan 2016	5142127	CRF 02-AG	KY786269
EHIV005	Female	38	RM	Saki	26 Jan 2016	5856363	CRF 02-AG	KY786270
EHIV006	Female	28	RM	Saki	26 Jan 2016	1742364	Recombinant GD	KY786271
EHIV007	Male	48	RM	Ibadan	26 Jan 2016	Undetectable	CRF 02-AG	KY786272
EHIV008	Male	38	RM	Saki	26 Jan 2016	Undetectable		Not Sequenced
EHIV009	Female	32	RM	Ibadan	20 Feb 2017	Undetectable		Not Sequenced
EHIV010	Male	31	RM	Ibadan	21 Feb 2017	6492033	A	MN943617
EHIV011	Male	26	RM	Ibadan	21 Feb 2017	5449249	A	MN943616
EHIV012 ^b	Female	29	RM	Ibadan	05 Oct 2017	5413537	G	MN943624
EHIV013 ^b	Female	32	RM	Ibadan	21 Aug 2017	6906290	A	MN943613
EHIV014	Male	25	RM	Ibadan	21 Jul 2017	10084640	CRF 02-AG	MN943628
EHIV015	Male	42	VBD	Ibadan	26 Jul 2017	5499245	A	MN943615
EHIV016 ^b	Male	29	VBD	Ibadan	02 Aug 2017	4013634	G	MN943627
EHIV017	Female	26	RM	Ibadan	11 Nov 2016	5613523	CRF 02-AG	MN943629
EHIV018	Female	29	RM	Ibadan	20 Feb 2017	5642093	CRF 02-AG	MN943633 ^a
EHIV019	Male	46	VBD	Ibadan	27 Jul 2017	Undetectable		Not Sequenced
EHIV020	Male	40	VBD	Ibadan	17 Aug 2017	4392180	CRF 02-AG	MN943634 ^a
EHIV021	Male	29	VBD	Ibadan	21 Jul 2019	Undetectable	CRF 02-AG	MN943635 ^a
EHIV022 ^b	Male	33	VBD	Ibadan	24 Jul 2017	6820582	G	MN943625
EHIV023 ^b	Male	22	VBD	Ibadan	17 May 2017	^c Undetectable	A	MN943619

Table 2: Analysis of samples collected during follow- up

S/N	Sample ID	Accession Numbers	Follow-up Schedule					Detection of Abs
			Baseline	3mths	6mths	9mths	1 year	
1	EHIV 012	MN943626; MN943624; MN943630	05/10/2017 ^a		✓ ^a			✓ ^a
2	EHIV 013	MN943613; MN943614; MN943631	21/08/2017 ^a		✓		✓	✓
3	EHIV 016	MN943627; MN943623	02/08/2017 ^a		✓			✓
4	EHIV 022	MN943625; MN943622	24/07/2017 ^a		✓			✓ ^a
5	EHIV 023	MN943618; MN943619; MN943620; MN943621	17/05/2017 ^a				✓ ^a	✓ ^a

^a HIV-1 GAG gene isolated and sequence

Table 3: Substitutions associated with highly conserved sites in HIV-1 GAG gene during early HIV infection

	EHIV accession no	CTL epitope region (162-183)	CTL epitope region (203-214aa)	CTL epitope region (243-251aa)	P-value
G	Ref Seq	KAFSPEVIPMFSALSEGATPQD	DTINEEAAEWDR	TSTLQEQIR	ns
	MN943624	Conserved	Conserved	Conserved	
	MN943625	Conserved	D203E	Conserved	
	MN943627	Conserved	Conserved	Conserved	
	% Substitution	Nil	2.7	Nil	
A	Ref Seq			*TSTPQEQIQWMT	ns
	MN943613	Conserved	Conserved	TSTTQEQIAWMT	
	MN943615	K162R; A163G	DTSPTR*PWAGI	*HXQELLVPLK	
	MN943616	A163G	DTSMRKLQNGTD	*HXQELLVPLK	
	MN943617	Conserved	Conserved	*HXQELLVPLK	
	MN943619	Conserved	DTIN*G*PXXGQ	*HNRSY*YPSRT	
	% Substitution	2.7	50	65	
CRF-02 AG	MN943628	Conserved	Conserved	Conserved	ns
	MN943629	Conserved	Conserved	Conserved	
	% Substitution	Nil	Nil	Nil	
P Value		ns	ns	0.00902	ns

