Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Kristine Dennis Date

Establishing Molecular Epidemiology in Concordant HIV-1 Positive Heterosexual Couples at CVCT Centers in Rwanda and Zambia

By

Kristine Dennis

Master of Public Health

Global Epidemiology

_____

Dr. Susan Allen, MD, MPH

Faculty Thesis Advisor

Establishing Molecular Epidemiology in Concordant HIV-1 Positive Heterosexual Couples at CVCT Centers in Rwanda and Zambia

By

Kristine Dennis

Bachelor of Arts

Luther College

2009

Faculty Thesis Advisor: Susan Allen, MD, MPH

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements of the degree of
Master of Public Health
in Global Epidemiology
2012

Abstract

Establishing Molecular Epidemiology in Concordant HIV-1 Positive Heterosexual Couples at CVCT Centers in Rwanda and Zambia

By Kristine Dennis

Background

Epidemiologic linkage techniques for the gp41 region of the HIV-1 proviral genome are established to characterize HIV-1 transmission dynamics among heterosexual couples with recent transmission events but have unknown validity for chronically infected couples with differing levels of viral diversification. To determine if the current pair-wise distance (PWD) cutoffs (Rwanda 5.4%; Zambia 4.6%) are valid for couples who were both HIV-1 positive (concordant positives) when first tested, a 399 base pair region of gp41 was sequenced from 12 longitudinally followed recent transmission pairs and 191 concordant positive couples from Rwanda and Zambia for PWD and neighbor-joining phylogenetic tree linkage analysis.

Results

The PWD of longitudinally followed couples demonstrated 0.33% and 0.28% PWD change/year across Rwanda and Zambia, respectively, confirming the PWD cutoff for each country remains valid for approximately 10 years. Linkage by PWD and time of cohabitation were compared with the gold-standard phylogenetic tree linkage status. Most concordant positive couples had matching linkage status between PWD and phylogenetic tree analysis with the exception of a few couples from Rwanda (9 of 101) and Zambia (12 of 90). After comparing median PWD for linked and unlinked couples between concordant positives and recent seroconverting pairs, linked concordant positives had significantly higher median PWD than recent seroconverters across both countries ($p<0.05$). Unlinked recent seroconverters had a significantly higher median PWD than unlinked concordant positives only in Zambia ($p<0.01$). The proportion of linked concordant positives was 44.6% in Rwanda and 47.8% in Zambia, as compared to a sample of recent transmission pairs enrolled in couples voluntary counseling and testing with 80.5% and 77.9% linked in Rwanda and Zambia, respectively.

Conclusions

Across the concordant positive couples from Rwanda and Zambia, the validity of the current gp41 PWD cutoff was confirmed for epidemiologic linkage analysis. There was a greater proportion of couples who were unlinked than recent transmission pairs. Through the characterization of these concordant positive couples who have not received couples voluntary counseling and testing, we have a sense of the linkage patterns present in the general population so future behavioral interventions can be more accurately targeted.

Establishing Molecular Epidemiology in Concordant HIV-1 Positive Heterosexual Couples at CVCT Centers in Rwanda and Zambia

By

Kristine Dennis

Bachelor of Arts

Luther College

2009

Faculty Thesis Advisor: Susan Allen, MD, MPH

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2012

Acknowledgements

I am grateful for the guidance I received from Susan Allen, Eric Hunter and Debi Boeras while completing this project. I would like to thank all the study participants, volunteers and staff of the Rwanda Zambia HIV research group, at the Projet San Francisco in Kigali, Rwanda and the Zambia Emory HIV Research Project in Lusaka, Zambia and the technicians and students of the Hunter Laboratory at Emory University for completing much of the genetic sequencing and transmission linkage.

To all my family and friends, your encouragement and support has been invaluable to me throughout this process.

Table of Contents

List of Figures

List of Supplemental Figures

Chapter I: BACKGROUND

Since the discovery of Human Immune Deficiency Virus (HIV) in 1981, the infection has impacted people across the world. Sub-Saharan Africa faces a significant burden with heterosexual couples as a primary risk group [1]. Molecular epidemiology serves as a primary tool to characterize patterns of recent transmission in order to target high-risk groups and prevention efforts. By applying epidemiologic linkage tools to chronically infected couples, their effectiveness for this population can be evaluated while simultaneously informing transmission patterns for couples who did not receive preventive counseling prior to HIV-1 infection.

HIV Prevalence

The worldwide impact of HIV-1 has been significant with over 33.3 million living with HIV-1 worldwide as of 2009 [1]. With around 22.5 million people infected, sub-Saharan Africa has the highest prevalence of HIV-1 in the world [1]. Rwanda and Zambia face a significant portion of the disease burden with around 140,000 to 170,000 and 670,000 to 790,000 infected individuals respectively [1, 2]. Through current efforts in prevention and treatment, HIV-1 incidence has begun to decline while the prevalence remains high as new treatments and better care allow people to live longer with the disease.

Heterosexual Transmission

In Africa, the majority of HIV-1 infections occur through heterosexual transmission between sero-discordant (one partner is HIV-1 negative and the other is HIV-1 positive) cohabiting couples [3, 4]. In sub-Saharan Africa, an estimated 44 percent of couples affected by HIV-1 are HIV-1 discordant [5]. With HIV-1 transmission rates

for African discordant couples with unknown serostatus ranging from 20-25%, CVCT interventions in Rwanda and Zambia serve to substantially reduce seroconversion to 3-7% per year [6-9]. Dunkle et al estimated between 55.1% to 92.7% of new heterosexually acquired HIV-1 infections in Rwanda and Zambia occur within sero-discordant cohabiting couples [3]. Since this demographic is particularly high risk, it provides an opportunity to prevent a large portion of transmissions within this population.

Rwanda-Zambia HIV Research Group

In 1986, the first site of the Rwanda-Zambia HIV Research Group (RZHRG) was created by Dr. Susan Allen in Kigali, Rwanda and expanded to Zambia in 1994. Currently there are two active sites located in Kigali, Rwanda and Lusaka, Zambia. Couples were identified as one of the largest risk groups in Africa with the opportunity to lower the risk of transmission through couples' voluntary counseling and testing [10]. RZHRG sites have maintained some of the largest heterosexual HIV-1 discordant couples' cohorts in the world [4, 11]. With CVCT, many transmissions among discordant couples can be prevented [3]. Despite these efforts, new HIV-1 infections still occur. Through tracking these couples over time, recent transmission events among HIV-1 discordant couples where a previously HIV-1 uninfected partner becomes HIV-1 positive during follow-up can be identified.

Epidemiologic Linkage

In order to understand the at-risk population and where the greatest proportion of transmission events are occurring, it is essential to correctly characterize whether transmission is happening between partners in cohabiting relationships or from an individual outside the relationship. Epidemiologic linkage is a tool to confirm linkage

status and eliminate any biases from self-reported behavior [12]. PCR amplification, sequencing, pair-wise distance, and phylogenetic tree analysis is the gold standard for establishing epidemiologic linkage of viruses from recent transmission pairs. Since there is a fairly high level of viral genomic diversity among quasispecies of HIV-1, relatedness of viruses between individuals can be determined with a relatively high level of certainty. Among cohabiting couples, epidemiologic linkage is determined through sequencing segments of the proviral genome [12]. gp41 is one gene that has enough variation to identify differences between viruses while also being fairly conserved to allow for reliable analysis of a couples' linkage status [13]. If the infection was acquired from within a cohabiting couple, the couple can be classified as "linked". Otherwise the infection was acquired from outside the couple and the transmission event can be classified as "unlinked".

Concordant Positive Couples

However, among concordant positive couples where transmission from one to the other may have occurred many years earlier, each individual's virus may evolve differently and after several years may differ by more than 4-6% [14]. Conversely, two individuals who were already HIV-1 infected when they married years ago may have 'exchanged viruses' which then may have combined. In couples who are both HIV-1 positive when first tested, the validity of our current approach for linkage analysis is unknown. Due to the continual divergence of the virus and the unknown impact of superinfection, hypermutation and recombination, there is the potential to misclassify linkage status [14-16].

HIV-1 Diversity: Recombination, Superinfection & Hypermutation

Recombination

Like other retroviruses, recombination can easily occur due to the replication mechanisms that occur with RNA viruses. Recombination has been observed to happen relatively efficiently at a frequency of around 0.2 recombination events per genome per replication cycle [16]. Since HIV infected cells usually carry at least two proviruses, there are frequent opportunities for recombination events to occur [17, 18].

Superinfection

For an individual already infected with HIV-1, there is the possibility of infection with another quasispecies of virus, resulting in what is termed superinfection [19, 20]. This differs from coinfection in a couple ways. Although both are dual infections, superinfection occurs only after the immune system has mounted a full response to the initial infecting viral strain before a second viral strain successfully infects the individual [19, 21]. Kraft *et al* 2012 observed HIV-1 superinfection occurring at rates similar to that of primary infection [22]. With levels of superinfection at this rate, it may have implications in the evaluation of linkage status, particularly when only one gene is sequenced for linkage assessment.

