

DISSERTATION

MODELING THE EVOLUTION OF SIV PROGENITOR VIRUSES TOWARDS HIV-1 AND
HIV-2 IN A HUMANIZED MOUSE SURROGATE MODEL

Submitted by

James Zachary Curlin

Graduate Degree Program in Cell and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2020

Doctoral Committee:

Advisor: Ramesh Akkina

Tawfik Aboellail
Mark Stenglein
Claudia Wiese

Copyright by James Zachary Curlin 2020

All Rights Reserved

ABSTRACT

MODELING THE EVOLUTION OF SIV PROGENITOR VIRUSES TOWARDS HIV-1 AND HIV-2 IN A HUMANIZED MOUSE SURROGATE MODEL

Human Immunodeficiency Virus Type 1 (HIV-1) and Type 2 (HIV-2), the causative agents of Acquired Immunodeficiency Syndrome (AIDS) first emerged in humans over the past century. Despite significant advances in treatment options, the pandemics continue with millions of new infections every year. Both HIV-1 and HIV-2 likely emerged through the cross-species transmission of primate lentiviruses originating from nonhuman primates (NHPs) including chimpanzees (SIVcpz), gorillas (SIVgor), and sooty mangabeys (SIVsm). SIVsm shares a remarkable degree of homology with HIV-2, while SIVcpz and SIVgor are most closely related to HIV-1. Nonhuman primates infected with these lentiviruses frequently come into contact with humans due to the prevalence of bushmeat hunting practices in various African countries. Other lentiviruses such as SIVmac239 represent independent instances of primate lentiviruses crossing into novel host species. The repeated exposure of primate lentiviruses to a human immune environment allowed the accumulation of adaptive genetic changes uniquely suited to overcoming the evolutionary pressures of a new host. Host-restriction factors such as tetherin, SAMHD1, APOBEC3G and SERINC3/5 exert species-specific antiviral activity and must be overcome for a virus to adapt to a new host cell. These evolutionary pressures could be a guiding force in the direction that these viruses adapt. In order to recapitulate these genomic cross-species adaptations, we used humanized mice engrafted with human hematopoietic stem cells (hu-HSC mice). These mice produce a full spectrum of human immune cells such as B cells, T cells, macrophages,

monocytes, and dendritic cells, and are susceptible to HIV infection. Representative progenitor viruses of both HIV-1 (SIVcpzEK505, SIVcpzMB897, and SIVcpzLB715) and HIV-2 (SIVsmE041) as well as other viruses of interest, namely, SIVmac239, SIVhu and SIV_{B670} lineages were intraperitoneally injected into hu-HSC mice. Following successful infections, the derivative viruses were subsequently passaged serially through multiple generations to simulate the repeated exposures that originally produced HIV-1 and HIV-2. Viral adaptation was assessed primarily through three different criteria. Plasma viral RNA levels were measured on a weekly basis using qRT-PCR to determine changes in viral replication kinetics over time. We found that the plasma viral loads of the viruses tested varied during serial passages, and mostly increased over time in many cases. Human CD4⁺ T cell engraftment decline as assessed by flow cytometry biweekly acts as a measure of AIDS progression in cases of human infection. CD4⁺ T cell levels declined over time with increasing rapidity upon further passaging in many cases. Additionally, viral RNA collected from the infected mice at multiple timepoints in each generation was used to generate overlapping amplicons spanning the length of the viral genome in order to be used with Illumina-based deep sequencing. Numerous nonsynonymous mutations arose in the first generation of passaging and were maintained across multiple sequential passages. While the mutations occurred throughout the viral genome, the bulk of the mutations were found in *env* and *nef*. Many of these mutations were present in known CD4⁺ binding sites, motifs involved in protein interactions, and other areas involved in host-restriction factor antagonism. While these results are revealing, further inquiry is needed to determine the true functionality of these genetic changes. These data showcase the value of using humanized mice to model lentiviral evolution and provide important insights into understanding the origin of HIVs.

ACKNOWLEDGEMENTS

The body of work presented here would not have been possible through the efforts of one person, and there are many individuals who I would like to acknowledge for enabling me to get to this point.

First of all, I would like to thank my advisor, Dr. Ramesh Akkina. Without your guidance, knowledge and resources, my development as a scientist and the work presented here would never have been possible. Your combination of optimism of the work we do and realism of what can be accomplished provides a balance as I strive to become a better scientist. Additionally, I would like to thank the members of my PhD committee, Tawfik Aboellail, Mark Stenglein, and Claudia Wiese. Each of you has been a valuable addition to my committee, and I greatly value the insight you have all provided for me during my time in graduate school. Mark in particular, I would like to thank for your assistance as a collaborator for this work and for providing excellent guidance as I began my journey into the world of sequencing and big data analysis.

I would like to thank the Cell and Molecular Biology Graduate Program here at Colorado State University for giving me the opportunity to earn my degree here. The shared experiences of the students in CMB helped remind me that I was not alone as I progressed through graduate school, and that many of the challenges of graduate school are a rite of passage that we all have to experience. Katie and Hannah, I want to thank you for being my friends and for providing an outlet when I needed to vent, or when I needed reassurance about my research.

I owe a large part of my success due to the members of the Akkina lab. Though most of you have moved on to other things, each of you helped contribute to my progress. Leila, thank you for your help with my training and guidance, particularly regarding the humanized mice. To the

various undergraduates, techs and post-docs in the lab over the years, including but not limited to: Dipu, Laurén, Lex, Molly, Callie, Betty, Joey, Jared, and Erin, thank you for creating an enjoyable atmosphere and workplace to be in. I am glad to have had you as my coworkers and friends.

Of course, I would like to thank my mom and sisters for all of their support both before and during graduate school. Knowing that I was in your thoughts during my comprehensive exam and final defense was a great comfort. I appreciate that you will be there for whatever is next on my journey to become a successful scientist.

The work reported here was supported primarily through NIH, USA grant R01 AI123234 to R. A. This project was also supported in part by the National Center for Research Resources and the Office of Research Infrastructure Programs (ORIP) of the NIH through a grant OD011104 at the Tulane National Primate Research Center and an NIH grant P51OD011106 at the Wisconsin National Primate Research Center. I would like to thank the Tulane National Primate Research Center Virus Characterization, Isolation and Production Core for providing the SIVsmE041 viral stocks. I would also like to thank Sri Boddeda for help during the initial phases of these studies and Stephanie Feely for her technical assistance. Computational resources were supported by NIH-NCATS Colorado CTSA Grant Number UL1 TR002535. The following reagents were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: SIV B670 Virus from Dr. Michael Murphey-Corb and NIH: SIVhu Virus from Dr. Thomas Folks.

Last, and certainly furthest from the least, I would like to thank Kimberly Schmitt. You have been my go-to person on a daily basis over the past 5 year any time I had a question or needed help with something. You have been the Virgil to my Dante, and without your guidance I am not sure I would have succeeded. I am genuinely grateful for everything you have done to help me.

Thank you for your years of expertise and friendship and for constantly assuring me that in fact, I can do this.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iv
CHAPTER 1: OVERVIEW OF LITERATURE.....	1
1.1 CURRENT STATE OF THE HIV PANDEMIC	1
1.2 ORIGIN OF THE HIV PANDEMIC.....	3
1.3 VIRAL STRUCTURE.....	9
1.3.1 Genomic Organization.....	11
1.3.2 LTR.....	11
1.3.3 Gag.....	12
1.3.4 Pol.....	14
1.3.5 Vif.....	16
1.3.6 Vpr.....	16
1.3.7 Vpx.....	18
1.3.8 Vpu.....	18
1.3.9 Tat.....	19
1.3.10 Rev.....	20
1.3.11 Env.....	21
1.3.12 Nef.....	22
1.3.13 ASP.....	23
1.4 REPLICATION CYCLE.....	23
1.5 RESTRICTION FACTORS	26
1.5.1 TRIM5 α /TRIMCYP	28
1.5.2 SAMHD1.....	29
1.5.3 APOBEC3	30
1.5.4 SERINC3/SERINC5.....	32
1.5.5 BST-2/Tetherin.....	33
1.6 ANIMAL MODELS FOR HIV/SIV INFECTION.....	34
CHAPTER 2: MODELING THE EVOLUTION OF SIV SOOTY MANGABEY PROGENITOR VIRUS TOWARDS HIV-2 USING HUMANIZED MICE.....	41
2.1 SUMMARY	41
2.2 MATERIALS AND METHODS.....	41
2.2.1 Generation of humanized mice.....	41
2.2.2 SIVsm primary isolate SIVsmE041	42
2.2.3 SIVsmE041 infection of humanized mice and viral determination by qRT-PCR	43
2.2.4 Determination of CD4 ⁺ T cell levels	44
2.2.5 Illumina sequencing and analysis.....	44

2.2.6 Cell culture	45
2.2.7 Viral propagation of SIV _{sm} <i>in vitro</i> by co-culturing with activated human PBMC....	45
2.2.8 Viral titration using GHOST cells	46
2.2.9 Calculation of variant frequencies in sequencing datasets	48
2.2.10 Assessment of amino acid frequencies in HIV-2 and SIV database sequences	48
2.2.11 Determining the d_N/d_S ratio of each gene from passage 5	48
2.3 RESULTS.....	49
2.3.1 SIV _{sm} E041 causes productive infection and chronic viremia in hu-mice.....	49
2.3.2 SIV _{sm} E041 infection leads to CD4 ⁺ T cell depletion in hu-HSC mice	49
2.3.3 Genetic evolution of SIV _{sm} E041 during sequential passages in hu-mice.....	52
2.4 DISCUSSION.....	55
CHAPTER 3: SIV PROGENITOR EVOLUTION TOWARD HIV: A HUMANIZED MOUSE SURROGATE MODEL FOR SIV_{SM} ADAPTATION TOWARD HIV-2	65
3.1 SUMMARY	65
3.2 MATERIALS AND METHODS.....	65
3.2.1 Generation of humanized mice.....	65
3.2.2 SIV _{sm} E041 Infection of humanized mice.....	65
3.2.3 Plasma viral load and CD4 ⁺ T cell level determination	66
3.2.4 Preparation of SIV _{sm} E041 for sequential passaging	66
3.3 RESULTS.....	66
3.4 DISCUSSION.....	67
CHAPTER 4: EVOLUTION OF SIV_{SM} IN HUMANIZED MICE TOWARDS HIV-2.....	71
4.1 SUMMARY	71
4.2 MATERIAL AND METHODS	71
4.2.1 Hu-HSC mouse preparation and Ethics.....	71
4.2.2 SIV _{sm} E041 infection of hu-HSC mice and serial passaging.....	71
4.2.3 CD4 ⁺ T-cell decline and plasma viral load detection	72
4.2.4 SNP detection and Illumina-based deep sequencing	72
4.3 RESULTS.....	73
4.4 DISCUSSION.....	73
CHAPTER 5: SIV_{CPZ} CROSS-SPECIES TRANSMISSION AND VIRAL EVOLUTION TOWARD HIV-1 IN A HUMANIZED MOUSE MODEL.....	76
5.1 SUMMARY	76
5.2 MATERIALS AND METHODS.....	76
5.2.1 Preparation of humanized mice	76
5.2.2 EK505 viral propagation and <i>in vivo</i> infection.....	76
5.2.3 Plasma viral load and CD4 ⁺ T-cell level evaluation.....	77
5.2.4 NGS and genomic analysis.....	77
5.3 RESULTS.....	78
5.4 DISCUSSION.....	78

CHAPTER 6: MIMICKING SIV CHIMPANZEE VIRAL EVOLUTION TOWARD HIV-1 DURING CROSS-SPECIES TRANSMISSION.....	82
6.1 SUMMARY	82
6.2 MATERIALS AND METHODS.....	82
6.2.1 Generation of humanized mice and ethics.....	82
6.2.2 LB715 viral propagation, <i>in vivo</i> infection and serial passage.....	82
6.2.3 Determination of PVL and CD4 ⁺ T-cell decline	83
6.2.4 Illumina-based deep sequencing and analysis	83
6.3 RESULTS.....	84
6.4 DISCUSSION.....	84
CHAPTER 7: CROSS-SPECIES TRANSMISSION AND EVOLUTION OF SIV CHIMPANZEE PROGENITOR VIRUSES TOWARDS HIV-1 IN HUMANIZED MICE.....	88
7.1 SUMMARY	88
7.2 MATERIALS AND METHODS.....	88
7.2.1 Preparation of the SIVcpz viral stocks	88
7.2.2 Generation of humanized mice.....	89
7.2.3 SIVcpz infection of humanized mice and viral load determination by qRT-PCR.....	90
7.2.4 Determination of CD4 ⁺ T cell levels	91
7.2.5 Illumina-based deep sequencing and analysis	92
7.2.6 Cell Culture.....	92
7.2.7 Viral propagation, titration and subsequent viral passaging of SIVcpz	93
7.2.8 Quantification of variant frequencies	93
7.2.9 Assessment of amino acid frequencies in HIV-1 and SIV database sequences	94
7.3 RESULTS.....	94
7.3.1 Chimpanzee native viral strains SIVcpzEK505, LB715, and MB897 can establish productive infection and chronic viremia in humanized mice	94
7.3 RESULTS.....	95
7.3.1 Chimpanzee native viral strains SIVcpzEK505, LB715, and MB897 can establish productive infection and chronic viremia in humanized mice	95
7.3.2 Serial passage of SIVcpz viruses in hu-mice leads to more rapid CD4 ⁺ T cell decline.....	97
7.3.3 SIVcpz viruses accumulate adaptive viral genetic changes during long-term serial passages in hu-mice	99
7.4 DISCUSSION.....	102
CHAPTER 8: ADAPTATION OF SIVMAC239 <i>IN VIVO</i> USING HUMANIZED MICE AS A PARALOG FOR SIVSM ADAPTATION INTO HIV-2.....	115
8.1 SUMMARY	115
8.2 MATERIALS AND METHODS.....	116
8.2.1 Creation of Hu-HSC and BLT mice.....	116
8.2.2 SIVmac239, SIVhu, and SIV _{B670} viral stock preparation and <i>in vivo</i> infection.....	117
8.2.3 TZM-bl based titrating of viral stocks	118
8.2.4 Serial passaging of SIVmac239, SIVhu, and SIV _{B670}	118

8.2.5 Plasma viral load and CD4 ⁺ T cell decline detection	119
8.2.6 Amplicon and Illumina-based deep sequencing preparation.....	120
8.2.7 Calculation of nonsynonymous single nucleotide polymorphism (SNP) frequencies	121
8.3 RESULTS	122
8.3.1 SIVmac239 is capable of productive and chronic infection in humanized mice and produces HIV-like CD4 ⁺ T cell depletion	122
8.3.2 SIV _{B670} is capable of productive and chronic infection in humanized mice and produces HIV-like CD4 ⁺ T cell depletion	123
8.3.3 SIVhu is capable of productive and chronic infection in humanized mice but does not produce HIV-like CD4 ⁺ T cell depletion	125
8.3.4 Evolutionary genetic adaptations arise during sequential passaging of SIVmac239, SIV _{B670} and SIVhu.....	126
8.4 DISCUSSION	127
CHAPTER 9: CONCLUSIONS AND FUTURE DIRECTIONS	141
9.1 CONCLUSIONS AND SUMMARY	141
9.2 FUTURE DIRECTIONS	148
REFERENCES	152
APPENDIX: SUPPLEMENTAL MATERIAL FOR CHAPTERS 2, 7, AND 8.....	178

1.1 Current state of the HIV Pandemic

In the early 1980s, clusters of homosexual men began presenting a number of opportunistic infections and conditions usually found in immunocompromised individuals, including pneumocystis pneumonia and Kaposi's sarcoma (Gottlieb et al., 1981). These conditions, which are typically not fatal in healthy individuals began resulting in numerous fatalities. By 1981, these symptoms in combination with a dramatically reduced level of circulating CD4⁺ T cells became classified as "Acquired Immunodeficiency Syndrome" (AIDS), and a retrovirus was soon determined to be the causative agent (Barre-Sinoussi et al., 1983; Popovic et al., 1984). AIDS was then further determined to be sexually transmitted, further cementing its origin as an infectious agent (Auerbach et al., 1984).

Initially restricted to high-risk groups, infections of the newly termed "Human Immunodeficiency Virus" (HIV-1) ballooned out of control within a few short years, and was particularly prevalent in sub-Saharan Africa by 1986 (Kreiss et al., 1986). A similar virus with roughly 50% homology to HIV-1, termed HIV-2, was identified in West Africa the same year (Clavel et al., 1986a; Clavel et al., 1986b). By 1990, millions of new infections of HIV were being reported each year ((UNAIDS), 2019). Since then, the resulting pandemic-causing pathogen has been the subject of extensive research. Despite over 40 years of research, HIV/AIDS remains a global health problem, though medical advances, like antiretroviral therapy (ART) have helped significantly. New infections with HIV peaked in 1997 at around 2.9 million with a subsequent decline to 1.7 million new infections in 2018, while AIDS-related deaths for individuals infected

with HIV peaked globally in 2004 with approximately 1.7 million deaths, and has declined to approximately 770,000 deaths in 2018 (UNAIDS 2019). Since 2010 there has been a 33% decline in mortality. The majority of this decline is due to progress in southern and eastern regions of Africa, where over half of the world's HIV-infected individuals reside. However, as of 2018 there are still approximately 37.9 million individuals living with HIV, thus highlighting the importance of additional basic science research, developing better therapeutics, and ultimately, creating a vaccine ((UNAIDS), 2019).

Currently, HIV-1 is organized into four major groups, M (major), N (non-M/non-O), O (outlier), and P (prototype) (Figure 1.1). Of these, group M is responsible for the vast majority of the global epidemic cases, followed by group O which is largely restricted to Cameroon and Gabon and makes up fewer than 1% of global HIV-1 cases (De Leys et al., 1990; Sharp and Hahn, 2011). Groups N and P make up only a handful of cases, and were all isolated in individuals from Cameroon (Plantier et al., 2009; Vallari et al., 2010). Group M, which constitutes the majority of HIV-1 cases, is further divided into the following subtypes: A, B, C, D, F, G, H, J, and K, as well as a special category known as CRFs, or circulating recombinant forms (Bell and Bedford, 2017). In contrast to the spread and divergence of HIV-1, HIV-2 has been relatively limited to West Africa with some cases in other countries with ties to West Africa, and is spread across nine groups representing independent transmissions: A, B, C, D, E, F, G, H, and I (Sharp and Hahn, 2011; Visseaux et al., 2016; Bell and Bedford, 2017). Approximately 90% of HIV-2 cases are from group A viruses across various West African countries, with the remaining cases occurring predominantly due to group B viruses in places such as Ghana and Burkina Faso (Sharp and Hahn, 2011; Visseaux et al., 2016). Due to the rapid evolution of these viruses, their phylogeny is being constantly updated, with certain groups such as HIV-2 group I being confirmed as recently as

2013, and HIV-1 group P as recently as 2009 (Smith et al., 2008; Plantier et al., 2009; Ayouba et al., 2013).

1.2 Origin of the HIV Pandemic

While the epidemic in the 1980s and the subsequent global pandemic catapulted HIV into infamy, there is some evidence that the origin of HIV began much earlier. Some reports have suggested that an individual was infected with HIV-1 as early as 1959 in the Democratic Republic of Congo, while others suggest that the major group of HIV-1 (Group M) has been diversifying since approximately 1930 (Zhu et al., 1998; Korber et al., 2000; Korber et al., 2001; Rowland-Jones, 2003). However, a study published in 2008 that analyzed viral sequences recovered from a biopsy collected in Kinshasa in 1960, along with previously known sequences, suggested a common ancestor may have arisen sometime between 1873 and 1933, with central estimates varying between 1902 and 1921 (Worobey et al., 2008). Due to the error-prone nature of the reverse transcriptase in HIV, it is susceptible to very rapid mutation (Bebenek et al., 1993). The presence of these earliest cases appearing predominantly in Africa where many NHP species are endemically infected with SIVs, rather than North America supports the theory that HIV may have originated through a cross-species transmission event.

HIV belongs to the *Retroviridae* family, in the genus *Lentivirus*, which contains RNA viruses known primarily for their use of reverse transcriptase to generate viral DNA from RNA. Lentiviruses have a wide variety of mammalian hosts across the globe, including cats, nonhuman primates, goats, horses, sheep, and humans (Dow et al., 1990; Gifford, 2012). Both HIV-1 and HIV-2 are similar lentiviruses that are responsible for epidemics in humans, but their theorized

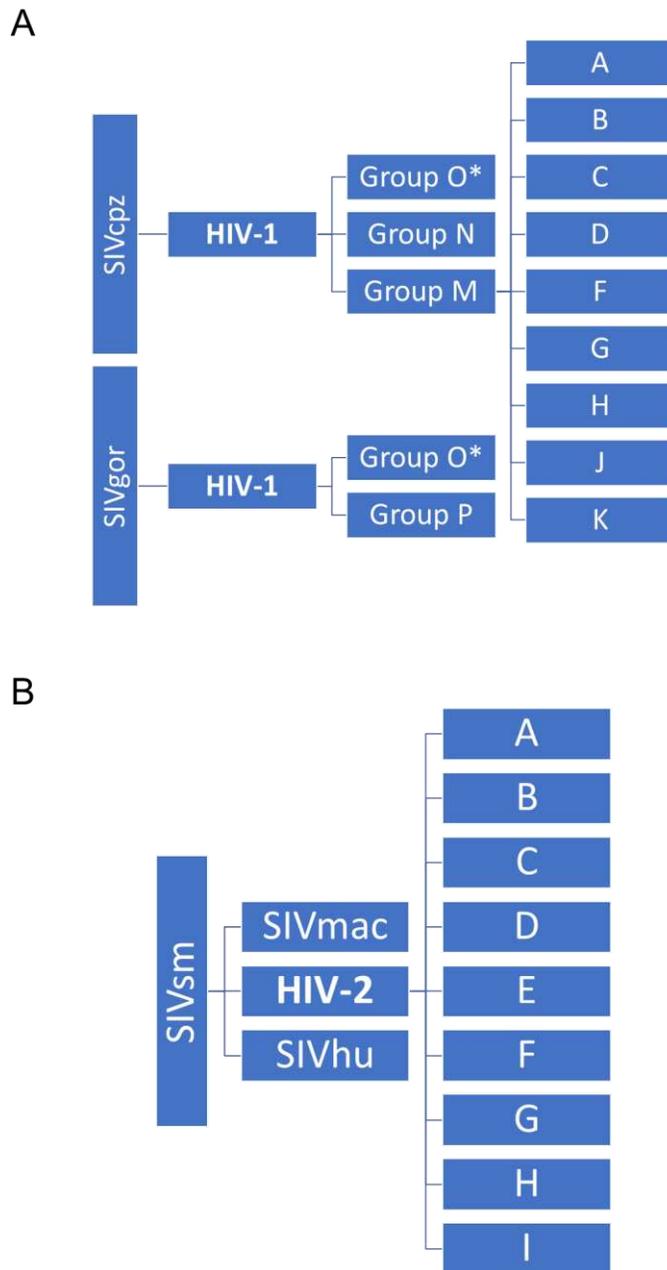


Figure 1.1: Major HIV families and their SIV ancestors. (A) HIV-1 is composed of Groups M, N, O, and P, which originated from SIVcpz and SIVgor. HIV-1 Group M is further divided into the subtypes shown. (B) HIV-2, which is divided into Groups A through I, originated from SIVsm, which is a shared ancestor for both SIVmac and SIVhu as well. *The origin of HIV-1 Group O is unclear and could be attributed to either SIVgor, or SIVcpz.

origins are unique. The lentivirus that is most genetically similar to HIV-1 groups M and N is a Simian Immunodeficiency Virus found in chimpanzees (SIVcpz). Additionally, HIV-1 groups O and P may have originated from gorillas (SIVgor) (Van Heuverswyn et al., 2006; Plantier et al., 2009; Sharp and Hahn, 2011). In comparison, HIV-2 was found to be even more genetically similar to an SIV originally found in sooty mangabeys (SIVsm) (Hirsch et al., 1989; Gao et al., 1992). Prior to this, a similar virus found to cause symptoms similar to AIDS was identified in captive rhesus macaques, and is believed to be a separate example of a cross-species transmission of a primate lentivirus resulting from the accidental transmission cohabitated infected sooty mangabeys (Daniel et al., 1985; Apetrei et al., 2005). It was initially believed that HIV-1 was drastically more diverged from SIVcpz than HIV-2 was from SIVsm, but further sampling discovered that the geographic origin of the respective SIV/HIVs and their genetic similarity to each other was highly correlated, suggesting that local strains of HIV greatly resembled the SIV strains of the surrounding NHP populations (Gao et al., 1999). This was further reinforced by evidence that showed certain strains of HIV-2 were more closely related to specific SIVsm strains than other HIV-2 strains, a finding that was also linked to the geographic distribution of the samples (Sharp and Hahn, 2011; Ayouba et al., 2013). It is currently believed that the four major HIV-1 groups (M, N, O, and P) and eight major HIV-2 groups arose as a result of independent cross-species transmission events (Gao et al., 1992; Gao et al., 1999; Bell and Bedford, 2017).

The exact mechanism that facilitated the repeated cross-species transmissions that eventually gave rise to the HIV epidemic is not well understood, but several predominant theories exist. As seen with the 2019 global pandemic of SARS-CoV-2, the intersection of human markets with wild animals provides a unique environment that lends itself to emerging zoonotic pathogens (Zhang and Holmes, 2020). Bushmeat consumption is still widespread across Africa, and many

instances of new HIV strains emerging have occurred in individuals confirmed to consume bushmeat (Ayouba et al., 2013). Contact with infected blood from nonhuman primates during the butchering process, combined with open wounds from bites or other injuries provide a plausible route of transmission (Hahn et al., 2000). Alternatively, it has also been proposed that HIV emerged as a consequence of unsterile vaccination campaigns, with SIV gradually accumulating adaptive mutations over time as infected needles were reused between patients (Marx et al., 2001)

There are over 40 currently identified species-specific SIVs found among native populations of old world monkeys and primates including, but not limited to: Chimpanzees, Sooty Mangabeys, African Green Monkeys, Red-capped Mangabeys, Gorillas, Mandrills, and even lemurs (Gilbert et al., 2009; Sharp and Hahn, 2011; Bell and Bedford, 2017). Interestingly, the prevalence of the SIVs vary vastly between different species, with less than 1% of the *Cercopithecus* populations infected, and greater than 50% of wild sooty mangabeys infected (Santiago et al., 2005; Aghokeng et al., 2006; Aghokeng et al., 2010; Sauter and Kirchhoff, 2019). Genetic evidence suggests that many of these nonhuman primates have co-existed with their respective SIVs for thousands of years, and they likely emerged following the divergence of Asian and African old world monkeys around 35 MYA, as the SIVs are absent in Asian Old World and New World monkeys (Schrage and Russo, 2003; Worobey et al., 2010; Sauter and Kirchhoff, 2019). This long period of co-evolution may partially explain why the majority of these species do not display AIDS-like symptoms, even when infected with the virus for years (Chahroudi et al., 2012; Sauter and Kirchhoff, 2019). In comparison, Chimpanzees have been shown to develop clinical AIDS over time, with devastating effects on chimp colonies that had higher rates of infection of the virus, suggesting that SIVcpz has not optimally adapted to its host species yet (Keele et al., 2006; Keele et al., 2009). Interestingly, SIV chimpanzee (SIVcpz) was determined

to be a mosaic virus, with components of its genome originating from different other SIVs. SIVrcm, which originated in red-capped mangabeys, constitutes the *nef* gene, 3'LTR and 5' half of the SIVcpz genome, and SIVs from *Cercopithecus* monkeys (SIVgsn/SIVmus/SIVmon) constitute the *vpu*, *tat*, *rev* and *env* regions of the SIVcpz genome (Bailes et al., 2003; Bell and Bedford, 2017). Both of these groups of primates are naturally preyed upon by chimpanzees and it has been previously shown that recombination between poorly adapted mutant strains from independent infections, in this instance SIV isolated from rhesus macaques (SIVmac), can produce a competent, pathogenic, combined virus (Wooley et al., 1997; Mitani and Watts, 1999). This hints at a potential mechanism through which chimpanzees likely acquired the constitutive viruses and lends credence to the theory of HIV's emergence in humans through bushmeat hunting. However, instances of cross-species transmissions of these viruses are still rare (Sauter and Kirchhoff, 2019).

HIV-1 and HIV-2 are not the only example of cross-species transmissions of lentiviruses. Captive sooty mangabeys that were infected with their endogenous virus (SIVsm) were found to have accidentally transmitted the virus to Asian-origin rhesus macaques that were co-housed in the same facility. It was subsequently determined that the virus (SIVmac) was not natively found in wild rhesus macaques and was unique to the animals in that particular facility, further confirming that this was an instance of cross-species transmission. This new virus, termed SIVmac, was found to be highly pathogenic in rhesus macaques unlike the majority of naturally occurring SIVs (Daniel et al., 1985; Apetrei et al., 2005). The ability of SIVmac to cause the rapid onset of AIDS-like symptoms in rhesus macaques has resulted in SIVmac becoming one of the earliest and most commonly used surrogates for understanding HIV infections (Naidu et al., 1988; Kestler et al., 1990).

SIV_{mac} is not the only recorded incidence of an SIV_{sm}-derived virus crossing the species barrier. In 1990, a laboratory worker was exposed to an SIV_{B670} via accidental needlestick. Although no positive viral loads were detected through PCR, and the blood from the worker did not transmit the virus into macaques, this individual underwent seroconversion and produced antibodies to *gag* and *env*. This was the first evidence of the potential for SIV to infect humans (Khabbaz et al., 1992). A second researcher was later identified who also produced antibodies to SIV and was likely infected in 1989 while working on SIV_{B670} with severe dermatitis and without gloves. While the patient never developed AIDS, nor passed the virus on to their partner, the virus was nevertheless detected via PCR and persisted for at least 3 years. This virus, termed SIV_{hu} was isolated and confirmed as the first documented case of a human infected with an SIV (Khabbaz et al., 1994).

Prior to the emergence of HIV-1, numerous genomic rearrangements needed to occur. The recombination of SIV_{rcm} and SIV_{gsn}/SIV_{mus}/SIV_{mon} into an adapted SIV_{cpz} resulted in the loss of *vpx*, as well as the elongation of *vif* and a spatial separation of the *env* and *nef* genes. Additionally, *vpu*, which was not found in SIV_{rcm} was preserved in SIV_{cpz}. As a consequence of these rearrangements, several functional changes also occurred, primarily interactions with host-restriction factors. Host-restriction factors are antiviral defense mechanisms found in many cells that serve to block viral replication at various points in the replication cycle. For example, the new virus lost the ability to counteract SAMHD1 (SAM domain and HD domain-containing protein 1) and tetherin/BST-2 with Vpr/Vpx and Vpu, respectively. To compensate, Nef took over as the primary means to counteract tetherin, while APOBEC3 (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3) activity was increasingly counteracted by Vif (Sauter and Kirchhoff, 2019). Upon transmission of SIV_{cpz} into humans, significantly fewer genomic rearrangements

were needed, likely due to the similarity between humans and chimpanzees. However, this cross-species transmission was accompanied by several changes in host-restriction factors interactions such as switching from Nef back to Vpu as the primary means of counteracting tetherin. Interestingly, SIVsm has seemingly had to undergo fewer broad genomic rearrangements when crossing into humans. In fact, SIVsm has been shown to replicate in human PBMC without adaptation *in vitro*, and several of its accessory proteins are already capable of counteracting human host restriction factors, such as SIVsm Env-mediated tetherin antagonism (Gautam et al., 2007; Le Tortorec and Neil, 2009; Heusinger et al., 2018).

1.3 Viral Structure

As a mature virion, HIV/SIV consists primarily of an enveloped capsid roughly 145 nm in size, that contains two copies of the positive sense single-stranded viral RNA (Briggs et al., 2003; Ganser-Pornillos et al., 2012). The phospholipid bilayer envelope originates from host cell's membrane that buds off during HIV/SIV replication. Embedded in the membrane are the envelope glycoprotein, consisting of the gp120 cap and gp41 stem that form trimeric spikes in order to bind to the host cell receptors (Figure 1.2). Within the envelope is a matrix composed of the p17 protein, which acts to preserve the integrity of the virion. Additionally, several protease molecules are present, which are necessary to facilitate the cleavage and maturation of the gag polyprotein. The viral capsid itself is composed of numerous copies of the p24 protein. Within the capsid, the RNA copies are attached to the p7 nucleocapsid proteins, and other necessary enzymes and proteins needed for viral replication such as the integrase, reverse transcriptase, Vpr, Vif and Nef are also present (Frankel and Young, 1998).

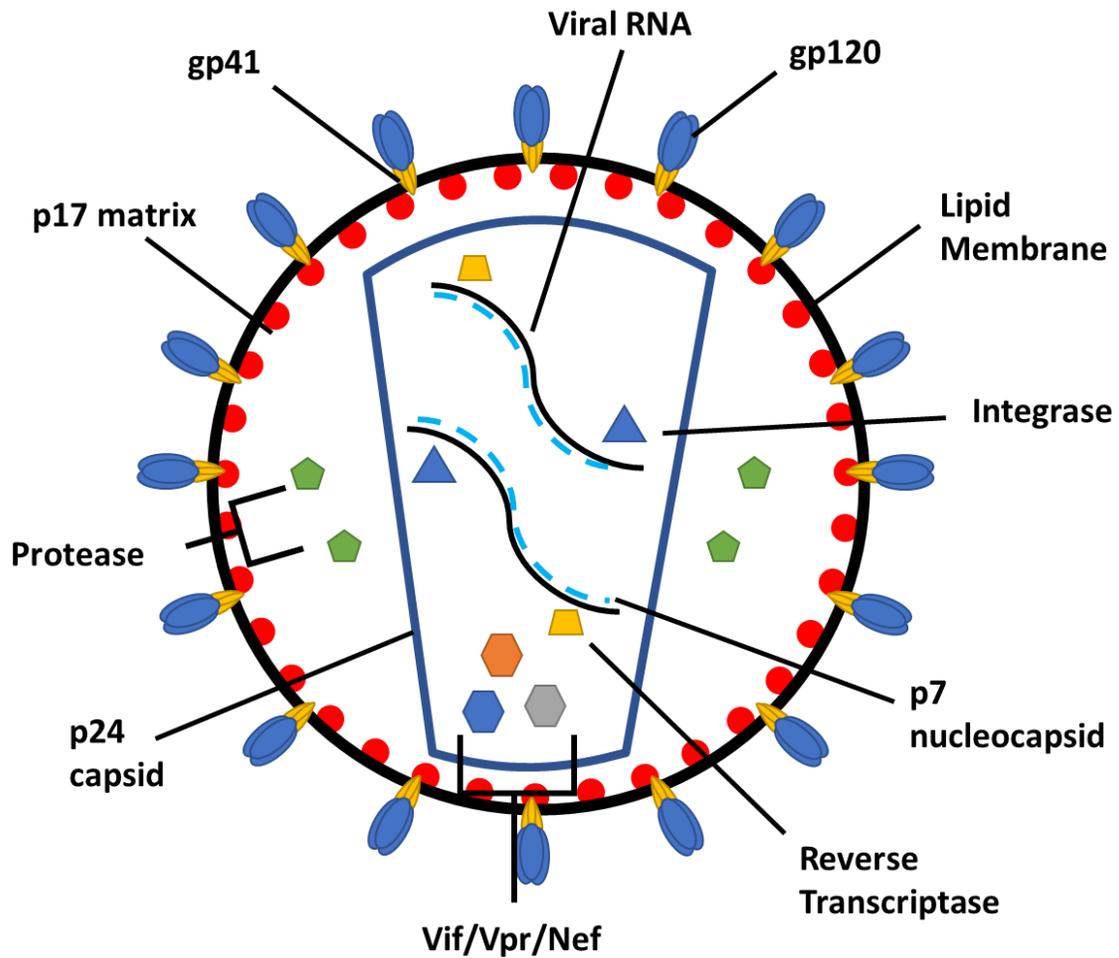


Figure 1.2: Schematic of an HIV-1 virion after maturation. An HIV capsid contains two single-stranded copies of the viral RNA genome and multiple viral proteins necessary for replication. The capsid is surrounded by a lipid envelope originating from the host cell that the virus budded from. This envelope is studded with viral proteins necessary for host-cell binding and entry.

1.3.1 Genomic Organization

Both HIV-1/SIVcpz and HIV-2/SIVsm have nine canonical genes, which make up the bulk of their roughly ~10 kb genomes (Figure 1.3). Three of these genes code for the structural proteins responsible for forming the mature virion and have a common lineage amongst retroviruses. These are proteins known as Gag (group-specific antigen), Pol (polymerase) and Env (envelope). The remaining six genes, which encode two regulatory proteins (Tat, Rev) and four accessory proteins (Vif, Vpr, Vpu/Vpx, Nef) are less conserved, but still maintain an important role in the fitness of the viruses (Strebel, 2013; Sauter and Kirchhoff, 2019). While most of these genes are the same between HIV-1 and HIV-2, *vpu* is unique to HIV-1, and *vpx* is unique to HIV-2. This difference can likely be attributed to the difference in origins between the two types of virus. Additionally, there is evidence to suggest the presence of a tenth “antisense protein” (ASP) based on an open reading frame that has been preserved amongst the majority of HIV-1 group M sequences, though it is absent from group O and SIVcpz sequences (Miller, 1988; Torresilla et al., 2015).

1.3.2 LTR

Long terminal repeats (LTR) are regions of the viral genome located both on the 5' and 3' ends of the coding region, with the 5' LTR flanking *gag*, and the 3' LTR overlapping with the *nef* gene. The LTR is typically divided into the U3, R (repeat domain), and U5 regions and is heavily involved in viral transcription. Several different transcription factors from the host cell including, NF- κ B, Elf-1 and Sp1, as well as the HIV protein Tat are all responsible for upregulating transcriptional activity mediated by the LTR (Markovitz et al., 1990; Markovitz et al., 1992; Hannibal et al., 1994; Rhim and Rice, 1994). These cellular transcription factors typically all bind to a small regulatory subregion of the LTR. HIV-1 has two NF- κ B binding sites, while HIV-2 has only a single site, necessitating the addition of several synergistic elements including, two purine

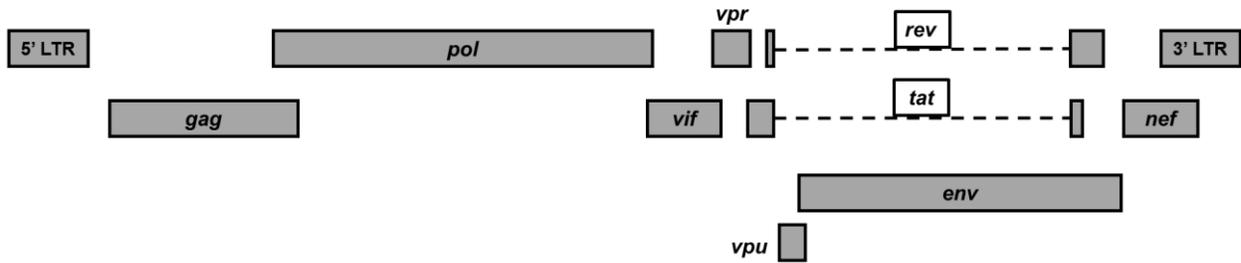
rich boxes (PUB1/PUB2) and a peri-ETS (pETS) site that can all bind Elf-1, and a peri- κ B site that is monocyte-specific (Markovitz et al., 1990; Tong-Starksen et al., 1990; Markovitz et al., 1992; Arya and Mohr, 1994; Hannibal et al., 1994). Additionally, it has been shown that genetic variability of this region affects the transcription factor binding sites and can result in drastic changes in viral transcription levels and the size of the viral reservoir (Qu et al., 2016; Le Hingrat et al., 2020). In both HIV-1 and HIV-2, there is significantly greater genetic diversity in the U3 region relative to the R and U5 regions (Le Hingrat et al., 2020). Within the R region of the LTR is the transactivation response (TAR) element, which forms a hairpin stem-loop structure and is the target used in viral transactivation and is bound by Tat along with various cellular proteins (Zheng et al., 2005; Foley, 2018).

1.3.3 Gag

The *gag* region of the viral genome is responsible for coding for the Gag polyprotein (p55/assemblin), which is further processed by the viral protease into the following domains starting at the amino-terminal end: matrix protein (MA, p17), the capsid protein (CA, p24), the nucleocapsid (NC, p7) and the p6 protein (Sundquist and Krausslich, 2012; Foley, 2018). These domains are connected with small, flexible regions that contain cleavage sites for the viral protease (Figure 1.4).

The primary role of MA is to recruit the envelope protein and bind to the cellular plasma membrane. After virion maturation, the MA lines the inside of the viral envelope (Figure 1.2). MA targets the membrane using an early basic-rich region of residues and an N-terminal myristate group. The trimeric myristate groups are inserted into the membrane, with the basic-region interacting with the phospholipid heads of the membrane (Lee and Linial, 1994; Frankel and Young, 1998). The R/K30 residue in Gag, has been the subject of interest due to its prevalence

HIV-1/SIVcpz



HIV-2/SIVsm

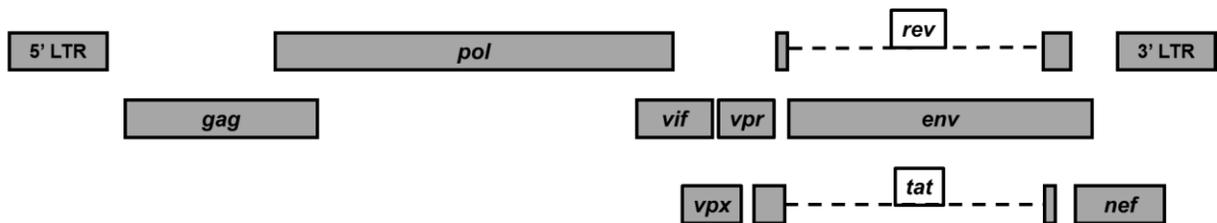


Figure 1.3: Genomic organization of HIV-1/SIVcpz and HIV-2/SIVsm viruses. Both groups of viruses have 9 genes. *Gag*, *pol* and *env* are the key structural and enzymatic genes. *Rev* and *tat* are regulatory genes characterized by the presence of an intron, while *vif*, *vpr*, *vpu/vpx*, and *nef* are accessory genes.

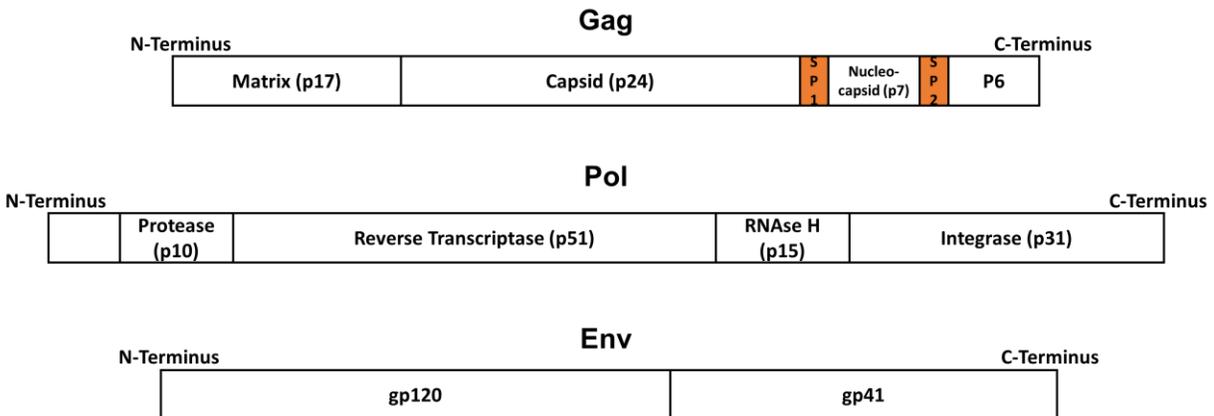


Figure 1.4: Genomic structure of *gag*, *pol*, and *env* genes. *Gag* encodes a polyprotein that is processed into the matrix (p17), capsid (p24), nucleocapsid (p7) and p6 structural proteins. *Pol* encodes a polyprotein that is processed into the protease (p10), reverse transcriptase (p51), RNase H (p15) and integrase (p31) enzymatic proteins. *Env* encodes the gp160 envelope protein that is processed into two subunits, gp120 and gp41.

amongst HIV-1 strains and absence amongst SIV strains (Wain et al., 2007; Sharp and Hahn, 2011). However, this residue was noticeably absent in HIV-1 group P strains, which had the methionine found in SIV strains (Plantier et al., 2009). Conversely, CA is involved in regulating the assembly of the immature virion through protein-protein interactions and constitutes the capsid within the virion that contains the viral RNA. Assembly of the capsid is mediated by the C-terminal domain, while the N-terminal domain of CA is more involved in viral uncoating through its interaction with cyclophilin A (Luban, 1996). CA also contains the major homology region (MHR), which is extremely conserved and forms a hydrophobic core. During reverse transcription of the viral genome, NC acts to chaperone the nucleic acids via two copies of a retroviral zinc finger motif, and helps to capture the viral genome during virion assembly by binding to the packaging signal, ψ . The NC acts nonspecifically, which allows it to completely coat the genomic RNA and may act as a shield against nucleases (Frankel and Young, 1998). Gag p6 is responsible for acting as a scaffold by binding viral Vpr and other proteins involved in the ESCRT pathway necessary for viral budding. The two spacer peptides (SP1, SP2) are involved in mediating viral maturation-based conformational changes (Sundquist and Krausslich, 2012).

1.3.4 Pol

The *pol* gene codes for the Pol (polymerase) portion of the Gag-Pol precursor polyprotein. *Pol* is separated from *gag* through the presence of a TTTTTT slippery site motif adjacent to a stem-loop structure (Foley, 2018). The stem-loop structure forces a delay with the translational machinery, which slips into a -1 ribosomal frameshift and begins the translation of Pol instead of the usual 3' end of the Gag protein. This occurs in approximately 5-10% of translating ribosomes (Frankel and Young, 1998). Following processing, Pol results in the protease (PR), reverse transcriptase (RT), RNase and integrase (IN) viral enzymes (Figure 1.4). PR is responsible for the

proteolytic processing of the viral gag and gag-pol polyproteins their cleavage sites and is needed for HIV virion maturation. Cleavage timing is highly regulated and can be influenced by the spacer peptides in *gag*, and as a result has been a significant target for therapeutic design (Wlodawer and Erickson, 1993; Pettit et al., 1994).

The RT is responsible for converting the viral RNA into a double-stranded DNA that is capable of being integrated into the host cell's genome. RT acts as both an RNA-dependent and DNA-dependent DNA polymerase at different points in order to synthesize the complementary DNA strand from the RNA template, and then generate the positive sense DNA strand using the first DNA strand as a template. Additionally, the RNase H domain acts as a ribonuclease by cleaving the RNA in the RNA-DNA double-helix prior to the DNA-dependent DNA polymerization. The RT is the major source of mutations for HIV/SIV due to its lack of proofreading, and low processivity (Jonckheere et al., 2000).

The integrase (IN) is necessary for the successful integration of the viral RNA into the host cell's genome and is a key component of the pre-integration complex (PIC), which also includes the Vpr and matrix proteins, among others. The PIC facilitates entry into the cell nucleus through the nuclear pore complex. From there, IN acts by removing two 3' nucleotides from both strands of the viral DNA and then joining the 3' ends to the 5' ends of the host cell DNA, while removing unpaired nucleotides from the 5' ends of the viral DNA (Katz and Skalka, 1994; Frankel and Young, 1998). Integration can occur repeatedly within a host cell, though it has been observed that integration does target certain genomic locations over others (Symons et al., 2018).

1.3.5 Vif

Known as the Viral infectivity factor, *vif* is found in the majority of lentiviruses, including bovine and feline lentiviruses and enhances the infectivity of the virus (Gifford, 2012; Foley, 2018). Additionally, Vif is involved in the recruitment of the p55^{gag} polyprotein to the plasma membrane at the final stages of viral replication (Simon et al., 1999). As a protein, it presents itself in two forms, a free-floating cytoplasmic form, and a membrane-associated form. One of the key functions of Vif is to counteract the activity of the APOBEC3 (A3) family of host restriction factors and as a result, can increase the infectivity of the virus a thousand-fold (Sheehy et al., 2002; Marin et al., 2003; Yu et al., 2003; Foley, 2018). The APOBEC3 family consists of A3A, A3B, A3C, A3D, A3F, A3G and A3H, with A3G, A3F, and A3H presenting the most potent antiviral activity against HIV-1 (Goila-Gaur and Strebel, 2008). Vif acts by binding A3G with its N-terminus and recruiting an E3 ubiquitin ligase complex to degrade it before APOBEC3G can be packaged into the newly forming virions (Sheehy et al., 2002; Marin et al., 2003; Yu et al., 2003). Much of the anti-host restriction factor activity of Vif is species-specific, as the Vif from SIVcpz and SIVgor is less effective against A3 proteins such as A3H (Etienne et al., 2013; Sauter and Kirchhoff, 2019). In the absence of Vif, A3 restriction factors are incorporated into the virions prior to budding, from which they can exert their viral restriction activity on the next cycle of infection (Walker et al., 2010). Vif is also essential for HIV-1 to replicate in macrophages and T cells which express high levels of A3 proteins (Chiu and Greene, 2006).

1.3.6 Vpr

Vpr, or Viral Protein R, is a smaller (14 kDa) accessory protein that enhances the replication and pathogenicity of HIV/SIV (Gonzalez, 2017). In HIV-1/SIVcpz, the *vpr* gene overlaps with the 3' end of *vif*, whereas in HIV-2/SIVsm, *vpr* overlaps with *vpx* (Figure 1.3). This

gene is conserved across all primate lentiviruses and bears a high degree of similarity to *vpx*. Vpr is incorporated into mature virions, where it associates with the p6 region of the Gag precursor polyprotein (Foley, 2018). The primary function of Vpr is to facilitate the importation of the viral pre-integration complex (PIC) by moving back and forth between the nucleus and cytoplasm (Kamata and Aida, 2000; Kogan and Rappaport, 2011). One of the other key functions of Vpr is to arrest the cell cycle of replicating cells at the G2 phase by interfering with DDB1 (DNA Damage-binding protein 1)-and-Cullin-4-associated Factor 1 (DCAF1), during which the viral promoter long terminal repeat (LTR) is very active. This can be induced with the oligomerizable or non-oligomerizable form of Vpr. The two functions of Vpr, namely enhancing macrophage infection and cell cycle arrest, appear to act independently, as evidenced by the fact that these functions are divided between Vpr and Vpx in HIV-2/SIVsm lineage viruses. The G2 cell cycle arrest is induced by an ATP deficiency precipitated by Vpr-mediated mitochondrial dysfunction (Macreadie et al., 1997; Masuda et al., 2016). This also results in decreased proteasomal function (Huang et al., 2010). Additionally, Vpr influences apoptosis, which occurs independently of the cell cycle arrest, and transcriptionally regulates immune function (Arunagiri et al., 1997; Ayyavoo et al., 1997). Vpr limits the proliferation of T cells and enhances the infection of primary macrophages, both of which can drastically contribute to CD4⁺ T cell depletion and the establishment of a latent reservoir (Romani et al., 2015). Additionally, Vpr is capable of crossing the blood-brain barrier and can contribute to the progression of neuroAIDS (Ferrucci et al., 2011). Binding to the Nuclear Factor of Activated T cells (NFAT) results in the increased susceptibility of T cells to infection with HIV (Hohne et al., 2016). Vpr also promotes the production of various inflammatory cytokines, including Tumor necrosis factor-alpha, IL-8 and IL-6, whereas, IL-12 production is inhibited (Kino et al., 1999; Sherman et al., 2000).

1.3.7 Vpx

Vpx is a unique protein that is only found in the HIV-2/SIVsm lineage of viruses, along with a few related SIVs, such as SIVrcm, SIVdrl and SIVmnd2 and is largely absent from the HIV-1 lineage of viruses (Yu et al., 1988; Sharp et al., 1996; Etienne et al., 2013; Foley, 2018). *Vpx* exhibits a high degree of sequence similarity to the *vpr* gene found in the African Green Monkey SIV (SIV_{AGM}) (Sharp et al., 1996; Etienne et al., 2013; Sauter and Kirchhoff, 2019). However, both *vpr* and *vpx* are preserved in the SIVsm lineage, despite their redundancy (Foley, 2018). Vpx is likely the result of a duplication event involving Vpr, possibly due to recombination (Sharp et al., 1996; Etienne et al., 2013; Foley, 2018; Sauter and Kirchhoff, 2019). Like Vpr, Vpx interacts with the Gag p6 region (Pancio and Ratner, 1998). While Vpx is needed for efficient replication of SIVsm in PBMCs, SIV-infected animals can progress to AIDS without it, likely due to its redundancy with Vpr (Wooley et al., 1997; Sauter and Kirchhoff, 2019). Additionally, Vpx plays an important role in antagonizing various host restriction factors, including SAMHD1, APOBEC3G and the HUSH complex (Laguette et al., 2011; Malim and Bieniasz, 2012; Strebel, 2013; Yurkovetskiy et al., 2018). Vpx binds to DCAF1 which starts the formation of an E3 ubiquitin ligase complex and in turn triggers the degradation of SAMHD1 (Srivastava et al., 2008; Bergamaschi et al., 2009; Hrecka et al., 2011; Laguette et al., 2011; Ahn et al., 2012; Laguette and Benkirane, 2012; Lim et al., 2012). As a result, without Vpx reverse transcription begins to fail due to the lack of available dNTPs (Laguette et al., 2011; Lahouassa et al., 2012).

1.3.8 Vpu

Viral protein U (VPU) is a protein that is only encoded in the HIV-1/SIVcpz lineage of viruses, as well as several other SIVs such as SIVgsn, SIVmus and SIVden (Bailes et al., 2003; Foley, 2018; Sauter and Kirchhoff, 2019). Vpu is wholly absent from the HIV-2/SIVsm lineage.

The functions of Vpu are split between degrading the CD4 molecule in the endoplasmic reticulum through ubiquitin ligase and adapter complex interactions that have largely been mapped to the C-terminal region of Vpu, and enhancing virion release from the infected cell by manipulating Env maturation (Frankel and Young, 1998; Neil et al., 2008). Additionally, Vpu has been found to downregulate MHC class 1 surface expression, as a means of escaping destruction by cytotoxic T cells (Kerkau et al., 1997). In the absence of Vpu, proper sorting and budding of virions can become disrupted (Neil et al., 2006). With HIV-1, Vpu serves an important role in counteracting the antiviral activity of tetherin, which is a feature also present in SIV_{gsn/mus/mon} but absent in SIV_{cpz} (Neil et al., 2008; Sauter et al., 2009; Sauter and Kirchhoff, 2019). Interestingly, the complex interactions mediating CD4 degradation are preserved in SIV_{cpz} Vpu, while HIV-1 group N viruses lost this function but preserved anti-tetherin activity, suggesting that these two functionalities are not directly linked (Sauter et al., 2012).

1.3.9 Tat

Tat (Trans-Activator of Transcription) acts as a trans-activator for viral gene expression and is one of the regulatory factors along with Rev. Unlike many other viral genes, Tat is characterized by two exon regions separated by an intron. Tat contains an arginine-rich domain that is involved in binding to TAR, with specific contact occurring between an arginine in Tat and a guanine in the TAR (Puglisi et al., 1992). By binding the TAR stem loop element, Tat enhances transcription and facilitates its initiation by blocking the action of the 5' LTR polyadenylation signal that normally results in premature termination (Selby et al., 1989; Rhim and Rice, 1994). This allows for a drastic increase in the production of viral transcripts due to the improved processivity of the polymerases, and is necessary to exponentially improve the output of mature virions (Zheng et al., 2005; Foley, 2018). Without Tat, viral transcription is typically limited to

several hundred base pairs (Frankel and Young, 1998). Due to its role in enhancing transcription, Tat is frequently one of the first viral proteins translated in HIV infection.

1.3.10 Rev

Similarly, to Tat, Rev (anti-Repression trans-activator) is a transactivating protein and regulatory factor that has two exons separated by an intron. Rev binds to the Rev responsive element (RRE) in *env* to stabilize viral mRNAs and facilitate nuclear export of both spliced and unspliced *gag* and *pol* transcripts, as well as the incompletely spliced *env*, *vif*, *vpr*, and *vpu* transcripts at the RRE by cycling between the cytoplasm and the nucleus (Favaro et al., 1998). This helps to overcome the cell's natural pathway in which splicing of transcripts occurs before being exported from the nucleus (Frankel and Young, 1998). By regulating the balance between early and late transcript quantities, an overall increase in virion production is observed (Hope, 1999). In the absence of Rev, HIV-1 late mRNA's are retained in the nucleus, which prevents their translation, and the expression of late stage structural proteins cannot occur until a sufficient amount of Rev is produced (Felber et al., 1990; Favaro et al., 1998). Interestingly, Rev's function is the most heavily conserved of the regulatory proteins in lentiviruses and is heavily involved in the precise shift of viral gene expression that dictates efficient protein synthesis (Pollard and Malim, 1998; Zheng et al., 2005; Foley, 2018).

Rev has been shown to contain a nuclear export signal (NES) that is rich in leucine and interacts with host proteins located at the nuclear pore (Fritz et al., 1995). Oligomerization of Rev allows for multiple NESs to localize with individual transcripts, thereby increasing the ability for the transcripts to be shuttled from the nucleus (Frankel and Young, 1998).

1.3.11 Env

The translation of *env* results in a precursor glycoprotein (gp160) that is processed into gp120, the external component of the glycoprotein, and gp41, the transmembrane component of the Env glycoprotein complex (Figure 1.4). These two components form a heterodimer and are noncovalently bound as functional trimers when expressed on the cell surface. As Env is the primary viral protein that extends beyond the viral envelope, it is frequently a target for novel therapeutics (Chan et al., 1997; Chan and Kim, 1998; Acharya et al., 2015; Foley, 2018). Gp120 is responsible for interacting with the CD4 receptor, and subsequently the CCR5 and CXCR4 chemokine receptors to facilitate new infections. Once gp120 has bound to the receptors during viral entry, gp41 undergoes the conformational change to enter a fusion-active state and facilitate fusion of the envelope with the cell membrane (Chan and Kim, 1998). The N-terminus of gp41 includes a glycine-rich, hydrophobic peptide crucial for successful fusion. Additionally, a looped region characterized by two cysteines is bordered by coiled coil heptad repeats. These coiled coils are formed into trimeric structures with the C helices forming the outer layer and the N helices forming an inner layer running in an antiparallel manner. This is the fusion-active conformation needed for viral entry (Chan et al., 1997; Chan and Kim, 1998). Env transcripts are translated at the endoplasmic reticulum, where the resulting polyproteins are complexed with CD4, the cell surface receptor that is downregulated by Vpu (Frankel and Young, 1998). Within the Env gene is the Rev responsive element (RRE), a motif that bridges the gp120 and gp41 regions and primarily interacts with Rev. This motif is not only found within HIV-1 and HIV-2, but has analogs within various other lentiviruses (Foley, 2018).