Table 4: Substitutions associated with highly conserved sites in HIV-1 GAG gene isolated from 5 individuals followed from early HIV infection till detection of antibodies

Subtype	Sample ID	EHIV accession no	CTL epitope region (162-183)	CTL epitope region (203-214aa)	CTL epitope region (243-251aa)	P-Value	
G	EHIV016	Ref Seq	KAFSPEVIPMFALSEGATPQD	DTINEEAAEWDR	TSTLQEQIR	ns	
		MN943627	Conserved	Conserved	Conserved		
		MN943623 ^a	Conserved	Conserved	Conserved		
		% Substitution	Nil	Nil	Nil		
	EHIV022	MN943625	Conserved		D203E	Conserved	ns
		MN943622 ^a	Conserved		E203D	Conserved	
		% Substitution	Nil	4	Nil		
	EHIV012	MN943626	Conserved		Conserved	Conserved	ns
		MN943624	Conserved		Conserved	Conserved	
		MN943630 ^a	M175k		D207E	R251G	
		% Substitution	2	2	2		
	A	EHIV023	Ref Seq			TSTPQEQQLQWMT	0.04929
MN943619			Conserved	DTIN*G*PXXGQ	*HNRSY*YPSRT		
MN943618			Conserved	Conserved	*HSRSY*YPSRT		
MN943620			Conserved	DTINEEEFG*SN	*HSRSY*YPSRT		
MN943621 ^a			Conserved	Conserved	*HSRSY*YPSRT		
% Substitution	Nil	31.2	8.3				

EHIV013	MN943613	Conserved	Conserved	TSTTQEQIAWMT	ns
	MN943614	Conserved	Conserved	TSTPQEIQWMT	
	MN943631 ^a	A166G	L216V	TSTLQEIQWMT	
	% Substitution	2.7	2.7	16.6	
P-Value	ns	ns	ns		

^aSequence isolated after detection of HIV-1 antibodies

Table 5: Frequency of non-synonymous substitutions in sites outside the MHR of HIV-1 GAG gene amino acids

G	aa position in HIV-1 GAG gene	16	21	34	68	70	103	107	%sub
	aa in Ref seq	L	V	Y	S	X	G	D	
	aa in EHIV samples	X	M	N	G	Y	C	V	2.3
CRF02 AG	aa position in HIV-1 GAG gene	92	97	105	116	176			
	aa in Ref seq	L	T	E	Q	T			
	aa in EHIV samples	Q	K	K	I	K			1.6
A*	aa position In HIV-1 GAG gene	89							
	aa in Ref seq								
	aa in EHIV samples	X							

These aa replacements were 100% substituted in the EHIV samples compared to Ref seq. *Indel in aa positions 85-88 introduced a frame shift

Table 6: Frequency of non-synonymous substitutions in variable sites of HIV-1 Subtype G GAG gene isolated from 3 individuals followed up from early infection till detection of antibodies

Sample ID		Accession no															
EHIV016		MN943627															
		MN943623 ^a		No variable site													
EHIV022 (2.3% sub)	aa position	106	107	108									110-113				
	MN943625	N	V	C									RREK				
	MN943622 ^a	K	I	Q									KSQE				
EHIV012 (8.6%sub)	aa position	31	62	66	69	82	86	90-96	101	104	106	110-114	121-2				
	MN943626	M	T	L	Q	V	Y	QEFWLKG	V	V	K	KSQQE	GN				
	MN943624	M	K	P	Q	V	Y	QRMGVKD	V	V	K	KSQQE	GN				
	MN943630 ^a	L	E	S	R	I	W	QRIDIRD	L	M	E	KSQKQ	GN				
	aa position	158	170	172	185	189	206	210				214	217	224	248		
	MN943626	V	M	T	T	T	D	E				L	TQ	A	R		
	MN943624	V	M	T	T	I	D	E				T	TQ	A	R		
MN943630 ^a	I	K	S	X	I	E	D				T	VH	P	G			

These aa replacements were 100% substituted in the early HIV-1 infected samples compared to ^asequences from same persons isolated after detection of antibodies