Hypermutation

The success of HIV-1 viral spread, transmission, and continued virulence has also been linked to the frequency of hypermutation. The virus has a preference for G to A hypermutation with this trait potentially benefiting the HIV-1 genomes' ability to evolve at a relatively high rate [15]. Berkhout *et al* 2001 propose that G to A mutation is more common given it is a property of reverse transcriptase [23].

Chapter II: MANUSCRIPT

Establishing Molecular Epidemiology in Concordant HIV-1 Positive Heterosexual Couples at CVCT Centers in Rwanda and Zambia

Authors:

Kristine Dennis, Debrah Boeras, Rebecca Johnson, Mackenzie Hurlston, Eric Hunter, Susan Allen

Abstract

Background
Epidemiologic linkage techniques for the gp41 region of the HIV-1 proviral genome are established to characterize HIV-1 transmission dynamics among heterosexual couples with recent transmission events but have unknown validity for chronically infected couples with differing levels of viral diversification. To determine if the current pair-wise distance (PWD) cutoffs (Rwanda 5.4%; Zambia 4.6%) are valid for couples who were both HIV-1 positive (concordant positives) when first tested, a 399 base pair region of gp41 was sequenced from 12 longitudinally followed recent transmission pairs and 191 concordant positive couples from Rwanda and Zambia for PWD and neighbor-joining phylogenetic tree linkage analysis.

Results
The PWD of longitudinally followed couples demonstrated 0.33% and 0.28% PWD change/year across Rwanda and Zambia, respectively, confirming the PWD cutoff for each country remains valid for approximately 10 years. Linkage by PWD and time of cohabitation were compared with the gold-standard phylogenetic tree linkage status. Most concordant positive couples had matching linkage status between PWD and phylogenetic tree analysis with the exception of a few couples from Rwanda (9 of 101) and Zambia (12 of 90). After comparing median PWD for linked and unlinked couples between concordant positives and recent seroconverting pairs, linked concordant positives had significantly higher median PWD than recent seroconverters across both countries ($p<0.05$). Unlinked recent seroconverters had a significantly higher median PWD than unlinked concordant positives only in Zambia ($p<0.01$). The proportion of linked concordant positives was 44.6% in Rwanda and 47.8% in Zambia, as compared to a sample of recent transmission pairs enrolled in couples voluntary counseling and testing with 80.5% and 77.9% linked in Rwanda and Zambia, respectively.

Conclusions
Across the concordant positive couples from Rwanda and Zambia, the validity of the current gp41 PWD cutoff was confirmed for epidemiologic linkage analysis. There was a greater proportion of couples who were unlinked than recent transmission pairs. Through the characterization of these concordant positive couples who have not received couples voluntary counseling and testing, we have a sense of the linkage patterns present in the general population so future behavioral interventions can be more accurately targeted.

Introduction

The global impact of HIV-1 has been considerable, particularly in sub-Saharan Africa with the highest prevalence of HIV-1 in the world and around 22.5 million people infected [1]. Rwanda and Zambia are no exception with HIV-1 prevalence among adults age 15 to 49 ranging from 2.5% to 3.3% and 12.3 to 16.1%, respectively [1, 24]. The majority of new HIV-1 infections in sub-Saharan Africa occur between heterosexual couples with cohabiting couples as a primary risk group [1, 25]. In Rwanda and Zambia, the majority of HIV-1 infections occur within sero-discordant (one partner HIV-1 negative and the other HIV-1 positive) cohabiting couples [1, 3, 4, 25].

Previous studies have shown the gp41 region in viruses from linked transmission differ less than 5.4% in Zambia and 4.6% in Rwanda [12]. However, among concordant positive couples where transmission from one partner to the other may have occurred many years earlier, each individual's virus may diverge differently and may differ by more than 4-6% after several years. Conversely, two individuals who were already HIV-1 infected upon initial cohabitation may have 'exchanged viruses' which may have combined. The effectiveness of the current approach for linkage analysis in couples who are both HIV positive when first tested is unknown, given the virus' continual divergence over a period of years (leading to a potential misclassification of 'linked' couples as 'unlinked') and the unknown impact of superinfection and recombination (leading to a potential misclassification of 'unlinked' couples as 'linked').

We observe progression of virus evolution over 5 to 7 years in twelve initially HIV-1 discordant couples, in whom the HIV-1 positive partner transmitted to the HIV-1 negative partner, and compare the pair-wise distance and phylogenetic relationships in

these couples with those in 191 cohabiting couples who were both HIV-1 positive when initially tested. Through assessing epidemiologic linkage of HIV-1 concordant positive cohabiting couples with the current gold standard techniques for recent transmissions, we can assess the translatability of these techniques to cross-sectional testing of concordant positive couples and the accuracy of current linkage cutoffs. For these concordant positive couples, length of marriage was recorded at time of HIV-1 testing. Given our understanding of HIV-1 prevalence, transmission patterns, and rates of viral genetic change, we can compare the observed and expected linkage patterns among concordant positive couples given length of cohabitation. The validity of the current linkage cutoffs will be analyzed to determine their effectiveness for cross-sectional testing of concordant positive couples and the potential for a more accurate estimate of the proportion of linked and unlinked HIV-1 transmission occurring among cohabiting heterosexual couples from the general population.

Methods

Longitudinal HIV-1 couples

The Zambia Emory HIV Research Project (ZEHRP) in Lusaka, Zambia and Projet San Francisco (PSF) in Kigali, Rwanda conducted couples HIV voluntary testing and counseling (CVCT). For the longitudinal analysis, six previously enrolled HIV-1 serodiscordant heterosexual couples from each CVCT cohort with documented seroconversion, confirmed epidemiologic linkage, and stored blood samples from at least 5 years post seroconversion were selected [3, 4]. Of the six Zambian couples, there were two female to male and four male to female transmission pairs and from the Rwandan couples, three female to male and three male to female transmission pairs.

Concordant HIV-1 positive couples

Concordant HIV-1 positive heterosexual couples were also identified at ZEHRP and PSF CVCT sites for a cross-sectional analysis of epidemiological linkage. Blood samples without identifiers were obtained from 90 and 101 couples from Lusaka and Kigali respectively. Self-reported length of cohabitation was collected for the analysis.

HIV-1 testing

Cohabitating heterosexual couples visiting CVCT sites were tested with rapid HIV-1 antibody qualitative assays as previously described [6, 8, 10]. Couples identified as serodiscordant were invited to enroll in the Heterosexual Transmission (HT) study. The HT study tested the HIV-1 negative partner quarterly with HIV-1 rapid antibody assays and an HIV-1 p24 antigen ELISA (Beckman Coulter). Upon seroconversion, the pair was followed longitudinally for follow-up studies. The couples identified as concordant positive at CVCT were not followed longitudinally.

Epidemiological linkage

All blood samples were centrifuged to separate plasma (PL) and buffy coat (BC). BC was stored in RNA later (Ambion, TX) and both PL and BC samples were stored at -80 C. Blood viral RNA and genomic DNA were extracted from PL and BC according to the manufacturers instructions using the QIAamp Viral RNA Mini Kit (Qiagen) and the QIAamp DNA Blood Midi Kit (Qiagen), respectively. Viral RNA and DNA from concordant positive and longitudinal HT individuals was used to amplify the same 399 base pair section from gp41. Amplification was conducted in triplicate for viral RNA with nested PCR and Reverse Transcriptase-PCR for the first round and for viral DNA

with two rounds of nested PCR in reaction conditions previously described [12]. The following primers were used for gp41 amplification:

*gp41*F1 – 5' TCTTAGGAGCAGCAGGAAGCACTATGGG 3'

*gp41*R1 - 5' AACGACAAAGGTGAGTATCCCTGCCTAA 3'

*gp41*F2 – 5' ACAATTATTGTCTGGTATAGTGCAACAGCA 3'

*gp41*R2 – 5' TTAAACCTATCAAGCCTCCTACTATCATTA 3'

Epidemiological linkage was determined for all concordant and longitudinal couples using phylogenetic analysis. A 399 base pair section from the HIV-1 envelope gp41 region was PCR amplified for each couple's time points from plasma and proviral DNA. Successfully amplified PCR products were purified and sent for population sequencing at Eurofins MWG Operon. Sequences were cleaned and trimmed with the program Sequencher 5.0 (Gene Codes Corporation, Ann Arbor, MI) and degenerate bases coded as necessary according to the IUPAC specifications. All sequences were aligned and compiled into neighbor-joining phylogenetic trees using Geneious Bioinformatics software (Biomatters Ltd, Auckland, New Zealand). The phylogenetic trees for each country were created with HIV-1 subtype reference sequences from the Los Alamos National Laboratory (LANL) HIV Sequence Database (http://www.hiv.lanl.gov/), known reference linkage pairs from the HT cohort and the respective concordant positive or longitudinal couple sequences. The neighbor-joining method with bootstrapping was used to determine linked and unlinked transmission pairs.

