Diversity and Tropism of HIV-1 Plasma Rebound Virus after Treatment Discontinuation

By:
Blake M. Hauser

Senior Honors Thesis
Department of Biology
College of Arts and Sciences
The University of North Carolina at Chapel Hill
April 20, 2016

Approved:

Ron Swanstrom

Dr. Ronald Swanstrom, Thesis Advisor
Distinguished Professor of Biochemistry and Biophysics

ABSTRACT

Modern antiretroviral therapies can confer effective suppression of HIV-1 infection. However, HIV-infected people discontinuing antiretroviral therapy experience a rebound of virus from a persistent reservoir. Characterizing this reservoir constitutes a crucial step towards developing a cure, and we hypothesized that assessing the diversity and tropism of rebound virus would provide insight into the types of cells that likely house the viral reservoir as well as reservoir diversity. We examined 10 rebound samples from a project within the large-scale AIDS Clinical Trials Group and used single genome amplification and Primer ID deep sequencing to assess the genetic diversity of the *env* gene, which encodes the HIV-1 envelope protein, among rebounding virus. Samples from the same patients with viral suppression due to antiretroviral therapy were also available. Most rebound virus populations showed significant diversity. All *env* genes examined were isoforms that would require cells to express high surface levels of the CD4 receptor for entry, consistent with the current hypothesis that virus is selectively replicated in CD4+ T cells. These results indicate that most people discontinuing therapy release a diverse population of virus, and this released virus targets CD4+ T cells rather than myeloid cells, which tend to last longer. Research indicates that a small proportion of viruses have evolved to efficiently enter myeloid cells. Current HIV-1 cure strategies focus on reactivating latent HIV-1 and targeting the resultant virus. It is essential to understand the features of rebound viruses because they are the first such targets.

INTRODUCTION

HIV-1 causes approximately 6,300 new infections daily. Successful antiretroviral therapy (ART) reduces HIV-1 viral load in the blood down to stable low levels; however ART does not eliminate virus. Developing a cure for HIV-1 constitutes one of the top priorities in the HIV-1 research field today, and such a treatment has the potential to end the dependency on ART that thousands currently have.

Since ART does not fully eliminate HIV-1 infection, most people discontinuing therapy experience a rebound of virus from a persistent reservoir [1-4]. The variable timing of rebound, where the virus is coming from, and the complexity of the rebound virus are poorly understood.

One potential source of virus is latently infected resting CD4+ T cells. These immune cells are well-established targets of HIV-1 infection. There is also a continual production of virus at low levels in spite of ART, and other cell types could be the source of rebound virus. This includes myeloid cells, which are differentiated from the bone marrow and can perform immune functions. In a largely quiescent reservoir, the rebound virus could be clonal as the result of virus from a single cell leaving a latent state. However, the number/complexity of the rebound virus within a person has recently been predicted to be much higher than predicted based on a largely quiescent reservoir [4].

To this end we have examined peripheral blood samples from ACTG study A5068 [5] to investigate the diversity and entry phenotype of rebound virus. Subjects in A5068 discontinued ART as part of an analytic treatment interruption (ATI) and had virologic rebound detectable in the blood on average between 2-4 weeks after discontinuation of therapy. To study the complexity of the rebound virus we assessed the genetic diversity of the viral *env* gene and also examined viral entry phenotype. Phylogenetic analysis showed a high degree of diversity in the

rebound virus population. In addition, all Env proteins examined required high levels of CD4 for efficient entry, consistent with prior selection for replication in CD4+ T cells. Our analysis suggests that a viral population replicating in myeloid cells is not a significant source of rebound virus found early in the blood.

METHODS

Study Population

The AIDS Clinical Trails Group (ACTG) study A5068 examined whether structured treatment interruptions and/or vaccination with an exogenous HIV vaccine that consisted of an ALVAC-HIV vector was associated with control of viral replication during a subsequent analytical treatment interruption. This study included a control group with no vaccination or structured interruption control for comparison (Arm A). The study showed a modest effect in reduced peak and set point viral load of the rebound virus in the arms that included the structured treatment interruptions [5]. Subjects with the analytical treatment interruption only (Arm A) had a median CD4+ T cell count of 779 copies/mL, and viral RNA levels of less than 50 copies/mL at entry into the study while still on therapy. For our study, we used blood plasma samples from 10 of these subjects early after the initiation of the analytical treatment interruption when viral load was at or above 1,000 copies/mL (Figure 6). A pre-therapy sample was also available for one subject.

Single Genome Amplification

Sequences of virion RNA were examined using single genome amplification (SGA) [6] of the full-length *env* gene. SGA aims to isolate cDNA specific to a single virion, allowing researchers to obtain individual sequences corresponding to single virions within a patient sample.

Prior to HIV-1 RNA isolation from blood plasma using the QIAmp Viral RNA Mini Kit (Qiagen), samples were pelleted by centrifugation at 25,000 xg for 2 hours at 4 ºC. Isolated vRNA was reverse transcribed using Superscript III Reverse transcriptase (Invitrogen) and an oligo-d(T) primer. SGA was accomplished by endpoint dilutions and semi-nested PCR using

Platinum Taq High Fidelity polymerase (Invitrogen) and primers B3F3 (5′-TGGAAAGGTGAAGGG GCAGTAGTAATAC-3′) LTR DN (5′-GACTCTCGAGAAGCACTCAAGGCAAGCTTTATTGAG-3′) and B5957 UP1 (5′-GATCAAGCTTTAGGCATCTCCTATGGCAGGAAGAAG-3′) and LTR DN for round one and two of PCR, respectively. Sequencing of individual PCR amplified *env* amplicons was performed, and it was confirmed that no double peaks were present in the sequencing output prior to phylogenetic analysis (as this would have indicated the simultaneous detection of more than one viral sequence).

Next Generation Sequencing using Primer ID

Sequences of virion RNA were further examined using deep sequencing with Primer ID [7] of the V1-V3 region of the *env* gene. Primer ID uses unique barcodes to tag each DNA template during the PCR process [7]. This allows investigators to validate the number of templates initially present in addition to lowering the sequencing error rate to 1 in 10,000 bases [7].

We used the previously described protocol to construct Primer ID sequencing libraries from viral RNA. [7] In brief, the cDNA primer with a block of 9 degenerate nucleotides (Primer ID) was used to tag each viral template with a unique Primer ID. After purification of cDNA using magnetic beads, we used two rounds of PCR to amplify Primer ID tagged cDNA and to add Illumina barcoded adaptor sequences for multiplexing. We used MiSeq 300 bp pair-end platform to sequence the constructed library. Raw sequencing reads were initially de-multiplexed by the Illumina pipeline. We used the previously described in-house pipeline to retrieve template consensus sequences [7].

Determination of Major Lineages

The minimum number of rebound viruses was estimated by analyzing distinct sequence lineages using both the phylogenetic tree analysis and highlighter plots (Figures 1D, 2A, 3, 4) of each subject. Both distinct lineages and complex clusters of linked sequence differences were used. This means that we initially counted major clusters on the phylogenetic trees, and we then used the highlighter plot to analyze the diversity of different regions (5 in total of about 600 basepairs per region) of the *env* gene and to counting clusters of similarly mutated regions. The lowest number of lineages was chosen. This was all done to avoid interpreting complexity that could have been the result of post-rebound recombination. These values are minimum estimates, as we took the lowest number of lineages counted per subject of the rebound virus complexity.

Virus Characterization

Phylogenetic analysis of the full-length *env* genes was used to select a representative sampling of amplicons for cloning to assess viral entry phenotype. *env* genes cloned in an expression vector were used in cotransfection with the pNL4-3.LucR-E reporter backbone plasmid into 293T cells to generate an Env-pseudotyped luciferase reporter virus as described previously [8-10]. By comparing the ability of the pseudotyped virus to infect Affinofile cells expressing high versus low levels of surface CD4, we are able to consistently assess viral CD4 entry phenotype. This allows an assignment of viral entry tropism as being largely restricted to CD4+ T cells when requiring high levels of CD4 for entry (T cell-tropic) versus being able to efficiently infect cells with a low density of CD4, as is found on macrophages (macrophage-tropic) [8]. Sensitivity to soluble CD4 was used to confirm the CD4 entry phenotype. Additionally, coreceptor usage was analyzed using Maraviroc inhibition of CCR5.

