UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Modeling the HIV Genetic Bottleneck in vitro

Permalink

https://escholarship.org/uc/item/1j68c5ng

Author

Caldwell, Thomas Allen

Publication Date

2023

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

Modeling the HIV Genetic Bottleneck *in vitro*

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Sciences

by

Thomas Allen Caldwell III

Dissertation Committee:
Professor Donald N. Forthal, Chair
Professor Elizabeth Head
Professor Klemens J. Hertel

2024

© 2024 Thomas Allen Caldwell III

DEDICATION

To

Seyed M. Sadjadi
May 27th 1928
October 12th 2012

Cynthia R. Caldwell
March 29th 1938
April 9th 2021

I am the person that I am because of their nurture and compassion throughout my life, and I owe a part of myself to them for their support and love, for which very few could match in quality.

And to all those who have been there for me as friends and family, thank you for all that you have done to help me get to where I am today.

Nouri Caldwell, Thomas A. Caldwell Jr., Seyed Sadjadi, Guoxin Hu,
Brandy E. Coats, Erica Treemarcki Vogel, Sophie Treemarcki Vogel,
Olivia Cooley, Michelle A. Mastro, Marianne "Zhetot Kwe" Almero,

"Science is a way of thinking much more than it is a body of knowledge."

Carl Sagan
Broca's Brain: Reflections on the Romance of Science

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to extend my deepest appreciation to my advisor, principal investigator, and committee chair, Professor Donald Forthal (Department of Molecular Biology and Biochemistry). He has been with my on this difficult journey of mine for years and has been incredibly patient as he nurtured and helped me develop my skills as a thinker and a scientist during the time I worked with him. Additionally, I extend similar gratitude to the members of the Forthal Lab, both present and previous: Dr. Johannes Gach, Dr. Swati Seth, Gwen Jordaan, Julianne Nguyen, and Khoa Vu.

I would like to thank my committee members, Professor Klemens Hertel (Department of Microbiology and Molecular Genetics, School of Medicine at UCI) and Professor Elizabeth Head (Department of Pathology and Laboratory Medicine, School of Medicine at UCI). Klemens Hertel has been by my side during the worst and best times of my graduate education, encouraging me along the way and giving tough love when I needed to hear it. Elizabeth Head has also encouraged me to keep fighting for my degree during some of my darkest times, and I cherish all of the intellectual conversations we have had regarding science.

I also wish to thank Leora Fellus (Director of Graduate Studies, School of Medicine at UCI) who has been by my side whenever I ran into difficult challenges—both personally and professionally. She has labored considerably to ensure that I keep working toward the rewards and goals I set out to fulfill, and I owe her immense gratitude for her sacrifices getting me to where I am today.

I also thank Stefani Ching (Graduate Student Coordinator, Department of Pathology and Laboratory Medicine) for always setting up our journal club and research-in-progress talks and scheduling events as well as seminars for the department. I thank Professors Edwin Monuki (Department of Pathology and Laboratory Medicine, School of Medicine at UCI) and Dan Mercola (Department of Pathology and Laboratory Medicine, School of Medicine at UCI) for their continued support and intellectual inputs to our scientific discussions.

I would like to thank the Center for Virus Research and all its members for providing me with a space to engage in scientific discussion as well as sponsoring me for the T32 Virology Training Grant Program during the academic year 2018 – 2019. I also wish to thank Graduate Division for awarding me the Graduate Division Completion Fellowship during the academic year 2022 – 2023. I also thank the National Institute of Allergy and Infectious Diseases (part of the National Institute of Health) for financial support through research funding (research grants 5R01AI118581-05 and 1R21AI149255-01).

VITA

Thomas Allen Caldwell III

Education:

2014 – 2023	**School of Medicine, University of California, Irvine** Experimental Pathology PhD Program PhD Candidate in Biomedical Sciences
2008 – 2011:	**University of California, Los Angeles** Major: Biochemistry, B.S. Academic Achievements: ISLP, Library Research Prize, Chancellor Service Award
2006 – 2008:	**Saddleback College** Major: Biology Academic Achievements: Dean's List, Biology Research Scholarship

Presentations and Conferences:

- 3rd Annual Neurology Science Poster Day — Jan. 2011
- 74th Annual ADA Scientific Sessions — Jun. 2014
- 75th Annual ADA Scientific Sessions — Jun. 2015
- 3rd Annual Cedars-Sinai Inter-Institution Graduate Student Symposium — Oct. 2015
- 9th Annual San Diego Beta Cell Society Meeting — Nov. 2015
- 2016 Associated Graduate Students Graduate Research Symposium — April 2016
- 2019 T32 Virology Training Grant Symposium — May 2019
- 2023 CVR Graduate Student and Post-Doctoral Research-in-Progress — Nov. 2023

Publications:

- **Caldwell, T.** (2007). Exhaustion due to Mental Stress and Metabolism of Sugar and Caffeine in Energy Drinks. *Saddleback Journal of Biology*. **7**: 66 – 71
- **Caldwell, T.** (2009). Diabetes: A review of current mechanisms, research studies, and treatments. *Undergraduate Science Journal*. **22**: 27 – 38

- **Caldwell, T.** (2010). NEWSBEAT: Genetic Regulation of Pancreatic Development. *Undergraduate Science Journal*. **23**: 3
- Yomogida, K., Wu, S., Baravati, B., Avendano, C., **Caldwell, T.**, Maniaci, B., Zhu, Y., and Cong-Qiu Chu. (2013). Cell penetrating recombinant Foxp3 protein enhances Treg function and ameliorates arthritis. *BBRC*. **434**(2): 263 – 267.
- Khanjyan, M., Yang, J., Kayali, R., **Caldwell, T.,** and Bertoni, C. (2013). A high-content, high-throughput siRNA screen identifies cyclin D2 as a potent regulator of muscle progenitor cell fusion and a target to enhance muscle regeneration. *Hum. Molec. Genet.* **22**(16): 3283 – 3295.
- Zhang, C., **Caldwell, T.**, Mirbolooki, M. R., Duong, D., Park E. J., Chi, N. W., and Chessler, S. D. (2016) Extracellular CADM1 interactions influence insulin secretion by rat and human islet beta-cells and promote clustering of syntaxin-1. *Am. J. Physiol. Endocrinol. Metab.* **310**(11): E874 – E885.

Work and Research Experience:

4/2018 – 12/2023 **Graduate Student Researcher, School of Biological Sciences, Department of Molecular Biology and Biochemistry, Forthal Lab, University of California, Irvine, CA**
I am currently studying the genetic bottleneck of HIV transmission: the pathologic phenomenon in which certain strains or genetic variants of HIV are favored for transmission from an infected host to an uninfected sexual partner.

6/2015 – 9/2017 **Graduate Student Researcher, School of Medicine, Department of Pathology, Chessler Lab, University of California, Irvine, CA**
Under the mentorship of Steven Chessler, I studied the function of a presynaptic transmembrane protein (Leukocyte common antigen-related receptor; LAR) in the formation of the insulin secretory complex in beta cells, as well as its role in regulating the expression of insulin gene expression.

7/2011 – 3/2014 **Research Associate II, Vivoscript Inc., Coasta Mesa, CA**
I helped design fusion peptides with extrinsic nuclear localization signals and recombinant transcription factors (i.e. OCT4, PDX1, etc.). I purified and refolded recombinant fusion peptides from bacterial inclusion bodies for cell reprogramming studies. I developed a new method

of identifying chemical factors (i.e. ionic strength, pH, viscosity, etc.) that influence the folding state of a given protein.

12/2009 – 2/2012 **Research Assistant in Neurology, University of California, Los Angeles, CA**
I studied the inhibitory effects of 767 functional kinases on muscle progenitor cell fusion by means of high-throughput screening in order to identify novel gene interactions that regulate muscle growth.

Volunteer Work:

1/2021 – 6/2021 **Mentor, HOSA, UCI, CA**
I mentored third and fourth year first-generation undergraduate students from underrepresented backgrounds and helped them plan their path to graduation, seek extracurricular and/or research opportunities, apply for additional funding sources, and find fulfilling careers in science and public health after graduation.

1/2017 – 2/2021 **Volunteer, Laguna Beach Animal Shelter, CA**
I serve as a volunteer member of the city-operated animal shelter in Laguna Beach. My responsibilities include, but are not limited to, cleaning animal living areas, walking/washing/feeding animals, facilitating pet adoptions and license renewals, and assist the public with their inquiries.

9/2016 – 6/2017 **Co-president, Secular Student Alliance, UCI, CA**
As president of SSA at UCI, I managed staff meetings, invited guest speakers, and helped foster an environment in which students interested in secularism and the defense of the Separation between Church and State could meet and discuss important issues facing our campus and the community at large. Our mission was to dispel the misconception that the ability to distinguish between right and wrong was not exclusive to religious belief and that atheists, agnostics, and humanists are just as capable of being just and moral as any individual of faith.

8/2015 – 6/2017 **Leadership Coach, DECADE PLUS, UCI, CA**
I mentored first-year, first-generation undergraduate students in their paths toward actualizing their academic

interests and goals during their first year at UCI. The goal of this program was to help students maintain satisfactory progress toward graduation as they navigated, confronted, and overcame the socioeconomic obstacles they faced while in school.

1/2015 – 3/2015 **Counselor, CampMed, UCI, CA**
As a member of CampMed, I was involved in community outreach to mentor high school students from under-represented socioeconomic backgrounds on college applications and to encourage their enrollment in science and pre-medical undergraduate programs. We also hosted an annual medical training camp where we gave instructional talks, demonstrations, and create fun activities for high school students to learn more about science, the medical field, and health-related issues.

11/2014 – 6/2015 **Staff Member, Secular Student Alliance, UCI, CA**
I help coordinate events, activities, and lectures by visiting scholars through the Secular Student Alliance chapter at UCI, which represents the non-religious community in Irvine. We seek to promote the separation between church and state, freedom of thought and enquiry, and protection of the right to practice religion (or to refuse to practice) freely and peacefully.

3/2011 – 8/2012 **Science Columnist, TrekNews.net**
I created and maintained a pro-science web series for the Star Trek fan news website, treknews.net, that discussed the technical plausibility (or implausibility) of futuristic technology as they were featured in the Star Trek canon. I also discussed contemporary scientific issues by reference to popular Star Trek themes (i.e. medical ethics, regenerative medicine, creationism vs evolution).

11/2010 – 6/2011 **Undergraduate Science Journal Staff Member (Review Board, Biological Sciences), Department of Life Sciences, UCLA, Los Angeles, CA**
I reviewed the suitability and scientific accuracy of submitted review articles for publication in the *Undergraduate Science Journal* at UCLA.

6/2010 – 6/2011 **CityLab Staff Member (Director of Module Development), Department of Molecular, Cell, and Developmental Biology, UCLA, Los Angeles, CA**

I started a new leadership position in CityLab @ UCLA to establish new educational material that educates and challenges high school students on scientific issues and to gauge their intellectual development.

11/2009 – 6/2010 **Undergraduate Science Journal Staff Member (Review Board, Biological Sciences), Department of Life Sciences, UCLA, Los Angeles, CA**
I reviewed the suitability and scientific accuracy of submitted review articles for publication in the *Undergraduate Science Journal* at UCLA.

6/2009 – 6/2010 **CityLab Staff Member (Co-Director of Curriculum Development), Department of Molecular, Cell, and Developmental Biology, UCLA, Los Angeles, CA**
I maintained and strengthened the program's curriculum by introducing new science related topics and expand older ones to participating high school students. I also helped film new educational videos and offered presentations/lectures to introduce high school students to science and college life.

1/2009 – 6/2009 **CityLab Member (volunteer), Department of Molecular, Cell, and Developmental Biology, UCLA, Los Angeles, CA**
I volunteered to educate and mentor high school students of the greater Los Angeles area about scientific topics concerning biotechnology, its applications and impact on society, in order to encourage them to pursue a college education. I also guided high school students in laboratory experiments.

6/2008 – 9/2008 **Laboratory Technician (Project Specialist), Chemistry Department, Saddleback College, Mission Viejo, CA**
I prepared and organized stock solutions and cleaned lab equipment in the chemistry laboratories at Saddleback College, as well as preparing student lab projects for chemistry courses and research.

8/2007 – 6/2008 **Surgery Waiting Room Attendant, Saddleback Memorial Medical Center, Laguna Hills, CA**
I provided comfort, support, and whatever aid necessary for those with friends and family members in surgery.

ABSTRACT OF THE DISSERTATION

Modeling the HIV Genetic Bottleneck *in vitro*

by

Thomas Allen Caldwell

Doctor of Philosophy in Biomedical Sciences

University of California, Irvine, 2024

Professor Donald N. Forthal, Chair

As of the year 2023, there has yet to be developed an effective prophylactic vaccine against HIV-1 infection. One likely reason for prior failures in vaccine trials is that vaccines have been designed against epitopes from non-transmitting variants of the virus. HIV-1 transmission results in a restrictive genetic bottleneck wherein a single viral strain, termed the transmitted/founder (T/F) virus, out of a diverse viral swarm leads to productive infection in recipients, and it is precisely these strains against which a successful vaccine could be developed. While the genetic bottleneck has been studied extensively, the rate of T/F discovery is severely limited by the availability of genetic records from clinical and animal studies, leading to scientific uncertainties regarding T/F selection. In order to accelerate such discoveries, an experimental model was developed to recapitulate the genetic bottleneck *in vitro*, garner insight into the mechanisms of T/F selection, and identify selected polymorphisms under particular physiological and immunological conditions that favor (or disfavor) them. Healthy donor cells (derived from peripheral blood mononuclear cells) were infected *in vitro* with various clinical isolates of HIV-1 in order to test the hypotheses that (1) transcytosis results in T/F selection, (2) *trans*-infection results in T/F selection, and (3) selection pressures can be modulated in favor of different T/Fs using antibodies, complement, or type-1 interferons. gp120 variable loops from

output viruses in these experiments were PCR-amplified, sequenced, and phylogenetically analyzed. Animal studies revealed that antibody may propagate positive selection in favor of SIV variants transmitted through the penile route with a more positively charged V4 loop, while *in vitro* studies showed that antibody did not result in selection during *trans*-infection. Furthermore, complement opsonization of virus led to differential selection patterns favoring viral variants from two different lineages. Finally, priming of unstimulated T-lymphocytes prior to direct infection resulted in profound, replicate-specific genetic bottlenecks of single viral strains that share some polymorphisms with known T/F viruses. The data here demonstrate that a clinical-like genetic bottleneck can be modeled *in vitro*, and such an approach may be an invaluable tool in order to understand the genetic bottleneck and provide useful insight toward the development of an effective vaccine.

CHAPTER 1: Background and Introduction

HIV Genetics and Life Cycle

Human immunodeficiency virus (HIV) is a member of the *Retroviridae* family, *Lentivirus* genus of viruses, all of which are characterized as having a positive-sense viral RNA (vRNA) genome that is reverse-transcribed to complementary DNA (cDNA) upon infection. This is a necessary step in its life cycle in order to incorporate its viral genome into the host genome of infected $CD4^+$ cells, establishing a reservoir of continual viral production. HIV is the primary cause of Acquired Immunodeficiency Syndrome (AIDS), which can be defined as a pathological condition in which the count of $CD4^+$ T-lymphocytes (the most vulnerable target of HIV) in an infected patient's whole blood has reached 350 cells per μL or lower [Barré-Sinoussi et al., 1983; Popovic et al., 1984; Busch, 1988; Schochetman, 1992; Seitz, 2016]. AIDS has also been associated with numerous other diseases, cancers, and opportunistic infections [Macher, 1988; Huang and Crothers, 2009; Ji and Lu, 2017; Longo 2018]. Understanding how HIV causes AIDS and gives rise to comorbidities requires an understanding of its genetics and life cycle in an infected host.

HIV is an enveloped virus containing two copies of its 9.2-kb vRNA genome, itself containing 3 large open reading frames (ORFs) (*gag, pol,* and *env*) and 6 smaller ORFs encoding host restriction factor inhibitors (*vif, vpr, vpu, tat, nef,* and *rev*) and is flanked on its 5' and 3' ends by long terminal repeats (LTRs) [Busch, 1988; Seitz, 2016] (**FIGURE 1A**). The three large ORFs (*gag, pol,* and *env*) each produce precursor proteins that must be cleaved by viral protease in order for each individual protein and enzyme to become functional. The *gag* ORF encodes the structural proteins p24 (capsid protein), p17 (matrix protein), p7 (nucleoprotein), and p6

(exocytotic release protein) as a single precursor polypeptide. The *pol* ORF encodes non-structural proteins or enzymes directly involved in genomic replication and integration into the host genome; these proteins are p10 (HIV protease), p51 (HIV reverse-transcriptase or RT), p15 (HIV ribonuclease H or RNase H), and p32 (HIV integrase). The *env* ORF encodes proteins important for viral spike assembly, and they serve as the infection machinery utilized by HIV against target T-lymphocytes. The precursor protein initially produced by *env* is glycoprotein 160 (gp160), which is cleaved into the mature gp120 and gp41 components of the trimeric spike protein [Pudney and Song, 1994; Zhu et al., 2008; Yuan et al., 2013; Earl et al., 2013; Yang et al., 2022] .