For longitudinal HT couples, nucleotide pair-wise distance (PWD) for each time point was determined and plotted in relation to years post seroconversion in Prism and fit with a linear regression line. Highlighter plots to identify accumulated point mutations

over time were created using the LANL HIV *Highlighter* analysis tool (http://www.hiv.lanl.gov/content/hiv-db/HIG-LIGHT/highlighter.html).

gp41 pair-wise distance cutoff

PWD for each couple was recorded after tree construction. PWD cutoff values for epidemiologic linkage were calculated with the mean sequence distances of gp41 from the reference set minus two standard deviations. This technique was established under previous methods for Zambian sequences and was separately utilized for Rwandan sequences [12]. The cutoffs for epidemiologic linkage are 5.4 percent for Zambia and 4.6 percent for Rwanda with any pair at or below this cutoff considered linked. Concordance of linkage status as determined by the branching of phylogenetic trees with the PWD cutoff was assessed graphically through scatter plot construction and statistical analysis in Prism. Median PWD was compared for linked concordant positive couples and recent transmission pairs for both countries using a one-sided Mann-Whitney test. Median PWD was compared for unlinked CP couples and recent transmission pairs for both countries using a Mann-Whitney test. Length of cohabitation in relation to PWD was plotted in Prism and assessed using Spearman rank correlation.

All research protocols were approved by the Emory University Institutional Review Board, the University of Zambia Schools of Medicine Research Ethics Committee, and the Rwanda Ethics Committee.

Results

Longitudinal follow-up of known transmission pairs confirms validity of linkage cutoff for CP couples

After amplifying the 399 bp segment of gp41 from the known transmission pairs at multiple time points, phylogenetic trees were constructed with all sequences for the

transmission pairs from each site. All couples cluster on their own branch of the tree, confirming they continue to appear linked for up to seven years (Figure 1a and 1b).

PWD was also determined for all male and female matching time points. PWD and years post day of seroconversion (DOS) were plotted against one another and fit with a linear regression line for the six couples from each site. From the sample of Zambian transmission pairs, the positive trend of PWD with years post DOS indicates about a 0.28 (95% CI 0.12, 0.45) percent PWD genetic change per year (Figure 1a). After seven years post DOS, no couples were near the 5.4% PWD linkage cutoff. From the sample of Rwandan transmission pairs, the positive trend of PWD with years post DOS indicates about a 0.33 (95% CI 0.16, 0.51) percent PWD genetic change per year (Figure 1b). Among Rwandan transmission pairs, one couple is nearing the 4.6% PWD cutoff at seven years post DOS but still remains below the cutoff.

Given an expected PWD change in the range of 0.28 to 0.33 percent per year, a concordant positive couple who has been cohabitating for ten years, assuming transmission occurred immediately, would on average experience about a 2.8 to 3.3 percent change in PWD over that time period. Given the accepted cutoff of 5.4 percent for Zambian linkage and 4.6 percent for Rwandan linkage, it is likely that concordant positive couples who were originally linked would only appear to be unlinked after about ten years of cohabitation. Therefore, the cutoffs established for recent transmission pairs are relatively accurate tools for assessing concordant positive couples' linkage status years after transmission has occurred.

In order to consider hypermutation, recombination and superinfection for these donor-recipient pairs, a highlighter plot was generated for the couple with the most

sequenced time points from Zambia and Rwanda. The Zambian recipient had more base pair changes with a total of ten base pair substitutions by year seven post DOS (Supplemental Figure 2a). There was also overall variability year to year with the location of base pair changes. The most common nucleotide change was to adenine for the Zambian recipient. The transition to adenine may fit with a hypothesis of hypermutation given G to A hypermutation is the most common base pair mutation. The Rwandan recipient had relatively few specific base pair changes with only three base pair substitutions when assessing a sample seven years after transmission occurred (Supplemental Figure 2b). The second most common nucleotide change was to cytosine for both the Rwandan and Zambian recipient.

Linkage assessment of concordant positive couples through PWD and phylogenetic tree analysis

Buffy coat and plasma samples were collected from 191 concordant positive couples in Zambia (90 couples) and Rwanda (101 couples) when the partners were identified as HIV-1 antibody positive. Once the 399bp region of gp41 was amplified and sequenced, two separate phylogenetic trees were constructed for the respective locations due to the predominance of different subtypes between Zambia (subtype C) and Rwanda (majority of samples subtype A) (Supplemental Figure 3a & 3b).

With the branching on the phylogenetic trees as the gold standard, linkage status was determined for all couples. When the PWD linkage status was compared to the linkage status according to the phylogenetic tree branching for Zambian concordant positives, seven couples appeared linked by PWD yet unlinked by branching and five couples appeared unlinked by PWD yet linked by branching (Figure 2a). For Rwandan

concordant positive couples, six appeared linked by PWD while unlinked by branching and three couples appeared unlinked by PWD yet linked by branching (Figure 3a).

For recent transmission pairs, one couple from Zambia appeared unlinked by PWD but linked by phylogenetic tree analysis (Figure 2a). Among Rwandan seroconverters, three couples appeared unlinked by PWD but linked by phylogenetic tree analysis and one linked by PWD but unlinked by phylogenetic tree analysis (Figure 3a).

The comparison of linkage status across PWD and phylogenetic tree branching indicates that the PWD cutoff primarily used for recent transmission pairs has a similar degree of accuracy for recent transmission pairs as for concordant positive couples with a few exceptional cases (Figure 2a & 3a). The distribution of discrepant cases among the concordant positives couples shows the majority of differences are due to PWD values indicating a couple is linked when the couple is actually unlinked according to phylogenetic tree branching.

In Zambia and Rwanda, the median PWD for linked concordant positives couples is significantly greater than linked recent seroconverter couples ($P<0.05$) by a one-sided Mann-Whitney test. In contrast, the median PWD of unlinked concordant positives in Zambia was significantly lower than recent seroconverter couples ($p<0.05$, Mann-Whitney test) but in Rwanda, unlinked concordant positives had a slightly higher median PWD than recent seroconverters ($p>0.05$, Mann-Whitney test).

For Zambian concordant positive couples, the PWD increases as the length of cohabitation increases ($P<0.05$, Spearman rank correlation; Figure 2b). When PWD and cohabitation time is compared for Zambian couples, two distinct linear clusters occur. One clusters along an approximate slope of 0.4% PWD per year similar to that observed

among longitudinal recent transmission pairs while the other group has a slope relatively close to zero. For Rwandan concordant positive couples, the PWD increases slightly with the length of cohabitation ($P<0.05$, Spearman rank correlation; Figure 3b).

Assessing transmission patterns in the general population without previous counseling

When couples were initially screened for enrollment in the HT study, estimations of concordant positive, concordant negatives and serodiscordant couples were calculated. In Lusaka, Zambia, 51% of couples were concordant negative, 23% were serodiscordant, and 26% were concordant positive [10]. In Kigali, Rwanda, 71% of couples were concordant negative, 9% were serodiscordant, and 16% were concordant positive [26].

From the concordant positive couples, 47.8% were linked in Zambia (52.2% unlinked) (Figure 2a) and 55.4% were linked in Rwanda (44.6% unlinked) (Figure 3a). From the HT study, linkage assessment up through July 2011 identified 77.9% of recent transmission pairs as linked in Zambia (22.1% unlinked) and 80.5% linked in Rwanda (19.5% unlinked).

Discussion

Given the majority of HIV-1 infections occur between cohabitating heterosexual couples, HIV-1 transmission patterns are particularly important to characterize in sub-Saharan Africa in order to reduce HIV-1 infection risk in this group [1]. Although HIV-1 serodiscordant couples enrolled in CVCT have been extensively studied, concordant positive couples provide an opportunity to assess transmission patterns that are more representative of the general population. For the longitudinally followed HIV-1 positive couples, genetic change within gp41 occurred at only 0.28 (0.12, 0.45) to 0.33 (0.16, 0.51) % PWD per year between Zambia and Rwanda, confirming the validity of the

current PWD linkage cutoffs for up to approximately twelve years of cohabitation in Zambia and nine years in Rwanda. When the PWD cutoffs were applied to the concordant positive couples, the results matched gold standard phylogenetic analysis with only a few discrepancies. Given gp41 PWD cutoffs are valid for Zambia (subtype C) and Rwanda (subtype A), our analysis supports the application of this epidemiologic linkage technique to concordant positive couples in cohorts of subtype A and C individuals.

As expected due to viral divergence over time, concordant positives across both countries had significantly higher PWD than recent transmission pair data. Among linked concordant positive couples, a clear positive correlation between PWD and time of cohabitation occurred as expected. The linked concordant positive couples remained well below the PWD linkage cutoffs despite up to approximately 45 and 30 years of cohabitation in which genetic divergence could occur for Zambian and Rwandan couples, respectively. We might expect that the longer a couple has been together the more likely they may be to have the same virus under a few scenarios.

For those couples who were initially concordant positive yet unlinked, they could experience superinfection. We know superinfection occurs at a rate similar to that of primary infection at approximately 13.6% per year, suggesting the concordant positive couples are experiencing similar patterns [22]. Since PCR amplifies only one or a few copies of RNA, a couple may appear linked if the amplified virus happened to be transmitted by the other partner. Couples who were initially serodiscordant and then transmitted to their partner would have greater similarity in their viruses, particularly if transmission did not occur until after a few years of cohabitation. For those couples who were initially concordant negative but one partner became infected outside the

partnership, a recent transmission event would lead to a concordant positive couple having similar viruses despite a long time of cohabitation.