Table 7: Frequency of non-synonymous substitutions in variable sites of HIV-1 Subtype A GAG gene isolated from 2 individuals followed up from early infection till detection of antibodies

Sample ID	Accession no														
EHIV023 (4% sub)	aa position	205	215-222	228-230											
	MN943619	I	LHPVHAGX	DKR											
	MN943618	I	QG*LYPRC	PAX											
	MN943620	T	NG*LYPRC	DKX											
	MN943621 ^a	I	LHPVHAGX	DKR											
EHIV013 (13.6%sub)	aa position	15	28	30	41	49	60	62-62	66	69	72	76	82	86	
	MN943613	K	T	K	M	G	V	EQ	S	K	T	R	V	Y	
	MN943614	S	K	R	L	S	I	DR	P	K	S	R	V	Y	
	MN943631 ^a	A	K	R	L	G	L	EQ	S	R	S	K	I	W	
	aa position	92-93	104	107	115	118	121	124	125-127	139	143	146-7	182	215	
	MN943613	EV	I	I	E	A	D	N	SKV	Q	V	AV	G	L	
	MN943614	DV	I	M	P	A	D	S	SKV	Q	V	TL	Q	L	
	MN943631 ^a	DI	M	I	V	T	A	S	***	R	T	SM	Q	V	
	aa position	218	223-224	228	243	248	252	260	263	267					
	MN943613	L	IA	I	T	A	G	D	R	I					
	MN943614	V	VP	M	P	G	S	D	K	I					
	MN943631 ^a	V	IP	M	L	G	S	E	K	V					

These aa replacements were 100% substituted in the EHIV samples compared to ^aSequence isolated after detection of antibodies.

Figure 4: Serum creatinine concentration during longitudinal follow up of three individuals from early infection till detection of antibodies

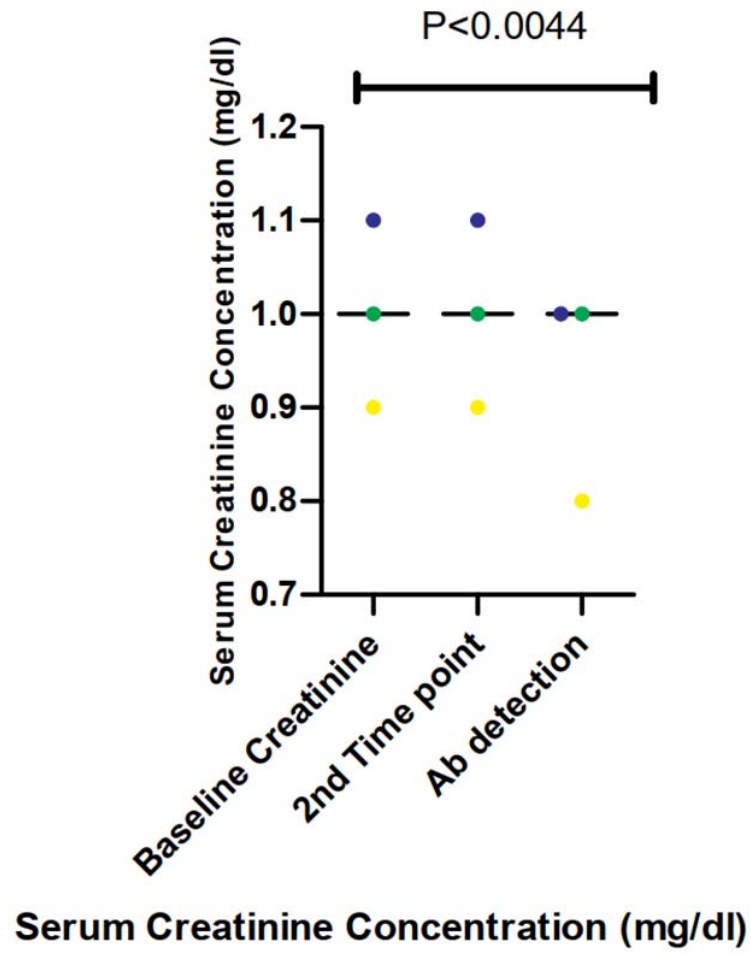


Figure Legend

Table 1: Summary on early HIV-1 infected Nigerians

Individuals comprised of persons referred for malaria parasite testing (RM) and Voluntary Blood Donors (VBD). All blood donors were from Ibadan. ^aThese samples were genotyped using the VIF gene as we did not obtain GAG gene for these individuals. ^bThese individuals were followed up until HIV-1 antibodies were detectable in their sera. ^cEHIV 023 HIV-1 RNA viral load became detectable after HIV-1 antibodies became detectable.