1.3.12 Nef

Located near the end of the genome, *nef* (Negative Factor) encodes an accessory protein that exists in myristylated and nonmyristylated forms in the cytoplasm near the plasma membrane, as well as the in the nucleus. While less critical for *in vitro* replication, Nef is necessary for effective viral spread *in vivo*, due to its role in the maintenance of high viral loads and progression to AIDS (Kestler et al., 1991; Chowers et al., 1994). Similarly to Vpu, Nef acts by downregulating CD4 and MHC class I molecules through clathrin-dependent endocytosis (Garcia and Miller, 1991; Aiken et al., 1994; Rhee and Marsh, 1994; Schwartz et al., 1996; Chaudhuri et al., 2007). This helps prevent the exposure of Env epitopes caused by CD4 that could cause the infected cell's degradation by cell-mediated cytotoxicity (Veillette et al., 2014). In primate lentiviruses causing non-pathogenic SIV infections, Nef has been shown to downregulate the T cell receptor complex to help mitigate cell death (Schindler et al., 2006). Additionally, Nef has PxxP motifs involved in Src kinase binding that enhance HIV growth, in addition to interacting with clathrin adapter protein (AP) complexes (Greenway and McPhee, 1997; Malim and Bieniasz, 2012; Foley, 2018). Nef is frequently packaged into newly produced virions in order to improve fusion of the virus with new host cells (Garcia and Miller, 1991). While Nef does not directly affect viral maturation, it has been shown that reverse transcription is drastically inhibited in the absence of Nef (Schwartz et al., 1995).

Additionally, in HIV-1, Nef plays a key role in mitigating the activity of several host-restriction factors such as SERINC3 and SERINC5 (Usami et al., 2015). It does this by preventing the SERINC molecules from being incorporated into the HIV virion (Rosa et al., 2015). In HIV-2, Nef is also involved in the inhibition of tetherin activity, though the nature of this action still

remains unclear and human tetherin is largely resistant to Nef proteins originating in SIVcpz and SIVgor (Jia et al., 2009; Zhang et al., 2009).

1.3.13 ASP

Within the *env* region is an antisense open reading frame (ORF) that could potentially produce a protein approximately 190 residues long. While no known function for ASP exists, its retention despite being in a region that is strongly selected for divergence suggests it may have a critical role for the virus (Foley, 2018). Interestingly, this ORF is not found in the majority of HIV-1 Group O and SIVcpz viruses, further obscuring its purpose (Miller, 1988; Torresilla et al., 2015).

1.4 Replication Cycle

One of the first phases of the HIV/SIV replication cycle is the entry into the host cells (Figure 1.5). The envelope glycoprotein extending from the mature virion's envelope uses the trimeric gp120 to bind to the host CD4 receptor, which induces a conformational shape change of the Env protein exposing the chemokine receptor binding domain of gp120 and facilitates further binding to seven-transmembrane G protein-coupled coreceptors (7TMRs) (Chan and Kim, 1998). CCR5 and CXCR4 are the two primary canonical coreceptors, with initial infections almost universally being established by CCR5-tropic (R5) viruses. However, once infection has successfully been established, it is common for the virus to switch to a CXCR4-tropism (X4), which is commonly associated with a rapid progression of disease and T cell loss (Coetzer et al., 2008; Gorry and Ancuta, 2011). This phenomenon appears to be limited to HIV, as most SIVs are found to rarely make use of CXCR4. Rather, it has been shown that SIVs such as SIVagm make use of a number of other 7TMR "orphan receptors" such as CXCR6, GPR15, and GPR1 to compensate (Wetzel et al., 2018). This is particularly true of SIVrcm, which primarily relies on CXCR6 for entry, and makes limited use of CCR5 due to a truncation mutation of CCR5

commonly found amongst red capped mangabeys. Additionally, CCR5-null sooty mangabeys were still found to be robustly infected with SIVsm, though the virus would still use CCR5 if present, further demonstrating the flexible use of coreceptors for SIVs in their natural hosts (Wetzel et al., 2018).

In contrast, pathogenic infections that occur in non-natural hosts, such as those seen in SIVmac were found to be crippled by blocking the canonical CCR5 receptor, suggesting that orphan receptor use is not as prevalent when compared to nonpathogenic natural host infections such as SIVsm. Interestingly, while SIVmac was found to be unable to use macaque CXCR6 efficiently, it could use the coreceptor from humans and other simians (Wetzel et al., 2018). This is attributed to a single amino acid change at S31R in the N-terminus of macaques CXCR6 (Wetzel et al., 2018). The inflexibility of non-natural infections is further reiterated by HIV-1's difficulty in infections individuals that have the CCR5 Δ 32 mutation (Liu et al., 1996). The N-terminal domain of CCR5 and CXCR4 primarily interact with the bridging sheet of HIV Env, through a number of sulfated tyrosine receptors, and the V3 loop of Env interacts with the second extracellular loop (ECL2) of the coreceptors (Wetzel et al., 2018).

Further conformational change of the gp41 subunit is induced by binding the chemokine receptors to gp120. This forms a pre-hairpin intermediate and the exposed fusion peptide is inserted into the host cell membrane. The extracellular region of gp41 constricts into a hairpin structure due to the interaction of the heptad repeat regions, and consequently the viral envelope is drawn to the host cell envelope, at which point fusion can occur. (Kowalski et al., 1987; Chan and Kim, 1998; Wyatt and Sodroski, 1998).

Upon fusing, the viral core consisting of viral RNA and proteins such as the reverse transcriptase, integrase and protease enter the cytoplasm through a newly formed pore and the viral

core moves to dock at the nuclear pore. From this point, the viral reverse transcriptase facilitates the production of double-stranded viral DNA from the RNA template after the bound viral proteins are shed, and the capsid is uncoated while the pre-integration complex (PIC) is formed from the viral DNA, integrase, matrix proteins, reverse transcriptase, Vpr, and the high-mobility group DNA-binding protein (HMGI) from the host cell (Miller et al., 1997; Zheng et al., 2005; Ganser-Pornillos and Pornillos, 2019). During replication, the reverse transcriptase degrades the viral RNA by acting as a ribonuclease, and then synthesizes the complementary strand of DNA in its place through DNA-dependent DNA polymerase activity. Reverse transcription is a frequent source of mutation for HIV/SIV, as the reverse transcriptase is highly error-prone and lacks proofreading capabilities (Bebenek et al., 1993). There is some evidence that suggests the capsid enables the reverse transcription by acting as a stabilizing force, and disruption of the capsid also disrupts reverse transcription (Forshey et al., 2002). At this point, the PIC is transported through the nucleus with the help of cellular microtubules and nuclear localization signals from the integrase, matrix, reverse transcriptase and Vpr proteins. Once inside the nucleus, integrase acts to incorporate the viral DNA into the host cell genome, usually in regions that are transcriptionally active (Schroder et al., 2002). This ensures that even if the cell is driven into latency, the virus has the potential to begin replicating again. Due to the presence of a promoter located within the 5' LTR, new viral RNA can be transcribed, and then is translated to form functional viral proteins in the cytoplasm, particularly the regulatory proteins Tat and Rev. For efficient transcription of the viral genome, Tat must bind to the TAR stem loop using positive transcription elongation factor b (P-TEFb) (Zheng et al., 2005). The resulting RNA transcripts are then shuttled from the nucleus into the cytoplasm by the Rev-RRE complex. While some copies of the RNA are used to form the genetic

material to be packaged in the virions, other copies are translated to produce the various additional viral proteins including Vif, Vpr and Vpu (Frankel and Young, 1998).

The Env proteins move through the host cell's secretory pathways after being cotranslationally inserted into the endoplasmic reticulum membranes. During this time, Env is assembled into trimeric complexes and then furin-cleaved into the gp41 and gp120 subunits (Rowell et al., 1995; Egan et al., 1996). At this point, Env arrives at the cellular membrane by vesicular transportation, while Nef enhances the degradation of surface CD4 molecules (Frankel and Young, 1998; Sundquist and Krausslich, 2012). In T cells, the newly synthesized viral components migrate to the cellular membrane, where they are packaged in lipid rafts and bud off from the membrane to form a new virion, whereas in macrophage and some subsets of dendritic cells, the packaging and assembly occurs at endosomal membranes prior to budding from the plasma membrane (Nguyen and Hildreth, 2000; Raposo et al., 2002). Within the newly budded virion, the packaged protease cleaves the immature polyproteins of Gag and Pol to create their constitutive components, including the matrix, capsid and nucleocapsid proteins. After processing by the protease, the capsid proteins assemble together to form the conical capsid around the viral RNA, and mature virion is now ready to infect a new host cell (Zheng et al., 2005).

1.5 Restriction Factors

There are many potential barriers to successful viral replication within a host, including a number of so-called host-restriction factors that have specifically evolved for an antiviral function. In a similar manner to differences in cellular receptor structure, many host restriction factors have evolved to be species-specific, and it has been speculated that viruses are engaged in an evolutionary arms race with their target hosts to evolve new accessory proteins and host restriction

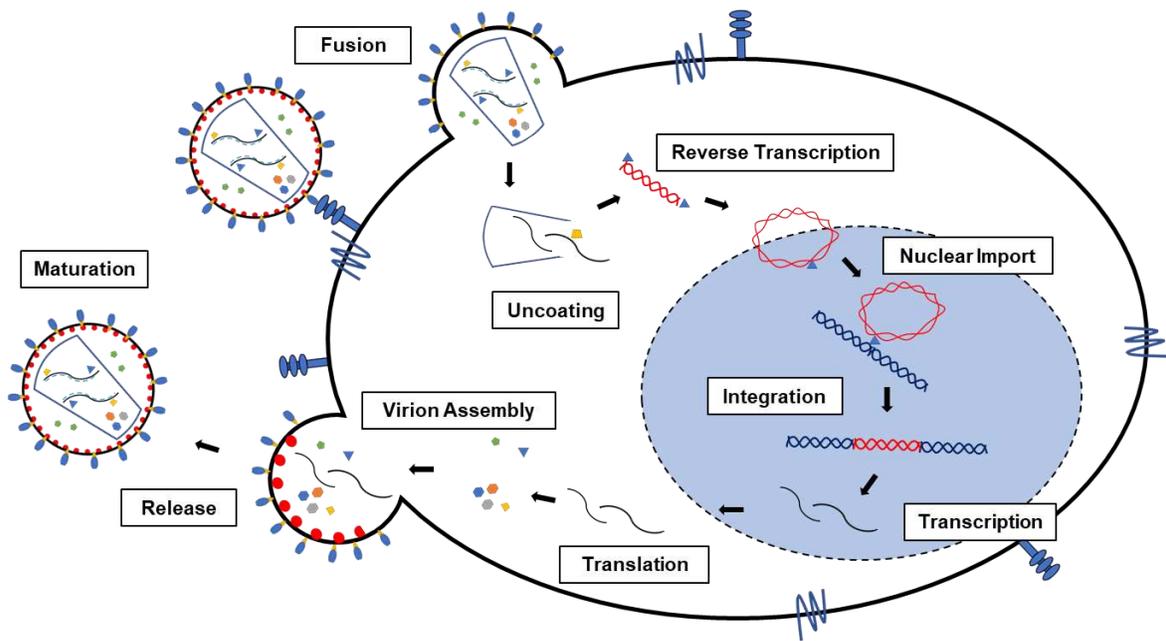


Figure 1.5: Replication cycle of HIV/SIV. Fusion: HIV/SIV Env binds to CD4/CCR5/CXCR4 to bring the viral envelope and cell membrane together. Uncoating: The capsid enters the cytoplasm and is disassembled, freeing the viral RNA. Reverse Transcription: Viral RT begins synthesizing cDNA from the viral RNA template. Nuclear Import: The PIC is formed to guide the viral cDNA into the cellular nucleus. Integration: Viral integrase inserts the cDNA into the cellular genome. Transcription: mRNA transcripts of the viral genome are produced and shuttled out of the nucleus. Translation: viral proteins are produced from the transcripts, beginning with Rev and Tat. Virion Assembly: Env and Gag molecules are shuttled to the cell surface, and the viral proteins and RNA are packed into a forming virion. Release: the cellular membrane is pinched, allowing the packaged virion to bud off of the cellular membrane. Proteolytic cleavage of the packaged polyproteins by the viral protease and assembly of the capsid occurs to produce a fully mature virion.

factors in an effort to gain an evolutionary advantage. For example, it is believed that while SIV_{mac} is pathogenic in rhesus macaques, it is inhibited in New World monkeys due to the presence of lentivirus susceptibility factor (LV1) (Besnier et al., 2002). As a result, prior to being capable of successfully infecting humans, the genomes of the various SIVs had to undergo various genetic changes to overcome these new host restriction factors (Sauter and Kirchhoff, 2019).

Over thousands of years, humans and other animals have developed a wide repertoire of host restriction factors that are capable of inhibiting retroviral replication at multiple points within the viral replication cycle. In response, HIV has finely tuned its accessory proteins to counteract these mechanisms of antiviral activity. While new host restriction factors such as ERM_{anI}, TSPO, GBP5 and ZAP are constantly being discovered, the most prominent examples and their viral interactions have been characterized below (Ghimire et al., 2018).

1.5.1 TRIM5 α /TRIMCYP

The tripartite motif-containing protein 5 (TRIM5) family containing both the alpha isoform (TRIM5 α) and cyclophilin A associated form (TRIMCyp) are a group of innate host restriction factors that serves as an early, post-entry, line of defense against HIV and other retroviruses in mammalian cells (Hatzioannou et al., 2004; Malim and Bieniasz, 2012; Ganser-Pornillos and Pornillos, 2019). Their primary function is to recognize and bind directly to the viral capsid using either a B30.2/SPRY domain or cyclophilin A prior to the uncoating of the capsid, which may result in a loss of stability of the capsid (Stremlau et al., 2006; Black and Aiken, 2010; Ganser-Pornillos and Pornillos, 2019). The viral capsid is responsible not only for protecting the nucleic acids within, but helps to facilitate reverse transcription and recruits the necessary cellular factors for the importation of nucleic acids into the nucleus (Ganser-Pornillos and Pornillos, 2019). TRIM5 α can only bind the viral capsid when the capsid is in an assembled form rather than the

soluble form and therefore acts as a stopgap immediately following viral entry (Dodding et al., 2005). The disruption of the capsid either through premature uncoating or total dissolution of the capsid results in abortive reverse transcription and consequently a failure to accumulate the viral RNAs needed to be packaged into new virions (Forshey et al., 2002; Stremlau et al., 2004; Black and Aiken, 2010). Additionally, TRIM5 can induce an antiviral state through a ubiquitin-dependent innate immune signaling pathway (Pertel et al., 2011).

As seen with other host-restriction factors, TRIM5 α has species-specific functionality (Sawyer et al., 2005). HIV-1 can be more effectively restricted by TRIM5 α originating from old world monkeys when compared to human-origin TRIM5 α (Stremlau et al., 2004). Conversely, SIVs were found to be less effectively blocked by human TRIM5 α , likely due to a difference originating in the capsid itself (Stremlau et al., 2006; Wu et al., 2013). In fact, inability of human TRIM5 α to counteract HIV-1 can largely be attributed to poor core recognition, that has been rectified experimentally through a single point mutation within the hypervariable region of the SPRY domain (Sawyer et al., 2005; Stremlau et al., 2005). Furthermore, selective pressures from cytotoxic T cells have been shown to give rise to human TRIM5 α -sensitive variants of HIV-1 (Granier et al., 2013). Therefore, it can be assumed that resistance to TRIM5 proteins is dictated by the viral capsid protein sequence (Hatzioannou and Evans, 2012). While TRIM5 α serves as a poor defense mechanism against naturally occurring retroviruses within a given host, it is quite effective against cross-species transmissions and is a significant barrier for SIVs attempting to replicate in human cells (Hatzioannou et al., 2006).

1.5.2 SAMHD1

Sterile α motif and HD domain-containing protein 1 (SAMHD1) is a host restriction factor that acts as a deoxynucleoside triphosphates (dNTPs) triphosphohydrolase that inhibits the reverse

transcription of viruses such as HIV-1 by reducing the dNTPs that are freely available within the cytoplasm of the host cell (Lahouassa et al., 2012). This becomes a particularly effective method of restriction for viruses with a reduced RT affinity for dNTPs, and the lack of direct interaction with a specific viral component means that it can theoretically function against a broad spectrum of viruses.

The viral protein Vpx has been found to be responsible for degrading SAMHD1, which leads to an increase in the available dNTP pool and improved reverse transcription (Laguette et al., 2011; Lahouassa et al., 2012). While both human and rhesus macaque SAMHD1 are counteracted by Vpx, it is not directly involved in the pyroptosis of bystander cells (Goujon et al., 2008; Lahouassa et al., 2012; Luo et al., 2019). The bulk of SAMHD1 activity occurs primarily in nondividing monocyte-derived macrophages and dendritic cells, because they do not require as high of a concentration of dNTPs as dividing cells (Mlcochova et al., 2017). In contrast to HIV-2 and its related SIVs, HIV-1 is heavily impacted by SAMHD-1 restriction, largely due to the lack of Vpx (Laguette et al., 2011). However, as a large proportion of HIV-1 viral replication occurs in T cells, it is possible that evolutionary pressures have not necessitated the development of a Vpx analog in HIV-1 as of yet.

1.5.3 APOBEC3

The apolipoprotein B-editing catalytic subunit-like 3 (APOBEC3, A3) family serve as host-restriction factors that function by causing the hypermutation of the viral genome with a cytidine deaminase (Chiu and Greene, 2006). In addition to this primary function, the A3 family has other less well studied cytidine deaminase independent effects that include the inhibition of tRNA-primed initiation of RT, RT-catalyzed DNA elongation, strand transfer, and DNA synthesis and integration (Dang et al., 2006; Guo et al., 2006; Iwatani et al., 2007; Luo et al., 2007; Mbisa

et al., 2007). Within this family, APOBEC3G has the most well characterized anti-HIV activity, but APOBEC3F and APOBEC3H have also been shown to suppress HIV-1, albeit to a lesser degree than APOBEC3G (Holmes et al., 2007; Albin and Harris, 2010). In the absence of Vif, A3 proteins interact with viral RNA and are packaged within budding virions, resulting in the extensive deamination of cytidine residues in the minus-single stranded viral DNA formed during reverse transcription in the next targeted host immune cell (Chiu and Greene, 2006; Koning et al., 2009). This results in the accumulation of a detrimental level of G→A mutations beginning at the 3' end that disrupts the proper production of viral proteins starting with Nef leading to a phenomena known as hypermutation (Chelico et al., 2006). APOBEC3G contains two domains to mediate its activity: An amino-terminal zinc-coordinating “Z domain” that facilitates packaging into the HIV virion in coordination with the Gag nucleocapsid, and the carboxy-terminal Z domain that dictates the catalytic deamination activity. Other A3 proteins contain a single deaminase domain (Bogerd and Cullen, 2008; Malim, 2009; Albin and Harris, 2010; Malim and Bieniasz, 2012).

APOBEC3G is robustly antagonized by the viral protein Vif. However, it has been suggested based on findings of hypermutated HIV-1 sequences in patients that A3 may be able to escape Vif inhibition in some circumstances (Malim, 2009; Compton et al., 2012). The degradation mechanism of A3 has been extensively studied and begins through the interaction of A3 with Vif. Following binding of the A3 and Vif proteins through specific domains, polyubiquitylation of the Vif-A3 complex results in the targeted destruction of A3 by the 26S proteasome (Marin et al., 2003; Sheehy et al., 2003; Yu et al., 2003). The interplay between A3 proteins and Vif is species specific, and is primarily dictated by several discontinuous elements at the N-terminus (Gaddis et al., 2004; Compton et al., 2012). A single mutation change at residue 128 dictates the ability of

Vif to antagonize APOBEC3G in SIV_{AGM} versus HIV-1, which further highlights that genetic changes are critical when adapting to a new host's restriction factors (Gaddis et al., 2004; Malim, 2009; Albin and Harris, 2010).

1.5.4 SERINC3/SERINC5

Recently discovered, serine incorporators (SERINC) 3 and 5 function as carrier proteins that primarily incorporate serine and help promote the production of sphingolipids and phosphatidylserine in the cellular membrane (Ghimire et al., 2018). The mechanism of action for SERINC3/5 is still being studied, but it appears to impact the translocation of the viral particle's contents to the cytoplasm and results in a reduction of successful reverse transcription. Budding virions have been shown to incorporate SERINC5 and have a reduced infectivity as a consequence (Rosa et al., 2015; Usami et al., 2015). It has also been suggested that SERINC3/5 may alter the shape of viral Env proteins, while disrupting fusion pores in the cellular membrane to block viral fusion (Sood et al., 2017).

Both SERINC5 and SERINC3 are naturally counteracted by HIV-1 Nef (Usami et al., 2015). SERINC5 expression on the cellular membrane becomes reduced when Nef is intact and is instead localized to endosomes. This antagonism is maintained through many primate lentiviruses, and is even more pronounced in certain SIVs (Rosa et al., 2015). The V1/V2 and V3 loops of Env have also been implicated in SERINC5 antagonism, though not to nearly the same degree as Nef (Beitari et al., 2017). So far, there is little evidence to suggest a prolonged evolutionary selection for SERINC3/5 as antiviral defenses when compared to other host restriction factors, which may explain the lack of species-specificity for it (Ghimire et al., 2018).

1.5.5 BST-2/Tetherin

Tetherin is a host restriction protein that inhibits the budding of virions from the host cell membrane and traps the virions at the cell surface (Neil et al., 2008). The protein itself is a single-pass transmembrane protein, with an extracellular portion that forms an α helix and a dimeric coiled-coil structure (Hinz et al., 2010; Schubert et al., 2010; Yang et al., 2010). This domain has a high tolerance for mutation and has been shown to remain functional even when the entire domain was replaced by a similar, nonhomologous, artificial domain (Perez-Caballero et al., 2009). Tetherin functions by colocalizing with the virus at the cell surface and then gets incorporated into the virion envelope as it begins to bud from the cellular membrane. It is likely that non-covalent interactions between the tetherin dimers on the viral envelope and on the cell membrane are the cause of the tethering activity which prevents virion release (Perez-Caballero et al., 2009).

The high degree of variability of the extracellular domain allows tetherin to maintain antiviral activity against a wide range of viruses (Jouvenet et al., 2009). As a consequence, different viruses have evolved independent responses to counteract tetherin activity. In cases of HIV-1 infection, tetherin is counteracted by the presence of the viral protein Vpu, which helps to limit the amount of tetherin present at the cellular membrane during viral budding (Neil et al., 2006; Van Damme et al., 2008). In contrast, the viral proteins Nef and Env have been found to counteract tetherin in a number of SIVs including SIVcpz and HIV-2, though their mechanism of action is less well understood (Jia et al., 2009; Le Tortorec and Neil, 2009; Zhang et al., 2009).

Similarly to other host restriction factors, the antagonism of tetherin is highly species-specific, as changes in the cytoplasmic tail of tetherin results in a degree of resistance to Nef (Jia et al., 2009; Zhang et al., 2009; Serra-Moreno et al., 2011; Serra-Moreno and Evans, 2012). It has

been speculated that the role of tetherin antagonism has jumped between the various viral proteins as the virus itself jumped between different host species, as evidenced by SIVcpz and SIVsm using Nef, while HIV-1 uses Vpu and HIV-2 uses Env to counteract tetherin (Lim et al., 2010; Sauter and Kirchhoff, 2019).

1.6 Animal models for HIV/SIV infection

HIV has a relatively narrow cellular tropism, which presents multiple challenges when modeling its infection. Human primary cells and cell lines can provide a means to study the viral replication kinetics, initial responses to therapeutics and other factors, but these *in vitro* models have a number of limitations. Primary cells are restricted by a short frame of viability in culture, and as most cell lines are derived from cancerous cells, they represent a degree of artifice that is less likely to be seen in natural cases of infection. Furthermore, despite our ever-increasing understanding of cellular and immune interactions within a living organism, there are countless other interactions remaining to be documented. These interactions are a key component that is absent from cell cultures containing a limited range of cell types (Denton and Garcia, 2011).

These limitations necessitate the development of animal models that can better replicate the conditions of natural SIV/HIV infections and enable the further development of targeted therapeutics. Due to the presence of their own native retrovirus known as feline immunodeficiency virus (FIV) that can cause AIDS-like symptoms, some researchers have found success in using domestic cats as an effective surrogate for modeling HIV infection (Pedersen et al., 1987; Hatzioannou and Evans, 2012). Advanced infections of FIV can produce CD4⁺ T cell depletion and neurological pathologies, both of which have been seen in human AIDS patients, and FIV has been extremely useful in the development of antiretroviral therapeutics (Dow et al., 1990; Torten et al., 1991; Arai et al., 2002). However, the use of FIV as a surrogate does have several limitations.

FIV is more diverged than HIV and SIV are, and as a result is missing several key accessory genes found in HIV (Hatzioannou and Evans, 2012). Additionally, FIV utilizes different cellular receptors for viral entry and therefore displays a different tropism than HIV (Shimojima et al., 2004). Altogether, this limits the usefulness of using FIV to study specific viral interactions unique to primate lentiviruses.

The most obvious model to study SIV and HIV would be the primates from which these lentiviruses originate, particularly the rhesus macaques, due to their ability to be infected by the pathogenic SIV_{mac}, which causes simian AIDS in rhesus and pigtailed macaques (Daniel et al., 1985; Murphey-Corb et al., 1986; Klatt et al., 2012). Historically, chimpanzees, the original hosts of SIV_{cpz}, have been used as a surrogate for studying a wide range of areas related to human health due to their closely related ancestry to humans and have been shown to develop an AIDS-like immunopathology in cases of SIV_{cpz} infection (Keele et al., 2009). However, research on chimpanzees today is limited because of the huge cost of maintaining the animals, and their endangered status (Denton and Garcia, 2011). Additionally, direct infection of chimps with HIV-1 rarely led to advanced AIDS development, which highlights that key differences in the immune environments of humans and chimps does exist (O'Neil et al., 2000; Hatzioannou and Evans, 2012). In fact, many of the naturally occurring SIVs do not produce a pathogenic infection, likely due to their presumed long history of co-evolution with their host (Chahroudi et al., 2012; Hatzioannou and Evans, 2012). However, sooty mangabeys and African green monkeys have both been used extensively as models for HIV because their natural SIVs recapitulate CD4⁺ T cell depletion of the mucosa and prolonged high levels of viral replication, though they fail to induce the same degree of chronic immune activation and lymph node destruction as seen in cases of HIV infection (Chahroudi et al., 2012). Differences in host restriction factors present a barrier for

directly translating research on SIV in nonhuman primates, to HIV in humans (Sauter and Kirchhoff, 2019). Nonhuman primates also fail to create the immune pressures unique to human cells that led to the emergence of HIVs. In order to determine the genetic changes that arose in order for HIV to replicate in humans, a different animal model is necessary.

Humanized mice are typically characterized by the presence of either human transgenes, or xenografted human tissue/cells within living mice (Denton and Garcia, 2011). Mice engrafted with human CD34⁺ hematopoietic progenitor cells are particularly useful for studying HIV, as these mice are capable of *de novo* generation of human immune cells that fall within HIV's tropism (Berges et al., 2006; Brehm et al., 2010; Akkina, 2013b; Brehm et al., 2014). It is possible to utilize multiple human donors to generate different immune cohorts of mice, which allows a better means of exploring the natural immune diversity present among humans and which could have a drastic effect on the dynamics of HIV infection (Denton and Garcia, 2011). Humanized mice are also capable of being infected with HIV through different routes of infection, which allows the mice to serve as a model for specific therapeutic interventions (Berges et al., 2008; Veselinovic et al., 2016).

Although extremely useful for studying a wide range of topics, humanized mice are not without their limitations. As most of these mice have an immunocompromised genetic background, they are prone to unrelated opportunistic infections. Additionally, due to the nature of the transplantation of human tissue, there can be a wide range in the level of human cell reconstitution, and the degree of sustained immune response depends on which model is being used (Denton and Garcia, 2011). Furthermore, mice have a shorter life-span relative to other models, which combined with the transient nature of the human cell/tissue engraftment makes long-term studies challenging.

One of the earliest attempts at generating a mouse with a human immune system involved the use of human peripheral blood lymphocytes (PBLs). When injected into immunodeficient mice (Hu-PBL-SCID), the human cells are able to successfully engraft primarily T cells, with a small number of B cells and other types of immune cells (Mosier et al., 1988; Ito et al., 2002). While useful for studying T cell interactions, this model is limited by the pervasive graft-versus-host disease that sets in within a few weeks following engraftment (King et al., 2009; Brehm et al., 2014).

Later mouse models began to incorporate human CD34⁺ hematopoietic stem cells (HSCs) as a means of engrafting human immune cells (Hu-SRC-SCID) (Lapidot et al., 1992; Brehm et al., 2014). These cells are typically derived from either human fetal liver, or umbilical cord blood, which provide a more robust engraftment than cells derived from an adult (Matsumura et al., 2003; Lepus et al., 2009). Early iterations involved injecting these cells into adult mice, but later variants involve intrahepatic injection of the cells into mouse neonates, with dramatically better human cell engraftment as a result (Figure 1.6) (Berges and Rowan, 2011; Akkina, 2013b). These hu-HSC mice generate a more complete complement of human immune cells relative to the Hu-PBL-SCID mice, but are still somewhat limited because the generated T cells undergo education in the mouse thymus (Watanabe et al., 2009). Regardless, this model has been a useful tool for studying human pathogens including HIV (Shultz et al., 2012). Further exploration of the genetic background of the mice led to the development of Non-obese diabetic/SCID (NOD/SCID) which lack the ability to generate T and B cells and a reduced capacity for NK cell production, and BALB/c-Rag2^{-/-}γc^{-/-} double knockout (NRG) mice which do not express the interleukin 2 receptor common gamma chain and cannot produce T, B and NK cells (Mazurier et al., 1999; Ueda et al., 2000). This allows a better reconstitution of human cells to fill the empty immune niche (McDermott et al., 2010).

The bone marrow/liver/thymus (BLT) humanized mice provide one of the most robust and comprehensive human immune models available and involves the physical transplantation of fetal liver and thymus tissue under the kidney capsule of immunocompromised adult mice, followed by intravenous injection of autologous bone marrow-derived CD34⁺ hematopoietic stem cells (Lan et al., 2006; Akkina, 2013b). Due to the presence of the xenografted human thymus tissue, the generated T cells are capable of undergoing education in the context of human leukocyte antigen (HLA), whereas in other models, education occurs in the mouse thymus, absent of HLA (Lan et al., 2006; Manz, 2007). Following engraftment, BLT mice have been shown to produce T cells, B cells, NK cells, macrophages, monocytes and dendritic cells (Lan et al., 2006; Melkus et al., 2006; Denton and Garcia, 2011; Garcia and Freitas, 2012; Ito et al., 2012; Shultz et al., 2012; Akkina, 2013b; a).

In addition to virological and flow-cytometry based analysis, investigators have looked at X4 and R5-tropic HIV-1 pathology in the thymus and lymph node tissues in hu-mice, wherein a clear depletion of CD4 T lymphocytes was observed (Berges et al., 2006). This was also verified in hu-mice that were infected by mucosal and rectal challenges (Berges et al., 2008). Additionally, viral presence in the lymph nodes, thymus, gut, and splenic tissues has been demonstrated using *in situ* hybridization and immunohistochemistry (Baenziger et al., 2006; Berges et al., 2006; Berges et al., 2008; Denton and Garcia, 2011). In some cases, cervical lymph node hyperplasia, CD8⁺ cell infiltration/expansion and the presence of pathomorphological changes in the structure of the lymph nodes have been observed in hu-mice infected with HIV (Gorantla et al., 2007). All of this further supports the idea that hu-mice can serve as an effective surrogate for studying the pathology of HIV infection.

To experimentally address the research aspects of the evolution of SIV into HIV as well as the cross-species transmission of nonhuman primate lentiviruses themselves, the following project was conceived as three independent but related investigations. The first investigation sought to understand the emergence of HIV-2 from SIVsm. Here, SIVsmE041 was used to inoculate humanized mice and then serially passage over the course of eight generations. Viral adaptation was assessed with a combination of plasma viral loads, CD4⁺ T cell levels, and deep sequencing of the viruses to determine adaptive mutations. The results of these experiments have been previously published and constitute chapters 2, 3 and 4 of this dissertation (Schmitt et al., 2017; Schmitt et al., 2018; Curlin et al., 2020a). The second investigation involved using three strains of SIVcpz (MB897, EK505 and LB715) to evaluate the evolution and transmission of SIVcpz into HIV-1 with serial passaging in humanized mice. Each virus was assessed for adaptation in the same manner as described for the first investigation. The experiments entailing the first two serial passages of these viruses have already been published in peer-reviewed journals, and make up the contents of chapters 5, 6 and 7 of this dissertation (Curlin et al., 2020b; Schmitt et al., 2020a; Schmitt et al., 2020b). The third investigation sought to understand the relationship between SIVsm, the progenitor of HIV-2, SIVmac, a virus that originated from SIVsm but crossed over into rhesus macaques, and SIVhu, a direct example of an SIV successfully infecting and replicating in a human being. We serially passaged each of these viruses in humanized and assessed their adaptation as described previously in an attempt to shed light on the commonalities and differences between these viruses as they undergo further cross-species adaptation. Chapter 8 of this dissertation details the results of these experiments.

Ultimately, by identifying the critical genetic changes that arise in the representative progenitor lentiviruses as they are continually exposed to a human immune environment, we hope

to provide a direction for future therapeutics to be explored in the search for better treatment and eventually a cure for HIV.

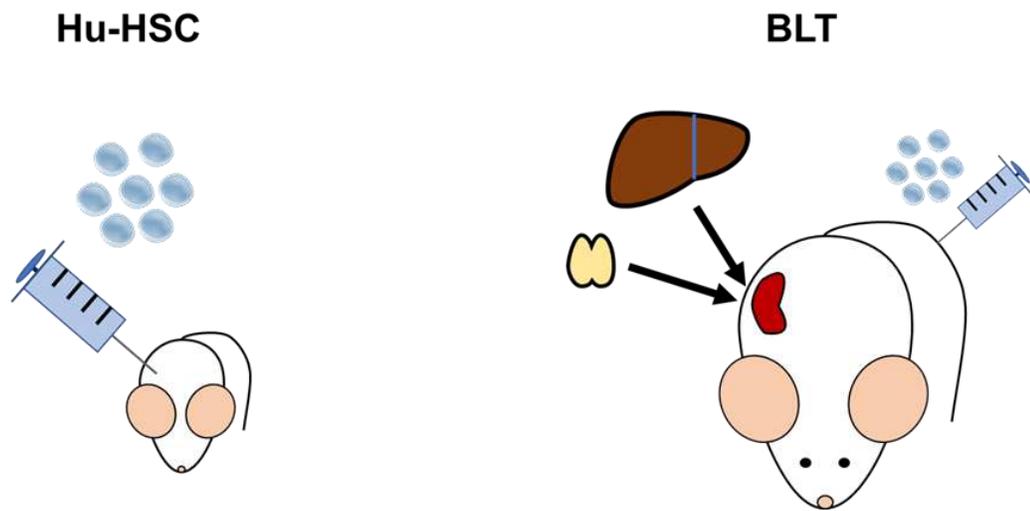


Figure 1.6: The Generation of Hu-HSC and BLT humanized mice. Hu-HSC mice are created by sublethally irradiating neonatal Balb/c $Rag1^{-/-}\gamma c^{-/-}$ or Balb/c $Rag2^{-/-}\gamma c^{-/-}$ mice which are then injected with human $CD34^{+}$ hematopoietic stem cells intrahepatically. BLT mice are generated by sublethally irradiating the mice and then surgically implanting fetal liver and thymic tissue under the kidney capsule. Each mouse then receives an intravenous injection of autologous $CD34^{+}$ hematopoietic stem cells. Both hu-HSC and BLT mice are then screened using flow cytometry after 8-12 weeks to determine human immune cell engraftment.

CHAPTER 2: MODELING THE EVOLUTION OF SIV SOOTY MANGABEY PROGENITOR VIRUS TOWARDS HIV-2 USING HUMANIZED MICE¹

2.1 Summary

HIV-2 is thought to have originated from an SIV progenitor native to sooty mangabeys. To model the initial human transmission and understand the sequential viral evolution, humanized mice were infected with SIVsm and serially passaged for five generations. Productive infection was seen by week 3 during the initial challenge followed by chronic viremia and gradual CD4⁺ T cell decline. Viral loads increased by the 5th generation resulting in more rapid CD4⁺ T cell decline. Genetic analysis revealed several amino acid substitutions that were nonsynonymous and fixed in multiple hu-mice across each of the 5 generations in the nef, env and rev regions. The highest rate of substitution occurred in the nef and env regions and most were observed within the first two generations. These data demonstrated the utility of hu-mice in modeling the SIVsm transmission to the human and to evaluate its potential sequential evolution into a human pathogen of HIV-2 lineage.

2.2 Materials and Methods

2.2.1 Generation of humanized mice

Human fetal liver-derived CD34 cells were isolated, column purified (Miltenyi Biotec, San Diego, CA) and cultured as previously described (Akkina et al., 1994; Bai et al., 2000; Veselinovic et al., 2016). CD34⁺ purity was assessed by flow cytometry. Neonatal Balb/c Rag1^{-/-}γc^{-/-} or Balb/c

¹ Curlin, J., Schmitt, K., Mohan Kumar, D., Remling-Mulder, L., Stenglein, M., O'Connor, S., Akkina, R. (2017). Modeling the evolution of SIV sooty mangabey progenitor virus towards HIV-2 using humanized mice. *Virology* 510, 175-184. doi: 10.1016/j.virol.2017.07.005.

Rag2^{-/-}γc^{-/-} mice were preconditioned by irradiation at 350 rads and injected intrahepatically with 0.5–1 x 10⁶ human CD34⁺ cells per mouse (Berges et al., 2008; Veselinovic et al., 2016). Transplanted mice were then screened at 10–12 weeks post-reconstitution for human cell engraftment. Peripheral blood was collected, and the red blood cells were lysed using the Whole Blood Erythrocyte Lysing Kit and the manufacturer's instructions (R & D Systems, Minneapolis, MN). Fractioned white blood cells were stained with mouse anti-human CD45 FITC (eBioscience), CD3 PE (eBioscience) and CD4 PE/Cy5 (BD Pharmingen, San Jose, CA) for FACS analysis to confirm human cell engraftment (Berges et al., 2006; Veselinovic et al., 2016). Mice were maintained at the Colorado State University Painter Animal Center. The studies conducted in this publication have been reviewed and approved by the CSU Institutional Animal Care and Use Committee.

2.2.2 SIVsm primary isolate SIVsmE041

SM E041 was a male sooty mangabey born in January 1981 at the Yerkes Primate Research Center (YNPRC) for experimental leprosy infection and transferred in 1983 to the Tulane National Primate Research Center (TNPRC) (Gormus et al., 1995a; Gormus et al., 1995b). E041 was infected with SIVsm by natural spread between group-housed mangabeys before leprosy experiments began at YNPRC and neither clinical nor histological evidence of leprosy was seen during 18 years of observation (Fultz et al., 1986; Murphey-Corb et al., 1986; Ling et al., 2004). SM E041 was 21 years old when euthanized with clinical signs of AIDS that included: high viral loads, low CD4⁺ T-cell counts, SIV giant cell disease and B-cell lymphoma (Ling et al., 2004). The SIVsmE041 used in this study (GenBank accession HM059825.1) was propagated in sooty mangabey PBMC, making it a true primary isolate that is comprised of a population of closely

related viruses. The tropism of this virus was determined to use the major co-receptor CCR5 (Ling et al., 2004).

2.2.3 SIVsmE041 infection of humanized mice and viral determination by qRT-PCR

Mice with high (>70% CD4⁺ T cells) human hematopoietic cell engraftment levels were used. At 16 weeks post engraftment, 200 µl of SIVsmE041 (TCID₅₀, 811) was injected into five hu-HSC mice intraperitoneally (i/p). All humanized mice used in subsequent passages were inoculated with the same amount of virus using methodology as described above. Peripheral blood was collected weekly using non-heparinized capillary tubes to assess plasma viral load by tail vein puncture and transferred immediately to EDTA-containing vacutainer tubes (BD Biosciences, San Jose, CA). PBS was added to the plasma, with a final volume of 150 µl, and centrifuged at 400 x g for 5 minutes. Plasma was removed and viral RNA was extracted from the plasma using the E.Z.N.A Viral RNA kit (OMEGA bio-tek, Norcross, GA). Viral loads were determined using the iScript One-Step RT-PCR kit with SYBR Green and the manufacturer's instructions (Bio Rad, Hercules, CA). Primers were designed based on a conserved region in SIVsmE041 *ltr* (GenBank accession number: HM059825.1). Forward (5'-CCACAAAGGGGATGTTATGGGG-3') and reverse (5'AACCTCCCAGGGCCAATCT-3') primers were used in a qRT-PCR reaction with the following cycling conditions: 50 °C for 10 min, 95 °C for 5 min, followed by 40 cycles of 95 °C for 15 sec and 60 °C for 30 sec in the Bio Rad C1000 Thermal Cycler with a CFX96 Real-Time System (Bio Rad, Hercules, CA). The standard curve was prepared using a series of 10-fold dilutions of viral SIVsmE041 *gag* at a known concentration. The sensitivity of this assay was 100 RNA copies per ml.

2.2.4 Determination of CD4⁺ T cell levels

Flow cytometry was used to assess human cell engraftment levels in humanized SIVsm infected and uninfected mice. Peripheral blood was collected in heparinized capillary tubes by tail vein puncture bimonthly. 5 ul of FcγR-block (Jackson ImmunoResearch Laboratories, Inc. West Grove, PA) was added to the blood cell pellet and then stained with fluorophore conjugated hCD45-FITC (eBioscience), hCD3-PE (eBioscience) and hCD4-PE/CY5 (BD Pharmingen, San Jose, CA) for 30 min. Erythrocytes were lysed using a RBC lysing kit (BD sciences). After lysing, stained cells were washed twice with washing buffer (BD Sciences) and fixed in 1% paraformaldehyde. Stained cells were analyzed using BD Accuri C6 FACS analyzer. In order to measure CD4⁺ T cell depletion in SIV infected mice, CD3⁺ T cell levels were calculated as a ratio of the entire CD45⁺ (lymphocyte common antigen marker) population. The CD4⁺ T cell level was then determined based on the ratio of the entire CD3⁺ T cell population. . Engraftment levels of CD45, CD3 and CD4 were measured before infection as a control. CD4⁺ T cell decline was assessed using a two-tailed Student's t-test (<0.05) to compare the two groups of mice, infected and uninfected. All flow data was analyzed using the FlowJo software package (FlowJo LLC, Ashland, Oregon).

2.2.5 Illumina sequencing and analysis

Plasma viral RNA from week 4, 16 and 28 were extracted using the E.Z.N.A Viral RNA kit (Omega bio-tek, Norcross, GA) and amplified for 40 cycles using gene specific primers which span the full-length genome of SIVsmE041 (Table 2.1). The amplicons from different primer sets were pooled together in equivalent amounts and subjected to next-generation sequencing (NGS) at the sequencing core facility at University of Wisconsin, Madison. All NGS sequencing samples were prepped using the Nextera XT DNA Library preparation kit and the manufacturer's

instructions. The amplicon library was sequenced using Nextera XT kit and MiSeq Illumina desktop sequencer (Invitrogen). The sequence reads were obtained in FASTQ format and analyzed as described below.

2.2.6 Cell culture

Whole blood filter packs were obtained from the Garth Englund Blood Center of Fort Collins, CO. Mononuclear cells were isolated from human peripheral blood by Ficoll-Plaque density centrifugation. PBMCs were grown in RPMI media containing 10% fetal bovine serum (FBS), 2x antibiotic-antimycotic mix (Thermo Fischer Scientific) and 1% L-glutamine. PBMCs were activated for 24 h with PHA at a final concentration of 2 µg per ml and cultured in medium supplemented with 0.25 ng per ml of recombinant human IL-2 (R & D Systems, Inc., Minneapolis, MN). The GHOST (3) cell line positive for CD4, CXCR4 and CCR5 were cultured in DMEM with 10% FBS supplemented with 500 µg/ml G418, 100 µg/ml hygromycin and 1 µg/ml puromycin (Morner et al., 1999).

2.2.7 Viral propagation of SIVsm *in vitro* by co-culturing with activated human PBMC

SIVsm infected mice with the highest plasma viral titer were sacrificed at the end of the study and different tissues such as bone marrow, thymus, spleen, mesenteric lymph nodes and whole blood obtained by cardiac puncture were harvested. Single cell suspensions were made, and leukocyte fraction isolated by Ficoll-Plaque density centrifugation. These cells were counted, plated at a density of 4-5 x 10⁶ cells/ml, and activated using 2 µg/ml PHA and 0.25 µg/ml recombinant human IL-2 for 24 hours. These cells were then co-cultured with 10 x 10⁶ freshly activated human PBMC from whole blood filter packs maintained in complete RPMI supplemented with 0.25 µg/ml recombinant human IL-2. Cell supernatant was harvested every third day and stored for use in future passages. Virus was quantified using qRT-PCR as mentioned

above as well as by GHOST cell titration and flow cytometry. Five hu-HSC mice were then subsequently injected with 200 μ l viral supernatant. This serial passage methodology was repeated for each consecutive generation.

2.2.8 Viral titration using GHOST cells

Ghost cells were seeded in a 48 well plate and infected with cell supernatant harvested from leukocyte/PBMC co-cultures in the presence of 8 μ g/ml of polybrene (Sigma-Aldrich, St. Louis, MO). After 4 h, cells were centrifuged, washed, and reseeded. The infected Ghost cells were

Table 2.1 Primers utilized to generate amplicons for next-generation sequencing.

Primer Location	Sequence	
<i>gag</i>	Forward	5'-CGCTCTGTATTCAGTCGCTCTG-3'
	Reverse	5'-GTCCCTCCTTTCCACAATTCCA-3'
<i>gag/pol</i>	Forward	5'-GCTCAAGGGTCTGGGTATGAATC-3'
	Reverse	5'-TGGAAAAATATGCATCACCTACA-3'
<i>pol</i>	Forward	5'-CCTCCAACCAATCCATATAACACC-3'
	Reverse	5'-GTCTCTGCCTCTGTCTGTCA-3'
<i>pol</i>	Forward	5'-CTCAGTCAAGAACAAGAAGGGTG-3'
	Reverse	5'-GCGATGTGAAGTTGGCACCATTATC-3'
<i>pol/vif/vpx</i>	Forward	5'-GGACTTGGCAAATGGACTGT-3'
	Reverse	5'-CTCGCCTACTGTTTCCTCTC-3'
<i>vif/vpx/vpr/tat/rev</i>	Forward	5'-GACACCAGAGAAAGGATGGCTC-3'
	Reverse	5'-GCTGATTCCCAAGACATCCCAT-3'
<i>vpr/tat/rev/env</i>	Forward	5'-GGGTAGTAGAAGTTCTGGAGGAAG-3'
	Reverse	5'-CTCTGTTTGTCTTTCTTAACCCTG-3'
<i>env</i>	Forward	5'-GGAACAACACAATGCTTGCCAG-3'
	Reverse	5'-CTTCACTTCTCGGATAGCCCCT-3'
<i>env</i>	Forward	5'-CCAGTCACCATTATGTCAGGGTTG-3'
	Reverse	5'-CTGAACATAAACGGGAGGGGAAGA-3'
<i>env/tat/rev/nef</i>	Forward	5'-GGATAGTGCAGCAACAGCAA-3'
	Reverse	5'-TTCCATGCCAGCACCTCTCC-3'

harvested at 48 h post-infection, GFP expression assessed using flow cytometry and the viral titer calculated in order to begin the next generation of infection.

2.2.9 Calculation of variant frequencies in sequencing datasets

To calculate variant frequencies in sequencing datasets, we first removed the adapter sequences and low quality bases using the cutadapt software v1.9.1 (Martin, 2011). We then used cd-hit v4.6 to collapse PCR duplicated sequences (Li and Godzik, 2006). Reads were mapped to the stock virus consensus sequence using the bowtie2 software v2.2.5 (Langmead and Salzberg, 2012). Bowtie2 BAM format output was used as input to lofreq software v2.1.2 to call variants (Wilm et al., 2012). We required >20 total aligned reads and >5% frequency to call a variant.

2.2.10 Assessment of amino acid frequencies in HIV-2 and SIV database sequences

To assess whether amino acids were ‘more HIV-2-like’ or ‘more SIV-like’, we collected multiple sequence alignments for each viral protein from the 2016 HIV Sequence Compendium (Foley, 2016). These alignments consisted of ~68 HIV-2 sequences and ~30 SIVsm sequences. We created separate position probability matrices for the HIV-2 and the SIV sequences in each alignment that tabulated the estimated probability of each amino acid at each alignment position, using the position-based pseudo-count method (Henikoff and Henikoff, 1996). Positions with fewer than 75% gap characters were evaluated.

2.2.11 Determining the d_N/d_S ratio of each gene from passage 5

Single Nucleotide Polymorphism (SNP) reports were generated as described above using lofreq. Using a FASTA file of the stock virus sequence obtained via NGS, a Gene Transfer Format (.gtf) file of the stock sequence with CDS information included as well as the SNP reports in VCF format (Format 2), SNPgenie (available at <https://github.com/hugheslab/snpgenie>) was used to

calculate the d_N and d_S values relative to the stock virus sequence. Additionally, the nonsynonymous (π_N) and synonymous (π_S) nucleotide diversity were estimated by a 9-codon sliding window analysis with SNPgenie according to Nei and Gojobori's (1986) methodology (Nei and Gojobori, 1986).

2.3 Results

2.3.1 SIVsmE041 causes productive infection and chronic viremia in hu-mice

To determine if a primary sooty mangabey SIV isolate can establish infection in hu-HSC mice, mice engrafted with HSC from two independent human donors were injected with SIVsmE041 by intraperitoneal route and serially passaged for a total of five generations (Figure 2.1A and 2.1B). Plasma viral loads were monitored on a weekly basis by qRT-PCR. Viral infection was evident by week 2, with the viral loads increasing three logs by 70 days post-inoculation. The successful infection of hu-mice by SIVsm showed its ability to cause productive infection of human cells *in vivo*. We next proceeded to carry out serial passages of the virus in hu-mice. Viral loads peaked earlier during the 5th passage compared to the 1st passage followed by a gradual decline over the subsequent weeks, suggesting an increased fitness of the virus to infect its target cells with serial passage (Figure 2.2A). No viral loads were detected as expected in control uninfected mice (data not shown).

2.3.2 SIVsmE041 infection leads to CD4⁺ T cell depletion in hu-HSC mice

Previous studies have shown that a central hallmark of HIV infection in humans and hu-HSC mice is the depletion of CD4⁺ T lymphocytes (Aldrovandi et al., 1993; Baenziger et al., 2006; Berges et al., 2006; Denton and Garcia, 2011). Peripheral blood samples collected bimonthly from SIVsm infected hu-HSC mice (Figure 2.1) were assessed for circulating CD4⁺ T cell levels using fluorophore conjugated antibodies for CD45, CD3 and CD4. Baseline CD4⁺ T cell levels prior to

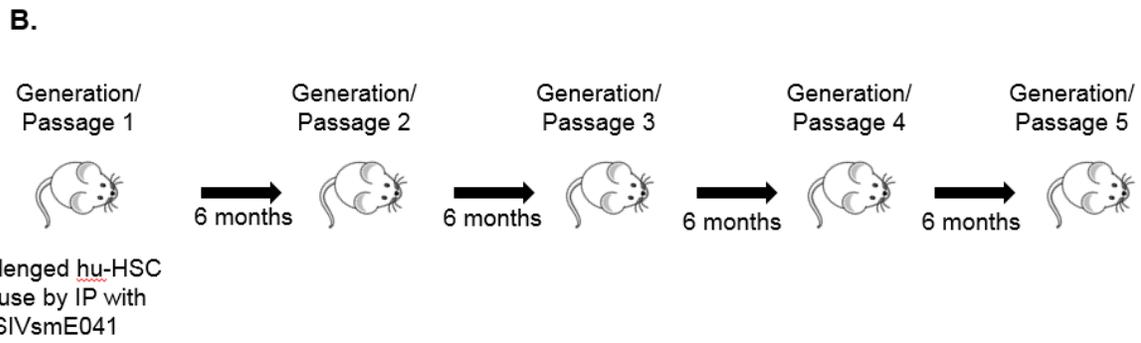
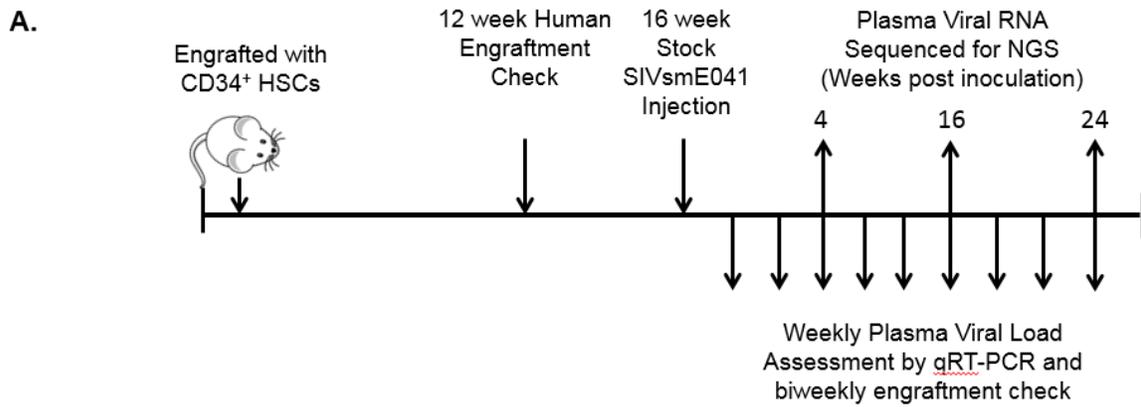


Figure 2.1 SIVsmE041 infection of humanized mice. Schematic representation of the (A) SIVsm infection and (B) serial passage in hu-HSC mice.

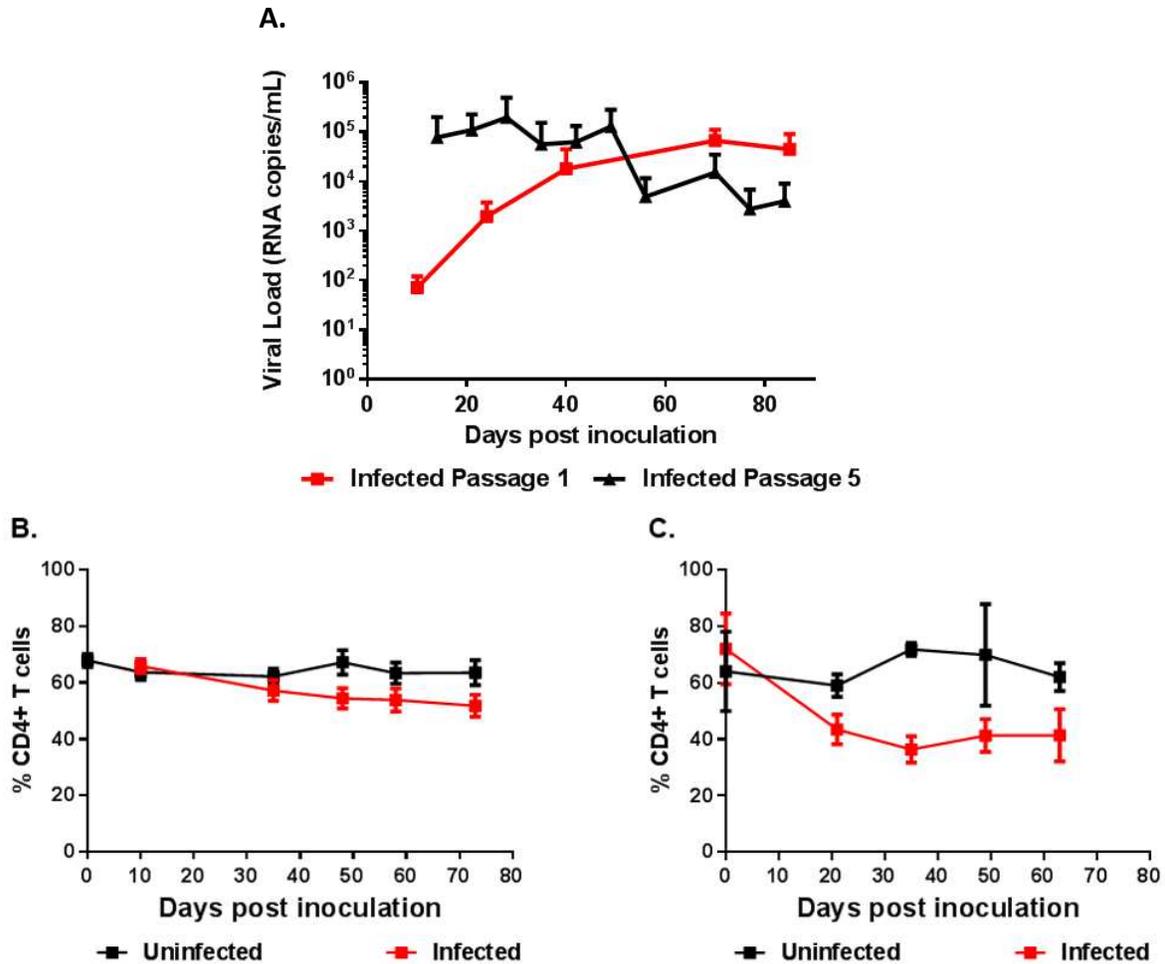


Figure 2.2 SIVsmE041 infection leads to chronic viremia and a gradual decline of CD4⁺ T cells in humanized mice. (A) First and fifth generation plasma viral loads. Hu-HSC mice were infected via i/p route and viral loads were monitored by qRT-PCR on a weekly basis. No viral loads were detected in the uninfected control mice (data not shown). The percentage of circulating CD4⁺ T cells relative to CD3⁺ cells in the first (B) and fifth (C) generation consisting of SIVsm infected and uninfected hu-HSC mice. Statistically significant depletion was seen for both the first and fifth generation in infected mice relative to the uninfected mice (two-tailed Student's t-test, $p < 0.05$).

infection were measured to be greater than 70% of all CD3⁺ cells in each mouse. Mice from the first viral passage showed a gradual decline of circulating CD4⁺ T cell levels, whereas the uninfected control mice did not show any marked CD4⁺ T-cell depletion (Figure 2.2B). In SIVsm infected mice, CD4⁺ T-cell decline started within 10 days post-infection. In comparison, the rate of CD4⁺ T cell depletion in passage 5 mice was more pronounced than in the earlier passages, suggesting increased pathogenicity of the virus to these cells (Figure 2.2C). This is likely caused by the higher plasma viral loads observed in passage 5 and is similar to the pathology observed in end stage human AIDS (Figure 2.2A). This data showed that SIVsm can establish viremia resulting in CD4⁺ T cell depletion in the hu-HSC mouse model.