In some cases, the longer a couple has been together the more likely we might be to see different viruses. Since linkage was assessed using only a 399 base pair segment of gp41, there are limitations in our ability to detect recombination, hypermutation and superinfection. The longer a person has been infected, the more likely some level of recombination has occurred. If two different yet related viral strains coinfect a cell, recombination can occur, changing the appearance of one partner's primary viral infection and making a couple appear unlinked [27]. If hypermutation occurred within one partner of a linked couple, viral change might have been enough to make the couple appear unlinked. With a high infection rate of cells, frequent replication and a relatively rapid mutation rate, mutation accumulations can occur relatively quickly and lead to viral change [28]. Immune escape also may pressure viral diversification faster in one partner than another depending on the specific host immune system. With a wide variety of mechanisms with the potential to increase viral diversity, initially linked couples may appear unlinked over time.

Among Zambian concordant positive couples there were two distinct linear clusters among the linked couples suggesting two subpopulations within the group (Figure 2b). The first group clusters along a linear slope of approximately 4 % PWD change per year similar to what we observed among longitudinally followed transmission pairs. The second group clusters along a line with a slope close to one, suggesting these transmission events may not have occurred when cohabitation began but after at least a few years of cohabitation. Among the second group, it is likely one partner was infected

outside the marriage and then infected their partner. Otherwise, we would have expected transmission to occur earlier given serodiscordant pairs who have not received counseling would be expected to transmit at 20-25% per year [6, 7].

Through applying epidemiologic linkage techniques to concordant positive couples, we are able to expand our understanding of transmission patterns across the general population that has not undergone CVCT. Among the concordant positive couples, a greater proportion of couples are unlinked (Zambia: 52.2%, Rwanda: 44.6%) than recent transmission pairs from the HT study (Zambia: 22.1% Rwanda: 19.5%). As discussed previously, there are a wide variety of mechanisms that might lead previously linked couples to appear unlinked. If this proportion of unlinked concordant positives is not an artifact of genetic diversification, then concordant positive couples may be a particularly important intervention target given superinfection is more likely to occur among unlinked couples, particularly those with continued extramarital sexual encounters [22].

The linkage patterns observed in the concordant positive couples might be understood given the known prevalence of HIV-1 in urban Zambia and Rwanda. The prevalence of HIV-1 infected in urban areas of Zambia in 2001 was 19.2% in men (ages 15-59) and 26.3% in women (ages 15-49) [2]. If more people in Zambia are independently infected before marriage occurs, it would make sense that the proportion of linked couples in the concordant positive individuals would be less than Rwanda since the prevalence in urban Rwanda in 2005 was much lower (8.6% in women and 5.8% in men) [1]. If fewer people are entering marriage infected, then serodiscordance within the

relationship may be having a greater influence on transmission and leading to a higher proportion of linked couples.

Through this study, we have established that epidemiologic linkage techniques utilizing a 399 base pair region of gp41 can be accurately applied to HIV-1 concordant positive couples in Zambia and Rwanda. By characterizing the linkage patterns of concordant positive couples, we have a sense of linkage patterns occurring in the general population and where future behavioral interventions might be targeted.

Acknowledgements

We would like to thank all volunteers and staff of the Rwanda Zambia HIV research group, at the Projet San Francisco in Kigali, Rwanda and the Zambia Emory HIV Research Project in Lusaka, Zambia and the technicians and students of the Hunter Laboratory at Emory University for completing genetic sequencing and transmission linkage.

References

1. UNAIDS, Global Report: UNAIDS Report on the Global AIDS Epidemic, 2010.
2. DHS, HIV/AIDS Data from the 2007 Zambia Demographic and Health Survey 2007, CSO and Macro International Inc.: Calverton, Maryland, USA.
3. Dunkle, K.L., et al., New heterosexually transmitted HIV infections in married or cohabiting couples in urban Zambia and Rwanda: an analysis of survey and clinical data. Lancet, 2008. 371(9631): p. 2183-91.
4. Allen, S., et al., Promotion of couples' voluntary counselling and testing for HIV through influential networks in two African capital cities. BMC Public Health, 2007. 7: p. 349.
5. Guthrie, B.L., G. de Bruyn, and C. Farquhar, HIV-1-discordant couples in sub-Saharan Africa: explanations and implications for high rates of discordancy. Curr HIV Res, 2007. 5(4): p. 416-29.
6. Allen, S., et al., Effect of serotesting with counselling on condom use and seroconversion among HIV discordant couples in Africa. BMJ, 1992. 304(6842): p. 1605-9.
7. Hira, S.K., et al., Epidemiology of human immunodeficiency virus in families in Lusaka, Zambia. J Acquir Immune Defic Syndr, 1990. 3(1): p. 83-6.
8. Fideli, U.S., et al., Virologic and immunologic determinants of heterosexual transmission of human immunodeficiency virus type 1 in Africa. AIDS Res Hum Retroviruses, 2001. 17(10): p. 901-10.
9. Kamenga, M., et al., Evidence of marked sexual behavior change associated with low HIV-1 seroconversion in 149 married couples with discordant HIV-1

serostatus: experience at an HIV counselling center in Zaire. AIDS, 1991. 5(1): p. 61-7.

10. Chomba, E., et al., Evolution of couples' voluntary counseling and testing for HIV in Lusaka, Zambia. J Acquir Immune Defic Syndr, 2008. 47(1): p. 108-15.
11. Rwanda Zambia HIV Research Group. 2005 [cited 2011 July 6]; Available from: http://www.rzhrg.org/index.html.
12. Trask, S.A., et al., Molecular epidemiology of human immunodeficiency virus type 1 transmission in a heterosexual cohort of discordant couples in Zambia. J Virol, 2002. 76(1): p. 397-405.
13. Yang, C., et al., Detection of phylogenetically diverse human immunodeficiency virus type 1 groups M and O from plasma by using highly sensitive and specific generic primers. J Clin Microbiol, 1999. 37(8): p. 2581-6.
14. Wolfs, T.F., et al., HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. Virology, 1992. 189(1): p. 103-10.
15. Vartanian, J.P., et al., Selection, recombination, and G----A hypermutation of human immunodeficiency virus type 1 genomes. J Virol, 1991. 65(4): p. 1779-88.
16. Hu, W.S. and H.M. Temin, Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proc Natl Acad Sci U S A, 1990. 87(4): p. 1556-60.
17. Kijak, G.H. and F.E. McCutchan, HIV diversity, molecular epidemiology, and the role of recombination. Curr Infect Dis Rep, 2005. 7(6): p. 480-8.
18. Jung, A., et al., Recombination: Multiply infected spleen cells in HIV patients. Nature, 2002. 418(6894): p. 144.

19. Smith, D.M., D.D. Richman, and S.J. Little, HIV superinfection. J Infect Dis, 2005. 192(3): p. 438-44.

20. Allen, T.M. and M. Altfeld, HIV-1 superinfection. J Allergy Clin Immunol, 2003. 112(5): p. 829-35; quiz 836.

21. van der Kuyl, A.C. and M. Cornelissen, Identifying HIV-1 dual infections. Retrovirology, 2007. 4: p. 67.

22. Kraft, C.S., et al., Timing and source of subtype-C HIV-1 superinfection in the newly infected partner of Zambian couples with disparate viruses. Retrovirology, 2012. 9(1): p. 22.

23. Berkhout, B., A.T. Das, and N. Beerens, HIV-1 RNA editing, hypermutation, and error-prone reverse transcription. Science, 2001. 292(5514): p. 7.

24. Central Statistical Office (CSO), M.o.H.M., Tropical Diseases Research Centre (TDRC), University of Zambia, and Macro International Inc. 2009. , Zambia Demographic and Health Survey 2007, CSO and Macro International Inc.: Calverton, Maryland, USA.

25. Skurnick, J.H., et al., Behavioral and demographic risk factors for transmission of human immunodeficiency virus type 1 in heterosexual couples: report from the Heterosexual HIV Transmission Study. Clin Infect Dis, 1998. 26(4): p. 855-64.

26. Roth, D.L., et al., Sexual practices of HIV discordant and concordant couples in Rwanda: effects of a testing and counselling programme for men. Int J STD AIDS, 2001. 12(3): p. 181-8.