The HIV spike protein is a trimer of heterodimeric pairs of gp120 and gp41 (**FIGURE 1B**) [Yang et al., 2022]. Prior to infection, the spike protein adopts a "closed" conformation. Upon engagement with the necessary host cell receptor (CD4) and a co-receptor (CCR5 or CXCR4), the spike then adopts an "open" conformation, enabling fusion between the host cell membrane and the viral envelope [Dalgleish et al., 1984; Kltazmann et al., 1984; Deng et al., 1996; Dragic et al., 1996; Trkola et al., 1996; Wu et al., 1996; Alexander et al., 2010; Earl et al., 2013; Yuan et al., 2013; Chen, 2019]. The binding interface between gp120 and its receptors is mediated by variable loops 1 (V1) and 2 (V2) with the D1 domain of CD4 and by variable loop 3 and 4 (V3 and V4) with extracellular loops (ECLs) of CCR5 and CXCR4 [Berger et al., 1998; Biscone et al., 2006; Rao et al., 2013; Tamamis and Floudas, 2013; Yuan et al., 2013; Tamamis and Floudas, 2014; Chen, 2019]. HIV's primary target T-lymphocytes are $CD4^+$ T-lymphocytes and macrophages (macrophages), T-lymphocytes being the more vulnerable to cell death than

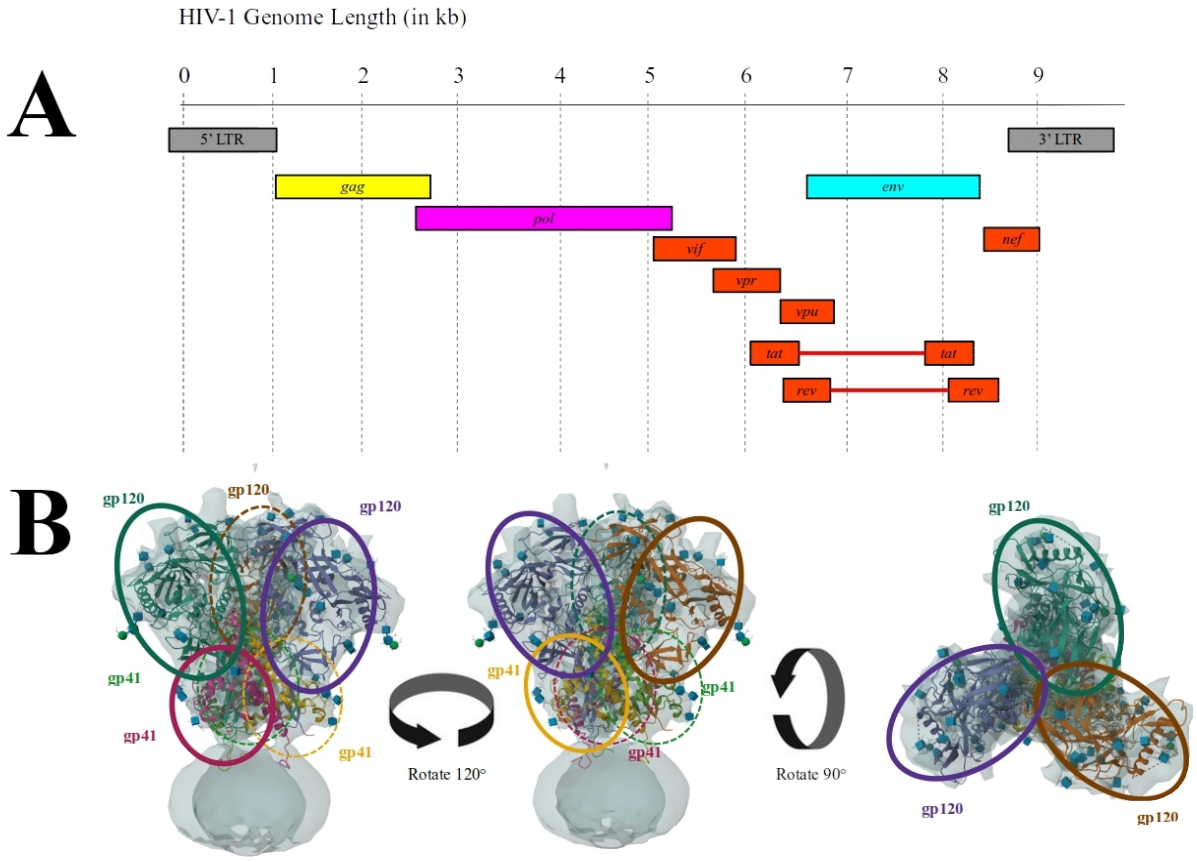


Figure 1: HIV Genetics and Spike Structure: (A) The HIV genome is composed of three major genes: *HIVgag*, *HIVpol*, and *HIVenv*. *HIVgag* encodes structural proteins important for protein capsid assembly, disassembly, and stability. *HIVpol* encodes non-structural proteins and enzymes that promote its replication and maturation to produce progeny viruses. *HIVenv* encodes the pre-processed polypeptide gp160, which is cleaved into gp120 and gp41, each forming the quaternary infection machinery, the HIV spike. (B) The HIV spike protein is constructed from three heterodimers of gp120 (circled in teal, brown, and purple) and gp41 (circled in red, yellow, and green), and it is the main structure responsible for binding to CD4 and co-receptors on the surface of target cells and mediates fusion between cellular and viral membranes for capsid delivery. *Sourced from rcsb.org: ref # 7SC5*. [Yang et al., 2022]

macrophages due to drastic changes in metabolism and immune clearance by $CD8^+$ T-lymphocytes [Kazaki et al., 1989; Shen et al., 2009; Chorin et al., 2014; Palmer et al., 2016]. After membrane fusion, the viral core enters the cytosol and begins disassembly [Arhel, 2010; Ramdas et al., 2020]. After uncoating of p24 from the viral core has begun, the vRNA genome is

then reverse-transcribed to cDNA by HIV RT using host- and virus-derived co-factors. Complementary viral DNA is imported to the nucleus and integrated with the host genome by HIV integrase; the integrated viral genome is commonly referred to as the "provirus" [Goff, 1992]. It is from the HIV provirus that the host cell transcribes viral mRNA, exports the mRNA to the cytoplasm for translation, and the viral core is assembled with structural proteins around replicated viral RNA and non-structural proteins [Qu et al., 2021]. This core then buds from the host cell with mature HIV spike protein and other host cell transmembrane proteins incorporated in the newly formed progeny virus [Huarte et al., 2016; Burnie and Guzzo, 2019; Ramdas et al., 2020; Qu et al., 2021]. The HIV life cycle is reviewed in greater detail (with information on restriction factor inhibitors and their contribution to the life cycle) by Coiras et al. [2010], Seitz [2016], and Ramdas et al. [2020].

Pathogenesis from Infection to AIDS

The progression of AIDS from initial transmission to death can be categorized into three phases: acute phase, latency phase, and AIDS with each phase bearing clinical hallmarks indicative of immunodeficiency [Seitz, 2016]. During the acute phase, HIV enters the host either through mucosal barriers during unprotected sex or during mother-to-child transmission or by bypassing mucosal and epidermal barriers through transfusion with HIV-contaminated blood and sharing HIV-contaminated needles. Symptoms are generally non-specific (not attributable to any certain physiological state or issue), which include fever, raised epidermal lesions, enlarged lymph nodes, and neuropathy, with an incubation period of two to three weeks [Lindbäck et al., 2000; Robb et al., 2016; Seitz, 2016]. During this time, the transmitted virus circulates to

draining lymph nodes either freely, by migrating infected T-lymphocytes, infected macrophages, or adhered to or in the endosomes of migrating dendritic cells (DCs) [Shen et al., 2009; Shen et al., 2010a; Bertram et al., 2019; Perez-Zsolt et al., 2019]; it is in the lymph nodes where the vast majority of resident $CD4^+$ T-lymphocytes become infected [Spira et al., 1996; Masurier et al., 1998; Shen et al., 2010a; Bracq et al., 2018] (reviewed by Manches et al. [2014], Dutartre et al. [2016], and Bracq et al. [2018]). $CD4^+$ T-lymphocyte counts also decrease during the acute phase, but their levels are partially, albeit incompletely, restored three weeks post infection and after cessation of symptoms [Seitz, 2016]. The decline in $CD4^+$ cells during the acute phase and later is largely due to immune clearance by $CD8^+$ T-lymphocytes, natural killer (NK) cells, and macrophages, and neutralizing Abs [Chorin et al., 2014; Manches et al., 2014; Salgado et al., 2014; Lu et al., 2016; Adeniji et al., 2021; Baiyegunhi et al., 2022]. Viral titers range typically from one hundred thousand to one billion vRNA copies/mL at peak viremia [Seitz, 2016]. One case measured a viral titer in one patient with a severe primary infection of HIV at 100 trillion copies/mL [des Roziers et al., 1995].

The latency phase, a period during which $CD4^+$ cell counts decreases progressively albeit at a slower rate than it did during the acute phase. The latency period is also associated with lower viral titers than what is usually measured during the acute phase, ranging from undetectable to as high as 25,000 vRNA copies/mL [Seitz, 2016]. This phase can last as long as several months to 25 years or more, often with the sporadic emergence of few and mild symptoms until the final phase is reached. During this period, $CD4^+$ T-lymphocytes undergo drastic cellular changes that impair their immune function and viability. These changes include metabolic exhaustion, dedifferentiation, chronic activation, attack by autoantibodies, and

metabolic reprogramming by both HIV-1 and $CD8^+$ T-lymphocytes [Bost et al., 1988; Salgado et al., 2014; Palmer et al., 2016].

Upon the onset of AIDS, the patient's immune system is severely compromised by a $CD4^+$ cell count below 200 cells per μL and a progressively expanding virus titer. During this phase, the impairment of immune recognition renders the patient highly vulnerable to opportunistic infections, cancers, reactivation of other latent viral infections, and neurodegeneration [Plata et al., 1987; Macher, 1988; ; Huang and Crothers, 2009; Dohgu et al., 2011; Ji and Lu, 2017; Yarchoan and Uldrick, 2018; Cantres-Rosario et al., 2019; Straehley et al., 2020]. Morality due to comorbidities is inevitable without therapeutic intervention through anti-retroviral therapy (ART) [Gallay et al., 2017; Zhang et al., 2021].

Factors Influencing Risk of Transmission and Disease Progression

Physiological and immunological outcomes, even the site of transmission, during the acute phase all play an important role in an infected patient's prognosis and progression toward AIDS; they can determine whether a patient may progress rapidly, slowly, or not at all. The risk of HIV transmission through vaginal, rectal, penile, epidermal, and oral mucosae have been reported to result in differences in measured risk and, in some cases, clinical outcomes. The oral and epidermal routes are considered to have the lowest risk of HIV transmission. However, due to the paucity of HIV infection cases through the oral and epidermal routes, it is difficult to accurately measure the relative risk of these two routes of exposure [Rothenberg et al., 1998; Yu and Vajdy, 2010]. It is likely that the risk is higher than previously considered. The risk of transmission through the oral route, for example, could be influenced by any number of factors

that include, but are not limited to, the donor's viral titer, immune status of the recipient, and co-occurring infection and inflammation (like HSV-1/HSV2 cold sores) [Rothenberg et al., 1998]. Another factor influencing HIV transmission risk through the oral route is the pH of the mouth and stomach. Aldunate et al. [2013] reported an acidic environment (< pH 4) reduced the infectivity of HIV by 99.8% compared to untreated controls. By deduction, it can be argued that acidity negatively influences HIV transmission, especially through the gastrointestinal (GI) tract, in so far as open sores in the mouth may positively influence it through the oral route. Animal studies have also shown that oral transmission, should it occur, may be mediated via transcytosis through the tonsils; although these results are discordant with other animal studies demonstrating that oral transmission was so rare it could not be studied (reviewed by Rothenberg et al. [1998]). The epidermal route is equally restrictive of HIV transmission to a new host. The epidermis is comprised of multiple layers of tight-junctioned cells, making cell-free viral transcytosis through the epidermis exceedingly more difficult than it could be through the oral route. However, a recent study by Bertram [2019] demonstrated that epidermal Langerhans cells (LCs) and $CD11c^+$ DCs potently enhance HIV passage through a human abdominal explants into the mucosa, mediated by non CD4-dependent and CCR5/CXCR4-dependent binding. This finding also corroborates earlier studies that identified LCs as a potential carrier and target of HIV infectious particles [Miller et al., 1999; Kawamura et al., 2008]. In light of these studies, the possibility of HIV transmission through the epidermal and oral routes cannot be discounted, even if their relative risks compared to other routes are immeasurably low. It can be argued, in fact, that it is because of the rarity of these cases that its risk may be higher than expected; more research is necessary before oral and epidermal HIV transmission risk could be completely understood.

The penile, vaginal, and rectal routes of exposure are, in increasing order, associated with much higher risks of transmission, with the rectal route remaining the mode of transmission associated with the highest measured risk to date [CDC HIV Surveillance Report, 2021]. The penile foreskin is one of the more vulnerable routes of HIV transmission on the male, aside from the rectal route [Pudney and Anderson, 1995; Ma et al., 2011; Dinh et al., 2015]. Male circumcision is associated with a 60% decrease in HIV infection risk [Tobian et al., 2014; Prodger and Kaul, 2017]. Bertram et al. [2019] identified epidermal LCs as being a potent carrier of HIV, not just in the epidermis but also in the inner and outer foreskin. In fact, compared to the epidermis, $CD11c^+$ DCs were more likely to passage HIV from the foreskin than from the epidermis [Bertram et al., 2019]. The data highlighted by Bertram et al. [2019] corroborates earlier findings in which the inner foreskin, glans penis, and urethral epithelium were highly vulnerable to HIV transmission in uncircumcised men [Pudney and Anderson, 1995; Patterson et al., 2002; Fischetti et al., 2009; Dinh et al., 2015]. Indeed, cell-mediated viral transfer across penile epithelium is not the only mechanism involved in the penile route of exposure. The squamous keratinized epithelium of the foreskin, glans penis, and urethra are thin enough to facilitate viral transcytosis from the exterior of the foreskin to the mucosae beneath it [Anderson et al., 2011].

The vagina is also highly vulnerable to HIV transmission, perhaps even more so than the uncircumcised penis is. In one of the earliest studies on HIV transcytosis, Hocini et al. [2001] demonstrated non-cell-mediated transcytosis across a model endometrium *in vitro*, suggesting one of several possible routes of transmission in the female genital tract. This transcytotic pathway is likely a real biological phenomenon of mucosal transmission and not just an artificial

consequence of utilizing abnormal cell lines in tissue culture [Kinlock et al., 2014]. In fact, comparative anatomy of the vagina, cervix, and penis reveals much thinner keratinized squamous epithelium in female urogenital tissues than in that of the male [Yu and Vajdy, 2010; Kaul et al., 2022]. This facilitates HIV transmission through the vaginal and cervical routes in two ways: (1) It increases the efficiency of viral transcytosis across the epithelium to the mucosa and (2) it potentiates interdigitation of LCs from the labia and DCs from the vagina and cervix, increasing the likelihood of cell-mediated viral transfer across the epithelium to the mucosa [Hu et al., 2000; Shen et al., 2010c; Bertram et al., 2019; Kaul et al., 2022]. Birth control, like medroxyprogesterone, has been shown to enhance HIV transcytosis by reducing lysosomolysis in cervical epithelial cells, a contributing risk factor making women more vulnerable to sexually transmitted HIV [Jia et al., 2021]. As was the case with oral transmission, acidity of the vagina also reduces HIV transmission risk by inactivating infectious virions, complicating the assessment of overall risk by counteracting factors [Aldunate 2013]. Despite this complexity, HIV transmission through female sexual tissues remains second highest among the known routes of sexual exposure [Kaul et al., 2022]. While the global incidence of HIV infections has decreased in the last 40 years, females carry a higher disease burden than men do globally, especially in sub-Saharan Africa [Kaul et al., 2015; De Cock, 2021; Kaul et al., 2022; Zhang et al., 2022]. This may not be due to clinical differences between men and women so much as it is likely due to inaccessibility to medical treatment and societal stigma related to HIV infection in women [Kaul et al., 2022; Zhang et al., 2022].

Among males and females, the penis and the vagina are not the only routes of exposure posing a significant risk to HIV transmission. The rectal route is currently the means of

transmission with the highest known risk of HIV infection, impacting both men and women [Shen et al., 2010a; Shen et al., 2010b; Kolodkin et al., 2013]. Two studies found that the higher transmission risk of penetrative anal sex could be attributed to tissue damage, higher pro-inflammatory cytokine production, and higher levels of activated T-lymphocytes in the rectal mucosa [Kelley et al., 2017; Yamada et al., 2021]. Viral penetration of rectal mucosal tissues likely occurs through caveolin-directed transcytosis, barring microabrasions caused by frictional sex [Anwar et al., 2022]. Another important feature of the rectal mucosa that might be associated with a higher risk of transmission is the proximity of iliac lymph nodes to the rectum, where $CD4^+$ T-lymphocytes normally reside [Yu and Vajdy, 2010]. This could explain why rapid disease progression is common among men who have sex with men (MSM) [Zhang et al., 2021]. Fortunately, pre-exposure prophylaxis (PrEP) and ART have demonstrated significant reductions in HIV transmission risk through the vaginal and rectal routes [Gallay et al., 2017].

The vaginal, rectal, and penile mucosae are patrolled by surveilling monocyte-derived cells like macrophages and DCs, and their purpose is to modulate effector functions in order to control the growth of commensal microbes and capture invasive pathogens for antigen processing and antigen presentation [Shen et al., 2010a; Shen et al ., 2010b; Cook and MacDonald, 2016; Lindenbergh and Stoorvogel, 2018; Stagg, 2018; Trifonova et al., 2018; Bertram et al., 2019; Kogut et al., 2020; Perez-Zsolt et al., 2019; Tezuka and Ohteki, 2019]. HIV is a pathogen distinct for its ability to “hijack” macrophages and DCs and exploit their migratory synaptic abilities to infect many more target T-lymphocytes through a process known as *trans*-infection (reviewed by Dutartre et al. [2016] and Bracq et al. [2018]). *Trans*-infection is the process by which a DC or macrophage containing infectious HIV particles on its surface or

inside endosomes then transfers infectious particles to $CD4^+$ T-lymphocytes more potently than through cell-free infection of target T-lymphocytes without antigen presentation from DCs and macrophages [Geijtenbeek et al., 2000; Hong et al., 2002; Nabatov et al., 2006; Bouhlal et al., 2007; Kolodkin et al., 2013; Trifonova et al., 2018]. *Trans*-infection in this regard is likely a contributing factor to dissemination of the virus [Koh et al., 2020]. In fact, Trifonova et al. [2018] reported evidence of potent *trans*-infection in mucosal explants. Of the five ectocervical explants the authors used, the HIV-1 clinical isolate JRCSF was capable of infecting $CD14^+$ DCs, and three of these explants contained $HIVgag^+$ $CD4^+$ T-lymphocytes. Furthermore, HIV-infected or -adhered DCs from explanted tissues were highly effective at transferring virus to PBMC-isolated $CD4^+$ T-lymphocytes. This suggests infection through the cell-associated *trans*-mechanism is a probable occurrence in HIV transmission through mucosal tissues, whether *trans*-infection occurs at the site of exposure or distal to it (i.e. lymphatic tissues).