2.3.3 Genetic evolution of SIVsmE041 during sequential passages in hu-mice

Viral adaptation in response to host innate and/or adaptive immune pressures may select for certain genomic changes in the transmitted SIVsmE041 founder virus in humanized mice. Viral fitness can be impacted by a number of viral attributes such as the affinity for CD4 receptor binding or efficient cellular membrane fusion. Initial infection can be due to the existence of a low frequency variant in the Env region of the stock virus that eventually becomes selected as the dominant viral phenotype over time. Completely novel mutations that are not initially represented in the viral swarm can arise due to adaptive pressure to counteract the effects of host restriction factors like tetherin or APOBEC3 in the *nef*, *env* or *vif* regions. To assess genomic changes throughout the experiment and to search for possible signatures of viral adaptation, next-generation sequencing (NGS) was performed on the SIVsmE041 stock virus used to infect hu-HSC mice, as well as plasma RNA isolated from the infected animals from all 5 generations. Samples were collected at various time points ranging from 4 weeks to 6 months post-infection. We determined the consensus sequences of the stock virus and passaged viruses from each time point. The

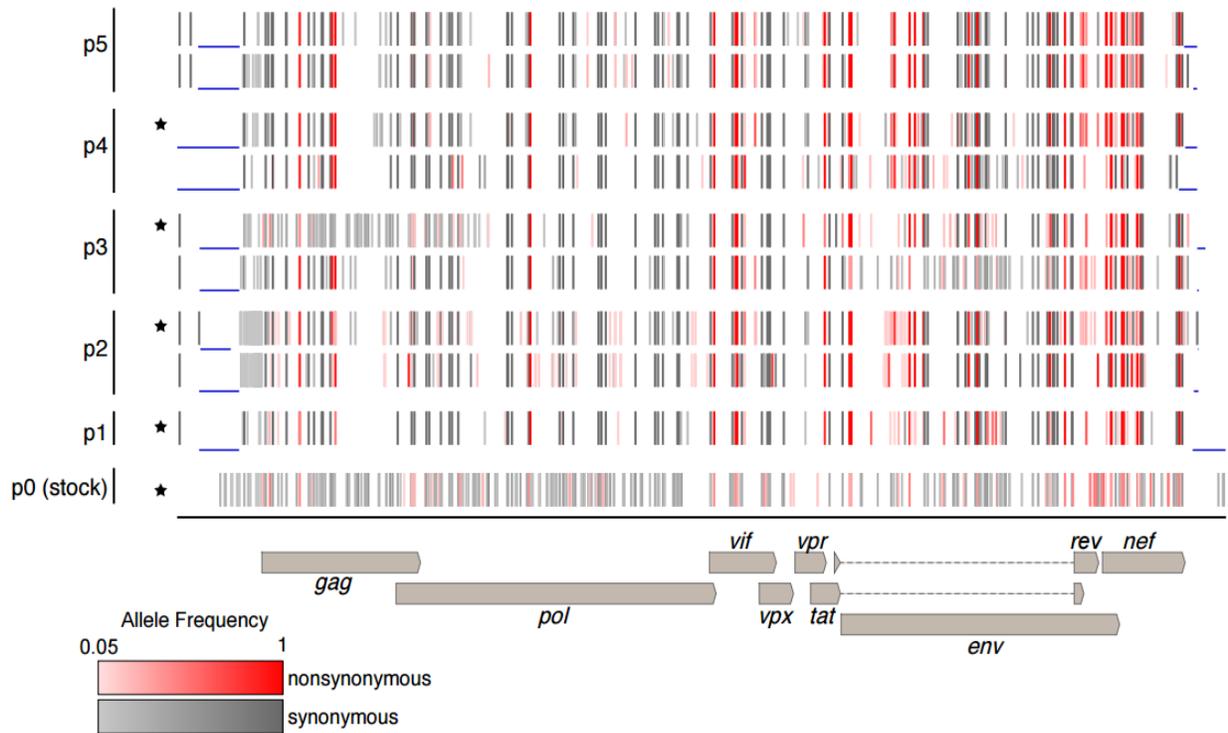


Figure 2.3 Global variant analysis in virus populations during sequential passage. Frequencies of synonymous (grey) and non-synonymous (red) variants in the representative SIV populations were determined from sequencing data (see 2.2 Materials and Methods). Populations from two mice per passage are shown. Regions with lower than 20x sequencing coverage are indicated with blue lines, as well as the positions of the virus coding regions. Mice used as a source for inoculum for the subsequent passage are indicated with a (*).

consensus sequence of the stock virus shared between 98.6 – 98.9% pairwise nucleotide identity in the coding region with the viruses from passage 1 – passage 5, which shared >99.5% pairwise identity amongst themselves, indicating that the largest bottleneck may have occurred upon initial infection of human cells in the first passage (Figure 2.3, Supplementary Table 2.1). We also calculated the frequency of all variants relative to the input virus with a frequency above 5% and determined whether variants were synonymous or nonsynonymous. The stock virus (p0) contained a number of variants with frequencies between 5% and 50%, as would be expected from a primary virus isolate (Figure 2.3, Supplementary Table 2.1). Later generations contained viral variants that ranged from 5% to 100% in allelic frequency. We searched for potentially adaptive variants according to several criteria. We reasoned that adaptive variants would be those that: (1) arose in the first passage and persisted at a high allelic frequency across subsequent passages, and/or (2) appeared in the first passage in multiple mice and increased in allelic frequency across the passages (Table 2.2, Supplementary Table 2.1; Figure 2.4). Variants that increased in allelic frequency across multiple passages were found throughout the genome, with very few found in the *pol*, *vif*, *vpr* or *rev* genes. Conversely, *gag*, *env* and *nef* were found to have a higher frequency of mutation with 3, 6 and 2 mutations, respectively (Figure 2.4).

We also used the sequence alignments in the HIV Sequence Compendium (Foley, 2016) to assess whether the non-synonymous variants produce a more “HIV-2-like” virus. Specifically, we were interested in determining whether the variant changed the encoded amino acid to one present in a higher fraction of HIV-2 sequences than in SIVsm sequences. One mutation in particular, Env T829A, appeared at a relatively low rate in the first passage, but subsequently became the dominant variant as passages were carried forward. Additionally, while some SIV

strains, such as SIVmac239, also contain the variant, it is frequently found amongst HIV-2 subtypes A, B and H (Table 2.2).

Patterns of variation were assessed for evidence of purifying (negative) or diversifying (positive) selection using SNPgenie (Nelson et al., 2015). The mean synonymous (d_S) and nonsynonymous (d_N) divergence from the reference sequence of the pooled NGS data was determined on a gene by gene basis (Figure 2.5). Each gene showed varying degrees of divergence, with *nef* and *vif* showing the highest degree of nonsynonymous differences. In contrast, *vpx*, *gag* and *pol* all showed a markedly lower degree of nonsynonymous divergence relative to synonymous divergence, which indicates that these genes are undergoing a greater degree of purifying selection.

2.4 Discussion

Compelling evidence suggests that HIV-1 and HIV-2 arose through cross-species transmission events from their respective ancestral SIVs that are native to chimpanzees/gorillas and sooty mangabeys. However, the viral genetic evolution that occurs and is reflective of host adaptive changes is still not well understood. In this study, we used humanized mice to model SIVsm infection in human cells using a physiological setting *in vivo* and performed serial passages of the virus to allow human adaptive changes, assessing both the phenotypic and genetic requirement for potential evolution into HIV-2. We used a primary isolate of SIVsmE041 stock virus, isolated from a 21-year-old captive sooty mangabey. Whether or not this isolate is able to cause AIDS in an accelerated manner in a naïve young sooty mangabey remains to be determined.

This virus was propagated in sooty mangabey PBMCs in order to infect hu-HSC mice and thus comprises a viral swarm. During our initial rounds of infection, we were able to successfully

Table 2.2 Amino Acid substitutions resulting from candidate adaptive mutations.

Protein	Position	Stock consensus ^a	P5 variant ^b	Stock freq ^c	P1 freq ^c	P5 freq. rep. 1/2 ^d	SIV fraction ^e	HIV-2 fraction. ^f
Gag	223	P	Q	0 ^g	0	1/1	0.00	0.00
Gag	233	R	T	0	0.26	1/1	0.02	0.04
Gag	482	R	K	0	0	0.56/0.58	0.84	0.65
Vif	136	K	R	0	0	0.57/0.68	0.29	0.07
Vpr	99	G	S	0	0	0.83/0.88	0.33	0.12
Env	158	N	D	0	0.06	1/1	0.03	0.07
Env	222	R	Q	0	0.13	1/1	0.87	0.16
Env	249	N	D	0	0.18	1/1	0.01	0.03
Env	393	V	I	0	0.28	1/1	0.01	0.04
Rev	34	R	K	0	0	0.9/0.76	0.00	0.01
Env	827	E	G	0	0	0.7/0.79	0.03	0.04
Env	829	T	A	0	0.15	1/0.98	0.35	0.86
Nef	92	D	N	0	0.09	1/0.99	0.02	0.03
Nef	94	D	N	0	0	0.59/0.44	0.03	0.12

^aConsensus amino acid from stock virus sequencing dataset

^bVariant amino acid in passage 5 (P5) viruses

^cFrequency of variant mutation in stock sequencing dataset

^dFrequency of variant mutation in passage 5 (P5) replicate 1 and 2 sequencing datasets

^eFraction of SIV sequences from HIV sequence compendium that contain the variant amino acid

^fFraction of HIV-2 sequences from HIV sequence compendium that contain the variant amino acid

^g0 indicates below the limit of detection (5%) of our variant identification pipeline

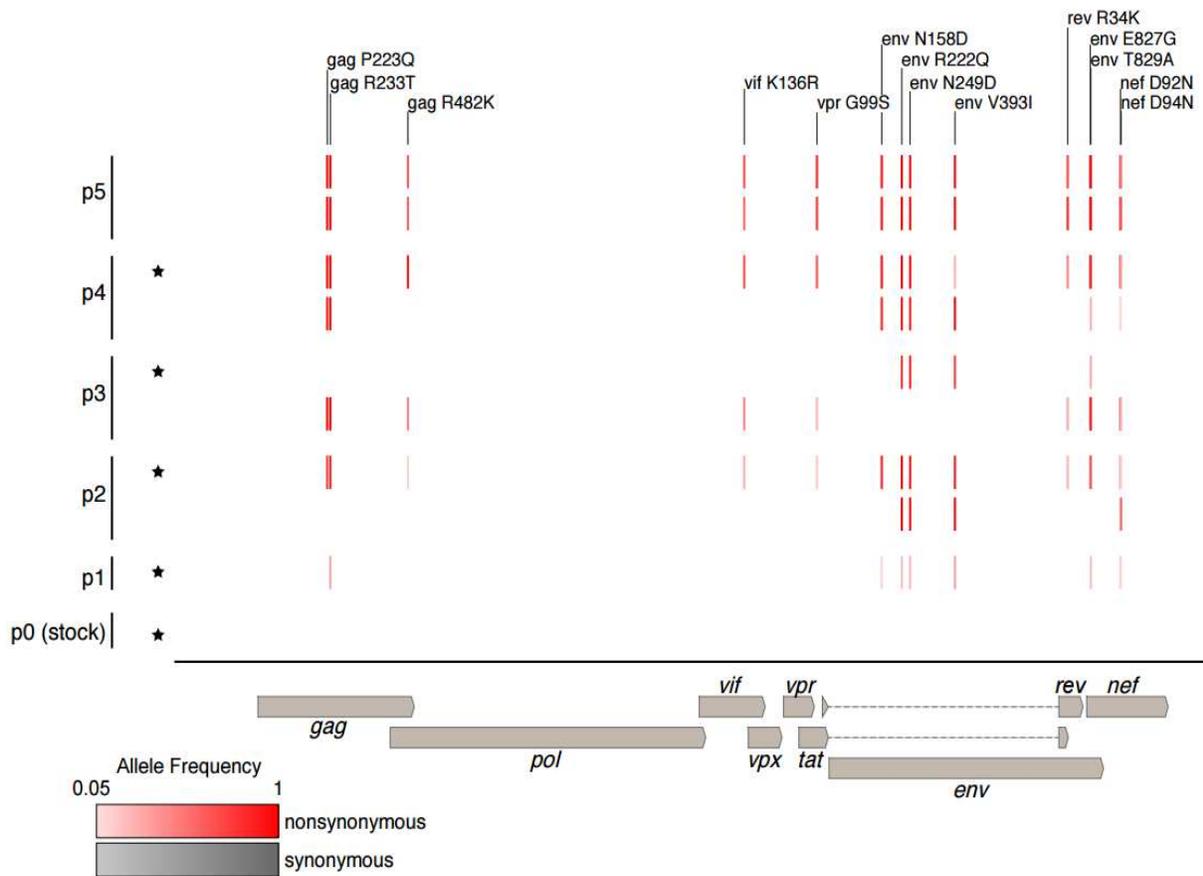


Figure 2.4 Viral variants that increased in frequency between generations 1 and 5. Frequencies of nonsynonymous (red) and synonymous (grey) variants whose frequency increased by more than 0.5 between passage 1 and 5 in representative SIV populations were determined from sequencing data (see 2.2 Materials and Methods). Mice used as the source of inoculum for the subsequent passage are indicated with a (*).

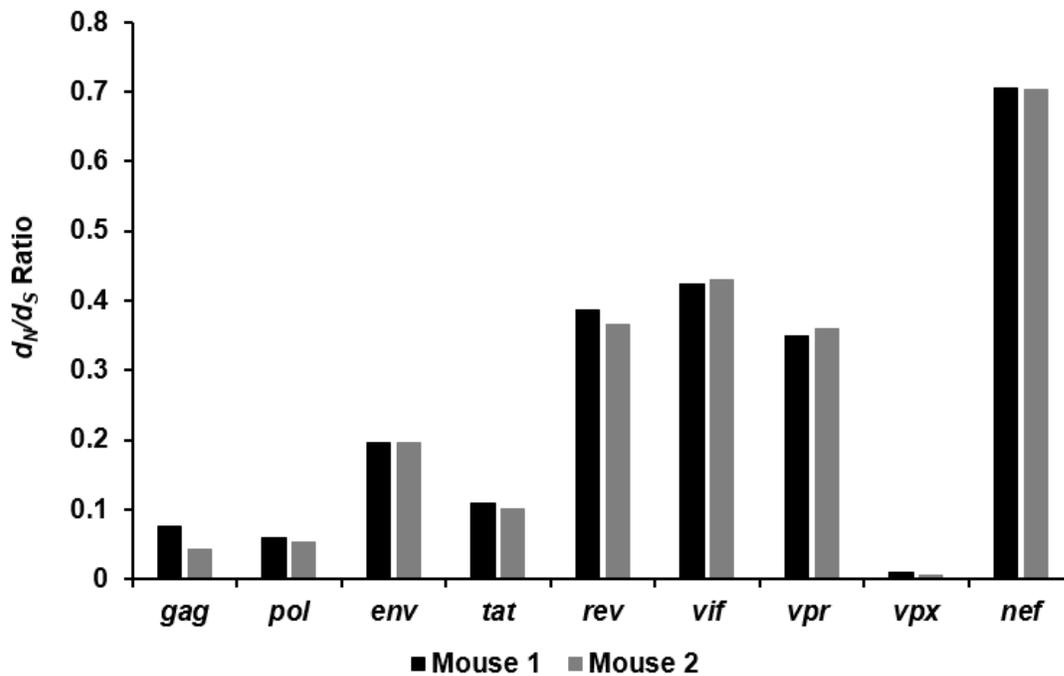


Figure 2.5 The d_N/d_S ratio of each gene from generation 5 calculated against the SIVsm reference sequence. The d_N value was calculated via SNPgenie as the mean number of nonsynonymous differences from the reference per nonsynonymous site in the input sequence with the d_S value being the mean number of synonymous differences from the reference per synonymous site in the input sequence. The *nef* and *vif* genes shows the greatest d_N/d_S ratios in the genome, while *vpx*, *gag*, and *pol* all have drastically reduced ratios, suggesting that more divergent evolution is occurring in *nef* and *vif*.

infect hu-HSC mice, which displayed viremia over the course of several months and thus confirmed that infection of human cells with a primary isolate of SIV_{sm} is fully feasible and that persistent viremia could be sustained.

These infected mice also displayed CD4⁺ T cell decline, recapitulating a key feature of human HIV infection. Comparison of viral growth and CD4⁺ T cell loss kinetics between the first and fifth viral passages revealed that infection onset occurs at a faster rate, viral loads are higher and CD4⁺ T cell loss is relatively more rapid during the fifth passage. This signifies enhanced viral adaptation and pathogenicity upon serial passage. This is consistent with the hypothesis that SIV_{sm} has evolved gradually during sequential infections to become more pathogenic to humans (Marx et al., 2001). The above data collectively indicate successful evolution of SIV_{sm} in becoming more fit to replicate *in vivo* in human cells and reflects the genetic changes in the parent virus during sequential passage. Accordingly, we analyzed the viral sequences from the first to the fifth passage virus by NGS and evaluated the sequence changes between these passages that encompass both synonymous and nonsynonymous mutations.

Among the important genetic changes that are predicted include the ability of the virus to gain entry into human cells efficiently and overcome host restriction factors. A number of host restriction factors, such as TRIM5 α , APOBEC3G and tetherin, pose a potential species-specific barrier for viruses to mount a successful infection. Of these restriction factors, tetherin, a transmembrane host protein, appears to have the most significant impact during the evolution of HIV-1 and HIV-2 progenitors (Compton and Emerman, 2013; Simon et al., 2015). Most SIVs use the Nef protein to counteract tetherin by targeting its cytoplasmic domain, while HIV-1 uses the Vpu protein in a similar manner (Sharp and Hahn, 2011; Simon et al., 2015). Nef has a range of roles in HIV and SIV infection that include: downregulation of surface CD4 and MHC-1

expression, regulation of apoptosis, counteracting tetherin antagonism, altering the cellular state of activation and enhancing virion infectivity (Garcia and Miller, 1991; Anderson et al., 1994; Miller et al., 1994; Finkel et al., 1995; Sandstrom et al., 1996; Schwartz et al., 1996; Hua et al., 1997; Greenberg et al., 1998; Arora et al., 2000; Simmons et al., 2001; Campbell et al., 2004; Wei et al., 2005; Atkins et al., 2008; Neil et al., 2008; Noviello et al., 2008; Wonderlich et al., 2008). Other SIVs and HIV-2 utilize their envelope proteins to target tetherin extracellular or cytoplasmic domains (Le Tortorec and Neil, 2009). In a *nef* deleted SIV mutant, compensatory changes were acquired by gp41 in order to counteract tetherin antagonism (Serra-Moreno et al., 2011). These varied anti-tetherin activities by lentiviruses appear to have evolved independently against host-specific selective pressures (Strebel, 2013).

We observed several nonsynonymous mutations in Nef and Env that became fixed in the population as seen in multiple mice across each generation (Figure 2.4; Supplementary Table 2.1). Consistently, we detected nonsynonymous mutations by passage 5 in Nef that corresponded to amino acid substitutions at positions R10H, K11R and H12R (Supplementary Table 2.1). These mutations occur in a region of *nef* that overlaps with *env* in a different reading frame. Additionally, these mutations are synonymous in the *env* reading frame and only appear to directly affect *nef*. However, the presence of these specific changes still warrants further investigation.

Homodimerization of Nef is important to create the PAK1/2 binding site that plays crucial role in exerting the anti-apoptotic activity of Nef by the phosphorylation of Bad (Wolf et al., 2001). In HIV-1, this dimerization is stabilized by a salt-bridge formed between amino acid residues D123 and the dibasic R105-R106 motif (Ye et al., 2004; Poe and Smithgall, 2009; Kwak et al., 2010). One of the substitutions that became fixed by passage 5 in our study was the substitution of a lysine

at position 105 into an arginine (K105R) which underscores the possible importance of this position in the host cell adaptation of SIV (Supplementary Table 2.1).

In SIVsm, Nef is used to counteract the action of tetherin, due to the absence of Vpu. However, HIV-2 alternatively utilizes Env to antagonize the human form of tetherin that contains a deleted cytoplasmic tail (Neil et al., 2008; Le Tortorec and Neil, 2009). Fixed mutations identified in the gp41 region of Env, E827G and T829A, may contribute to important adaptive changes (Figure 2.4). Additionally, the Gag M30K/R substitution found in a previous study was observed in two mice from the second passage at very low allelic frequencies and was subsequently lost in future passages (data not shown) (Ye et al., 2004; Poe and Smithgall, 2009). This suggests that the Gag M30K/R substitution may not be critical for species-specific adaptation. However, more functional studies are needed to assess the importance of these mutations.

Further analysis of the genome as a whole, as well as on a gene by gene basis suggests that while certain regions are becoming more adapted, other components of the genome are much less tolerant of change. Using SNPgenie to assess the NGS data, π_N represents the mean number of pairwise nonsynonymous differences per nonsynonymous site and π_S is the same variable calculated for synonymous changes. These variables were able to be elucidated across the genome in a 9-codon sliding window analysis. The disparity between π_N and π_S values indicate how strong purifying selection is acting on a certain part of the genome (Nei and Gojobori, 1986). There were noticeable π_S peaks that were present in each gene, which suggests that those regions may be subject to purifying selection and are less likely to tolerate amino acid substitutions (Figure 2.6). However, in several locations on the *vif* and *nef* genes, this trend is reversed and the π_N peaks are much higher than π_S at the same genomic locations. This could reflect positive (diversifying) selection and is consistent with the higher d_N/d_S ratio associated with these genes (Figure 2.6). The

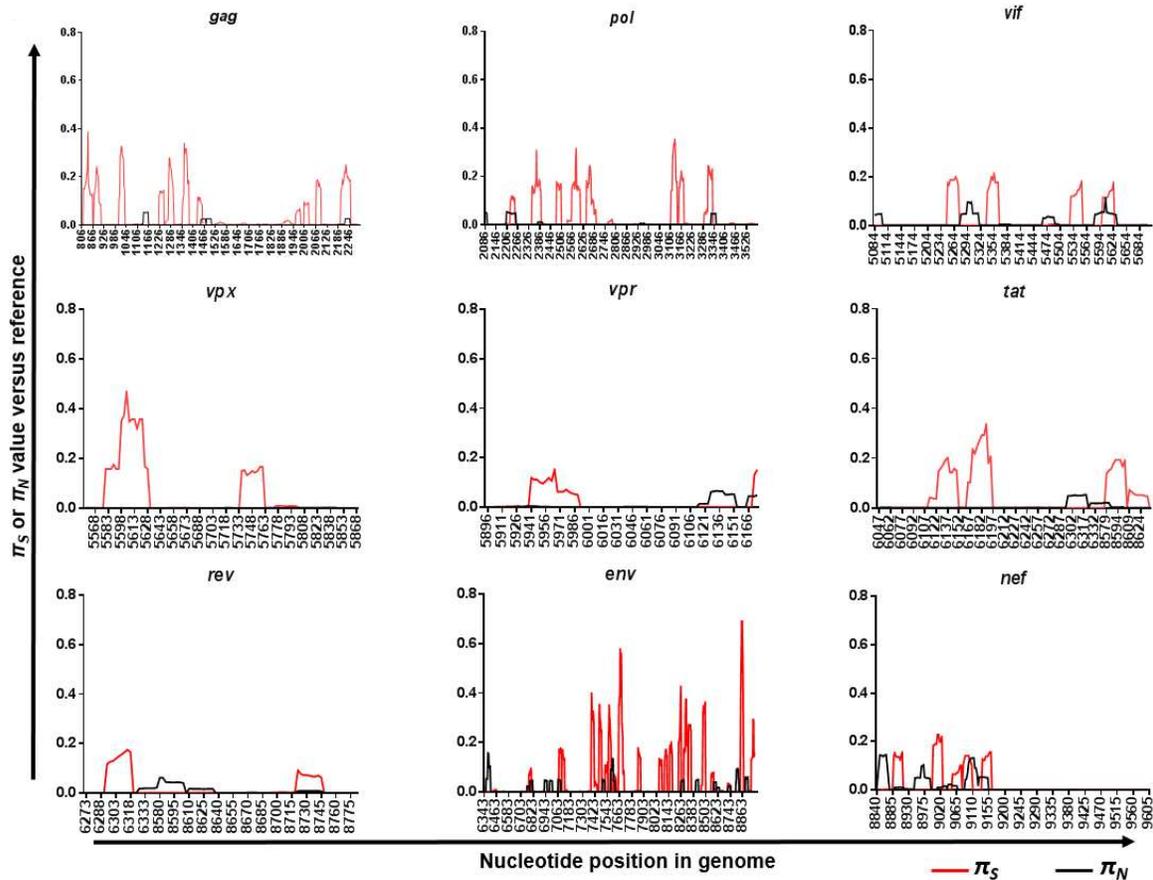


Figure 2.6 9-codon Sliding Window Analysis of Passage 5 Viral Genes. π_N (black) and π_S (red) of the passage 5 mice were calculated for each gene using a 9-codon sliding window analysis by SNPgenie. Multiple π_S peaks were identified in each gene, with relatively low respective π_N values, which indicate regions that may be crucial to the fitness of SIVsm. Alternatively, in other locations across multiple genes, particularly *vif* and *nef*, there are regions with markedly higher π_N values relative to π_S , which is indicative of diversifying selection.

π_N peaks do not cover the full length of each genome but are instead representative of small 9 codon sliding windows of analysis. It is reasonable to conclude that while the genes as a whole favor purifying selection to prevent deleterious changes, individual regions within the genes may be more tolerant of variation and indeed may be the key drivers of viral adaptation into new hosts (Nelson and Hughes, 2015). These results provide a rational basis for follow-up studies to assess regions that control cross-species fitness.

A recent report by Yuan et al. utilized BLT hu-mice to assess the infection with various chimpanzee SIVcpz strains to detect adaptive changes towards HIV-1. Interestingly, only two non-synonymous mutations in *env* were found in both in SIVcpzMB897 and SIVcpzBF1167 (Yuan et al., 2016). This is likely due to the short 14-week period of evaluation during a single passage and that the whole genome was not analyzed by next-generation sequencing. Also, the chimpanzee strains used were grown from molecular clones representing homogeneous virus populations. In contrast, our study on the evolution of HIV-2 employed a primary SIVsm isolate, was serially passaged 5 times in hu-mice and the whole viral genomes were assessed for mutations on a global scale to detect notable viral genetic changes that are selected to become fixed during long term viral adaptation.

In summary, our results showed that SIVsm can infect hu-HSC mice causing chronic viremia leading to CD4⁺ T cell decline. These properties are augmented between passage one and five indicating sequential viral adaptation. Sequence changes noted during viral adaptation may be indicative of increased fitness to the human host by passage 5; however, functional studies of these mutations are needed. While we recognize that the progeny virus from passage 5 has not fully evolved into HIV-2, it is likely that further passages are necessary, and experiments are currently underway to evaluate this question. Nevertheless, our data demonstrates that SIVsm evolution in

a human host can be modeled in hu-HSC mice and that key adaptation mutations will be identified and further characterized.

CHAPTER 3: SIV PROGENITOR EVOLUTION TOWARD HIV: A HUMANIZED MOUSE SURROGATE MODEL FOR SIV_{SM} ADAPTATION TOWARD HIV-2²

3.1 Summary

How SIV progenitors evolved into deadly HIV-1 and HIV-2 following initial cross-species transmission still remains a mystery. Here we used humanized mice as a human surrogate system to evaluate SIV_{sm} evolution into HIV-2. Increased viral virulence to human CD4⁺ T cells and adaptive genetic changes were observed during serial passages.

3.2 Materials and Methods

3.2.1 Generation of humanized mice

Humanized mice were generated as previously described (Figure 3.1A) (Akkinä et al., 1994; Bai et al., 2000; Berges et al., 2008; Veselinovic et al., 2016; Schmitt et al., 2017). Mice were maintained at the Colorado State University Painter Animal Center. The studies conducted in this publication have been reviewed and approved by the CSU Institutional Animal Center and Use Committee.

3.2.2 SIV_{sm}E041 Infection of humanized mice

The SIV_{sm}E041 (GenBank accession HM059825.1) used in this study was obtained from a 21 year old sooty mangabey euthanized for displaying clinical signs of AIDS and then propagated in sooty mangabey PBMC creating a true primary isolate comprised of a population of closely

² Curlin, J., Schmitt, K., Kumar, D.M., Remling-Mulder, L., Feely, S., Stenglein, M., O'Connor, S., Marx, P., Akkinä, R. (2018). SIV progenitor evolution toward HIV: A humanized mouse surrogate model for SIV_{sm} adaptation toward HIV-2. *J Med Primatol* 47(5), 298-301. doi: 10.1111/jmp.12380.

related viruses (Fultz et al., 1986; Murphey-Corb et al., 1986; Gormus et al., 1995a; Gormus et al., 1995b; Ling et al., 2004).

3.2.3 Plasma viral load and CD4⁺ T cell level determination

Five hu-HSC mice were injected intraperitoneally (i/p) with 200 μ L of SIVsmE041 (TCID₅₀ 811). Peripheral blood was collected by tail vein puncture weekly for viral loads and bimonthly for engraftment (Figure 1B). Viral RNA was extracted using the E.Z.N.A. Viral RNA kit (Omega bio-tek) and viral loads determined using the iScript One-Step RT-PCR kit with SYBR Green and the manufacturer's instructions (Bio Rad) as described previously.¹³ Whole blood was stained using fluorophores conjugated to mouse anti-human CD45-FitC (eBioscience), CD3-PE (eBioscience) and CD4-PE/Cy5 (BD Pharmingen) and assessed using the BD Accuri C65 FACS Analyzer as described previously (Schmitt et al., 2017). CD4⁺ T cell levels were calculated within the CD45⁺CD3⁺ double positive population and decline was assessed using a two-tailed Student's *t*-test ($p < 0.05$).

3.2.4 Preparation of SIVsmE041 for sequential passaging

At the end of each generation (Figure 3.1C), SIVsm-infected mice with the highest titer were euthanized and various tissues such as the spleen, bone marrow, lymph nodes and whole blood were harvested and used to propagate virus as previously described (Schmitt et al., 2017). For each subsequent generation, five hu-HSC mice were injected with 200 μ L of viral supernatant.

3.3 Results

Following the initial challenge with SIVsmE041, hu-mice showed detectable viremia within 2 weeks. The viremia levels gradually increased by three logs peaking at 6.69×10^4 copies/ml at 10 weeks post-inoculation (Figure 3.2A). Virus from this first passage was then

subsequently passaged six more times in hu-HSC mice to mimic the evolution of SIV_{sm} into HIV-2. By the seventh generation, viral loads were found to be 2.5-logs higher at 2 weeks post-inoculation compared to that of initial first passage indicating an increase in viral fitness (Figure 3.2B) to human cell infection.

One of the hallmarks of HIV infection is the depletion of CD4⁺ T cells. There was no significant human CD4⁺ T cell decline during the initial few weeks of infection with SIV_{sm}E041. However, a gradual decline is noted during subsequent weeks relative to the uninfected controls (Figure 3.2A). By comparison, during the seventh passage, SIV_{sm} displayed a more rapid, statistically significant ($p < 0.05$) CD4⁺ T cell decline by day 10 (Figure 3.2B), suggesting an increased pathogenicity of the serially passaged seventh-generation viral strain.

3.4 Discussion

Many theories exist regarding the origin and transmission of HIV pathogens in the human population. Compelling genomic evidence suggests that HIV-2 arose through cross-species transmission events from sooty mangabeys (Marx et al., 2001). Viral adaptive changes needed for this successful cross-species evolution are not well understood due to the lack of an ideal *in vivo* system for experimental evaluation. In this study, we used a hu-HSC mouse model to test this hypothesis. Our results showed that SIV_{sm} can readily infect hu-mice and give rise to chronic viremia consistent with HIV-2 infection in humans. Increased viral fitness to human cells as evidenced by the higher viral loads observed by the seventh generation resulting in a more rapid CD4⁺ T cell decline suggests selection and evolution of the progenitor SIV_{sm} *in vivo*. This is consistent with the hypothesis that SIV_{sm} evolved gradually during sequential infections in the human host to become more pathogenic (Marx et al., 2001). At the genomic level, we previously identified 14 nonsynonymous substitutions in *gag*, *vif*, *vpr*, *rev*, *env* and *nef* that became fixed in

the human adapted viral population. This was observed in multiple mice across five passages/generations (Schmitt et al., 2017). These substitutions are likely indicative of increased viral fitness to the human host. Studies are currently underway to elucidate if these nonsynonymous substitutions indeed determine cross-species transmissibility.

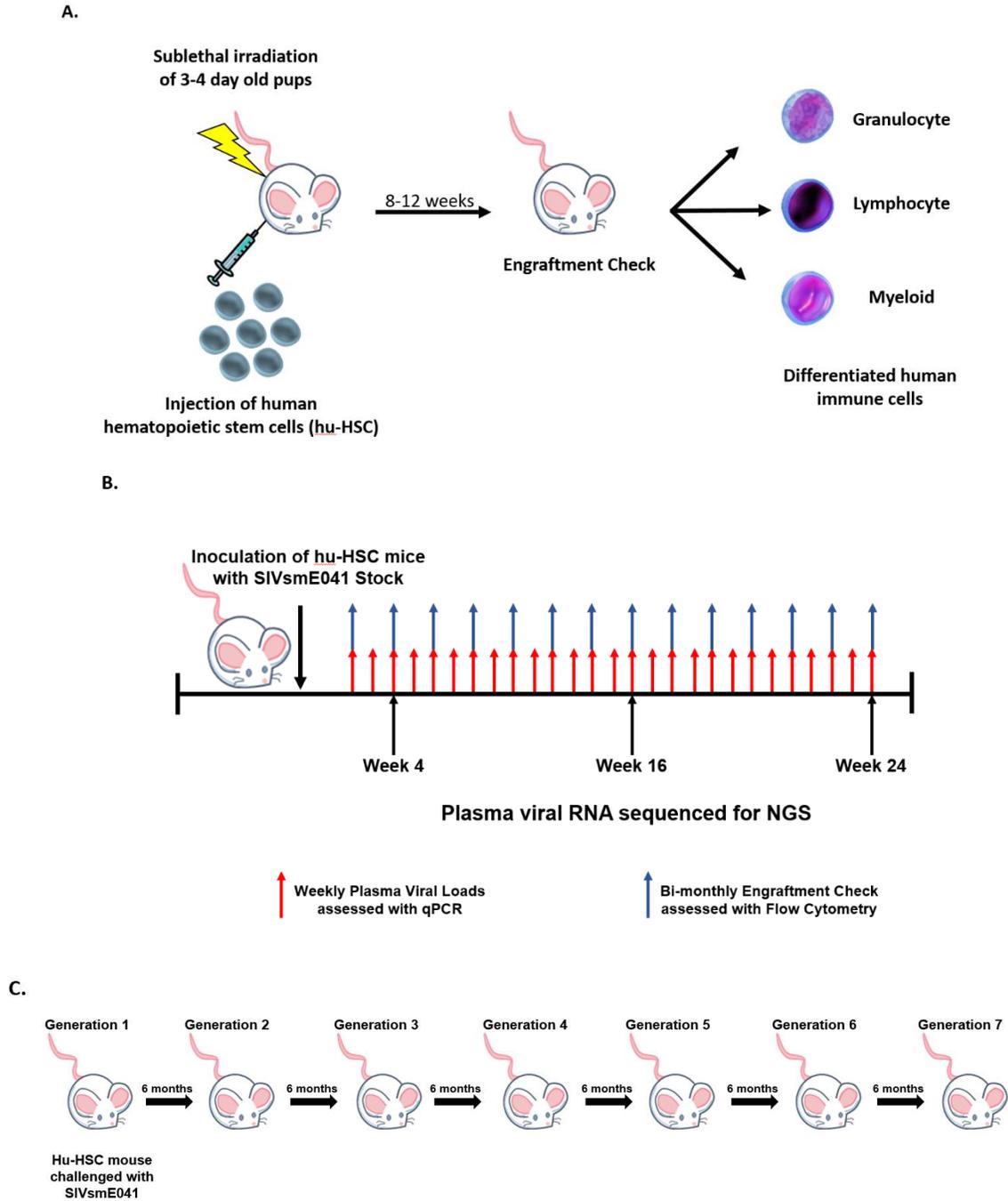


Figure 3.1 Hu-HSC mice generation and experimental scheme with SIVsmE041 progenitor virus. (A) Neonatal mice were sublethally irradiated and injected intrahepatically with human CD34⁺ hematopoietic stem cells. Mice are screened 8-12 weeks post-reconstitution for human immune cell engraftment. (B) Schematic representation of SIVsmE041 infection. (C) Serial passaging methodology in hu-HSC mice.

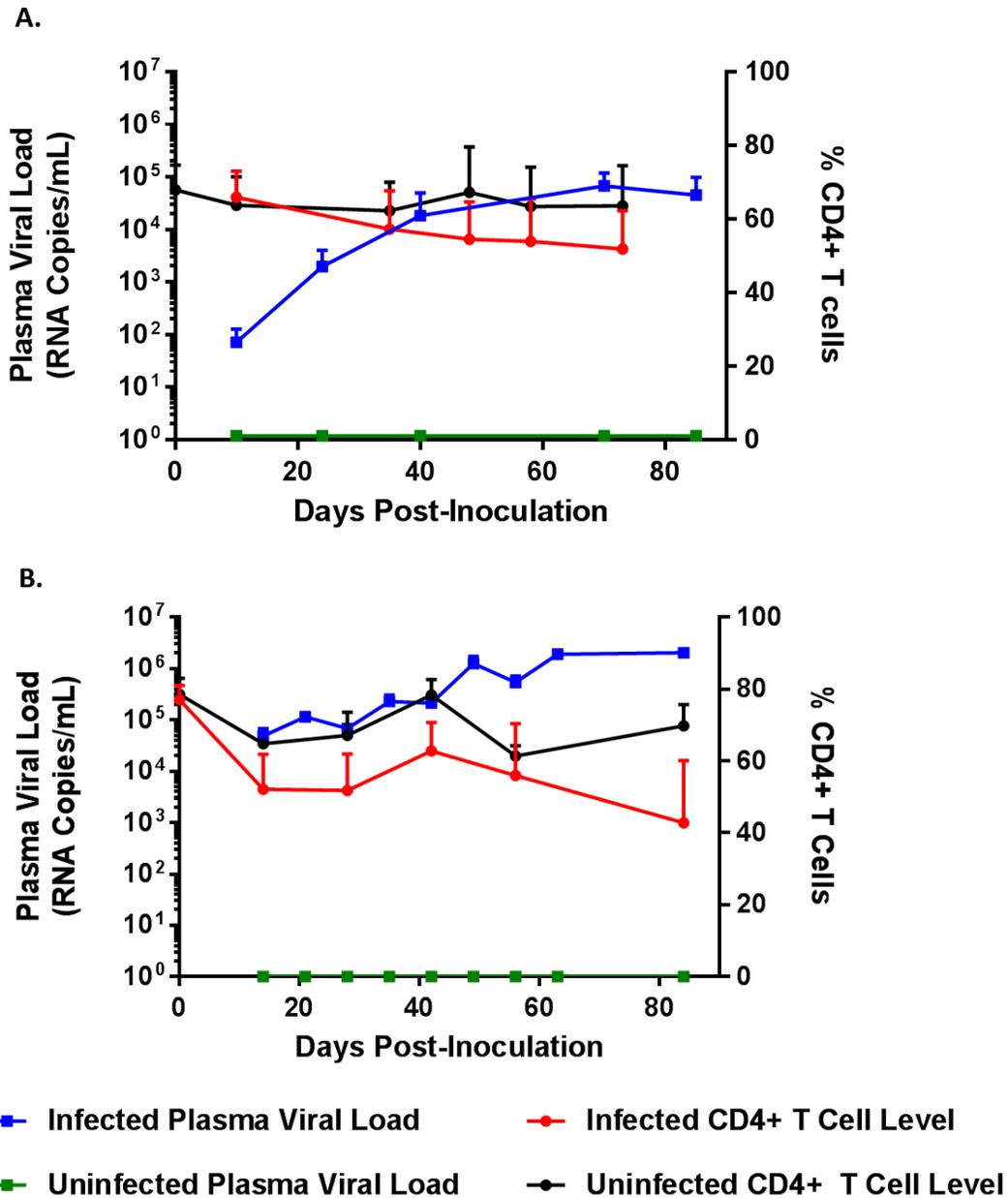


Figure 3.2 Kinetics of SIVsmE041 infection and CD4⁺ T cell decline during *in vivo* passages in hu-HSC mice. (A) First and (B) seventh generation plasma viral loads and CD4⁺ T-cell decline seen in SIVsmE041 infected and uninfected hu-HSC mice. Statistically significant CD4⁺ T-cell depletion was seen in both the first and seventh generation hu infected mice relative to the uninfected mice (two-tailed Student's *t*-test, $p < 0.05$).

CHAPTER 4: EVOLUTION OF SIV_{SM} IN HUMANIZED MICE TOWARDS HIV-2³

4.1 Summary

Through the accumulation of adaptive mutations, HIV-2 originated from SIV_{sm}. To identify these evolutionary changes, a humanized mouse model recapitulated the process that likely enabled this cross-species transmission event. Various adaptive mutations arose, as well as increased virulence and CD4⁺ T-cell decline as the virus was passaged in humanized mice.

4.2 Material and Methods

4.2.1 Hu-HSC mouse preparation and Ethics

Humanized hu-HSC mice were prepared as previously described (Akkina et al., 1994; Bai et al., 2000; Berges et al., 2008; Garcia and Freitas, 2012; Shultz et al., 2012). Experimental animals were maintained at the Colorado State University Painter Animal Center. All studies were approved and reviewed by the CSU Institutional Animal Center and Use Committee.

4.2.2 SIV_{sm}E041 infection of hu-HSC mice and serial passaging

SIV_{sm}E041 (GenBank accession HM059825.1), isolated from sooty mangabey PBMC, was used to infect hu-HSC mice as previously described (Schmitt et al., 2017; Schmitt et al., 2018). Approximately 24 weeks after inoculation, mice with the highest plasma viral loads were

³ Curlin, J., Schmitt, K., Remling-Mulder, L., Moriarty, R., Goff, K., O'Connor, S., Stenglein, M., Marx, P., Akkina, R. (2020) Evolution of SIV_{sm} in humanized mice towards HIV-2. *J Med Primatol.* 00:1–4. <https://doi.org/10.1111/jmp.12486>

euthanized to propagate the virus as previously described (Schmitt et al., 2017; Schmitt et al., 2018; Curlin et al., 2020b). This process was repeated for 8 sequential passages.

4.2.3 CD4⁺ T-cell decline and plasma viral load detection

Peripheral blood was collected weekly by tail vein puncture and viral RNA was isolated from plasma using the E.Z.N.A Viral RNA kit (Omega bio-tek, Norcross, CA). RNA was quantified using qRT-PCR and SYBR Green with the iScript One-Step RT-PCR kit (BioRad) according to the manufacturer's instructions to determine plasma viral loads. Bimonthly, whole blood collected from the mice was stained with mouse anti-human CD45-APC (eBioscience), CD3-FITC (eBioscience), and CD4-PE (BD Pharmingen, San Jose, CA) antibodies. CD4⁺ T-cell levels were assessed as a percentage of CD45⁺/CD3⁺ cells with the BD Accuri C6 cytometer as described previously (Schmitt et al., 2017; Schmitt et al., 2018; Curlin et al., 2020b). Comparison of CD4⁺ T-cell decline between control and infected mice was assessed with a two-tailed Student's *t*-test ($p < 0.001$).

4.2.4 SNP detection and Illumina-based deep sequencing

Viral RNA from two infected mice at weeks 3, 11, 19, and 23 post-infection in the eighth passage were used to generate amplicons with primers designed using Primal Scheme software.(Quick et al., 2017) Amplicons were prepared using the TruSeq Nano DNA HT Library Preparation Kit and the MiSeq Illumina desktop sequencer (Invitrogen). Nonsynonymous SNPs were identified by aligning reads to the SIVsm stock virus consensus sequence using bowtie2 software v2.2.5, and then calling variants at >100 coverage depth and >1% frequency using lofreq software v2.1.2.(Langmead and Salzberg, 2012; Wilm et al., 2012) Genome plots were generated

using R and ggplot2 (ISBN: 0387981403). R scripts can be found at https://github.com/stenglein-lab/viral_variant_explorer.

4.3 Results

Eight serial passages in hu-mice resulted in a human adapted SIVsmE041 capable of rapid infection within one week of inoculation that was sustained for over 100 days and maintained above 10^5 RNA copies/ml (Figure 4.1A). CD4⁺ T-cell levels showed significant ($p < 0.001$) decline that commenced around 60 days post-inoculation and persisted through the remainder of the passage (Figure 4.1B). This is a markedly drastic decline compared to earlier passages with SIVsm in humanized mice. Numerous nonsynonymous single nucleotide polymorphisms (SNPs) were identified using Illumina-based deep sequencing throughout the viral genome at high frequencies (Figure 4.2).

4.4 Discussion

Serial passaging in humanized mice provided an effective model for recapitulating the genetic changes that allowed SIVsm to adapt to human immune cells similarly to HIV-2. After eight sequential passages, SIVsm is consistently capable of readily infecting humanized mice lasting for the duration of the passage. Additionally, the CD4⁺ T-cell decline displayed after eight passages is drastically more pronounced than that seen in earlier passages of SIVsmE041 (Schmitt et al., 2017; Schmitt et al., 2018). This indicates that the virus has continued to develop a greater affinity for CD4⁺ T-cells than in earlier passages.

Many previously characterized SNPs could still be seen even after the eighth passage, suggesting that these mutations became fixed (Schmitt et al., 2017; Schmitt et al., 2018). Furthermore, several additional substitutions such as Pol I172M and Nef F65I that were only present at less than 30% frequency early in the passage and/or previously unidentified began to

increase dramatically to over 70% frequency. Further studies are needed to ascertain their function and determine why these mutations with potential advantage have begun to arise during later passages of SIVsmE041.

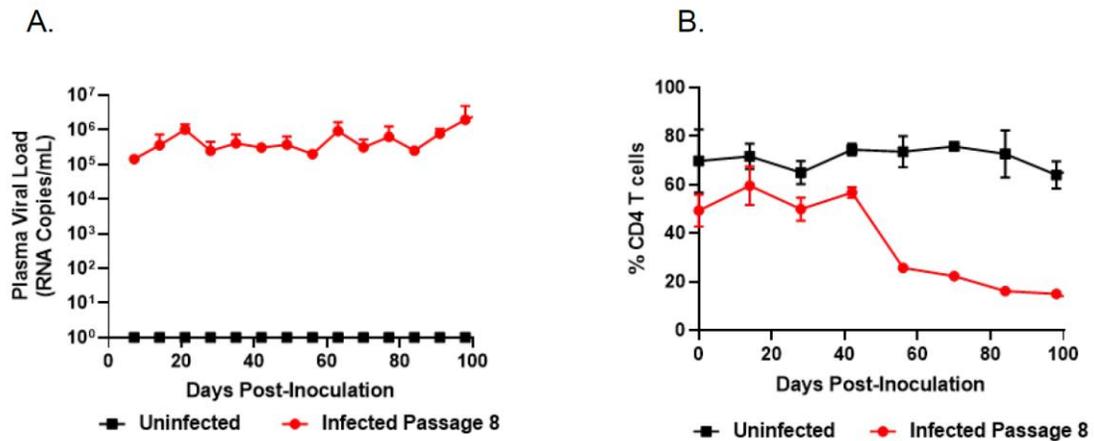


Figure 4.1 Plasma Viral Loads and CD4⁺ T-cell decline following eight serial passages of SIVsmE041 in hu-HSC mice. (A) Plasma viral loads collected from the eighth passage of SIVsmE041 infected mice (B) CD4⁺ T-cell depletion following SIVsmE041 infection. CD4⁺ T cell depletion was rapid and significant within 60 days post-inoculation (two-tailed Student's t-test, $p < 0.001$).

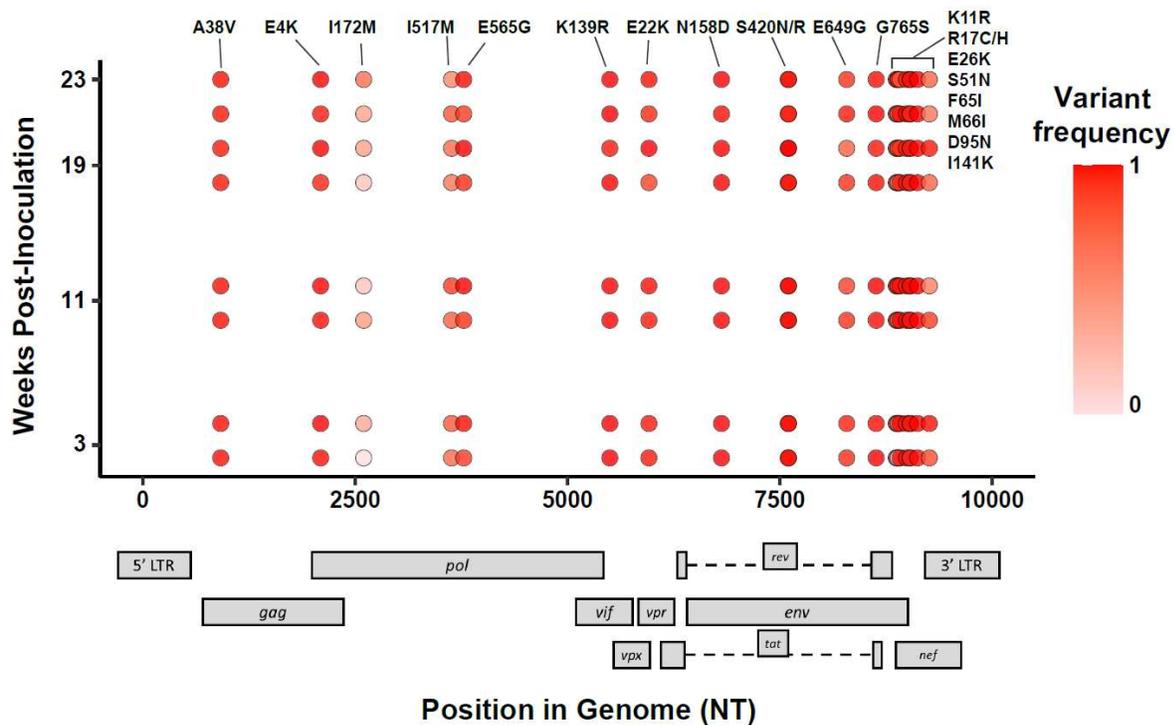


Figure 4.2 SNP frequencies identified after eight serial passages in hu-HSC mice that reached >50% of the viral population at the last timepoint. Lofreq v2.1.2 was used to identify SNP frequencies following read alignment by bowtie2 v2.2.5. Mutations were present throughout the genome, with the highest concentration in *env* and *nef*.

CHAPTER 5: SIV_{CPZ} CROSS-SPECIES TRANSMISSION AND VIRAL EVOLUTION TOWARD HIV-1 IN A HUMANIZED MOUSE MODEL⁴

5.1 Summary

HIV-1 evolved from its progenitor SIV strains, but details are lacking on its adaptation to the human host. We followed the evolution of SIVcpz in humanized mice to mimic cross-species transmission. Increasing viral loads, CD4⁺ T-cell decline, and non-synonymous mutations were seen in the entire genome reflecting viral adaptation.

5.2 Materials and Methods

5.2.1 Preparation of humanized mice

Hu-HSC mice were prepared as described previously (Figure 5.1A) (Akkina et al., 1994; Bai et al., 2000; Berges et al., 2008; Veselinovic et al., 2016; Schmitt et al., 2017). All animals were cared for according to the CSU IACUC approved protocols.

5.2.2 EK505 viral propagation and *in vivo* infection

HEK 293T cells were transfected with the SIVcpzEK505 plasmid, kindly provided by Dr. Preston Marx. After 48 hours, the supernatant was ultracentrifuged using a 20% sucrose cushion and 200 μ L (TCID₅₀ 3.2 x 10⁵) of the concentrated virus was injected intraperitoneally into 5 well-engrafted (>60% CD45⁺, >50% CD4⁺) hu-HSC mice (Figure 5.1A). Viral Titer was determined using TZM-bl reporter cells as previously described (Platt et al., 1998; Derdeyn et al., 2000; Wei et al., 2002; Takeuchi et al., 2008; Platt et al., 2009).

⁴ Curlin, J., Schmitt, K., Remling-Mulder, L., Moriarty, R., Stenglein, M., O'Connor, S., Marx, P., Akkina, R. (2020) SIVcpz cross-species transmission and viral evolution toward HIV-1 in a humanized mouse model. *J Med Primatol.* 49:40–43. <https://doi.org/10.1111/jmp.12440>

5.2.3 Plasma viral load and CD4⁺ T-cell level evaluation

To determine the PVLs and CD4⁺ T-cell engraftment, peripheral blood was obtained on a weekly and bimonthly basis respectively, by tail vein puncture (Figure 5.1). The E.Z.N.A Viral RNA kit (Omega bio-tek) was used to extract viral RNA from the plasma, which was quantified as previously described using the iScript One-Step RT-PCR kit with SYBR Green per the manufacturer's guidelines (Bio Rad) (Schmitt et al., 2017). Mouse anti-human CD45-FITC (eBioscience), CD3-PE (eBioscience), and CD4-PE/Cy5 (BD Pharmingen, San Jose, CA) antibodies were used to stain whole blood samples. CD4⁺ T-cell levels were determined within the CD45⁺CD3⁺ population using the BD Accuri C6 cytometer as described previously (Schmitt et al., 2017). CD4⁺ T-cell decline between the infected and uninfected mice was assessed using a two-tailed Student's *t*-test ($p < 0.05$).

5.2.4 NGS and genomic analysis

Amplicons for sequencing were generated using RNA samples from timepoints 3-, 11-, 19- and 22-weeks post-inoculation. Multiplexed whole-genome spanning primer pools were designed using the Primal Scheme software (Quick et al., 2017). Samples were prepared using the Nextera XT DNA Library preparation kit and sequenced at the University of Wisconsin, Madison sequencing core facility using a MiSeq Illumina desktop sequencer (Invitrogen).

Reads were mapped to the SIVcpzEK505 stock virus (GenBank accession Number: DQ373065.1) using bowtie2 software v2.2.5 (Langmead and Salzberg, 2012). These alignments were combined with lofreq software v2.1.2 to call variants (Wilm et al., 2012). Each variant required >20 reads depth of coverage and >20% frequency increase relative to the stock virus.

5.3 Results

Successful infection resulting in viremia was seen in hu-mice challenged with a number of SIVcpz isolates, with the SIVcpzEK505 results presented here. Virus was detected in plasma within one week of inoculation. Viral loads increased gradually by approximately 2 orders of magnitude, peaking around 80 days post-challenge before declining slightly (Figure 5.1B). There was slow CD4⁺ T-cell decline from the starting point in the infected mice versus those of uninfected controls (Figure 5.1C). Viral genomic RNA obtained from plasma samples from viremic mice at 3, 11, 19 and 22 weeks was subjected to next-generation sequencing (NGS). Numerous synonymous and non-synonymous single-nucleotide changes (SNPs) were seen throughout the genome. Non-synonymous SNPs that increased in frequency within the viral population by at least 20% are presented in Table 1.

5.4 Discussion

Here we utilized the humanized mouse model to mimic the initial cross-species transmission of SIVcpz and understand the viral phenotypic and genetic adaptation in an *in vivo* context. SIVcpz readily infected hu-mice and was viremic in a week confirming successful transmission to humans. Gradual viral load increases suggest improved viral fitness in human cells. CD4⁺ T-cell decline indicates viral predilection to these cells comparable to observed immunosuppression in humans. In addition to other shortcomings, previous studies did not provide enough time to adequately determine the variants needed for cross-species adaptation (Yuan et al., 2016; Sato et al., 2018; Yuan et al., 2018). We analyzed the viral variants at different time points spanning 22 weeks. Multiple SNPs were identified throughout the viral genome in contrast to only a few variants identified in the envelope gene previously. While many mutations were seen at individual early time points, some disappeared subsequently, others persisted across multiple time

points, with some even increasing in frequency (Table 5.1). SNPs displaying a 20% increase in occurrence are presented. These encompass *gag*, *pol*, *vif*, *env*, and *nef* genes suggesting SIV adaptation to humans requires many allelic changes. We are currently investigating the genomes of later passages of the virus in an effort to identify the major changes that confer improved fitness to overcome human restriction factors.