27. Burke, D.S., Recombination in HIV: an important viral evolutionary strategy. Emerg Infect Dis, 1997. 3(3): p. 253-9.

28. Coffin, J.M., HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science, 1995. 267(5197): p. 483-9.

Figures

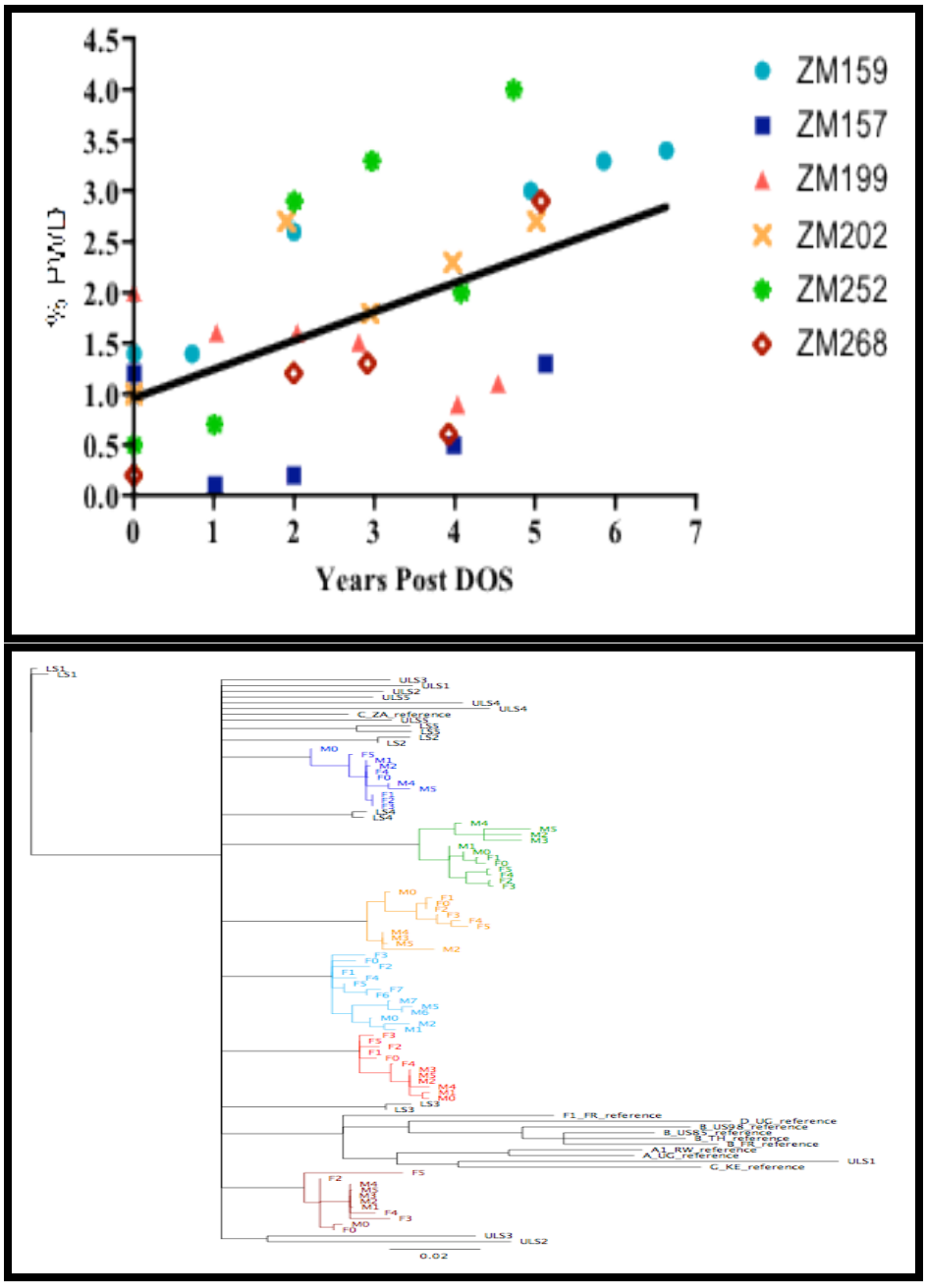


Figure 1a: Donor and recipient gp41 pair-wise distance (PWD) comparison after day of seroconversion (DOS) for Zambian seroconverters followed up to 7 years. PWD values increased over time with PWD values never closely approaching the 5.4% linkage cutoff. For every 1 year, there is an approximate increase in PWD of 0.28 (95% CI: 0.12, 0.45) percent. Linear Regression: y=0.28x+0.97. (top) The neighbor-joining phylogenetic tree for the Zambian seroconverters identifies all 6 transmission pairs (color coded) and years post DOS. All couples remain clustered throughout the entire period of follow-up. (bottom)

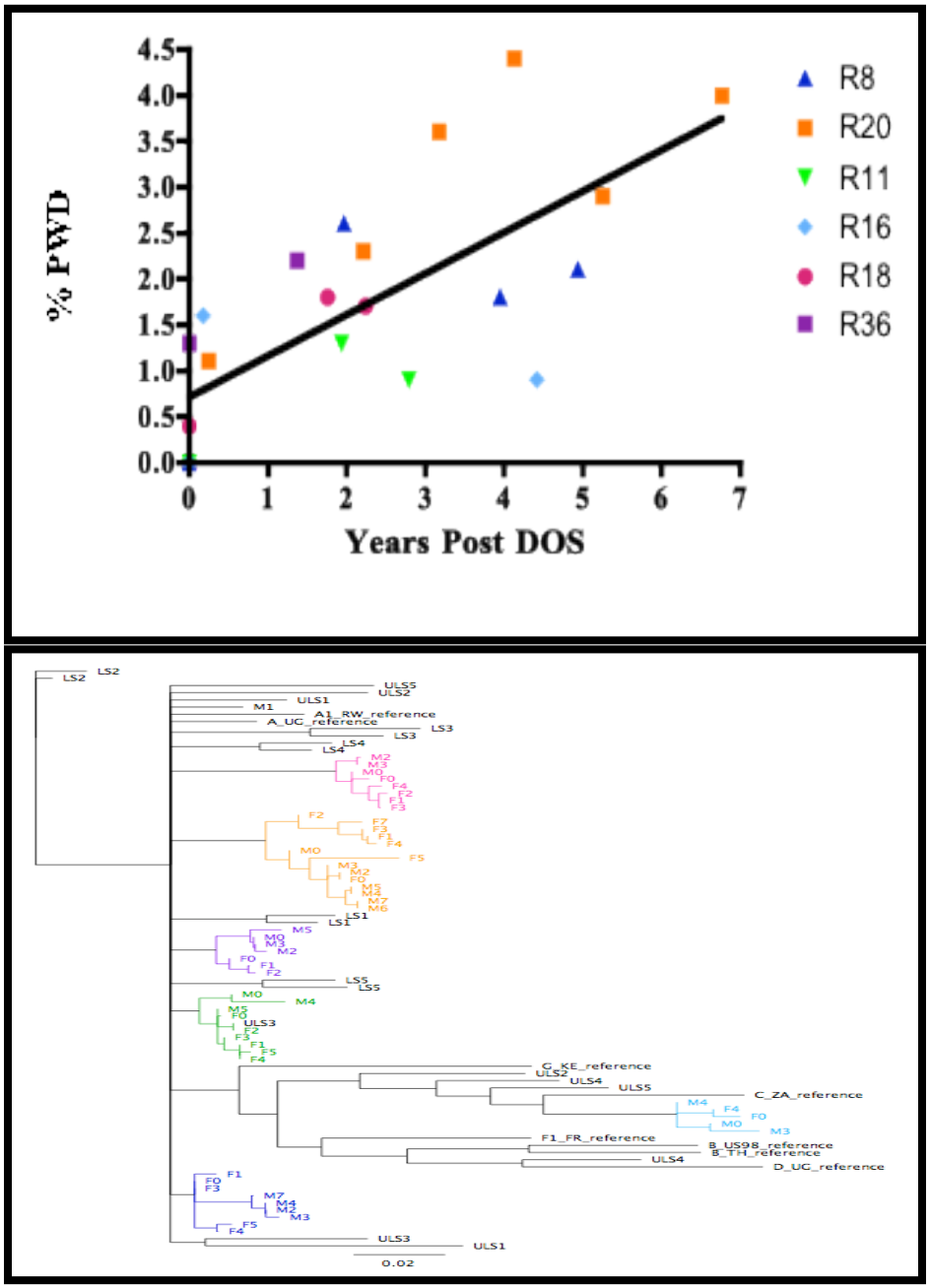


Figure 1b: Donor and recipient gp41 pair-wise distance (PWD) comparison as the viruses diversified after day of seroconversion (DOS) for Rwandan seroconverters followed up to 7 years. PWD values increased over time with PWD closely approaching the 4.6% linkage cutoff. For every 1 year, there is an approximate increase in PWD of 0.33 (95% CI: 0.16, 0.51) percent. Linear Regression: y=0.33x+0.78 (top) The neighbor-joining phylogenetic tree for the Rwandan seroconverters identifies all 6 transmission pairs (color coded) and years post DOS. All couples remain clustered throughout the entire period of follow-up. (bottom)