RESULTS

We analyzed eleven plasma samples from ten subjects (subject 43656 had a pre-therapy (R_0) and rebound (R1) sample available; Figure 1A). We used plasma from the first available time point with a viral load above 1,000 copies/mL to focus on the early rebound virus. Most samples had viral loads less than 10,000 copies/mL (Figure 1B). Eight of ten subjects had nadir CD4+ T cell counts prior to therapy recorded as between 201-500 cells/mL, while one (sample R10) was <201 cells/mL.

In order to determine the viral sequence diversity of the rebound samples, we employed deep sequencing using Primer ID to sequence the V1-V3 region of the *env* gene (Figure 1D), in addition to doing SGA of the full length *env* gene. Deep sequencing results were achieved for eight of ten rebound samples and the pre-therapy sample (however, the matching rebound sample R1 was unsuccessful). We estimated the minimum number of rebound viruses as distinct sequence lineages in the phylogenetic tree analysis and highlighter plots (Figures 1D, 2A, 3, 4). We used both distinct lineages and complex clusters of linked sequence differences in forming this estimate while trying to avoid interpreting complexity that could have been the result of post-rebound recombination. We consider these values as minimum estimates of the rebound virus complexity since the viruses considered as potential recombinants could have pre-existed in the reservoir and been independently reactivated. Analysis of the minimum number of lineages versus time (in weeks) to rebound demonstrated no obvious correlation between time to rebound and complexity of the rebound virus population (Figure 1C). For virus that rebounded in the first few weeks we estimated between 1 and 5 distinct lineages by deep sequencing, with a mean of 3.1 lineages, for these early rebound viruses. R2 took the longest to rebound (10 weeks) and showed relatively low diversity with only 2 lineages. SGA results of full length *env* genes

showed between 1 and 4 distinct lineages, with a mean of 3.1 lineages, which is similar to the results obtained with deep sequencing (Figure 1C). The mean intrapatient diversity was also determined by pairwise analysis using full-length *env* and ranged from 0.2% to 3.1%. Additional analyses of subjects early in infection (10 subjects) [11] and chronically infected subjects (7 subjects, unpublished observation) were also done and ranged between 0.16% to 1.2% for early infection subjects and 1.1% to 4.1% for chronically infected subjects not on therapy. These numbers suggest that the rebound virus is more similar to virus in chronically infected subjects where multiple viral lineages form a genetically complex population.

We subsequently cloned individual *env* genes (51 total; average 4 amplicons per subject) from the major lineages in each participant to determine the entry phenotype of the rebound virus using the Affinofile cell line (Figures 2A, 4). The pre-therapy sample (R_0) was also included; no clones were obtained for sample R5. Affinofile cells allow controlled expression of CD4 and are currently the most consistent method to assess CD4-dependent entry phenotype [8]. Entry phenotype was evaluated based on the ability of viruses pseudotyped with the Env proteins from the rebound virus to infect Affinofile cells expressing either high or low levels of surface CD4. The phenotype was confirmed by sensitivity to inhibition by soluble CD4 at a concentration of 5.0 μg/mL when infecting TZM-bl cells [12], a level of soluble CD4 that neutralizes macrophage-tropic viruses but not R5 T cell-tropic viruses. None of the Env proteins tested showed any ability to infect cells expressing low levels of CD4 compared to their infectivity on cells expressing high levels of CD4 as all exhibited less than 2% relative infectivity (Figure 2B). All Env proteins tested were also resistant to inhibition by soluble CD4 (Figure 2D). In addition, the CCR5 antagonist Maraviroc inhibited all pseudotyped viruses. Therefore, all isolated Env

proteins used CCR5 as the coreceptor. Taken together, all Env proteins tested from the rebound virus of all subjects had a phenotype consistent with R5 T cell-tropic virus.

DISCUSSION

In this report we describe entry phenotype and population complexity of the initial virus that rebounds from the latent or low-level viremia reservoir after therapy discontinuation. These features are relevant to developing strategies of HIV-1 eradication because they are features of the first virus that must be targeted. We evaluated entry phenotype in order to determine what cell type this virus had been predominantly replicating in at the time therapy was initiated and/or when the virus entered the latent reservoir. In addition, we examined the complexity of the rebound population as a measure of the minimum number of viruses that rebound from the reservoir and can be detected in the blood. The genetic diversity of the viral reservoir and rebound virus has been debated, but it is perceived to have low diversity and to be primarily clonal or multiclonal [13, 14]. Our results are consistent with the observation of Rothenberger et al.[4] who proposed a multifocal model of virus reactivation. We observed a mean of 3.1 distinct lineages when counting major lineages with both deep sequencing and SGA. This points to a minimum of 3 viruses rebounding within a short period after treatment interruption. While this is significantly less than the 5-25 rebound/founder viruses suggested by Rothenberger et al. [4], when we reanalyzed the sequence data from that study for the 3 subjects with SGA of virion RNA using the same criteria as used in this report, we found the mean number of distinct lineages to be 3.3. This is similar to what we observed. However, since we did not count potential recombinants in our analysis, this is a minimum estimate.

The number of rebound viruses could be affected by factors that affect viral reservoir size, such as duration of ART and time after infection of ART initiation. These data are not available for the subjects included in this report. Reservoir size, as measured by cell-associated viral RNA and DNA, has been shown to be correlated with set point after therapy

discontinuation [15]. Sample R2 did show one of the lowest sequence diversities (0.59%) and took the longest to have a measurable rebound virus population, but we only analyzed one subject who took longer than four weeks for rebound to be detected.

The use of viral outgrowth assays from subjects on therapy after isolation of resting CD4+ T cells and reactivation in culture has established this cell type as a reservoir for HIV-1 latency [1-3]. We sought to evaluate the rebound virus population by using entry phenotype. Distinct lineages of HIV-1 can evolve the ability to use low levels of CD4 for entry, but these have been found primarily in the cerebral spinal fluid [8,11]; presumably, this localized evolution is due to selective pressure to enhance replication in cells such as macrophages, which express low levels of CD4. Here, we assessed the CD4 entry phenotype in conjunction with soluble CD4 sensitivity (a feature of low CD4-using Envs) to determine that the rebound viruses needed high levels of CD4 for efficient entry (Figure 2B). Sensitivity to inhibition by Maraviroc of all Env proteins showed that all of the viruses were CCR5-using (Figure 2C), making these viruses R5 T cell-tropic. Based on this phenotypic characterization, we can conclude that the rebound virus was adapted to replicating in CD4+ T cells at the time the reservoir was formed. These results do not define the cell type that harbored the latent virus prior to rebound (T cells or macrophages); however, these results exclude the possibility that the rebound virus was replicating extensively in a myeloid cell lineage prior to latency since this would have selected for the ability to use low levels of CD4 for entry. It should be noted that viral rebound in isolated compartments of the body, such as the brain, or viruses released later from smaller reservoirs, could have phenotypes that are distinct from the initial rebound virus that predominates in the blood.

In summary, we have shown that the early rebounding virus is more genetically diverse than previously thought. Additionally, we were able to show that all virus isolates from the plasma used the CCR5 co-receptor in the cell entry process and required high levels of CD4 for entry. Therefore, these virus isolates are R5 T cell-tropic. This excludes a population of virus replicating in myeloid cells as a significant contributor to the viral reservoir that repopulates blood early after therapy discontinuation. Elucidating the nature of this viral reservoir constitutes a crucial prerequisite to eliminating the latent reservoir, which will be required in order to cure HIV. By providing considerable insight into the genetic diversity of rebounding virus populations in addition to characterizing the tropism phenotypes of this virus, our findings also lay the groundwork for future research aimed at targeting the rebound virus population for eradication.