Table 2: Analysis of samples collected during follow up

Blood samples collected from the five individuals were analyzed for HIV-1 GAG gene. All samples were analyzed at baseline, after HIV-1 antibodies detection and at every time point indicated with superscript a. Each analyzed gene was sequenced and successfully submitted to GenBank with the accession numbers provided. Each accession number corresponds to a GAG gene sequenced at a particular time point. Two samples were sequenced at two and three time points each while a sample was sequenced at four time points. The time at which HIV-1 antibodies were detected does not correspond to seroconversion date as most of these individuals are likely to have seroconverted before this period.

Table 3: Substitutions associated with highly conserved sites in HIV-1 GAG gene during early HIV infection

Amino acid lengths of CTL epitope regions are indicated in parenthesis. Reference HIV-1 GAG sequences for subtypes G, A and CRF02-AG were downloaded from Los Alamos national HIV sequence database (www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html) and were used for comparisons with similar subtypes obtained from persons at the early stages of HIV infection. Sequences that are not different from reference sequences (presence of the consensus amino acid) are indicated with the word **Conserved**. If an amino acid position is variable to the reference sequence, the letter and the position is indicated in the normal convention, e.g MN943625 sequence is variable at position 203 and it is indicated as D203E which means at position 203, D is substituted with E. When more than two amino acid positions are variable, the entire sequence is written against the accession number of the subtype. *Reference sequence for subtype A CTL epitope region 243-251 is different from those of other subtypes and it is indicated. %Substitution was calculated by deducing the percentage of the ratio of variable positions against total amino acid positions for each sequence. Statistical tests of significance within and across CTL epitope regions for the subtypes studied were calculated using Kruskal-Wallis test, with P set at 0.05. ns=not significant. Significant values are indicated in bold values.

Table 4: Substitutions associated with highly conserved sites in HIV-1 GAG gene isolated from 5 individuals followed from early HIV infection till detection of antibodies

Amino acid lengths of CTL epitope regions are indicated in parenthesis. Reference HIV-1 GAG sequences for subtypes G, A and CRF02-AG were downloaded from Los Alamos national HIV sequence database (www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html) and were used for comparisons with similar subtypes obtained from persons at the early stages of HIV infection. ^aIndicated accession number for sequences obtained from samples after detectable HIV-1 antibodies. Sequences that are not different from reference sequences (presence of the consensus amino acid) are indicated with the word **Conserved**. If an amino acid position is

variable to the reference sequence, the letter and the position is indicated in the normal convention, e.g MN943625 sequence is variable at position 203 and it is indicated as D203E which means at position 203, D is substituted with E. When more than two amino acid positions are variable, the entire sequence is written against the accession number of the subtype. ^aReference sequence for subtype A CTL epitope region 243-251 is different from those of other subtypes and it is indicated. %Substitution was calculated by deducing the percentage of the ratio of variable positions against total amino acid positions for each sequence. Statistical tests of significance within and across CTL epitope regions for the subtypes studied were calculated using Kruskal-Wallis test, with P set at 0.05. ns=not significant. Significant values are indicated in bold values.

Table 5: Frequency of non-synonymous substitutions in variable sites of HIV-1 GAG gene amino acids

Reference amino acid HIV-1 GAG sequences for subtypes G, A and CRF02-AG were downloaded from Los Alamos national HIV sequence database (www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html) and were used as background sequences to query similar subtypes obtained from persons at the early stages of HIV infection using VESPA program (<http://www.hiv.lanl.gov/content/sequence/VESPA/vespa.html>) for variable signature amino acid residues. ^aIndicated amino acid replacements were 100% substituted in the HIV samples compared to reference sequences. These substituted residues are more likely explored as possible transmission and pathogenesis signatures. Residue replacements were restricted to these signatures with 100% substitutions to reduce, if not eliminate false discovery rate-adjusted P-values (Benjamin-Hochberg procedure). *indel in amino acid positions 85-88 introduced a frame shift mutation that terminated the computation of signatures for subtype A.