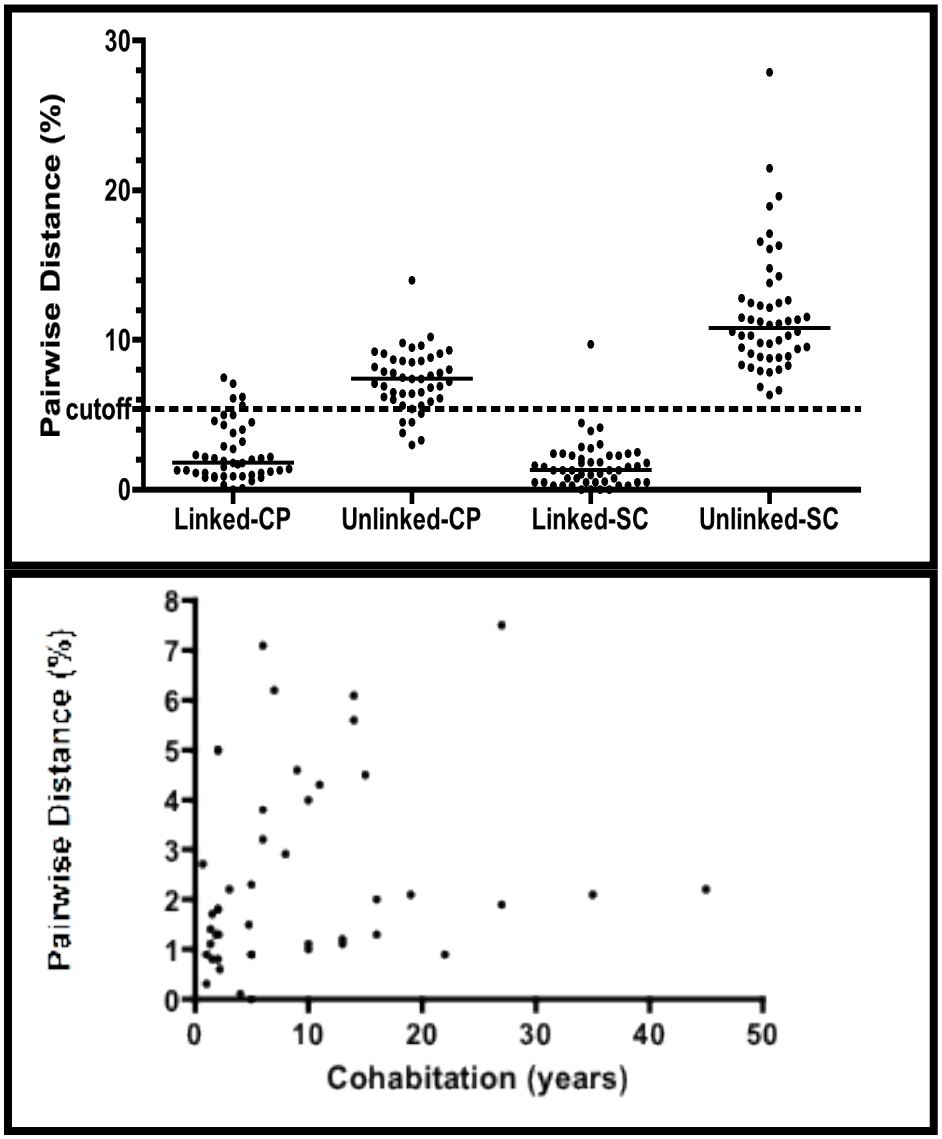


Figure 2a: Comparison of linked (N=43) and unlinked (N=47) Zambian HIV-1 concordant positives (CP) with linked (N=50) and unlinked (N=50) seroconverting (SC) Zambian transmission pairs in relation to the 5.4% PWD linkage cutoff for transmission pairs. The median PWD for the linked CP (1.80%) and SC (1.30%) transmission pairs is significantly different (p=0.014, one-sided Mann-Whitney test) and the median PWD of unlinked CP (7.40%) and SC (10.78%) transmission pairs is also significantly different (p<0.0001, Mann-Whitney test).

Figure 2b: The PWD and cohabitation time comparison for linked Zambian concordant positive couples showed a slight increase in PWD over time with a significant correlation between PWD and cohabitation time (p=0.023, Spearman rank correlation).

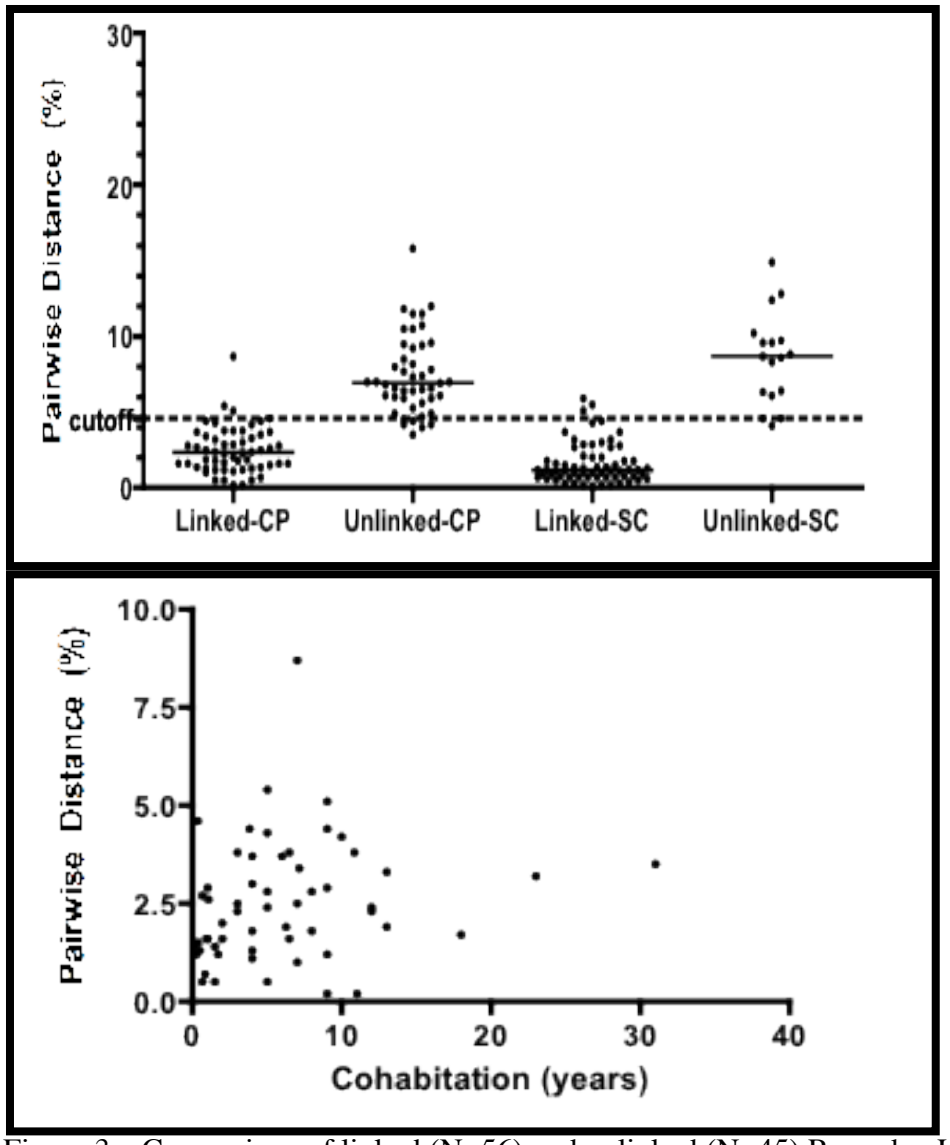
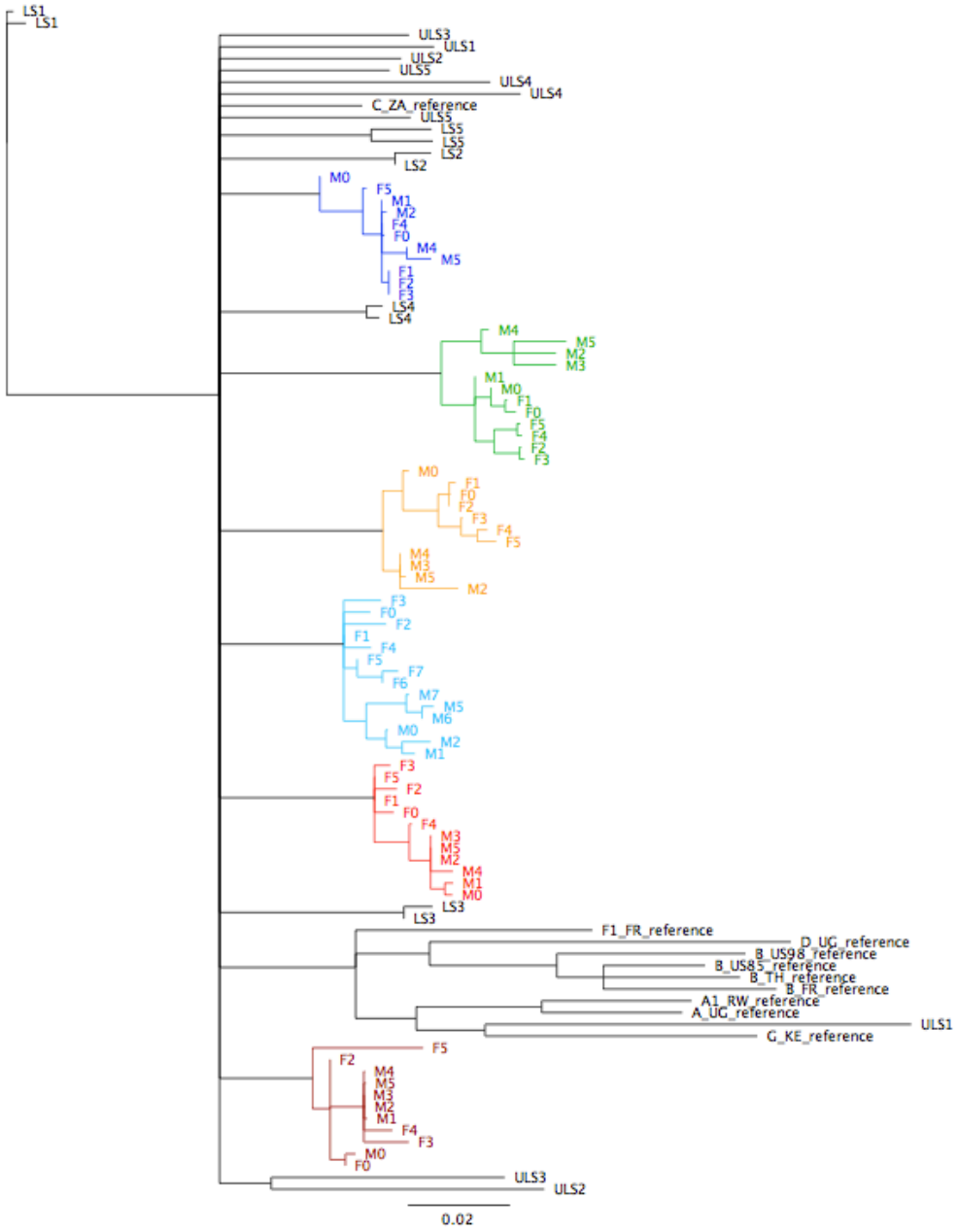