References

1. Finzi D, Hermankova M, Pierson T, et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science **1997**; 278:1295-300.

2. Wong JK, Hezareh M, Gunthard HF, et al. Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. Science **1997**; 278:1291-5.

3. Finzi D, Blankson J, Siliciano JD, et al. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. Nat Med **1999**; 5:512-7.

4. Rothenberger MK, Keele BF, Wietgrefe SW, et al. Large number of rebounding/founder HIV variants emerge from multifocal infection in lymphatic tissues after treatment interruption. Proc Natl Acad Sci U S A **2015**; 112:E1126-34.

5. Jacobson JM, Pat Bucy R, Spritzler J, et al. Evidence That Intermittent Structured Treatment Interruption, but Not Immunization with ALVAC-HIV vCP1452, Promotes Host Control of HIV Replication: The Results of AIDS Clinical Trials Group 5068. Journal of Infectious Diseases **2006**; 194:623-32.

6. Salazar-Gonzalez JF, Bailes E, Pham KT, et al. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. Journal of Virology **2008**; 82:3952-70.

7. Zhou S, Jones C, Mieczkowski P, Swanstrom R. Primer ID Validates Template Sampling Depth and Greatly Reduces the Error Rate of Next-Generation Sequencing of HIV-1 Genomic RNA Populations. Journal of Virology **2015**; 89:8540-55.

8. Joseph SB, Arrildt KT, Swanstrom AE, et al. Quantification of entry phenotypes of macrophage-tropic HIV-1 across a wide range of CD4 densities. J Virol **2014**; 88:1858-69.

9. Ping LH, Cohen MS, Hoffman I, et al. Effects of genital tract inflammation on human immunodeficiency virus type 1 V3 populations in blood and semen. J Virol **2000**; 74:8946-52.

10. Keele BF, Giorgi EE, Salazar-Gonzalez JF, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A **2008**; 105:7552-7.

11. Sturdevant CB, Joseph SB, Schnell G, Price RW, Swanstrom R, Spudich S. Compartmentalized replication of R5 T cell-tropic HIV-1 in the central nervous system early in the course of infection. PLoS Pathog **2015**; 11:e1004720.

12. Montefiori DC. Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. Curr Protoc Immunol **2005**; Chapter 12:Unit 12 1.

13. Kearney MF, Spindler J, Shao W, et al. Lack of detectable HIV-1 molecular evolution during suppressive antiretroviral therapy. PLoS Pathog **2014**; 10:e1004010.

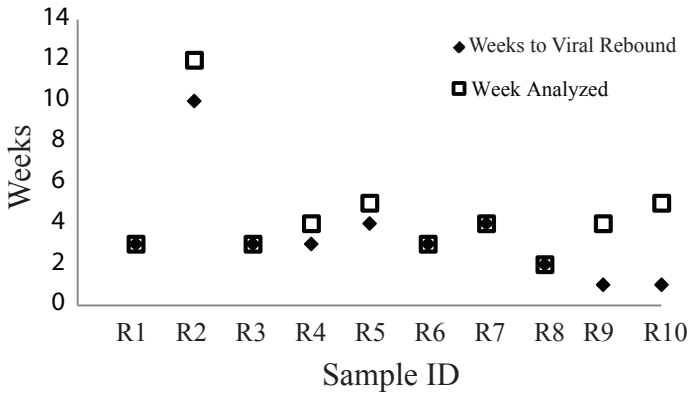
14. Joos B, Fischer M, Kuster H, et al. HIV rebounds from latently infected cells, rather than from continuing low-level replication. Proceedings of the National Academy of Sciences **2008**; 105:16725-30.

15. Li JZ, Heisey A, Ahmed H, et al. Relationship of HIV reservoir characteristics with immune status and viral rebound kinetics in an HIV therapeutic vaccine study. AIDS **2014**; 28:2649-57.

A. Sample and Patient information

Sample ID	PID	Nadir CD4 Cell Count (categorized)	CD4 at Time of Sample	VL of Sample
R_0	43656	201-500		11337
R1	43656	201-500	471	1226
R2	291131	201-500	807	8267
R3	621260	201-500	798	1738
R4	141496	201-500	539	1911
R5	621243	201-500	455	3437
R6	621312	201-500	432	1335
R7	291180	201-500	1000	2525
R8	83111	Data Not Avalible	1287	1317
R9	111899	201-500	578	4691
R10	621388	<201	640	10068

B. Sample Analyzed Compared to Rebound Occurance



C. Minimum Number of Viral Lineages vs. Time to Rebou

Sample ID	Weeks After Therapy Interuption	Minimum Number of Major Lineages in Deep Sequencing Tree	Minimum Number of Major Lineages in SGA Tree
R_0	N/A	5	5
R1	3		3
R2	10	1	2
R3	3	5	5
R4	3	3	3
R5	4	5	4
R6	3	1	1
R7	4		3
R8	2	3	3
R9	1	3	3
R10	1	4	4

D. Deep Sequencing Phylogenetic Trees

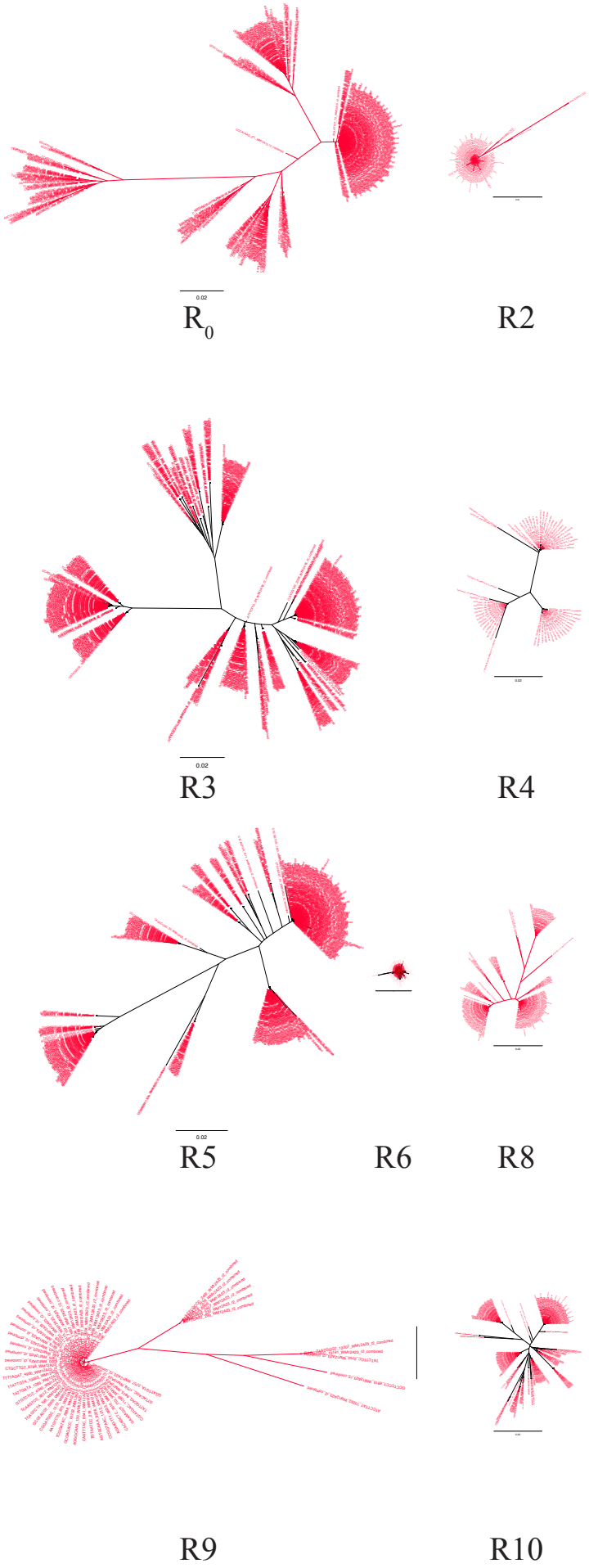


Figure 1: Sample/subject information and phylogenetic analysis of *env* sequences using a deep sequencing approach.