A recent study has demonstrated that *trans*-infection is so potent that it occurs even in the presence of darunavir and maraviroc *in vitro*, an HIV-1 protease inhibitor and CCR5-binding competitor, respectively [Rappocciolo et al., 2019]. It remains unclear whether the finding reported by Rappocciolo et al. [2019] bears any significant clinical relevance for HIV-seropositive individuals undergoing ART to transmit HIV to uninfected individuals, but cell-associated transmission, more so than cell-free transmission, across mucosal barriers remains a reality of enhanced HIV transmission risk in unsafe sexual practices, either through the rectal mucosa [Kolodkin et al., 2013] or the female genital tract [Hocini et al., 2001]. Given that the rate of disease progression is directly related to the decrease in $CD4^+$ T-lymphocyte counts in acutely infected individuals, it can be inferred that *trans*-infection of T-lymphocytes would be a

probable factor in accelerating disease progression more so than direct infection of T-lymphocytes by cell-free viral particles alone, provided that all other factors are equal [Nabatov et al., 2006; Robb et al., 2016; Zhang et al., 2021].

It should be noted, however, that although numerous studies report monocyte-derived cells as being the initial cell type that disseminates the virus to lymph nodes [Hu et al., 2000; Arthos et al., 2008; Koh et al., 2020], and in some cases infected cells from the donor travel to the lymph nodes, achieving the same outcome [Ibata et al., 1997], this is contradicted by other findings that demonstrate T-lymphocytes are initially infected by SIV *in vivo* inoculation of the vagina and that the virus is only transferred to monocytes upon phagocytosis of infected cells [Zhang et al., 1999; Stieh et al., 2016]. There may be some doubt as to the clinical relevance of the paper by Stieh et al. [2016] as the authors used non-human primates (NHPs) as the animal model for a virus similar, but not identical to, HIV. Another shortcoming in this study is the lack of data comparing virus-bound $CD11c^+$ monocytes to $CD4^+$ T-lymphocytes: Just because monocytes weren't shown to be infected initially doesn't mean they didn't artificially enhance infection of T-lymphocytes in *trans*. An additional confounding factor is that pre-existing inflammation at the sites of exposure in some of the animals were reported, and their causes and contribution to transmission could not be accounted for. Even if *trans*-infection did not occur in these animals, inflammation could have recruited and stimulated T-lymphocytes at the site of exposure (reviewed by Mueller and Strange [2004]), which would artificially influence the results and doubts would remain as to the probability of this occurring in every recipient of HIV transmission. In the study performed by Zhang et al. [1999] a similar methodological issue was found, even in the data obtained from HIV-1-infected individuals. Furthermore, tissues from

human subjects were obtained by biopsy at the onset of symptoms, approximately twelve or more days post infection. The tissue biopsies used in that study were from lymph nodes, not the tissues at the potential site of transmission in question. The HIV-1-infected T-lymphocytes from biopsied lymph nodes do not strictly represent the initial cells that were infected at the time of transmission, and even if they were data acquired so late in the course of infection does not rule out the possibility that a subset of those cells were infected by *trans*-infection.

Another prior study found that LCs and DCs, not T-lymphocytes, were preferentially infected early during SIV infection using similar methods and model system [Hu et al., 2000]. Regardless, it is nugatory to argue whether one cell type is preferentially infected over others early in acute infection; the point of the matter here is that disease outcome and progression, as well as transmission selectivity (this concept will be introduced in a later section), may be influenced depending on how, where, and the degree to which a productive infection by HIV is established in a patient. Unfortunately, these are not well understood currently, and at best one can only conjecture how tissue-specific transmission and cell-specific infection modulate disease progression and transmission selectivity.

Another important factor influencing transmission risk is a concurrent sexually transmitted infection (STIs) . Inflammation and open sores at the site of HIV transmission have been reported to enhance susceptibility to HIV infection [Haaland et al., 2009]. A case study even showed that non-ulcerative STIs were transmission-enhancing factors for HIV-1 infection, suggesting that open sores and ulcers are not necessary to increase HIV transmission risk during an STI [Laga et al., 1993]. According to data analyzed from the 2017 DC Cohort Longitudinal HIV Study, approximately 40% of study participants with at least one confirmed diagnosis of

chlamydia, gonorrhea, and/or syphilis also had detectable levels of plasma HIV within several months of diagnosis [Lucar et al., 2018]. This suggests that concurrent bacterial STIs are associated with a higher HIV transmission risk. A similar finding was also found among HIV-infected people receiving care in Thailand [Tunthanathip et al., 2009]. Dohgu et al. [2011] identified bacterial lipopolysaccharide (LPS) pro-inflammatory cytokines as agents that enhance HIV transcytosis across the blood-brain barrier. Not only could this explain the dementia and neurodegeneration among individuals living with HIV, it very well could be mechanism that facilitates HIV transcytosis across genital and rectal mucosal barriers in individuals with a concurrent bacterial STI. The topic of concurrent infections influencing HIV transmission risk is also reviewed extensively by Macher et al. [1988], Cone [2014], Kaul et al. [2015], and Passmore et al.[2016].

Pudney et al. [2019] reported that anogenital warts caused by low-risk human papillomaviruses (HPVs) harbored higher concentrations of DCs, macrophages, and $CD4^+$ T-lymphocytes than what is normally contained in non-inflammed epidermis. A meta-analysis from 2015 showed that concurrent genital infection by herpes simplex virus 2 (HSV-2) was associated with a population attributable risk of nearly 50% among HIV-infected women in Kenya from 2008 – 2012 [Masese et al., 2015]. This report also showed that HSV-2 infection contributes the highest increase HIV transmission risk compared to other concurrent infections like gonorrhea and HPV [Masese et al., 2015]. Non-STIs, like yeast infection and bacterial vaginosis, are also associated with a 1.7-fold increase in the risk of HIV transmission [Atashili et al., 2008; Low et al., 2011; Masese et al., 2015]. This is likely due to an increase in vaginal pH, recruitment of target T-lymphocytes to vaginal and cervical epithelium triggered by tissue

inflammation, and interdigitated leukocytes that could carry HIV from the donor's cell-associated virions [Atashili et al., 2008; van de Wijgert et al., 2008; Masese et al., 2015].

It is interesting to note that in addition to concurrent STIs, microabrasions caused by sexual activity, even in healthy subjects, increases the risk of HIV transmission [Kawamura et al., 2008; Kelley et al., 2017]. This is likely due to structurally damaged mucosal and epithelial barriers in the penis and especially in the vagina [Fraser et al., 1999]. Such damage is sufficient to induce a pro-inflammatory response and recruit immune cells, such as $CD4^+$ T-lymphocytes, Th17 cells, LCs, and DCs, to the site of the damage in order to repair it and prevent bacterial infections, where HIV could either infect or adhere [Kawamura et al., 2008; Kaul et al., 2015; Passmore et al., 2016; Stieh et al., 2016; Kelley et al., 2017].

Epidemiology and Demographics of HIV/AIDS

As of 2022, 39 million people worldwide live with HIV with approximately 1.3 million new cases and 630,000 HIV-related deaths [UNAIDS/WHO, 2023]. During 2022 in the United States alone, 36,136 new cases of HIV were found by CDC surveillance programs [Zhang et al., 2022; CDC, 2023]. Of those new cases, 18% were women and 79% were men, the majority of whom contracted HIV through male-to-male sex. Black/African American are disproportionately the racial group impacted the most by the HIV epidemic, comprising nearly 40% of new HIV cases in 2021; Hispanice/Latino populations are the second most impacted racial group in the United States at 29% of new cases. [CDC, 2023]. Racial discrimination, social stigma, poverty, mass incarceration, and homophobia were identified as being persistent barriers to healthcare access for racial/ethnic minorities in the United States and abroad

[Pitisuttithum et al., 2008; CDC, 2023]. Meta-analysis by Andersson et al. [2020] also showed that removing such barriers would significantly improve the health-related quality of life for those living with HIV that also face discrimination and stigma.

While unprotected male-to-male sex was found to be the most common mode of transmission in the United States (accounting for approximately 71% of new infections in 2021), the same statistic does not apply to other countries worldwide [Patel et al., 2014; CDC, 2023]. In nations in sub-Saharan Africa, heterosexual HIV transmission is the most common mode of HIV acquisition, accounting up to three quarters of the global incidence of HIV [Trask et al., 2002; Kaul et al., 2015; Monaco et al., 2017; Bbosa et al., 2019b; De Cock et al., 2021]. In Southeast Asia, HIV acquisition through intravenous drug use, transfusion with contaminated blood, and sex work are significant contributors to the occurrence of new cases [Tunthanathip et al., 2009; Patel et al., 2014; Monaco et al., 2017].

HIV most likely evolved from a Chimpanzee-specific simian immunodeficiency virus (SIVcpz) [Heeney et al., 2006; Sharp and Hahn, 2011] There are two species of HIV: HIV type 1 (HIV-1) and HIV type 2 (HIV-2). HIV-1 is the most common species of HIV worldwide, and it is comprised of multiple subtypes, two of which (B and C) are the most common [Bbosa et al., 2019a]. HIV-1 subtype B is predominant in North America, South America, northern Africa, Europe, the Middle East, and Australia, while HIV-1 subtype C is predominant in sub-Saharan Africa, Southeast Asia, and the Pacific Basin [Bbosa et al., 2019a]. **FIGURE 2** shows a phylogenetic tree of patient HIV-1 sequences using SIVcpz as an outgroup to demonstrate the phylogenetic relationship of the various subtypes of HIV-1 with its most common ancestor.

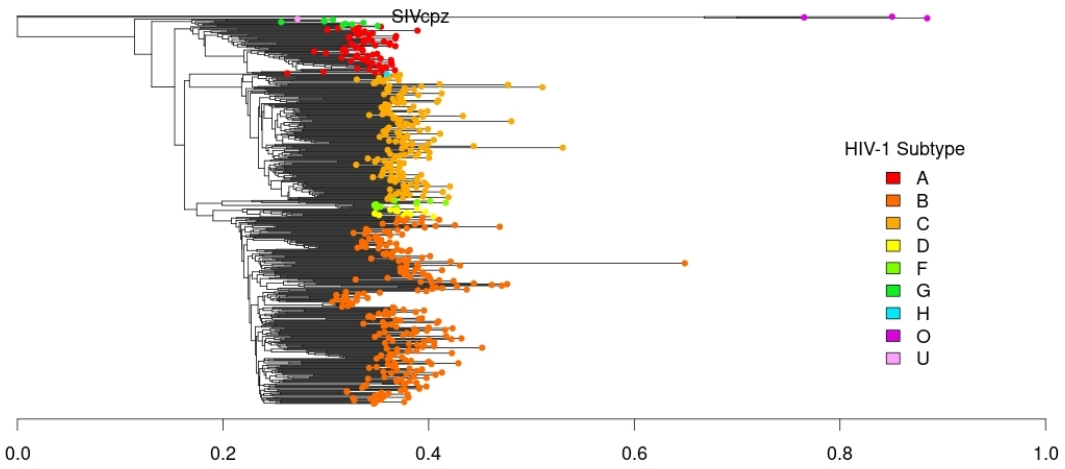


Figure 2: Phylogenetic Tree of Clinical Isolates by HIV-1 Subtypes: Maximum-likelihood tree for sequences from 4,859 randomly selected group M clinical isolates in the HIV-1 sequence database from hiv.lanl.gov. Clinical isolates cluster according to their subtypes; SIVcpz was chosen as an outgroup. Units in the x-axis represents numbers of amino acid changes per site.

Prevention and Therapeutic Strategies

Since the 1990s, the yearly incidence of new HIV infections and HIV-related mortality have dropped significantly. This improvement was enabled by public health measures aimed at expanding access to healthcare and education to those with the highest HIV acquisition risk [De Cock et al., 2011; De Cock et al., 2021]. In fact, De Cock et al. [2011] recommended that continuing the downward trend in new HIV cases would require innovating new clinical interventions, employ HIV surveillance and testing measures for the public, expanding and societal access to ART and PrEP, and collaborative international efforts to work toward a vaccine and a cure.

The goal of ART is two-fold: (1) treatment of HIV-infected individuals in order to improve quality of life and prevent the onset of AIDS and (2) prevent the spread of HIV from seropositive individuals to uninfected individuals. Long-term adherence to ART is associated with viral suppression and prevention of AIDS in HIV-infected individuals [Li et al., 2011]. Treating HIV involves treatment regimen designed to inhibit viral enzymes necessary for HIV infection and replication. The most common drug targets are HIV reverse-transcriptase, protease, and integrase [Pau and George, 2014]. FDA-approved and commercially available nucleoside drug inhibitors of HIV reverse-transcriptase include abacavir, emtricitabine, lamivudine, tenofovir, and zidovudine; non-nucleoside inhibitors include efavirenz, etravirine, nevirapine, and rilpivirine [Pau and George, 2014]. Rilpivirine, for example, was shown to effectively suppress HIV replication in nearly 93% of study participants receiving the long-acting form of the drug, requiring only monthly injections instead of daily oral doses [Swindells et al., 2020]. Inhibitors of HIV protease include darunavir, fosamprenavir, lopinavir, saquinavir, tipranavir, and atazanavir [Pau and George, 2014]. Lastly, inhibitors of integrase include dolutegravir, elvitegravir, and raltegravir [Pau and George, 2014]. There is also maraviroc, which blocks CCR5 on target T-lymphocytes to prevent HIV from binding to them [Pau and George, 2014]. Maraviroc can only be used to treat patients infected with CCR5-tropic HIV; a dual infection with CCR5- and CXCR4-tropic viruses could produce a selective pressure favoring the CXCR4-tropic virus during treatment [Matume et al., 2020].

These drugs are extremely effective at suppressing HIV production in seropositive individuals, even in cases where an individual contracted Chlamydia or Gonorrhea during ART [Davies et al., 2017; Storim et al., 2019]. They were also effective at abrogating a higher

morality in HIV-infected individuals caused by endothelial damage [Affi et al., 2022]. Until recently, it wasn't known whether controllers required treatment; controllers are asymptomatic individuals living with HIV that have a mostly unimpaired immune system and are able to suppress HIV replication on their own without treatment. The INSIGHT START Study Group [2015] found that ART is safe for controllers and recommended that they begin treatment even if their $CD4^+$ cell counts are higher than 350 cells per μL, now part of standard treatment for asymptomatically infected individuals. Despite a recent finding that *trans*-infection of $CD4^+$ T-lymphocytes is possible with ART *in vitro* [Rappocciolo et al., 2019], the clinical relevance of which is still in question, ART has played a very important role in reducing the global spread of HIV. Indeed, consistent treatment of infected individuals is crucial to curbing the HIV pandemic, as there is evidence that episodic ART rather than continuous ART and unsafe sexual practices increases the risk of transmitting HIV from seropositive donors to even low-risk recipients [Burman et al., 2008; Tunthanathip et al., 2009].

As clinical research and drug trials continue, researchers are discovering new potential therapies for HIV suppression. A randomized phase 2 trial on the clinical efficacy of lenacapavir found that subcutaneous delivery of lenacapavir semi-annually was associated with a 90% suppression of HIV replication in seropositive individuals [Gupta et al., 2023]. Lenacapavir is a p24 blocker that prevents capsid assembly, nuclear import, and HIV budding. Another study involving a phase 3 randomized drug trial found that injectable cabotegravir, a third generation integrase inhibitor with a low likelihood of resistance, was equally as effective at suppressing HIV production from hosT-lymphocytes as rilpivirine [Swindells et al., 2020]. Data reported by Swindells et al. [2020] corroborated results from a study one year prior finding that cabotegravir

can enhance the effectiveness of oral tonofovir and emtricabine [Glidden et al., 2019]. Drug research and design led to many other candidate compounds with the potential to treat infected individuals but have not become commercially available for HIV treatment (likely because viral enzyme inhibitors are the most effective forms of therapy to date). In 2004, Balzarini et al. reported on the anti-HIV microbial activity of mannose-binding plant lectins from members of the *Amaryllidaceae* family. They were shown to adhere to HIV gp120, effectively blocking its contact with CD4, CCR5, CXCR4, and DC-SIGN (a cell surface receptor of DCs) [Balzarini et al., 2004; Saïdi et al., 2007]. While their anti-viral property had been demonstrated, mannose-binding lectins did not enhance the therapeutic efficacy of FDA-approved ART drugs [Balzarini et al., 2004]. Another drug under study with anti-viral potential was cyanovirin-N. Cyanovirin-N is a small anti-viral protein produced by *Nostoc ellipsosporum* (a species of cyanobacteria), likely as a defense against bacteriophages and other viruses. Boyd et al. [1997] showed that at 3 nM, cyanovirin-N was sufficient to reduce RT activity in tissue culture. The mechanism of action was proposed to be the blockade of gp120 at the interface between gp120 and CD4 and CXCR4 [Mori and Boyd, 2001]. However, the therapeutic efficacy of cyanovirin-N was not strong enough to enhance currently available ART drugs, as was the case for plant lectins; furthermore, cyanovirin-N was insufficient at blocking HIV adherence to DCs, which enhances viral transfer to $CD4^+$ T-lymphocytes through *trans*-infection [Hong et al., 2002]. At the very least, these candidate drugs, although insufficient to block HIV infectivity and replication, proved invaluable as research tools to study mechanisms of virus-host interaction and are still used to this day for that purpose.

PrEP, which is a medical prescription for at-risk individuals in order to protect them from acquiring HIV, has been equally instrumental in reducing HIV spread in the North American population. In one study trial using FDA-approved ART drugs as form of prophylaxis, maintaining a regimen of anti-retroviral drugs in HIV-uninfected individuals resulted in a 44% reduction in HIV incidence among MSM [Grant et al., 2010]. This finding was corroborated with similar results in more recent studies, and the reduction in transmission risk was fine-tuned to > 90% [Molina et al., 2015; McCormack et al., 2016; Pitisuttithum and Marovich, 2020]. PrEP was also equally effective among heterosexual men and women [Cohen et al., 2011; Baeten et al., 2012].