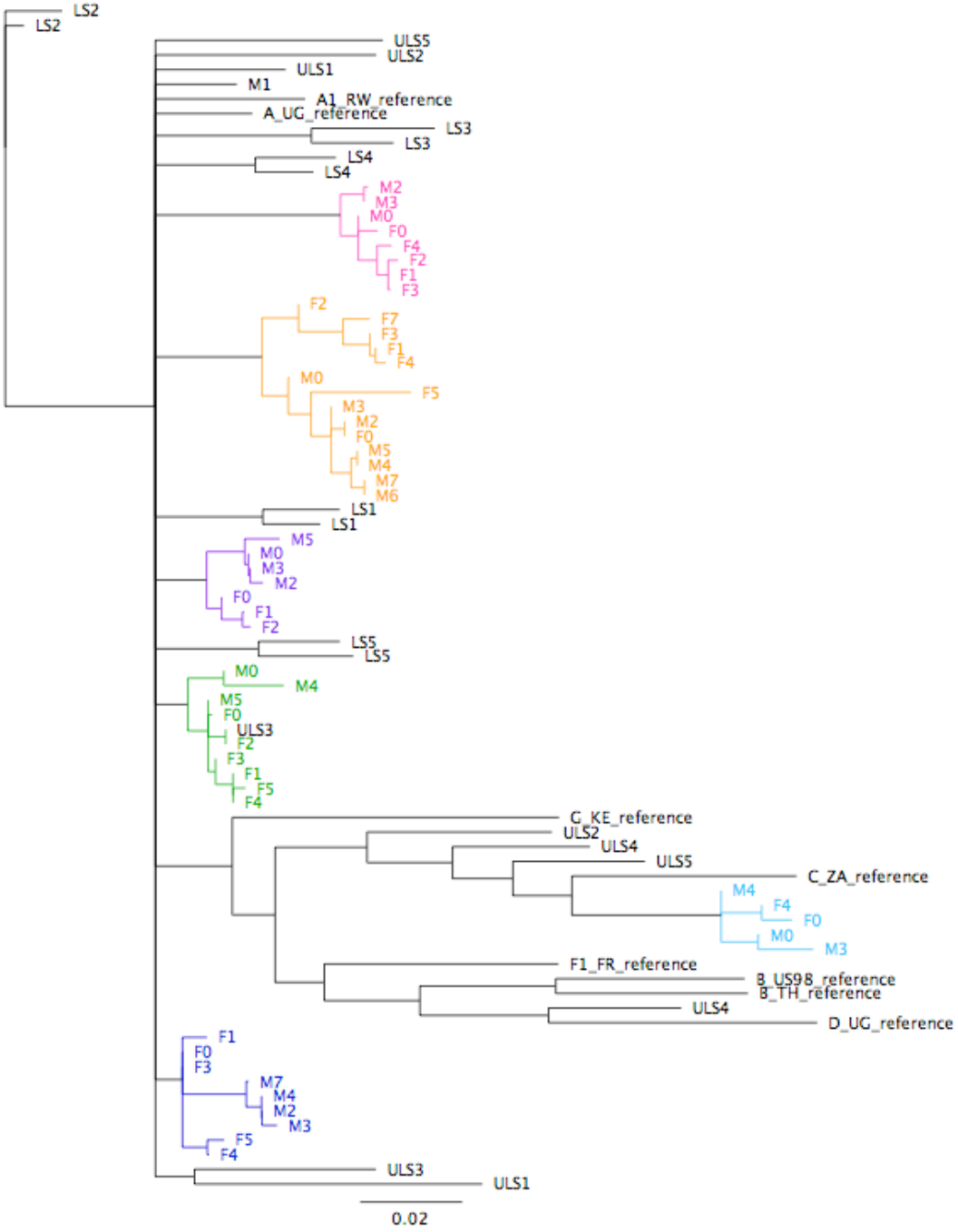
Figure 3a: Comparison of linked (N=56) and unlinked (N=45) Rwandan HIV-1 concordant positives (CP) with linked (N=66) and unlinked (N=17) seroconverting (SC) Rwandan transmission pairs in relation to the 4.6% PWD linkage cutoff for transmission pairs. The median PWD for the linked CP (2.35%) pairs is significantly different than the SC (0.70%) transmission pairs ($p<0.0001$, one-sided Mann-Whitney test). The median PWD for the unlinked CP (6.90%) pairs is slightly higher but not significantly different than the SC (6.20%) transmission pairs ($p>0.05$, Mann-Whitney test).

Figure 3b: The PWD and cohabitation time comparison for Rwandan concordant positive couples showed a slight increase in PWD over time with a significant correlation between PWD and cohabitation time ($p=0.037$, Spearman rank correlation).