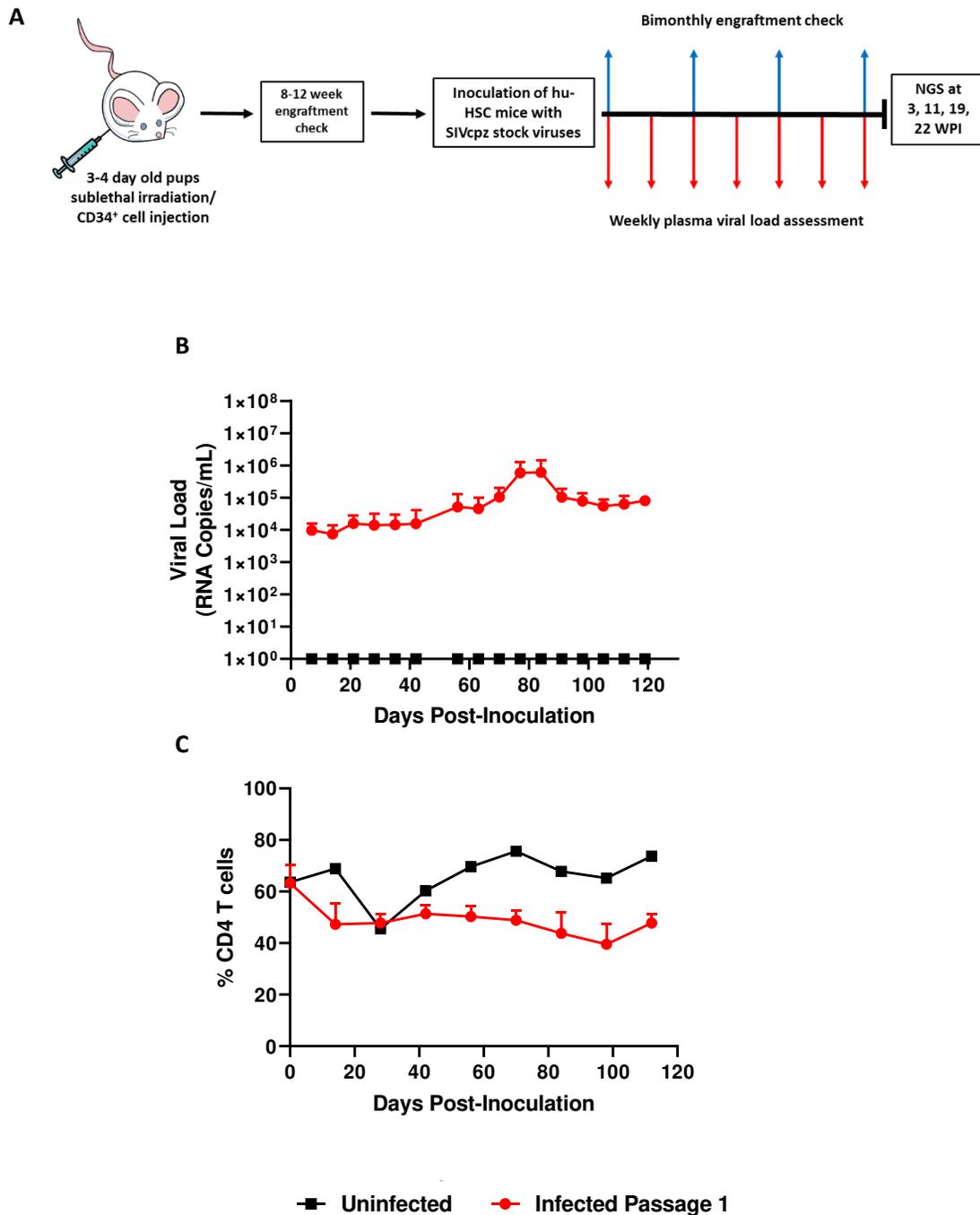


Figure 5.1 Schematic representation of the hu-HSC mouse generation, viral infection, SIVcpzEK505 plasma viral loads and CD4⁺ T-cell decline following inoculation. (A) Neonatal mouse pups were intrahepatically injected with human CD34⁺ hematopoietic stem cells following sublethal irradiation. Mice were screened 8-12 weeks post-reconstitution to determine engraftment levels of human immune cells. (B) Plasma viral loads and (C) CD4⁺ T-cell depletion of SIVcpzEK505-infected and uninfected hu-HSC mice. Statistically significant CD4⁺ T-cell depletion was seen in the infected mice relative to the uninfected mice (two-tailed Student's *t*-test, $p < 0.05$).

Table 5.1 Nonsynonymous single-nucleotide polymorphisms (SNPs) that showcase a >20% increase in viral population frequency by 22 weeks post-inoculation.

Gene	Codon Number^a	AA change^b	Highest Endpoint Frequency^c
Gag	35	V→I	0.88
Gag	105	R→K	0.23
Pol	175	P→Q	0.34
Pol	364	E→K	0.22
Vif	13	V→A	0.41
Vif	55	I→T	0.61
Vif	118	S→A	0.92
Env	402	V→G	0.56
Env	403	G→E	0.99
Env	414	P→T	0.6
Env	601	Q→K	0.77
Env	630	K→R	0.74
Env	668	R→K	0.3
Env	703	E→K	0.6
Env	773	I→K	0.58
Env	829	G→S	0.2
Nef	79	M→L	0.35

^aCodon number based on the beginning of the respective gene

^bThe amino acid change resulting from the nonsynonymous mutation

^cThe highest detected frequency of the change between samples at the last time point collected

CHAPTER 6: MIMICKING SIV CHIMPANZEE VIRAL EVOLUTION TOWARD HIV-1 DURING CROSS-SPECIES TRANSMISSION⁵

6.1 Summary

HIV-1 evolved from SIV during cross-species transmission events, though viral genetic changes are not well understood. Here, we studied the evolution of SIVcpzLB715 into HIV-1 Group M using humanized mice. High viral loads, rapid CD4⁺ T-cell decline, and non-synonymous substitutions were identified throughout the viral genome suggesting viral adaptation.

6.2 Materials and Methods

6.2.1 Generation of humanized mice and ethics

Humanized mice were generated according to methods previously described (Figure 6.1A) (Akkina et al., 1994; Bai et al., 2000; Berges et al., 2008; Garcia and Freitas, 2012; Shultz et al., 2012). All mice were maintained at the Colorado State University Painter Animal Center. The studies conducted here have all been reviewed and approved by the CSU Institutional Animal Care and Use Committee.

6.2.2 LB715 viral propagation, *in vivo* infection and serial passage

The SIVcpzLB715 plasmid was obtained from Dr. Preston Marx and used in cell transfections to generate stock virus as described previously (Curlin et al., 2020b). For the initial infection, 200 μ l (TCID₅₀ 2.0x10⁵) of the concentrated virus was injected intraperitoneally into

⁵ Curlin, J., Schmitt, K., Remling-Mulder, L., Moriarty, R., Goff, K., O'Connor, S., Stenglein, M., Marx, P., Akkina, R. Mimicking SIV chimpanzee viral evolution toward HIV-1 during cross-species transmission. *J Med Primatol.* 2020;00:1–4. <https://doi.org/10.1111/jmp.12485>

five (>75% CD45⁺, >60% CD4⁺) hu-HSC mice. SIVcpzLB715-infected mice with the highest viral titer after 6 months were euthanized, and infected tissues were cultured to harvest the first passage stock virus as described previously (Schmitt et al., 2017; Schmitt et al., 2018). For the next generation, a new cohort of five hu-HSC mice was injected with 200 µl of first passage virus.

6.2.3 Determination of PVL and CD4⁺ T-cell decline

To evaluate PVL and CD4⁺ T-cell decline, peripheral blood was obtained on a weekly and bimonthly basis, respectively. Plasma RNA was extracted utilizing the E.Z.N.A. Viral RNA kit (Omega bio-tek, Norcross, CA) and the manufacturer's instructions. Viral load was quantified using the iScript One-Step RT-PCR kit with SYBR Green (BioRad) per the manufacturer's instructions as described previously (Schmitt et al., 2017; Curlin et al., 2020b). CD4⁺ T-cell decline was determined by staining whole blood with fluorophore-conjugated anti-human CD45-FitC (eBioscience), CD3-PE (eBioscience), and CD4-PE/Cy5 (BD Pharmingen, San Jose, CA) antibodies. BD Accuri C6 FACS Analyzer was used to determine cell counts as described previously (Schmitt et al., 2017; Schmitt et al., 2018). The CD4⁺ T-cell decline between the infected and uninfected mice was assessed using a two-tailed Student's *t*-test ($p < 0.001$).

6.2.4 Illumina-based deep sequencing and analysis

Viral RNA samples from timepoints 3, 11, 19, and 24 weeks post-inoculation were used to generate amplicons for sequencing. Whole-genome primer pools were made with Primal Scheme (Quick et al., 2017). Amplicons were prepared using the TruSeq Nano DNA HT Library Preparation kit and sequenced utilizing the MiSeq Illumina sequencer (Invitrogen, Carlsbad, CA). Reads were mapped to the SIVcpzLB715 stock virus (GenBank accession number: KP861923.1) using bowtie2 software v2.2.5 (Langmead and Salzberg, 2012). This output was used as input for

lofreq software v2.1.2 to call variants (Wilm et al., 2012). Each variant required >20 reads depth of coverage, >20% last timepoint frequency, a change of at least 30% frequency over time and were present in at least 5 datasets.

6.3 Results

Virus from the first hu-mice passage was used to infect a second cohort of five hu-mice to evaluate the potential evolution of SIVcpzLB715 into HIV-1 Group M (Figure 6.1A). Viral loads showed an approximate 4.5-log increase throughout the duration of infection (Figure 6.1B). CD4⁺ T-cell depletion occurred within 14 days and significantly declined ($p < 0.001$) throughout the duration of the second-generation infection (Figure 6.1B). These data show SIVcpzLB715 can establish chronic viremia upon serial passage that produces significant CD4⁺ T-cell decline by the end of the second generation.

Viral RNA was extracted from plasma of viremic mice at weeks 3, 11, 19 and 24 and subjected to deep sequencing. Detected amino acid substitutions were scored using a BLOSUM62 matrix to determine the likelihood of any given residue substitution based on known protein alignments (Figure 6.2). Mutations in *gag*, *pol*, *nef* and *env* were identified as more favorable for viral protein structure.

6.4 Discussion

We assessed the phenotypic and genetic changes that facilitated the evolution of SIVcpzLB715 toward HIV-1 Group M utilizing a humanized mouse model to recapitulate cross-species transmission. Hu-mice became infected with SIVcpzLB715 within 2 weeks after initial viral challenge. Enhanced CD4⁺ T-cell loss was observed during the second serial passage suggesting increased pathogenicity and viral fitness. Additionally, one of the mutations in the Env

gene (E523Q) located in the CD4-binding site appears to have a favorable BLOSUM62 score (Figure 6.2) (Chen et al., 2005; Yu et al., 2010). This implies that the CD4-binding site is adapting to improve virus-receptor binding. With increased knowledge of how SIVcpzLB715 first adapted to the human host, we may be able to gain better insight on how SIVs cross over into the human population.

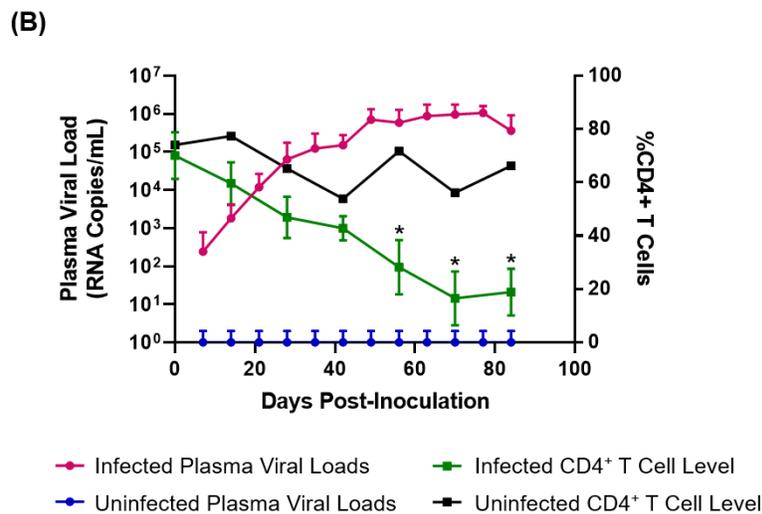
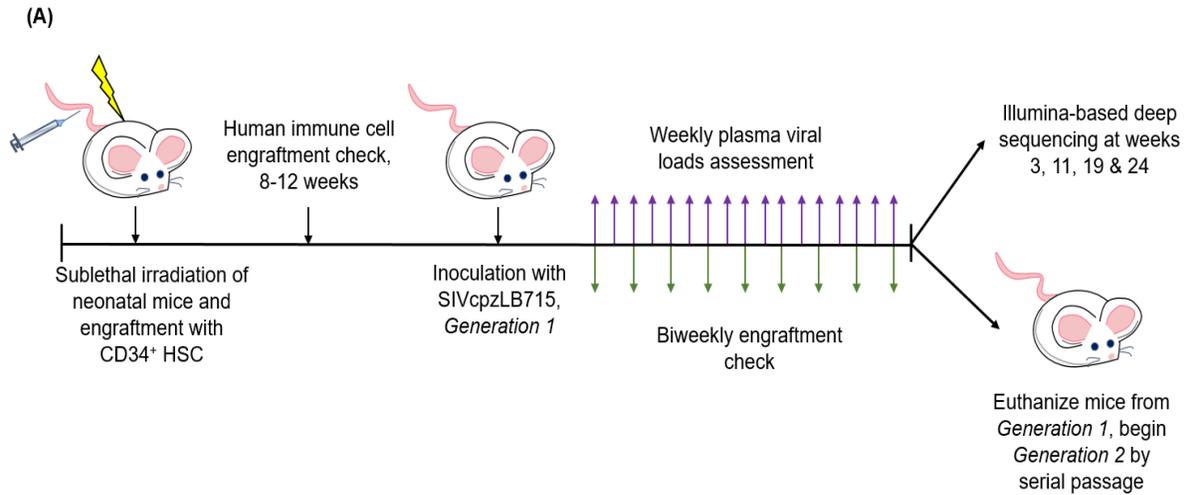


Figure 6.1 SIVcpzLB715 infection kinetics during the second generation in hu-HSC mice. (A) Experimental scheme for SIVcpzLB715 infection of hu-HSC mice. Neonatal mice are sublethally irradiated and inoculated intrahepatically with CD34⁺ hematopoietic stem cells. Following SIVcpzLB715 inoculation, PVL were assessed weekly and CD4⁺ T-cell decline bimonthly. (B) Second generation PVL and CD4⁺ T-cell decline seen in SIVcpzLB715 infected hu-HSC mice. Statistically significant CD4⁺ T-cell depletion occurred by the end of the second generation in infected mice relative to the uninfected mice (two-tailed Student's *t*-test, **p* < 0.001).

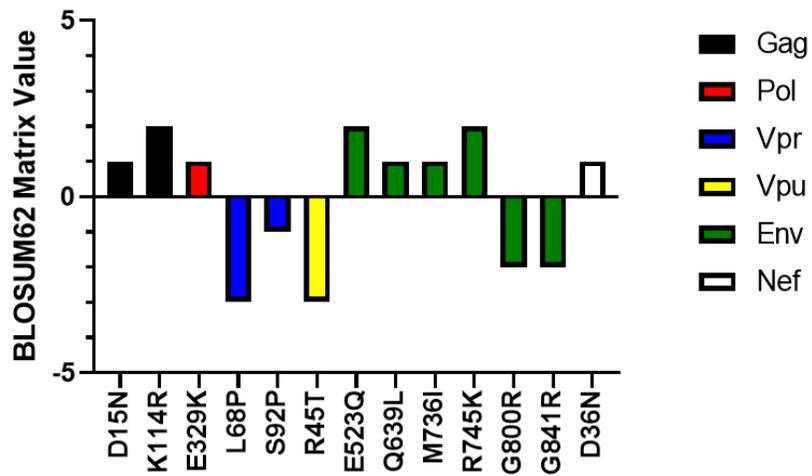


Figure 6.2 BLOSUM62 matrix scores of identified residue changes. Variant residues were identified as described in the results and methods. Some variants in Vpr, Vpu and Env appear to be disfavored, yet still became more frequent in the viral population.

CHAPTER 7: CROSS-SPECIES TRANSMISSION AND EVOLUTION OF SIV CHIMPANZEE PROGENITOR VIRUSES TOWARDS HIV-1 IN HUMANIZED MICE⁶

7.1 Summary

The genetic evolution of HIV-1 from its progenitor virus SIV following cross-species transmission is not well understood. Here we simulated the SIVcpz initial transmission to humans using humanized mice and followed the viral evolution during serial passages lasting more than a year. All three SIVcpz progenitor viruses used, namely LB715 and MB897 (group M) as well as EK505 (group N) readily infected hu-mice resulting in chronic viremia. Viral loads increased progressively to higher set-points and the CD4⁺ T cell decline became more pronounced by the end of the second serial passage indicating viral adaptation and increased pathogenicity. Viral genomes sequenced at different time points revealed many nonsynonymous variants not previously reported that occurred throughout the viral genome, including the gag, pol, env and nef genes. These results shed light on the potential changes that the SIVcpz genome had undergone during the initial stages of human infection and subsequent spread.

7.2 Materials and Methods

7.2.1 Preparation of the SIVcpz viral stocks

In order to generate virus, full-length infectious molecular clones of pSIVcpzMB897, pSIVcpzLB715 and pSIVcpzEK505 were transfected into 293T cells. Briefly, 15 µg of plasmid

⁶ Curlin, J., Schmitt, K., Remling-Mulder, L., Moriarty, R., Goff, K., O'Connor, S., Stenglein, M., Marx, P., Akkina, R. (2020). Cross-species transmission and evolution of SIV chimpanzee progenitor viruses towards HIV-1 in humanized mice. *Frontiers in Microbiology*.

DNA and 30 μ l of TurboFect transfection reagent (Thermo Fisher Scientific, Waltham, MA) was diluted in serum-free DMEM to transfect 293T cells seeded in a T150 flask. At 48 h, virus supernatants were harvested and clarified by low speed centrifugation. To concentrate the viral inoculum, clarified supernatant was ultracentrifuged on a 20% sucrose (w/v) cushion in Ultra-Clear tubes (Beckman Coulter, Pasadena, CA). Ultracentrifugation was conducted utilizing a L8-70M ultracentrifuge and SW 28 Ti rotor (Beckman Coulter, Pasadena, CA) at 27,000 rpm for 2 h at 4°C. Supernatant was decanted and pelleted virus was resuspended gradually on ice for 20 min in serum-free DMEM. Resuspended, cell-free virus was used to directly infect humanized mice, as described below, and an aliquot was used for titration on TZM-bl cells. Briefly, virus was serially diluted from 10^{-1} to 10^{-6} and used to inoculate TZM-bl reporter cells. At 48 h, media was removed, cells were washed with PBS and the monolayer was fixed using 1% formaldehyde-0.2% glutaraldehyde in PBS for 5 min. Cells were washed and incubated in a solution of 4 mM potassium ferrocyanide, 4 mM potassium ferricyanide, 4 mM magnesium chloride and 0.4 mg X-gal per ml, for 2 h at 37°C. The reaction was stopped and the tissue culture infectious dose (TCID₅₀) per ml was calculated to determine virus titer (Derdeyn et al., 2000; Wei et al., 2002).

7.2.2 Generation of humanized mice

Humanized mice were prepared using fetal liver-derived CD34 cells that were isolated, column purified (Miltenyi Biotec, San Diego, CA), cultured and assessed for purity utilizing flow cytometry as previously described (Akkina et al., 1994; Bai et al., 2000; Veselinovic et al., 2016; Schmitt et al., 2017). Preconditioned by irradiation with 350 rads, neonatal Balb/c Rag1^{-/-}γc^{-/-} or Balb/c Rag2^{-/-}γc^{-/-} mice were intrahepatically injected with 0.5-1 x 10⁶ human CD34⁺ cells per mouse (Berges et al., 2008; Veselinovic et al., 2016). Human cell engraftment was then determined by flow cytometry in transplanted mice at 10-12 weeks post-reconstitution by collecting peripheral

blood. The red blood cells were lysed using the Whole Blood Erythrocyte Lysing Kit according to the manufacturer's instructions (R & D Systems, Minneapolis, MN). Fractioned white blood cells were stained with mouse anti-human CD45 FITC (eBioscience), CD3 PE (eBioscience) and CD4 PE/Cy5 (eBioscience) for flow cytometry in order to confirm human cell engraftment (Berges et al., 2006; Veselinovic et al., 2016; Schmitt et al., 2017). All mice were maintained at the Colorado State University Painter Animal Center. The studies conducted in this publication have been reviewed and approved by the CSU Institutional Animal Care and Use Committee.

7.2.3 SIVcpz infection of humanized mice and viral load determination by qRT-PCR

Mice with high (>60% CD45⁺, >50% CD4⁺) human hematopoietic cell engraftment levels were used. At 16 weeks post-engraftment, 200 µl of cell-free SIVcpzMB897 (2 x 10⁵ TCID₅₀), SIVcpzEK505 (3.2 x 10⁵ TCID₅₀) and SIVcpzLB715 (2 x 10⁵ TCID₅₀) were used to inject five hu-HSC mice intraperitoneally (i/p). To assess plasma viral loads, peripheral blood was collected weekly by tail vein puncture using non-heparinized capillary tubes and transferred immediately to EDTA-containing vacutainer tubes (BD Biosciences, San Jose, CA). The peripheral blood was mixed with PBS for a final volume of 150 µl and centrifuged for 5 min at 400 x *g*. Plasma was collected, and viral RNA was extracted from the plasma using the E.Z.N.A. Viral RNA kit according to the manufacturer's instructions (OMEGA bio-tek, Norcross, GA). Viral loads were determined using the iScript One-Step RT-PCR kit with SYBR green and the manufacturer's instructions (Bio Rad, Hercules, CA). Virus-specific primers were designed based on a conserved region in the LTR of SIVcpzEK505, SIVcpzMB897 and SIVcpzLB715 (GenBank accession numbers: DQ373065, EF535994.1 and KP861923.1, respectively). The primers designed for qRT-PCR were as follows: 1.) SIVcpzEK505: forward 5'-TAGTGTGTGCCCATCCATTCG-3' and reverse 5'-CACCGCCAGTCAAAATTGCG-3', 2.) SIVcpzMB897 forward (5'-

CCTCAGATATTAAGTGTCTGTGCGG-‘3) and reverse (5’-GCTAGTCAAAAATTAGGCGTACTCACC-3’), and 3.) SIVLB715 forward (5’-TGCTCGGACTCTGGTAACTA-3’) and reverse (5’-CCGCTACTTCTGGTTTCACTTTCACTT-3’). All primer sets listed above were used in a qRT-PCR reaction with the following cycling conditions: 50°C for 10 min, 95°C for 5 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 30 sec in the Bio Rad C1000 Thermo Cycler with the CFX96 Real-Time System (Bio Rad, Hercules, CA). The standard curve was prepared using a series of 10-fold dilutions of viral SIVcpzEK505, SIVcpzMB897 or SIVcpzLB715 LTR at a known concentration. The sensitivity of this assay was 1,000 copies per ml (1 copy per µl).

7.2.4 Determination of CD4⁺ T cell levels

Peripheral blood was collected bi-monthly from infected and control mice by tail vein puncture. Human cell engraftment levels were assessed by flow cytometry. 5 µl of FcγR-block (Jackson ImmunoResearch Laboratories, Inc. West Grove, PA) was added to the blood for 5 min. The blood was then stained with fluorophore conjugated hCD45-FITC, hCD3-PE and hCD4-PE/Cy5 (BD Pharmingen, San Jose, CA) for 30 min. Erythrocytes were lysed using the Whole Blood Erythrocyte Lysing kit according to the manufacturer’s instructions (R&D Systems, Minneapolis, MN). The stained cells were then fixed in 1% paraformaldehyde and 0.45 µm-filtered. To assess CD4⁺ T cell depletion in uninfected and infected mice, the CD3⁺ T cells levels were calculated as a ratio of the entire CD45⁺ (lymphocyte common antigen) population. The CD4⁺ T cell population levels were then determined as a percentage of the entire CD3⁺ T cell population. Baseline levels of the CD45⁺, CD3⁺ and CD4⁺ cells were measured prior to infection as a control. All flow cytometry data was analyzed using the FlowJo v10.0.7 software package

(FlowJo LLC, Ashland, OR). CD4⁺ T cell decline was assessed utilizing a two-tailed Student's *t*-test ($p < 0.001$) to compare the two groups of mice, infected and uninfected.

7.2.5 Illumina-based deep sequencing and analysis

For two mice of each generation, viral RNA from the plasma collected approximately at weeks 3, 11, 19 and 24 post-infection from two mice from each passage were used to synthesize the cDNA using SuperScript IV and the manufacturer's instructions (Invitrogen, Carlsbad, CA). Two separate multiplexed primer pools containing overlapping regions between, but not within the pools, were designed utilizing the Primal Scheme software (<https://primal.zibraproject.org>) in order to generate overlapping 400 base pair amplicons that spanned our entire viral genomes as seen in Supplementary Tables 7.1, 7.2 and 7.3 (Quick et al., 2017). The amplicons were prepped for Illumina-based deep sequencing using the Nextera XT DNA Library preparation kit and the manufacturer's instructions (Illumina, San Diego, CA). The amplicon library was deep sequenced at the sequencing core facility at the University of Wisconsin, Madison, utilizing a MiSeq Illumina desktop sequencer (Invitrogen, Carlsbad, CA). All sequence reads were prepared as FASTQ file format and analyzed as described below.

7.2.6 Cell Culture

Whole blood filter packs were obtained from the Garth Englund Blood Center of Fort Collins, CO. Mononuclear cells were isolated by Ficoll-Plaque density centrifugation. PBMC were grown and maintained in RPMI media containing 10% heat inactivated fetal bovine serum (HI FBS), 2x antibiotic-antimycotic mix (Thermo Fisher Scientific, Waltham, MA) and 20 ng/mL IL-2 (R&D Systems, Inc., Minneapolis, MN). For viral propagation, the PBMC were CD8 depleted by positive selection and then stimulated using 100 ng/mL of anti-CD3 and anti-CD28 soluble

antibody (Miltenyi Biotec Inc., Auburn, CA) for 48 h. The TZM-bl reporter and 293 cell lines were cultured in DMEM media containing 10% HI FBS, 1% antibiotic-antimycotic mix (ThermoFisher Scientific, Waltham, MA) and 1% L-glutamine.

7.2.7 Viral propagation, titration and subsequent viral passaging of SIVcpz

To propagate the virus from the first passage to the next, bone marrow, thymus, spleen, mesenteric and axillary lymph nodes, and whole blood obtained through cardiac puncture were harvested from the SIVcpz infected mice with the highest plasma viral titer. Lymphocyte fractions isolated by Ficoll-Plaque density centrifugation were seeded at a density of $2-3 \times 10^6$ cells/mL and activated for 48 h with 100 ng/mL of anti-CD3 and anti-CD28 soluble antibody (Miltenyi Biotec Inc., Auburn, CA). To assure viral infection of the next generation mice, these cells were then co-cultured for 48hrs with fresh splenocytes obtained from the new hu-mice cohort used for serial passage. These cultured cells together with culture supernatants containing the virus were inoculated intraperitoneally to the next hu-mice batch of 5 mice.

7.2.8 Quantification of variant frequencies

To calculate the variant frequencies in the sequencing datasets, a strategy similar to that we previously described in Schmitt et al. was used (Schmitt et al., 2018). Briefly, we removed the primer sequences first by trimming 30 bases off the ends of the reads, then removed the adapter sequences and low-quality bases using the cutadapt software v1.9.1 (Martin, 2011). Filtered reads were mapped to stock virus consensus sequences using the bowtie2 software v2.2.5 (Langmead and Salzberg, 2012). Bowtie2 BAM format output was used as input to lofreq software v2.1.2 to call variants (Wilm et al., 2012). To qualify for a variant, we required >100 coverage and >1% frequency. The impact of variants was determined, and variants were plotted using R and ggplot2

(ISBN: 0387981403). Scripts are available at GitHub repositories, https://github.com/stenglein-lab/SIV_to_HIV-1, and https://github.com/stenglein-lab/viral_variant_explorer.

7.2.9 Assessment of amino acid frequencies in HIV-1 and SIV database sequences

To quantify the frequencies of specific amino acids in SIV and HIV sequences for use as a comparison available from the HIV Sequence Database Compendium (<http://www.hiv.lanl.gov/>), multiple protein sequence alignments for each HIV-1/SIVcpz protein were downloaded from the 2017 compendium, which were the most recent alignments available for retrieval in April of 2019 (Foley, 2017). From the total HIV-1/SIVcpz alignments, subsets containing only HIV-1 or SIV sequences were extracted. These were used to create a position frequency matrix by tabulating, for each position (each column) of the alignment, the number of observations of each amino acid or gap character. A mapping of the amino acid position in the SIV protein to the position (column) in the multiple sequence alignment was then established for each protein of each of the three SIVcpz strains (EK505, MB897 and LB715). The position frequency matrix plus this mapping helped determine the frequency in these sets of HIV and SIV sequences of any variant that arose during the passage/generation experiments. The scripts and data files used to conduct this analysis are available at: https://github.com/stenglein-lab/SIV_to_HIV-1.

7.3 Results

7.3.1 Chimpanzee native viral strains SIVcpzEK505, LB715, and MB897 can establish productive infection and chronic viremia in humanized mice

To determine if SIVcpzEK505, MB897 and LB715 viruses of chimpanzee origin can establish productive infection, cohorts of five hu-HSC mice were inoculated with each of these and later serially passaged through different cohorts for a total of two generations (Figure 7.1).

Plasma viral loads were assessed on a weekly basis using qRT-PCR with strain-specific primers and probes. Viral infection was evident within 14 days after the initial exposure with all the three SIVcpz progenitor strains demonstrating their ability to productively infect human cells *in vivo* (Figure 7.2). Plasma viral loads in SIVcpzEK505-infected hu-mice increased by 2-logs, peaking by 84 days post-inoculation. In comparison, viral loads peaked 14 days sooner during the second passage and were 1-log higher indicating enhanced replicative fitness of the virus (Figure 7.2A).

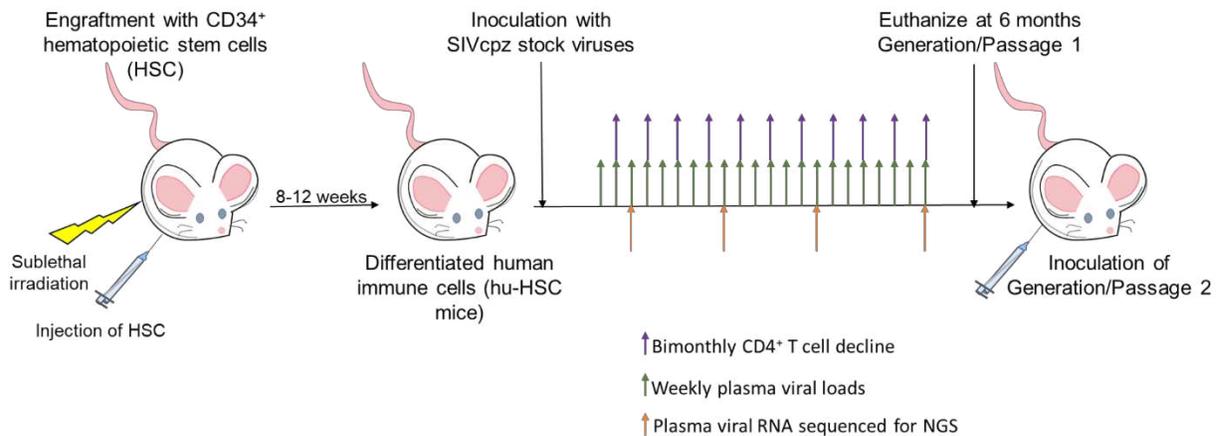


Figure 7.1 Viral passaging experimental scheme. Neonatal mice were sublethally irradiated and injected intrahepatically with human CD34⁺ hematopoietic stem cells as described in the materials and methods. Mice were screened 8-12 weeks post-reconstitution for human immune cell engraftment. Five hu-HSC mice with high human hematopoietic cell engraftment levels from a single cohort were infected intraperitoneally (*i/p*) with either cell-free SIVcpzEK505, LB715 or MB897 and monitored for plasma viral loads on a weekly basis using qRT-PCR. CD4⁺ T cell decline was assessed bimonthly by flow cytometry and plasma viral RNA was used for Illumina-based deep sequencing at 3, 11, 19- and 24-weeks post-inoculation. The infected hu-HSC mice were euthanized at around 24 weeks and the isolated, infected cells were subsequently injected intraperitoneally into five naïve hu-HSC mice to begin the second passage/generation

7.3 Results

7.3.1 Chimpanzee native viral strains SIVcpzEK505, LB715, and MB897 can establish productive infection and chronic viremia in humanized mice

To determine if SIVcpzEK505, MB897 and LB715 viruses of chimpanzee origin can establish productive infection, cohorts of five hu-HSC mice were inoculated with each of these

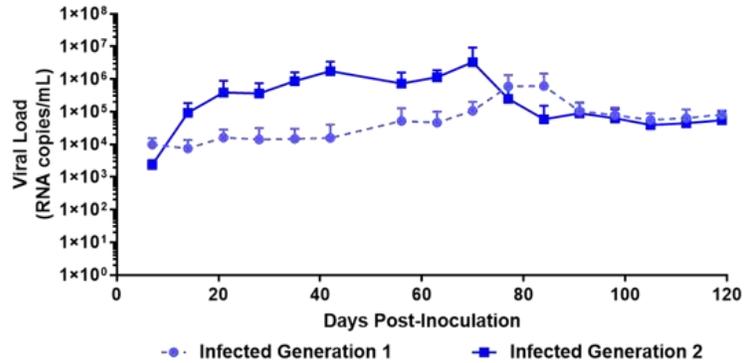
and later serially passaged through different cohorts for a total of two generations (Figure 7.1). Plasma viral loads were assessed on a weekly basis using qRT-PCR with strain-specific primers and probes. Viral infection was evident within 14 days after the initial exposure with all the three SIVcpz progenitor strains demonstrating their ability to productively infect human cells *in vivo* (Figure 7.2). Plasma viral loads in SIVcpzEK505-infected hu-mice increased by 2-logs, peaking by 84 days post-inoculation. In comparison, viral loads peaked 14 days sooner during the second passage and were 1-log higher indicating enhanced replicative fitness of the virus (Figure 7.2A). With the SIVcpzLB715 viral strain, first passage viral loads increased slowly increasing through 161 days ($>10^5$ RNA copies/mL) post-inoculation (data not shown). However, the viral loads became more robust during the second generation peaking at 2-logs higher ($>10^7$ RNA copies/mL) and 35 days earlier (112 days post-inoculation) (Figure 7.2B). Thus, the SIVcpzLB715 virus also appears to have gained growth fitness in human immune cells over time. While still successful in initiating infection, viral loads with the SIVcpzMB897 were found to be lower than the other strains tested. Viral loads stagnated slightly above the limit of detection for most of the first passage period and peaked ($\geq 10^4$ RNA copies/mL) at 119 days post-inoculation (Figure 7.2C). Unlike the other two strains above, the increase in second-generation plasma viral loads was slow for SIVcpzMB897 until around day 77. With the SIVcpzMB897 virus there was clearly a delay in viral adaptation to human immune cells during the first generation, and the viral growth appeared to evolve more rapidly during the middle of the second generation indicating eventual increased viral fitness. As expected, no plasma viral loads were detected in the control uninfected hu-HSC mice (data not shown). Previous data on HIV-1 BaL infections in hu-HSC mice showed that peak viral loads reach at around 8 weeks post-inoculation (Berges et al., 2010). This is unlike the first generation of SIVcpz inoculations into hu-HSC mice where SIVcpzEK505, SIVcpzMB897 do not

display peak viral loads till days 84 and 119, respectively, and SIVcpzLB715 does not peak during the first generation. In contrast, the viral loads peaked much earlier during the second generation, reflecting a greater adaptation towards human immune cells.

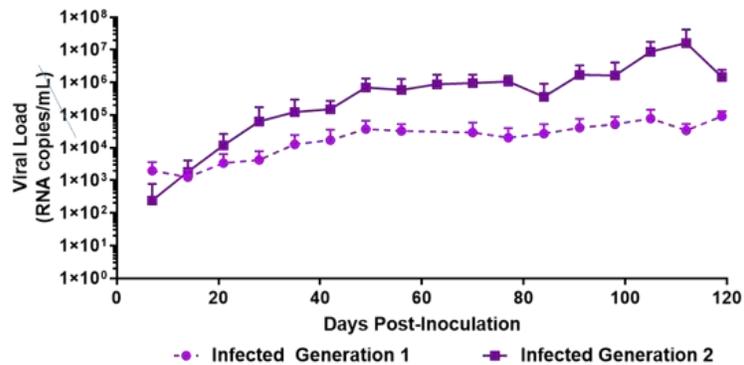
7.3.2 Serial passage of SIVcpz viruses in hu-mice leads to more rapid CD4⁺ T cell decline

CD4⁺ T cell decline is a key aspect of HIV infection (Aldrovandi et al., 1993; Baenziger et al., 2006; Berges et al., 2006; Denton and Garcia, 2011; Akkina, 2013b). Accordingly, we followed the CD4⁺ T cell levels in SIVcpz infected mice over time. The baseline human CD4⁺ T cell levels prior to infection in all mice was greater than 50% of all human CD3⁺ cells. By the end of second generation, CD4⁺ T cell decline became apparent in mice infected with any of the three viruses tested ($p < 0.001$), with varying levels of decline depending on the virus (Figure 7.3). The only SIVcpz strain with a noticeable decline in CD4⁺ T cells during the first passage was SIVcpzEK505. This decline was observed beginning at day 14 and was maintained over the course of 6 months (Figure 7.3A). However, it became slightly more pronounced during the second passage wherein decline was also seen within 14 days post-inoculation and persisted throughout the course of infection. SIVcpzLB715 and SIVcpzMB897 strains displayed no notable CD4⁺ T cell decline during the first passage, but SIVcpzLB715 infected mice showed rapid CD4⁺ T cell loss 14 days post-inoculation during the second passage (Figure 7.3B). While the SIVcpzMB897 second generation showed a much less precipitous drop in CD4⁺ T cells relative to the other

(A) SIVcpzEK505



(B) SIVcpzLB715



(C) SIVcpzMB897

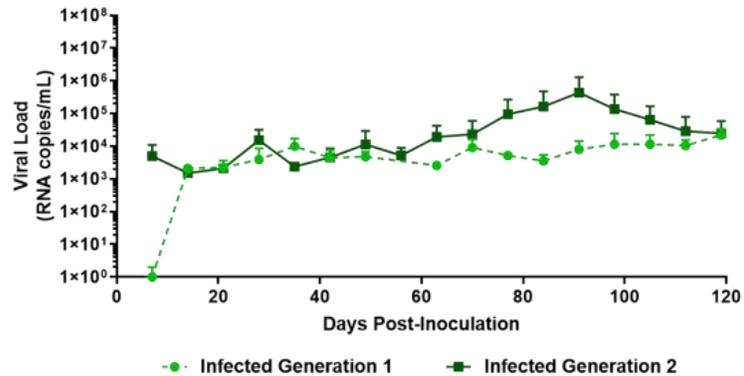


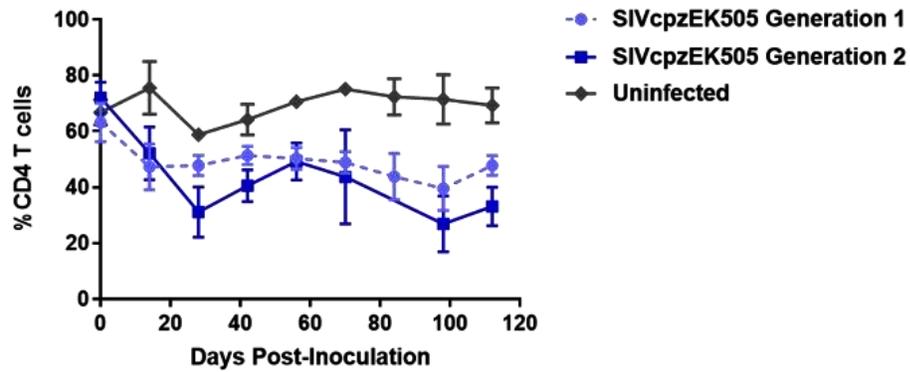
Figure 7.2 SIVcpz infection leads to detectable chronic viremia in humanized mice. (A) EK505, (B) LB715 and (C) MB897 first- and second-generation plasma viral loads. Five humanized mice were infected via i/p injections and the plasma viral loads were monitored by qRT-PCR on a weekly basis for the duration of the infection. No viral loads were detected in the uninfected control mice (data not shown).

SIVcpz strains, significant CD4⁺ T cell decline was still observed 98 days post-inoculation (Figure 7.3C). Overall, these results showing rapid decline of CD4⁺ T cells during the second passage versus the first suggest enhanced viral adaptation and pathogenicity. More importantly, this data demonstrates that all three progenitor SIVcpz viruses could establish productive viremia in hu-mice resulting in CD4⁺ T cell decline.

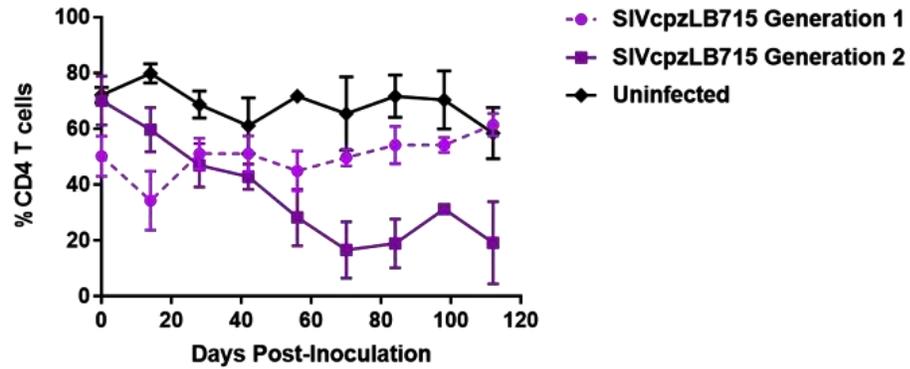
7.3.3 SIVcpz viruses accumulate adaptive viral genetic changes during long-term serial passages in hu-mice

Selection pressures imposed by the human immune cell environment in hu-mice are likely to promote adaptive changes in SIV progenitor viruses. This “genetic bottleneck” arises from both physical and immunological conditions that prevent most variants within the founder viral population from establishing infection within the human host (Joseph et al., 2015; Kariuki et al., 2017). The adaptive changes to viral fitness may be mediated through changes that improve viral entry, replication, epistasis, immune escape, host restriction factors and/or pathogenicity. Furthermore, novel mutations in the virus may arise due to these host selection pressures on genes such as *vif*, *env* or *nef* that counteract host restriction factors like TRIM5 α , APOBEC3 proteins and tetherin. To assess the genomic changes occurring in SIVcpz viral strains during passages in humanized mice through two generations, we utilized Illumina-based deep sequencing to identify possible signatures of viral adaptation. This was performed on the genomes of SIVcpzEK505, LB715 and MB897 clonal stock viruses used for initial infections and the viruses sampled at roughly 3, 11, 19- and 24-weeks post-inoculation during the first and second passages. The consensus sequence of each viral stock was used as a reference point for aligning the sequence reads of passaged viruses at each time point. Variants that occurred at a frequency of greater than 1% of the population relative to the reference sequence were identified and then characterized as

(A)



(B)



(C)

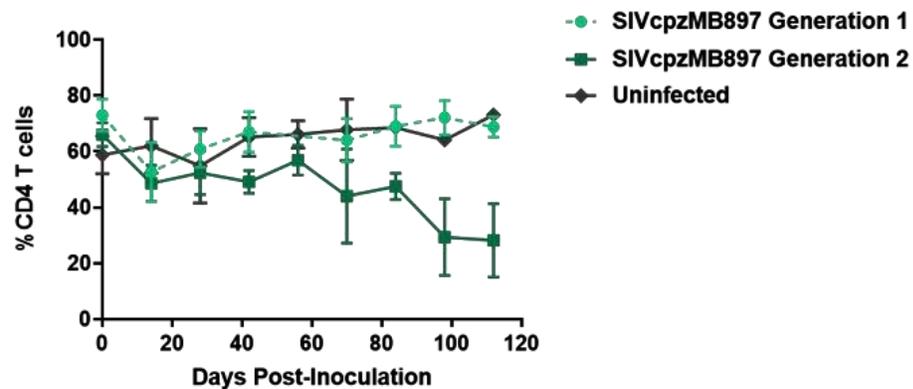


Figure 7.3 SIVcpz infection leads to a gradual decline in CD4⁺ T cells in humanized mice. The (A) EK505, (B) LB715 and (C) MB897 first- and second-generation percentage of circulating CD4⁺ T cells relative to the total CD45⁺/CD3⁺ cell populations were assessed on a bimonthly basis. Statistically significant depletion was seen in the second generation in infected hu-HSC mice relative to the uninfected control (two-tailed Student's *t*-test, $p < 0.001$).

either synonymous or nonsynonymous. We looked for potentially adaptive nonsynonymous variants based on several criteria including: 1.) variants that arose during the first passage and increased in frequency amongst the viral population over time and 2.) variants that appeared in multiple mice within a given passage across both passages. All three SIVcpz strains tested gave rise to many new nonsynonymous variants that increased in frequency across both serial passages which were absent in the original founder stock viral sequences (Tables 7.1, 7.2 and 7.3). Nine variants matching this criterion were identified in SIVcpzEK505 strain, seven in SIVcpzLB715, and with the greatest number of qualifying variants, 16, identified in SIVcpzMB897. All of these nonsynonymous variants were absent in their respective stock virus population. The majority of substitutions occurring with at least 50% frequency at the last time point analyzed from the viruses sampled during the second passage were detected in the *env* gene, although other variants were found throughout the rest of the viral genome. SIVcpzMB897 had eleven nonsynonymous changes in *env*, two within *gag* and *nef*, and one within *vif* meeting these criteria (Table 7.3). Based on their relative size, the *gag* and *pol* genes displayed very few substitutions when compared to the *env* gene, which is consistent with the previous knowledge that the *gag* and *pol* genes are highly conserved amongst retroviruses. Several of these substitutions identified that arose during the first passage were maintained at almost 100% frequency during the second passage in both replicates of mice from which the sequenced viral samples were obtained (Supplementary Figure 7.1). Relatively few substitutions were seen in *nef*, *vif*, *vpu* and *vpr* genes in any of the SIVcpz strains tested which was surprising, given the role of these proteins in counteracting human restriction factors (Tables 7.1, 7.2 and 7.3).

Utilizing the sequence alignments in the HIV Sequence Compendium, we next sought to assess whether or not each nonsynonymous variant observed produced a more “HIV-1-like” or

“SIV-like” virus (Foley, 2017). We determined if a variant amino acid detected is present in a higher fraction of HIV-1 or SIV sequences (Tables 7.1, 7.2, and 7.3). Two amino acid substitutions identified in Env for both SIVcpzMB897 (K414R and F702L) and SIVcpzEK505 (K291N and S611T) strains were found in a higher fraction of HIV-1 sequences than in those of SIVs’, indicating a change more representative of HIV than SIV. Interestingly, with the SIVcpzLB715 strain, the detected substitutions were neither more “SIV-like” nor “HIV-1-like” in nature. Additional passages *in vivo* of these SIVcpz viruses during a much longer time span may be necessary to observe the evolution towards HIV-1.

7.4 Discussion

Here we evaluated both the genetic and phenotypic changes that occurred during the evolution of progenitor SIV viruses towards HIV-1 using a humanized mouse model. In the first set of experiments, clonal stocks of three NHP SIV strains SIVcpz LB715, MB897 and EK505 belonging to groups M and N; respectively, were inoculated into hu-HSC mice to determine whether they can establish infection in the human surrogate host and affect CD4⁺ T cell levels. Later, the viruses were serially passaged in different cohorts of hu-mice to mimic serial human transmission and evaluate the potential increase in viral fitness and pathogenicity. While no animal model is perfect in fully mimicking the human system, the hu-HSC mouse model used here has several advantages over using human PBMC in *in vitro* experiments. These include the provision of a physiologic *in vivo* environment, immune pressures and prolonged time frame like six months or longer for serial viral propagation/adaptation. Each of the viral strains was able to infect hu-mice validating that SIVs originating from chimpanzees are capable of readily infecting human hosts as might have happened in the past. We followed the kinetics of viremia for several months in infected hu-mice and found there was a slow but steady increase in viremia over time (Figure

Table 7.1 Amino acid substitutions resulting from candidate adaptive mutations in SIVcpzEK505 passaged virus.

Protein	Position	Stock Residue ^a	Variant Residue ^b	P1 Endpoint Frequency ^c	P1 Endpoint Frequency ^d	P2 Endpoint Frequency ^e	P2 Endpoint Frequency ^f	SIV Fraction ^g	HIV-1 Fraction ^h
Gag	35	V	I	0.88	0.14	1.00	1.00	0.00	0.04
Vif	110	Y	N	0.00 ⁱ	0.00 ⁱ	0.72	0.30	0.00	0.00
Env	291	K	N	0.00 ⁱ	0.00 ⁱ	0.85	0.76	0.78	0.96
Env	402	V	G	0.56	0.00 ⁱ	0.95	0.98	0.00	0.01
Env	414	P	S	0.00 ⁱ	0.00 ⁱ	0.90	0.64	0.00	0.04
Env	611	S	T	0.16	0.00 ⁱ	0.81	0.94	0.61	0.95
Env	616	G	E	0.00 ⁱ	0.00 ⁱ	0.87	0.78	0.04	0.07
Env	668	R	K	0.30	0.00 ⁱ	0.88	0.31	0.04	0.02
Env	703	E	K	0.60	0.39	1.00	0.99	0.00	0.00

^aConsensus amino acid from the stock virus sequencing dataset.

^bVariant amino acid in passage 2 (P2) viruses.

^cFrequency of the variant mutation in passage 1 (P1) replicate 1 sequencing datasets.

^dFrequency of the variant mutation in passage 1 (P1) replicate 2 sequencing datasets.

^eFrequency of the variant mutation in passage 2 (P2) replicate 1 sequencing datasets.

^fFrequency of the variant mutation in passage 2 (P2) replicate 2 sequencing datasets.

^gFraction of SIV sequences from the HIV sequence compendium that contain the variant amino acid.

^hFraction of the HIV-1 sequences from the HIV sequence compendium that contain the variant amino acid.

ⁱThe 0 indicates below the limit of detection (1%) of our variant identification pipeline.

Table 7.2 Amino acid substitutions resulting from candidate adaptive mutations in SIVcpzLB715 passaged virus.

Protein	Position	Stock Residue ^a	Variant Residue ^b	P1 Endpoint Frequency ^c	P1 Endpoint Frequency ^d	P2 Endpoint Frequency ^e	P2 Endpoint Frequency ^f	SIV Fraction ^g	HIV-1 Fraction ^h
Pol	329	E	K	0.00 ⁱ	0.00 ⁱ	0.84	0.55	0.00	0.01
Vpr	68	L	P	0.00 ⁱ	0.00 ⁱ	0.85	0.97	0.00	0.00
Vpu	45	R	T	0.00 ⁱ	0.00 ⁱ	0.66	0.79	0.00	0.00
Env	523	E	Q	0.00 ⁱ	0.00 ⁱ	0.50	0.76	0.00	0.00
Env	800	G	R	0.55	0.25	1.00	1.00	0.00	0.00
Env	841	G	R	0.00 ⁱ	0.00 ⁱ	0.47	0.60	0.00	0.00
Nef	36	D	N	0.00 ⁱ	0.00 ⁱ	0.83	0.93	0.00	0.00

^aConsensus amino acid from the stock virus sequencing dataset.

^bVariant amino acid in passage 2 (P2) viruses.

^cFrequency of the variant mutation in passage 1 (P1) replicate 1 sequencing datasets.

^dFrequency of the variant mutation in passage 1 (P1) replicate 2 sequencing datasets.

^eFrequency of the variant mutation in passage 2 (P2) replicate 1 sequencing datasets.

^fFrequency of the variant mutation in passage 2 (P2) replicate 2 sequencing datasets.

^gFraction of SIV sequences from the HIV sequence compendium that contain the variant amino acid.

^hFraction of the HIV-1 sequences from the HIV sequence compendium that contain the variant amino acid.

ⁱThe 0 indicates below the limit of detection (1%) of our variant identification pipeline.

Table 7.3 Amino acid substitutions resulting from candidate adaptive mutations in SIVcpzMB897 passaged virus.

Protein	Position	Stock Residue ^a	Variant Residue ^b	P1 Endpoint Frequency ^c	P1 Endpoint Frequency ^d	P2 Endpoint Frequency ^e	P2 Endpoint Frequency ^f	SIV Fraction ^g	HIV-1 Fraction ^h
Gag	35	V	I	0.00 ⁱ	0.00 ⁱ	0.99	0.64	0.00	0.04
Gag	105	E	K	0.00 ⁱ	0.00 ⁱ	0.90	0.74	0.69	0.02
Vif	47	E	K	0.88	1.00	1.00	1.00	0.00	0.01
Env	149	N	Y	0.54	0.68	0.97	0.75	0.00	0.01
Env	346	R	Q	0.00 ⁱ	0.50	1.00	0.99	0.00	0.02
Env	351	S	N	0.00 ⁱ	0.28	1.00	0.99	0.04	0.00
Env	413	G	R	0.83	0.64	0.99	0.99	0.00	0.00
Env	413	G	E	0.14	0.66	0.99	0.99	0.00	0.01
Env	414	K	R	0.33	0.00 ⁱ	0.84	0.91	0.17	0.30
Env	442	N	I	0.00 ⁱ	0.51	0.99	0.99	0.04	0.06
Env	475	L	V	0.08	0.60	0.74	0.52	0.87	0.50
Env	596	K	N	0.49	0.07	0.40	0.87	0.26	0.00
Env	702	F	L	0.17	0.10	0.90	0.61	0.04	0.29
Env	838	G	S	0.12	0.94	1.00	1.00	0.00	0.00
Nef	126	N	K	0.89	0.08	0.34	0.74	0.00	0.00
Nef	163	V	M	0.57	0.22	0.99	0.98	0.00	0.00

^aConsensus amino acid from the stock virus sequencing dataset.

^bVariant amino acid in passage 2 (P2) viruses.

^cFrequency of the variant mutation in passage 1 (P1) replicate 1 sequencing datasets.

^dFrequency of the variant mutation in passage 1 (P1) replicate 2 sequencing datasets.

^eFrequency of the variant mutation in passage 2 (P2) replicate 1 sequencing datasets.

^fFrequency of the variant mutation in passage 2 (P2) replicate 2 sequencing datasets.

^gFraction of SIV sequences from the HIV sequence compendium that contain the variant amino acid.

^hFraction of the HIV-1 sequences from the HIV sequence compendium that contain the variant amino acid.

ⁱThe 0 indicates below the limit of detection (1%) of our variant identification pipeline.

7.2A-C) during the first passage with each of the viral strains used. When the viruses from the first passage were used to infect new cohorts of hu-mice, viremia levels rapidly increased with the viral load set points being relatively higher than in the first passage. This showed that the SIVcpz strains were adapting to propagate better in the hu-mouse system as a consequence of increased fitness. Since helper CD4⁺ T cell decline is a hallmark of HIV-1 infection that results in immunosuppression, we also assessed this trait between the first and second passage of the hu-mouse adapted SIV strains. Whereas a slower pace of CD4⁺ T cell decline was noted during the first passage with all three viral strains over time, the cell decline became more pronounced and rapid during the second passage (Figure 7.3A-C). Both the increased amplitude of the viral loads and faster pace of CD4⁺ T cell decline seen with each strain during the second passage is indicative of viral evolution resulting in increased human pathogenicity.

While all the three SIVcpz viral strains tested above were capable of initiating infection in hu-mice, the SIVcpzMB897 strain displayed less robust viral loads than the others during the first passage. This contrasts with the conclusions of Sato et al., which reported that SIVcpzMB897 was more preadapted towards human immune cells compared to the other SIVcpz strains (Sato et al., 2018). Our findings are further supported by the longer timeframe required for noticeable CD4⁺ T cell decline by SIVcpzMB897 when compared to SIVcpzLB715 and SIVcpzEK505 strains. Furthermore, sequence data showed a greater number of variants in SIVcpzMB897 relative to the other two SIVcpz strains during the serial passage indicating it requires further adaptation for efficient propagation in humans.

Based on sequence differences documented between SIVs and HIVs, it is clear that numerous genetic changes must have taken place in the progenitor SIV to give rise to the four known HIV-1 lineages (Keele et al., 2006; Van Heuverswyn et al., 2006; Sharp and Hahn, 2011;

Hemelaar, 2012; Sauter and Kirchhoff, 2019). During this process, the progenitor SIVs had to evade or adjust to a number of native human cell barriers capable of interference with many steps in the viral life cycle. In addition to cell surface receptors involved in viral attachment, other barriers consist of restriction factors. These currently include, but are not limited to, TRIM5 α , APOBEC3 proteins, SAMHD1, SERINC3/5, and tetherin which function by various mechanisms (Sheehy et al., 2002; Sayah et al., 2004; Stremlau et al., 2004; Neil et al., 2008; Le Tortorec and Neil, 2009; Berger et al., 2011; Goldstone et al., 2011; Serra-Moreno and Evans, 2012; McNatt et al., 2013; Etienne et al., 2015; Rosa et al., 2015; Usami et al., 2015; Ghimire et al., 2018; Meyerson et al., 2018). The list of restriction factors is continually expanding pointing to many other yet undiscovered intrinsic host defense mechanisms (Ghimire et al., 2018). These factors impose a strong selective pressure on non-native viruses to acquire adaptive changes. To identify potential genetic changes, we analyzed whole viral genomes from the first and second passages at different time points with a focus on nonsynonymous genetic changes. Viral variants that increased in population frequency from the first to the second generation are presented in Figure 7.4. Some variants occurring in low frequency at the early time points became more prominent/dominant during later time points and became fixed (Tables 7.1-7.3). We also noticed several variants of transient nature that were seen at early time points which disappeared subsequently (data not shown).

Among the recent reports on SIV infection and evolution in humanized mice (Yuan et al., 2016; Sato et al., 2018; Yuan et al., 2018), only limited regions of the viral genomes were evaluated by sequence analysis in one report and only variants from short-term infection (14-16 weeks of infection) at a single terminal time point were analyzed in all three studies (Yuan et al., 2016; Sato et al., 2018; Yuan et al., 2018). To extend and improve upon these early studies and

increase the breadth of analysis, here we examined the viral variants sampled at multiple time points during a much longer period spanning more than a year. Sequential passages performed in different cohorts of hu-mice also allowed the recapitulation of serial viral transmission and subsequent spread to a new human host. Not surprisingly, numerous changes could be identified throughout the viral genome beyond the two mutations described in the envelope gene in the studies mentioned above (Yuan et al., 2016; Sato et al., 2018).

Among the many nonsynonymous changes seen, of note is the V35I substitution in Gag. This arose independently in both the EK505 (Group N) and MB897 (Group M) strains. While its exact function needs to be determined, that it arose in two divergent progenitor viral strains points to an essential role for it in human immune cell adaptation. However, this change was not seen in the LB715 strain (also group M, like MB897). More passages of LB715 may be needed for it to appear at a later time point or alternatively, other mutations in the vicinity might be playing a compensatory role. We also looked for the potential emergence of an important Gag M30K/R substitution, described previously in elegant studies by Hahn, Sharp and colleagues detailing the origins of HIVs from NHP viruses (Wain et al., 2007; Sharp and Hahn, 2011; Bibollet-Ruche et al., 2012). This mutation was shown to be highly conserved in HIV-1, but absent in most SIVs, suggesting it being a species-specific signature. However, it was not detected in our present studies. Corroborating our findings here, results from recent reports on SIVcpzMB897, SIVcpzEK505 and other chimpanzee-derived strains passaged in hu-mice also did not detect this Gag M30K/R substitution (Yuan et al., 2016; Sato et al., 2018). Thus, this substitution may not be as critical as previously thought for human cell adaptation, at least to overcome intracellular barriers during the early stages of human propagation. Supporting this hypothesis is the observation by Sato et al., that introduction of this substitution into SIVcpzMB897 strain did not

augment viral fitness and replication kinetics in hu-mice (Sato et al., 2018). Alternatively, another variant the V35I substitution namely seen here may be playing an equivalent role similar to the Gag M30K/R substitution.