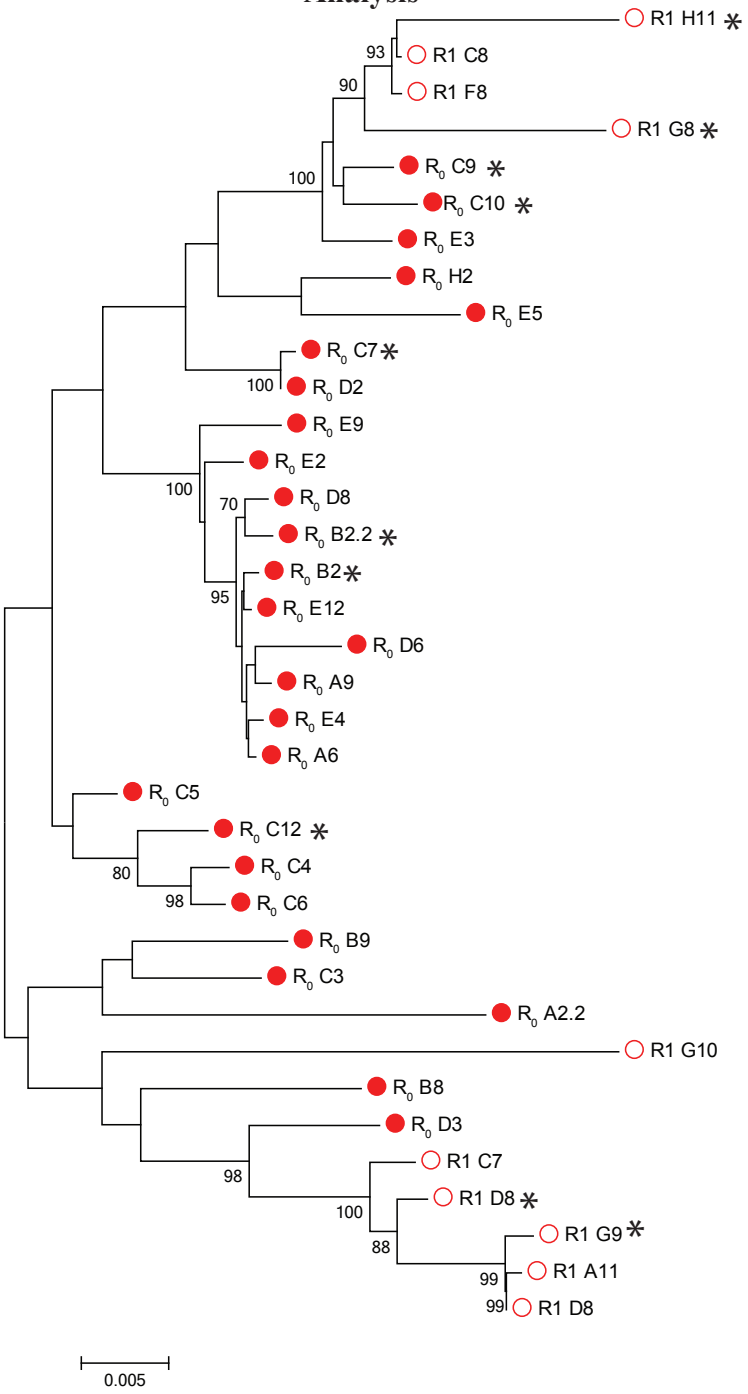
A. Sample and subject information includes a key for which samples were from which subjects, as well as nadir CD4 over the course of subjects care categories where available, CD4 levels at the time of sampling, and viral loads.

B. Comparison of the sample analyzed versus time to viral rebound was recorded. In the majority of subjects, we were able to analyze the first sample in which virus was detected or a sample very close to rebound detection.

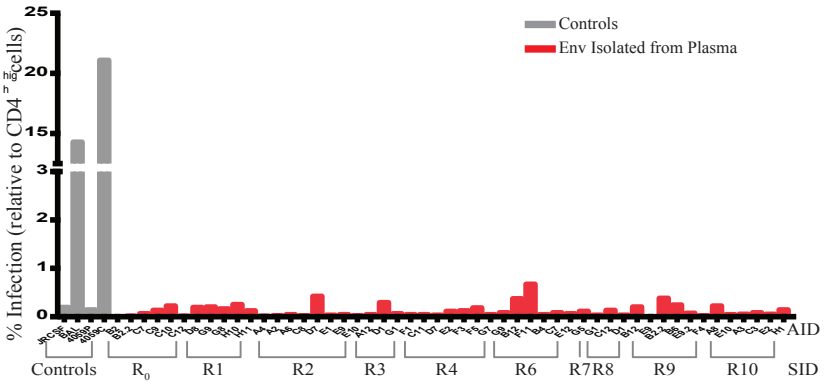
C. A comparison of the number of the minimum number of major lineages is represented in both deep sequencing phylogenetic analysis and SGA phylogenetic analysis.

D. Phylogenetic trees, created in FigTree, of the sequenced samples produced using deep sequencing with Primer ID, and presented so all samples are on the same scale.

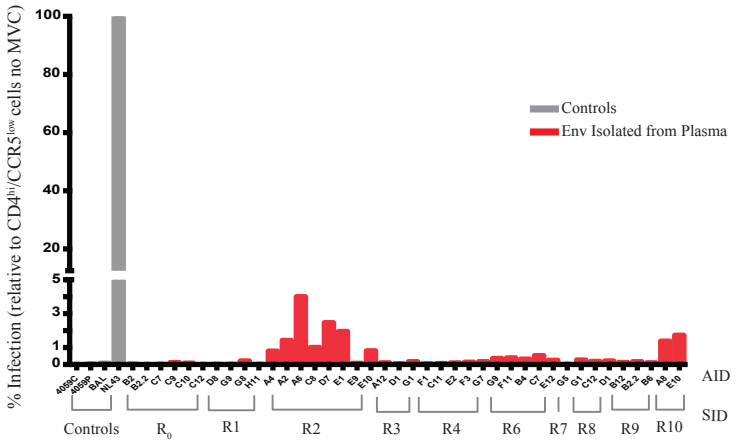
A. Single Genome Amplification Phylogenetic Analysis



B. CD4 Usage Assay Results



C. Coreceptor Usage Assay Results



D. Sensitivity to sCD4 Assay Results

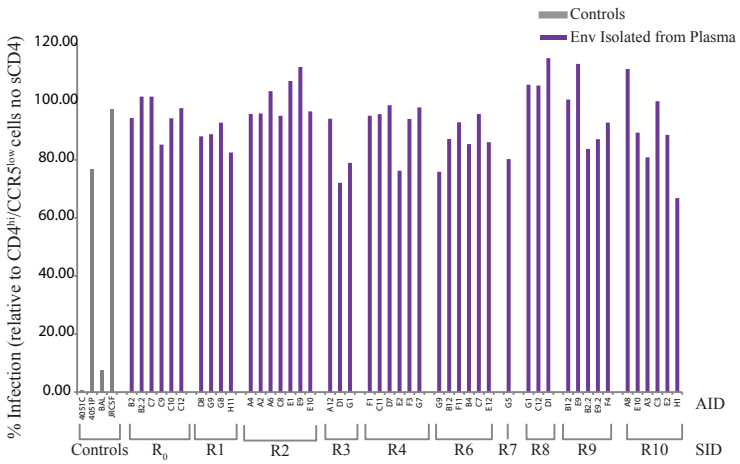


Figure 2: Phylogenetic analysis of *env* sequences and tropism determination results.