Regular testing is also an important factor in curtailing the yearly incidence of HIV (standard laboratory diagnostic tests for HIV are reviewed in detail by Busch et al. [1988] and Schochetman et al. [1992]). From 1987 to 2000, HIV antibody tests have become reliably more sensitive and faster (results could now be acquired in minutes instead of hours) that routine, in-clinic testing and reporting is now standard practice for HIV clinics and diagnostic centers [Lindbäck et al., 2000]. Current knowledge on the correlation between viral load, diagnostic measures of infection, and disease progression has enabled researchers and clinicians to adopt a diagnostic staging system that can directly inform prognosis on the basis of categorical test results, facilitating care for HIV patients and preventing further spread [Fiebig et al., 2003]. Lowering the spread of HIV is not only a matter of accurate tests and healthcare accessibility; it also requires reporting new diagnoses to regulatory agencies and using computer models to describe HIV spread in a population. Open source software exists to model such spread; the HIV Transmission Cluster Engine (HIV-TRACE) was developed to allow researchers to study HIV

epidemiology from a molecular perspective [Pond et al., 2018]. Programs like this would make documenting and modeling the spread of new HIV variants (such as the highly virulent VB variant discovered in the Netherlands by Wymant et al. [2022]) more cost-effective and easier.

These research and public health measures have played important roles in controlling the HIV pandemic and they continue to do so. However, the ultimate goal of HIV prevention research is the development of a prophylactic vaccine to protect seronegative individuals from contracting the virus. While it is not a direct aim of the current topic of this dissertation, the data shown and discussed here may one day provide valuable insight into HIV vaccinology.

Vaccine Trials and Their Failures

As stated previously, a regimen of ART with periods of nonadherence increases the risk of HIV transmission from seropositive individuals [Burman et al., 2008]. It has even been shown to allow reactivation of latent viruses that exhibit resistance against the host's antiviral cytokines in cells wherein cytokine signaling is already impaired [Ranganath et al., 2016; Gondim et al., 2021]. Forgetfulness and even a lack of willpower adhering to ART at the individual level could compromise public health initiatives in reducing HIV spread. The same can also be true for PrEP in uninfected individuals [Rolle et al., 2019]. Meta-analysis on 156 peer-reviewed articles showed that education, income level, and awareness of commercially available prophylaxis were the main determinants of whether MSM sought to use PrEP [Sun et al., 2022]. Although PrEP effectively reduced HIV acquisition, two of three research participants in one study that acquired HIV, despite being on PrEP, developed resistance mutations to emtricitabine [McCormack et al., 2016]. Given that selective pressure favoring resistance

mutations in individuals on PrEP, however rare HIV acquisition might be, there is still a possibility of resistance emerging in those individuals infected with HIV while adhering to PrEP regimen.

For those that become infected with HIV, the life expectancy of people living with HIV while adhering to ART is still lower than that of the general population [Samji et al., 2013]. The percentage of people living with HIV in the United States was only 62% as of 2017, only slightly higher than the global percentage at 59% that same year [D'Souza et al., 2019]. The rate of decrease in global incidence of HIV has waned in the period from 2015 – 2020, especially in Southeast Asia [Brown and Peerapatanapokin, 2019]. In order to continue the trend of incidence reduction as seen during 1990s and early 2000s, it is imperative for medical science to achieve a prophylaxis beyond anti-retroviral drugs.

From 2000 to 2020, the literature on vaccine targets and strategies against HIV expanded considerably, offering researchers invaluable insights into vaccine design with high potential to generate broadly-neutralizing antibodies (bnAbs). Efficacious immunity against HIV is contingent upon a vaccinated individual generating bnAbs that bind multiple epitopes of the HIV envelope spike in order to block its infection mechanism. More importantly, the plasma of HIV-infected individuals that show superior viremic control serves as an invaluable resource of bnAbs that could be used to identify novel immunogenic epitopes and formulating immune complexes that would greatly improve vaccine development [Haaren et al., 2017], though this may be a point of contention given that rational design is potentially reductive and overly simplistic [Regenmortel et al., 2018]. Despite the classification problems inherent with certain approaches to vaccine designs, there is currently a large body of evidence on bnAbs and various vaccine

designs that give rise to them in clinical and preclinical experiments, many of which have been instrumental in vaccine trials (reviewed by Liu et al. [2020]).

Immunoglobulin-G (IgG) and immunoglobulin-A (IgA) antibodies against the HIV spike protein were shown to block transcytosis [Bomsel et al., 1998; Hocini and Bomsel, 1999; Matoba et al., 2004; Shen et al., 2010b; Jain and Rosenthal, 2011], impede the viral fusion mechanism [Devito et al., 2004], and mediate antibody-dependent cellular cytotoxicity (ADCC) [Koup et al., 2010; Smalls-Mantey et al., 2012; Chorin et al., 2014; Lu et al., 2016; Huang et al., 2016a; Naiman et al., 2019; Doepker et al., 2021], which is an effector function carried out by $CD8^+$ T-lymphocytes and natural killer (NK) cells to destabilize the plasma membrane of infected cells and induce autophagy by chemical signalers. Broadly neutralizing antibodies (bnAbs) against gp160 epitopes isolated from elite controllers and vaccinated subjects exhibit viremic control primarily through enhancement of enhance ADCC responses and secondarily through blockade of CD4- and CCR5-/CXCR4-binding sites [Devito et al., 2004; Koup et al., 2010; Smalls-Mantey et al., 2012; Rao et al., 2013; Chorin et al., 2014; Caskey et al., 2015; Huang et al., 2016a; Lu et al., 2016; Heß et al., 2019; Naiman et al., 2019; Doepker et al., 2021]. In fact, NK cells and $CD8^+$ T-lymphocytes from HIV-infected controllers (individuals infected with HIV that have successfully suppressed viral outgrowth without ART) show robust ADCC activity against HIV-infected cells [Taborda et al., 2015], suggesting that ADCC may be an important effector function that not only significantly impairs virus in infected individuals but also may be important for lasting immunity against HIV in potentially vaccinated individuals.

There is also literature on an unconventional approach to an HIV vaccine using human cytomegalovirus (hCMV). Hansen et al. [2019] demonstrated that an engineered live, attenuated

Rhesus CMV ($RhCMV_{68\text{-}1}$) could be used in combination with an SIV challenge to provide vaccinated animals up to 60% protection against SIV through $CD8^+$ T-lymphocyte-mediated ADCC. These results, including confirmation of the effector function giving rise to this protection, have since been corroborated by two studies [Malouli et al., 2021; Verweij et al., 2021]. Given that SIV is a close evolutionary homolog of HIV, this unique vaccine approach was considered as a possible research avenue in HIV vaccinology, and in follow-up Yang et al. [2021] reported that human $CD8^+$ T-lymphocytes could be primed with $RhCMV_{68\text{-}1}$/HIV*gag* proteins to kill HIV-infected cells through the same effector function *in vitro*. While the main effector function elicited by this vaccine doesn't directly neutralize the virus, it does enhance ADCC in a manner that it potently and rapidly destroys infected $CD4^+$ T-lymphocytes before productive infection is established. Koup et al. [2010] and Fitzgerald et al. [2011] also showed vaccination with adenoviral vectors expressing HIV gp120 and gp41 resulted in robust ADCC responses against HIV infection *in vitro*.

Another novel vaccine design involves a trimeric HIV envelope spike protein designated as SOSIP.664 gp140 [Sanders et al., 2013]. SOSIP.664 gp140 was designed to mimic the native conformation of the spike on virions and to generate broadly neutralizing antibodies that recognize the spike protein in its native state [Sanders et al., 2013; Sanders et al., 2015]. The HIV-1 epitope associated with the strongest adaptive immune response was derived from an infant (BG505) infected with HIV-1 through mother-to-child transmission that later developed broadly neutralizing antibodies [Goo et al., 2014]. Since then, additional SOSIP vaccines have been generated that target epitopes from multiple viral strains across multiple HIV clades, and they maintain the spike protein subunits in similar quaternary structures [Schorcht et al., 2020].

Animal studies using this design showed evidence of antibodies that were broadly neutralizing against subtypes A, B, and C, without non-neutralizing byproducts (the presence of which could enhance viral infectivity through ADE) [Gorlani and Forthal, 2013; Sanders et al., 2013; Julien et al., 2015; Sanders et al., 2015; Tumba et al., 2022]. To date, the adenoviral delivery system and SOSIP design are two of the most promising avenues for an HIV vaccine.

Despite the current knowledge on HIV vaccine targets, there is, as of yet, no working vaccine available: Where several working vaccines against SARS-CoV-2 had been developed within a year, nearly non had been developed against HIV for nearly 40 years. There has been 13 HIV vaccine trials and only one of them generated some efficacy but was still insufficient for commercial use [Ng'uni et al., 2020]. Indeed, even animal models using SIV epitopes to vaccinate *Rhesus macaques* failed to produce antibodies with sufficient neutralizing capacity [Letvin et al., 2007; Qureshi et al., 2012]. The first vaccine trial (VaxSyn) was in 1987, and it utilized a vaccine made from recombinant HIV gp160 (rgp160) [Dolin et al., 1991] While antibody production occurred, they were hardly neutralizing and failed to grant vaccinated subjects long-term protection. [Dolin et al., 1991]. Another vaccine was developed 10 years later called AIDSVAX. The immunogen used for this vaccine was bivalent rgp120 from CCR5-tropic and CXCR4-tropic HIV-1 B and E, and it was reasoned that failure in the last trial could have been due to selection pressure in vaccinated subjects favoring one form of tropism over another [Ng'uni et al., 2020]. AIDSVAX was tested in two other phase 3 vaccine trials from 1998 – 2003 in Thailand (Vax003) and North America (Vax004) [Ng'uni et al., 2020]. Data from both trials, however, produced a similar failure [The rgp120 HIV Vaccine Study Group, 2005; Pitisuttithum et al., 2006]. Shortly thereafter, another vaccine was implemented called ALVAC-

HIV, and it employed a pox virus as a delivery system for DNA expressing HIV *gag*, *pol*, and *env* genes from HIV samples isolated during the 1990s [Ng'uni et al., 2020]. Utilized in a clinical trial (RV144) in Thailand from 2003 – 2009, it is the only time in which an HIV vaccine produced some protection against HIV, albeit only at 31% efficacy [Rerks-Ngarm et al., 2009]. Since then, there has been numerous vaccine trials using either ALVAC-HIV in combination with AIDSVAX or ALVAC-HIV alone, and all have resulted in failures [Ng'uni et al., 2020].

The literature on HIV has produced over 20 years of immunological research toward a vaccine against HIV, yet a prevailing question remains: why haven't any of the HIV vaccine trials been successful? There are several proposed reasons. Kijak et al. [2013] calculated phylogenetic distances between the HIV genes used in the RV144 vaccine trial to HIV genes in strains circulating Thailand. They found that the genes from HIV strains circulating Thailand at the time the vaccine was tested (2003 – 2009) had diverged considerably from HIV strains at the time the vaccine was developed (1990s) [Kijak et al., 2013]. A consequence of this genetic divergence is weak antibody binding and inconsistent switching between IgG subclasses, partly explaining the vaccine's low efficacy [Karnasuta et al., 2017].

From the standpoint of antibody production, Heß et al. [2019] showed that N-glycosylation of asparagine residues on gp120 blocks IgG2a antibody recognition in favor of weaker effector responses from the IgG1 subclass. This very well could be another possible mechanism as chronic strains of HIV are heavily glycosylated. Another possible explanation is significant reduction in ADCC [Huang et al., 2016a]. Compared to HIV-infected individuals, antibodies isolated from vaccinated subjects in the VAX004 trial exhibited a 50% relative reduction in ADCC [Huang et al., 2016a]. Gach et al. [2017] studied antibody-dependent

phagocytosis (ADP), an effector function of macrophages in which they are recruited by antibodies to "eat and digest" infected cells or viruses. Antibodies normally enhance phagocytosis of opsonized virus and cells, but in the case of HIV there was no enhancement [Gach et al., 2017]. This was most likely due to the scarcity of HIV spikes on the surface of the viral envelope [Zhu et al., 2008], impairing the ability of antibodies to cluster for Fc gamma receptor (FcγR) binding on macrophages [Gach et al., 2017]. In fact, chronic HIV strains tend to carry half the concentration of spikes on the surface compared to their earlier transmitted variants [Parrish et al., 2013]. This is likely an adaptation to evade ADP. Non-enhancement of ADP might be another likely explanation for vaccine failures.

While antibodies are produced to fight infections, in the context of certain viruses they could have the opposite effect, especially if they are incapable of neutralizing the reactive pathogen. Low-titer antibody responses, whether neutralizing or non-neutralizing, propagate a selective pressure favoring escape mutations [Bar et al., 2012; Wu et al., 2012]. Moscoso et al. [2014] and Andrews et al. [2018] reported 27 amino acid sites variable loops of HIV gp120 that undergo mutations during immune selection; conserved regions of gp120 are enclosed in the interior of the gp120 trimer, making it inaccessible to antibodies. In the context of non-neutralizing antibodies especially, they may enhance infectivity, a phenomenon known as antibody-dependent enhancement (ADE) [Robinson et al., 1987; Takeda et al., 1988; Takeda and Ennis, 1990; Willey et al., 2011; Forthal et al., 2012; Gorlani and Forthal, 2013]. This occurs, usually, by infecting macrophages through antibody bridges between FcγR and gp120, enabling HIV to fuse with the endosomal membrane and enter the cytosol [Trischmann et al., 1995]. Given that *trans*-infection is more efficient than cell-free direct infection of $CD4^+$ T-

lymphocytes, enhancement of viral infectivity would be mediated through immunological synapses and vesicle signaling between T-lymphocytes and infected macrophages. In fact, results from the VaxSyn indicated that ADE occurred in some vaccinated subjects [Dolin et al., 1991]. Immune complexes, a method often employed to enhance vaccine efficacy, between SOSIP.664 and monoclonal antibodies reduced the breadth of the neutralizing antibodies generated in animal models [Gach et al., 2019]. While this does not necessarily rule out the SOSIP system as a promising strategy, it does limit how such a vaccine may be formulated to enhance the generation of broadly neutralizing antibodies. Depending on the epitope used for the vaccine, ADCC could also enhance viral infectivity via mother-to-child transmission through breastfeeding and lead to higher infant mortality [Naiman et al., 2019].

In addition to ADE, a complement cascade, an immunological function that enhance antibody-mediated effector functions, is also capable of enhancing infectivity through a process known as complement-dependent enhancement (CDE) (reviewed in detail by Yu et al. [2010]). CDE is a well understood feature of immune interactions with HIV, as it has been documented both *in vitro* experiments and from the VaxSyn trial [Dolin et al., 1991; Thieblemont et al., 1993b; Bouhlal et al., 2002; Bouhlal et al., 2007; Nijmeijer et al., 2021].

The mechanisms listed and briefly described above are well-documented rationales for the apparent failures of HIV vaccine trials. Continued research on epitopes and antibody-mediated effector functions would be invaluable toward filling in knowledge gaps and producing more effective vaccines in the future. However, there is one underappreciated avenue with regard to HIV transmission that was not considered until 2008 and 2009, and it is now a vaguely understood phenomenon. The literature regarding this aspect of HIV transmission has produced

important information, unexplained contradictions, and seemingly many more questions than before it was first studied [Monaco et al., 2017]. This is the phenomenon of the genetic bottleneck of HIV transmission, and it may be the single most important area of virology pertinent to vaccine development.

The Genetic Bottleneck and Transmitted/Founder (T/F) Variants

In the vast majority of sexual or maternal-to-child transmissions of HIV-1, there is a tight genetic bottleneck that allows only one of the many genetic variants of virus in the infected, transmitting partner or mother to establish infection in the recipient partner or infant. Eighty percent of newly infected patients are infected by a single HIV virus (termed the T/F strain) from a donor while the remaining 20% infected by two to five viruses from the donor [Keele et al., 2008; Salazar-Gonzalez et al., 2008; Abrahams et al., 2009; Fischer et al., 2010] (reviewed by Monaco et al. [2017]). The same phenomenon was also observed in an animal model using SIV [Keele et al., 2009; Stone et al., 2010; Ma et al., 2011; Yuan et al., 2017; Chen et al., 2018]. Infection by a single virus (or multiple identical clones of a single virus) contributes to genetic homogeneity of viral variants that established productive infection early during the acute phase [Edo-Matas et al., 2010; Piantadosi et al., 2019], which gradually diversifies overtime into genetically divergent chronic quasispecies during the process of immune selection [Edo-Matas et al., 2010] (reviewed by Joseph et al. [2015]). This intriguing feature of HIV infection is of critical importance in HIV prevention because it is precisely this transmitted/founder (T/F) strain that must be the target of a vaccine-elicited immune response (reviewed by Monaco et al. [2017]). The T/F strain is rarely the majority strain in the transmitting host, thus implying that

specific characteristics of T/F strains, other than stochastic probability, determine its transmissibility [Keele et al., 2008; Salazar-Gonzalez et al., 2008; Carlson et al., 2014]. This implies that during HIV transmission there is selective pressure favoring (or disfavoring) certain viral phenotypes over others, and that this "transmission selectivity" may be propagated by immunological and physiological factors in interaction with the T/F virus. In fact, Abrahams et al. [2009] showed that even in cases of HIV-1 infection, whether from clades B or C, that resulted in the transmission of multiple T/F variants, the distribution of viral genotypes of transmitted T/Fs in each individual was non-poisson, implying a non-random process of selection. Several phenotypic characteristics of T/F strains that may be important for transmission selectivity in the genetic bottleneck have been described [Parrish et al., 2013] (these chracteristics are also reviewed by Monaco et al. [2017]).