The viral envelope protein is involved in host cell receptor binding and dictates the efficiency of initial viral host cell interactions. Importantly, a nonsynonymous change in Env identified in strain MB897 in two previous reports and found to improve viral pathogenicity, the Env G413R/(E) mutation, was also detected in our studies (Yuan et al., 2016; Sato et al., 2018). This variant appeared at 11 weeks post-inoculation and later became fixed. Its detection in three independent studies from different laboratories makes a compelling case for its critical nature in the SIVcpzMB897 strain's adaptation to human cells. During the course of our two serial passages, we also observed two additional mutations of significance, namely, the Env N442I substitution in MB897 strain. These increased in frequency during the first passage and remained stable through the second passage. Env N442I is located adjacent to a well-characterized CD4 binding site in the V5 loop of envelope and potentially aids in facilitating viral attachment (Xiang et al., 2002; Chen et al., 2005; Decker et al., 2005; Yu et al., 2010). Another mutation, the Env K414R change, located in the CD4 binding site itself appeared later and began to increase in frequency during the second passage. It is possible that the initial adjacent mutations "prime" the site itself to better accommodate more drastic changes to the site, which may explain the delayed appearance of the Env K414R mutation relative to the others. The other two SIV strains tested also showed adaptive changes in receptor interacting sites. In SIVcpzEK505, an Env variant V402G in the CD4 binding site appeared during the late first passage (week 19) and eventually made up 100% of the viral population by the second passage. Another substitution directly adjacent to the chemokine binding site, the Env P414S, was detected at the beginning of the second passage (Xiang et al., 2002; Chen

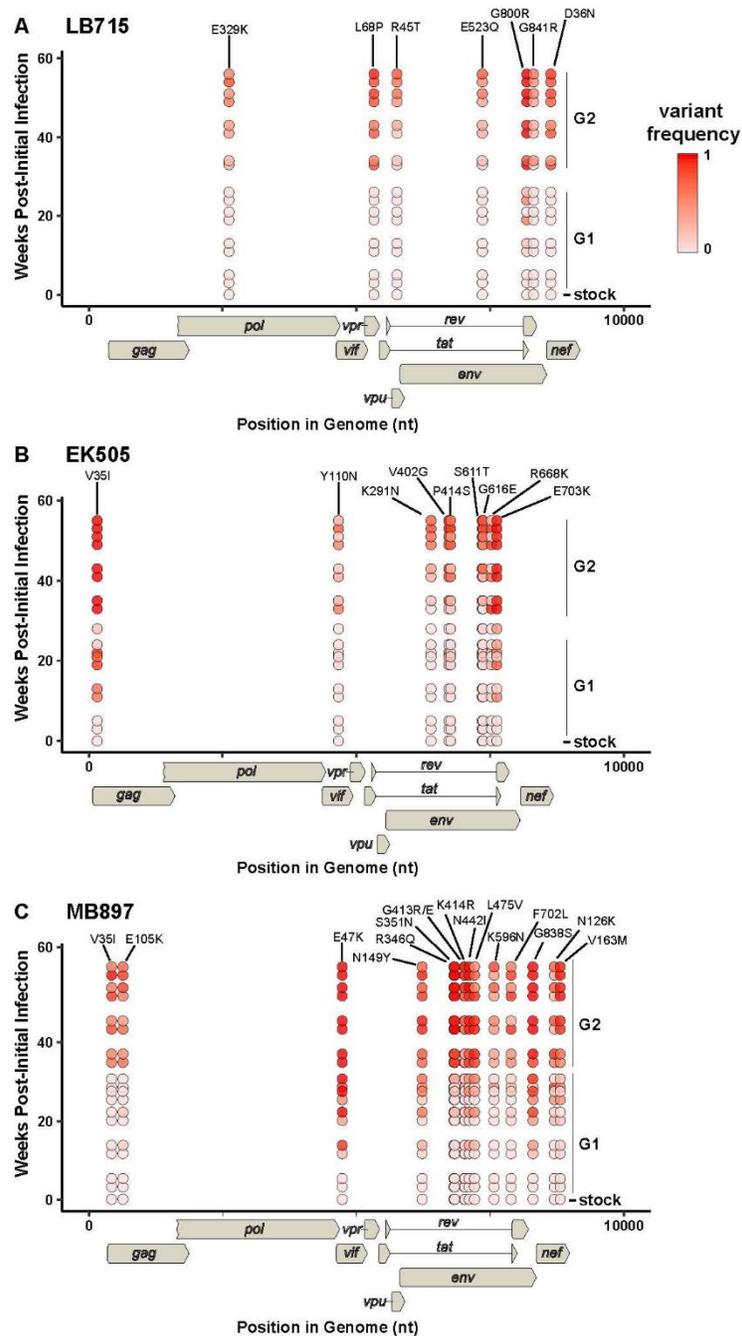


Figure 7.4 Viral variants increasing in population frequency from the first to the second generation. RNA samples from two mice per generation were sequenced. The frequencies of individual variants within a given timepoint sample from SIVcpz (A) EK505, (B) LB715 and (C) MB897 are plotted as a function of the weeks post-initial infection. Diagrams of the viral genomes are indicated below the X-axis. Non-synonymous variants that had a mean frequency of >0.1 in the final time point are shown. The viral variant frequency is indicated by the red color scale shown. Week 0 corresponds to the frequency in the viral stock pre-infection and the second generation was subsequently plotted beginning at week 30. For each time point, replicates are offset vertically from each other. The residue changes and numbers for each position are listed above their respective locations.

(Xiang et al., 2002; Chen et al., 2005; Decker et al., 2005; Yu et al., 2010). These changes may have contributed to increased pathogenicity of SIVcpzEK505 as evidenced by both increased viral loads and more rapid CD4⁺ T cell decline that commenced at the beginning of the second passage. Similarly, the Env E523Q mutation seen in SIVcpzLB715 is located in the CD4 binding site and appeared during the second passage (Xiang et al., 2002; Chen et al., 2005; Decker et al., 2005; Yu et al., 2010). By evolving the CD4 and chemokine receptor binding sites, these viruses may be gaining improved engagement of the cell receptors and an overall higher fitness to the human host.

Several of the variants we detected did not appear until relatively late during the first human passage (19-24 weeks post-infection) and did not become predominant until the second passage (Supplementary Figure 7.1). A number of other mutations were shown to arise at single time points and then disappear entirely indicating their transient nature and irrelevance in the context of viral evolution. These observations underscore the importance of longer *in vivo* viral passages and serial propagation to ascertain viral variants at multiple time points as they arise.

In addition to incompatible cell receptors barriers for viral entry, other major barriers for virus transmission is the myriad of host restriction factors such as TRIM5 α , APOBEC3 proteins and tetherin. One of the functions of Vpu is to overcome tetherin. In some cases only a few genetic changes appear to be necessary to allow chimpanzee Vpu to be effective against human tetherin (Lim et al., 2010; Vigan and Neil, 2010; Kluge et al., 2013; Sauter and Kirchhoff, 2019). While we have not yet detected any variants directly connected to the known human restriction factors at these serial passage levels, several other significant variants with increased frequency are noted as shown in Tables 7.1-7.3, including the Vpu R45T substitution identified in SIVcpzLB715. We also identified two amino acid substitutions in both SIVcpzEK505 Y110N and SIVcpzMB897 E47K, which appeared fixed by the end of the second generation. Both of these variants were

below the limit of detection in the starting stock viruses (Tables 7.1 and 7.3). Additional Vif variants may continue to arise over repeated passaging in our humanized mouse model to further confirm candidate adaptive mutations that may contribute to counteracting the action of APOBEC3G (or APOBEC3 proteins) in human immune cells. These mutations and others need to be further characterized to determine their role in viral adaptation, pathogenesis and fitness. Two variants in Pol and Vpr, Pol E329K and Vpr L68P, that were identified in viral strain SIVcpzLB715 arose at the beginning of the second passage and then were maintained in over half of the viral population indicating their potential role for viral adaptation (Table 7.2). The Pol E329K mutation is located in the p51 reverse transcriptase subunit of the viral polymerase (Nagata et al., 2017) and the Vpr L68P substitution is found within the “LxxLL” motif (Kino et al., 1999; Kamata and Aida, 2000; Sherman et al., 2000; Kogan and Rappaport, 2011; Gonzalez, 2017). While, the function of the Pol E329K substitution is still unknown, the “LxxLL” motif is required for glucocorticoid-dependent activation of transcription to maintain resting and stress-related homeostasis as well as being directly implicated in the nuclear localization of the PIC complex (Kino et al., 1999; Kamata and Aida, 2000; Sherman et al., 2000; Gonzalez, 2017). This same mutation (Vpr L68P) was also identified as a proline (LxxLP) in HIV-1 NL4-3. This proline change has been noted in a participant sample but was not tested experimentally to confirm its effect on viral pathogenicity. However, previous studies have found that direct mutation of this residue leads to rapid progression of AIDS (Sherman et al., 2000; Gonzalez, 2017). Additionally, it is possible that some of these mutations may confer no additional fitness, but rather become fixed stochastically. However, the persistence of these mutations across multiple generations and immune environments make it unlikely that non-favorable mutations would increase in frequency within

the viral population over time. Further studies and passages of these viruses are needed to determine the direct function of these amino acid substitutions.

An important question is whether the nonsynonymous variants seen here during serial passages in a human surrogate system reflect evolution towards HIV-1 as a whole or singly align with those of existing variants in the SIV genomes. Sequence alignments were done using the HIV Sequence Compendium (Foley, 2017). Specifically, we assessed whether the observed amino acid variants are present in a higher fraction of HIV-1 sequences than in SIV sequences. Most of the mutations we identified were present in very few SIV or HIV strains. However, two variants, Env K291N and Env S611T observed in the EK505 progenitor strain were found approximately 20% more frequently in HIV sequences than in SIV (Foley, 2017). Additionally, these substitutions have no known function. The variant Env F702L seen in MB897 strain is located in the gp41 of Env adjacent to the Kennedy Epitope but no function is known (Steckbeck et al., 2011). None of the variants seen in the LB715 strain appeared to be common in either SIV or HIV strains. Several other variants seen were found to be more common in SIVs than in HIVs though the functionality of these changes remains unclear.

In summary, our results showed that HIV progenitor SIVcpz viral strains, from the pandemic Group M (MB897 and LB715) and non-pandemic Group N (EK505), can readily infect humanized mice leading to chronic viremia and upon further passage, marked CD4⁺ T cell decline. These findings reinforce the utility of the hu-mouse human surrogate system in modeling cross-species transmission and viral adaptation. Whereas the nonsynonymous sequence changes seen in the CD4 and chemokine receptor binding Env regions of these evolving viruses show one aspect of enhancing viral affinity and fitness to the human host, other notable changes might contribute to other steps in viral replication. While the sequences of the variant viruses at the second passage

spanning several months do not appear to be more like HIV, the SIV progenitors have responded to the selection pressure imposed by the *in vivo* human system by acquiring sequence changes. It remains to be seen if additional passages in hu-mice would impose further selection giving rise to viral strains more like HIV-1. Nevertheless, the ease with which these viruses infected humanized mice is remarkable, suggesting that many SIV strains currently circulating in the wild continue to pose a significant zoonosis risk to the human and have the potential to generate dangerous new HIV strains in the future. With the knowledge gained on how SIVcpz first adapted to human immune cells and gained virulence we will be able to better predict if, and how, other NHP derived lentiviruses may cross over into humans in the future.

CHAPTER 8: ADAPTATION OF SIVMAC239 *IN VIVO* USING HUMANIZED MICE AS A PARALOG FOR SIVSM ADAPTATION INTO HIV-2

8.1 Summary

HIV-2, though less pathogenic than HIV-1, is responsible for millions of infections worldwide. This virus is believed to have emerged as a result of the accumulation of adaptive mutations in a simian immunodeficiency virus originating in sooty mangabeys (SIVsm). While there is evidence that SIVsm has crossed into humans on at least nine separate occasions, these are not the only examples of SIVsm crossing over into another species. Captive sooty mangabeys infected with SIVsm were housed with rhesus macaques and inadvertently infected them with SIVsm. The resulting pathogenic infection gave rise to a new virus termed SIVmac, which has served as a popular model for studying HIV infections. Additionally, SIVhu was a virus isolated from a human patient who was exposed to SIV_{B670} through a laboratory accident. SIV_{B670} was originally a strain of SIVsm that was passaged twice in rhesus macaques. Therefore, we serially passaged SIVmac239, SIV_{B670} and SIVhu over three generations in humanized mice in order to determine the genetic changes that arise as these viruses adapt to an *in vivo* human immune environment. Plasma viral loads changed across these passages for each respective virus. CD4⁺ T cell decline, reminiscent of HIV infection, was observed in SIVmac239 and SIV_{B670}, though it was not observed with SIVhu. Analysis of viral sequences at various timepoints identified numerous nonsynonymous mutations that may be involved in the adaptation of these viruses to human immune cells. These data showcase that other SIVsm-derived viruses are capable of human infection and may shed light on what changes are necessary for these viruses to adapt to a new host.

8.2 Materials and Methods

8.2.1 Creation of Hu-HSC and BLT mice

All mice used in these studies were cared for in the Colorado State University Painter Animal Center. Fetal liver-derived human CD34⁺ cells were purified by magnetic column (Miltenyi Biotec, San Diego, CA), cultured as previously described and then assessed for purity via flow cytometry (Akkina et al., 1994; Bai et al., 2000; Schmitt et al., 2017; Schmitt et al., 2018; Curlin et al., 2020a; Curlin et al., 2020b). To create humanized hematopoietic stem cell (hu-HSC) mice, neonatal (3-4 day old) Balb/c Rag1^{-/-}γc^{-/-} or Balb/c Rag2^{-/-}γc^{-/-} mouse pups received sublethal irradiation (350 rads) in order to precondition the mice prior to CD34⁺ cell engraftment (Berges et al., 2008; Veselinovic et al., 2016; Schmitt et al., 2017). Approximately 0.5-1.0 x 10⁶ human CD34⁺ cells were then intrahepatically injected into each mouse. Ten to twelve weeks after engraftment, peripheral blood was collected via tail vein puncture and used to assess the degree of human immune cell engraftment. According to the manufacturer's instructions, red blood cells were lysed with the Whole Blood Erythrocyte Lysing Kit (R & D Systems, Minneapolis, MN). The remaining white blood cells were stained for flow cytometry with mouse anti-human CD45-FITC (eBioscience), CD3-PE (eBioscience) and CD4-PE/Cy5 (eBioscience). In order to produce humanized bone marrow, liver, thymus (BLT) mice, adult Balb/c Rag1^{-/-}γc^{-/-} or Balb/c Rag2^{-/-}γc^{-/-} mice were surgically engrafted with a combination of human fetal liver and thymic tissue under the kidney capsule. This was followed by a tail-vein injection of autologous CD34⁺ human hematopoietic stem cells from the same source as the liver and thymic tissue (Lan et al., 2006; Akkina, 2013b). Human immune cell engraftment was determined using the same method as was used for the hu-HSC mice.

8.2.2 SIV_{mac239}, SIV_{hu}, and SIV_{B670} viral stock preparation and *in vivo* infection

Using a molecular clone plasmid provided by Christine Fennessey, SIV_{mac239} was transfected into 8 flasks of 293T cells as described previously (Curlin et al., 2020b; Schmitt et al., 2020a). Briefly, 15 µg of the viral DNA and 30 µL of the Turbofect transfection reagent (Thermo Fisher Scientific, Waltham, MA) were combined with serum-free DMEM to transfect a single T150 flask. The supernatants were collected from the cells at 48 h and cell debris was removed with low speed centrifugation. This supernatant was subsequently ultracentrifuged using a L8-70M ultracentrifuge/SW 28 Ti rotor (Beckman Coulter, Pasadena, CA) on top of a 20% sucrose cushion layer in Ultra-Clear tubes (Beckman Coulter, Pasadena, CA). Following ultracentrifugation at 27,000 rpm for 2 hours at 4°C, the supernatant was decanted, and the viral pellet was slowly reconstituted on ice for 20 min using serum-free DMEM. A small aliquot was set aside for outgrowing the virus in cell culture and to be titration using TZM-bl cells as described below. SIV_{hu} and SIV_{B670} viral supernatants were obtained from the NIH AIDS Reagent Program and were cultured according to their recommended guidelines (Murphey-Corb et al., 1986; Baskin et al., 1988; Khabbaz et al., 1994). Groups of hu-HSC or BLT mice of the same donor background/cohort were inoculated with ~200 µL of viral supernatant via intraperitoneal (i/p) injection with one of the respective progenitor viruses (SIV_{mac239}: $1 \times 10^{5.5}$ TCID₅₀/mL, SIV_{hu}: 2.00×10^8 TCID₅₀/mL, SIV_{B670}: 5.78×10^7 TCID₅₀/mL). Due to the lower titer of SIV_{mac239}, a series of three additional i/p injections were given to the infected mice over the course of two weeks. 293T cells were maintained with DMEM media that contained 1% L-glutamine, 10% heat-inactivated fetal bovine serum, and 1% antibiotic-antimycotic mix (Thermo Fisher Scientific, Waltham, MA, United States).

8.2.3 TZM-bl based titrating of viral stocks

TZM-bl cells were used to determine the functional titer of the viruses prior to injection and their respective outgrowths (Platt et al., 1998; Derdeyn et al., 2000; Wei et al., 2002; Takeuchi et al., 2008; Platt et al., 2009; Curlin et al., 2020b). Briefly, each viral supernatant was diluted from 10^0 - 10^{-5} serially and used to infect TZM-bl cells on a 96-well plate that were at a density of approximately 10,000 cells/well. After 72 h, cells were washed with 1x PBS, and were then fixed with 1% formaldehyde/0.2% glutaraldehyde in 1x PBS for ~5 minutes. Cells were then washed 3x with 1x PBS and then stained with 4 mM potassium ferrocyanide, 4 mM potassium ferricyanide, 4 mM magnesium chloride and 0.4 mg X-gal per mL while incubated at 37°C for 2 h. Cells were then washed with 1x PBS, and the tissue culture infectious dose per mL (TCID₅₀/mL) was determined. TZM-bl cells were maintained with DMEM media that contained 1% L-glutamine, 10% heat-inactivated fetal bovine serum, and 1% antibiotic-antimycotic mix (Thermo Fisher Scientific, Waltham, MA, United States).

8.2.4 Serial passaging of SIV_{mac239}, SIV_{hu}, and SIV_{B670}

After approximately 24 weeks post-inoculation the infected mice that showed the highest plasma viral loads were euthanized, their whole blood was collected via cardiac puncture and their spleen, lymph nodes and bone marrow were collected as previously described (Schmitt et al., 2020a). These tissues were also collected from an uninfected mouse of the same immune cohort that was being used for the subsequent generation. Tissues were enzymatically digested concurrent with mechanical tissue dissociation to produce a single cell suspension. Leukocyte fractions of the cells were collected using Ficoll-Plaque density centrifugation. Cells were then plated at roughly $2-3 \times 10^6$ cells/mL and activated for roughly 48 hours using 100 ng/mL of anti-hCD3 and anti-hCD28 soluble antibody (Miltenyi Biotec Inc., Auburn, CA). The activated cells were then washed

2x, combined with the uninfected cells and co-cultured at a roughly equal ratio of infected to uninfected mouse cells for another 48 hours. Finally, these co-cultured cells were centrifuged at 1500 rpm for 10 minutes at RT and subsequently injected via i/p into the new cohort of hu-HSC or BLT mice. This process was used to initiate the second and third generations of infection. Primary cells were maintained in RPMI media with 1x antibiotic/antimycotic mix (Thermo Fisher Scientific, Waltham, MA, United States), 10% heat-inactivated fetal bovine serum, and 20 ng/mL IL-2 (R&D Systems, Inc., Minneapolis, MN, United States) as previously described (Schmitt et al., 2020a).

8.2.5 Plasma viral load and CD4⁺ T cell decline detection

During each passage, peripheral blood was collected on a weekly basis to assess plasma viral loads (PVL). Plasma was collected from blood via centrifugation, and then viral RNA was extracted with an E.Z.N.A. Viral RNA kit as outlined by the manufacturer's recommendations (OMEGA bio-tek, Norcross, GA). This RNA was then used to determine PVL via qRT-PCR using the iScript One-Step RT-PCR kit with SYBR green (Bio Rad, Hercules, CA). Primers were designed for SIVmac239 based on the LTR (GenBank accession: M33262.1), while SIVhu and SIV_{B670} used primers that were originally designed for a conserved region of the LTR in SIVsmE041 (GenBank accession: HM059825.1). The primers used for PCR are as follows: 1.) SIVmac239: forward 5'- GCAGGTAAGTGCAACACAAA-3' and reverse 5'- CCTGACAAGACGGAGTTTCT-3', T_m = 54°C 2.) SIV_{B670} and SIVhu: forward 5'- CCACAAAGGGGATGTTATGGGG -3' and reverse 5'- AACCTCCCAGGGCTCAATCT-3', T_m = 60°C. These primers were used with the following cycling reactions for qRT-PCR quantification: 50°C for 10 min, 95°C for 5 min, followed by 40 cycles of 95°C for 15 sec and 54°C or 60°C for 30 sec using a Bio Rad C1000 Thermo Cycler and CFX96 Real-Time System

(Bio Rad, Hercules, CA). The limit of detection was 1,000 copies/mL and was determined using a standard curve consisting of viral SIVmac239, and SIVsmE041 LTR 10-fold dilutions generated at a known concentration.

Peripheral blood was collected bi-monthly in order to assess human CD4⁺ T cell engraftment levels via flow cytometry. Briefly, 5 μ L of Fc γ R-block (Jackson ImmunoResearch Laboratories, Inc. West Grove, PA) was added to the blood for 5 min. Each blood sample was then stained with fluorophore conjugated hCD45-FITC, hCD3-PE and hCD4-PE/Cy5 (BD Pharmingen, San Jose, CA) for roughly 30 min. Erythrocytes were lysed with the Whole Blood Erythrocyte Lysing kit based on the manufacturer's guidelines (R&D Systems, Minneapolis, MN). Antibody-stained cells were then fixed in 1% paraformaldehyde/PBS and passed through a 0.45 μ m filter. Samples were then run on the BD Accuri C6 Flow Cytometer (BD Biosciences, San Jose, CA). CD4⁺ T cell levels were assessed as a percentage of the CD3⁺/CD45⁺ cell population. Flow cytometry data was analyzed using FlowJo v10.0.7 software package (FlowJo LLC, Ashland, OR). A two-tailed Student's *t*-test was used to determine CD4⁺ T cell decline between the uninfected and infected mouse groups.

8.2.6 Amplicon and Illumina-based deep sequencing preparation

To identify potential adaptive mutations, viral RNA from two mice per passage at multiple timepoints from across the duration of the passage, i.e. 3, 11, 19, 23 weeks post-inoculation, etc. as well as the starting stock viruses were used for sequencing. Briefly, cDNA was synthesized from the viral RNA with a SuperScript IV kit according to the manufacturer's guidelines (Invitrogen, Carlsbad, CA). Primal Scheme was used to design two multiplexed inter-overlapping primer pools that span the coding region of the viral genomes based on the SIVmac239 reference sequence (GenBank accession: M33262.1) and an independently sequenced SIVhu reference

sequence for SIV_{hu} and SIV_{B670} (<https://primal.zibraproject.org>) (Quick et al., 2017). These primer pools were then used with the viral cDNA and Q5 Hot Start High Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA) as described by Quick et al., to produce roughly 400 bp overlapping amplicons of the viral genomes (Supplementary Tables 8.1 and 8.2). Agencourt AMPure XP (Beckman Coulter Life Sciences, Indianapolis, IN) magnetic beads were used to purify the amplicons, which were further prepped for Illumina-based deep sequencing with the TruSeq Nano DNA HT Library Preparation Kit according to the manufacturer's guidelines (Illumina, San Diego, CA). Sequencing of the amplicon library was performed on a MiSeq Illumina desktop sequencer (Invitrogen, Carlsbad, CA) at the core sequencing facility of University of Wisconsin, Madison.

8.2.7 Calculation of nonsynonymous single nucleotide polymorphism (SNP) frequencies

The deep sequencing reads were prepared for analysis by filtering out low quality reads, and by trimming the ~30 bp primer sequences and the adapter sequencing off of the reads with cutadapt software v1.9.1 (Martin, 2011). Filtered reads were aligned to the stock virus sequence with bowtie2 software v2.2.5 (Langmead and Salzberg, 2012). The resulting BAM format output was used as the input to call variants with lofreq software v2.1.2 (Wilm et al., 2012). Variants identified here had >100 read depth and at least 1% frequency. Genome plots were generated using R and ggplot2 (ISBN: 0387981403). R scripts used to graph the plots can be found at https://github.com/stenglein-lab/viral_variant_explorer.

8.3 Results

8.3.1 SIVmac239 is capable of productive and chronic infection in humanized mice and produces HIV-like CD4⁺ T cell depletion

In order to determine if SIVmac239 can establish productive infection, cohorts of five hu-HSC mice were inoculated with SIVmac239 and serially passaged for a total of three generations. Within one week after inoculation, RNA copies were detected in the plasma using qRT-PCR in all five hu-HSC mice (Figure 8.1A). During the first generation, the plasma viral loads continued to climb gradually until approximately 64 days post-inoculation and peaked at over 3.4×10^4 RNA copies/mL. This viremia was maintained for the majority of the generation, with a slight decline beginning around 120 days post-inoculation. This first generation proved successful chronic infection of human immune cells in hu-HSC mice by SIVmac239. After the virus was serially passaged into a second generation of hu-mice, the viral loads displayed a remarkably similar pattern to the first generation. Positive viral loads were detected within one week and continued to climb, with an earlier peak in viremia (2.4×10^4 RNA copies/mL) relative to the first generation at 21 days post-inoculation. This was followed by a slight decline for the next several timepoints, and then the viral loads remained consistently just under 1×10^4 RNA copies/mL through 120 days post-inoculation. The third generation plasma viral loads were noticeably lower than the first generation from 28 to roughly 63 days post-inoculation, after which the viral loads rose to around 1×10^4 RNA copies/mL and remained somewhat level to the end of the generation. Interestingly, several individual mice in the third generation displayed drastically higher peak plasma viral loads than the others, reaching 1×10^6 RNA copies/mL within 21 days post-inoculation, suggesting that the virus had fundamentally adapted to human immune cells in these hu-mice. Unfortunately, these

mice also died approximately 30 days after inoculation most likely due to their high viral loads. As expected, no viral loads were detected in the control uninfected mice (Figure 8.1A).

In the first generation, SIVmac239 displayed a slight decline in CD4⁺ T cell levels within the first 100 days post-inoculation, which was followed by a gradual increase near the end of the generation (Figure 8.2A). This resulted in no real net decrease of CD4⁺ T cell levels. In contrast, the second generation showed an immediate decline in CD4⁺ T cell levels through 70 days post-inoculation, followed by a small recovery, producing a total decline in CD4⁺ T cells of almost 20%. Finally, in the third generation, the infected mice displayed a rapid decline beginning around 28 days post-inoculation as a result of the rapid increase in plasma viral loads, with a relatively large rebound around 84 days post-inoculation, followed by another drastic decline through the end of the generation. By the end of the third generation, there was a statistically significant difference in the CD4⁺ T cell decline when compared to the uninfected control hu mice (two-tailed, Student's *t*-test, $p < 0.01$) demonstrating that serially passaging SIVmac239 results in CD4⁺ T cell depletion in the humanized mouse model.

8.3.2 SIV_{B670} is capable of productive and chronic infection in humanized mice and produces HIV-like CD4⁺ T cell depletion

To determine if SIV_{B670} can establish productive infection in humanized mice, hu-mice were inoculated with SIV_{B670} and serially passaged for a total of 3 generations. SIV_{B670} displayed a pattern of infection somewhat similar to SIVmac239. Following the initial inoculation, the virus was detected within the first week at over 1×10^4 RNA copies/mL (Figure 8.1B). The viral loads showed a very slight increase peaking around 8×10^4 RNA copies/mL at 35 days post-inoculation, at which point the viral loads showed a steady but persistent decline to just over 1×10^3 RNA copies/mL until the end of the first generation. This demonstrated that SIV_{B670} could produce

productive and chronic infection in human immune cells *in vivo*. In contrast, the second generation was not detected above the limit of detection until approximately 21 days post-inoculation and fluctuated rapidly with sporadic dips in viral loads occurring throughout the rest of the second generation. Viral loads peaked lower than the first generation at just over 3×10^4 RNA copies/mL around 49 days post-inoculation. This was followed by a sporadic downward trend and an eventual leveling off near the end of the generation. The third generation was similar to the second, as the plasma viral loads were initially over a log lower than those observed in the first generation. However, the third generation showed drastically less fluctuation than the second generation and showed a gradual increase to peak at 3.64×10^4 RNA copies/mL by 63 days post-inoculation, and a general, persistent decline through the rest of the third generation. No plasma viral loads were detected in the control uninfected hu-mice.

The first generation of SIV_{B670} displayed a drastic decline in CD4⁺ T cells compared to the controls within 14 days post-inoculation, and then remained relatively steady, with a slight decline towards the end of the generation (Figure 8.2B). In contrast, the CD4⁺ T cell levels in the second generation remained initially stable through 42 days post-inoculation, followed by a dip in CD4⁺ T cell levels, before recovering around day 126 post-inoculation. By the third generation, SIV_{B670} displayed a very consistent and gradual decline in CD4⁺ T cell levels until 56 days post-inoculation, followed by rapid CD4⁺ T cell decline that continued until leveling out around 112 days post-inoculation, at which point the CD4⁺ T cell engraftment was around 10% of the CD45⁺/CD3⁺ population. From 56 days post-inoculation through the end of the third generation, the CD4⁺ T cell engraftment had significantly declined relative to the uninfected controls (two-tailed, Student's *t*-test, $p < 0.05$). This data showed that SIV_{B670} could result in CD4⁺ T cell depletion in the humanized mouse model by the third generation.

8.3.3 SIVhu is capable of productive and chronic infection in humanized mice but does not produce HIV-like CD4⁺ T cell depletion

Like the other two viruses, SIVhu was serially passaged over a total of three generations in humanized mice in order to determine if SIVhu was capable of productive chronic infection. Within the first generation of infection, SIVhu was initially detectable in the plasma viral loads by the first week at around 10^4 RNA copies/mL and remained within a log of that before peaking at 6×10^4 RNA copies/mL around 63 days post-inoculation (Figure 8.1C). After this peak, the viral loads rapidly dropped to around the limit of detection (1×10^3 RNA copies/mL) until the end of the generation. This demonstrated that SIVhu could readily infect human immune cells leading to productive viremia. After serial passaging, the virus struggled to replicate, and did not become detectable in the second generation until around 49 days post-inoculation. From there, the PVL continued to rise until roughly 91 days post-inoculation around 9×10^3 RNA copies/mL and then followed with a slight, sporadic decline through the end of the generation. The third generation was a continuation of the trend seen in the second generation, with the virus remaining undetected until around 63 days post-inoculation before peaking shortly after at 1.5×10^5 RNA copies/mL. This was followed by a sharp decline to the limit of detection by 105 days post-inoculation (1×10^3 RNA copies/mL) and a gradual recovery in viral loads towards the end of the generation. Human CD4⁺ T cell engraftment levels, as assessed by flow cytometry showed that there was little to no significant CD4⁺ T cell decline across any of the three generations of serial passaging among mice infected with SIVhu. In fact, the CD4⁺ T cell levels remained relatively consistent across the entire duration of each generation (Figure 8.2C).

8.3.4 Evolutionary genetic adaptations arise during sequential passaging of SIVmac239, SIV_{B670} and SIVhu

In addition to the *in vivo* pathogenicity of these viruses, adaptation was assessed by identifying nonsynonymous mutations that rose in frequency over time that could change the phenotype of the virus. Therefore, we used Illumina-based deep sequencing on plasma viral RNA from two mice (replicates) per generation collected at early, mid, and late timepoints across three serial passages. Paired end reads of the viruses were mapped to the consensus sequence of the starting viral stock. While there were numerous mutations that arose briefly and disappeared, the following criteria were used to identify truly significant variants: 1.) Mutations must be nonsynonymous and present in the coding sequence (CDS), 2.) The mutations must be present in at least four of the data sets which means each variant is present beyond a single replicate in a single generation, 3.) Mutations must increase in viral frequency over the subsequent generations, and 4.) The frequency of the mutation in the viral population must average at least 50% at the last sequenced timepoint. Eight mutations were found in SIVmac239, 32 mutations were found in SIV_{B670} and 11 mutations were found in SIVhu that matched these criteria (Figure 8.3, Tables 8.1-8.3).

A number of nonsynonymous mutations were identified in SIVmac239 after three generations of passaging across a number of genes, with the largest concentration occurring in *env* and *nef* (Figure 8.3A, Table 8.1). Of these mutations, Env D676E and Nef D258N had individual replicate samples in the third generation that had persistently low frequencies, while all the other identified mutations were near 100% population frequency in both replicates. SIV_{B670} had by far the largest number of qualifying mutations, with a large number occurring in *gag*, *pol*, *env*, and *nef* (Figure 8.3B, Table 8.2). Interestingly, though a number of these mutations did increase to a

frequency of nearly 100%, quite a few of the mutations were still at lower frequencies, albeit they had still increased in frequency relative to the stock virus. SIV_{hu} had fewer qualifying mutations than SIV_{B670}, with a relatively even distribution that was slightly weighted towards the 5' end of the genome with *env* and *nef* having the most mutations (Figure 8.3C, Table 8.3). Interestingly, all of the mutations found in *env* had 100% frequency of the viral population of both replicates of the third generation. One mutation in particular at residue Gag 216, was present in all three viral strains, though the residue change was L→I in SIV_{hu} and L→V in SIV_{B670} and SIV_{mac}, which may suggest some shared function in adaptation.

8.4 Discussion

With these experiments, we sought to understand the genetic and phenotypic changes that would arise after serially passaging SIV_{hu}, SIV_{B670} and SIV_{mac239} in humanized mice and determine if any commonalities in the adaptation of these viruses could emerge given their shared ancestry. To do this, viral stocks of SIV_{mac239}, SIV_{hu}, and SIV_{B670} were inoculated into hu-HSC and hu-BLT mice and allowed to propagate over the course of approximately 24 weeks. The viruses were then isolated from the infected mouse tissues and used to infect a new immune cohort of humanized mice. This simulated the accumulation of adaptive mutations that may have given rise to HIV-2 from SIV_{sm}, as has been previously modeled in humanized mice with other primate lentiviruses (Schmitt et al., 2017; Schmitt et al., 2018; Curlin et al., 2020a; Schmitt et al., 2020a).

In comparison to SIV_{hu}, SIV_{mac239} and SIV_{B670} had more consistently positive viral loads across all three passages. While SIV_{B670} did not show much change in viral loads between generations, one hu-mouse in the third passage of SIV_{mac239} infection showed very robust plasma viral loads at least a log higher than previously seen (Figure 8.1A), though this mice died within 30 days after inoculation with the virus, possibly due to the high viral loads and CD4⁺ T

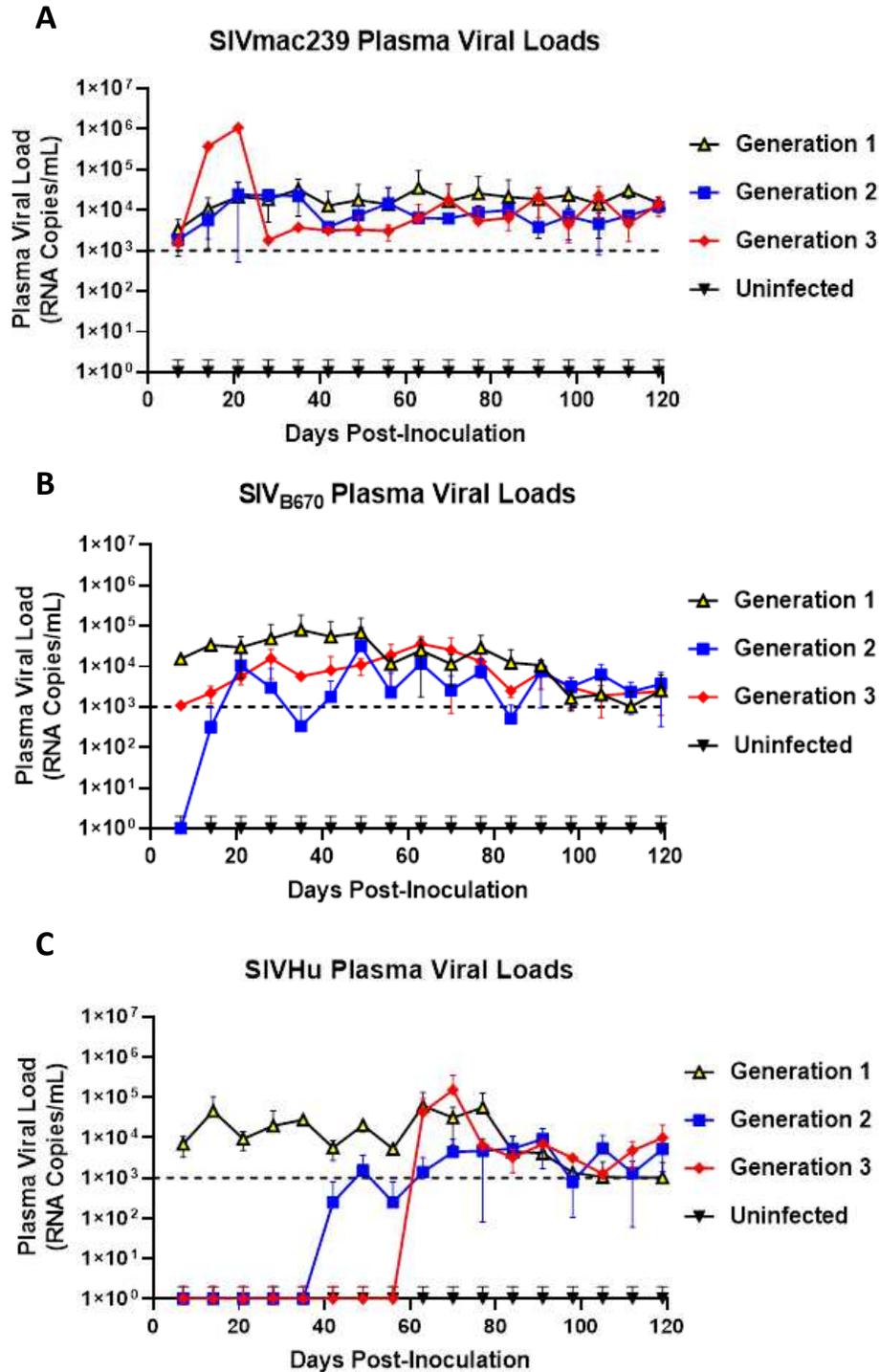


Figure 8.1 SIVmac239, SIV_{B670}, and SIVhu are all capable of producing a sustained and transferrable viral infection in humanized mice. (A) SIVmac239, (B) SIV_{B670}, and (C) SIVhu plasma viral loads during three generations of serial passaging. qRT-PCR was used to assess the plasma viral loads on a weekly basis for the duration of each generation. No positive viral loads were identified in the uninfected control hu-mice. The limit of detection was 1×10^3 RNA copies/mL.

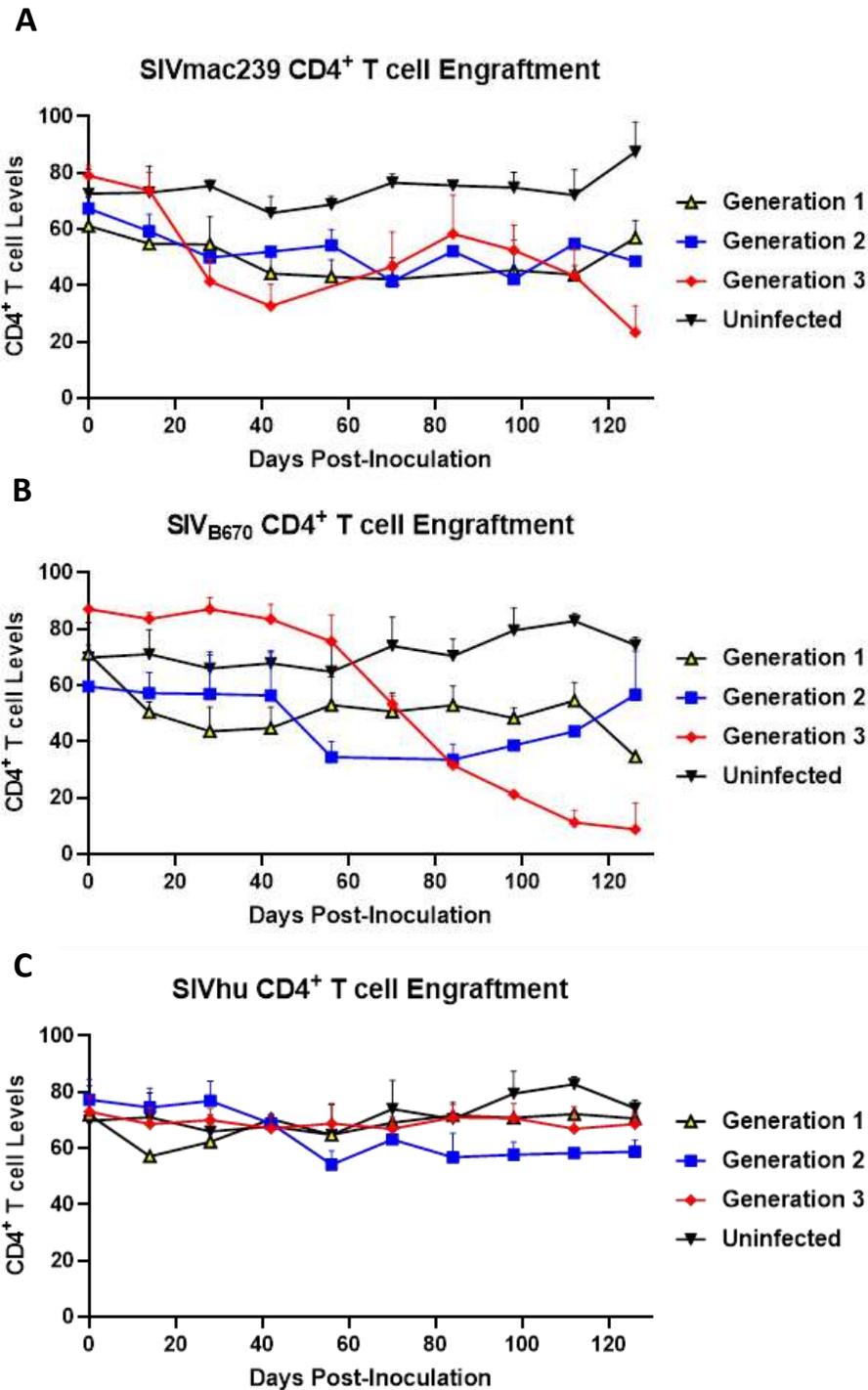
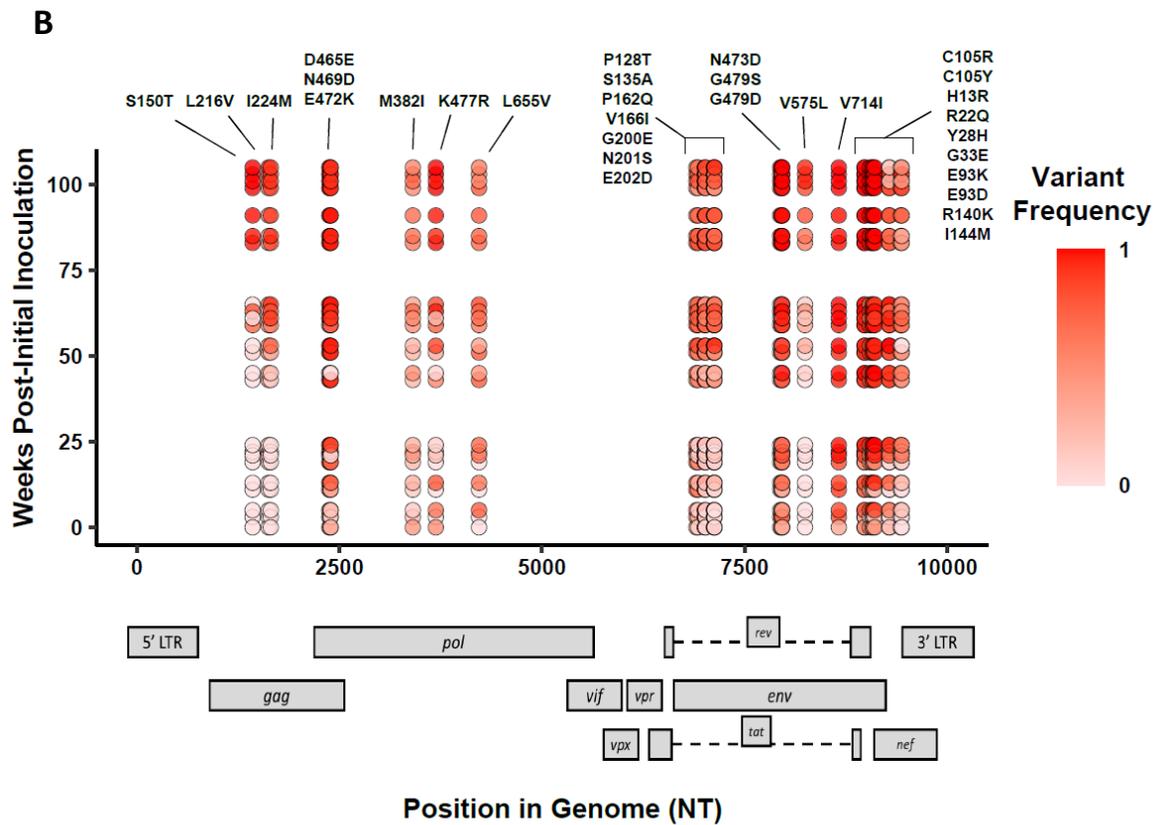
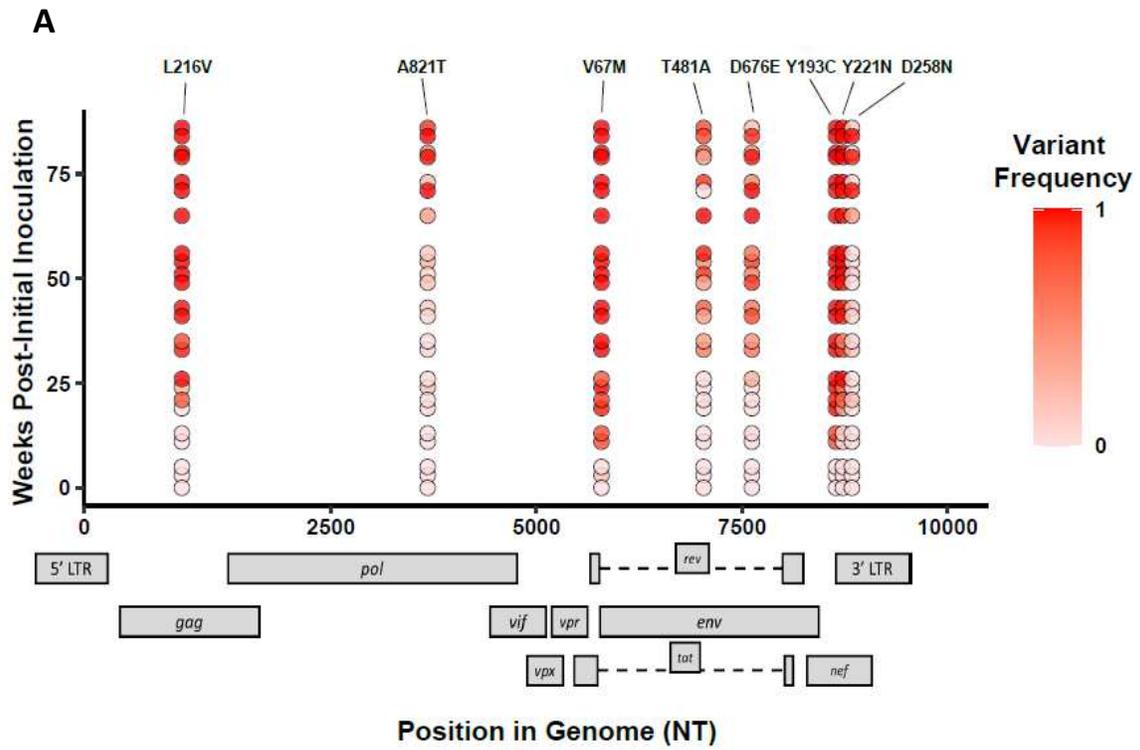


Figure 8.2 CD4⁺ T cell levels across three generations of serial passaging SIVmac239, SIV_{B670}, and SIVhu in humanized mice. (A) SIVmac239, (B) SIV_{B670}, and (C) SIVhu CD4⁺ T cell levels of the course of 3 generations. CD4⁺ T cell decline was monitored bimonthly using flow cytometry. The percent of circulating CD4⁺ T cell levels is relative to the percentage of CD45⁺/CD3⁺ cell populations.



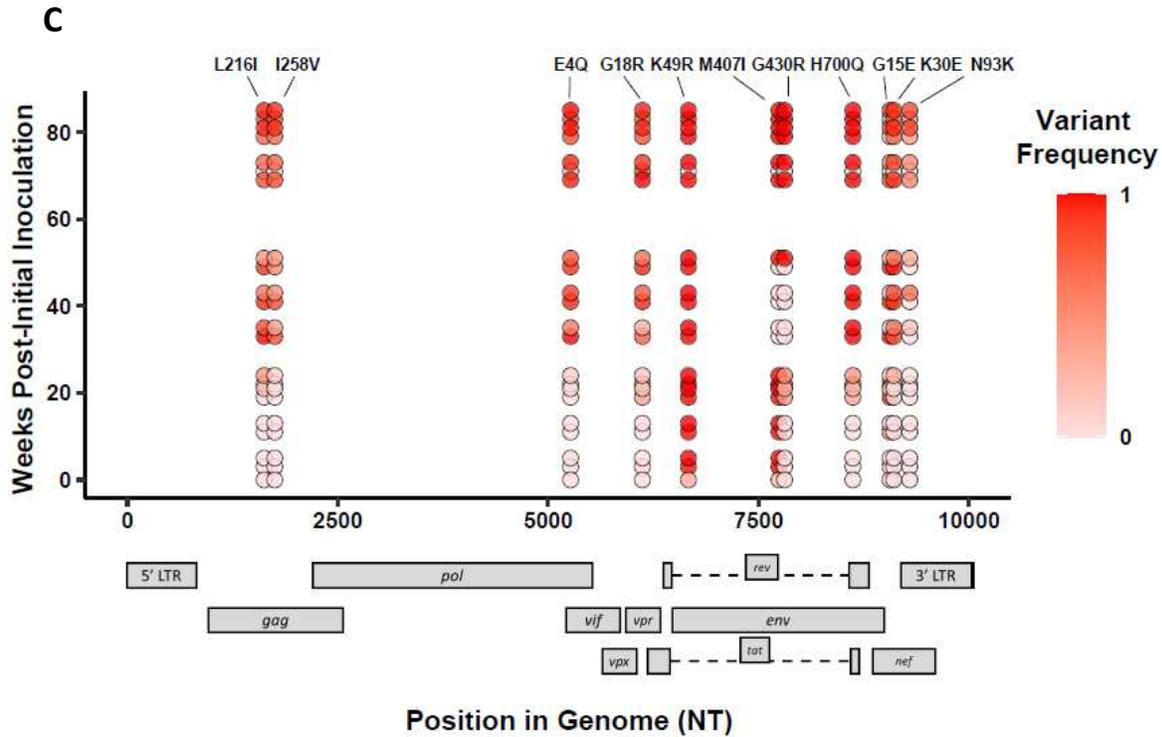


Figure 8.3 Single nucleotide polymorphisms that increased in population frequency over three serial passages. Viral RNA from two hu-mice per strain per generation at early, middle, and late timepoints were sequenced. Mutations identified were nonsynonymous, present in the CDS, had an average endpoint frequency of at least 50%, and were found in at least four timepoints. (A) SIVmac239, (B) SIV_{B670}, (C) SIVhu variants are plotted based on their position in the genome along the x-axis. Their y-axis location is based on the overall time of collection after the original infection. Variant population frequency is denoted by the shading color scale for each point. Week 0 corresponds to the frequency of the viral stock pre-inoculation, with the second and third passages being plotted beginning at week 30 and 60, respectively. For each time point, replicates are offset vertically from each other. Both the residue changes and number for each position are listed above their respective locations.

Table 8.1 Amino acid substitutions arising from three passages of SIVmac239.

Protein	Position	Stock Residue^a	Variant Residue^b	Stock Frequency^c	P3 Endpoint Frequency^d	P3 Endpoint Frequency^e
Gag	216	L	V	0.00 ^f	1.00	1.00
Pol	821	A	T	0.00 ^f	1.00	0.94
Env	67	V	M	0.00 ^f	1.00	1.00
Env	481	T	A	0.00 ^f	0.85	0.84
Env	676	D	E	0.00 ^f	0.95	0.23
Nef*	193	Y	C	0.00 ^f	1.00	1.00
Nef*	221	Y	N	0.00 ^f	1.00	1.00
Nef	258	D	N	0.00 ^f	1.00	0.15

^aConsensus amino acid residue from sequenced stock virus.

^bVariant amino acid residue from passaged virus.

^cFrequency of variant residue in the stock virus population.

^dReplicate 1 frequency of variant residue at the end of the third passage.

^eReplicate 2 frequency of variant residue at the end of the third passage.

^fThe 0.00 indicates that the variant frequency was below the limit of detection of the variant identification pipeline.

*Variant residue is present at the same location in the other stock virus sequences.

Table 8.2 Amino acid substitutions arising from three passages of SIV_{B670}

Protein	Position	Stock Residue ^a	Variant Residue ^b	Stock Frequency ^c	P3 Endpoint Frequency ^d	P3 Endpoint Frequency ^e
Gag	150	S	T	0.00 ^f	1.00	1.00
Gag	216	L	V	0.00 ^f	0.70	0.96
Gag	224	I	M	0.00 ^f	0.88	0.91
Gag	465	D	E	0.11	0.99	0.99
Gag*	469	N	D	0.34	1.00	1.00
Gag	472	E	K	0.30	0.95	0.91
Pol*	382	M	I	0.43	0.80	0.54
Pol	477	K	R	0.48	1.00	1.00
Pol	655	L	V	0.00 ^f	0.55	0.59
Env	128	P	T	0.37	0.74	0.64
Env	135	S	A	0.09	0.72	0.65
Env*	162	P	Q	0.44	1.00	1.00
Env*	166	V	I	0.00 ^f	0.71	0.82
Env*	200	G	E	0.49	0.69	0.91
Env*	201	N	S	0.49	0.78	0.94
Env	202	E	D	0.00 ^f	0.66	0.91
Env	473	N	D	0.29	0.50	1.00
Env	479	G	S	0.37	0.50	1.00
Env	479	G	D	0.37	0.50	1.00
Env*	481	R	Q	0.30	1.00	1.00
Env	575	V	L	0.00 ^f	0.89	0.95
Env*	714	V	I	0.34	1.00	1.00
Rev	105	C	R	0.50	1.00	0.99
Rev	105	C	Y	0.49	1.00	1.00
Nef*	13	H	R	0.42	1.00	1.00
Nef*	22	R	Q	0.46	1.00	1.00
Nef	28	Y	H	0.39	1.00	1.00
Nef*	33	G	E	0.43	1.00	1.00
Nef	93	E	K	0.24	0.93	0.15
Nef*	93	E	D	0.20	0.91	0.14
Nef	140	R	K	0.00	0.93	0.09
Nef	144	I	M	0.00 ^f	0.87	0.43

^aConsensus amino acid residue from sequenced stock virus.

^bVariant amino acid residue from passaged virus.

^cFrequency of variant residue in the stock virus population.

^dReplicate 1 frequency of variant residue at the end of the third passage.

^eReplicate 2 frequency of variant residue at the end of the third passage.

^fThe 0.00 indicates that the variant frequency was below the limit of detection of the variant identification pipeline.

*Variant residue is present at the same location in the other stock virus sequences.

Table 8.3 Amino acid substitutions arising from three passages of SIVhu

Protein	Position	Stock Residue^a	Variant Residue^b	Stock Frequency^c	P3 Endpoint Frequency^d	P3 Endpoint Frequency^e
Gag	216	L	I	0.00 ^f	0.77	0.96
Gag	258	I	V	0.00 ^f	0.79	0.95
Vif	4	E	Q	0.00 ^f	0.85	0.99
Vpr	18	G	R	0.00 ^f	0.82	0.97
Env*	49	K	R	0.32	1.00	1.00
Env	407	M	I	0.25	1.00	1.00
Env	430	G	R	0.00 ^f	1.00	1.00
Env*	700	H	Q	0.00 ^f	1.00	1.00
Nef	15	G	E	0.00 ^f	0.31	0.90
Nef	30	K	E	0.00 ^f	0.68	0.94
Nef	93	N	K	0.00 ^f	0.51	0.85

^aConsensus amino acid residue from sequenced stock virus.

^bVariant amino acid residue from passaged virus.

^cFrequency of variant residue in the stock virus population.

^dReplicate 1 frequency of variant residue at the end of the third passage.

^eReplicate 2 frequency of variant residue at the end of the third passage.

^fThe 0.00 indicates that the variant frequency was below the limit of detection of the variant identification pipeline.

*Variant residue is present at the same location in the other stock virus sequences.

cell decline. The CD4⁺ T cell levels were also consistent with the plasma viral loads, in that SIVmac239-infected mice showed significant ($p < 0.01$) decline compared to uninfected mice. However, certain differences in CD4⁺ T cell decline between generations was slight. SIV_{B670} displayed only a slight decline in the first two generations, with the third generation showing a dramatic decline in human cell engraftment by the end of the generation. Overall, both SIVmac239 and SIV_{B670} had a drastically larger CD4⁺ T cell decline when compared to SIVhu.