Table 6: Frequency of non-synonymous substitutions in variable sites of HIV-1 Subtype G GAG gene isolated from 3 individuals followed up from early infection till detection of antibodies

Amino acid sequences obtained during early stages of HIV-1 infection were used as background sequences to query ^asequences obtained from same persons after antibodies detection using VESPA program (<http://www.hiv.lanl.gov/content/sequence/VESPA/vespa.html>) for variable signature amino acid residues. Indicated amino acid replacements were 100% substituted in the early HIV infected samples compared to sequences obtained after antibodies detection. These substituted residues are more likely explored as possible transmission and pathogenesis signatures. Residue replacement was restricted to these signatures with 100% substitutions to reduce, if not eliminate false discovery rate-adjusted P-values (Benjamin-Hochberg procedure).

Table 7: Frequency of non-synonymous substitutions in variable sites of HIV-1 Subtype A GAG gene isolated from 2 individuals followed up from early infection till detection of antibodies

Amino acid sequences obtained during early stages of HIV-1 infection were used as background sequences to query ^asequences obtained from same persons after antibodies detection using VESPA program (<http://www.hiv.lanl.gov/content/sequence/VESPA/vespa.html>) for variable signature amino acid residues. Indicated amino acid replacements were 100% substituted in the early HIV infected samples compared to

sequences obtained after antibodies detection. These substituted residues are more likely explored as possible transmission and pathogenesis signatures. Residue replacement was restricted to these signatures with 100% substitutions to reduce, if not eliminate false discovery rate-adjusted P-values (Benjamin-Hochberg procedure).

Figure 1: Phylogenetic tree of the P17/P24 regions of the GAG gene of HIV-1

Reference subtypes are indicated with Ref. before their accession numbers. Other sequences are indicated with their accession numbers and country of isolation. Subtypes obtained from samples in this study are indicated with their accession numbers only. Subtypes A, G and CRF02_AG were identified with green diamond, blue circle and pink square symbols respectively. Multiple sequences alignment and phylogenetic tree were constructed using MAFFTS and Maximum Parsimony algorithm in MEGA 6 software. Statistical significance of the tree topology was tested by 1000 bootstrap replication.

Figure 2: Distribution of mutations associated with escape in EHIV012 and EHIV013 before detection of antibodies