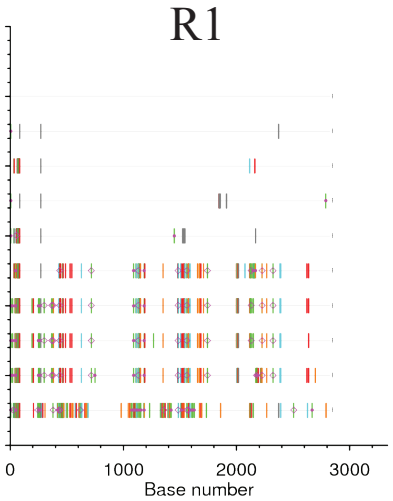
A. Phylogenetic tree, created using MEGA, of the sequences of subject 43656 pre-therapy, R_0 (●) sample and rebound sample, R1 (o) generated using SGA. Cloned and tropism-analyzed *env* sequences are marked with a * beside each label.

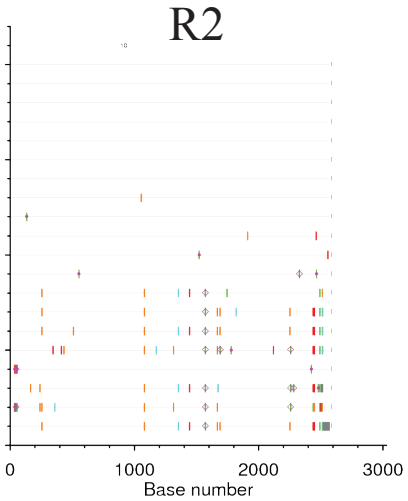
B-D. Tissue culture assay results using pseudotyped virus to define viral entry tropism phenotypes

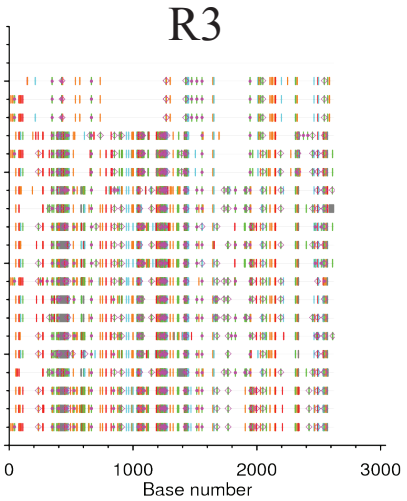
B. Relative infectivity of pseudoviruses (represented by red bars, while gray bars are controls) was determined using Affinofile cells induced to expressing low levels of CD4 compared to infection at high levels of CD4. Relative infectivity of all pseudotyped virus was below 2% and indicates T cell-tropism.

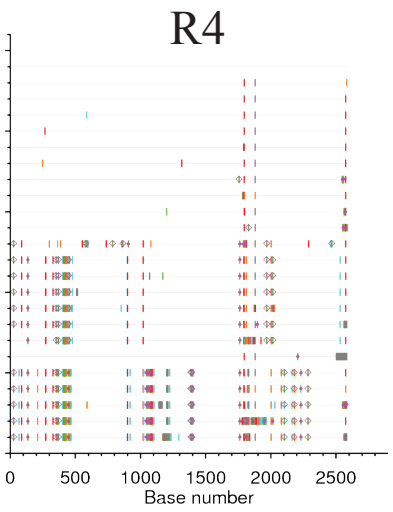
C. Relative infectivity of pseudovirus in the presence of CCR5 antagonist Maraviroc (represented by red bars while gray bars are controls) showed nearly complete inhibition of all pseudotyped viruses and indicates usage of CCR5.

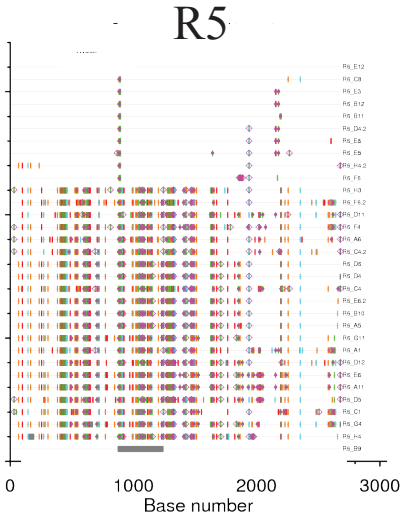
D. Relative infectivity of TZM-bl cells by pseudotyped virus in the presence of soluble CD4 (represented by purple bars while gray bars are controls) showed resistance to inhibition by all pseudotype viruses, which further confirms T cell-tropic phenotype.