In addition to the data derived from the plasma viral loads and CD4⁺ T cell engraftment, the various nonsynonymous mutations that have arisen after three serial passages provide some interesting avenues to explore how these viruses may be adapting. Relative to SIVhu and SIVmac239, SIV_{B670} overwhelmingly had the largest number of qualifying nonsynonymous mutations throughout the viral genome (Table 8.3). Similarly to what has been found in previous studies, the bulk of the nonsynonymous mutations were found in the 3' end of the viral genome (Schmitt et al., 2017; Schmitt et al., 2020a). Several of these mutations identified actually cause an amino acid change that brings the residue sequence more in line with the other starting stock viral proteins sequences. For example, the residue Gag N469D in SIV_{B670} is already an aspartic acid (D) at this position in the stock SIVmac239 and the SIVsmE041 sequences. However, one of the most striking features amongst the three passaged viruses was the mutation at residue 216 in Gag that appeared within all three different strains of viruses in a highly conserved part of the genome (Foley, 2018; Sauter and Kirchhoff, 2019). This residue is part of the gene that encodes the capsid protein and is located within the cyclophilin binding loop. Furthermore, it is possible that this residue is involved in TRIM5 α antagonism and escape, which is an important barrier to overcome for successful cross-species transmission (Bukovsky et al., 1997; Wu et al., 2013). The dramatic increase in viral population frequency of this mutation further supports the idea that this

gene is conferring improved fitness on these viruses (Sanjuan et al., 2004; Cuevas et al., 2012; Acevedo et al., 2014).

While functional studies are still necessary to fully understand the exact nature of these mutations, known motifs and binding interactions can potentially shed some light of the role of these changes. For example, the amino acid residue changes at Gag 465 and 469 in SIV_{B670} are located within a dileucine-motif that has been shown to be responsible for the particle association and uptake of Vpx and Vpr in SIVmac (Pancio and Ratner, 1998; Accola et al., 1999). This interaction between the DXAXXLL motif and Vpr allows Vpr to be diverted away from the proteasome-degradative pathway, while deletion of this particular motif was shown to result in a total lack of incorporation of Vpx in SIVmac (Accola et al., 1999). The fact that this motif is fundamentally changed by the D465E mutation with no marked decline in pathogenicity for SIV_{B670} is interesting. It could be due to the fact that the LXXLF motif at the C terminus of the p6^{gag} region remained intact, and that they share some overlap in functionality (Accola et al., 1999). Just as compelling is that the Gag N469D mutation brings SIV_{B670} closer to amino acid sequences found in both SIVsmE041 and SIVmac239, which were already aspartic acids at this particular location (Foley, 2018). Very few qualifying mutations were identified in *vif*, *vpr*, and *vpx* in all three viruses studied. However, the SIVhu Vpr G18R residue change is immediately adjacent to a previously described DCAF-1 binding motif canonically found in both Vpr and Vpx, and has been implicated in G2 arrest and SAMHD1 degradation, respectively (Wei et al., 2012).

In contrast, both *env* and *nef* acquired a large number of mutations in all three viruses that increased in frequency over time and may ultimately have functional implications regarding virus-host restriction factor interactions. The Env V67M mutation identified in SIVmac239 not only increased from 0% frequency in the starting stock virus to 100% frequency by the end of the third

generation, but was previously shown to increase infectivity of the virus (Sato et al., 2008). The mechanism of this increased infectivity was not tied to neutralizing antibody escape, however. Additionally, this mutation has been implicated in conferring macrophage tropism (Mori et al., 1992). Amino acid residue changes in the Env P128T and S135A of SIV_{B670} occur within the V1 hypervariable loop region, which has been implicated in neutralizing antibody susceptibility (Johnson et al., 2002; Saunders et al., 2005). Prolines have a high degree of conformational rigidity, and are less commonly involved in binding sites, so its substitution for a Threonine, which is significantly more reactive, could suggest a shift in functionality for this region of the protein (Betts, 2003). Other mutations such as Env G430R in SIV_{hu} occurs in the V4 loop and could potentially affect neutralizing antibody responses. Additional mutations that were identified, but did not meet the previously described criteria, such as Env I421K, R425G, and Q428R all occurred in the V4 loop of SIV_{B670} and could be involved in neutralizing antibody responses, with R425G at approximately the equivalent P421 residue in SIV_{mac239}, that has been directly implicated in neutralizing antibody escape and increased infectivity when mutated into P421Q (Sato et al., 2008). Env N201S, and R481Q in SIV_{B670} and T481A in SIV_{mac239} occur within N-linked glycosylation motifs of the V2 and V5 loops. The effect of N-linked glycosylation motifs in the envelope proteins of HIV and SIV on antibody neutralization have been well-characterized. These motifs allow a glycan shield to form, which protects exposed epitopes from being bound by neutralizing antibodies (Chackerian et al., 1997; Wei et al., 2003). The Rev-response element (RRE) is a critical binding motif necessary for the shuttling of transcripts from the nucleus to the cytoplasm. It is possible that Env V575L in SIV_{B670}, which falls within the RRE could be advantageously affecting this interaction, given that this mutation increased in frequency to make up around 90% of the viral population by the third generation.

Despite the smaller size of *nef* relative to *env*, *nef* has a disproportionate frequency of nonsynonymous mutations. However, given that APOBEC3 deaminase activity begins on the 5' end of the viral genome, it is possible that a restriction factor like APOBEC3 could play a role in the abundance of nonsynonymous mutations occurring in *nef*, though this requires further analysis (Chelico et al., 2006). Similarly, to mutations seen in other genes, many of the substitutions in *nef* occur near previously characterized binding and interaction sites. For example, the Nef residue 93 changes that occur in both SIVhu and SIV_{B670} occur within a known PACS-1 binding site (Heusinger et al., 2018). The Nef Y193C substitution identified in SIVmac239 has a low evolutionary probability (Supplementary Figure 8.1) (Dayhoff, 1972). This residue is found within a ExxxL(L/M) (di-leucine) motif in the $\alpha 4$ region of Nef, that has been implicated as an AP interaction site and is involved in downregulating CD3 and CD4 (Manrique et al., 2017). Additionally, this exact residue has also been implicated in tetherin antagonism (Serra-Moreno et al., 2013). At the equivalent location in the stock sequences of SIVsmE041 and SIV_{B670} this residue is already a cysteine. In HIV-2 and SIVmac, it has been shown that the di-leucine motif is flexible enough to form an alpha helix by binding to a hydrophobic crevice between $\alpha 2$ and $\alpha 3$ with Glu 190 and Leu 194. Changing amino acid residue 193 to a cysteine could have a drastic impact on the ability of the di-leucine motif to bind to the hydrophobic crevice. Furthermore, Nef I144M from SIV_{B670} is located within the $\alpha 3$ region of Nef and is adjacent to a number of residues that have been shown to interact with the amino acid residues of the di-leucine sorting motif (Hirao et al., 2020). However, this substitution is more biochemically favored than the others and may have less of an impact on binding for this region. The Nef Y221N substitution in SIVmac is located at the end of $\alpha 5$ and is also highly disfavored biochemically. It is also adjacent to residue Y223 and Y226, which are involved in MHC-I downregulation and be functionally disrupted through

substitutions (Swigut et al., 2000). Nef D258N from SIV_{mac} is located after $\alpha 9$ near the end of Nef and likely has minimal effect on the function of Nef. In SIV_{B670} residue Nef Y28H occurs at a conserved N-proximal Tyrosine residue that has been found to be part of an adapter protein (AP) binding site (Heusinger et al., 2018). Nef K30E in SIV_{hu} also occurs at this same AP binding site.

Despite being the only strain of SIV isolated directly from a human patient, SIV_{hu} appeared to be the slowest to adapt to human cells the most among the three tested viruses. While the first generation showed viral loads comparable to the other viruses, the distinct decline in PVL and prolonged delay in detectable virus suggests that SIV_{hu} is still not fully adapted to human immune cells. Furthermore, the lack of significant CD4⁺ T cell decline indicates that SIV_{hu} is not yet capable of producing AIDS-like symptoms in humanized mice after three passages. However, the increased peaks and general upward trend of the third generation of plasma viral loads suggest that SIV_{hu} is still adapting, albeit slowly. The noted difficulty that SIV_{hu} displays could be at least partially attributed to the documented frameshift deletion in *nef* that encodes a truncated protein (Khabbaz et al., 1994). Nef has been shown to be involved in downregulating the CD4 and MHC class I molecules and helps to mitigate cell-mediated cytotoxicity (Aiken et al., 1994; Rhee and Marsh, 1994; Chaudhuri et al., 2007; Veillette et al., 2014). Additionally, Nef is involved in counteracting the action of host restriction factors such as tetherin and SERINC3/5, and is therefore critical in the proper functioning of the virus (Jia et al., 2009; Zhang et al., 2009; Usami et al., 2015). No mutations meeting our selection criteria were present around the frameshift or premature truncation, which indicates that after three generations of serial passaging, the *nef* gene had not reverted, which is supported by the consistently low plasma viral loads.

Overall, it is clear that despite a common origin in SIV_{sm}, each of these viruses is undergoing numerous unique changes over the course of three generations in an *in vivo* human

immune environment. Certain individual nonsynonymous changes have arisen, which hint at a shared functionality, though these changes are rare. It is possible that further direct similarities have not been seen due to the random nature of the mutations, and changes at different residues could be serving similar functional purposes despite different sequence locations. Based on the mutations seen here, it is clear that neutralizing antibody escape and host restriction factor antagonism are important features for overcoming the species-barrier.

9.1 Conclusions and Summary

The unabated epidemics of HIV-1 and HIV-2 remain a great public health challenge even after 30 years. With over forty different simian immunodeficiency viruses currently circulating in various nonhuman primate species, and the ever increasing number of nonhuman pathogens being discovered with human pandemic potential, it is clear that a better understanding of the evolution of these pathogens is needed (Gilbert et al., 2009; Keusch et al., 2009; Sharp and Hahn, 2011; Bell and Bedford, 2017).

The overarching goal of the research described here was to utilize a human surrogate mouse model to recreate the immune environment and evolutionary pressures of a human host and recapitulate the genetic changes that were required for HIV to emerge from its progenitor SIVs. We have shown here that multiple strains of both SIVcpz and SIVsm lineages can effectively infect and replicate to produce viremia in humanized mice. This proves that despite the differences in immune environments between humans and NHPs, SIVs are capable of direct infection of human immune cells supporting the idea that HIV-1 and HIV-2 emerged as a direct result of SIVcpz and SIVsm crossing into humans. Predictably, the ability to infect hu-mice with certain viral strains was tied to their unique genetic quirks. For example, SIVhu was previously shown to have a frameshift deletion that resulted in a truncated Nef protein due to a premature stop codon (Khabbaz et al., 1994). Given the well described importance of Nef for maintaining high viral loads, in downregulating CD4 and MHC, and counteracting multiple host restriction factors, this lack of Nef resulted in predictably lower viral loads and greater difficulty in maintaining a persistent

infection relative to other viral strains such as SIVcpzEK505, SIVmac239 and SIV_{B670} that originated in nonhuman primates, in spite of SIVhu being directly isolated from a human (Garcia and Miller, 1991; Kestler et al., 1991; Chowers et al., 1994; Jia et al., 2009; Zhang et al., 2009; Rosa et al., 2015).

Additionally, we have shown that these viruses are able to be serially passaged across different hu-mice cohorts, with changing pathogenicity over time. In spite of their pathogenicity in other species, we saw only minor differences in plasma viral loads over three generations for SIVmac239 and SIV_{B670}. In contrast, large increases in plasma viral loads were seen in multiple strains of SIVcpz within two generations of serial passaging. This could be partially explained by the differences in the natural hosts of the viruses. Rhesus macaques and sooty mangabeys are significantly more diverged from a genomic standpoint to humans when compared to chimpanzees (Chatterjee et al., 2009). Ultimately, it is clear that multiple serial passages are necessary to recapitulate the virulence of HIV, as the differences between any two sequential generations are slight. When viewed cumulatively, such as when comparing the first generation of SIVsm with the seventh generation, the increase in plasma viral loads as the viruses adapt further becomes clear.

The adaptation of these SIV strains was not measured solely by plasma viral loads. In clinical cases of HIV infection, the natural pathology is the eventual decline of CD4⁺ T cells when left untreated (Rowland-Jones, 2003). The nature of the humanized mouse model allows this pathology to be effectively reproduced, and served as an important tool for our assessment of viral adaptation (Berges et al., 2006). We found that CD4⁺ T cell decline was a common consequence of infection with different SIV strains, even those that have not been proven to cause AIDS-like symptoms in their natural hosts, such as SIVsm (Schmitt et al., 2017; Schmitt et al., 2018; Curlin et al., 2020a; Curlin et al., 2020b; Schmitt et al., 2020a). As with the plasma viral loads, this

pathology in many cases became more pronounced upon further serial passaging. The one exception to this pattern was SIVhu, which did not display any significant CD4⁺ T cell decline even after three generations of serial passaging. It has been shown that the Nef-mediated downmodulation of CD3 is related to high CD4⁺ T-cell levels in cases of HIV-2 infection (Khalid et al., 2012; Markle et al., 2013). Therefore, it is possible that the absence of CD4⁺ T cell decline could be attributed to how SIVhu Nef interacts with CD3, because the CD3-mediated activation and death of infected T cells is blocked by SIV and HIV-2 Nef (Munch et al., 2005; Schindler et al., 2006; Khalid et al., 2012). However, given that the number of serial passages needed for CD4⁺ T cell decline to occur also varied between viral strains, it is quite possible that given enough time and with enough serial passages, SIVhu could eventually display CD4⁺ T cell decline over time as well.

Perhaps most importantly, we have identified a number of mutations that arose within the viral genomes and were maintained across multiple serial passages. It has been suggested that lower frequencies of minor alleles in nonsynonymous changes is indicative of their deleterious nature, and conversely alleles that increase in frequency may indicate positive selection (Cargill et al., 1999). These mutations are not limited to any one gene in the virus, but are found throughout the viral genome, with some of the mutations arising on multiple occasions. The genetic changes we have observed do not necessarily result in a 1:1 exact recreation of HIV-1 or HIV-2 from SIVcpz or SIVsm, respectively. Rather, the resulting virus bears a degree of similarity to the existing human viruses but remains a distinct product of the random mutations that have accumulated over the sequential passages. One of the difficulties of defining if an SIV has evolved into an HIV is the very diverse nature of HIVs to begin with. In fact, the independent emergence of nine groups of HIV-2 suggest that there are multiple avenues to reach the same endpoint

(Ayoub et al., 2013). Due to the lentiviral reverse transcriptase being very error-prone, there is more diversity amongst strains of HIV-2 than there is between individual strains HIV-2 and their progenitor SIVsm strains (Visseaux et al., 2016; Foley, 2018). Couple this with the random nature of genetic mutation to begin with, and evolutionary pressures can only get so far in producing the identical changes that are needed for an SIV to become an HIV. However, it is possible to broadly identify genomic changes that have facilitated a greater affinity for human cells in the tested viruses, and could be involved in functional changes ranging from improved CD4⁺ binding affinity to an increase in viral entry, to changes in accessory proteins that may help counteract host restriction factors. Ultimately, by identifying the mutations that facilitated successful replication of the virus in human immune cells, we hope to provide a robust group of new targets for therapeutics to be designed.

Given the distinct differences in their lineages, it is fair to treat the evolution of SIVsm into HIV-2 and SIVcpz into HIV-1 independently of each other. While both groups of viruses must overcome the same environmental challenges in crossing into human immune cells, their points of origin, namely sooty mangabeys and chimpanzees, are very distinct, and have likely shaped the viruses in a subtle manner that affects their reaction to human cells. This also makes specific amino acid comparisons difficult as protein lengths and motifs can vary wildly between HIV-1 and HIV-2 sequences. However, viruses from a specific lineage, such as SIVmac239, SIVhu and SIVsmE041 can more readily be compared given their degree of sequence homology and point of origin.

One of the goals of this research was to identify any common changes between different strains as they gradually adapted to human immune cells, as their repeated appearance would suggest a crucial role in overcoming the inherent species-barriers. For example, the Gag V35I

substitution was identified in the MB897 (Group M) and EK505 (Group N) at the exact same residue location in a conserved region of Gag. While it appeared relatively quickly in EK505 and was greater than 50% frequency in the population within 11 weeks of the first generation, it occurred slightly later in MB897, staying near 0% frequency until the beginning of the second generation. This is consistent with the higher viral loads in EK505 relative to MB897, but also indicates that these strains may acquire adaptive mutations at differing rates. Furthermore, the V→I change has been identified as a potential protease inhibitor resistance mutation (Clark et al., 2007). Another example of common mutations is the conserved residue 216 of Gag in SIV_{mac239}, SIV_{hu} and SIV_{B670}. Interestingly, this mutation is absent from the stock or passaged SIV_{smE041}, even after eight generations, which could indicate that SIV_{smE041} acquired an alternative mutation along the way that served a similar purpose to the 216 mutation.

There are some broad trends that have been observed and apply to all of the viruses tested, regardless of lineage. Primarily, despite being a similar size to *gag* and *pol*, *env* had an overwhelmingly high frequency of mutations. To some degree, this could be attributed to APOBEC3G-mediated hypermutation, as *env* is located at the 3' of the genome, whereas *gag* and *pol* are located at the 5' end (Chelico et al., 2006). However, this would only explain the guanine to adenine mutations that would not otherwise be favorably selected. Another possibility is that the mutations in *env* are being driven by antibody-mediated selection, given that humanized mice have been shown to be capable of human antibody production (Akkina, 2013b; a). Additionally, mutations in the binding sites could also improve the viruses' affinity for human CD4, which could result in an increased rate of infection for the adapted viruses. It is also necessary to consider that in a mature virion, the Env protein is the only viral protein exposed beyond the lipid envelope and is critical for binding CD4, CXCR4 and CCR5 in viral entry. This makes it a ripe target for host

defenses, and consequently makes variability of the *env* sequence an advantageous trait compared to the more conserved *gag* and *pol* sequences (Veillette et al., 2014; Foley, 2018). The role of Env in HIV-2 and different SIVs for antagonizing host restriction factors such as tetherin could be a significant factor in driving the accumulation of adaptational changes (Sauter et al., 2012; Heusinger et al., 2018). Furthermore, among each of the tested viruses, there is a noticeable lack of persistent mutations that increase in frequency found in the *rev* and *tat* regulatory genes. This could be attributed to the fact that sections of their introns overlap in the genome not only with each other, but with other genes in different reading frames. While utilizing overlapping reading frames is an efficient means of increasing the complexity of a viral genome, it also means that mutations that would have previously been silent can now result in amino acid substitutions. The possibility of disrupting multiple protein sequences at once could certainly have a suppressive effect on the number of mutations that are tolerated at these sites.

Along with *env*, the other gene that acquired the greatest number of nonsynonymous mutations almost universally was *nef*. One trend regularly shown in these adapted viral strains is the accumulation of substitutions at important Nef binding sites. As seen by the fifth generation of serial passaging of SIV_{sm}, substitutions were found at sites involved with PAK1/2 binding, which mediates anti-apoptotic activity (Wolf et al., 2001). Similarly, mutations were identified in SIV_{hu} and SIV_{B670} that affect both PACS-1 binding sites and AP-1 binding sites, both of which are involved in endocytic sorting and the downregulation of MHC-I (Crump et al., 2001; Roeth et al., 2004; Heusinger et al., 2018). By altering these binding sites, it is possible that the SIV Nef is adapting to more effectively downregulate human MHC-I.

Of note is the occurrence of mutations involving dileucine motifs, LxxL(L/F), ExxxL(L/M), and DXAXXLL in these passaged viral strains. In SIV_{cpz}-lineage viruses, the Vpr

dileucine motif has been shown to be involved in transcriptional activation and the nuclear localization of the PIC complex. Previously, substitutions in this motif have generally resulted in advanced disease progression to AIDS (Sherman et al., 2000; Gonzalez, 2017). In comparison, substitutions in SIV_{B670} may have affected a dileucine motif in Gag that regulates the association and uptake of Vpx/Vpr to prevent their premature degradation, while SIV_{mac} produced a dileucine motif mutation in Nef that is involved in downregulating CD3 and CD4 (Aiken et al., 1994; Accola et al., 1999; Sherman et al., 2000; Gonzalez, 2017). The fact that these substitutions arose in dileucine motifs in both SIV_{cpz}- and SIV_{sm}-lineage viruses suggests that this may be a common target for adaptation when evolving towards a new host to increase viral infectivity.

Additionally, the interactions between host restriction factors and their viral antagonists are one of the primary drivers of host-virus evolution. Studies have shown that the presence of viral Vif and Vpr/Vpx have enacted a positive selection of counteractive mutations in APOBEC3G and SAMHD1, respectively. Coincidentally, the reverse has been shown to be true, with the actions of these host-restriction factors driving positive selection of mutations in the accessory viral genes (Compton et al., 2012). This is perhaps most clearly observed in the Nef Y193C substitution acquired by SIV_{mac239}, that increased to 100% frequency within the viral population despite being absent in the stock virus population. In addition to being found in a di-leucine motif, it has been found previously to directly mediate tetherin antagonism (Serra-Moreno et al., 2013). Interestingly, the SIV_{cpz} strains that were serially passaged did not appear to produce any mutations that affected tetherin or other host-restriction factors interactions. However, numerous mutations such as Env V402G, K414R, P414S, and N442I were identified in different SIV_{cpz} strains that altered known CD4 and chemokine binding sites. While this could be due to the genetic

similarity between chimpanzees and humans compared to rhesus macaques and sooty mangabeys, this question warrants further investigation.

9.2 Future Directions

Despite the preponderance of information that these experiments have generated, many questions regarding the genetic changes that occur during cross-species transmission of these viruses remain unanswered. First, several of these viral strains are not yet at the level of change that they can be confidently described as HIV. Rather, these adapted SIV strains appear to be in an intermediate point between HIV and SIV, where the ability to infect human immune cells is improving, but they are unlikely to be at the level of pathogenicity of the more pandemic strains of HIV-1. A more apt comparison would be with HIV-2 group F viruses, which have displayed positive infection in humans, but has shown little evidence of human to human transmission (Smith et al., 2008).

The most logical progression of this research is to continue to serially passage each of these viral strains and sequence the resulting timepoints. As we have shown with SIV_{sm}, these viruses continue to accumulate, and in a few cases, lose nonsynonymous mutations even after seven or eight serial passages. For example, Pol I172M, was absent until the eight generation and was present in over 50% of the viral population by the end of the generation, while multiple mutations in Nef including R17C/H, F65I, I141K made up over 60% of the viral population. In contrast, mutations in SIV_{sm} such as Vif K136R, Env E827G, and Nef D94N which were identified as potentially adaptive in the fifth generation were absent or below our frequency threshold by the eighth generation. While many of the most critical mutations arise within the first few generations, further serial passages help to reduce the noise and trim away those mutations arising purely by chance rather than selection. Maintaining these viruses through at least ten serial passages, or even

longer based on the results, would provide a more complete picture of their adaptation over time. Additionally, while the virus' evolution may initially focus on improving the chances of entry, or evading host restriction factors, other factors that mediate the fitness of the virus may come into play and affect viral evolution. With additional serial passages, it may be possible to see the preference for which genomic areas evolve over time.

One of the most common routes of HIV infection is by sexual transmission through mucosal vaginal and rectal tissues. This presents a different immune environment in mucosa with unique barriers to transmission when compared to an intraperitoneal challenge (Berges et al., 2008; Veselinovic et al., 2016). Therefore, it is reasonable to assume that an SIV that has adapted to an equivalent level of HIV should also be capable successful transmission through mucosal challenge. This challenge would also provide the unique opportunity for a comparative analysis of the evolutionary pressures exerted by these different entry routes. In theory, the mucosal challenge more closely resembles natural HIV infections, and would consequently produce genetic changes that more closely resemble existing HIV strains when compared to an i/p challenged virus. If the mutations that arose from a mucosal challenge were identical to the changes that arose from the i/p challenge, it would further strengthen the validity of i/p serial passaging as a surrogate model for cross-species transmissions, and would strengthen the claim that those mutations are critical to the ability of the virus to function in a novel host environment. Additionally, if the variant residues are not maintained when the adapted viruses are no longer cultured in human immune cells, it would further suggest their specific role in human-cell adaptation. This could easily be tested by serially back passaging the adapted viruses in nonhuman primate PBMC. If the variant residues reverted back to the residues found in the un-adapted stock virus sequences, it would highlight their species-specific role.

The studies described here focused on assessing viral adaption through plasma viral loads, CD4⁺ T cell decline and identification of adaptive mutations. However, additional areas of interest can be further explored. For example, a deeper investigation of the changing pathology of the passaged viruses could reveal valuable information. Immunohistochemical analysis of the thymus and lymph nodes of the humanized mice during infection with the adapted viruses could be compared to known pathologies of HIV infection to provide a more robust metric for assessing if these viruses are becoming more HIV-like.

While the identification of mutations that arise as SIVs become adapted to human immune cells is an important first step, comparing these mutations with known motifs and binding sites can only generate hypotheses regarding their functional importance. Therefore, it is critical that direct functional studies be conducted on the most important genetic changes identified. By introducing the mutations on an individual basis and testing their effect on growth and viral replication, we would be able to better understand how these changes have allowed the SIVs to adapt. This would also allow us to elucidate if these mutations function independently, or in tandem with each other.

In our work described here, the viruses tested were limited to those that either directly gave rise to HIV (SIVcpzEK505, SIVcpzLB715, SIVcpzMB897, SIVsmE041), or were specifically derived from these progenitors (SIVhu, SIVmac239, SIV_{B670}). To further explore the genetic requirements for successful transmission in human immune cells, additional primate lentiviruses could be serially passaged in humanized mice. SIVgor, which originated in gorillas and may have been transmitted from chimpanzees, would be of particular interest due to its potential role in the development of HIV-1 Groups P and O, and could serve as an effective foil alternative evolutionary model for SIVcpz in the same way that SIVmac is for SIVsm (Van Heuverswyn et al., 2006; Plantier et al., 2009; Sharp and Hahn, 2011). Other lentiviruses from different primates

such as mandrills (SIVmnd), red-capped mangabeys (SIVrcm), and African green monkeys (SIVagm) could also be used to further explore the genetic changes necessary for human transmission. Furthermore, some groups of HIV-2 are limited to just a few individual cases, with no evidence of human to human transmission and a lack of obvious pathology, such as in HIV-2 group F (Smith et al., 2008; Ayouba et al., 2013; Visseaux et al., 2016). These less pathogenic strains of HIV-2 may be acting as an intermediate and could resemble the more pathogenic HIV-2 groups A and B if given the chance to evolve. With HIV-2 group F and SIVhu each acting as separate intermediates for the transition of SIVsm into HIV-2, it may be possible to determine the “missing genetic link” that bridges the virus across two different species.

Though many questions about the nature of the origin of HIV remain, the work presented here has begun to shed some light on the process. By characterizing the genetic changes that gave rise to HIV-1 and HIV-2, more effective therapeutics can eventually be developed to specifically inhibit these lentiviruses and allow us to be better prepared in the event of further emergence of new lentiviruses pathogenic to humans.

References

- (UNAIDS), J.U.N.P.o.H.A. (2019). UNAIDS Data 2019. *Geneva: UNAIDS*.
- Accola, M.A., Bukovsky, A.A., Jones, M.S., and Gottlinger, H.G. (1999). A conserved dileucine-containing motif in p6(gag) governs the particle association of Vpx and Vpr of simian immunodeficiency viruses SIV(mac) and SIV(agm). *J Virol* 73(12), 9992-9999.
- Acevedo, A., Brodsky, L., and Andino, R. (2014). Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 505(7485), 686-690. doi: 10.1038/nature12861.
- Acharya, P., Lusvardi, S., Bewley, C.A., and Kwong, P.D. (2015). HIV-1 gp120 as a therapeutic target: navigating a moving labyrinth. *Expert Opin Ther Targets* 19(6), 765-783. doi: 10.1517/14728222.2015.1010513.
- Aghokeng, A.F., Ayouba, A., Mpoudi-Ngole, E., Loul, S., Liegeois, F., Delaporte, E., et al. (2010). Extensive survey on the prevalence and genetic diversity of SIVs in primate bushmeat provides insights into risks for potential new cross-species transmissions. *Infect Genet Evol* 10(3), 386-396. doi: 10.1016/j.meegid.2009.04.014.
- Aghokeng, A.F., Liu, W., Bibollet-Ruche, F., Loul, S., Mpoudi-Ngole, E., Laurent, C., et al. (2006). Widely varying SIV prevalence rates in naturally infected primate species from Cameroon. *Virology* 345(1), 174-189. doi: 10.1016/j.virol.2005.09.046.
- Ahn, J., Hao, C., Yan, J., DeLucia, M., Mehrens, J., Wang, C., et al. (2012). HIV/simian immunodeficiency virus (SIV) accessory virulence factor Vpx loads the host cell restriction factor SAMHD1 onto the E3 ubiquitin ligase complex CRL4DCAF1. *J Biol Chem* 287(15), 12550-12558. doi: 10.1074/jbc.M112.340711.
- Aiken, C., Konner, J., Landau, N.R., Lenburg, M.E., and Trono, D. (1994). Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell* 76(5), 853-864. doi: 10.1016/0092-8674(94)90360-3.
- Akkina, R. (2013a). Human immune responses and potential for vaccine assessment in humanized mice. *Curr Opin Immunol* 25(3), 403-409. doi: 10.1016/j.coi.2013.03.009.
- Akkina, R. (2013b). New generation humanized mice for virus research: comparative aspects and future prospects. *Virology* 435(1), 14-28. doi: 10.1016/j.virol.2012.10.007.
- Akkina, R.K., Rosenblatt, J.D., Campbell, A.G., Chen, I.S., and Zack, J.A. (1994). Modeling human lymphoid precursor cell gene therapy in the SCID-hu mouse. *Blood* 84(5), 1393-1398.
- Albin, J.S., and Harris, R.S. (2010). Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. *Expert Rev Mol Med* 12, e4. doi: 10.1017/S1462399409001343.
- Aldrovandi, G.M., Feuer, G., Gao, L., Jamieson, B., Kristeva, M., Chen, I.S., et al. (1993). The SCID-hu mouse as a model for HIV-1 infection. *Nature* 363(6431), 732-736. doi: 10.1038/363732a0.

- Anderson, S.J., Lenburg, M., Landau, N.R., and Garcia, J.V. (1994). The cytoplasmic domain of CD4 is sufficient for its down-regulation from the cell surface by human immunodeficiency virus type 1 Nef. *J Virol* 68(5), 3092-3101.
- Apetrei, C., Kaur, A., Lerche, N.W., Metzger, M., Pandrea, I., Hardcastle, J., et al. (2005). Molecular epidemiology of simian immunodeficiency virus SIVsm in U.S. primate centers unravels the origin of SIVmac and SIVstm. *J Virol* 79(14), 8991-9005. doi: 10.1128/JVI.79.14.8991-9005.2005.
- Arai, M., Earl, D.D., and Yamamoto, J.K. (2002). Is AZT/3TC therapy effective against FIV infection or immunopathogenesis? *Vet Immunol Immunopathol* 85(3-4), 189-204. doi: 10.1016/s0165-2427(01)00426-3.
- Arora, V.K., Molina, R.P., Foster, J.L., Blakemore, J.L., Chernoff, J., Fredericksen, B.L., et al. (2000). Lentivirus Nef specifically activates Pak2. *J Virol* 74(23), 11081-11087. doi: 10.1128/jvi.74.23.11081-11087.2000.
- Arunagiri, C., Macreadie, I., Hewish, D., and Azad, A. (1997). A C-terminal domain of HIV-1 accessory protein Vpr is involved in penetration, mitochondrial dysfunction and apoptosis of human CD4+ lymphocytes. *Apoptosis* 2(1), 69-76. doi: 10.1023/a:1026487609215.
- Arya, S.K., and Mohr, J.R. (1994). Conditional regulatory elements of human immunodeficiency virus type 2 long terminal repeat. *J Gen Virol* 75 (Pt 9), 2253-2260. doi: 10.1099/0022-1317-75-9-2253.
- Atkins, K.M., Thomas, L., Youker, R.T., Harriff, M.J., Pissani, F., You, H., et al. (2008). HIV-1 Nef binds PACS-2 to assemble a multikinase cascade that triggers major histocompatibility complex class I (MHC-I) down-regulation: analysis using short interfering RNA and knock-out mice. *J Biol Chem* 283(17), 11772-11784. doi: 10.1074/jbc.M707572200.
- Auerbach, D.M., Darrow, W.W., Jaffe, H.W., and Curran, J.W. (1984). Cluster of cases of the acquired immune deficiency syndrome. Patients linked by sexual contact. *Am J Med* 76(3), 487-492. doi: 10.1016/0002-9343(84)90668-5.
- Ayouba, A., Akoua-Koffi, C., Calvignac-Spencer, S., Esteban, A., Locatelli, S., Li, H., et al. (2013). Evidence for continuing cross-species transmission of SIVsmm to humans: characterization of a new HIV-2 lineage in rural Cote d'Ivoire. *AIDS* 27(15), 2488-2491. doi: 10.1097/01.aids.0000432443.22684.50.
- Ayyavoo, V., Mahboubi, A., Mahalingam, S., Ramalingam, R., Kudchodkar, S., Williams, W.V., et al. (1997). HIV-1 Vpr suppresses immune activation and apoptosis through regulation of nuclear factor kappa B. *Nat Med* 3(10), 1117-1123. doi: 10.1038/nm1097-1117.
- Baenziger, S., Tussiwand, R., Schlaepfer, E., Mazzucchelli, L., Heikenwalder, M., Kurrer, M.O., et al. (2006). Disseminated and sustained HIV infection in CD34+ cord blood cell-transplanted Rag2-/- gamma c-/- mice. *Proc Natl Acad Sci U S A* 103(43), 15951-15956. doi: 10.1073/pnas.0604493103.
- Bai, J., Gorantla, S., Banda, N., Cagnon, L., Rossi, J., and Akkina, R. (2000). Characterization of anti-CCR5 ribozyme-transduced CD34+ hematopoietic progenitor cells in vitro and in a SCID-hu mouse model in vivo. *Mol Ther* 1(3), 244-254. doi: 10.1006/mthe.2000.0038.

- Bailes, E., Gao, F., Bibollet-Ruche, F., Courgnaud, V., Peeters, M., Marx, P.A., et al. (2003). Hybrid origin of SIV in chimpanzees. *Science* 300(5626), 1713. doi: 10.1126/science.1080657.
- Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., et al. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 220(4599), 868-871. doi: 10.1126/science.6189183.
- Baskin, G.B., Murphey-Corb, M., Watson, E.A., and Martin, L.N. (1988). Necropsy findings in rhesus monkeys experimentally infected with cultured simian immunodeficiency virus (SIV)/delta. *Vet Pathol* 25(6), 456-467. doi: 10.1177/030098588802500609.
- Bebenek, K., Abbotts, J., Wilson, S.H., and Kunkel, T.A. (1993). Error-prone polymerization by HIV-1 reverse transcriptase. Contribution of template-primer misalignment, miscoding, and termination probability to mutational hot spots. *J Biol Chem* 268(14), 10324-10334.
- Beitari, S., Ding, S., Pan, Q., Finzi, A., and Liang, C. (2017). Effect of HIV-1 Env on SERINC5 Antagonism. *J Virol* 91(4). doi: 10.1128/JVI.02214-16.
- Bell, S.M., and Bedford, T. (2017). Modern-day SIV viral diversity generated by extensive recombination and cross-species transmission. *PLoS Pathog* 13(7), e1006466. doi: 10.1371/journal.ppat.1006466.
- Bergamaschi, A., Ayinde, D., David, A., Le Rouzic, E., Morel, M., Collin, G., et al. (2009). The human immunodeficiency virus type 2 Vpx protein usurps the CUL4A-DDB1 DCAF1 ubiquitin ligase to overcome a postentry block in macrophage infection. *J Virol* 83(10), 4854-4860. doi: 10.1128/JVI.00187-09.
- Berger, A., Sommer, A.F., Zwarg, J., Hamdorf, M., Welzel, K., Esly, N., et al. (2011). SAMHD1-deficient CD14+ cells from individuals with Aicardi-Goutieres syndrome are highly susceptible to HIV-1 infection. *PLoS Pathog* 7(12), e1002425. doi: 10.1371/journal.ppat.1002425.
- Berges, B.K., Akkina, S.R., Folkvord, J.M., Connick, E., and Akkina, R. (2008). Mucosal transmission of R5 and X4 tropic HIV-1 via vaginal and rectal routes in humanized Rag2^{-/-} gammac^{-/-} (RAG-hu) mice. *Virology* 373(2), 342-351. doi: 10.1016/j.virol.2007.11.020.
- Berges, B.K., Akkina, S.R., Remling, L., and Akkina, R. (2010). Humanized Rag2^(-/-)gammac^(-/-) (RAG-hu) mice can sustain long-term chronic HIV-1 infection lasting more than a year. *Virology* 397(1), 100-103. doi: 10.1016/j.virol.2009.10.034.
- Berges, B.K., and Rowan, M.R. (2011). The utility of the new generation of humanized mice to study HIV-1 infection: transmission, prevention, pathogenesis, and treatment. *Retrovirology* 8, 65. doi: 10.1186/1742-4690-8-65.
- Berges, B.K., Wheat, W.H., Palmer, B.E., Connick, E., and Akkina, R. (2006). HIV-1 infection and CD4 T cell depletion in the humanized Rag2^{-/-}-gamma c^{-/-} (RAG-hu) mouse model. *Retrovirology* 3, 76. doi: 10.1186/1742-4690-3-76.
- Besnier, C., Takeuchi, Y., and Towers, G. (2002). Restriction of lentivirus in monkeys. *Proc Natl Acad Sci U S A* 99(18), 11920-11925. doi: 10.1073/pnas.172384599.

- Betts, M., Russell, R. (2003). "Amino Acid Properties and Consequences of Substitutions," in *Bioinformatics for Geneticists*, ed. M. Barnes.), 289-316.
- Bibollet-Ruche, F., Heigele, A., Keele, B.F., Easlick, J.L., Decker, J.M., Takehisa, J., et al. (2012). Efficient SIVcpz replication in human lymphoid tissue requires viral matrix protein adaptation. *J Clin Invest* 122(5), 1644-1652. doi: 10.1172/JCI61429.
- Black, L.R., and Aiken, C. (2010). TRIM5alpha disrupts the structure of assembled HIV-1 capsid complexes in vitro. *J Virol* 84(13), 6564-6569. doi: 10.1128/JVI.00210-10.
- Bogerd, H.P., and Cullen, B.R. (2008). Single-stranded RNA facilitates nucleocapsid: APOBEC3G complex formation. *RNA* 14(6), 1228-1236. doi: 10.1261/rna.964708.
- Brehm, M.A., Shultz, L.D., and Greiner, D.L. (2010). Humanized mouse models to study human diseases. *Curr Opin Endocrinol Diabetes Obes* 17(2), 120-125. doi: 10.1097/MED.0b013e328337282f.
- Brehm, M.A., Wiles, M.V., Greiner, D.L., and Shultz, L.D. (2014). Generation of improved humanized mouse models for human infectious diseases. *J Immunol Methods* 410, 3-17. doi: 10.1016/j.jim.2014.02.011.
- Briggs, J.A., Wilk, T., Welker, R., Krausslich, H.G., and Fuller, S.D. (2003). Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J* 22(7), 1707-1715. doi: 10.1093/emboj/cdg143.
- Bukovsky, A.A., Weimann, A., Accola, M.A., and Gottlinger, H.G. (1997). Transfer of the HIV-1 cyclophilin-binding site to simian immunodeficiency virus from *Macaca mulatta* can confer both cyclosporin sensitivity and cyclosporin dependence. *Proc Natl Acad Sci U S A* 94(20), 10943-10948. doi: 10.1073/pnas.94.20.10943.
- Campbell, E.M., Nunez, R., and Hope, T.J. (2004). Disruption of the actin cytoskeleton can complement the ability of Nef to enhance human immunodeficiency virus type 1 infectivity. *J Virol* 78(11), 5745-5755. doi: 10.1128/JVI.78.11.5745-5755.2004.
- Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., et al. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22(3), 231-238. doi: 10.1038/10290.
- Chackerian, B., Rudensey, L.M., and Overbaugh, J. (1997). Specific N-linked and O-linked glycosylation modifications in the envelope V1 domain of simian immunodeficiency virus variants that evolve in the host alter recognition by neutralizing antibodies. *J Virol* 71(10), 7719-7727. doi: 10.1128/JVI.71.10.7719-7727.1997.
- Chahroudi, A., Bosinger, S.E., Vanderford, T.H., Paiardini, M., and Silvestri, G. (2012). Natural SIV hosts: showing AIDS the door. *Science* 335(6073), 1188-1193. doi: 10.1126/science.1217550.
- Chan, D.C., Fass, D., Berger, J.M., and Kim, P.S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 89(2), 263-273. doi: 10.1016/s0092-8674(00)80205-6.
- Chan, D.C., and Kim, P.S. (1998). HIV entry and its inhibition. *Cell* 93(5), 681-684. doi: 10.1016/s0092-8674(00)81430-0.

- Chatterjee, H.J., Ho, S.Y., Barnes, I., and Groves, C. (2009). Estimating the phylogeny and divergence times of primates using a supermatrix approach. *BMC Evol Biol* 9, 259. doi: 10.1186/1471-2148-9-259.
- Chaudhuri, R., Lindwasser, O.W., Smith, W.J., Hurley, J.H., and Bonifacino, J.S. (2007). Downregulation of CD4 by human immunodeficiency virus type 1 Nef is dependent on clathrin and involves direct interaction of Nef with the AP2 clathrin adaptor. *J Virol* 81(8), 3877-3890. doi: 10.1128/JVI.02725-06.
- Chelico, L., Pham, P., Calabrese, P., and Goodman, M.F. (2006). APOBEC3G DNA deaminase acts processively 3' --> 5' on single-stranded DNA. *Nat Struct Mol Biol* 13(5), 392-399. doi: 10.1038/nsmb1086.
- Chen, B., Vogan, E.M., Gong, H., Skehel, J.J., Wiley, D.C., and Harrison, S.C. (2005). Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature* 433(7028), 834-841. doi: 10.1038/nature03327.
- Chiu, Y.L., and Greene, W.C. (2006). Multifaceted antiviral actions of APOBEC3 cytidine deaminases. *Trends Immunol* 27(6), 291-297. doi: 10.1016/j.it.2006.04.003.
- Chowers, M.Y., Spina, C.A., Kwoh, T.J., Fitch, N.J., Richman, D.D., and Guatelli, J.C. (1994). Optimal infectivity in vitro of human immunodeficiency virus type 1 requires an intact nef gene. *J Virol* 68(5), 2906-2914.
- Clark, S., Calef, C., and Mellors, J.W. (2007). Mutations in retroviral genes associated with drug resistance.
- Clavel, F., Brun-Vezinet, F., Guetard, D., Chamaret, S., Laurent, A., Rouzioux, C., et al. (1986a). [LAV type II: a second retrovirus associated with AIDS in West Africa]. *C R Acad Sci III* 302(13), 485-488.
- Clavel, F., Guyader, M., Guetard, D., Salle, M., Montagnier, L., and Alizon, M. (1986b). Molecular cloning and polymorphism of the human immune deficiency virus type 2. *Nature* 324(6098), 691-695. doi: 10.1038/324691a0.
- Coetzer, M., Nedellec, R., Salkowitz, J., McLaughlin, S., Liu, Y., Heath, L., et al. (2008). Evolution of CCR5 use before and during coreceptor switching. *J Virol* 82(23), 11758-11766. doi: 10.1128/JVI.01141-08.
- Compton, A.A., and Emerman, M. (2013). Convergence and divergence in the evolution of the APOBEC3G-Vif interaction reveal ancient origins of simian immunodeficiency viruses. *PLoS Pathog* 9(1), e1003135. doi: 10.1371/journal.ppat.1003135.
- Compton, A.A., Hirsch, V.M., and Emerman, M. (2012). The host restriction factor APOBEC3G and retroviral Vif protein coevolve due to ongoing genetic conflict. *Cell Host Microbe* 11(1), 91-98. doi: 10.1016/j.chom.2011.11.010.
- Crump, C.M., Xiang, Y., Thomas, L., Gu, F., Austin, C., Tooze, S.A., et al. (2001). PACS-1 binding to adaptors is required for acidic cluster motif-mediated protein traffic. *EMBO J* 20(9), 2191-2201. doi: 10.1093/emboj/20.9.2191.

- Cuevas, J.M., Domingo-Calap, P., and Sanjuan, R. (2012). The fitness effects of synonymous mutations in DNA and RNA viruses. *Mol Biol Evol* 29(1), 17-20. doi: 10.1093/molbev/msr179.
- Curlin, J., Schmitt, K., Remling-Mulder, L., Moriarty, R., Goff, K., O'Connor, S., et al. (2020a). Evolution of SIVsm in humanized mice towards HIV-2. *J Med Primatol*. doi: 10.1111/jmp.12486.
- Curlin, J., Schmitt, K., Remling-Mulder, L., Moriarty, R., Stenglein, M., O'Connor, S., et al. (2020b). SIVcpz cross-species transmission and viral evolution toward HIV-1 in a humanized mouse model. *J Med Primatol* 49(1), 40-43. doi: 10.1111/jmp.12440.
- Dang, Y., Wang, X., Esselman, W.J., and Zheng, Y.H. (2006). Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. *J Virol* 80(21), 10522-10533. doi: 10.1128/JVI.01123-06.
- Daniel, M.D., Letvin, N.L., King, N.W., Kannagi, M., Sehgal, P.K., Hunt, R.D., et al. (1985). Isolation of T-cell tropic HTLV-III-like retrovirus from macaques. *Science* 228(4704), 1201-1204. doi: 10.1126/science.3159089.
- Dayhoff, M.O. (1972). *Atlas of protein sequence and structure*. National Biomedical Research Foundation.
- De Leys, R., Vanderborcht, B., Vanden Haesevelde, M., Heyndrickx, L., van Geel, A., Wauters, C., et al. (1990). Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *J Virol* 64(3), 1207-1216. doi: 10.1128/JVI.64.3.1207-1216.1990.
- Decker, J.M., Bibollet-Ruche, F., Wei, X., Wang, S., Levy, D.N., Wang, W., et al. (2005). Antigenic conservation and immunogenicity of the HIV coreceptor binding site. *J Exp Med* 201(9), 1407-1419. doi: 10.1084/jem.20042510.
- Denton, P.W., and Garcia, J.V. (2011). Humanized mouse models of HIV infection. *AIDS Rev* 13(3), 135-148.
- Derdeyn, C.A., Decker, J.M., Sfakianos, J.N., Wu, X., O'Brien, W.A., Ratner, L., et al. (2000). Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. *J Virol* 74(18), 8358-8367.
- Dodding, M.P., Bock, M., Yap, M.W., and Stoye, J.P. (2005). Capsid processing requirements for abrogation of Fv1 and Ref1 restriction. *J Virol* 79(16), 10571-10577. doi: 10.1128/JVI.79.16.10571-10577.2005.
- Dow, S.W., Poss, M.L., and Hoover, E.A. (1990). Feline immunodeficiency virus: a neurotropic lentivirus. *J Acquir Immune Defic Syndr (1988)* 3(7), 658-668.
- Egan, M.A., Carruth, L.M., Rowell, J.F., Yu, X., and Siliciano, R.F. (1996). Human immunodeficiency virus type 1 envelope protein endocytosis mediated by a highly conserved intrinsic internalization signal in the cytoplasmic domain of gp41 is suppressed in the presence of the Pr55gag precursor protein. *J Virol* 70(10), 6547-6556. doi: 10.1128/JVI.70.10.6547-6556.1996.

- Etienne, L., Bibollet-Ruche, F., Sudmant, P.H., Wu, L.I., Hahn, B.H., and Emerman, M. (2015). The Role of the Antiviral APOBEC3 Gene Family in Protecting Chimpanzees against Lentiviruses from Monkeys. *PLoS Pathog* 11(9), e1005149. doi: 10.1371/journal.ppat.1005149.
- Etienne, L., Hahn, B.H., Sharp, P.M., Matsen, F.A., and Emerman, M. (2013). Gene loss and adaptation to hominids underlie the ancient origin of HIV-1. *Cell Host Microbe* 14(1), 85-92. doi: 10.1016/j.chom.2013.06.002.
- Favaro, J.P., Borg, K.T., Arrigo, S.J., and Schmidt, M.G. (1998). Effect of Rev on the intranuclear localization of HIV-1 unspliced RNA. *Virology* 249(2), 286-296. doi: 10.1006/viro.1998.9312.
- Felber, B.K., Drysdale, C.M., and Pavlakis, G.N. (1990). Feedback regulation of human immunodeficiency virus type 1 expression by the Rev protein. *J Virol* 64(8), 3734-3741. doi: 10.1128/JVI.64.8.3734-3741.1990.
- Ferrucci, A., Nonnemacher, M.R., and Wigdahl, B. (2011). Human immunodeficiency virus viral protein R as an extracellular protein in neuropathogenesis. *Adv Virus Res* 81, 165-199. doi: 10.1016/B978-0-12-385885-6.00010-9.
- Finkel, T.H., Tudor-Williams, G., Banda, N.K., Cotton, M.F., Curiel, T., Monks, C., et al. (1995). Apoptosis occurs predominantly in bystander cells and not in productively infected cells of HIV- and SIV-infected lymph nodes. *Nat Med* 1(2), 129-134. doi: 10.1038/nm0295-129.
- Foley, B.L., T.; Apetrei, C.; Hahn, B.; Mizrachi, I.; Mullins, J.; Rambaut, A.; Wolinsky, S.; Korber, B. (2016). "HIV Sequence Compendium 2016". Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico).
- Foley, B.L., T.; Apetrei, C.; Hahn, B.; Mizrachi, I.; Mullins, J.; Rambaut, A.; Wolinsky, S.; Korber, B. (2018). "HIV Sequence Compendium 2018". Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico).
- Foley, B.L., Thomas; Apetrei, Christian; Hahn, Beatrice; Mizrachi, Ilene; Mullins, James; Rambaut, Andrew; Wolinsky, Steven; Korber, Bette (2017). "HIV Sequence Compendium 2017". (Los Alamos, New Mexico: Los Alamos National Laboratory, Theoretical and Biophysics).
- Forshey, B.M., von Schwedler, U., Sundquist, W.I., and Aiken, C. (2002). Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J Virol* 76(11), 5667-5677. doi: 10.1128/jvi.76.11.5667-5677.2002.
- Frankel, A.D., and Young, J.A. (1998). HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* 67, 1-25. doi: 10.1146/annurev.biochem.67.1.1.
- Fritz, C.C., Zapp, M.L., and Green, M.R. (1995). A human nucleoporin-like protein that specifically interacts with HIV Rev. *Nature* 376(6540), 530-533. doi: 10.1038/376530a0.
- Fultz, P.N., McClure, H.M., Anderson, D.C., Swenson, R.B., Anand, R., and Srinivasan, A. (1986). Isolation of a T-lymphotropic retrovirus from naturally infected sooty mangabey monkeys (*Cercocebus atys*). *Proc Natl Acad Sci U S A* 83(14), 5286-5290. doi: 10.1073/pnas.83.14.5286.

- Gaddis, N.C., Sheehy, A.M., Ahmad, K.M., Swanson, C.M., Bishop, K.N., Beer, B.E., et al. (2004). Further investigation of simian immunodeficiency virus Vif function in human cells. *J Virol* 78(21), 12041-12046. doi: 10.1128/JVI.78.21.12041-12046.2004.
- Ganser-Pornillos, B.K., and Pornillos, O. (2019). Restriction of HIV-1 and other retroviruses by TRIM5. *Nat Rev Microbiol* 17(9), 546-556. doi: 10.1038/s41579-019-0225-2.
- Ganser-Pornillos, B.K., Yeager, M., and Pornillos, O. (2012). Assembly and architecture of HIV. *Adv Exp Med Biol* 726, 441-465. doi: 10.1007/978-1-4614-0980-9_20.
- Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., et al. (1999). Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* 397(6718), 436-441. doi: 10.1038/17130.
- Gao, F., Yue, L., White, A.T., Pappas, P.G., Barchue, J., Hanson, A.P., et al. (1992). Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature* 358(6386), 495-499. doi: 10.1038/358495a0.
- Garcia, J.V., and Miller, A.D. (1991). Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. *Nature* 350(6318), 508-511. doi: 10.1038/350508a0.
- Garcia, S., and Freitas, A.A. (2012). Humanized mice: current states and perspectives. *Immunol Lett* 146(1-2), 1-7. doi: 10.1016/j.imlet.2012.03.009.
- Gautam, R., Carter, A.C., Katz, N., Butler, I.F., Barnes, M., Hasegawa, A., et al. (2007). In vitro characterization of primary SIVsmm isolates belonging to different lineages. In vitro growth on rhesus macaque cells is not predictive for in vivo replication in rhesus macaques. *Virology* 362(2), 257-270. doi: 10.1016/j.virol.2006.12.037.
- Ghimire, D., Rai, M., and Gaur, R. (2018). Novel host restriction factors implicated in HIV-1 replication. *J Gen Virol* 99(4), 435-446. doi: 10.1099/jgv.0.001026.
- Gifford, R.J. (2012). Viral evolution in deep time: lentiviruses and mammals. *Trends Genet* 28(2), 89-100. doi: 10.1016/j.tig.2011.11.003.
- Gilbert, C., Maxfield, D.G., Goodman, S.M., and Feschotte, C. (2009). Parallel germline infiltration of a lentivirus in two Malagasy lemurs. *PLoS Genet* 5(3), e1000425. doi: 10.1371/journal.pgen.1000425.
- Goila-Gaur, R., and Strebel, K. (2008). HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology* 5, 51. doi: 10.1186/1742-4690-5-51.
- Goldstone, D.C., Ennis-Adeniran, V., Hedden, J.J., Groom, H.C., Rice, G.I., Christodoulou, E., et al. (2011). HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* 480(7377), 379-382. doi: 10.1038/nature10623.
- Gonzalez, M.E. (2017). The HIV-1 Vpr Protein: A Multifaceted Target for Therapeutic Intervention. *Int J Mol Sci* 18(1). doi: 10.3390/ijms18010126.

- Gormus, B.J., Xu, K., Baskin, G.B., Martin, L.N., Bohm, R.P., Blanchard, J.L., et al. (1995a). Experimental leprosy in monkeys. I. Sooty mangabey monkeys: transmission, susceptibility, clinical and pathological findings. *Lepr Rev* 66(2), 96-104. doi: 10.5935/0305-7518.19950012.
- Gormus, B.J., Xu, K., Cho, S.N., Baskin, G.B., Bohm, R.P., Martin, L.N., et al. (1995b). Experimental leprosy in monkeys. II. Longitudinal serological observations in sooty mangabey monkeys. *Lepr Rev* 66(2), 105-125. doi: 10.5935/0305-7518.19950013.
- Gorry, P.R., and Ancuta, P. (2011). Coreceptors and HIV-1 pathogenesis. *Curr HIV/AIDS Rep* 8(1), 45-53. doi: 10.1007/s11904-010-0069-x.
- Gottlieb, M.S., Schroff, R., Schanker, H.M., Weisman, J.D., Fan, P.T., Wolf, R.A., et al. (1981). Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med* 305(24), 1425-1431. doi: 10.1056/NEJM198112103052401.
- Goujon, C., Arfi, V., Pertel, T., Luban, J., Lienard, J., Rigal, D., et al. (2008). Characterization of simian immunodeficiency virus SIVSM/human immunodeficiency virus type 2 Vpx function in human myeloid cells. *J Virol* 82(24), 12335-12345. doi: 10.1128/JVI.01181-08.
- Granier, C., Battivelli, E., Lecuroux, C., Venet, A., Lambotte, O., Schmitt-Boulanger, M., et al. (2013). Pressure from TRIM5alpha contributes to control of HIV-1 replication by individuals expressing protective HLA-B alleles. *J Virol* 87(18), 10368-10380. doi: 10.1128/JVI.01313-13.
- Greenberg, M.E., Iafrate, A.J., and Skowronski, J. (1998). The SH3 domain-binding surface and an acidic motif in HIV-1 Nef regulate trafficking of class I MHC complexes. *EMBO J* 17(10), 2777-2789. doi: 10.1093/emboj/17.10.2777.
- Greenway, A., and McPhee, D. (1997). HIV1 Nef: the Machiavelli of cellular activation. *Res Virol* 148(1), 58-64. doi: 10.1016/s0923-2516(97)81915-2.
- Guo, F., Cen, S., Niu, M., Saadatmand, J., and Kleiman, L. (2006). Inhibition of tRNA(3)(Lys)-primed reverse transcription by human APOBEC3G during human immunodeficiency virus type 1 replication. *J Virol* 80(23), 11710-11722. doi: 10.1128/JVI.01038-06.
- Hahn, B.H., Shaw, G.M., De Cock, K.M., and Sharp, P.M. (2000). AIDS as a zoonosis: scientific and public health implications. *Science* 287(5453), 607-614.
- Hannibal, M.C., Markovitz, D.M., and Nabel, G.J. (1994). Multiple cis-acting elements in the human immunodeficiency virus type 2 enhancer mediate the response to T-cell receptor stimulation by antigen in a T-cell hybridoma line. *Blood* 83(7), 1839-1846.
- Hatzioannou, T., and Evans, D.T. (2012). Animal models for HIV/AIDS research. *Nat Rev Microbiol* 10(12), 852-867. doi: 10.1038/nrmicro2911.
- Hatzioannou, T., Perez-Caballero, D., Yang, A., Cowan, S., and Bieniasz, P.D. (2004). Retrovirus resistance factors Ref1 and Lv1 are species-specific variants of TRIM5alpha. *Proc Natl Acad Sci U S A* 101(29), 10774-10779. doi: 10.1073/pnas.0402361101.