Since 1996, one of the earliest phenotypes known to be associated with HIV variants during acute HIV infection was CCR5-tropism, a characteristic of the majority of T/F strains that has since been corroborated by more recent studies [Deng et al., 1996; Dragic et al., 1996; Wu et al., 1996; Keele et al., 2008; Salazar-Gonzalez et al., 2009; Mild et al., 2010; Parker et al., 2013; Parrish et al., 2013; Monaco et al., 2017]. Further investigations of T/F strains revealed other phenotypic properties. They typically have a higher replicative capacity compared to their chronic counterparts [Ochsenbauer et al., 2012; Sugrue et al., 2022; Madlala et al., 2023], likely due to a loss of fitness in mutations accumulated during disease progression [Zanini et al., 2015; Ashokkumar et al., 2020; Sugrue et al., 2022]. T/F variants also preferentially infect $CD4^+$ T-lymphocytes rather than macrophages. T/F variants from the HIV-1 B lineage contain fewer N-glycosylation sites and shorter variable loops [Derdeyn et al., 2004; Chohan et al., 2005; Frost et

al., 2005; Liu et al., 2008; Gnanakaran et al., 2011]. These two specific characteristics modulate different functions of the gp120 spike on T/F viruses in clade B. Lower N-glycosylation of the variable loops provides higher transcytosis function [Shen et al., 2014], higher affinity for cell-free infection of $CD4^+$ T-lymphocytes [Ochsenbauer et al., 2012], higher binding affinity to $\alpha4\beta7$ integrins on $CD4^+$ T-lymphocytes [Arthos et al., 2008; Cicala et al., 2009], greater vulnerability to antibody binding [Derdeyn et al., 2004; Frost et al., 2005; Wilen et al., 2011; Moscoso et al., 2014], and lower adherence to monocytes like dendritic cells and macrophages [Keele et al., 2008; Salazar-Gonzalez et al., 2009; Ochsenbauer et al., 2012; Parrish et al., 2013; Kafando et al., 2019; Lamers et al., 2019]. During the course of disease progression, these clade B T/F-specific characteristics undergo intrapatient evolution through the process of immune selection, in which the emergence neutralizing antibodies select for longer variable loops [Moscoso et al., 2014] and more N-glycosylation sites to block epitope access [Frost et al., 2005; Keele et al., 2008; Salazar-Gonzalez et al., 2008; Curlin et al., 2010; Fischer et al., 2010; Gnanakaran et al., 2011; Gonzalez et al., 2015; Zanini et al., 2015; Park et al., 2016; Palmer and Poon, 2019]. Although many of these changes (such as variable loop lengthening) may reduce the replication capacity (RC) and infectivity of the evolving strains, it could be that the selection pressure for immune evasion is strong enough that there is still a net benefit to the virus's fitness even with a functionally compromised gp120 spike or replication-promoting factors [Ganusov et al., 2011; Gnanakaran et al., 2011; Ochsenbauer et al., 2012; Hamoudi et al., 2013; Gonzalez et al., 2015; Zanini et al., 2015; Lamers et al., 2019]. Co-receptor switching (i.e. the emergence of CXCR4-utilizing strains from an initially homogenous population of CCR5-utilizing T/F population) also occurs during immune selection [Edo-Matas et al., 2010], which implies HIV adapts to infect

more cell types in order to maintain its viability during the chronic phase of infection. Selection pressure favoring such mutations enable the virus to evade immune responses and is a probable factor in its persistence during disease progression.

Characteristics of T/F viruses from the HIV-1 C lineage, on the other hand, carry vastly different phenotypes, conferring their own unique functions to the virus. These viruses tend to have a greater number of N-glycosylation sites on gp120 than those from the B lineage, and as a result C-lineage T/F viruses exhibit different virus-host interactions during acute infection compared to B-lineage viruses [Derdeyn et al., 2004; Salazar et al., 2009; Parrish et al., 2013; Monaco et al., 2017; Lumngwena et al., 2019]. C-lineage T/F viruses are more sensitive to interferon-stimulated immune responses than those from the B lineage [Deymier et al., 2015]. The higher glycosylation of gp120 variable loops on these viruses confers resistance to antibody-binding, and immune selection disfavors deletion of N-glycosylation sites and strengthens CCR5-utilization during chronic infection [Parker et al., 2013; Heß et al., 2019; Lumngwena et al., 2019]. One study by Li et al. [2006] showed that clade C HIV-1 T/Fs were still sensitive to neutralizing antibodies, although clade C viruses were also significantly more resistant to neutralization than were clade B viruses in general. Moreover, N-glycosylation sites enhance binding to DC-SIGN [Geijtenbeek et al., 2000; Hong et al., 2002; Kwon et al., 2002; Saïdi et al., 2007; Shan et al., 2007; Shen et al., 2010a; Lumngwena et al., 2020], Siglec-1 [Zou et al., 2011; Sewald et al., 2015; Perez-Zsolt et al., 2019], and -7 [Varchetta et al., 2013] carbohydrate receptors on the surface of DCs, opening a significant opportunity for C-lineage T/F viruses to infect a host through cell-mediated migration across epithelial barriers (rather than transcytosis) and *trans*-infection [Geijtenbeek et al., 2000;; Kwon et al., 2002; Shen et al., 2010a; Shen et al.,

2010c;; Shen et al., 2014; Sewald et al., 2015; Trifonova et al., 2018; ; Perez-Zsolt et al., 2019; Lumngwena et al., 2020]. Unlike B-lineage T/F viruses, the greater N-glycosylation of C-lineage T/F viruses impedes α4β7 integrin binding with $CD4^+$ T-lymphocytes [Parrish et al., 2012]. Clade C T/Fs, like clade B T/Fs, exhibit higher RC in infected cells compared to chronic strains [Madlala et al., 2023], suggesting that genetic characteristics of T/Fs from either clade may be responsible for worsening disease outcomes. Current knowledge on T/F viruses from both lineages and their known characteristics are reviewed in detail by Monaco et al. [2017].

As a consequence of functional phenotypic differences between B- and C-lineage T/F viruses, there is not a known universal characteristic defining all T/F viruses to date other than CCR5-tropism in most T/F variants [Keele, 2010]. As consequence, known characteristics provide little accuracy in predicting which strain out of a swarm of viruses in a transmitting host would become the T/F strain in the recipient host. In fact, Frange et al. [2013] showed that inferring a T/F variant *a priori* from a donor is is not likely given the multitude of variants in the plasma and infected peripheral blood mononuclear cells (PBMCs) of transmitting donors. Results from other reports revealed completely different T/F variants infecting *Rhesus macaques* in an animal-specific manner even from a single inoculum stock of SIVmac251, also implying difficulty in predicting selection of T/F variants in animals [Keele et al., 2009; Ma et al., 2011; Yuan et al., 2017]. Interestingly, the T/F virus that ultimately establishes productive infection in a recipient may not be the only T/F virus that initially gets transmitted, as has been reported previously [Frange et al., 2013]. According to data from animal models, cohort studies, and meta-analyses, the transmission bottleneck very early during acute infection (perhaps within a few hours to a few days) results in a transient population of multiple viral strains exhibiting little

genetic diversity [Abrahams et al., 2009; Li et al., 2010; Gnanakaran et al., 2011; Chen et al., 2018; Piantodosi et al., 2019; Macharia et al., 2020; Baxter et al., 2023]. While the multiplicity of transmitted variants suggests that the genetic bottleneck is not as restrictive as shown by Keele et al. [2008] and Salazar-Gonzalez et al. [2008], the initial viral population reduces over a short time to a single variant in most cases [Piantadosi et al., 2019]. However, the issue of multiplicity here confounds efforts to infer T/F variants further and to ascertain whether the genetic bottleneck is strictly a selective process or if there may be a stochastic element before or after transmission.

The following sections cover what is currently known (and also what is not known) about biological features in the host or donor that may influence the selective transmission of T/F viruses over others. Selection pressures influencing T/F transmission in each biological feature are also reviewed by Nijmeijer and Geijtenbeek [2019].

Epithelial and Mucosal Barriers

HIV transmission during sexual contact requires passage or transport of virions across the epithelial layer to the submucosa of the recipient where the virus can infect target T-lymphocytes. Thus, epithelial cells of the penis, cervix, vagina, rectum, and GI tract represent a physical barrier to transmission to a new host and may favor the passage of select T/F strains over non-T/F strains from an HIV swarm. In fact, rectal [Keele et al., 2009] and vaginal [Stone et al., 2010] inoculation of *Rhesus macaques* with either an SIVmac251 or SIVsmE660 swarm permits the transmission of significantly fewer T/F strains than intravenous inoculation. Indeed, in one study that examined the prevalence of HIV viral variants in the female genital tract and blood

plasma of 3 acutely infected women, each single-nucleotide viral variant made up no larger than 6% of the total population of viral genomes in both compartments at 4 days after initial detection of HIV [Piantadosi et al., 2019]. Most of these variants were no longer present in samples on later days, suggesting that the genetic bottleneck was established very early during infection and that numbers of variants continued to be reduced during the progress of acute infection, as stated previously.

From a more mechanistic standpoint, the glycosylation of asparagine residues in gp120 of HIV contributes to enhancement of transmission through epithelial cells. In a study by Shen et al. [2014], viral clones with a greater number of mannose oligomers and fewer complex glycans were more efficiently transcytosed through epithelial cells both *in vitro* using a transcytosis assay and *ex vivo* using rectal mucosal explants. Notably, N-glycosylation is a determining characteristic of not only C-lineage T/F viruses, but also of those in A- and D-lineages (reviewed by Monaco, Ende and Hunter, 2017). Based on these results and others showing that sexual transmission of HIV results in a single T/F strain in 80% of new recipient hosts [Keele et al., 2008; Salazar-Gonzalez et al., 2008; Fischer et al., 2010], there is reason to consider the epithelial barrier as a physiologically important factor in a bottleneck that significantly reduces the number of T/F strains in the recipient after sexual exposure.

The migration of virions across the epithelial barrier likely doesn't just depend on interaction between the virus and the epithelium. One report found evidence that $CD11c^{+}$ epidermal DCs and LCs interact with HIV and may enhance transmission across the epithelium while CD33-deficient dendritic cells do not [Bertram et al., 2019]. A similar result was found in a 2018 study using human cervical explants where myeloid cells are principally the first cell types

to become infected with HIV and transfer them to susceptible T-lymphocytes [Trifonova et al., 2018]. These data could explain to some extent the bias of genital tissue compartmentalization toward M-tropic variants during acute infection [Sutthent et al., 2001]. This is contradicted by an earlier study, however, that showed Th17 cells were the initial cell types to become infected by SIV in monkeys and thus may be responsible for viral translocation to the mucosa [Stieh et al., 2016]. As stated previously, this finding may not be as applicable to our understanding of the HIV-1 genetic bottleneck due to the study's shortcomings.

It is important to note here that HIV may transmit more easily through sites of local inflammation or ulcers from a concurrent sexually transmitted infection. The environment in which HIV transmits across the epithelial barrier may artificially affect the transmission selectivity of T/F strains under these special circumstances. Indeed, reports have shown that genital inflammation increases the susceptibility of the recipient due to a compromised epithelial structure and the recruitment of target T-lymphocytes (reviewed by Passmore et al., 2016 and Kaul et al., 2018). Sexual intercourse often causes microabrasions in the vaginal and cervical epithelia that prompt a minor inflammatory response in order to repair the tissue and protect the site from pathogens [Fraser et al., 1999]. This would likely recruit $CD4^+$ T-lymphocytes to the site of the abrasion or ulceration, more easily exposing susceptible cells to HIV infection, especially if micro-abraded tissue were infected with other pathogens [Yamada et al., 2021]. Such was found to be the case in a study where Th17 cells were predominantly the initial cell type infected by an SIV challenge in female monkeys with pre-existing genital inflammation [Stieh et al., 2016].

In addition to cellular and immunological features of the epithelium and mucosa that may contribute to T/F selection, there is evidence of tissue compartmentalization of T/F variants in both animal and clinical studies. One of the earliest reported indications of phenotype-dependent tissue compartmentalization was the observation that macrophage-tropic, non-syncytium-inducing HIV-1 quasispecies were associated with genital fluids of both male and female subjects, despite the existence of syncytium-inducing strains permeating the plasma during acute infection [Sutthent et al., 2001]. Although the authors at the time may not have surmised this difference in tissue-associated phenotypes was a result of T/F selection, compartmentalization may be an underappreciated feature of the genetic bottleneck that has only recently garnered the attention of virologists. More recently, Lamers et al. [2019] autopsied deceased HIV-infected patients with either detectable or undetectable plasma viral loads and found that tissue-compartmentalized HIV in both groups exhibited discordant *env* phenotypes, which implies that tissue-associated viral phenotypes correlate with disease progression an that it is not a phenomenon restricted to early acute infection. Chen et al. [2018] found that certain tissues in the GI tract of animals were more susceptible to SIV transmission than others. Differences in transmission susceptibility not only resulted in differences in viral loads of each challenged tissue but also in numbers and differences of *env* phenotypes of the tissue-resident T/F variants. Klein et al. [2021] reported that the transmitted HIV variants from the donor are retained by the genital mucosae even throughout chronic infection, but that a single variant capable of breaking out of its mucosal compartment would establish productive infection early during the acute phase. Differences in phenotypes of the "breakout" virus could be a determinant of transmission selectivity and may inform vaccine development [Klein et al., 2021]. Similar results were found

by Tortorec et al. [2008] animals intravenously challenged with SIVmac251. Neurovirulence of SIVmac239 (an SIV swarm isolated from an infected animal) is a function of particular phenotypes [Anderson et al., 1993], and compartmentalization of select viral phenotypes in the central nervous system appears to be tissue-specific [Chen et al., 2006]. Tissue-specific compartmentalization of particular variants was also observed in the male genital tract of infected animals [Tortorec et al., 2008], suggesting that variant-selected tissue compartmentalization is not limited to the central nervous system. Such compartmentalization may be important to transmission selectivity as well as persistence of latent virus in infected cells, and knowledge regarding tissue-resident viruses during both acute and chronic stages of infection as well as information on the mechanisms of selection governing them are currently limited.

Trans-infection by DCs

After HIV passes through the epithelial barrier into the mucosa and submucosa, virus must disseminate throughout the body in order to productively infect the new host. A feature of this dissemination is the adherence of virions to DCs, which may disseminate the virus to distal lymphoid tissue and selectively transfer T/F strains to susceptible $CD4^+$ T-lymphocytes. To date, there is a growing body of evidence suggesting that cell-to-cell interactions play a prominent role in enhancing HIV infectivity. Regardless of any role that they may play in transmission selectivity or the genetic bottleneck, dendritic cells have been shown to not only draw virions from the epithelial barrier but can also enhance trans-infection of susceptible $CD4^+$ T-lymphocytes [Bertram et al., 2019; Trifonova et al., 2018; Perez-Zsolt et al., 2019] (reviewed by

Bracq et al. [2018] and Dutartre et al. [2016]). In fact, a recent report found evidence that trans-infection (transfer of virus from a myeloid carrier to a susceptible T cell) can still occur even during treatment with antiretroviral drugs, purported to be highly effective in the prevention of AIDS progression and transmission to new hosts [Rappocciolo et al., 2019]. Thus the need to investigate the role of *trans*-infection in HIV transmission and its potential role in the genetic bottleneck cannot be overstated.

One mechanism by which $CD4^+$ T-lymphocytes can be *trans*-infected is through the immunological synapse (reviewed by Bracq et al., 2018 and Dutartre et al., 2016). The immunological synapse is formed when a cell, such as a dendritic cell, physically interfaces with a lymphocyte (i.e. T cell) in order to present antigens to and stimulate the T cell. Under normal conditions, this is an important immune function that instructs T-lymphocytes and B cells to recruit other immune cells and generate antibodies against pathogens, respectively. There are likely other mechanisms by which *trans*-infection can occur (reviewed by Bracq et al. [2018]. However, *trans*-infection through contact with surface-adhered virus on dendritic cells and through exchange of vesicle-enclosed virus in the synapse are the most likely routes of infection by cell-to-cell contact in a recipient host.

Not only is *trans*-infection capable of enhancing infection of $CD4^+$ T-lymphocytes, the interaction between target T-lymphocytes and DCs may mediate selective transfer of particular genetic variants of HIV from a viral swarm. To date, there is published data suggesting that such a mode of selection exists, even though it remains unclear whether it contributes to a bottleneck. In a report by Hong et al. [2002], gp120 binding to DC-SIGN (a cell surface lectin expressed on DCs) was shown to be mannosylation-dependent; more specifically, the glycosylation of specific

asparagine residues on gp120 promotes this interaction with DCs. Although the authors did not determine whether these phenotypes were associated with a bottleneck, let alone transfer to T-lymphocytes, it implies that glycosylation of asparagine residues enhances adherence to DC-SIGN on primary DCs and could thus possibly play a role in *trans*-infection selectivity to susceptible cells. Indeed, another group has demonstrated evidence of N-glycosylation-mediated *trans*-infection of $CD4^+$ T-lymphocytes by monocyte-derived DCs [Shen et al., 2014]. A higher degree of mannose oligomers and less complex glycans on the viral envelope of HIV-1 contributed to stronger adherence to DCs and, consequently, higher *trans*-infectivity to susceptible T-lymphocytes. This finding is consistent with a more recent publication showing that a higher degree of mannosylation on the gp120 protein of HIV-1 promotes antigen routing outside of degradative endosomes to invaginated pockets just under the cell membranes of DCs [Jarvis et al., 2019]; in theory, this poises such viruses to infect T-lymphocytes through the immunological synapse more effectively. Nabatov et al. [2006] reported that the number of N-glycosylated residues, length of V1/V2 loop, and overall charge of V3 residues in HIV gp120 are factors involved in enhancing binding affinity to DCs. Enhancement of trans-infection by DCs was also observed with viral variants containing either a higher positive charge on V3 and greater V1/V2 length or lower positive charge on V3 and shorter V1/V2 length. It can be inferred from these studies that DC-mediated trans-infection of T-lymphocytes depends on a combination of viral phenotypes, and some T/F strains with these phenotypes probably enhance both adherence to DCs and trans-infection from DCs to T-lymphocytes. Such was found to be the case by Parrish et al. [2013]. However, whether these phenotypes are involved in a genetic bottleneck is not fully known.

Some of these phenotypes (strains with a higher degree of N-glycosylation on gp120) were found to be consistent with some of those observed in T/F strains [Nabatov et al., 2006], which is indicative of a role in the genetic bottleneck. Interestingly, viruses with this phenotype were poor at direct infection of T-lymphocytes [Shen et al., 2014], which may suggest that T/F selection during acute infection depends on the prevalence of cell types in the mucosa at the time of infection. For example, an abundance of DCs at the site of infection could favor dissemination by T/F strains with more glycosylated residues in the viral envelope while an abundance of Th17 cells at the site of infection could favor infection by fewer glycosylated strains. The determining mechanism of selection here and how it may contribute to a restrictive bottleneck are unknown but may depend on different physiological and immunological conditions that shape this selectivity. Characterizing how those conditions result in T/F strain selection is the goal of the second hypothesis of this proposal.

Antibodies and Complement

During HIV transmission, it is likely that nearly all of the virions in the swarm from the donor are opsonized by complement and antibody. Coating of the virus may influence transmission of select T/F strains across the mucosa and/or in T cell infection, either directly or in trans by DCs. Indeed, studies have shown that complement and antibody both play separate although similar roles in enhancing HIV infectivity at the level of (1) transcytosis across the epithelial barrier and (2) trans-infection. Virus-host interactions through complement and antibodies and how they influence viral infectivity and neutralization are reviewed by Yu et al. [2010], Gorlani and Forthal [2013], and Posch et al. [2020].