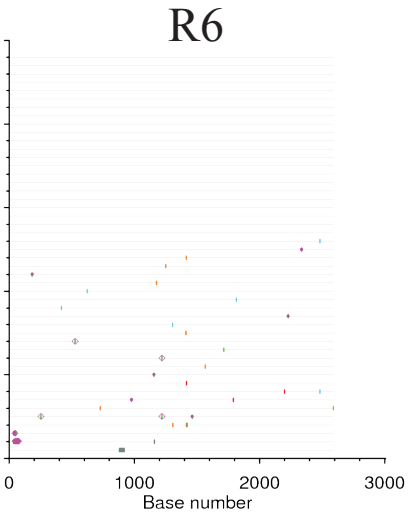

R1
0
1000
2000
3000
Base number

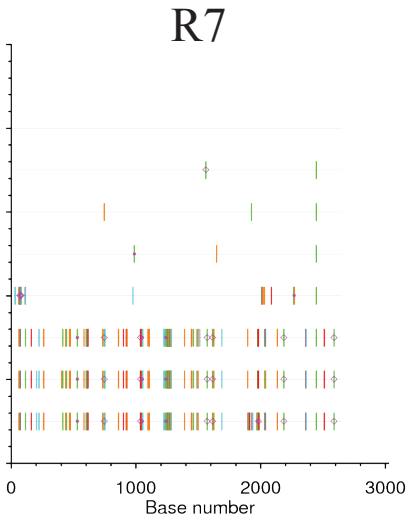

R2
0
1000
2000
3000
Base number

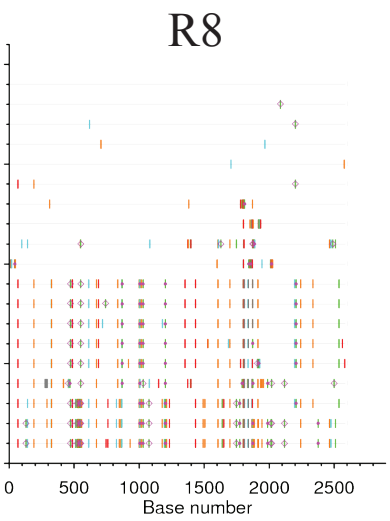

R3
0
1000
2000
3000
Base number

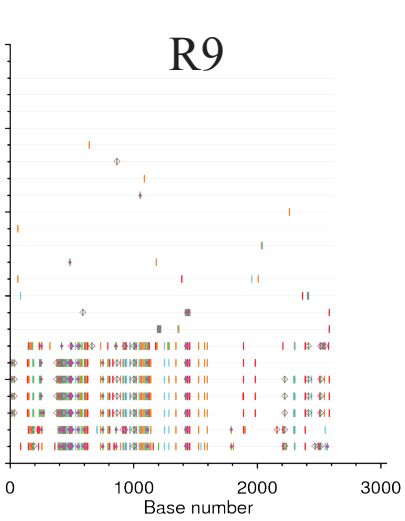

R4
0
500
1000
1500
2000
2500
Base number


R5
0
1000
2000
3000
Base number


R6
0
1000
2000
3000
Base number


R7
0
1000
2000
3000
Base number


R8
0
500
1000
1500
2000
2500
Base number


R9
0
1000
2000
3000
Base number

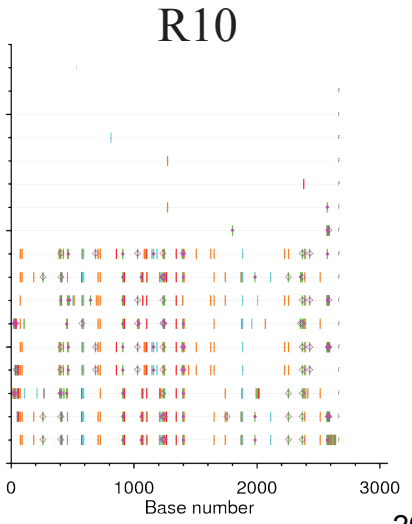
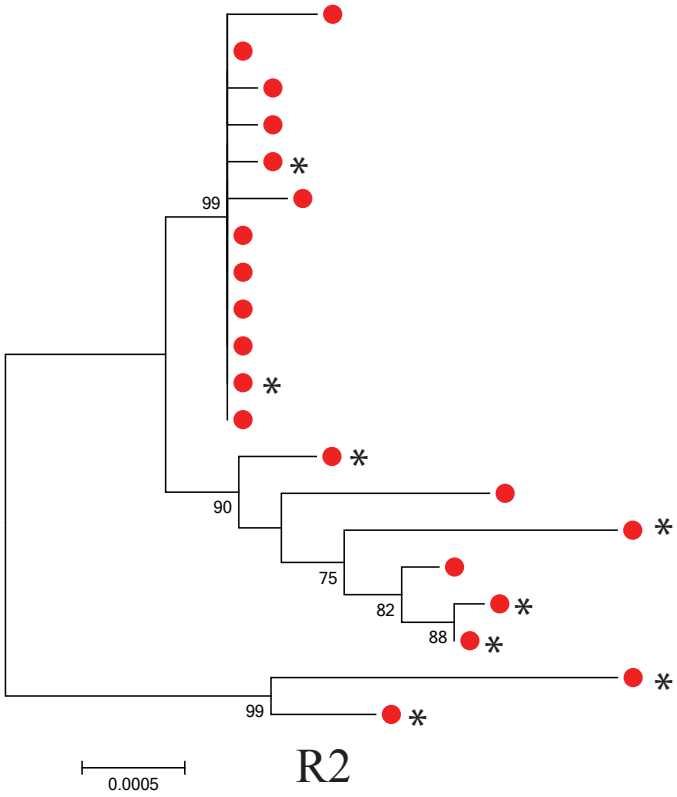

R10
0
1000
2000
3000
Base number

Figure 3: Highlighter plot analysis of *env* sequences

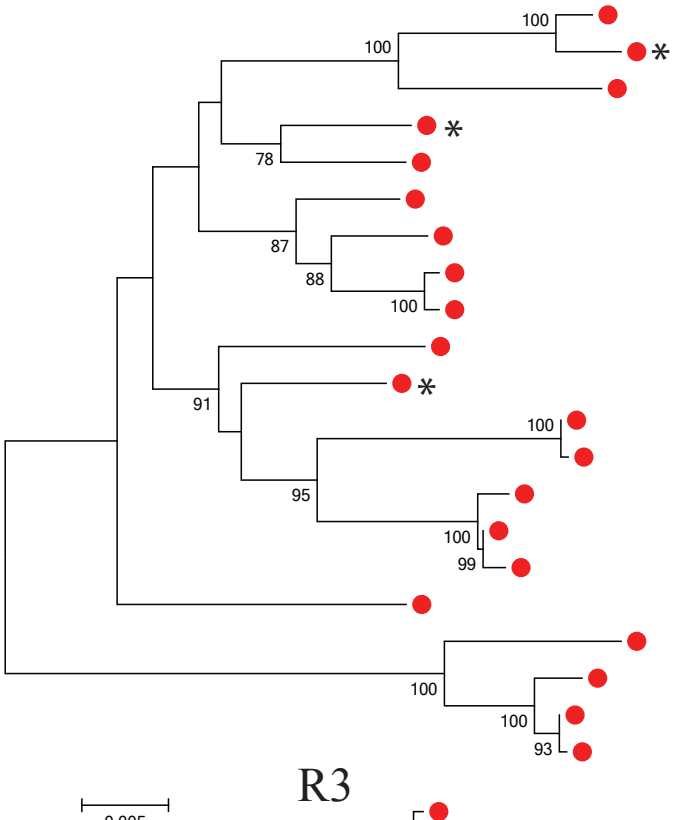
Highlighter plots were generated from SGA *env* sequences using the Los Alamos National Laboratory Highlighter generator program. Highlighter plots compare each amplicon sequence to a master sequence designated for that patient. Differences between each sequence and the master sequence are denoted by colored marks corresponding to the location in which this variation is located as well as the nature of the mutation.

Key:

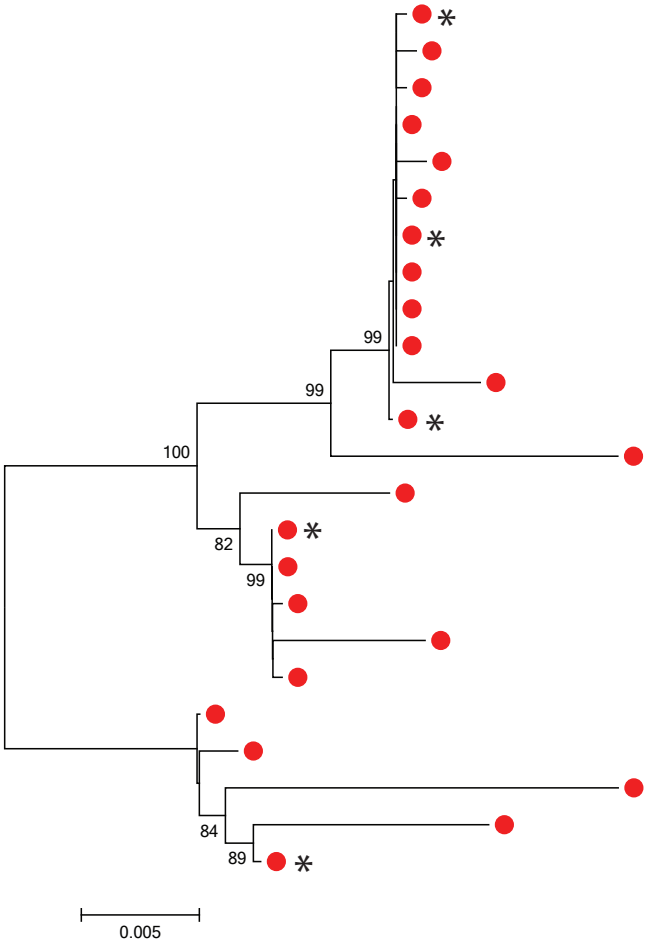
A: green; T: red; C: light blue; IUPAC: dark blue; Gaps: grey; Circle: Apobec; Diamond: G→A