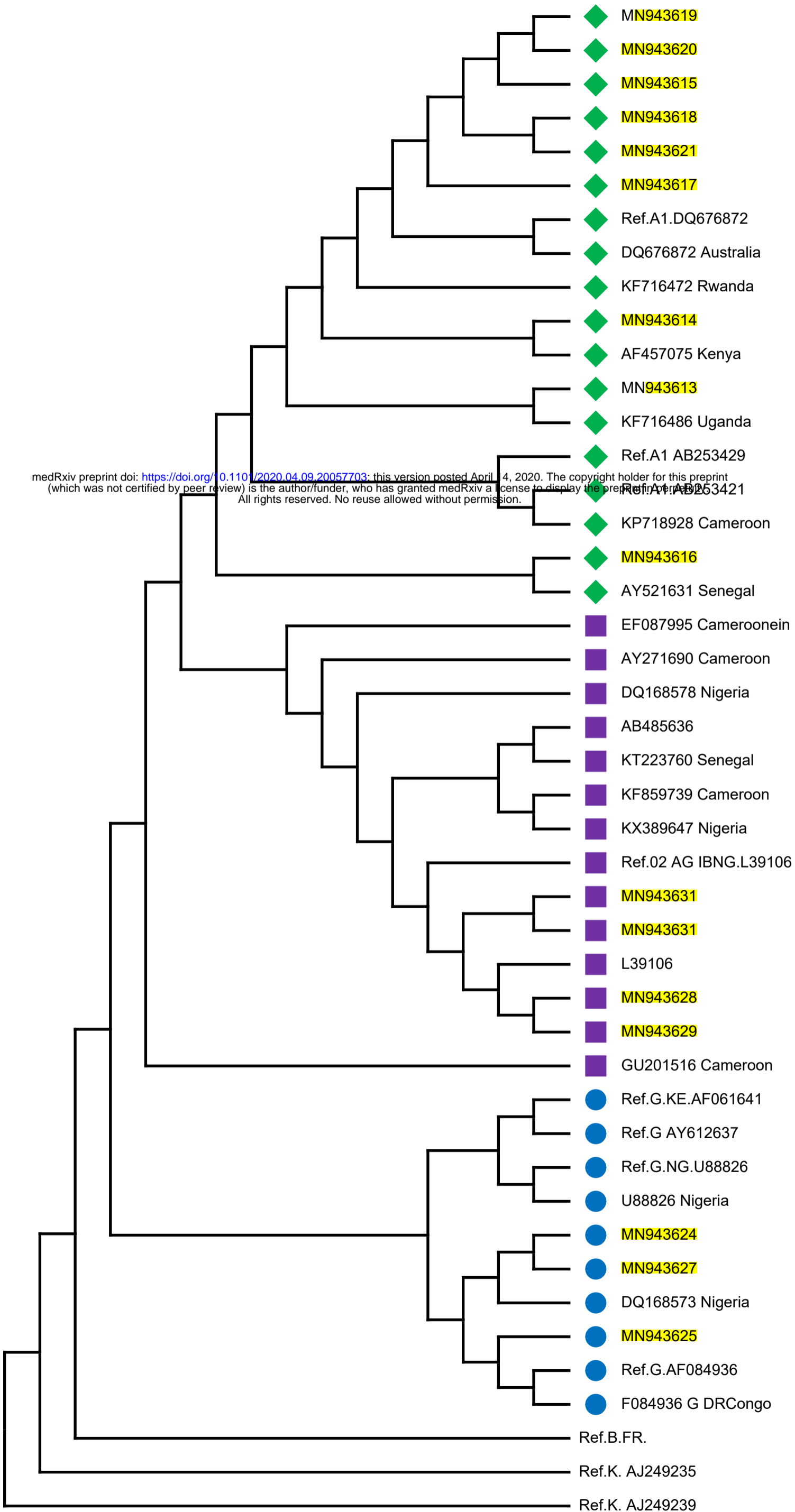
Three individuals had mutations previously associated with escape, diminished responses, non-susceptible forms etc. These mutations were compiled from the Los Alamos National Laboratory HIV Immunology Database for CTL/CD8+ Epitope Variants and Escape Mutations (https://www.hiv.lanl.gov/content/immunology/variants/ctl_variant.html). The list of all the identified mutations are presented in the Supplementary Table 1. Two individuals (EHIV012 and EHIV013) had mutations outside the GAG MHR while EHIV022 had a mutation corresponding to escape (Murakoshi) – E203D. Figure 3 shows the distribution of mutations associated with escape in EHIV 012(Red Bars) and EHIV013 (Green Bars). The number of occurrences of the mutations in GAG sequences are shown in the Y-axis while the X-axis shows aa mutations.

Figure 3: Distribution of mutations associated with escape in EHIV012 and EHIV013 after detection of antibodies

Three individuals had mutations previously associated with escape, diminished responses, non-susceptible forms etc. These mutations were compiled from the Los Alamos National Laboratory HIV Immunology Database for CTL/CD8+ Epitope Variants and Escape Mutations (https://www.hiv.lanl.gov/content/immunology/variants/ctl_variant.html). The list of all the identified mutations are presented in the Supplementary Table 1. Two individuals (EHIV012 and EHIV013) had mutations outside the GAG MHR while EHIV022 had a mutation corresponding to escape (Murakoshi) – E203D. Figure4 shows the distribution of mutations associated with escape in EHIV 012(Red Bars) and EHIV013 (Green Bars). The number of occurrences of the mutations in GAG sequences are shown in the Y-axis while the X-axis shows aa mutations.

Figure 4: Serum creatinine concentration during longitudinal follow up of three individuals from early infection till detection of antibodies

The three individuals followed up are represented in blue, yellow and green. Serum creatinine levels for these individuals were measured in three time points, Baseline, second time point and after antibodies detection. Significant differences in the levels of serum creatinine (measured in mg/dl) observed across the three time points ($P < 0.0044$) were calculated using 1-way ANOVA.