- Hatzioannou, T., Princiotta, M., Piatak, M., Jr., Yuan, F., Zhang, F., Lifson, J.D., et al. (2006). Generation of simian-tropic HIV-1 by restriction factor evasion. *Science* 314(5796), 95. doi: 10.1126/science.1130994.
- Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends Mol Med* 18(3), 182-192. doi: 10.1016/j.molmed.2011.12.001.
- Henikoff, J.G., and Henikoff, S. (1996). Using substitution probabilities to improve position-specific scoring matrices. *Comput Appl Biosci* 12(2), 135-143. doi: 10.1093/bioinformatics/12.2.135.
- Heusinger, E., Deppe, K., Sette, P., Krapp, C., Kmiec, D., Kluge, S.F., et al. (2018). Preadaptation of Simian Immunodeficiency Virus SIVsmm Facilitated Env-Mediated Counteraction of Human Tetherin by Human Immunodeficiency Virus Type 2. *J Virol* 92(18). doi: 10.1128/JVI.00276-18.
- Hinz, A., Miguet, N., Natrajan, G., Usami, Y., Yamanaka, H., Renesto, P., et al. (2010). Structural basis of HIV-1 tethering to membranes by the BST-2/tetherin ectodomain. *Cell Host Microbe* 7(4), 314-323. doi: 10.1016/j.chom.2010.03.005.
- Hirao, K., Andrews, S., Kuroki, K., Kusaka, H., Tadokoro, T., Kita, S., et al. (2020). Structure of HIV-2 Nef Reveals Features Distinct from HIV-1 Involved in Immune Regulation. *iScience* 23(1), 100758. doi: 10.1016/j.isci.2019.100758.
- Hirsch, V.M., Olmsted, R.A., Murphey-Corb, M., Purcell, R.H., and Johnson, P.R. (1989). An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature* 339(6223), 389-392. doi: 10.1038/339389a0.
- Hohne, K., Businger, R., van Nuffel, A., Bolduan, S., Koppensteiner, H., Baeyens, A., et al. (2016). Virion encapsidated HIV-1 Vpr induces NFAT to prime non-activated T cells for productive infection. *Open Biol* 6(7). doi: 10.1098/rsob.160046.
- Holmes, R.K., Koning, F.A., Bishop, K.N., and Malim, M.H. (2007). APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G. *J Biol Chem* 282(4), 2587-2595. doi: 10.1074/jbc.M607298200.
- Hope, T.J. (1999). The ins and outs of HIV Rev. *Arch Biochem Biophys* 365(2), 186-191. doi: 10.1006/abbi.1999.1207.
- Hrecka, K., Hao, C., Gierszewska, M., Swanson, S.K., Kesik-Brodacka, M., Srivastava, S., et al. (2011). Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* 474(7353), 658-661. doi: 10.1038/nature10195.
- Hua, J., Blair, W., Truant, R., and Cullen, B.R. (1997). Identification of regions in HIV-1 Nef required for efficient downregulation of cell surface CD4. *Virology* 231(2), 231-238. doi: 10.1006/viro.1997.8517.
- Huang, H., Zhang, X., Li, S., Liu, N., Lian, W., McDowell, E., et al. (2010). Physiological levels of ATP negatively regulate proteasome function. *Cell Res* 20(12), 1372-1385. doi: 10.1038/cr.2010.123.
- Ito, M., Hiramatsu, H., Kobayashi, K., Suzue, K., Kawahata, M., Hioki, K., et al. (2002). NOD/SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells. *Blood* 100(9), 3175-3182. doi: 10.1182/blood-2001-12-0207.

- Ito, R., Takahashi, T., Katano, I., and Ito, M. (2012). Current advances in humanized mouse models. *Cell Mol Immunol* 9(3), 208-214. doi: 10.1038/cmi.2012.2.
- Iwatani, Y., Chan, D.S., Wang, F., Stewart-Maynard, K., Sugiura, W., Gronenborn, A.M., et al. (2007). Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. *Nucleic Acids Res* 35(21), 7096-7108. doi: 10.1093/nar/gkm750.
- Jia, B., Serra-Moreno, R., Neidermyer, W., Rahmberg, A., Mackey, J., Fofana, I.B., et al. (2009). Species-specific activity of SIV Nef and HIV-1 Vpu in overcoming restriction by tetherin/BST2. *PLoS Pathog* 5(5), e1000429. doi: 10.1371/journal.ppat.1000429.
- Johnson, W.E., Morgan, J., Reitter, J., Puffer, B.A., Czajak, S., Doms, R.W., et al. (2002). A replication-competent, neutralization-sensitive variant of simian immunodeficiency virus lacking 100 amino acids of envelope. *J Virol* 76(5), 2075-2086. doi: 10.1128/jvi.76.5.2075-2086.2002.
- Jonckheere, H., De Clercq, E., and Anne, J. (2000). Fidelity analysis of HIV-1 reverse transcriptase mutants with an altered amino-acid sequence at residues Leu74, Glu89, Tyr115, Tyr183 and Met184. *Eur J Biochem* 267(9), 2658-2665. doi: 10.1046/j.1432-1327.2000.01272.x.
- Joseph, S.B., Swanstrom, R., Kashuba, A.D., and Cohen, M.S. (2015). Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. *Nat Rev Microbiol* 13(7), 414-425. doi: 10.1038/nrmicro3471.
- Jouvenet, N., Neil, S.J., Zhadina, M., Zang, T., Kratovac, Z., Lee, Y., et al. (2009). Broad-spectrum inhibition of retroviral and filoviral particle release by tetherin. *J Virol* 83(4), 1837-1844. doi: 10.1128/JVI.02211-08.
- Kamata, M., and Aida, Y. (2000). Two putative alpha-helical domains of human immunodeficiency virus type 1 Vpr mediate nuclear localization by at least two mechanisms. *J Virol* 74(15), 7179-7186. doi: 10.1128/jvi.74.15.7179-7186.2000.
- Kariuki, S.M., Selhorst, P., Arien, K.K., and Dorfman, J.R. (2017). The HIV-1 transmission bottleneck. *Retrovirology* 14(1), 22. doi: 10.1186/s12977-017-0343-8.
- Katz, R.A., and Skalka, A.M. (1994). The retroviral enzymes. *Annu Rev Biochem* 63, 133-173. doi: 10.1146/annurev.bi.63.070194.001025.
- Keele, B.F., Jones, J.H., Terio, K.A., Estes, J.D., Rudicell, R.S., Wilson, M.L., et al. (2009). Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. *Nature* 460(7254), 515-519. doi: 10.1038/nature08200.
- Keele, B.F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M.L., et al. (2006). Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* 313(5786), 523-526. doi: 10.1126/science.1126531.
- Kerkau, T., Bacik, I., Bennink, J.R., Yewdell, J.W., Hunig, T., Schimpl, A., et al. (1997). The human immunodeficiency virus type 1 (HIV-1) Vpu protein interferes with an early step in the biosynthesis of major histocompatibility complex (MHC) class I molecules. *J Exp Med* 185(7), 1295-1305. doi: 10.1084/jem.185.7.1295.

- Kestler, H., Kodama, T., Ringler, D., Marthas, M., Pedersen, N., Lackner, A., et al. (1990). Induction of AIDS in rhesus monkeys by molecularly cloned simian immunodeficiency virus. *Science* 248(4959), 1109-1112. doi: 10.1126/science.2160735.
- Kestler, H.W., 3rd, Ringler, D.J., Mori, K., Panicali, D.L., Sehgal, P.K., Daniel, M.D., et al. (1991). Importance of the nef gene for maintenance of high virus loads and for development of AIDS. *Cell* 65(4), 651-662. doi: 10.1016/0092-8674(91)90097-i.
- Keusch, G.T., Pappaioanou, M., Gonzalez, M.C., Scott, K.A., Tsai, P., and Council, N.R. (2009). "Drivers of Zoonotic Diseases," in *Sustaining Global Surveillance and Response to Emerging Zoonotic Diseases*. National Academies Press (US)).
- Khabbaz, R.F., Heneine, W., George, J.R., Parekh, B., Rowe, T., Woods, T., et al. (1994). Brief report: infection of a laboratory worker with simian immunodeficiency virus. *N Engl J Med* 330(3), 172-177. doi: 10.1056/NEJM199401203300304.
- Khabbaz, R.F., Rowe, T., Murphey-Corb, M., Heneine, W.M., Schable, C.A., George, J.R., et al. (1992). Simian immunodeficiency virus needlestick accident in a laboratory worker. *Lancet* 340(8814), 271-273.
- Khalid, M., Yu, H., Sauter, D., Usmani, S.M., Schmokel, J., Feldman, J., et al. (2012). Efficient Nef-mediated downmodulation of TCR-CD3 and CD28 is associated with high CD4+ T cell counts in viremic HIV-2 infection. *J Virol* 86(9), 4906-4920. doi: 10.1128/JVI.06856-11.
- King, M.A., Covassin, L., Brehm, M.A., Racki, W., Pearson, T., Leif, J., et al. (2009). Human peripheral blood leucocyte non-obese diabetic-severe combined immunodeficiency interleukin-2 receptor gamma chain gene mouse model of xenogeneic graft-versus-host-like disease and the role of host major histocompatibility complex. *Clin Exp Immunol* 157(1), 104-118. doi: 10.1111/j.1365-2249.2009.03933.x.
- Kino, T., Gragerov, A., Kopp, J.B., Stauber, R.H., Pavlakakis, G.N., and Chrousos, G.P. (1999). The HIV-1 virion-associated protein vpr is a coactivator of the human glucocorticoid receptor. *J Exp Med* 189(1), 51-62. doi: 10.1084/jem.189.1.51.
- Klatt, N.R., Canary, L.A., Vanderford, T.H., Vinton, C.L., Engram, J.C., Dunham, R.M., et al. (2012). Dynamics of simian immunodeficiency virus SIVmac239 infection in pigtail macaques. *J Virol* 86(2), 1203-1213. doi: 10.1128/JVI.06033-11.
- Kluge, S.F., Sauter, D., Vogl, M., Peeters, M., Li, Y., Bibollet-Ruche, F., et al. (2013). The transmembrane domain of HIV-1 Vpu is sufficient to confer anti-tetherin activity to SIVcpz and SIVgor Vpu proteins: cytoplasmic determinants of Vpu function. *Retrovirology* 10, 32. doi: 10.1186/1742-4690-10-32.
- Kogan, M., and Rappaport, J. (2011). HIV-1 accessory protein Vpr: relevance in the pathogenesis of HIV and potential for therapeutic intervention. *Retrovirology* 8, 25. doi: 10.1186/1742-4690-8-25.
- Koning, F.A., Newman, E.N., Kim, E.Y., Kunstman, K.J., Wolinsky, S.M., and Malim, M.H. (2009). Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol* 83(18), 9474-9485. doi: 10.1128/JVI.01089-09.

- Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., and Detours, V. (2001). Evolutionary and immunological implications of contemporary HIV-1 variation. *Br Med Bull* 58, 19-42. doi: 10.1093/bmb/58.1.19.
- Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., et al. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science* 288(5472), 1789-1796. doi: 10.1126/science.288.5472.1789.
- Kowalski, M., Potz, J., Basiripour, L., Dorfman, T., Goh, W.C., Terwilliger, E., et al. (1987). Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. *Science* 237(4820), 1351-1355. doi: 10.1126/science.3629244.
- Kreiss, J.K., Koech, D., Plummer, F.A., Holmes, K.K., Lightfoote, M., Piot, P., et al. (1986). AIDS virus infection in Nairobi prostitutes. Spread of the epidemic to East Africa. *N Engl J Med* 314(7), 414-418. doi: 10.1056/NEJM198602133140704.
- Kwak, Y.T., Raney, A., Kuo, L.S., Denial, S.J., Temple, B.R., Garcia, J.V., et al. (2010). Self-association of the Lentivirus protein, Nef. *Retrovirology* 7, 77. doi: 10.1186/1742-4690-7-77.
- Laguette, N., and Benkirane, M. (2012). How SAMHD1 changes our view of viral restriction. *Trends Immunol* 33(1), 26-33. doi: 10.1016/j.it.2011.11.002.
- Laguette, N., Sobhian, B., Casartelli, N., Ringeard, M., Chable-Bessia, C., Segeral, E., et al. (2011). SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* 474(7353), 654-657. doi: 10.1038/nature10117.
- Lahouassa, H., Daddacha, W., Hofmann, H., Ayinde, D., Logue, E.C., Dragin, L., et al. (2012). SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat Immunol* 13(3), 223-228. doi: 10.1038/ni.2236.
- Lan, P., Tonomura, N., Shimizu, A., Wang, S., and Yang, Y.G. (2006). Reconstitution of a functional human immune system in immunodeficient mice through combined human fetal thymus/liver and CD34+ cell transplantation. *Blood* 108(2), 487-492. doi: 10.1182/blood-2005-11-4388.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4), 357-359. doi: 10.1038/nmeth.1923.
- Lapidot, T., Pflumio, F., Doedens, M., Murdoch, B., Williams, D.E., and Dick, J.E. (1992). Cytokine stimulation of multilineage hematopoiesis from immature human cells engrafted in SCID mice. *Science* 255(5048), 1137-1141. doi: 10.1126/science.1372131.
- Le Hingrat, Q., Visseaux, B., Bertine, M., Chauveau, L., Schwartz, O., Collin, F., et al. (2020). Genetic Variability of Long Terminal Repeat Region between HIV-2 Groups Impacts Transcriptional Activity. *J Virol* 94(7). doi: 10.1128/JVI.01504-19.
- Le Tortorec, A., and Neil, S.J. (2009). Antagonism to and intracellular sequestration of human tetherin by the human immunodeficiency virus type 2 envelope glycoprotein. *J Virol* 83(22), 11966-11978. doi: 10.1128/JVI.01515-09.

- Lee, P.P., and Linial, M.L. (1994). Efficient particle formation can occur if the matrix domain of human immunodeficiency virus type 1 Gag is substituted by a myristylation signal. *J Virol* 68(10), 6644-6654. doi: 10.1128/JVI.68.10.6644-6654.1994.
- Lepus, C.M., Gibson, T.F., Gerber, S.A., Kawikova, I., Szczepanik, M., Hossain, J., et al. (2009). Comparison of human fetal liver, umbilical cord blood, and adult blood hematopoietic stem cell engraftment in NOD-scid/gammac^{-/-}, Balb/c-Rag1^{-/-}gammac^{-/-}, and C.B-17-scid/bg immunodeficient mice. *Hum Immunol* 70(10), 790-802. doi: 10.1016/j.humimm.2009.06.005.
- Li, W., and Godzik, A. (2006). VISSA: a program to visualize structural features from structure sequence alignment. *Bioinformatics* 22(7), 887-888. doi: 10.1093/bioinformatics/btl019.
- Lim, E.S., Fregoso, O.I., McCoy, C.O., Matsen, F.A., Malik, H.S., and Emerman, M. (2012). The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. *Cell Host Microbe* 11(2), 194-204. doi: 10.1016/j.chom.2012.01.004.
- Lim, E.S., Malik, H.S., and Emerman, M. (2010). Ancient adaptive evolution of tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses. *J Virol* 84(14), 7124-7134. doi: 10.1128/JVI.00468-10.
- Ling, B., Apetrei, C., Pandrea, I., Veazey, R.S., Lackner, A.A., Gormus, B., et al. (2004). Classic AIDS in a sooty mangabey after an 18-year natural infection. *J Virol* 78(16), 8902-8908. doi: 10.1128/JVI.78.16.8902-8908.2004.
- Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., et al. (1996). Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 86(3), 367-377. doi: 10.1016/s0092-8674(00)80110-5.
- Luban, J. (1996). Absconding with the chaperone: essential cyclophilin-Gag interaction in HIV-1 virions. *Cell* 87(7), 1157-1159. doi: 10.1016/s0092-8674(00)81811-5.
- Luo, K., Wang, T., Liu, B., Tian, C., Xiao, Z., Kappes, J., et al. (2007). Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation. *J Virol* 81(13), 7238-7248. doi: 10.1128/JVI.02584-06.
- Luo, X., Herzig, E., Doitsh, G., Grimmer, Z.W., Munoz-Arias, I., and Greene, W.C. (2019). HIV-2 Depletes CD4 T Cells through Pyroptosis despite Vpx-Dependent Degradation of SAMHD1. *J Virol* 93(24). doi: 10.1128/JVI.00666-19.
- Macreadie, I.G., Thorburn, D.R., Kirby, D.M., Castelli, L.A., de Rozario, N.L., and Azad, A.A. (1997). HIV-1 protein Vpr causes gross mitochondrial dysfunction in the yeast *Saccharomyces cerevisiae*. *FEBS Lett* 410(2-3), 145-149. doi: 10.1016/s0014-5793(97)00542-5.
- Malim, M.H. (2009). APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philos Trans R Soc Lond B Biol Sci* 364(1517), 675-687. doi: 10.1098/rstb.2008.0185.
- Malim, M.H., and Bieniasz, P.D. (2012). HIV Restriction Factors and Mechanisms of Evasion. *Cold Spring Harb Perspect Med* 2(5), a006940. doi: 10.1101/cshperspect.a006940.

- Manrique, S., Sauter, D., Horenkamp, F.A., Lulf, S., Yu, H., Hotter, D., et al. (2017). Endocytic sorting motif interactions involved in Nef-mediated downmodulation of CD4 and CD3. *Nat Commun* 8(1), 442. doi: 10.1038/s41467-017-00481-z.
- Manz, M.G. (2007). Human-hemato-lymphoid-system mice: opportunities and challenges. *Immunity* 26(5), 537-541. doi: 10.1016/j.immuni.2007.05.001.
- Marin, M., Rose, K.M., Kozak, S.L., and Kabat, D. (2003). HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat Med* 9(11), 1398-1403. doi: 10.1038/nm946.
- Markle, T.J., Philip, M., and Brockman, M.A. (2013). HIV-1 Nef and T-cell activation: a history of contradictions. *Future Virol* 8(4). doi: 10.2217/fvl.13.20.
- Markovitz, D.M., Hannibal, M., Perez, V.L., Gauntt, C., Folks, T.M., and Nabel, G.J. (1990). Differential regulation of human immunodeficiency viruses (HIVs): a specific regulatory element in HIV-2 responds to stimulation of the T-cell antigen receptor. *Proc Natl Acad Sci U S A* 87(23), 9098-9102. doi: 10.1073/pnas.87.23.9098.
- Markovitz, D.M., Smith, M.J., Hilfinger, J., Hannibal, M.C., Petryniak, B., and Nabel, G.J. (1992). Activation of the human immunodeficiency virus type 2 enhancer is dependent on purine box and kappa B regulatory elements. *J Virol* 66(9), 5479-5484.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal* 17(1).
- Marx, P.A., Alcibes, P.G., and Drucker, E. (2001). Serial human passage of simian immunodeficiency virus by unsterile injections and the emergence of epidemic human immunodeficiency virus in Africa. *Philos Trans R Soc Lond B Biol Sci* 356(1410), 911-920. doi: 10.1098/rstb.2001.0867.
- Masuda, F., Ishii, M., Mori, A., Uehara, L., Yanagida, M., Takeda, K., et al. (2016). Glucose restriction induces transient G2 cell cycle arrest extending cellular chronological lifespan. *Sci Rep* 6, 19629. doi: 10.1038/srep19629.
- Matsumura, T., Kametani, Y., Ando, K., Hirano, Y., Katano, I., Ito, R., et al. (2003). Functional CD5+ B cells develop predominantly in the spleen of NOD/SCID/gammac(null) (NOG) mice transplanted either with human umbilical cord blood, bone marrow, or mobilized peripheral blood CD34+ cells. *Exp Hematol* 31(9), 789-797. doi: 10.1016/s0301-472x(03)00193-0.
- Mazurier, F., Fontanellas, A., Salesse, S., Taine, L., Landriau, S., Moreau-Gaudry, F., et al. (1999). A novel immunodeficient mouse model--RAG2 x common cytokine receptor gamma chain double mutants--requiring exogenous cytokine administration for human hematopoietic stem cell engraftment. *J Interferon Cytokine Res* 19(5), 533-541. doi: 10.1089/107999099313983.
- Mbisa, J.L., Barr, R., Thomas, J.A., Vandegraaff, N., Dorweiler, I.J., Svarovskaia, E.S., et al. (2007). Human immunodeficiency virus type 1 cDNAs produced in the presence of APOBEC3G exhibit defects in plus-strand DNA transfer and integration. *J Virol* 81(13), 7099-7110. doi: 10.1128/JVI.00272-07.
- McDermott, S.P., Eppert, K., Lechman, E.R., Doedens, M., and Dick, J.E. (2010). Comparison of human cord blood engraftment between immunocompromised mouse strains. *Blood* 116(2), 193-200. doi: 10.1182/blood-2010-02-271841.

- McNatt, M.W., Zang, T., and Bieniasz, P.D. (2013). Vpu binds directly to tetherin and displaces it from nascent virions. *PLoS Pathog* 9(4), e1003299. doi: 10.1371/journal.ppat.1003299.
- Melkus, M.W., Estes, J.D., Padgett-Thomas, A., Gatlin, J., Denton, P.W., Othieno, F.A., et al. (2006). Humanized mice mount specific adaptive and innate immune responses to EBV and TSST-1. *Nat Med* 12(11), 1316-1322. doi: 10.1038/nm1431.
- Meyerson, N.R., Warren, C.J., Vieira, D., Diaz-Griffero, F., and Sawyer, S.L. (2018). Species-specific vulnerability of RanBP2 shaped the evolution of SIV as it transmitted in African apes. *PLoS Pathog* 14(3), e1006906. doi: 10.1371/journal.ppat.1006906.
- Miller, M.D., Farnet, C.M., and Bushman, F.D. (1997). Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol* 71(7), 5382-5390.
- Miller, M.D., Warmerdam, M.T., Gaston, I., Greene, W.C., and Feinberg, M.B. (1994). The human immunodeficiency virus-1 nef gene product: a positive factor for viral infection and replication in primary lymphocytes and macrophages. *J Exp Med* 179(1), 101-113. doi: 10.1084/jem.179.1.101.
- Miller, R.H. (1988). Human immunodeficiency virus may encode a novel protein on the genomic DNA plus strand. *Science* 239(4846), 1420-1422. doi: 10.1126/science.3347840.
- Mitani, J.C., and Watts, D.P. (1999). Demographic influences on the hunting behavior of chimpanzees. *Am J Phys Anthropol* 109(4), 439-454. doi: 10.1002/(SICI)1096-8644(199908)109:4<439::AID-AJPA2>3.0.CO;2-3.
- Mlcochova, P., Sutherland, K.A., Watters, S.A., Bertoli, C., de Bruin, R.A., Rehwinkel, J., et al. (2017). A G1-like state allows HIV-1 to bypass SAMHD1 restriction in macrophages. *The EMBO journal* 36(5), 604-616.
- Mori, K., Ringler, D.J., Kodama, T., and Desrosiers, R.C. (1992). Complex determinants of macrophage tropism in env of simian immunodeficiency virus. *J Virol* 66(4), 2067-2075. doi: 10.1128/JVI.66.4.2067-2075.1992.
- Morner, A., Bjorndal, A., Albert, J., Kewalramani, V.N., Littman, D.R., Inoue, R., et al. (1999). Primary human immunodeficiency virus type 2 (HIV-2) isolates, like HIV-1 isolates, frequently use CCR5 but show promiscuity in coreceptor usage. *J Virol* 73(3), 2343-2349.
- Mosier, D.E., Gulizia, R.J., Baird, S.M., and Wilson, D.B. (1988). Transfer of a functional human immune system to mice with severe combined immunodeficiency. *Nature* 335(6187), 256-259. doi: 10.1038/335256a0.
- Munch, J., Schindler, M., Wildum, S., Rucker, E., Bailer, N., Knoop, V., et al. (2005). Primary sooty mangabey simian immunodeficiency virus and human immunodeficiency virus type 2 nef alleles modulate cell surface expression of various human receptors and enhance viral infectivity and replication. *J Virol* 79(16), 10547-10560. doi: 10.1128/JVI.79.16.10547-10560.2005.
- Murphey-Corb, M., Martin, L.N., Rangan, S.R., Baskin, G.B., Gormus, B.J., Wolf, R.H., et al. (1986). Isolation of an HTLV-III-related retrovirus from macaques with simian AIDS and its possible origin in asymptomatic mangabeys. *Nature* 321(6068), 435-437. doi: 10.1038/321435a0.

- Nagata, S., Imai, J., Makino, G., Tomita, M., and Kanai, A. (2017). Evolutionary Analysis of HIV-1 Pol Proteins Reveals Representative Residues for Viral Subtype Differentiation. *Front Microbiol* 8, 2151. doi: 10.3389/fmicb.2017.02151.
- Naidu, Y.M., Kestler, H.W., 3rd, Li, Y., Butler, C.V., Silva, D.P., Schmidt, D.K., et al. (1988). Characterization of infectious molecular clones of simian immunodeficiency virus (SIVmac) and human immunodeficiency virus type 2: persistent infection of rhesus monkeys with molecularly cloned SIVmac. *J Virol* 62(12), 4691-4696.
- Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3(5), 418-426. doi: 10.1093/oxfordjournals.molbev.a040410.
- Neil, S.J., Eastman, S.W., Jouvenet, N., and Bieniasz, P.D. (2006). HIV-1 Vpu promotes release and prevents endocytosis of nascent retrovirus particles from the plasma membrane. *PLoS Pathog* 2(5), e39. doi: 10.1371/journal.ppat.0020039.
- Neil, S.J., Zang, T., and Bieniasz, P.D. (2008). Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* 451(7177), 425-430. doi: 10.1038/nature06553.
- Nelson, C.W., and Hughes, A.L. (2015). Within-host nucleotide diversity of virus populations: insights from next-generation sequencing. *Infect Genet Evol* 30, 1-7. doi: 10.1016/j.meegid.2014.11.026.
- Nelson, C.W., Moncla, L.H., and Hughes, A.L. (2015). SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. *Bioinformatics* 31(22), 3709-3711. doi: 10.1093/bioinformatics/btv449.
- Nguyen, D.H., and Hildreth, J.E. (2000). Evidence for budding of human immunodeficiency virus type 1 selectively from glycolipid-enriched membrane lipid rafts. *J Virol* 74(7), 3264-3272. doi: 10.1128/jvi.74.7.3264-3272.2000.
- Noviello, C.M., Benichou, S., and Guatelli, J.C. (2008). Cooperative binding of the class I major histocompatibility complex cytoplasmic domain and human immunodeficiency virus type 1 Nef to the endosomal AP-1 complex via its mu subunit. *J Virol* 82(3), 1249-1258. doi: 10.1128/JVI.00660-07.
- O'Neil, S.P., Novembre, F.J., Hill, A.B., Suwyn, C., Hart, C.E., Evans-Strickfaden, T., et al. (2000). Progressive infection in a subset of HIV-1-positive chimpanzees. *J Infect Dis* 182(4), 1051-1062. doi: 10.1086/315823.
- Pancio, H.A., and Ratner, L. (1998). Human immunodeficiency virus type 2 Vpx-Gag interaction. *J Virol* 72(6), 5271-5275.
- Pedersen, N.C., Ho, E.W., Brown, M.L., and Yamamoto, J.K. (1987). Isolation of a T-lymphotropic virus from domestic cats with an immunodeficiency-like syndrome. *Science* 235(4790), 790-793. doi: 10.1126/science.3643650.
- Perez-Caballero, D., Zang, T., Ebrahimi, A., McNatt, M.W., Gregory, D.A., Johnson, M.C., et al. (2009). Tetherin inhibits HIV-1 release by directly tethering virions to cells. *Cell* 139(3), 499-511. doi: 10.1016/j.cell.2009.08.039.

- Pertel, T., Hausmann, S., Morger, D., Zuger, S., Guerra, J., Lascano, J., et al. (2011). TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* 472(7343), 361-365. doi: 10.1038/nature09976.
- Pettit, S.C., Moody, M.D., Wehbie, R.S., Kaplan, A.H., Nantermet, P.V., Klein, C.A., et al. (1994). The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J Virol* 68(12), 8017-8027. doi: 10.1128/JVI.68.12.8017-8027.1994.
- Plantier, J.C., Leoz, M., Dickerson, J.E., De Oliveira, F., Cordonnier, F., Lemeé, V., et al. (2009). A new human immunodeficiency virus derived from gorillas. *Nat Med* 15(8), 871-872. doi: 10.1038/nm.2016.
- Platt, E.J., Bilaska, M., Kozak, S.L., Kabat, D., and Montefiori, D.C. (2009). Evidence that ecotropic murine leukemia virus contamination in TZM-bl cells does not affect the outcome of neutralizing antibody assays with human immunodeficiency virus type 1. *J Virol* 83(16), 8289-8292. doi: 10.1128/JVI.00709-09.
- Platt, E.J., Wehrly, K., Kuhmann, S.E., Chesebro, B., and Kabat, D. (1998). Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *J Virol* 72(4), 2855-2864.
- Poe, J.A., and Smithgall, T.E. (2009). HIV-1 Nef dimerization is required for Nef-mediated receptor downregulation and viral replication. *J Mol Biol* 394(2), 329-342. doi: 10.1016/j.jmb.2009.09.047.
- Pollard, V.W., and Malim, M.H. (1998). The HIV-1 Rev protein. *Annu Rev Microbiol* 52, 491-532. doi: 10.1146/annurev.micro.52.1.491.
- Popovic, M., Sarngadharan, M.G., Read, E., and Gallo, R.C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* 224(4648), 497-500. doi: 10.1126/science.6200935.
- Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D., and Williamson, J.R. (1992). Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* 257(5066), 76-80. doi: 10.1126/science.1621097.
- Qu, D., Li, C., Sang, F., Li, Q., Jiang, Z.Q., Xu, L.R., et al. (2016). The variances of Sp1 and NF-kappaB elements correlate with the greater capacity of Chinese HIV-1 B'-LTR for driving gene expression. *Sci Rep* 6, 34532. doi: 10.1038/srep34532.
- Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., et al. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 12(6), 1261-1276. doi: 10.1038/nprot.2017.066.
- Raposo, G., Moore, M., Innes, D., Leijendekker, R., Leigh-Brown, A., Benaroch, P., et al. (2002). Human macrophages accumulate HIV-1 particles in MHC II compartments. *Traffic* 3(10), 718-729. doi: 10.1034/j.1600-0854.2002.31004.x.

- Rhee, S.S., and Marsh, J.W. (1994). Human immunodeficiency virus type 1 Nef-induced down-modulation of CD4 is due to rapid internalization and degradation of surface CD4. *J Virol* 68(8), 5156-5163.
- Rhim, H., and Rice, A.P. (1994). Exon2 of HIV-2 Tat contributes to transactivation of the HIV-2 LTR by increasing binding affinity to HIV-2 TAR RNA. *Nucleic Acids Res* 22(21), 4405-4413. doi: 10.1093/nar/22.21.4405.
- Roeth, J.F., Williams, M., Kasper, M.R., Filzen, T.M., and Collins, K.L. (2004). HIV-1 Nef disrupts MHC-I trafficking by recruiting AP-1 to the MHC-I cytoplasmic tail. *J Cell Biol* 167(5), 903-913. doi: 10.1083/jcb.200407031.
- Romani, B., Shaykh Baygloo, N., Aghasadeghi, M.R., and Allahbakhshi, E. (2015). HIV-1 Vpr Protein Enhances Proteasomal Degradation of MCM10 DNA Replication Factor through the Cul4-DDB1[VprBP] E3 Ubiquitin Ligase to Induce G2/M Cell Cycle Arrest. *J Biol Chem* 290(28), 17380-17389. doi: 10.1074/jbc.M115.641522.
- Rosa, A., Chande, A., Ziglio, S., De Sanctis, V., Bertorelli, R., Goh, S.L., et al. (2015). HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature* 526(7572), 212-217. doi: 10.1038/nature15399.
- Rowell, J.F., Stanhope, P.E., and Siliciano, R.F. (1995). Endocytosis of endogenously synthesized HIV-1 envelope protein. Mechanism and role in processing for association with class II MHC. *J Immunol* 155(1), 473-488.
- Rowland-Jones, S.L. (2003). Timeline: AIDS pathogenesis: what have two decades of HIV research taught us? *Nat Rev Immunol* 3(4), 343-348. doi: 10.1038/nri1058.
- Sandstrom, P.A., Pardi, D., Goldsmith, C.S., Chengying, D., Diamond, A.M., and Folks, T.M. (1996). bc1-2 expression facilitates human immunodeficiency virus type-1 mediated cytopathic effects during acute spreading infections. *J Virol* 70(7), 4617-4622.
- Sanjuan, R., Moya, A., and Elena, S.F. (2004). The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci U S A* 101(22), 8396-8401. doi: 10.1073/pnas.0400146101.
- Santiago, M.L., Range, F., Keele, B.F., Li, Y., Bailes, E., Bibollet-Ruche, F., et al. (2005). Simian immunodeficiency virus infection in free-ranging sooty mangabeys (*Cercocebus atys atys*) from the Tai Forest, Cote d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. *J Virol* 79(19), 12515-12527. doi: 10.1128/JVI.79.19.12515-12527.2005.
- Sato, K., Misawa, N., Takeuchi, J.S., Kobayashi, T., Izumi, T., Aso, H., et al. (2018). Experimental Adaptive Evolution of Simian Immunodeficiency Virus SIVcpz to Pandemic Human Immunodeficiency Virus Type 1 by Using a Humanized Mouse Model. *J Virol* 92(4). doi: 10.1128/JVI.01905-17.
- Sato, S., Yuste, E., Lauer, W.A., Chang, E.H., Morgan, J.S., Bixby, J.G., et al. (2008). Potent antibody-mediated neutralization and evolution of antigenic escape variants of simian immunodeficiency virus strain SIVmac239 in vivo. *J Virol* 82(19), 9739-9752. doi: 10.1128/JVI.00871-08.

- Saunders, C.J., McCaffrey, R.A., Zharkikh, I., Kraft, Z., Malenbaum, S.E., Burke, B., et al. (2005). The V1, V2, and V3 regions of the human immunodeficiency virus type 1 envelope differentially affect the viral phenotype in an isolate-dependent manner. *J Virol* 79(14), 9069-9080. doi: 10.1128/JVI.79.14.9069-9080.2005.
- Sauter, D., and Kirchhoff, F. (2019). Key Viral Adaptations Preceding the AIDS Pandemic. *Cell Host Microbe* 25(1), 27-38. doi: 10.1016/j.chom.2018.12.002.
- Sauter, D., Schindler, M., Specht, A., Landford, W.N., Munch, J., Kim, K.A., et al. (2009). Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. *Cell Host Microbe* 6(5), 409-421. doi: 10.1016/j.chom.2009.10.004.
- Sauter, D., Unterweger, D., Vogl, M., Usmani, S.M., Heigele, A., Kluge, S.F., et al. (2012). Human tetherin exerts strong selection pressure on the HIV-1 group N Vpu protein. *PLoS Pathog* 8(12), e1003093. doi: 10.1371/journal.ppat.1003093.
- Sawyer, S.L., Wu, L.I., Emerman, M., and Malik, H.S. (2005). Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proc Natl Acad Sci U S A* 102(8), 2832-2837. doi: 10.1073/pnas.0409853102.
- Sayah, D.M., Sokolskaja, E., Berthoux, L., and Luban, J. (2004). Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* 430(6999), 569-573. doi: 10.1038/nature02777.
- Schindler, M., Munch, J., Kutsch, O., Li, H., Santiago, M.L., Bibollet-Ruche, F., et al. (2006). Nef-mediated suppression of T cell activation was lost in a lentiviral lineage that gave rise to HIV-1. *Cell* 125(6), 1055-1067. doi: 10.1016/j.cell.2006.04.033.
- Schmitt, K., Curlin, J., Kumar, D.M., Remling-Mulder, L., Feely, S., Stenglein, M., et al. (2018). SIV progenitor evolution toward HIV: A humanized mouse surrogate model for SIVsm adaptation toward HIV-2. *J Med Primatol* 47(5), 298-301. doi: 10.1111/jmp.12380.
- Schmitt, K., Curlin, J., Remling-Mulder, L., Moriarty, R., Goff, K., O'Connor, S., et al. (2020a). Cross-Species Transmission and Evolution of SIV Chimpanzee Progenitor Viruses Toward HIV-1 in Humanized Mice. *Frontiers in Microbiology* 11(1889). doi: 10.3389/fmicb.2020.01889.
- Schmitt, K., Curlin, J., Remling-Mulder, L., Moriarty, R., Goff, K., O'Connor, S., et al. (2020b). Mimicking SIV chimpanzee viral evolution toward HIV-1 during cross-species transmission. *Journal of Medical Primatology*.
- Schmitt, K., Mohan Kumar, D., Curlin, J., Remling-Mulder, L., Stenglein, M., O'Connor, S., et al. (2017). Modeling the evolution of SIV sooty mangabey progenitor virus towards HIV-2 using humanized mice. *Virology* 510, 175-184. doi: 10.1016/j.virol.2017.07.005.
- Schrager, C.G., and Russo, C.A. (2003). Timing the origin of New World monkeys. *Mol Biol Evol* 20(10), 1620-1625. doi: 10.1093/molbev/msg172.
- Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110(4), 521-529. doi: 10.1016/s0092-8674(02)00864-4.

- Schubert, H.L., Zhai, Q., Sandrin, V., Eckert, D.M., Garcia-Maya, M., Saul, L., et al. (2010). Structural and functional studies on the extracellular domain of BST2/tetherin in reduced and oxidized conformations. *Proc Natl Acad Sci U S A* 107(42), 17951-17956. doi: 10.1073/pnas.1008206107.
- Schwartz, O., Marechal, V., Danos, O., and Heard, J.M. (1995). Human immunodeficiency virus type 1 Nef increases the efficiency of reverse transcription in the infected cell. *J Virol* 69(7), 4053-4059. doi: 10.1128/JVI.69.7.4053-4059.1995.
- Schwartz, O., Marechal, V., Le Gall, S., Lemonnier, F., and Heard, J.M. (1996). Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. *Nat Med* 2(3), 338-342. doi: 10.1038/nm0396-338.
- Selby, M.J., Bain, E.S., Luciw, P.A., and Peterlin, B.M. (1989). Structure, sequence, and position of the stem-loop in tar determine transcriptional elongation by tat through the HIV-1 long terminal repeat. *Genes Dev* 3(4), 547-558. doi: 10.1101/gad.3.4.547.
- Serra-Moreno, R., and Evans, D.T. (2012). Adaptation of human and simian immunodeficiency viruses for resistance to tetherin/BST-2. *Curr HIV Res* 10(4), 277-282. doi: 10.2174/157016212800792496.
- Serra-Moreno, R., Jia, B., Breed, M., Alvarez, X., and Evans, D.T. (2011). Compensatory changes in the cytoplasmic tail of gp41 confer resistance to tetherin/BST-2 in a pathogenic nef-deleted SIV. *Cell Host Microbe* 9(1), 46-57. doi: 10.1016/j.chom.2010.12.005.
- Serra-Moreno, R., Zimmermann, K., Stern, L.J., and Evans, D.T. (2013). Tetherin/BST-2 antagonism by Nef depends on a direct physical interaction between Nef and tetherin, and on clathrin-mediated endocytosis. *PLoS Pathog* 9(7), e1003487. doi: 10.1371/journal.ppat.1003487.
- Sharp, P.M., Bailes, E., Stevenson, M., Emerman, M., and Hahn, B.H. (1996). Gene acquisition in HIV and SIV. *Nature* 383(6601), 586-587. doi: 10.1038/383586a0.
- Sharp, P.M., and Hahn, B.H. (2011). Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med* 1(1), a006841. doi: 10.1101/cshperspect.a006841.
- Sheehy, A.M., Gaddis, N.C., Choi, J.D., and Malim, M.H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 418(6898), 646-650. doi: 10.1038/nature00939.
- Sheehy, A.M., Gaddis, N.C., and Malim, M.H. (2003). The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* 9(11), 1404-1407. doi: 10.1038/nm945.
- Sherman, M.P., de Noronha, C.M., Pearce, D., and Greene, W.C. (2000). Human immunodeficiency virus type 1 Vpr contains two leucine-rich helices that mediate glucocorticoid receptor coactivation independently of its effects on G(2) cell cycle arrest. *J Virol* 74(17), 8159-8165. doi: 10.1128/jvi.74.17.8159-8165.2000.
- Shimajima, M., Miyazawa, T., Ikeda, Y., McMonagle, E.L., Haining, H., Akashi, H., et al. (2004). Use of CD134 as a primary receptor by the feline immunodeficiency virus. *Science* 303(5661), 1192-1195. doi: 10.1126/science.1092124.

- Shultz, L.D., Brehm, M.A., Garcia-Martinez, J.V., and Greiner, D.L. (2012). Humanized mice for immune system investigation: progress, promise and challenges. *Nat Rev Immunol* 12(11), 786-798. doi: 10.1038/nri3311.
- Simmons, A., Aluvihare, V., and McMichael, A. (2001). Nef triggers a transcriptional program in T cells imitating single-signal T cell activation and inducing HIV virulence mediators. *Immunity* 14(6), 763-777. doi: 10.1016/s1074-7613(01)00158-3.
- Simon, J.H., Carpenter, E.A., Fouchier, R.A., and Malim, M.H. (1999). Vif and the p55(Gag) polyprotein of human immunodeficiency virus type 1 are present in colocalizing membrane-free cytoplasmic complexes. *J Virol* 73(4), 2667-2674. doi: 10.1128/JVI.73.4.2667-2674.1999.
- Simon, V., Bloch, N., and Landau, N.R. (2015). Intrinsic host restrictions to HIV-1 and mechanisms of viral escape. *Nat Immunol* 16(6), 546-553. doi: 10.1038/ni.3156.
- Smith, S.M., Christian, D., de Lame, V., Shah, U., Austin, L., Gautam, R., et al. (2008). Isolation of a new HIV-2 group in the US. *Retrovirology* 5, 103. doi: 10.1186/1742-4690-5-103.
- Sood, C., Marin, M., Chande, A., Pizzato, M., and Melikyan, G.B. (2017). SERINC5 protein inhibits HIV-1 fusion pore formation by promoting functional inactivation of envelope glycoproteins. *J Biol Chem* 292(14), 6014-6026. doi: 10.1074/jbc.M117.777714.
- Srivastava, S., Swanson, S.K., Manel, N., Florens, L., Washburn, M.P., and Skowronski, J. (2008). Lentiviral Vpx accessory factor targets VprBP/DCAF1 substrate adaptor for cullin 4 E3 ubiquitin ligase to enable macrophage infection. *PLoS Pathog* 4(5), e1000059. doi: 10.1371/journal.ppat.1000059.
- Steckbeck, J.D., Craig, J.K., Barnes, C.O., and Montelaro, R.C. (2011). Highly conserved structural properties of the C-terminal tail of HIV-1 gp41 protein despite substantial sequence variation among diverse clades: implications for functions in viral replication. *J Biol Chem* 286(31), 27156-27166. doi: 10.1074/jbc.M111.258855.
- Strebel, K. (2013). HIV accessory proteins versus host restriction factors. *Curr Opin Virol* 3(6), 692-699. doi: 10.1016/j.coviro.2013.08.004.
- Stremlau, M., Owens, C.M., Perron, M.J., Kiessling, M., Autissier, P., and Sodroski, J. (2004). The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* 427(6977), 848-853. doi: 10.1038/nature02343.
- Stremlau, M., Perron, M., Lee, M., Li, Y., Song, B., Javanbakht, H., et al. (2006). Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. *Proc Natl Acad Sci U S A* 103(14), 5514-5519. doi: 10.1073/pnas.0509996103.
- Stremlau, M., Perron, M., Welikala, S., and Sodroski, J. (2005). Species-specific variation in the B30.2(SPRY) domain of TRIM5alpha determines the potency of human immunodeficiency virus restriction. *J Virol* 79(5), 3139-3145. doi: 10.1128/JVI.79.5.3139-3145.2005.
- Sundquist, W.I., and Krausslich, H.G. (2012). HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med* 2(7), a006924. doi: 10.1101/cshperspect.a006924.

- Swigut, T., Iafrate, A.J., Muench, J., Kirchhoff, F., and Skowronski, J. (2000). Simian and human immunodeficiency virus Nef proteins use different surfaces to downregulate class I major histocompatibility complex antigen expression. *J Virol* 74(12), 5691-5701. doi: 10.1128/jvi.74.12.5691-5701.2000.
- Symons, J., Cameron, P.U., and Lewin, S.R. (2018). HIV integration sites and implications for maintenance of the reservoir. *Curr Opin HIV AIDS* 13(2), 152-159. doi: 10.1097/COH.0000000000000438.
- Takeuchi, Y., McClure, M.O., and Pizzato, M. (2008). Identification of gammaretroviruses constitutively released from cell lines used for human immunodeficiency virus research. *J Virol* 82(24), 12585-12588. doi: 10.1128/JVI.01726-08.
- Tong-Starksen, S.E., Welsh, T.M., and Peterlin, B.M. (1990). Differences in transcriptional enhancers of HIV-1 and HIV-2. Response to T cell activation signals. *J Immunol* 145(12), 4348-4354.
- Torresilla, C., Mesnard, J.M., and Barbeau, B. (2015). Reviving an old HIV-1 gene: the HIV-1 antisense protein. *Curr HIV Res* 13(2), 117-124. doi: 10.2174/1570162x12666141202125943.
- Torten, M., Franchini, M., Barlough, J.E., George, J.W., Mozes, E., Lutz, H., et al. (1991). Progressive immune dysfunction in cats experimentally infected with feline immunodeficiency virus. *J Virol* 65(5), 2225-2230. doi: 10.1128/JVI.65.5.2225-2230.1991.
- Ueda, T., Yoshino, H., Kobayashi, K., Kawahata, M., Ebihara, Y., Ito, M., et al. (2000). Hematopoietic repopulating ability of cord blood CD34(+) cells in NOD/Shi-scid mice. *Stem Cells* 18(3), 204-213. doi: 10.1634/stemcells.18-3-204.
- Usami, Y., Wu, Y., and Gottlinger, H.G. (2015). SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature* 526(7572), 218-223. doi: 10.1038/nature15400.
- Vallari, A., Bodelle, P., Ngansop, C., Makamche, F., Ndembi, N., Mbanya, D., et al. (2010). Four new HIV-1 group N isolates from Cameroon: Prevalence continues to be low. *AIDS Res Hum Retroviruses* 26(1), 109-115. doi: 10.1089/aid.2009.0178.
- Van Damme, N., Goff, D., Katsura, C., Jorgenson, R.L., Mitchell, R., Johnson, M.C., et al. (2008). The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. *Cell Host Microbe* 3(4), 245-252. doi: 10.1016/j.chom.2008.03.001.
- Van Heuverswyn, F., Li, Y., Neel, C., Bailes, E., Keele, B.F., Liu, W., et al. (2006). Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature* 444(7116), 164. doi: 10.1038/444164a.
- Veillette, M., Desormeaux, A., Medjahed, H., Gharsallah, N.E., Coutu, M., Baalwa, J., et al. (2014). Interaction with cellular CD4 exposes HIV-1 envelope epitopes targeted by antibody-dependent cell-mediated cytotoxicity. *J Virol* 88(5), 2633-2644. doi: 10.1128/JVI.03230-13.
- Veselinovic, M., Charlins, P., and Akkina, R. (2016). Modeling HIV-1 Mucosal Transmission and Prevention in Humanized Mice. *Methods Mol Biol* 1354, 203-220. doi: 10.1007/978-1-4939-3046-3_14.

- Vigan, R., and Neil, S.J. (2010). Determinants of tetherin antagonism in the transmembrane domain of the human immunodeficiency virus type 1 Vpu protein. *J Virol* 84(24), 12958-12970. doi: 10.1128/JVI.01699-10.
- Visseaux, B., Damond, F., Matheron, S., Descamps, D., and Charpentier, C. (2016). Hiv-2 molecular epidemiology. *Infect Genet Evol* 46, 233-240. doi: 10.1016/j.meegid.2016.08.010.
- Wain, L.V., Bailes, E., Bibollet-Ruche, F., Decker, J.M., Keele, B.F., Van Heuverswyn, F., et al. (2007). Adaptation of HIV-1 to its human host. *Mol Biol Evol* 24(8), 1853-1860. doi: 10.1093/molbev/msm110.
- Walker, R.C., Jr., Khan, M.A., Kao, S., Goila-Gaur, R., Miyagi, E., and Strebel, K. (2010). Identification of dominant negative human immunodeficiency virus type 1 Vif mutants that interfere with the functional inactivation of APOBEC3G by virus-encoded Vif. *J Virol* 84(10), 5201-5211. doi: 10.1128/JVI.02318-09.
- Watanabe, Y., Takahashi, T., Okajima, A., Shiokawa, M., Ishii, N., Katano, I., et al. (2009). The analysis of the functions of human B and T cells in humanized NOD/shi-scid/gammac(null) (NOG) mice (hu-HSC NOG mice). *Int Immunol* 21(7), 843-858. doi: 10.1093/intimm/dxp050.
- Wei, B.L., Arora, V.K., Raney, A., Kuo, L.S., Xiao, G.H., O'Neill, E., et al. (2005). Activation of p21-activated kinase 2 by human immunodeficiency virus type 1 Nef induces merlin phosphorylation. *J Virol* 79(23), 14976-14980. doi: 10.1128/JVI.79.23.14976-14980.2005.
- Wei, W., Guo, H., Han, X., Liu, X., Zhou, X., Zhang, W., et al. (2012). A novel DCAF1-binding motif required for Vpx-mediated degradation of nuclear SAMHD1 and Vpr-induced G2 arrest. *Cell Microbiol* 14(11), 1745-1756. doi: 10.1111/j.1462-5822.2012.01835.x.
- Wei, X., Decker, J.M., Liu, H., Zhang, Z., Arani, R.B., Kilby, J.M., et al. (2002). Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob Agents Chemother* 46(6), 1896-1905.
- Wei, X., Decker, J.M., Wang, S., Hui, H., Kappes, J.C., Wu, X., et al. (2003). Antibody neutralization and escape by HIV-1. *Nature* 422(6929), 307-312. doi: 10.1038/nature01470.
- Wetzel, K.S., Elliott, S.T.C., and Collman, R.G. (2018). SIV Coreceptor Specificity in Natural and Non-Natural Host Infection: Implications for Cell Targeting and Differential Outcomes from Infection. *Curr HIV Res* 16(1), 41-51. doi: 10.2174/1570162X15666171124121805.
- Wilm, A., Aw, P.P., Bertrand, D., Yeo, G.H., Ong, S.H., Wong, C.H., et al. (2012). LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res* 40(22), 11189-11201. doi: 10.1093/nar/gks918.
- Wlodawer, A., and Erickson, J.W. (1993). Structure-based inhibitors of HIV-1 protease. *Annu Rev Biochem* 62, 543-585. doi: 10.1146/annurev.bi.62.070193.002551.
- Wolf, D., Witte, V., Laffert, B., Blume, K., Stromer, E., Trapp, S., et al. (2001). HIV-1 Nef associated PAK and PI3-kinases stimulate Akt-independent Bad-phosphorylation to induce anti-apoptotic signals. *Nat Med* 7(11), 1217-1224. doi: 10.1038/nm1101-1217.

- Wonderlich, E.R., Williams, M., and Collins, K.L. (2008). The tyrosine binding pocket in the adaptor protein 1 (AP-1) μ 1 subunit is necessary for Nef to recruit AP-1 to the major histocompatibility complex class I cytoplasmic tail. *J Biol Chem* 283(6), 3011-3022. doi: 10.1074/jbc.M707760200.
- Wooley, D.P., Smith, R.A., Czajak, S., and Desrosiers, R.C. (1997). Direct demonstration of retroviral recombination in a rhesus monkey. *J Virol* 71(12), 9650-9653. doi: 10.1128/JVI.71.12.9650-9653.1997.
- Worobey, M., Gemmel, M., Teuwen, D.E., Haselkorn, T., Kunstman, K., Bunce, M., et al. (2008). Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* 455(7213), 661-664. doi: 10.1038/nature07390.
- Worobey, M., Telfer, P., Souquiere, S., Hunter, M., Coleman, C.A., Metzger, M.J., et al. (2010). Island biogeography reveals the deep history of SIV. *Science* 329(5998), 1487. doi: 10.1126/science.1193550.
- Wu, F., Kirmaier, A., Goeken, R., Ourmanov, I., Hall, L., Morgan, J.S., et al. (2013). TRIM5 α drives SIVsmm evolution in rhesus macaques. *PLoS Pathog* 9(8), e1003577. doi: 10.1371/journal.ppat.1003577.
- Wyatt, R., and Sodroski, J. (1998). The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* 280(5371), 1884-1888. doi: 10.1126/science.280.5371.1884.
- Xiang, S.H., Doka, N., Choudhary, R.K., Sodroski, J., and Robinson, J.E. (2002). Characterization of CD4-induced epitopes on the HIV type 1 gp120 envelope glycoprotein recognized by neutralizing human monoclonal antibodies. *AIDS Res Hum Retroviruses* 18(16), 1207-1217. doi: 10.1089/08892220260387959.
- Yang, H., Wang, J., Jia, X., McNatt, M.W., Zang, T., Pan, B., et al. (2010). Structural insight into the mechanisms of enveloped virus tethering by tetherin. *Proc Natl Acad Sci U S A* 107(43), 18428-18432. doi: 10.1073/pnas.1011485107.
- Ye, H., Choi, H.J., Poe, J., and Smithgall, T.E. (2004). Oligomerization is required for HIV-1 Nef-induced activation of the Src family protein-tyrosine kinase, Hck. *Biochemistry* 43(50), 15775-15784. doi: 10.1021/bi048712f.
- Yu, B., Fonseca, D.P., O'Rourke, S.M., and Berman, P.W. (2010). Protease cleavage sites in HIV-1 gp120 recognized by antigen processing enzymes are conserved and located at receptor binding sites. *J Virol* 84(3), 1513-1526. doi: 10.1128/JVI.01765-09.
- Yu, X., Yu, Y., Liu, B., Luo, K., Kong, W., Mao, P., et al. (2003). Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* 302(5647), 1056-1060. doi: 10.1126/science.1089591.
- Yu, X.F., Ito, S., Essex, M., and Lee, T.H. (1988). A naturally immunogenic virion-associated protein specific for HIV-2 and SIV. *Nature* 335(6187), 262-265. doi: 10.1038/335262a0.
- Yuan, Z., Kang, G., Daharsh, L., Fan, W., and Li, Q. (2018). SIVcpz closely related to the ancestral HIV-1 is less or non-pathogenic to humans in a hu-BLT mouse model. *Emerg Microbes Infect* 7(1), 59. doi: 10.1038/s41426-018-0062-9.

- Yuan, Z., Kang, G., Ma, F., Lu, W., Fan, W., Fennessey, C.M., et al. (2016). Recapitulating Cross-Species Transmission of Simian Immunodeficiency Virus SIVcpz to Humans by Using Humanized BLT Mice. *J Virol* 90(17), 7728-7739. doi: 10.1128/JVI.00860-16.
- Yurkovetskiy, L., Guney, M.H., Kim, K., Goh, S.L., McCauley, S., Dauphin, A., et al. (2018). Primate immunodeficiency virus proteins Vpx and Vpr counteract transcriptional repression of proviruses by the HUSH complex. *Nat Microbiol* 3(12), 1354-1361. doi: 10.1038/s41564-018-0256-x.
- Zhang, F., Wilson, S.J., Landford, W.C., Virgen, B., Gregory, D., Johnson, M.C., et al. (2009). Nef proteins from simian immunodeficiency viruses are tetherin antagonists. *Cell Host Microbe* 6(1), 54-67. doi: 10.1016/j.chom.2009.05.008.
- Zhang, Y.Z., and Holmes, E.C. (2020). A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. *Cell*. doi: 10.1016/j.cell.2020.03.035.
- Zheng, Y.H., Lovsin, N., and Peterlin, B.M. (2005). Newly identified host factors modulate HIV replication. *Immunol Lett* 97(2), 225-234. doi: 10.1016/j.imlet.2004.11.026.
- Zhu, T., Korber, B.T., Nahmias, A.J., Hooper, E., Sharp, P.M., and Ho, D.D. (1998). An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* 391(6667), 594-597. doi: 10.1038/35400.

APPENDIX: Supplemental material for chapters 2, 7, and 8

Supplementary Table 2.1. Nucleotide variant alleles that increased in frequency from passage 1 to passage 5.