99
90
75
82
88
99
0.0005

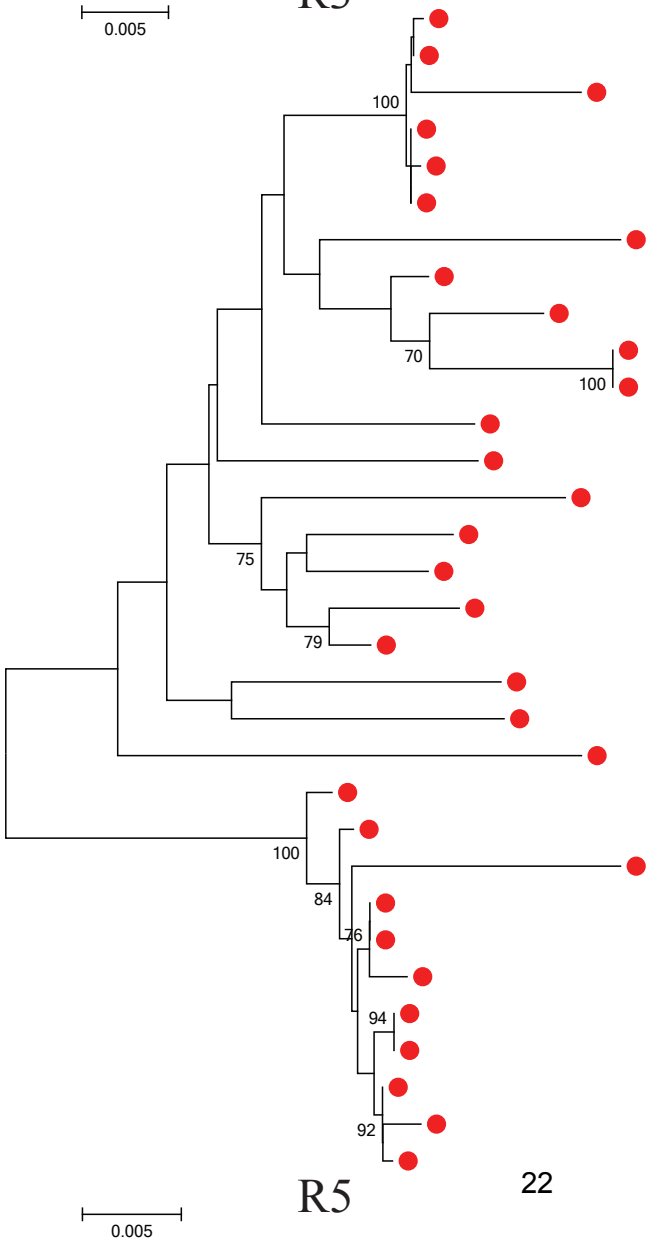
R2


100
100
78
87
88
100
91
95
100
99
100
100
100
93
0.005

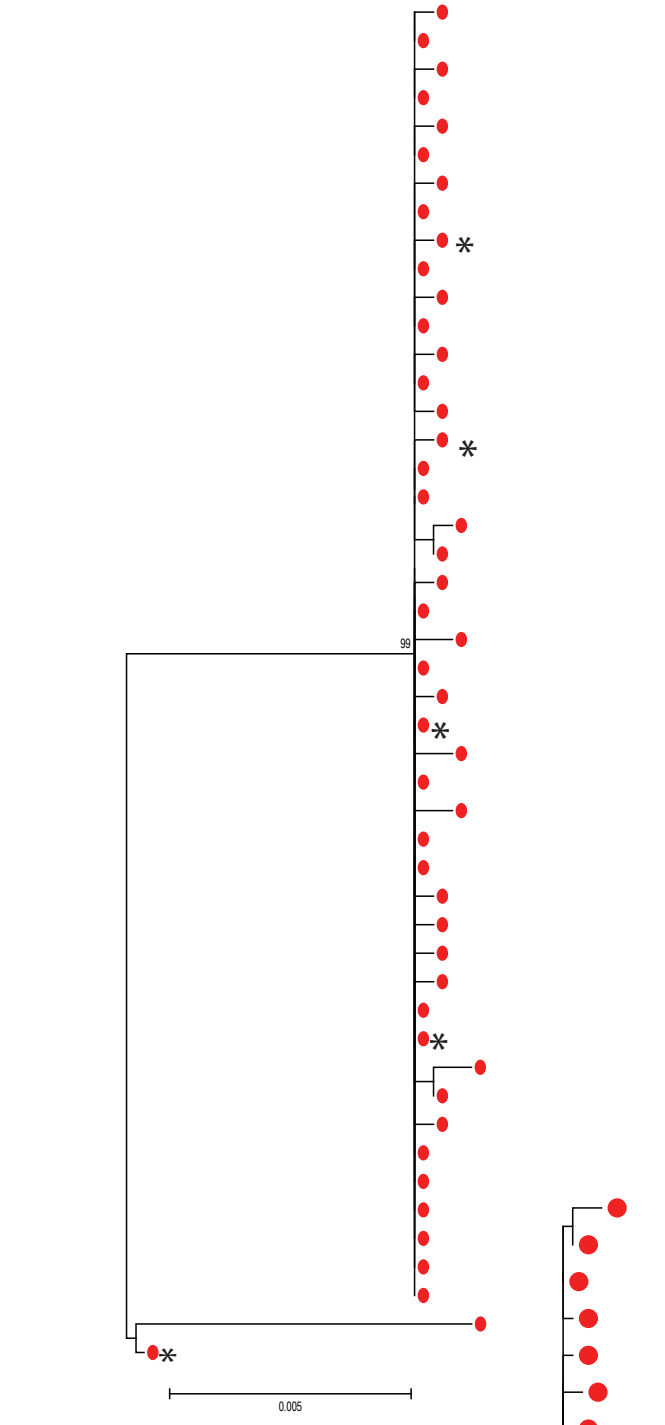
R3


99
99
100
82
99
84
89
0.005

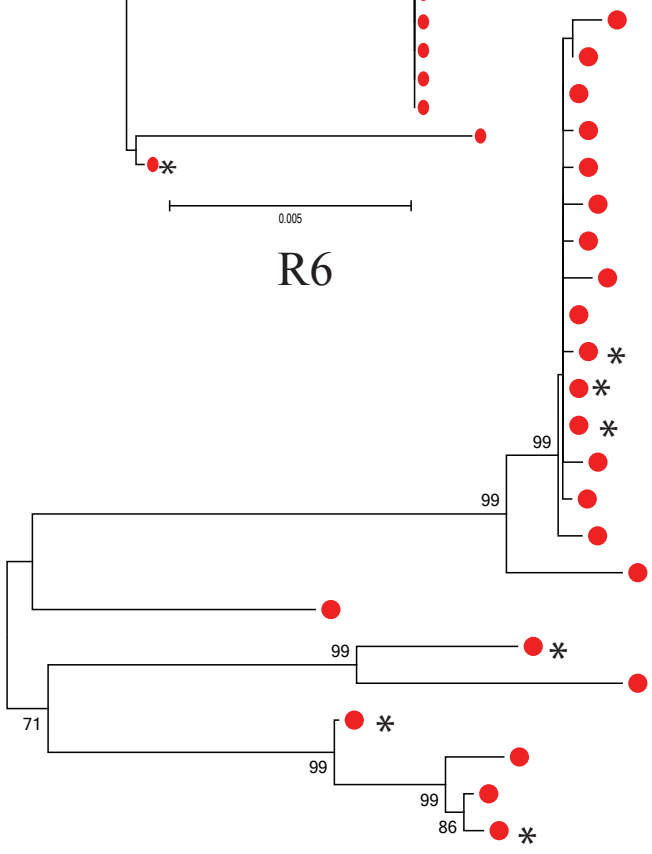
R4


100
70
100
75
79
100
84
76
94
92
0.005

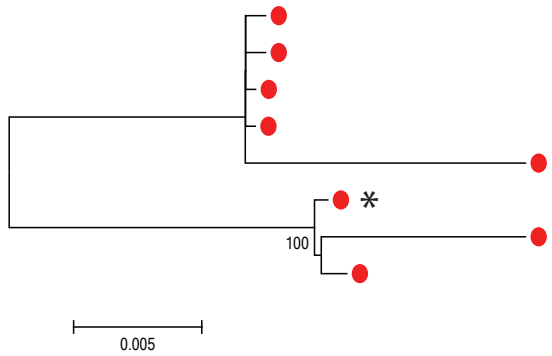
R5



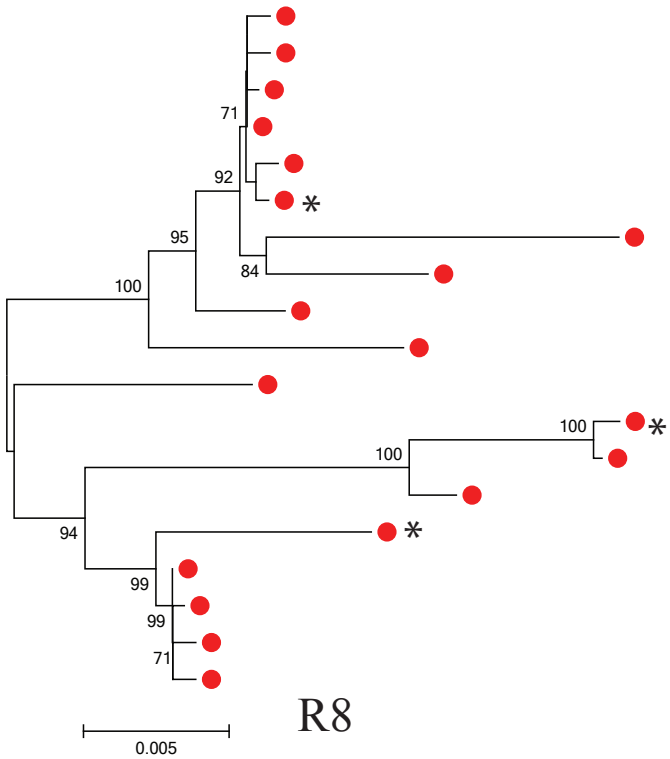
R6



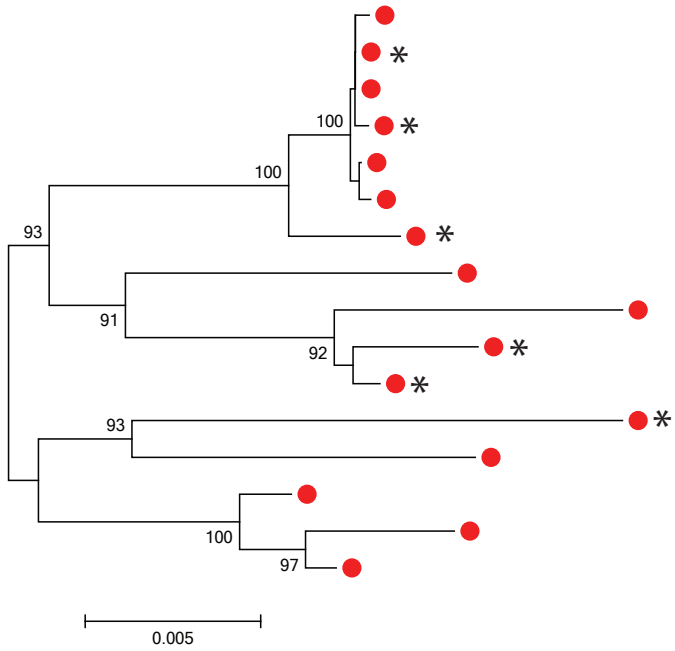
R9



R7



R8



R10

Figure 4: Phylogenetic analysis of *env* sequences. Phylogenetic trees, created using MEGA, of the sequences from all samples (●) sample generated using SGA. Cloned and tropism analyzed *env* sequences are marked with a * beside each label.

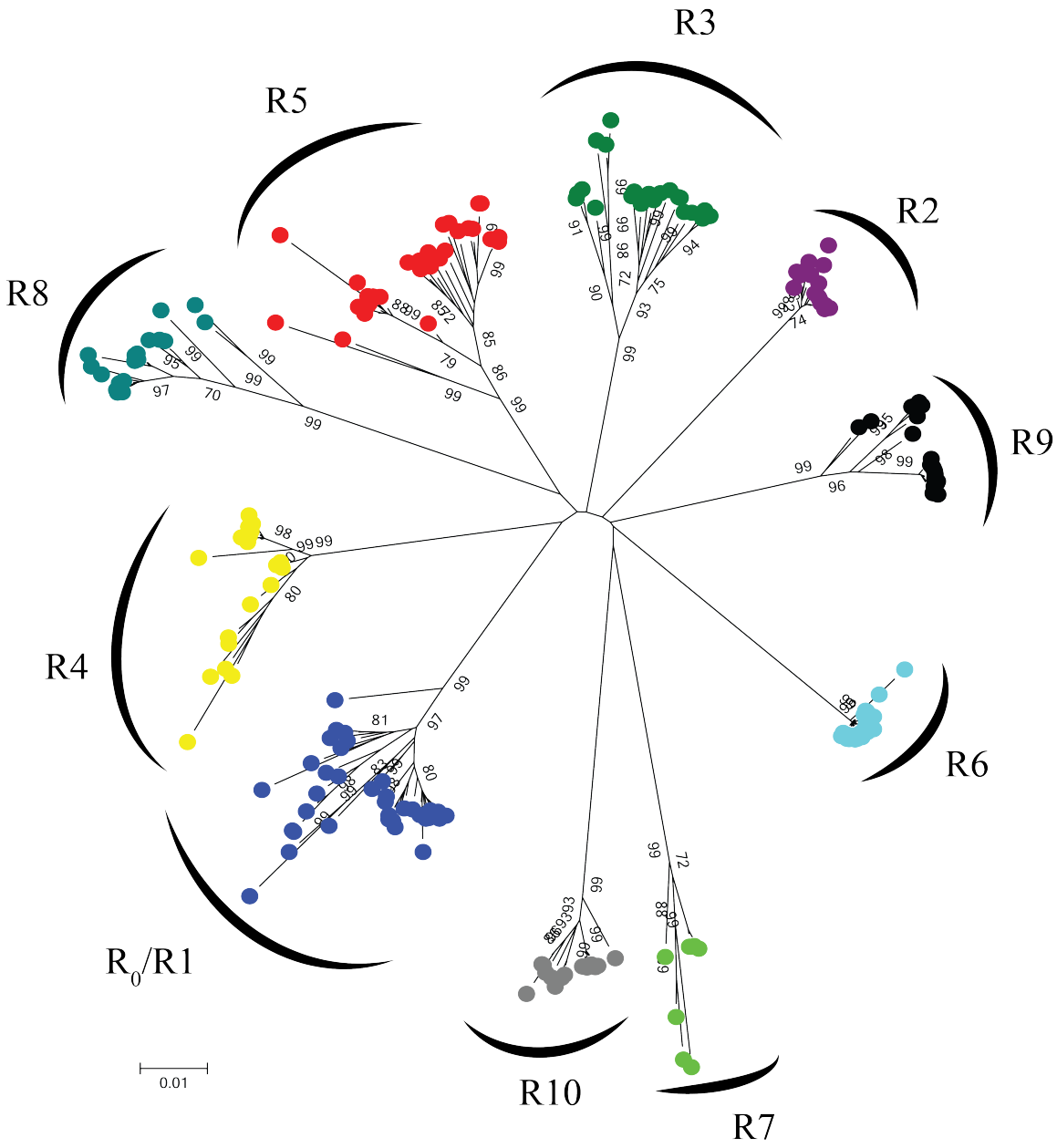

R3
R5
R2
R8
R9
R4
R6
R0/R1
R10
R7
0.01

Figure 5: Phylogenetic analysis of all *env* sequences. Phylogenetic tree, created using MEGA, of all of the sequences from all samples generated using SGA to test for any contamination. No contamination was detected. If it were present, contamination would have been revealed when sequences from two or more patients appeared to have identical or near-identical sequences as indicated by proximity on this phylogenetic tree.