Currently, it is unclear whether antibody or complement may be involved in the selective transcytosis of particular strains of HIV across the epithelial/mucosal barrier. But there is a body of evidence suggesting their roles in enhancement (or inhibition) of viral transcytosis and that interactions between HIV and these immunologically important molecules may result in some form of transcytosis selectivity. In the case of antibodies, neutralizing antibodies have been shown to impair transcytosis of both HIV T/F and chronic variants across the epithelial barrier to a similar degree [Gonzalez and Sagar, 2016]. In another study by Shen et al. [2010b], gp41-specific antibodies also impaired transcytosis, not only in a model epithelium but in rectal mucosal explants as well. The findings here suggest that antibody does not favor the transmission of T/F strains over the majority of chronic strains since pooled plasma from chronically infected individuals [Gonzalez and Sagar, 2016] and specific antibodies [Shen et al., 2010] inhibit transcytosis. However, in the absence of phylogenetic analysis of viral variants that successfully transcytosed across the barrier, it is unclear whether antibody selects for individual strains in addition to restricting the numbers of viruses that successfully penetrate the mucosa.

In the case of complement, however, there is an indication of a more complicated role in modulating transcytosis. HIV infection often triggers a complement activation cascade that leads to the opsonization of virus in either an antibody-dependent or antibody-independent fashion. In most cases, HIV virions from a chronically infected donor are bound by activated complement components [Saïdi et al., 2012]; while opsonization by complement is intended to reduce viability of the virus in the chronically infected host, it can paradoxically enhance transmission to a new recipient (reviewed by Yu et al. [2010]). Some of the earliest re ports indicated enhanced direct infection of $CD4^+$ T-lymphocytes by virus through a complement mediated

mechanism [June et al., 1991], and that it does so without engagement of CD4 on target cells [Boyer et al., 1991]. Other studies around the same time even showed that HIV infection of monocytes was enhanced through a complement-mediated mechanism [Thieblemont et al., 1993a], with implications for *trans*-infection. In a recent study, opsonization of HIV by activated complement nullified plant lectin-mediated inhibition of transcytosis through a monolayer of HEC-1A cells (an endometrial cell line) [Jenabian et al., 2010]. Mannose-binding plant lectins expressed in members of the *Amaryllidaceae* family bind to glycosylated residues in the HIV envelope protein and disrupt viral adherence to T cell CD4 and CCR5 receptors [Balzarini et al., 2004]. This finding is troubling in that it suggests complement impairs the effect of certain anti-HIV microbials [Saïdi et al., 2007] in preventing new HIV infections. More importantly, it could be indicative of a molecular interaction between virus and complement that enhances transcytosis under certain conditions and may enhance direct or trans-infection under others.

In an earlier study, HIV transcytosis through colorectal epithelial cells could theoretically be enhanced by activated complement [Bouhlal et al., 2002]. Indeed, CXCR4- and CCR5-tropic viruses opsonized by semen complement enhance infection of HT-29 cells in a model epithelium in a CD18/CD11b-dependent manner (which are heterodimeric receptors for the complement C3b) [Bouhlal et al., 2002]. Whether HIV is capable of infecting epithelial cells remains controversial, as there are contradictory reports on the susceptibility of epithelial cells to infection [Chenine et al., 1998; Collins et al., 2000]. Furthermore Bouhlal et al. [2002] utilized a lab strain of HIV-1, abnormally more infectious than clinically isolated HIV-1, and a colorectal adenocarcinoma cell line that may be different from normal human colorectal epithelia. On these grounds, there would be reason to doubt whether complement would facilitate infection of

"normal" epithelial cells by "normal" HIV-1, let alone transcytosis through the mucosa. However, complement receptor 3 (CR3, formed by heterodimerization of CD11b and CD18) are expressed on the surface of not only normal human rectal epithelium but also myeloid cells that may interdigitate between epithelial cells [Hussain et al., 1995]. As a matter of fact, complement does enhance transcytosis and infection of DCs through this interaction [Bouhlal et al., 2007]. In this case, it is still plausible that complement promotes HIV passage either between or through epithelial cells to the mucosa by interaction with CD11b and CD18 on either epithelial cells or dendritic cells. Infection of epithelial cells may still occur in this fashion, albeit transiently in the early stages of acute infection. The interaction between semen complement and HIV in the enhancement of transcytosis alone may play a crucial role modulating the selection of T/F strains during transmission, though this is as of yet not known. However, it remains unclear whether complement binding, in the presence or absence of antibody, affects transmission selectivity across the epithelial barrier to the mucosa.

The notion that semen components may enhance HIV-1 transmission from donor to recipient is muddled by other reports [Balandya et al., 2010]. While it is certainly the case that seminal complement [Bouhlal et al., 2007; Nijmeijer et al., 2021] seminal heparan sulfate [Ceballos et al., 2009] may enhance transcytosis, direct infection of DCs/LCs, and *trans*-infection of $CD4^+$ T-lymphocytes, mucin-6 [Stax et al., 2009] and other positively-charged proteins [Martellini et al., 2009] in human semen counteract this effect. It is interesting to note that components of the seminal plasma block CXCR4-tropic viruses but not CCR5-tropic viruses from penetrating the mucosa [Balandya et al., 2010], suggesting that seminal components is probable factor in modulating transmission selectivity of T/Fs.

Some reports indicate that complement and antibodies from the donor host may modulate infectivity of T/F strains, either negatively or positively (reviewed by Gorlani and Forthal [2013]). ADE usually occurs, but not exclusively, when an antibody bound to a pathogen interacts with FcγRs on cells such as macrophages or dendritic cells, providing an alternate means of entering the cell (reviewed by Gorlani and Forthal [2013] and Takeda and Ennis [1990]). In the example of HIV, an antibody directed against the CCR5-binding region of gp120 could act as a bridge between the virion's infection machinery and FcγR, an unintended consequence of particular antibodies. While cells of the myeloid lineage are not particularly the target of HIV during acute infection [Ochsenbauer et al., 2012], antibodies could drive HIV infection of these cell types [Trischmann et al., 1995]. Despite the published literature on ADE, it has yet to be studied whether antibodies are capable of selecting particular HIV variants from a viral swarm.

In addition to ADE, there is also complement-mediated enhancement of viral infectivity, which may or may not occur in an antibody-dependent manner (reviewed by Gorlani and Forthal [2013] and Thieblemont et al. [1993b]). Opsonization of HIV-1 virions with a non-neutralizing antibody complexed with activated complement significantly enhanced infection of T-lymphocytes in a complement receptor 2 (CR2)-dependent manner [Willey et al., 2011] and enhanced adherence to and infection of dendritic cells in a CR3-dependent manner [Jenabian et al., 2010]. More importantly, virus opsonized by semen complement alone enhanced infection of dendritic cells as well as trans-infection of $CD4^+$ T-lymphocytes by DCs [Bouhlal et al., 2007], which may be an important factor in complement-mediated transmission selectivity through sexual exposure. These results are particularly important in HIV transmission because

autologous non-neutralizing antibodies are often expressed very early during acute infection and transmitted virions from a chronically infected donor to a recipient are likely bound by complement and antibodies. While there is evidence that complement from the donor may enhance viral infectivity, it is unclear whether complement opsonization, with or without antibody, also modulates infection selectivity in direct infection or trans-infection. Studying how opsonization by complement and/or antibody would contribute to the genetic bottleneck or transmission selectivity, at least, from the standpoint of DC-mediated *trans*-infection, would greatly improve efforts toward vaccine design.

Interferons

Antibody and complement are not likely the only immunological factors that modulate infection selectivity of $CD4^+$ T-lymphocytes. Antiviral cytokines, like interferon-gamma (Ifn-γ) and interferon-alpha (Ifn-α) are secreted by myeloid cells in order to stimulate $CD4^+$ T-lymphocytes to a proinflammatory state (reviewed by Schneider et al. [2014], Schoggins [2019], and Walter [2020]). HIV infection in target cells potently induce interferon production in myeloid cells like DCs and macrophages [Fong et al., 2002; Yonezawa et al., 2003; Schmidt et al., 2005; Pierini et al., 2021]. One prominent feature of this activation is that T-lymphocytes adopt an antiviral defense mechanism that allows it to resist HIV-1 infection (reviewed by Utay and Douek [2016]). The interplay between host interferons and infectious HIV is of critical importance as HIV not only expresses enzymes that block interferon's effector functions but T/F virus selection is also influenced by them. The current literature on this topic, however, lacks

sufficient information to develop an unambiguous theory explaining such interplay as well as exploiting interferons to develop new generation prophylactic drugs and vaccines.

Interferons are divided into three familes: Type 1 Ifns, Type 2 Ifns, and Type 3 Ifns. The Type 1 Ifns comprise of a subfamily of interferon-alphas (Ifnα), interferon-beta (Ifnβ), interferon-epsilon (Ifnε, interferon-kappa (Ifnκ), and interferon-omega (Ifnω). There is only one Type 2 Ifn: interferon-gamma (Ifnγ). And interferons under the Type 3 family are interferon-lambda (Ifnλ), interleukin-10 (IL10), interleukin-22 (IL22), and interleukin-26 (IL26) [Walter, 2020]. These interferons are typically secreted by myeloid cells, like plasmacytoid DCs, in order to activate anti-viral programs in infected cells or potentially infected cells, as well as other immune cells to enhance proinflammatory responses [Yonezawa et al., 2003; Schmidt et al., 2005; Li et al., 2009; Schwartz et al., 2018; Yun et al., 2021].

Type 1, Type 2, and Type 3 interferons elicit their functions by binding to three different receptors: IFNAR, IFNGR, and IFNLR, respectively [Schoggins, 2019]. This binding activates the transcription of interferon-stimulated genes (ISGs) through the JAK/STAT signaling pathway, ultimately to express proteins that inhibit viral proteins and enzymes in infected cells and amplify cytokine production in nearby cells [Schoggins, 2019]. The ISGs activated by interferons and interleukins work in concert to establish a proinflammatory condition to render affected tissues hostile to viral replication and infection.

The ISGs known to target HIV-specific processes are IFITMs (inhibits HIV fusion mechanism), SERINC5 (alters target cell and viral membranes to resist viral endocytosis), TRIM5α (inhibits capsid disassembly), APOBEC-3G (induces G-to-A hypermutations in HIV cDNA during reverse transcription), SAMHD1 (hydrolyzes dNTPs in order to prevent their

incorporation into viral cDNA during reverse transcription) MX2 (blocks nuclear import of the HIV genome), GBP5 (abrogates post-translational modification of HIV gp160), and tetherin (blocks HIV virion budding from infected cells) [Stremalu et al., 2004; Neil et al., 2008; Wang et al., 2008; Sato et al., 2010; Goldstone et al., 2011; Lu et al., 2011; Kane et al., 2013; Krapp et al., 2016; Wang et al., 2017; Ikeda et al., 2019; Pierini et al., 2021]. HIV, on the other hand, expresses proteins that antagonizes interferon-stimulated restriction factors. HIV Vpu blocks GBP5 from impairing gp160 post-translational modification and tetherin from impairing viral exocytosis [Neil et al., 2008; Krapp et al., 2016]. HIV Nef enhances infectivity of SERINC5-altered virions and sequesters virions from SERINC5 [Trautz et al., 2016] HIV Vif promotes APOBEC3G degradation through ubiquitination and arrests the cell cycle at G_2 in order to shield reverse transcription and enhance proviral gene expression, respectively [Sato et al., 2010; Evans et al., 2018]. HIV-1-specific Vpr and HIV-2-specific Vpx (both paralogs of one another) arrests the cell cycle to maximize proviral transcription and degrade SAMHD1 also through ubiquitination [Planelles and Barker, 2010; Goldstone et al., 2011; Laguette et al., 2011]. The perpetual action between viral regulatory proteins and host restriction factors contributes to a runaway effect or "arms race" that could result in potentially fatal cytokine storms among infected patients and lead to worsening disease outcomes, as shown in human cohort and animal studies [Roberts et al., 2010; Huang et al., 2016b; Keating et al., 2016; Muema et al., 2020]. There are other restriction factors activated by Ifns that attenuate HIV infectivity, and there may be additional functions of HIV regulatory proteins that protect it from those restriction factors (reviewed by Coiras et al. [2010] and Soper et al. [2018]). Our understanding of those biological functions, however, are currently limited.

With regard to the genetic bottleneck of HIV, interferon-resistance, as well as resistance to myeloid-cell-mediated restriction, is a phenotypic feature unique to T/F viruses [Fenton-May et al., 2013; Iyer et al., 2017; Hertoghs et al., 2019]. In fact, Ifn-interference may be one of many characteristics that distinguishes it from chronic (non-transmitted) variants and may play an important role in establishing the transmission bottleneck [Ashokkumar et al., 2020; Gondim et al., 2021; Sugrue et al., 2022] (reviewed by Nijmeijer and Geijtenbeek [2019]). Two reports found that the *env* gene of T/F strains resists the antiviral program by $CD4^+$ T-lymphocytes through Ifn-α- [Iyer et al., 2017; Rihn et al., 2017], Ifn-β- [Iyer et al., 2017] and Ifn-γ-mediated responses [Rihn et al., 2017]. Another report found that secretion of Ifn-α by plasmacytoid DCs is less responsive to T/F strains than it is to chronic (non-transmitted) HIV-1 strains [Schwartz et al., 2018]. T/F viruses are also resistant to restriction factors produced by infected LCs [Hertoghs et al., 2019]. More importantly, molecular clones of T/F strains replicated at higher efficiency in $CD4^+$ T-lymphocytes than chronic strains in the presence of Ifn-α, even though the replication activities of both strains of HIV-1 were similar [Parrish et al., 2013; Iyer et al., 2017]. Data recently reported by Gondim et al. [2021] showed that variants during acute infection (likely T/F viruses) were highly resistant to Type 1 interferons, and that viruses from latently infected cells reactivated by suspension of ART were also resistant to Type 1 interferons.

One of the ways in which T/F viruses are more resistant to interferons than their chronic variants is by gp120 V3 phenotypes that circumvent IFITM-mediated fusion inhibition [Wang et al., 2017]. Interferon-resistance also results in higher replicative capacity, rendering a likely role it plays in establishing a productive infection during the acute phase [Zanini et al., 2015; Iyer et al., 2017; Ashokkumar et al., 2020; Sugrue et al., 2022]. There may be other T/F-associated

mutations in viral protein factors (like Vif and Vpu) that confer the interferon resistance phenotype; due to the focus of current research on the *env* gene of T/F viruses to improve efforts toward vaccine development, literature on such mutations specific to T/F viruses is limited. All in all, the findings described here suggest that transmission selectivity and susceptibility of T-lymphocytes to be infected by T/F strains also depends on a response to interferons. While interferons do in fact reduce HIV infection in T-lymphocytes, they may also select for infection by T/F strains.

Gaps, Contradictions, and Conjectures

Prior attempts at developing a working vaccine against HIV resulted in insufficient efficacy at best and outright failure at worst, and the reasons resulting in such failures are still under investigation. Some of the most important factors involved in these failures are ADE, CDE, non-enhancement of antibody-dependent phaogocytosis, genetic dissimilarities between vaccine epitopes and circulating T/F sequences, co-receptor switching, and non-native presentation of epitopes [Dolin et al., 1991; Mild et al., 2010; Forthal et al., 2012; Kijak et al., 2013; Sanders et al., 2015; Gach et al., 2017]. Despite the scant successes of the RV144 trial and the generation of next generation vaccines that have shown promise, there is still a great deal about vaccine design and epitopes that remain to be understood. The next logical step in this understanding is the very concept of the genetic bottleneck and how it is produced through selection. Thus it is important to address knowledge gaps and contradictions in the literature, especially in regard to characteristics of T/F viruses.

Many of the characteristics of T/F viruses have been investigated and even documented to considerable detail. Some of these include CCR5 tropism [Isaacman-Beck, 2009; Alexander et al., 2010; Parker et al., 2013; Ping et al., 2013; Song et al., 2016], sensitivity to neutralizing antibodies [Derdeyn et al., 2004; Gnanakaran et al., 2011; Wilen et al., 2011], subtype-specific N-glycosylation patterns [Chohan et al., 2005; Liu et al., 2008; Curlin et al., 2010; Wilen et al., 2011; Ping et al., 2013], shorter variable loops [Liu et al., 2008; Curlin et al., 2010; Gnanakaran et al., 2011], and utilization of α4β7 integrins as an alternative coreceptor [Arthos et al., 2008; Cicala et al., 2009]. However, many of these characteristics have come into question with recent publications showing that they are either less of a distinguishing characteristic of T/F viruses compared to chronic variants or that they may be experimental artefacts. T/F viruses do not particularly have higher binding affinity to α4β7 integrins than their chronic counterparts, neither does blocking α4β7 integrins inhibit the infectivity of T/F strains [Parrish et al., 2012]. T/F and chronic strains are also indistinguishable from their preference for primary infection of different subsets of $CD4^+$ T cells [Wilen et al., 2011], which presents doubt regarding previous data indicating preferential, initial infection of a particular subtype at the site of transmission [Zhang et al., 1999; Stieh et al., 2016] While CCR5 tropism is certainly a characteristic of most, if not all, T/F strains, this characteristic does not necessarily distinguish them from chronic variants that are also highly CCR5-tropic [Isaacman-Beck et al., 2009; Parrish et al., 2012]. Alas, the problem of identifying T/F characteristics deepen even further: T/F-specific phenotypes are also different among viral clades.