	Codon Number ^a	Stock Consensus ^b	P5 variant ^c	N or S ^d	Stock virus ^e	P1 Mouse 1385 ^f	P1 Mouse 1304 ^f	P5 Mouse 1949 ^g	P5 Mouse J2436 ^g
<i>gag</i>	12	K	K	S	0.39	0.99	1.00 ⁱ	1.00 ⁱ	0.98
<i>gag</i>	20	R	R	S	0.00 ^h	0.00 ^h	0.00 ^h	0.91	0.83
<i>gag</i>	31	L	L	S	0.00 ^h	0.00 ^h	0.00 ^h	0.66	0.69
<i>gag</i>	35	V	V	S	0.28	0.72	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>gag</i>	76	K	K	S	0.00 ^h	0.00 ^h	0.00 ^h	0.69	0.72
<i>gag</i>	78	L	L	S	0.34	0.65	1.00 ⁱ	1.00 ⁱ	0.97
<i>gag</i>	120	V	A	N	0.41	0.79	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>gag</i>	148	P	P	S	0.26	0.67	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>gag</i>	165	K	K	S	0.26	0.00 ^h	1.00 ⁱ	0.47	0.68
<i>gag</i>	166	F	F	S	0.00 ^h	0.00 ^h	0.00 ^h	0.47	0.62
<i>gag</i>	190	L	L	S	0.26	0.75	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>gag</i>	193	V	V	S	0.29	0.74	1.00 ⁱ	1.00 ⁱ	0.98
<i>gag</i>	217	Q	Q	S	0.27	0.74	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>gag</i>	223	R	T	N	0.00 ^h	0.00 ^h	0.26	1.00 ⁱ	1.00 ⁱ
<i>gag</i>	233	R	T	N	0.00 ^h	0.00 ^h	0.26	1.00 ⁱ	1.00 ⁱ
<i>gag</i>	408	R	R	S	0.00 ^h	0.00 ^h	0.00 ^h	0.52	0.64
<i>gag</i>	430	E	E	S	0.34	0.66	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>gag</i>	474	Y	Y	S	0.21	0.16	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>gag</i>	482	R	K	N	0.00 ^h	0.00 ^h	0.00 ^h	0.56	0.58
<i>gag</i>	483	E	E	S	0.23	0.24	0.99	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	95	L	L	S	0.40	0.11	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	103	I	I	S	0.41	0.13	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	140	K	K	S	0.38	0.11	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	168	L	L	S	0.38	0.15	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	175	V	V	S	0.37	0.10	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	195	P	P	S	0.42	0.73	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	350	L	L	S	0.31	0.31	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	353	Y	Y	S	0.28	0.31	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	365	D	D	S	0.44	0.06	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	418	Q	Q	S	0.41	0.00 ^h	0.99	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	424	V	I	N	0.41	0.00 ^h	0.99	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	516	K	K	S	0.42	0.09	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	527	K	K	S	0.43	0.08	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	560	L	L	S	0.40	0.09	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	639	E	E	S	0.47	0.16	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	648	L	L	S	0.46	0.14	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	662	A	A	S	0.43	0.09	1.00 ⁱ	1.00 ⁱ	0.99

<i>pol</i>	693	E	K	N	0.00 ^h	0.00 ^h	0.00 ^h	0.18	0.23
<i>pol</i>	756	G	G	S	0.52	0.27	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	818	I	I	S	0.35	0.07	0.98	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	828	L	L	S	0.35	0.08	0.97	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	843	H	H	S	0.33	0.11	0.83	0.88	0.87
<i>pol</i>	862	A	T	N	0.00 ^h	0.00 ^h	0.00 ^h	0.14	0.12
<i>pol</i>	889	T	T	S	0.28	0.46	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>pol</i>	894	I	I	S	0.27	0.46	1.00 ⁱ	0.89	1.00 ⁱ
<i>pol</i>	921	G	G	S	0.00 ^h	0.00 ^h	0.18	0.99	0.99
<i>pol</i>	994	A	A	S	0.53	0.32	0.98	1.00 ⁱ	1.00 ⁱ
<i>vif</i>	6	S	N	N	0.39	0.10	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>vif</i>	64	E	E	S	0.28	0.11	0.98	1.00 ⁱ	1.00 ⁱ
<i>vif</i>	74	N	S	S	0.38	0.10	0.98	1.00 ⁱ	1.00 ⁱ
<i>vif</i>	79	K	R	N	0.27	0.06	0.98	1.00 ⁱ	1.00 ⁱ
<i>vif</i>	94	R	R	S	0.28	0.06	0.98	1.00 ⁱ	1.00 ⁱ
<i>vif</i>	136	K	R	N	0.00 ^h	0.00 ^h	0.00 ^h	0.57	0.68
<i>vif</i>	157	S	S	S	0.30	0.09	0.97	1.00 ⁱ	1.00 ⁱ
<i>vpx</i>	17	T	T	S	0.31	0.09	0.97	1.00 ⁱ	1.00 ⁱ
<i>vpx</i>	23	D	D	S	0.26	0.06	0.98	1.00 ⁱ	1.00 ⁱ
<i>vpx</i>	25	L	L	S	0.28	0.07	0.96	0.99	0.96
<i>vpx</i>	69	Y	Y	S	0.46	0.42	0.97	1.00 ⁱ	1.00 ⁱ
<i>vpr</i>	24	L	L	S	0.00 ^h	0.06	0.06	0.69	0.59
<i>vpr</i>	32	L	L	S	0.00 ^h	0.00 ^h	0.00 ^h	0.28	0.39
<i>vpr</i>	82	I	V	N	0.00 ^h	0.00 ^h	0.00 ^h	0.27	0.29
<i>vpr</i>	86	G	R	N	0.00 ^h	0.00 ^h	0.81	0.99	1.00 ⁱ
<i>vpr</i>	99	G	S	N	0.00 ^h	0.00 ^h	0.00 ^h	0.83	0.88
<i>vpr</i>	101	L	L	S	0.30	0.17	0.98	1.00 ⁱ	1.00 ⁱ
<i>env</i>	15	L	V	N	0.25	0.00 ^h	1.00 ⁱ	0.98	1.00 ⁱ
<i>env</i>	20	I	T	N	0.25	0.08	1.00 ⁱ	1.00 ⁱ	0.99
<i>env</i>	21	S	Y	N	0.25	0.09	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	148	V	I	N	0.00 ^h	0.00 ^h	0.00 ^h	0.42	0.46
<i>env</i>	154	P	P	S	0.00 ^h	0.00 ^h	0.00 ^h	0.43	0.36
<i>env</i>	158	N	D	N	0.00 ^h	0.00 ^h	0.06	1.00 ⁱ	1.00 ⁱ
<i>env</i>	205	N	K	N	0.37	0.62	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	222	R	Q	N	0.00 ^h	0.00 ^h	0.13	1.00 ⁱ	1.00 ⁱ
<i>env</i>	249	N	D	N	0.00 ^h	0.08	0.18	1.00 ⁱ	1.00 ⁱ
<i>env</i>	252	N	N	S	0.28	0.58	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	262	K	K	S	0.28	0.50	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	357	G	G	S	0.31	0.64	0.99	1.00 ⁱ	1.00 ⁱ
<i>env</i>	358	A	A	S	0.32	0.65	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	382	N	N	S	0.00 ^h	0.00 ^h	0.21	1.00 ⁱ	1.00 ⁱ
<i>env</i>	383	N	N	S	0.34	0.71	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	393	V	I	N	0.00 ^h	0.00 ^h	0.28	1.00 ⁱ	1.00 ⁱ

<i>env</i>	402	G	G	S	0.00 ^h	0.00 ^h	0.00 ^h	0.57	0.51
<i>env</i>	413	L	L	S	0.35	0.71	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	420 ^J	S	N	N	0.45	0.75	0.99	0.99	0.99
<i>env</i>	420 ^J	S	R	N	0.45	0.75	0.99	0.99	0.99
<i>env</i>	426	R	R	S	0.41	0.75	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	446	I	I	S	0.44	0.80	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	447	N	N	S	0.44	0.80	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	451	R	R	S	0.43	0.81	0.99	1.00 ⁱ	1.00 ⁱ
<i>env</i>	455	N	N	S	0.00 ^h	0.00 ^h	0.00 ^h	0.50	0.51
<i>env</i>	509	E	E	S	0.43	0.39	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	578	D	D	S	0.00 ^h	0.00 ^h	0.00 ^h	0.53	0.80
<i>env</i>	595	K	K	S	0.33	0.08	1.00 ⁱ	0.99	1.00 ⁱ
<i>env</i>	612	A	A	S	0.41	0.12	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	641	N	N	S	0.27	0.07	1.00 ⁱ	0.99	1.00 ⁱ
<i>env</i>	649	E	G	N	0.00 ^h	0.00 ^h	0.73	1.00 ⁱ	1.00 ⁱ
<i>env</i>	658	N	N	S	0.26	0.07	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	663	L	L	S	0.26	0.07	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	673	N	N	S	0.40	0.19	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	697	R	K	N	0.44	0.80	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	715	V	V	S	0.00 ^h	0.00 ^h	0.00 ^h	0.17	0.25
<i>env</i>	719	V	V	S	0.27	0.57	0.97	1.00 ⁱ	1.00 ⁱ
<i>env</i>	720	Q	Q	S	0.32	0.06	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>env</i>	827	E	G	N	0.00 ^h	0.00 ^h	0.00 ^h	0.70	0.79
<i>env</i>	829	T	A	N	0.00 ^h	0.00 ^h	0.15	1.00 ⁱ	0.98
<i>rev</i>	16	H	H	S	0.28	0.12	0.96	0.98	1.00 ⁱ
<i>rev</i>	27	G	R	N	0.00 ^h	0.00 ^h	0.00 ^h	0.37	0.37
<i>rev</i>	34	R	K	N	0.00 ^h	0.00 ^h	0.00 ^h	0.90	0.76
<i>rev</i>	44	R	K	N	0.00 ^h	0.00 ^h	0.68	0.33	0.40
<i>rev</i>	81	E	E	S	0.00 ^h	0.00 ^h	0.00 ^h	0.46	0.47
<i>nef</i>	10	R	H	N	0.31	0.55	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>nef</i>	11	K	R	N	0.34	0.56	1.00 ⁱ	0.99	1.00 ⁱ
<i>nef</i>	12	H	R	N	0.31	0.56	0.97	1.00 ⁱ	1.00 ⁱ
<i>nef</i>	25	G	G	S	0.38	0.64	0.25	1.00 ⁱ	0.99
<i>nef</i>	26	E	K	N	0.00 ^h	0.00 ^h	0.00 ^h	0.09	0.3
<i>nef</i>	45	G	V	N	0.49	0.11	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>nef</i>	51	S	N	N	0.47	0.11	0.96	1.00 ⁱ	1.00 ⁱ
<i>nef</i>	62	E	E	S	0.49	1.00 ⁱ	0.99	1.00 ⁱ	1.00 ⁱ
<i>nef</i>	66	E	K	N	0.00 ^h	0.00 ^h	0.00 ^h	0.41	0.24
<i>nef</i>	76	E	K	N	0.00 ^h	0.00 ^h	0.00 ^h	0.41	0.24
<i>nef</i>	80	L	L	S	0.00 ^h	0.00 ^h	0.00 ^h	0.36	0.48
<i>nef</i>	90	V	V	S	0.00 ^h	0.39	0.15	0.56	0.39
<i>nef</i>	92	D	N	N	0.00 ^h	0.00 ^h	0.09	1.00 ⁱ	0.99
<i>nef</i>	94	D	N	N	0.00 ^h	0.00 ^h	0.00 ^h	0.59	0.44

<i>nef</i>	95	D	N	N	0.36	0.89	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>nef</i>	96	D	N	N	0.00 ^h	0.00 ^h	0.00 ^h	0.32	0.45
<i>nef</i>	105	K	R	N	0.34	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>nef</i>	108	L	L	S	0.22	0.31	1.00 ⁱ	1.00 ⁱ	0.99
<i>nef</i>	110	T	T	S	0.29	0.29	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>nef</i>	218	E	E	S	0.37	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>nef</i>	225	A	A	S	0.41	0.00 ^h	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>nef</i>	228	K	M	N	0.40	0.00 ^h	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ
<i>nef</i>	236	S	S	S	0.42	0.00 ^h	1.00 ⁱ	1.00 ⁱ	1.00 ⁱ

^aCodon number starting at the beginning of each gene

^bConsensus amino acid from stock virus sequencing dataset

^cVariant amino acid in passage 5 (P5) viruses

^dNonsynonymous (N) or Synonymous (S) mutation

^eFrequency of variant mutation in stock virus

^fFrequency of variant mutation relative to stock virus sequence for passage 1 (P1) mice 1385 and 1304

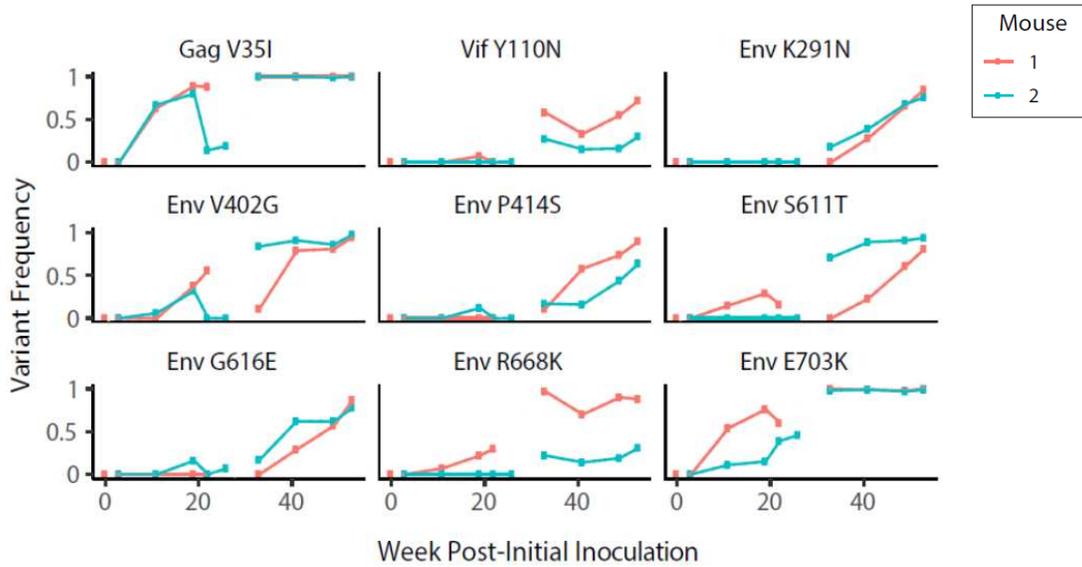
^gFrequency of variant mutation relative to stock virus sequence for passage 5 (P5) mice 1949 and J2436

^h0.00 represents below the limit of detection (5%) of our variant identification pipeline

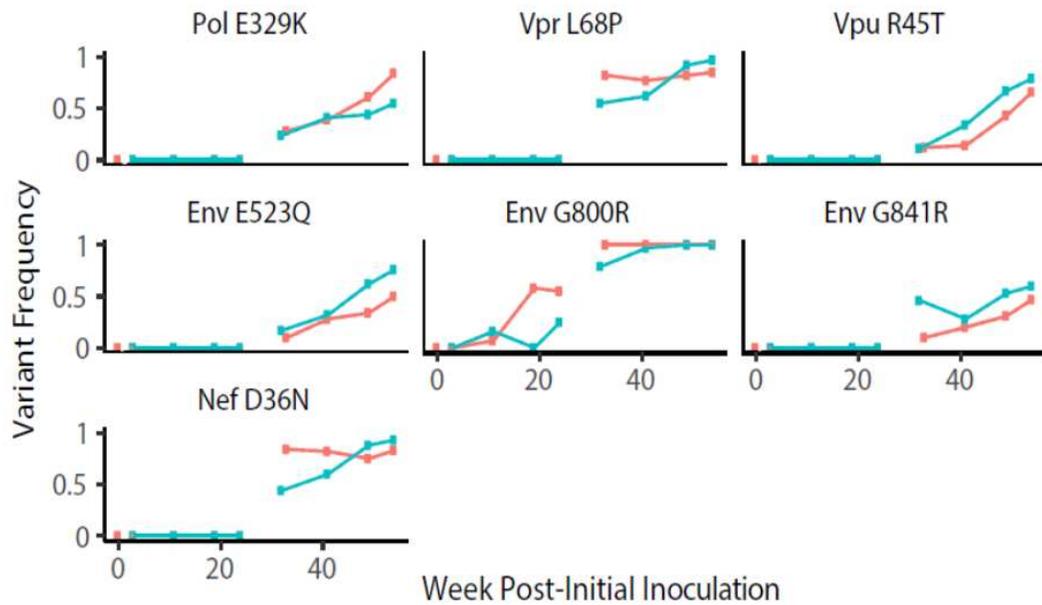
ⁱ1.00 indicates that the frequency of the mutation in the sequence is 100%

^jDinucleotide mutation in the same codon, individual variants are linked.

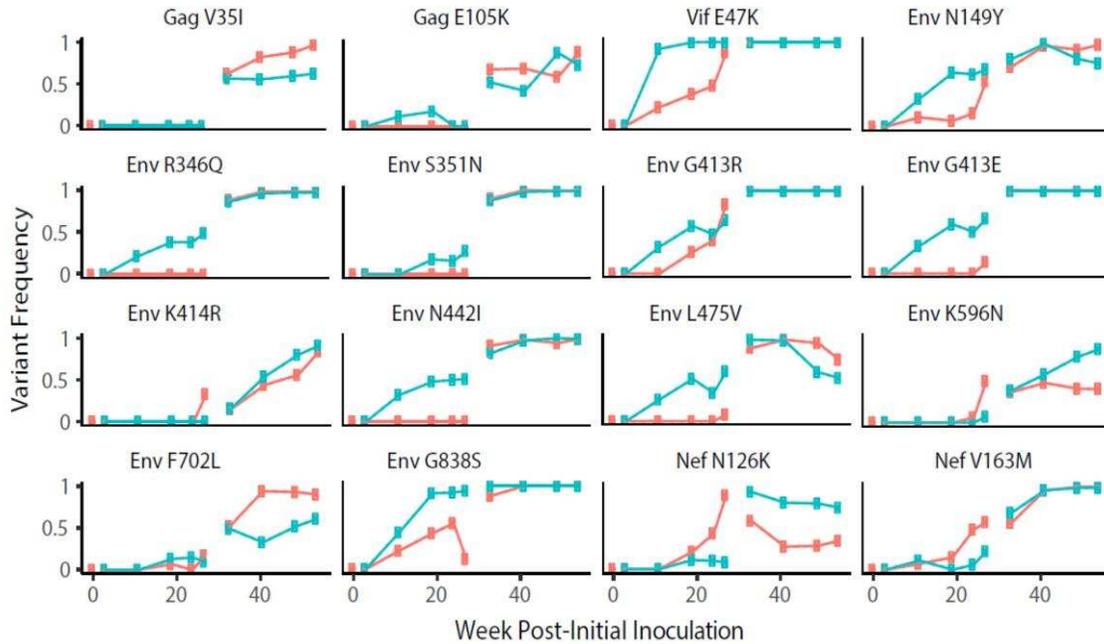
A.



B.



C.



Supplementary Figure 7.1. Consensus changing variants at the end of the second generation. The frequencies of individual variants from SIVcpz (A) EK505, (B) LB715 and (C) MB897 are plotted as a function of weeks post-initial infection. Non-synonymous variants that had a mean frequency of >0.1 by the final time points are shown. The different colors correspond to the frequencies in duplicate mice within the specified generation. Week 0 corresponds to the frequencies in the viral stock pre-infection and the second generation was plotted to start at week 30 (gap in the graph).

Supplementary Table 7.1. SIVcpzMB897 primal scheme primers used for Next-generation sequencing.

Primer Name	Oligonucleotide Sequence (5' to 3')	Oligo-nucleotide Pool	Nucleotide Start Site	Nucleotide End Site
1-LEFT	TCTTGTTAGACCAGATCTGAGCCT	1	5	29
1-RIGHT	TCCTCCCGTTAATACTGACGCT	1	367	345
2-LEFT	GAGAAGTTCTCTCGACGCAGGA	2	218	240
2-RIGHT	TTGATGTACGCACCAGAGGACT	2	604	582
3-LEFT	GGAGCTGGAAAGATTCGCTTGTA	1	467	489
3-RIGHT	AAACATGGGTATGACTTCTGGGC	1	887	865
4-LEFT	GAAACTTCCCGATAGTAACAAATGCTC	2	756	778
4-RIGHT	TGTAAATGTCTCCTACCGGGGT	2	1,121	1,099
5-LEFT	CCAATTCCTCCAGGCCAAATGA	1	997	1,019
5-RIGHT	TCTTCTAAGGTTGCCCTGGAC	1	1,368	1,346
6-LEFT	TGGACAGATTTTACAAAATCTAAGAGCT	2	1,223	1,252
6-RIGHT	GGTGTCTTCTTGCCACATTT	2	1,592	1,570
7-LEFT	TGATGCAAAGAGGTAATGTTAAGATTCAG	1	1,457	1,486
7-RIGHT	TCCTCCTATTCTTGCTGCGACA	1	1,869	1,847
8-LEFT	TCAGGAACGAGGAACAGACAGA	2	1,740	1,770
8-RIGHT	ACCGTCCATCCCTGGCTTTAAT	2	2,169	2,147
9-LEFT	ATTAGTAGGGCCCACACCTGTG	1	2,042	2,064
9-RIGHT	CCCTACGTCCAGCACTGTACT	1	2,451	2,429
10-LEFT	AGCACAAAATGGAGAAAATTAGTAGACTT	2	2,316	2,345
10-RIGHT	TTCCACCTTTTCTCTATGTTGCTCT	2	2,724	2,699
11-LEFT	CCAGCTATTTTTCAAAGCAGTATGACT	1	2,583	2,610
11-RIGHT	CTTCCTGTGTGAGGGCAACTAC	1	3,010	2,988
12-LEFT	CAGCTGGACTGTCAATGACATACA	2	2,864	2,888
12-RIGHT	TGGGCAGTCTAAACTTGGGTGT	2	3,292	3,270
13-LEFT	AGGAAAATATGCAAAAATGAGAGCTGC	1	3,167	3,194
13-RIGHT	AGCTTGCAGTCTGCTCTTTGA	1	3,558	3,536
14-LEFT	ACCTTTTATGTAGATGGGGCAGC	2	3,429	3,452
14-RIGHT	GGACTTTCCTGATTCCTGCACT	2	3,793	3,771
15-LEFT	AGTAAATCAGATAATAGAGCAGCTAGTGAA	1	3,665	3,695
15-RIGHT	GTCTCTTGCCCTGTTTCAGCAG	1	4,085	4,063
16-LEFT	AGACTGTAGTCCAGGAATATGGCA	2	3,956	3,980
16-RIGHT	GCCATTTGCACTGCTGTCTTTA	2	4,331	4,309
17-LEFT	CCAACAGGAATTTGGAATTCCTACA	1	4,199	4,225
17-RIGHT	TTAGCCTTTCTGCGTGGGACTA	1	4,592	4,570
18-LEFT	TTACAGGGACAGCAGAGAACCT	2	4,472	4,494
18-RIGHT	CTCCTGGCATGAGTCCCCAATA	2	4,832	4,810
19-LEFT	AGGCACGAGGTTGGTTTTACAG	1	4,706	4,728
19-RIGHT	AGCAGTGCTGAGATTGCCAAAT	1	5,067	5,045
20-LEFT	TGCATTATTTGATTGTTTTGCAGAGTCT	2	4,931	4,960
20-RIGHT	TGCTATGTTGGCAGCCAATTCT	2	5,362	5,340
21-LEFT	AGTGGCTTCATGATTTAGGACAGC	1	5,233	5,257

21-RIGHT	ATCTTGATGATCCTTACTGCTCTCAG	1	5,593	5,567
22-LEFT	CCAGCCTAAAACCCCCTGTAAC	2	5,442	5,464
22-RIGHT	CCACAAGCACATGACGATCCTC	2	5,864	5,842
23-LEFT	AAGACAAGAAAGGAGAGATCAATTGATTG	1	5,733	5,762
23-RIGHT	ACATGGCTTTAGGCTTTGATCCC	1	6,152	6,129
24-LEFT	GTGTGCCTACAGATCCTAACCT	2	6,015	6,038
24-RIGHT	TGTTTTGGACAAGCCTGGGTT	2	6,398	6,376
25-LEFT	AGCAAGTGTACTCCTTATTTTATAGAGAGG	1	6,270	6,300
25-RIGHT	TGCTGTAGCTAGCTGCACAATG	1	6,647	6,625
26-LEFT	GGTATTAAGCCTGTAGTCTCTACCCA	2	6,521	6,547
26-RIGHT	ACTTCTAGATCTCCGCCGGATG	2	6,895	6,873
27-LEFT	GGACAGCCTGGAATGAGACTCT	1	6,774	6,796
27-RIGHT	TCTCATGTCTCCTCCTGTAGGTCT	1	7,160	7,136
28-LEFT	GGAAAAGGAATCTTTGCACCTCCT	2	7,031	7,055
28-RIGHT	GCTGTTGCGCCTCTATAGCTTT	2	7,422	7,400
29-LEFT	CTTGATTTCATGGGAGCAGCAG	1	7,289	7,311
29-RIGHT	TCTTGCTCATTGACTTCCTGTTGA	1	7,714	7,690
30-LEFT	TTAGGAACAAACCTATCCTTTTGGGA	2	7,583	7,609
30-RIGHT	CCAGGCAAGTGCTAAGAATCCG	2	8,009	7,987
31-LEFT	CTCACCTTTGTCGTTCCAGACC	1	7,873	7,895
31-RIGHT	ATGCCTCTGCCAACTCTTTGTG	1	8,278	8,256
32-LEFT	TGAAGAGCATTCTCCAGTATTGGG	2	8,145	8,169
32-RIGHT	TCCCTTTGCTCCTCTAACCAGG	2	8,519	8,497
33-LEFT	TTAGAAGGACACCTCCAGCAGC	1	8,389	8,411
33-RIGHT	TGCTTCCTCCACATCTGTTGGA	1	8,799	8,777
34-LEFT	TCATACACAAGGCTTCTTCCCTGA	2	8,675	8,699
34-RIGHT	AGAGAGACCCAGTACAAGCGAG	2	9,080	9,058

Supplementary Table 7.2. SIVcpzEK505 primal scheme primers used for Next-generation sequencing.

Primer Name	Oligonucleotide Sequence (5' to 3')	Oligo-nucleotide Pool	Nucleotide Start Site	Nucleotide End Site
1-LEFT	AGCAAGGAACCCACTGCTTAAG	1	3	29
1-RIGHT	CCTGCTTGCCATACTAAGTGT	1	410	387
2-LEFT	GCGGAAAGGTGAGTACGCAATT	2	274	296
2-RIGHT	GCTGCTTGTTCTGGTTCAGGTT	2	690	668
3-LEFT	CGCTCTCTATTTAACACCCTGGC	1	553	576
3-RIGHT	CCACTCTGCTGCTTCCTCATTG	1	973	951
4-LEFT	GTTTATGGCCTTGTCAGAAGGGG	2	849	872
4-RIGHT	GGCCTGCTCTGCTCTTAGAGTT	2	1,264	1,242
5-LEFT	TGCCAATCCTCTGTTCAGTA	1	1,089	1,111
5-RIGHT	TTGCTGTTTGTGCCTGTGTCAT	1	1,457	1,435
6-LEFT	TGAGAGCATTAGGACCAGGAGC	2	1,337	1,359
6-RIGHT	TCTCTATTGGAGGGGCTGTTGG	2	1,733	1,711
7-LEFT	CTGCAGAAGTGGAGAAAGGCAG	1	1,611	1,633
7-RIGHT	AACCCTCCAATTCCCCCTATCA	1	1,982	1,960
8-LEFT	GCGACCCATCCTTACAGTAAAAATAGG	2	1,844	1,871
8-RIGHT	TCTTCTGTCAAAGGCCACTGTT	2	2,207	2,185
9-LEFT	TGGAAGAAATTTTCTGACACAAATTGGT	1	2,078	2,106
9-RIGHT	AGCATCTCCTACATCTAGCACTGT	1	2,463	2,439
10-LEFT	CCAGAATAGGGCCTGAAAATCCA	2	2,263	2,286
10-RIGHT	TCAGAGCCTACATAGAGATCATCCA	2	2,696	2,671
11-LEFT	CACAAGGCTGGAAAGGATCACC	1	2,569	2,591
11-RIGHT	ATCCCTGGGTAGATCTGGCTAG	1	2,942	2,920
12-LEFT	GGGATATGAACTCCATCCAGATAAATGG	2	2,810	2,838
12-RIGHT	ACCTCATTTTTGCATATTTTCCTGTCT	2	3,196	3,169
13-LEFT	AGAACCTGTACATGGGGCCTAT	1	3,053	3,075
13-RIGHT	TTTCTGCCCCGCTAATTGGTTC	1	3,436	3,414
14-LEFT	CCAGTGCAAAAAGAAGTATGGGAGA	2	3,294	3,319
14-RIGHT	GCAGGTACCCAGGAAAGATATATTCT	2	3,734	3,708
15-LEFT	GGCCTTACAAGATGCAGGACAA	1	3,572	3,594
15-RIGHT	GCCATACCCCTGGACTACAGTT	1	3,985	3,963
16-LEFT	GCAATGGCCAGTGATTTTAATCTACC	2	3,861	3,887
16-RIGHT	ACTGCACCCTGACTTTGAGGAT	2	4,253	4,231
17-LEFT	TGGAAGATGGCCAGTAAAAGTCAT	1	4,115	4,139
17-RIGHT	GGGCTCTCTGCTGTCTCTGTAA	1	4,500	4,478
18-LEFT	GGGATACTGCAGGGGAAAGA	2	4,376	4,398
18-RIGHT	GTCTATAGAACCAACCTCGTGCC	2	4,736	4,713
19-LEFT	TGGAAAACAGATGGCAGGTGATG	1	4,613	4,636
19-RIGHT	CCTGATGGCAGAGGCTGTAAAA	1	4,975	4,953
20-LEFT	AAGGCCATGGCATTTAGGACAT	2	4,839	4,861

20-RIGHT	TTGCCCTAGTCCATGTAGCCAA	2	5,262	5,240
21-LEFT	TGGAACAGGCACCAGAAGATCA	1	5,131	5,153
21-RIGHT	TCCGCTTCTTCTGCCATATGA	1	5,555	5,533
22-LEFT	ATTAGAGCCCTGGAATCACCCA	2	5,418	5,440
22-RIGHT	GCCAATTCTTCTGCATCTCCGT	2	5,825	5,803
23-LEFT	GAGACTCTAATAGTAATAGTAGTTTGGGCA	1	5,673	5,703
23-RIGHT	TCCACATATCAAAACGTTCTGTCACA	1	6,062	6,036
24-LEFT	CTGTGCTTCAGATGCTAAGGCC	2	5,931	5,953
24-RIGHT	ACCACTGCTATTATTAGTGTCTAAGGGT	2	6,331	6,303
25-LEFT	ACCAAACAACCACGTACAAATGC	1	6,197	6,221
25-RIGHT	TGTTGCTAAGCTGCCATTTAGGA	1	6,574	6,551
26-LEFT	GCACCACCAGGATTTGCCATTA	2	6,427	6,449
26-RIGHT	TTCTAGGTCTCCGCCTGGATTG	2	6,856	6,834
27-LEFT	CACTAGGCAAGCATATTGTGAGGT	1	6,732	6,756
27-RIGHT	GGTAGTCATGTCTCCCCCTGT	1	7,136	7,114
28-LEFT	ATGCTCCGCCAATTAAGGACC	2	7,013	7,035
28-RIGHT	GTTGTTGCGCCTCTATTGCTCT	2	7,397	7,375
29-LEFT	TTCTTGGGTTTCTTGAGCAGC	1	7,262	7,284
29-RIGHT	TTGCTCTTGCTTGCTCTAGC	1	7,666	7,644
30-LEFT	GCCTTGGAATACCACTTGAGC	2	7,530	7,552
30-RIGHT	TGTTCTTTCTGGCCTGTCTGGT	2	7,909	7,887
31-LEFT	GTAGTAGCAGGCATTGTAGGCA	1	7,771	7,793
31-RIGHT	TTAGCTCCCTTCCCCAGTAAGC	1	8,153	8,131
32-LEFT	GACAGACTGTCTCTCAATCCTGAG	2	8,022	8,046
32-RIGHT	CCTTCTGCTGGATCTTGCTGTC	2	8,389	8,367
33-LEFT	ACCTAGGAGAATCAGACAAGGTCT	1	8,259	8,283
33-RIGHT	TGCCAGTCAGGGAAGATTCCTT	1	8,677	8,655
34-LEFT	ACAAAGAAGCTTTTGATCTTTCCTTCTT	2	8,544	8,572
34-RIGHT	TCTGGTAGAACTCCGGATGCAG	2	8,916	8,894
35-LEFT	AGTACCCTGTACCTGAGGAG	1	8,737	8,759
35-RIGHT	GGCTTAAGCAGTGGGTTCCCTTG	1	9,105	9,083

Supplementary Table 7.3. SIVcpzLB715 primal scheme primers used for Next-generation sequencing.

Primer Name	Oligonucleotide Sequence (5' to 3')	Oligo-nucleotide Pool	Nucleotide Start Site	Nucleotide End Site
1-LEFT	TTTCAAGCAGAAGACGGCATAACG	1	9	32
1-RIGHT	GAAGGAGAGAGAATGGGTGCGA	1	397	375
2-LEFT	GCTCCCTGCTTGCCCACTAA	2	264	286
2-RIGHT	CGAGAGATCTGCTCCGCTACTT	2	624	602
3-LEFT	CTGCAAGTAGTGTGTGCCATT	1	526	548
3-RIGHT	ACAAGCAAATCTGTCCAGCTCC	1	940	918
4-LEFT	GGGAGCTGGACAGATTTGCTTG	2	843	865
4-RIGHT	GGCTCCTTCTGCTAATGCCATG	2	1,256	1,234
5-LEFT	GGTGCATCAGCCATTATCACCC	1	1,144	1,166
5-RIGHT	GCTTTATGTCTAGAATACTGACAGGGC	1	1,578	1,551
6-LEFT	AGTAACCCTCCCATTCCAGTGG	2	1,472	1,494
6-RIGHT	GCATTTGCTTGACTCATGGCCT	2	1,834	1,812
7-LEFT	AGGAGCCACCCTAGAGGAAATG	1	1,735	1,757
7-RIGHT	AACCCAAAGCTCTCCATTGGTG	1	2,107	2,085
8-LEFT	TTTTTAGGGAAAATCTGGCCTCTCA	2	2,009	2,034
8-RIGHT	TGTGGGTCCTACTAAAACGTACCT	2	2,431	2,406
9-LEFT	AGGGGGAAAATGGAAACCAAAAATG	1	2,304	2,329
9-RIGHT	TGGTGTATTGTATGGATTTTCAGGCC	1	2,665	2,639
10-LEFT	ACCAAGAGTCAAACAATGGCCA	2	2,541	2,563
10-RIGHT	GCTGGTGATCCTTTCCATCCCT	2	2,961	2,939
11-LEFT	TGCCCCCTAGATGAAAATTTTAGGA	1	2,839	2,864
11-RIGHT	GACAGTCCAGCTCTCCTTGCT	1	3,250	3,228
12-LEFT	ACAAGAAACATCAGAAGGAACCACC	2	3,140	3,165
12-RIGHT	ATGTCCACTGACTCTGTCCCTG	2	3,503	3,481
13-LEFT	GGGAGATATTGGCAAACCCAGT	1	3,407	3,429
13-RIGHT	ACCAGAGTTTTACTAGGGGAGGG	1	3,767	3,744
14-LEFT	AAAGGAAACATGGGAAGCCTGG	2	3,669	3,691
14-RIGHT	TCACTTTTATCTGGCTGGCCTTG	2	4,029	4,006
15-LEFT	GCTATTCAGCTAGCTTTACAGGATTCA	1	3,928	3,955
15-RIGHT	GGACTGCAGTCTACCTGTCCAT	1	4,341	4,319
16-LEFT	GGCCAGTGACTTTAATATACCCCC	2	4,233	4,257
16-RIGHT	CCTTGACTTTGGGGATTGTAGGG	2	4,614	4,591
17-LEFT	ACACACTGATAATGGCAGCAACT	1	4,506	4,529
17-RIGHT	ACTACTGCTCCTTCACCCTTCC	1	4,917	4,895
18-LEFT	AGCAACAGACATACAAACACTACAGAACT	2	4,779	4,806
18-RIGHT	TGCCAGTCTCTTTCTCCTGGTT	2	5,218	5,196
19-LEFT	AGTACAAGTCCAAAAGTAAGCTCAGA	1	5,114	5,140
19-RIGHT	ACACTAGGCAAGGGTGGTCTTT	1	5,476	5,454
20-LEFT	GCCCTAAGTGTGAATATCCAGCA	2	5,367	5,390
20-RIGHT	CCTTCTTCGAACCACAATGCCT	2	5,761	5,739
21-LEFT	GCGGGAGTTGAAGCCATAATAAGG	1	5,659	5,683

21-RIGHT	TGCTACTATCCAAGCTCCTATTGCT	1	6,075	6,050
22-LEFT	AGATTCTATACCAGAGCAGTAAGTATGC	2	5,962	5,990
22-RIGHT	GTGCCTCTGTTTCTTGTGCCTT	2	6,350	6,328
23-LEFT	TAAGGAGGAACTGTGGGTGACA	1	6,243	6,265
23-RIGHT	GCTGCAGTTTCTCAAGTCTTCCC	1	6,619	6,596
24-LEFT	GTGTAAAGTTGACTCCTCTCTGTGT	2	6,509	6,534
24-RIGHT	AAAGAAGTTTTTGGGCAGGCCT	2	6,870	6,848
25-LEFT	ATAGCAATGACACATGCAGCCC	1	6,773	6,795
25-RIGHT	TGTCCTGGCCCTATAGGAATGC	1	7,179	7,157
26-LEFT	ACAAAAACCAAAGTACTGACACCATT	2	7,067	7,093
26-RIGHT	TCCCAGGTCGTCTCATTCAAGTG	2	7,461	7,439
27-LEFT	AACACTACTTCAATTGTGGAGGAGAA	1	7,352	7,378
27-RIGHT	TCATTTCTCCTCCTGCAGGTCT	1	7,733	7,711
28-LEFT	AAGCAGATAGTGAGAACAAAACAAGG	2	7,634	7,660
28-RIGHT	GGAGCTGTTTAATGCCCCAGAC	2	8,036	8,014
29-LEFT	GGCAATTATTGACTGGCATAAGTGC	1	7,931	7,955
29-RIGHT	CAAACCAATTCCACAGGCTTGC	1	8,327	8,305
30-LEFT	TGGGACAAAGAAGTCAGAAATTACACA	2	8,197	8,224
30-RIGHT	TCTCAAGCGGTGGTAGAGGAAG	2	8,623	8,601
31-LEFT	GAAGGTGGAGAGCAAGACAACG	1	8,512	8,534
31-RIGHT	AGGCTGCTCTTTGACCATTTGG	1	8,929	8,907
32-LEFT	AGCACAAAGATTTGGTAGAGGCA	2	8,823	8,046
32-RIGHT	CCTTCCAGTCCCCCTCTTTCTT	2	9,196	9,174
33-LEFT	AGGTTTCCCAGTAAGGCCTCAG	1	9,097	9,119
33-RIGHT	TACTCCGGATGTTGCTCTCTGG	1	9,505	9,483
34-LEFT	AGGTTTCCCAGTAAGGCCTCAG	2	9,097	9,119
34-RIGHT	TACTCCGGATGTTGCTCTCTGG	2	9,505	9,483

Supplementary Table 8.1. SIVmac239 primal scheme primers used for deep sequencing

Primer Name	Oligonucleotide Sequence (5' to 3')	Oligo-nucleotide Pool	Nucleotide Start Site	Nucleotide End Site
SIVmac239_1_LEFT	GCTGTCTTTTATCCAGGAAGGGG	1	112	135
SIVmac239_1_RIGHT	CGCTAGATGGTGCTGTTGGTCT	1	542	520
SIVmac239_2_LEFT	ACACACTGAGGAAGCAAAACAGA	2	438	461
SIVmac239_2_RIGHT	GATCCTGACGGCTCCCTAAGTT	2	855	833
SIVmac239_3_LEFT	AGCGGCTATGCAGATTATCAGAGA	1	744	768
SIVmac239_3_RIGHT	GCACTAGCTTGCAATCTGGGTT	1	1157	1135
SIVmac239_4_LEFT	AGACAGGTTCTACAAAAGTTAAGAGCA	2	1047	1075
SIVmac239_4_RIGHT	CCACATTTCCAGCATCCCTGTC	2	1407	1385
SIVmac239_5_LEFT	GGGACCAAGAAAGCCAATTAAGTG	1	1308	1332
SIVmac239_5_RIGHT	GGCTGTCTTCAATATGAGCAGT	1	1712	1689
SIVmac239_6_LEFT	GCAGAGAGAAAGCAGAGAGAAGC	2	1602	1625
SIVmac239_6_RIGHT	TGGCCACTGCTTCAATTTGGT	2	2028	2006
SIVmac239_7_LEFT	TTGCTAACAGCTCTGGGGATGT	1	1920	1942
SIVmac239_7_RIGHT	GAGGTATGGAGAAATATGCATCACCT	1	2311	2285
SIVmac239_8_LEFT	AGGGTCACTCAGGACTTTACGG	2	2196	2218
SIVmac239_8_RIGHT	ACTACCCTGTCATGTTCCAGGT	2	2558	2536
SIVmac239_9_LEFT	ACATGTGCTAGAACCCCTCAGGA	1	2447	2470
SIVmac239_9_RIGHT	TCTGCTTCTGCCATCTCAGTCC	1	2858	2836
SIVmac239_10_LEFT	GGGCAGCTCAAATTTATCCAGGT	2	2749	2772
SIVmac239_10_RIGHT	TGCCAATAGTCTGTCCACCACT	2	3170	3148
SIVmac239_11_LEFT	ACAGAAAATAGGAAAGGAAGCAATAGTGA	1	3065	3094
SIVmac239_11_RIGHT	TGCTCTCTGATTCTGTAGGGCA	1	3496	3474
SIVmac239_12_LEFT	CATGGCATTGACAGACTCAGGG	2	3398	3420
SIVmac239_12_RIGHT	CTGCCATGTATAGCCTCTCCT	2	3789	3767
SIVmac239_13_LEFT	AGCCAGCACAAGAAGAACATGA	1	3646	3668
SIVmac239_13_RIGHT	GCCACCATGCAACCATCTTTA	1	4028	4006
SIVmac239_14_LEFT	GTAAAAATTGGCAGGCAGATGGC	2	3932	3955
SIVmac239_14_RIGHT	GGGTCCCTCCACAGTTGATCT	2	4344	4322
SIVmac239_15_LEFT	ACATGATCACTACAGAACAAGAGATACA	1	4234	4262
SIVmac239_15_RIGHT	GGCTTCCTCCTGTAGTGGGAA	1	4637	4615
SIVmac239_16_LEFT	AGAGGCTAGAGAGGTGGCATAG	2	4490	4512
SIVmac239_16_RIGHT	CCCTGGGATCTGACATCGCTTA	2	4929	4907
SIVmac239_17_LEFT	AGGGGAGAACAACCTGCTGTCTT	1	4819	4841
SIVmac239_17_RIGHT	TTATGCTAGTCTGGAGGGGGA	1	5252	5230
SIVmac239_18_LEFT	TGCATTGCAAGAAAGGCTGTAGA	2	5154	5177
SIVmac239_18_RIGHT	TGCTTCTAGAGGGCGGTATAGC	2	5550	5528
SIVmac239_19_LEFT	AACGAGCGCTCTTCATGCATTT	1	5448	5470
SIVmac239_19_RIGHT	GCATTCCTCCAAGCTGGTACAC	1	5818	5796
SIVmac239_20_LEFT	TGGGATGTCTTGGGAATCAGCT	2	5706	5728
SIVmac239_20_RIGHT	TGCTGATGCTGTCGTTGATGTT	2	6125	6103

SIVmac239_21_LEFT	TCCCCATTATGCATTACTATGAGATGC	1	6017	6044
SIVmac239_21_RIGHT	TCTAAGCAAAGCATAACCTGGAGG	1	6431	6407
SIVmac239_22_LEFT	ACATGAACCACTGTAACTTCTGT	2	6324	6349
SIVmac239_22_RIGHT	TTCCTCCAAACCAACCCATGC	2	6750	6728
SIVmac239_23_LEFT	ACCAGGAAATAAGACAGTTTTACCAGT	1	6643	6670
SIVmac239_23_RIGHT	TGCCTACTTTATGCCAAGTGTTGA	1	7041	7017
SIVmac239_24_LEFT	GGGTAGAAGATAGGAATACAGCTAACCA	2	6933	6961
SIVmac239_24_RIGHT	TTGCGAGAAAACCCAAGAACCC	2	7317	7295
SIVmac239_25_LEFT	AGTAGAGATCACTCCAATTGGCTTG	1	7201	7226
SIVmac239_25_RIGHT	TCCCACTCTTGCCAAGTCTCAT	1	7627	7605
SIVmac239_26_LEFT	ATGCTTGGGGATGTGCGTTTAG	2	7524	7546
SIVmac239_26_RIGHT	GGGGAAGAGAACACTGGCCTAT	2	7888	7866
SIVmac239_27_LEFT	TGGCAATTGGTTTGACCTTGCT	1	7738	7760
SIVmac239_27_RIGHT	GAATCCTCTGTAGGGTCGCAGA	1	8142	8120
SIVmac239_28_LEFT	CTGATACGCCTCTTGACTTGGC	2	8030	8052
SIVmac239_28_RIGHT	TTCAGCTGGGTTTCTCCATGGA	2	8403	8381
SIVmac239_29_LEFT	CCAGGAGGATTAGACAAGGGCT	1	8304	8326
SIVmac239_29_RIGHT	GGGACTAATTTCCATAGCCAGCC	1	8714	8691
SIVmac239_30_LEFT	AGAAAAGGAAGAAGGCATCATACCA	2	8615	8640
SIVmac239_30_RIGHT	ATCCCCTTGTGGAAAGTCCCTG	2	8997	8975

Supplementary Table 8.2. SIVhu and SIV_{B670} primal scheme primers used for deep sequencing

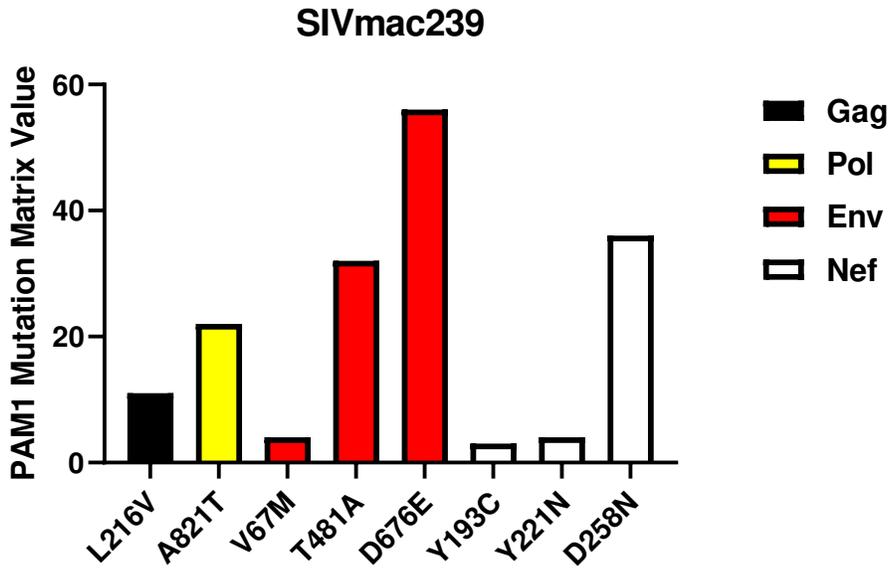
Primer Name	Oligonucleotide Sequence (5' to 3')	Oligo-nucleotide Pool	Nucleotide Start Site	Nucleotide End Site
SIVhu_1_LEFT	CACACCAGGACCAGGAATCAGA	1	17	39
SIVhu_1_RIGHT	TAAGTAGGCGTTCCAACCAGCT	1	406	384
SIVhu_2_LEFT	AGAGAAGGCTAACCGCAAGAGG	2	274	296
SIVhu_2_RIGHT	ACCAGGGTCTTCTTATGTCTCGA	2	693	670
SIVhu_3_LEFT	GGCTCCATGCTTGCTTGCTTAA	1	564	586
SIVhu_3_RIGHT	CCCACTCTCTACTCTTTCCTCA	1	977	954
SIVhu_4_LEFT	TCCTGAGTACGGCTGAGTGAAG	2	783	805
SIVhu_4_RIGHT	ACCTGTCGGAACAAATGGAGCT	2	1192	1170
SIVhu_5_LEFT	GAAACATGTAGAGTGGGCAGCA	1	1071	1093
SIVhu_5_RIGHT	GCACTACCTCTGCCCAAATTT	1	1493	1471
SIVhu_6_LEFT	GGAGGAAATTACCCAGTGCAGC	2	1372	1394
SIVhu_6_RIGHT	ACCAAGTTGGATCCATCTCCT	2	1793	1771
SIVhu_7_LEFT	GCCAAGAGGATCAGCTATTGCA	1	1671	1693
SIVhu_7_RIGHT	CCCTTCAATGCCTCTGCCATT	1	2092	2070
SIVhu_8_LEFT	CCCAGATTATAAATTGGTACTCAAGGGT	2	1965	1993
SIVhu_8_RIGHT	ATTCACAGCTGGATCCTCTGGG	2	2386	2364
SIVhu_9_LEFT	CAAGGCTGCTGGAAATGTGGAA	1	2215	2237
SIVhu_9_RIGHT	TCTCTGGTGTAAATGAAGCCTCCT	1	2645	2621
SIVhu_10_LEFT	GCCTACATTGAAGAACAGCCCA	2	2508	2530
SIVhu_10_RIGHT	GCTCCCTCTAACTGGCCATCTT	2	2918	2896
SIVhu_11_LEFT	TGGAGCCTATAAAAAGTAACACTAAAACCA	1	2782	2811
SIVhu_11_RIGHT	TCCTGGTCTGCATTGTTTACTGA	1	3192	3168
SIVhu_12_LEFT	CCCTGCAGGACTAGCAAAAAGG	2	3056	3078
SIVhu_12_RIGHT	CATCCACTGGAATGGGGGATCT	2	3459	3437
SIVhu_13_LEFT	TGACATCCTAATAGCTAGTGATAGAACAGA	1	3323	3353
SIVhu_13_RIGHT	TCTTCATATTCTGCCCTGCCA	1	3683	3661
SIVhu_14_LEFT	TAGGAGTACTCAACTGGGCAGC	2	3550	3572
SIVhu_14_RIGHT	GAAATTTTGGCACCTGTCCCA	2	3934	3912
SIVhu_15_LEFT	CTGAAAGTAGGCAAATTCGCAAAGA	1	3810	3835
SIVhu_15_RIGHT	AGGTTTTGCTTTGTCTCTGCCT	1	4170	4148
SIVhu_16_LEFT	AGTCTTCAACCTGGTAAAAGAACCT	2	4040	4065
SIVhu_16_RIGHT	GGTGATCTACCTCTTGATTCCTCC	2	4417	4392
SIVhu_17_LEFT	GCAGGTCAACCCACTGAATCAG	1	4284	4306
SIVhu_17_RIGHT	TCCACTAGCCACATGTACTGCA	1	4692	4670
SIVhu_18_LEFT	AATGCCACCAGAAAGGAGAAGC	2	4570	4592
SIVhu_18_RIGHT	TGTGGGTTTTTAGATGATGATTCATTGC	2	4930	4902
SIVhu_19_LEFT	TGGTGCCAATTTACATCACAAGA	1	4796	4820
SIVhu_19_RIGHT	ACAATAGCTCACGGGTCCTTT	1	5166	5144
SIVhu_20_LEFT	AGTCAATATGATCACCACAGAACAAGA	2	5038	5065
SIVhu_20_RIGHT	TACTCTGCTGCAAGTCCACCAT	2	5421	5399

SIVhu_21_LEFT	ATACCGGAGAGGCTAGAGAGGT	1	5292	5314
SIVhu_21_RIGHT	ACTTGCAAGACAGCAACTT	1	5659	5637
SIVhu_22_LEFT	AGAAATTTCTGGACAGATGTAACACCA	2	5529	5556
SIVhu_22_RIGHT	TGCCAGTATCCCAGGACCTTT	2	5892	5870
SIVhu_23_LEFT	GGAGAAGAGACAATAGGAGAGGCC	1	5761	5785
SIVhu_23_RIGHT	TCTACTACCCATCCATCCCACG	1	6127	6105
SIVhu_24_LEFT	TTAGGAGAAGGGCATGGGGTAG	2	5989	6011
SIVhu_24_RIGHT	CCTCTAGAGGGCGGTATAGTTGA	2	6358	6335
SIVhu_25_LEFT	CCCTGAAGGAGCAGGAGAACT	1	6222	6244
SIVhu_25_RIGHT	CACAGAAGAGGGGAATTGTCGC	1	6648	6626
SIVhu_26_LEFT	TGGGATGTCTTGGGAATCAGCT	2	6516	6538
SIVhu_26_RIGHT	GCTGCTGTTGTTGCTGTTGTTG	2	6922	6900
SIVhu_27_LEFT	CTTTGAGACCTCTATAAACCTGTGT	1	6793	6820
SIVhu_27_RIGHT	GCATCCAATAATGCTTGTCACAG	1	7204	7180
SIVhu_28_LEFT	CAGGGTAAAAAGGGACAAAAAGAAAGA	2	7044	7072
SIVhu_28_RIGHT	AGACTGTCTTATCCCTGGCCT	2	7482	7460
SIVhu_29_LEFT	GGTTTGGTTTCAATGGTACTAGAGCA	1	7344	7370
SIVhu_29_RIGHT	AAACTCTCTCTACAATTTGTCCACA	1	7721	7695
SIVhu_30_LEFT	GGCCATCCAGGAGGTAAGGAA	2	7579	7601
SIVhu_30_RIGHT	GTCCGGTTGCCATTGATCCAAT	2	7972	7950
SIVhu_31_LEFT	CATATTAGACAGATAATCAACATGGCA	1	7835	7864
SIVhu_31_RIGHT	TGTTGCTGTTGCTGCACTATCC	1	8239	8217
SIVhu_32_LEFT	AGAGAGGGGTCTTTGTGCTAGG	2	8109	8131
SIVhu_32_RIGHT	TGCCTCTAGGAAGTCAACCTGT	2	8486	8464
SIVhu_33_LEFT	GGGATGTGCTTTTAGGCAGGTC	1	8365	8387
SIVhu_33_RIGHT	CATAAGCGGGAGGGGAAGAGAA	1	8733	8711
SIVhu_34_LEFT	CAATTGGTTTACCTTACTTCTTGAT	2	8575	8602
SIVhu_34_RIGHT	AAGAGCCTCTGGAACACTGGTT	2	8956	8934
SIVhu_35_LEFT	TGGCCTTGGCAGATAGAATATATCA	1	8822	8848
SIVhu_35_RIGHT	TCCTCACTGTATCCCTGTCCT	1	9201	9179
SIVhu_36_LEFT	CAAGCGCGTGGAGAGACTTATG	2	9075	9097
SIVhu_36_RIGHT	GCCAACTCTGGATGCCTTCTTCC	2	9477	9455
SIVhu_37_LEFT	GCCATGCATACAAATTGGCGAT	1	9339	9362
SIVhu_37_RIGHT	CCTCTTCTCTGACAAGCCTGA	1	9738	9716
SIVhu_38_LEFT	GCTTAGTGCATCCAGCTCAGAC	2	9585	9607
SIVhu_38_RIGHT	CTGCTAGTGTGGAGAGAACCT	2	9971	9949
SIVhu_39_LEFT	GGCTGACAAGAAGGAAACAAGCT	1	9775	9798
SIVhu_39_RIGHT	ACCAGGTCTTCTTATGTCTCGA	1	10160	10137
Revised Primer Pairs for improved coverage				
SIVhu_7b_Fwd	GCAGGACAACCTTAGAGAGCCAAGA	1	1655	1678
SIVhu_7b_Rev	CCTTTCAATGCCTCTGCCATTAGC	1	2091	2068
SIVhu_13b_Fwd	GACCTGATCCAATACATGGATGACA	1	3303	3328
SIVhu_13b_Rev	ACACCCTTCTGTTCTTGCTGA	1	3720	3698
SIVhu_16b_Fwd	TGTATCAACACCTCCCTTAGTCAGATTAG	2	4014	4042
SIVhu_16b_Rev	ACTAACTAGGTGATCTACCTTTGATTCC	2	4425	4396
SIVhu_30b_Fwd	AACAGGCCTGGTGTGGTTTAG	2	7543	7564
SIVhu_30b_Rev	ATGTTCTGTCGGTTGCCATTGA	2	7978	7957
SIVhu_36b_Fwd	ACTCTGGGAAGGGTTGGAAGATG	2	9102	9124

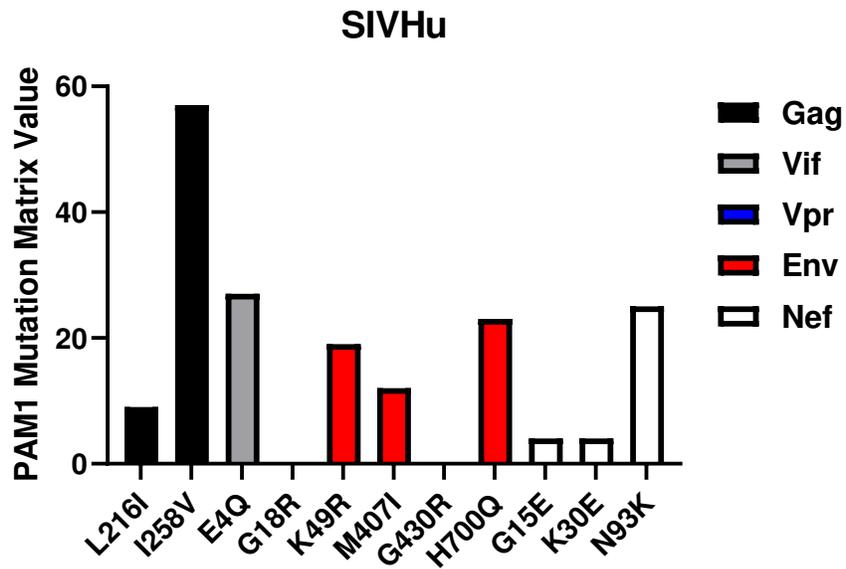
SIVhu_36b_Rev	CTGATTCCTGGTCCTGGTGTGTAA	2	9505	9482
SIVhu_37b_Fwd	CCAGTACACCCAAAGGTGCCATTA	1	9313	9336
SIVhu_37b_Rev	CCTCTTGCGGTTAGCCTTCTCTTAC	1	9763	9738

Supplementary Figure 8.1

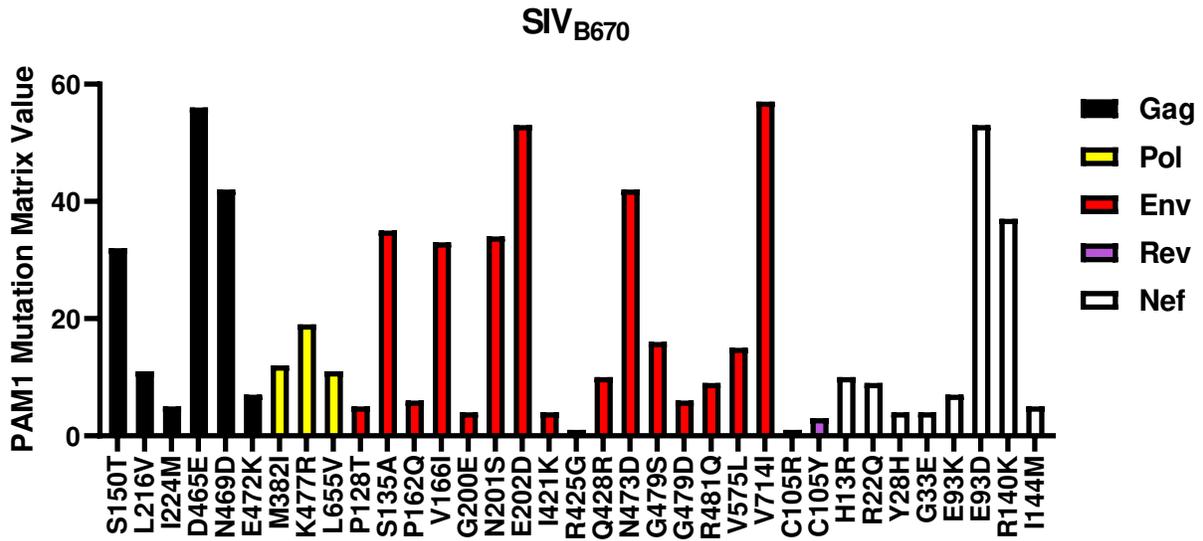
A.



B.



C.



Supplementary Figure 8.1 PAM1 matrix values of identified variant amino acid residues. (A) SIV_{mac239}, (B) SIV_{hu}, (C) SIV_{B670} variants with their respective PAM1 matrix values. Residues were grouped by genome order and colored based on gene. Columns with no bars had PAM1 values of 0. Both favored and disfavored amino acid changes occurred throughout the viral genomes of all three viruses.