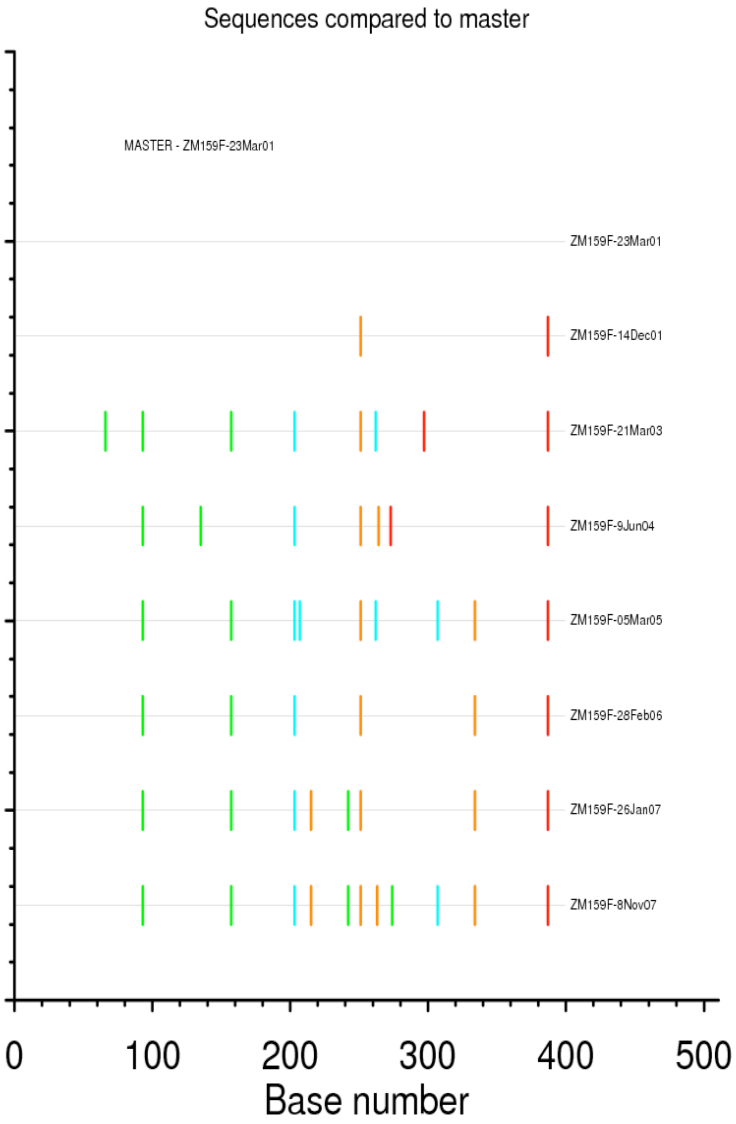
Supplemental Figures



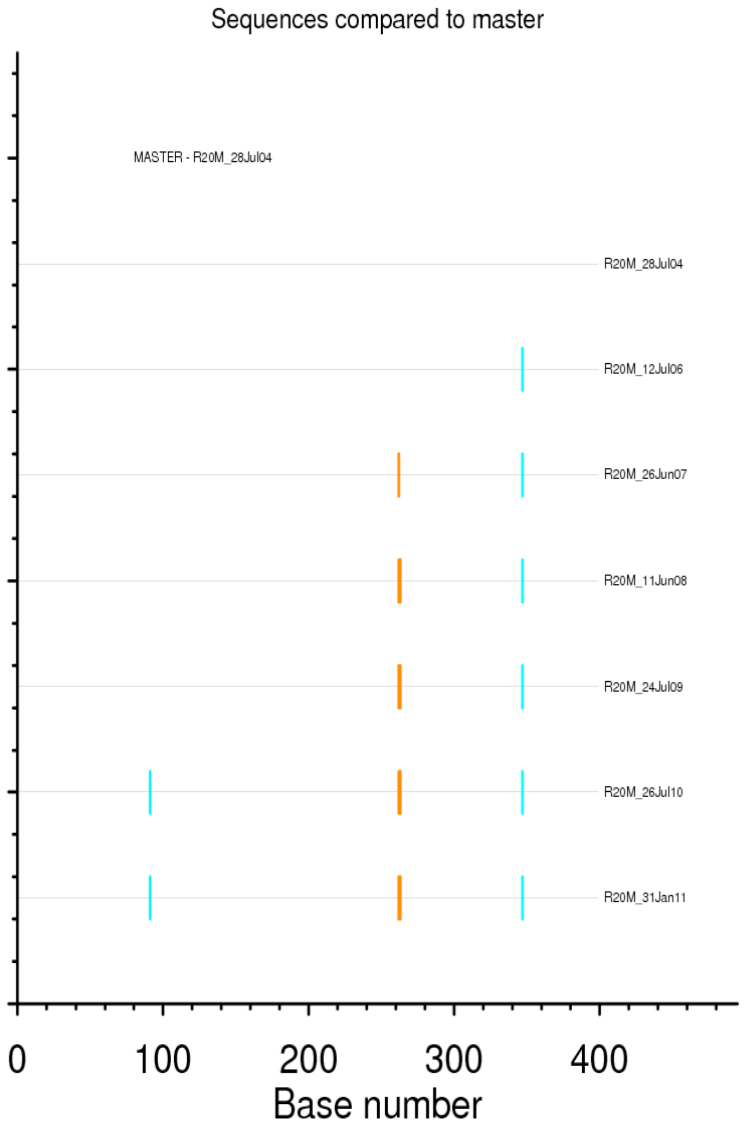
Supplemental Figure 1a: The neighbor-joining phylogenetic tree for the Zambian seroconverters identifies all 6 transmission pairs (color coded with Figure 1a) and years post DOS. All couples remain clustered throughout the entire period of follow-up. Horizontal branch lengths are drawn to scale (the scale bar represents 0.02 nucleotide substitutions per site).



Supplemental Figure 1b: The neighbor-joining phylogenetic tree for the Rwandan seroconverters identifies all 6 transmission pairs (color coded with Figure 1b) and years post DOS. All couples remain clustered throughout the entire period of follow-up. Horizontal branch lengths are drawn to scale (the scale bar represents 0.02 nucleotide substitutions per site).

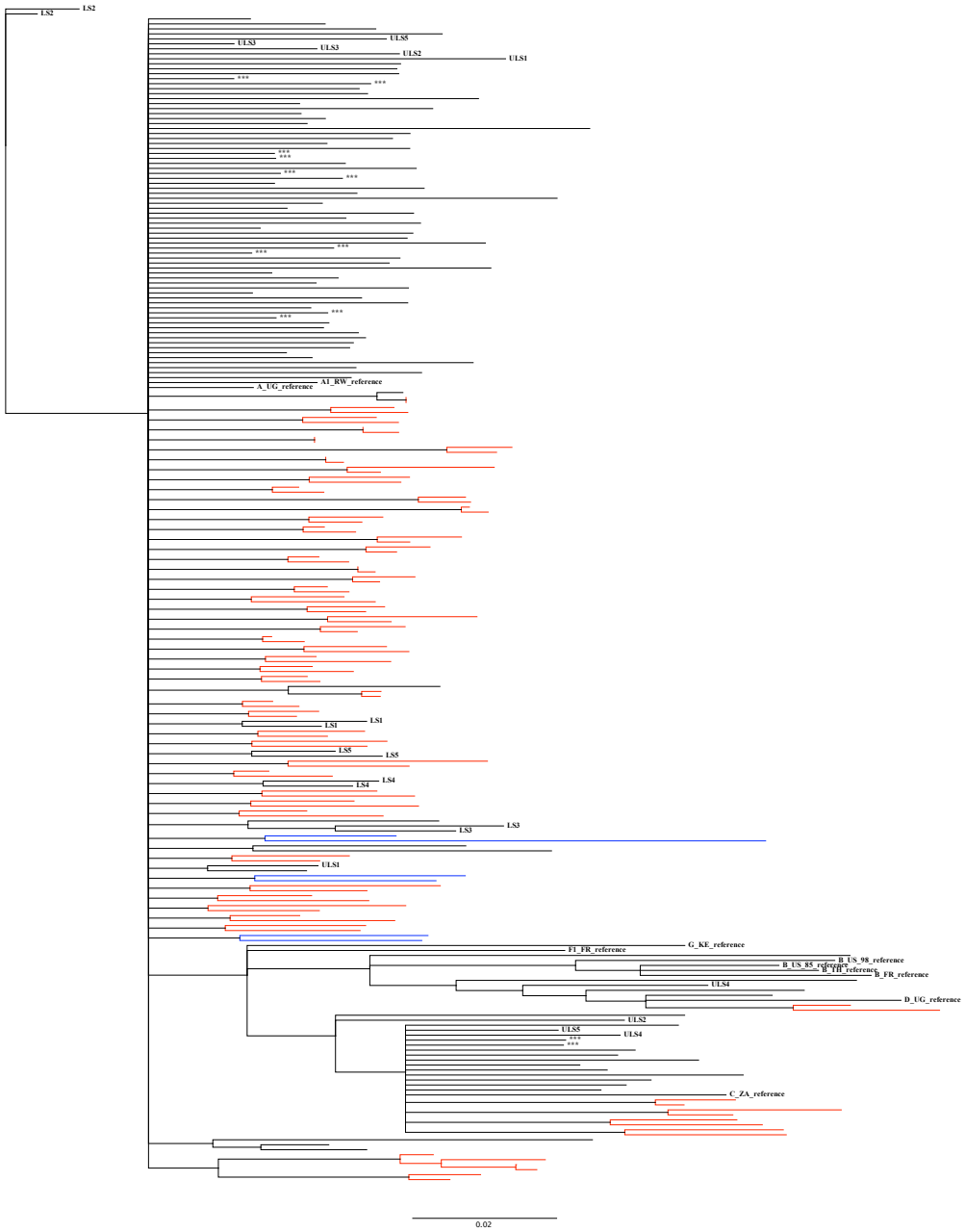


Supplemental Figure 2a: Highlighter plot of a Zambian recipient followed over seven years. Tick marks indicate a nucleotide change from the earliest time point sequence at the respective base number location. (Blue: Cytosine, Red: Thymine, Orange: Guanine, Green: Adenine) (http://www.hiv.lanl.gov/content/hiv-db/HIGH-LIGHT/highlighter.html)



Supplemental Figure 2b: Highlighter plot of a Rwandan recipient followed over seven years. Tick marks indicate a nucleotide change from the earliest time point sequence at the respective base number location. (Blue: Cytosine, Red: Thymine, Orange: Guanine, Green: Adenine) (http://www.hiv.lanl.gov/content/hiv-db/HIGH-LIGHT/highlighter.html)

Supplemental Figure 3a: Phylogenetic tree of Zambian concordant positive couples' (N=90) partial gp41 sequences (399bp) by using the neighbor-joining method. Horizontal branch lengths are drawn to scale (the scale bar represents 0.02 nucleotide substitutions per site). Red branches represent linked concordant positive couples by the phylogenetic tree and the PWD cutoff. Blue branches represent couples linked by the phylogenetic tree but unlinked by the PWD cutoff. Branch tips with three stars indicate couples unlinked by the phylogenetic tree but linked by the PWD cutoff. The 10 clade reference sequences were obtained from the Los Alamos Sequence Database (http://hiv-web.lanl.gov/HRML/alignments.html). Reference sequences for 10 known linkage pairs (five unlinked and five linked) couples from the CVCT cohort in Zambia were included.



Supplemental Figure 3b: Neighbor-joining phylogenetic tree of Rwandan concordant positive couples' (N=101) partial gp41 sequences (399bp). Horizontal branch lengths are drawn to scale (the scale bar represents 0.02 nucleotide substitutions per site). Red branches represent linked concordant positive couples by the phylogenetic tree and the PWD cutoff. Blue branches represent couples linked by the phylogenetic tree but unlinked by the PWD cutoff. Branch tips with three stars indicate couples unlinked by the phylogenetic tree but linked by the PWD cutoff. The 10 clade reference sequences were obtained from the Los Alamos Sequence Database (http://hiv-web.lanl.gov/HRML/alignments.html). Reference sequences for 10 known linkage pairs (five unlinked and five linked) couples from the CVCT cohort in Rwanda were included.