Distribution of mutations associated with escape in EHIV012 and EHIV013 before detection of antibodies

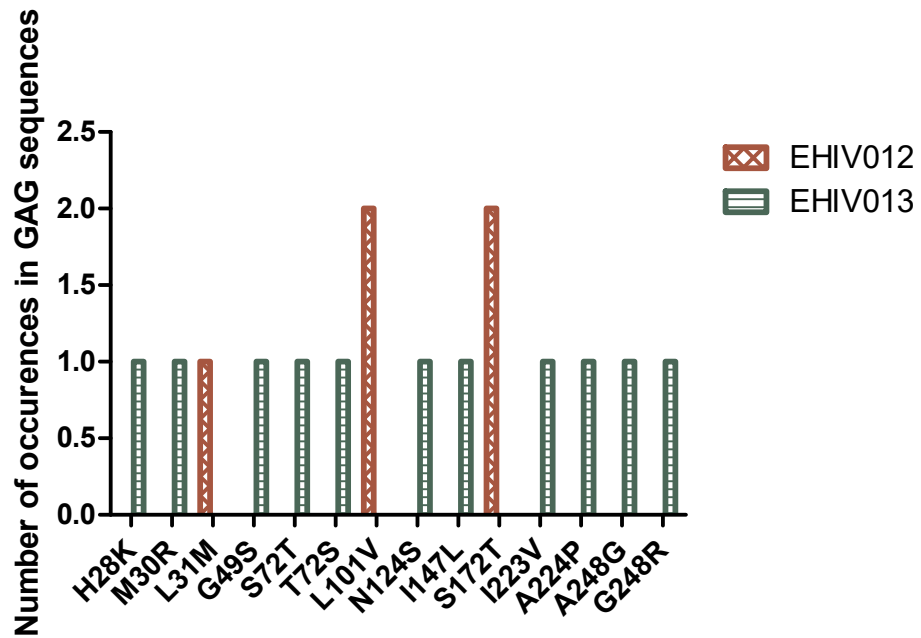


Figure 2 Mutations associated with escape is sites outside the MHR

Distribution of mutations associated with escape in EHIV012 and EHIV013 after detection of antibodies

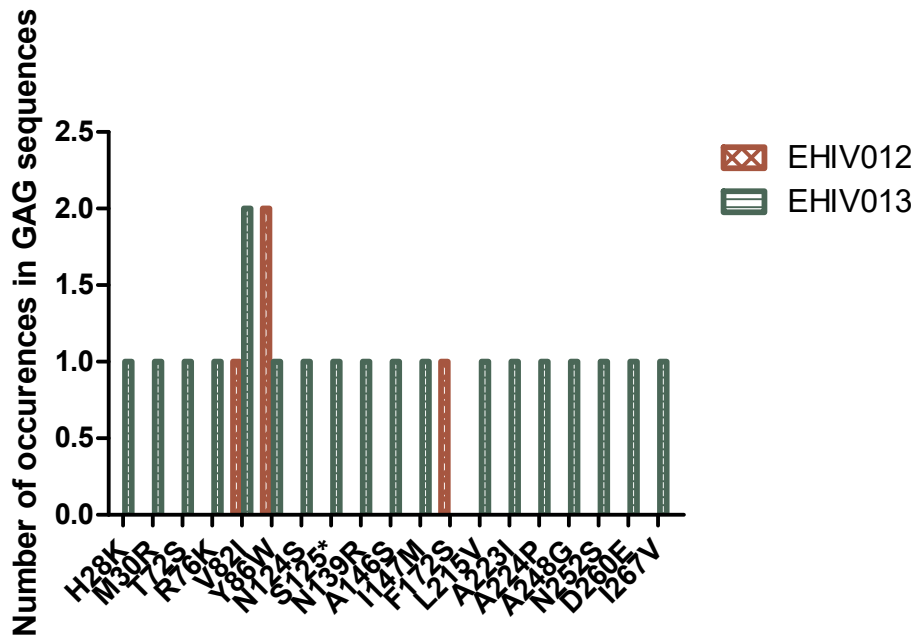


Figure 3: Mutations associated with escape is sites outside the MHR