Characteristic differences between clade B and clade C viruses may lead to dissimilar selection pressures, which would complicate vaccine design. It already complicates identifying

universal characteristics of T/F viruses and virus-host interactions that select for or against them. Glycoprotein-120 of T/F viruses from clade B are under-glycosylated compared to those from clade C (reviewed by Monaco et al. [2017]). Given this difference in N-glycosylation patterns, one may surmise that T/F viruses from clade B and clade C would have higher preference for cell-free direction infection and *trans*-infection of $CD4^+$ T-lymphocytes, respectively. This is contradicted by an earlier study showing that viruses from clades A and D (which contain N-glycan-enriched variable loops) transcytosed more efficiently through non-cell-mediated means than did clade B viruses, however the likely reasoning here is that they bind through lectins on the surface of the model endometrium that the study utilized [Hocini et al., 2001]. Clade C T/F viruses would also have greater infection opportunities through mucosal barriers as previous evidence showed N-glycans facilitate binding to lectins on the surface of dendritic cells and epithelial cells [Hocini et al., 2001; Hong et al., 2002; Shen et al., 2014]. This interaction would not likely be inhibited by antibodies as the greater amount of N-glycosylation in V1 and V2 of gp120 from clade C T/F viruses evades antibody recognition, even though prior studies showed robust inhibition of transcytosis and binding to interdigitated myeloid cells of clade B T/F viruses by IgG and IgA antibodies [Hocini and Bomsel, 1999; Shen et al., 2010a; Shen et al., 2010b; Willey et al., 2011; Shen et al., 2014]. Even on the topic of N-glycosylation and its observed resistance to antibody-binding, Wagh et al. [2018] showed that while heavily glycosylated spike proteins were initially resistant to antibody-binding, their presence correlated with accelerated production of neutralizing antibodies, probably due to mutation of non-glycosylated regions that were initially recognized during acute and latent phases of infection. Paradoxically, N-glycans, which may be important for transmission and early immune evasion

by clade C T/Fs, may in fact be more prone to neutralization during chronic infection. This contradiction is further entrenched by studies demonstrating that concentration of N-glycans on gp120 from clade C are increased in some cases [Kerina et al., 2011] and unchanged in others [Parrish et al., 2013], while results by Wagh et al. [2018] suggest immune selection might favor the exact opposite for clade C viruses. While the fact that higher N-glycosylation of clade C viruses may become a serendipitous factor in an effective vaccine design [Wagh et al., 2018], the crux of the issue is this: T/Fs exhibit a wide range of characteristics that are cannot be uniformly synthesized into an all-encompassing T/F description. Furthermore, how the N-glycosylation profile of gp120 changes due to selection during early or late disease progression remains a mystery. Identifying universal characteristics between B-lineage and C-lineage T/F viruses, let alone identifying common immunogenic epitopes, has already presented a serious scientific challenge.

Clade B and clade C T/F viruses also differ in their relative resistances to interferons and their downstream signaling pathways. Clade B T/F viruses are equally resistant to interferons as clade C T/F viruses are [Parrish et al., 2013]. Over the course of immune selection, however, the progeny of clade B T/F viruses lose that resistance in favor of greater N-glycosylation and longer variable loops for a net enhancement of replicative fitness during immune selection, while those from clade C do not [Parrish et al., 2013]. The reasoning for this may be that clade C T/F viruses are already heavily N-glycosylated compared to T/Fs from clade B, thus potentially deleterious mutations to evade antibody recognition during late acute infection would not be necessary in clade C.

Many of these findings present a scientific controversy, however, as they have been contradicted by other reports in the literature. In one study investigating the characteristics of clade C T/Fs, Ashokkumar et al. [2018] showed that T/Fs from this lineage are not as replete in N-glycosylation sites as originally considered. They also bear shorter V1 and V2 loops compared to chronic viruses, just as clade B T/Fs do [Ashokkumar et al., 2018]. T/Fs from clade C still carry more N-glycans than T/Fs from clade B, but the degree to which clade C viruses are N-glycosylated and how such glycosylation patterns are influenced by selection are not fully known. Clade C T/Fs also utilize an additional co-receptor: CXCR6 [Ashokkumar et al., 2018]. In a later report, the same authors found that clade C T/F viruses were significantly more resistant to Ifnα than non-transmitted chronic viruses, as previously considered [Ashokkumar et al., 2020]. The studies conducted by Ashokkumar et al. [2018; 2020] is not without experimental criticisms, however: They cloned clade C gp120 from a single chronically infected mother and her children into an *env* backbone derived from a clade B virus. This chimera may have altered the phenotypes that Ashokkumar et al. [2018; 2020] measured.

Despite the limitations of the previously mentioned study, it is not the only one that contradicted what is currently accepted about clade C T/Fs. Kerina et al. [2011] found that clade C T/Fs do undergo changes during disease progression as do those from clade B: V5 length increases due to insertions and other variable loops, like V4, accumulate more potential N-glycosylation sites. This disagrees with data reported by Parrish et al. [2013] indicating that clade C T/Fs undergo minimal changes, including mutations that produce greater N-glycosylation sites, during disease progression. Interestingly, Deymier et al. [2015] and Song et al. [2016] showed that transmission selectivity of clade C T/Fs were not influenced by Ifnα or

enhanced RC, in agreement with Parrish et al. [2013]. In sum on the basis of T/F characteristics among clade B and C HIV, reported findings in the literature complicate our understanding of clade C T/F phenotypes.

The mode of transmission and risk factors associated with them also produce dissimilar selection pressures, even for the same HIV subtype. The probability that a reported exposure would lead to multiplicity of infection (i.e. infection by multiple T/Fs) was found to be about13%, 21%, 30%, and 37% for female-to-male, male-to-female, male-to-male, and intravenous routes of transmission, respectively [Baxter et al., 2023]. The probability of multiplicity for mother-to-child transmissions in the same study was calculated to be 17%, 27%, and 18% for pre-partum, intra-partum, and post-partum periods, respectively [Baxter et al., 2023]. Although infection by a single T/F variant was more likely in most routes of transmission, the risk of infection by multiple variants varied by transmission route and often resulted in more adverse disease outcomes [Macharia et al., 2020]. The increase in probability of infection by multiple variants correlated with an increase in the practice of unsafe sexual practices and worsening disease outcomes [Li et al., 2010; Patel et al., 2014; Kelley et al., 2017]. While this correlation is not surprising and that safer sexual practice would likely result in a lower transmission risk, it is not clear from these data whether the route of exposure itself may influence transmission selectivity of T/F variants over other variants. A prevailing question remains: Is there any special physiological or immunological features of each tissue at the site of exposure that may favor (or disfavor) infection by certain *env* phenotypes? There is data in the literature that may provide insight, however limited, as to what those selection pressures entail.

Rectal transmission, especially involving condomless anal sex, is associated with one of the highest risks of HIV transmission [Patel et al., 2014; Baxter et al., 2023]. Keele et al. [2009] showed evidence in an animal model designed to replicate HIV transmission and pathogenesis as they occur in humans that SIV rectal inoculation resulted in the transmission of far fewer T/F viruses than did the intravenous route. This implies that rectal mucosa, perhaps other tissues as well, propagate a selective pressure that restricts the number of T/Fs during exposure. As of yet, it isn't known how rectal tissues influence transmission selectivity. One possibility could be in the characteristics of the cells associated with the rectal mucosa. Rectal and colonic mucosa are populated by myeloid cells, like DCs, and interaction between these cell types and HIV T/Fs would likely be mediated through complement opsonization and lectin binding [Bouhlal et al., 2007; Shen et al., 2010a; Shen et al., 2014]. HIV from the sexual fluids of chronically infected donor, whether male or female, is usually opsonized by complement, with or without antibodies [Hussain et al., 1995; Bouhlal et al., 2002; Bouhlal et al., 2007; Jenabian et al., 2010]. Complement forms a "bridge" between bound regions on the HIV spike protein and CR2 of $CD4^+$ T-lymphocytes and CR3 of DCs and rectal tissues [Hussain et al., 1995; Bouhlal et al., 2007; Willey et al., 2011]. Lectins on the surface of rectal epithelial cells also enhance transcytosis of HIV through the mucosa by binding the N-glycans of HIV gp120 [Saïdi et al., 2007; Shen et al., 2014]. Not only does this virus-host interaction enhance transmission through mucosal barriers and infection of target cells, it may also play a role in transmission selectivity of variants with favored phenotypes. For example, clade C T/Fs are considered to carry a greater concentration of N-glycans [Salazar-Gonzalez et al., 2008; Salazar-Gonzalez et al., 2009; Parrish et al., 2013]. One may infer that clade C T/Fs might be preferrentially selected over clade B T/Fs

through rectal transmission; one may also infer that depending on selection pressures extant at the time of transmission, certain clade C T/Fs may be favored merely on the basis of binding affinity with either complement receptors or lectins. However, there is very little data available to determine how such selection pressures might be modulated to favor one variant over another, if they occur at all.

The vaginal route also presents a unique environment for transmission. Vaginal transmission is not incredibly dissimilar from rectal transmission, in that both tissues contain lectin- and complement receptor-expressing myeloid cells and epithelial cells that transmit HIV either through (1) vaginal/cervical/endometrial transcytosis or (2) capture by interdigitated myeloid cells to transfer them to the mucosa and/or lymph nodes [Hocini et al., 2001; Shen et al., 2010c; Trifonova et al., 2018; Perez-Zsolt et al., 2019; Day et al., 2022]. This mechanism had been demonstrated earlier in animal studies by vaginal inoculation with SIV [Hu et al., 2000]. As was shown by Keele et al. [2009] in the case of rectal SIV inoculation of animals, Piantodosi et al. [2019] showed that HIV-1 transmission through female genital mucosa resulted in a low diversity of viral variants isolated from the cervicovaginal lavage by 4 days post infection, leading ultimately to a single viral variant by 11 days post infection. This result indirectly corroborated a prior finding that bypassing the vaginal mucosa through severe concurrent STIs and inflammation of the female genital tract resulted in transmission of multiple T/F variants [Haaland et al., 2009]. As was the case in transmission through rectal tissues, transmission of multiple SHIV (SIV/HIV hybrid) T/Fs through the vaginal route in *Rhesus macaques* correlated with worsening disease progression [Tsai et al., 2014]. Mucosa of the female genital tract is

composed of similar types of cells as that of rectal tissues, and likely establish a similar restriction bottleneck during viral transmission on the basis of its epithelial architecture.

However, key differences in the anatomy, cell surface profile, and cytokines at either site could produce vastly different selection pressures that influence the transmission of HIV so as to favor one T/F variant over others from a donor. Unlike the rectal mucosa, vaginal macrophages are more permissive to HIV-1 infection, representing one possible avenue by which selection for T/Fs with a higher N-glycosylation profile on the viral spike protein [Shen et al., 2009]. On the other hand, where rectal epithelial cells and vaginal macrophages may favor transmission of T/F viruses from clade C over clade B, prior animal studies showed that the initial cell type infecting during vaginal SIV challenge is the $CD4^+$ Th17 cell [Stieh et al., 2016]. Criticisms over the methodology and conclusions of this paper have already been addressed, but it does not overscore the possibility of a distinct, vaginal/cervical/endometrial-specific selection pressure. Microabrasions and tissue damage caused by frictional vaginal sex could recruit $CD4^+$ T-lymphocytes, Th17 cells, and other mononuclear cells to the site of exposure and release proinflammatory cytokines [Fraser et al., 1999]. This particular inflammatory environment may select for interferon-resistant (or other cytokine-resistant) T/Fs that directly infect $CD4^+$ T-lymphocytes over heavily N-glycosylated non-transmitting viruses that adhere more strongly to DCs, LCs, and macrophages. Additionally, the acidic nature of vaginal secretions, shown previously by Aldunate et al. [2013] to inhibit HIV infectivity, may propagate an entirely different selection pressure than tissue inflammation alone. T/F viruses that are naturally acid resistant might be selected in this scenario. Conversely, concurrent STIs are known to alter the vaginal microbiota, resulting in a more basic environment [Cone et al., 2014]. Not only is such

alteration and inflammation associated with a higher risk of HIV transmission, it may modulate the selection pressures influencing T/F transmission, perhaps favoring viruses that would normally have exhibited reduced fitness in the acidic environment.

While the evidence currently reported indicates transmission selectivity of certain T/F phenotypes is the most likely cause of the genetic bottleneck [Abrahams et al., 2009; Salazar-Gonzalez et al., 2009; Parrish et al., 2013], the predictability of the transmission of T/F variants from a donor remains a serious challenge for researchers. This point is exemplified by the statistical improbability of predicting the T/Fs in a recipient from the plasma of a transmitting donor [Frange et al., 2013]. This is to say: There is very little that can be learned from viruses contained in the blood compartment of a chronically infected individual that could inform us on the characteristics of a potential T/F strain. For example, mother-to-child transmission in one case study resulted in vastly different T/F phenotypes in the infants than what had been found through horizontal transmission between sexual partners, despite the same donor source [Ashokkumar et al., 2016]. These phenotypes differed greatly in the length and N-glycosylation patterns of the V1 and V2 regions of gp120, even when all recipients in the study were infected by the same donor. In another report involving transmission of multiple T/Fs, the characteristics of Ifnα-resistance and high RC were not uniformly shared by all T/Fs, even within the same multiply infected individual [Song et al., 2016]. This would imply either stochastic transmission (less likely) or a combination of selection pressures that would favor the transmission of different characteristics exhibited separately by different T/Fs (more likely). Likelihood on mechanism aside, there is still insufficient information on how selective pressures could have influenced the transmission of multiple variants with different phenotypes even in the same individual. While

this is largely a problem in only a fifth of sexually transmitted HIV cases, the majority of cases are caused by a single virus, studied by researchers often months or years after initial infection [Keele et al., 2008; Salazar-Gonzalez et al., 2008; Fischer et al., 2010]. Animal and clinical studies have shown that T/Fs tend to be "animal-specific": In cases where a recipient is infected by a single T/F strain of SIV or HIV, the T/F strain of one individual is rarely the same as the strain infecting another, even from the same donor [Keele et al., 2009; Salazar-Gonzalez et al., 2008; Ma et al., 2011; Yuan et al., 2017]. Transmission by specific T/Fs that do not share uniformly the same characteristics to different recipients even by the same donor complicates our understanding of what genetically defines a T/F strains.

Our understanding of selection pressures is not yet complete enough to characterize T/F viruses. In fact, selection pressures are not likely the same across all potential sites of exposure. HIV vRNA is detectable in different tissue and plasma compartments of an infected individual [Shepard et al., 2000]; there may be selection pressures influencing either the transmission to or from these tissues. The low pH of vaginal secretions, for example, are generally hostile to HIV, and pH-altering vaginosis may increase the risk of HIV transmission through vaginal, cervical, or endometrial epithelia [Atashili et al., 2008; van de Wijgert et al., 2008; Aldunate et al., 2013]. Contrasting this from the nearly neutral pH of the rectum and colon [Nunes et al., 2014], not only might there be a higher risk of transmission associated with penetrative anal sex compared to penetrative vaginal sex there could be vastly different selective pressures favoring one T/F strain in one compartment over another merely on the basis of differences in tissue pH. An acid-resistant strain, should one exist, might be favored for transmission through healthy cervical tissue while that particular pressure might be absent in the rectum with an entirely different

pressure favoring a different strain altogether. Thus, transmission selectivity by tissues has been a topic largely understudied since the discovery of the genetic bottleneck; only in recent years have researchers begun to examine how tissues might compartmentalize particular strains of HIV and how such compartmentalization could influence or be influenced by transmission selectivity.

Boeras et al. [2011] asked whether T/F viruses could be transmitted from a viral population of low genetic diversity compartmentalized in the male genital tract. They found that despite the proximity of seminal fluids to tissue-compartmentalized HIV there was no evidence tissue-specific T/F viruses from male donors in the female recipients [Boeras et al., 2011]. While this seems to suggest that the T/F virus came from the plasma compartment, the lack of evidence confirming one pathway is not evidence against it. The question remains open as to whether a T/F virus may be sampled from tissue compartments versus the plasma compartment. Transmission to women, however, paints a slightly different picture.

Bull et al. [2013] showed that cervical and plasma compartments of eight women contained one or two HIV-1 T/F strains. The strains tended to be compartment specific (i.e. one T/F variant would have diversified into tissue-specific and plasma-specific lineages) until about two years post infection; beyond which time, the tissue became populated with variants from one or the other lineage, but rarely both. Furthermore, the diversity found in the tissue compartment reflected that in the plasma compartment, suggesting that T/Fs, even in the cervix, are not immutable during the course of disease progression. Piantadosi et al. [2019] also examined the female genital tract and plasma of women infected with HIV-1. They found a transient population of viral strains in the cervical and vaginal secretions with up to than 6% single nucleotide polymorphism diversity as late as one day since positive detection in one individual

and four days since in another. This diversity, albeit low by HIV genetic standards, reduced over the course of one to two weeks, rendering more a more homogenous viral population in both compartments. A third female did not have any detectable single nucleotide variant from the transmitted virus in either compartment during the observation period. Though this sample size may be too small to come to a conclusion regarding tissue specificity of particular variants, it did show that tissue compartmentalization is possible and that diversity of those variants decrease over the course of acute infection. Similar findings also occur in the male genital tract, but unlike the female genital tract, compartmentalization of male tissue-specific viral variants persists throughout chronic infection [Tortorec et al., 2008]. Even initial transmission by multiple T/Fs early during acute infection could resolve to a single, productively infecting T/F strain [Kijak et al., 2017], even if it doesn't occur in every case; this certainly occurs in plasma, and it could very well occur in vaginal, cervical, or endometrial tissues. Thus T/Fs likely form a transient population in cervical tissues for one week at the earliest and a year to two years at the latest.

On the other hand, Yuan et al. [2017] showed no rectal compartmentalization of SIVmac251 T/Fs in rectally inoculated animals. Though this study did not use HIV, its results remain applicable to transmission dynamics of HIV, given that SIV infection of *Rhesus macaques* can be achieved and manipulated to replicate similar pathologies as observed in humans infected with HIV [Keele et al., 2009]. A possible explanation for this could be direct infection of $CD4^+$ T-lymphocytes through the gut and rectal mucosa via specific receptors, which then migrate to gut-endemic lymph nodes through infected cells. Arthos et al. [2008] demonstrated that to be the case: α4β7 integrin on the surface of gut-associated $CD4^+$ T-

lymphocytes bound HIV-1 efficiently through the rectal and colonic mucosa. Infected cells could then carry the virus to lymphatic tissues where the virus may spread to vulnerable target cells and disseminate throughout the body. While Parrish et al. [2012] showed that α4β7 integrin-binding is not strictly a specific characteristic of T/F viruses, it does not discount the probability that HIV-1 T/F strains could infect T-lymphocytes through the gut or urethral mucosa [Pudney and Anderson, 1995]. Another study showed that rectal transmission of HIV-1 T/Fs in humanized mice disseminated to the bone marrow, spleen, thymus, and lymph nodes [Chateau et al., 2013], which supports the hypothesis that HIV compartmentalizes in tissues outside of the rectum during rectal transmission. Perhaps the lack of evidence for SIV tissue compartmentalization as reported by Yuan et al. [2017] might be due to this phenomenon, and that T/Fs may have disseminated to proximal lymphoid tissues. In fact, a later study showed that to be the case. Chen et al. [2018] reported that animals intrarectally infected with a high viral load of SIVmac251 also did not contain detectable SIV variants in the rectal compartment; instead, multiple T/F viruses were found compartmentalized in other tissues higher in the GI tract, the spleen, and the plasma. The same very well could be true for HIV in humans during rectal transmission.