Sample ID	PID	Gender	Age (at study entry)	Infection Type	Duration of time on ART (years)	Treatment prior to interruption	HIV-1 RNA level prior to interruption (copies/mL)
R1	43656	Male	37	Chronic	3.95	cmbv efv	<50
R2	291131	Male	57	Chronic	2.87	cmbv efv	<50
R3	621260	Male	50	Chronic	3.06	cmbv efv	<50
R4	141496	Male	45	Chronic	3.06	d4t 3tc kal	<50
R5	621243	Male	56	Chronic	4.71	d4t 3tc efv	<50
R6	621312	Male	52	Chronic	3.53	cmbv efv	<50
R7	291180	Male	48	Chronic	5.71	cmbv efv	<50
R8	83111	Male	49	Chronic	7.07	d4t 3tc idv rtv	<50
R9	111899	Male	48	Chronic	8.99	cmbv idv rtv	<50
R10	621388	Female	44	Chronic	4.68	3tc tdf efv	<50

Figure 6: ACTG Study A5068 Patient Data. Demographic data, treatment prior to interruption, and viral load prior to treatment interruption were collected for ACTG study A5068 patient records.

Acknowledgements

I would like to thank everyone in the Swanstrom Lab for modeling a commitment to excellence in research and for mentoring me throughout my time at UNC. In particular, I would like to thank Dr. Maria Bednar for supervising me and for allowing me to share her projects, including this one, over the past few years. Additionally, I would like to thank Dr. Shuntai Zhou for performing the deep sequencing work for this project.