In order to illustrate the state of ambiguity and unpredictability regarding information on the genetic bottleneck and T/F selection, consider an example in which an individual with a concurrent rectal STI is exposed to a viral swarm. It is possible that selection pressures in the rectum can be sufficient to select for T/F viruses over any other non-transmitting chronic variant during a concurrent STI due to hightened cytokine levels and recruitment of $CD4^+$ T-lymphocytes. This is certainly plausible given the natural interferon-resistance of T/F strains,

and their lower N-glycosylation profile enabling higher infectivity of $CD4^+$ T-lymphocytes through direct, non cell-mediated means. By the very same circumstance of a concurrent STI, the epithelial damage caused by the inflammatory response may allow chronic variants to bypass selection pressures favoring T/F variants and contribute to multiplicity of infection. This is not to say that inflamed tissues can both select for and against chronic strains at the same time in equal measure, nor that there exists a contradiction in our understanding of the genetic bottleneck. Rather, it is a problem of limited understanding. The transmission bottleneck is the result of a net sum of all competing selection pressures extant in the tissue at the time of exposure, much like a balance containing weights on both arms. In the case of T/F selection, one arm may contain the selective pressures (for example, cytokines) favoring one T/F variant while the other contains selective pressures (for example, dendritic cells and complement) favoring a different variant. The case of competing seminal components as discussed earlier may in fact represent such a balancing act: Where semen contains components that enhance [Bouhlal et al., 2007; Ceballos et al., 2009; Nijmeijer et al., 2021] and impede HIV-1 transmission [Martellini et al., 2009; Stax et al., 2009; Balandya et al., 2010], the favorable transmission of certain viral strains in the semen may depend on the relative affinities of HIV-1 glycoproteins to bind/interact with such components in whatever magnitude each factor might be present at the time of transmission. The bottleneck, in this analogy, is nothing more than the tipping of the scales toward on T/F variant over another on the basis of how much weight or strength in the form selective pressure is given to that T/F variant, and it is a view shared by others in the field [Mosa, 2021]. To date, the literature is deplete in sufficient mechanistic explanations for how the genetic bottleneck works. More troubling still is the fact that studies hitherto have focused too

heavily on T/F viruses that have been discovered as a representation of most (if not all) likely T/F viruses, even those yet to be discovered. The impact of this categorical fallacy might be a probable cause of the inefficacy of HIV vaccines and the inaccuracy of T/F prediction.

All information garnered on T/F viruses have been obtained either by direct examination of transmitted viruses in newly infected individuals or by *in vitro* studies using chimeras and pseudoviruses, and this is part of the problem. Parrish et al. [2013] even highlighted this as a persistent problem in T/F research. The poor predictability of identifying T/F strains, therefore, may be due to the fact that they are often compared to non-T/F strains, chosen from the swarm of virus in the sexual partner or mother: There is no probability given to the likelihood that a non-transmitted strain could have been transmitted in a selective manner. There is a tendency to ascribe the characters of the few T/F viruses currently documented as universal characters of all potential, yet undiscovered T/F viruses. The epistemological issues regarding T/F characterization and identification is analogous to the problem associated with rational vaccine design [Regenmortel et al., 2018], wherein there is an implicit propensity to ascribe the high breadth of antibody neutralization in one individual to robust vaccine-elicited immunization in a different individual. Such approaches, while promising in most respects, are not immune (pun intended) to inductive fallacies or researcher bias. Just as rational vaccine design doesn't always take into account the immunological and physiological conditions of an elite controller that enabled bnAb generation during acute infection, so do investigations that attempt to identify T/F characteristics without reference to the selective pressures that favored their transmission, often days, months, or years before patient samples were acquired and studied. Thus, there exists a "misclassification" problem wherein non-T/F strains are probably wrongly classified as such. In

addition to the problem of misclassification, the very fundamental questions of where and how the bottleneck occurs and, in particular, what immunological or other factors result in selection of a certain T/F strain over all other possible strains are not known. Understanding the mechanisms that contribute to the genetic bottleneck and the phenotypic characteristics of T/F strains is crucial in the development of a vaccine against HIV since it is the T/F strains that must be targeted by vaccine-elicited immunity. Most, if not all, vaccine attempts have not factored epitopes from T/F strains, and if it were it would be uncertain whether the chosen epitope from one or even several discovered T/F strains would elicit a properly tailored immunity against any other undiscovered T/F strains. Therefore, a faster, more scientific approach to T/F discovery would be to use models in an *a priori*, rather than *a posteriori*, setting. Indeed, the lack of predictive experimental models is a serious epistemological problem in T/F research [Frange et al., 2013].

Central Hypotheses and Specific Aims

My research was focused both on developing methods to better characterize T/F strains to differentiate them from non-transmitted strains and on pinpointing immunological and physiological features underlying the genetic bottleneck. My objectives here were to develop *in vitro* assays that can predict *a priori* which variants could be T/F strains from a swarm of HIV-1 or SIV. Given evidence from the literature that the bottleneck occurs at the level of the mucosa and/or dissemination, I tested the following hypotheses with the following specific aims:

#1 Viral transcytosis through epithelial barriers results in selection of T/F strains.

Specific Aim 1: To determine if transcytosis through an epithelial barrier results in selection of HIV-1 strains, I performed transcytosis assays on immortalized cell lines and human mucosal explants using an HIV-1 swarm (inoculum). Transcytosed viruses was quantified by reverse transcription-quantitative PCR (RT-qPCR). To characterize the strains selected by transcytosis, viral RNA (vRNA) isolated from transcytosed viruses was reverse-transcribed, PCR-amplified, and sequenced for phylogenetic analysis.

#2 Transfer of T/F strains from dendritic cells to $CD4^+$ T cells results in selection of T/F strains.

Specific Aim 2: To determine if DC-to-lymphocyte transfer results in selection of HIV-1 strains, I performed *in vitro* infection assays in which I challenged dendritic cells and/or $CD4^+$ T cells with an HIV-1 swarm (inoculum). The post-infection viral output was quantified by RT-qPCR. To characterize the strains selected by DC-lymphocyte transfer, vRNA was isolated from supernatants post-infection, reverse-transcribed, PCR-amplified, and sequenced for phylogenetic analysis.

#3 Antibody, complement, or cytokines contribute to the selection of the T/F strain at the level of transcytosis or transfer from dendritic cells to $CD4^+$ T cells.

Specific Aim 3: At the time of transmission, T/Fs may be opsonized by immunologically important molecules, like antibodies and/or complement. Furthermore, HIV transmission and/or concurrent infections may also rouse nearby myeloid cells to produce cytokines, like interferons, and may influence the transmission selectivity of interferon-resistant T/F strains. Thus, for this Specific Aim, I repeated Specific Aims 1 and 2 with prior opsonization of the viral inoculum with antibody or complement. I also repeated Specific Aim 2 with Ifnα2-activated T-lymphocytes, cocultured with or without dendritic cells, and challenged with HIV-1 inocula.

The Relevance of These Hypotheses

Prior studies on HIV transmission selectivity have inferred the sequence identities of T/F strains by phylogenetic analysis of isolated specimens from animal models or clinical samples [Keele et al., 2008; Salazar-Gonzalez et al., 2008; Keele et al., 2009; Salazar-Gonzalez et al., 2009; Fischer et al., 2010; Stone et al., 2010; Wu et al., 2012; Yuan et al., 2017; Piantadosi et al., 2019]. While published reports have thus far characterized T/F strains in detail, the mechanism of selection itself is not well understood. Indeed, the current research methodologies used to study T/F strains from clinical isolates are limited in that specimens are only examined days to

weeks post infection. One can only garner a "before-and-after" view (assuming the researcher has access to specimens from donor and recipient pairs) of infection while the important information regarding the process of establishing infection may be lost. This problem can be circumvented by using *in vivo* animal models, such as *Rhesus macaques* infected with SIV or humanized mice infected with HIV. Such models may enable researchers to study the genetic bottleneck as it occurs *in vivo*. However, such investigations would be far more expensive and less experimentally analogous to humans than using live cells isolated from human specimens. Using *in vitro* models, like those utilized in this study, are less financially restrictive and may allow one to directly gauge specific cellular and molecular interactions known to occur during acute infection for their potential roles in transmission selectivity. The use of an *in vitro* model may provide precision in determining genetic and phenotypic characteristics of T/F strains, something that is unachievable using data generated from human or animal studies, where there is a literal "host" of physiological and immunological factors establishing selection pressures and they cannot all be identified *in vivo*.

It was the scope and aim of this study to develop a novel model system that would enable future researchers to study the bottleneck further and exploit this knowledge to develop more effective vaccines against HIV. Studying T/F variants is important for vaccine design as these are usually the transmitted variants from which the immunized individual must be protected. Studies involving T/F epitopes as immunogens have demonstrated robust protection in animal studies [Liao et al., 2013; Wagh et al., 2018; Schorcht et al., 2020; Tumba et al., 2022], showing unambiguously that T/F viruses are crucial to the development of a vaccine (also reviewed by Mosa [2021]). However, the field of study regarding T/F viruses is limited to only the ones that

have been discovered. This new model system is designed to replicate or simulate *in vitro* an infection with HIV as it would normally occur in a living person in two ways: (1) isolate any specific physiological or immunological factor that has previously been shown to exert some form of T/F-specific selection pressure and determine what new variants this pressure could favor from any clinical swarm of HIV-1 and (2) discover *de novo* selection pressures that have not been discovered hitherto using PBMCs and/or tissues from healthy human donors challenged with a clinical swarm of HIV-1. In doing so, it is the explicit goal of this experimental assay to greatly expand the search for T/F viruses from clinical isolates and improve efforts toward effective vaccine designs. Hence why this study was undertaken to facilitate those discoveries further.

CHAPTER 2: Methods and Experimental Approach

Testing Hypothesis #1

In vitro Transcytosis through a Model Epithelium

The transcytosis assay is an experimental procedure that assess the efficiency by which an infectious viral particle or other pathogen penetrates a model epithelium from the apical surface on top of a cellular monolayer to the basolateral region between the cellular monolayer; it does so by quantitating the percentage of viral particles from an inoculum that passed through the monolayer into the subnatant fluid beneath. This is a well-accepted assay that is performed in HIV research to study transmission dynamics of viral particles [Bomsel, 1997; Bomsel et al., 1998; Hocini and Bomsel, 1999; Hocini et al., 2001; Shen et al., 2010b; Dohgu et al., 2011; Jain and Rosenthal, 2011; Gupta et al., 2013; Kinlock et al., 2014; Shen et al., 2014; Gonzalez and Sagar, 2016; Anwar et al., 2022].

HEC-1A (endometrial) or T84 (colorectal) cells were be grown in cell culture inserts with a growth area of 0.3 cm^2 on a membrane containing 0.4-µm or 3-µm pores until an intact monolayer had formed. This was assessed by measuring trans-epithelial electrical resistance (TEER) across the membrane and visually estimating the percentage of growth area occupied by cells (confluence) (**FIGURE 3**). Once a monolayer had formed and TEER reached an appropriate level recommended by literature for the particular cell line [Gupta et al., 2013; Srinivasan et al., 2015], the apical portion of the monolayer was challenged with an inoculum of a clinical isolate of HIV-1 (92US657) containing 4-ng equivalents or 10-ng equivalents of p24. This inoculum was either pre-incubated with or without antibody and/or complement in order to

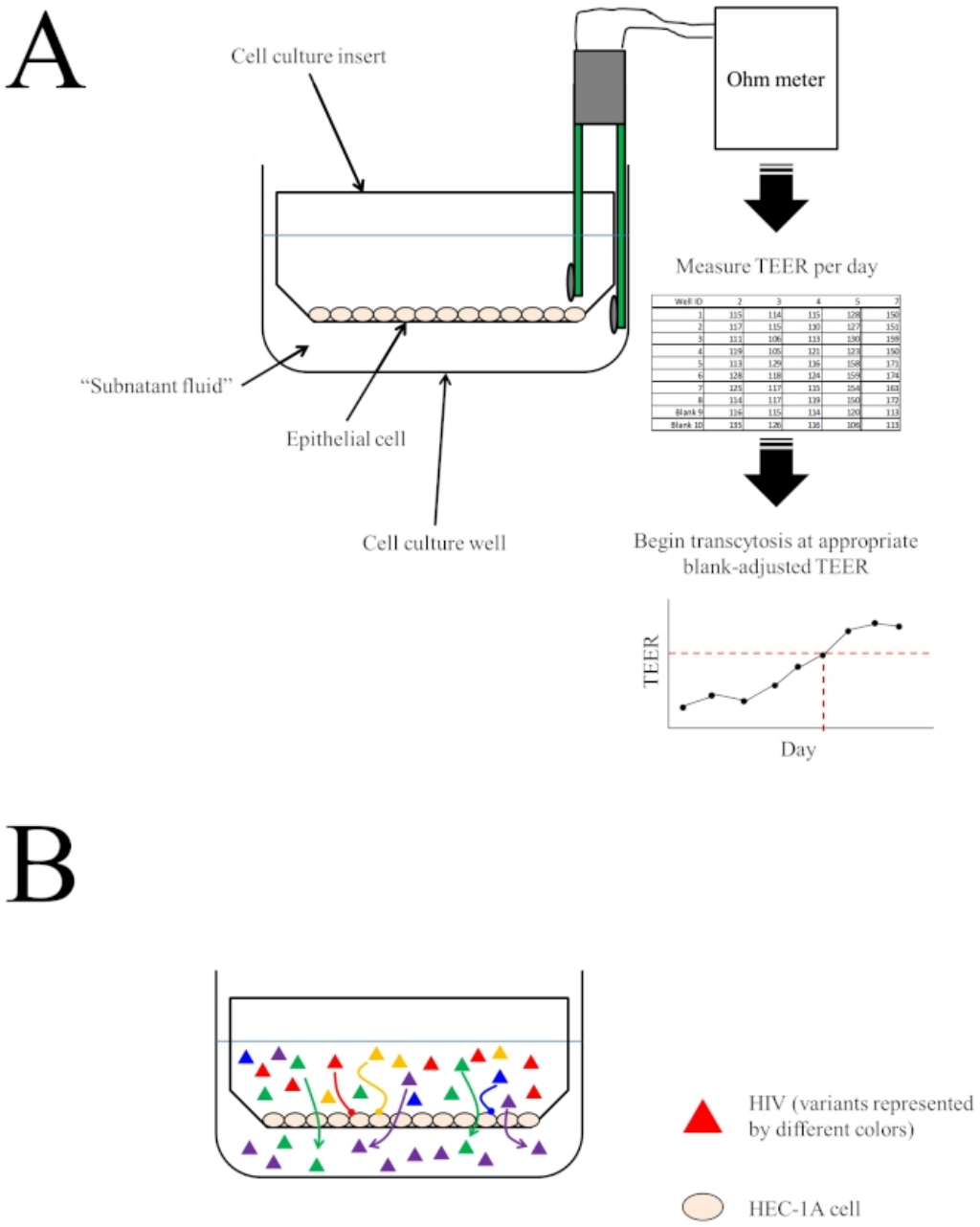


Figure 3: Experimental Setup for the *in vitro* Transcytosis Model: (A) The transcytosis experiment starts with seeding a trans-well insert with ~ 50,000 cells (of a particular epithelial cell line) and growing it until the trans-epithelial electrical resistance (TEER) approaches a value that normally indicates the formation of tight junctions. (B) Then the apical surface of the established monolayer is challenged with a viral inoculum, and the subnatant (culture media under the trans-well) would be sampled to identify viral variants that transcytosed through the cells.

determine the effect of these factors on viral transcytosis selectivity. After HIV-1 challenge, the cells were incubated at 37 ºC and 5% CO_2 for 12 hours. Then the subnatant fluid and inoculum was sampled and lysed. vRNA was isolated from the lysates and their copy numbers were measured by RT-qPCR to determine the percentage of the original inoculum that successfully transcytosed to the subnatant fluid.

Ex vivo Transcytosis through Human Colonic Tissue

Mucosal transmission of HIV-1 is likely a very important factor that influences transmission selectivity, which is evidenced in previously published reports [Arthos et al., 2008; Keele et al., 2009; Li et al., 2009; Stone et al., 2010; Klein et al., 2021]. The experimental designed described above is a generally well-accepted model to study viral transmission through a model that simulates epithelial tissue. However, this model is typically a homogenous culture of a single cell type, which presents several challenges: (1) it utilizes a culture of immortalized cancer cells that are morphologically and epigenetically dissimilar to normal cells that comprise epithelial barriers in a healthy individual and (2) it lacks the immune cells that likely play crucial roles in modulating T/F selection. Thus, the transcytosis system using cell lines grown on trans-well inserts is severely limited in that it assays T/F transmission selectivity through mucosal tissue isolated from any modulation by immune cells, like DCs, macrophages, and $CD4^+$ T-lymphocytes. A modification of the transcytosis assay allows researchers to include other cell types in coculture or interdigitated with epithelial cells across a trans-well [Collins et al., 2000; Noel et al., 2017]. While this is certainly an improvement of the previous design and isn't so isolated from the contributions made by immune cells populating mucosal tissues, there is still

yet another model system that enables examination of the genetic bottleneck and T/F selection through rectal, vaginal, and penile mucosae: an explant of the mucosa itself.

Studies of lentiviral transmission through cervicovaginal [Hu et al., 2000; Patterson et al., 2002; Shen et al., 2009; Trifonova et al., 2018; Perez-Zsolt et al., 2019; Klein et al., 2021], penile [Pudney and Anderson, 1995; Patterson et al., 2002; Fischetti et al., 2009; Dinh et al., 2015; Klein et al., 2021],epidermal [Kawamura et al., 2008; Bertram et al., 2019; Nijmeijer et al., 2021] and colorectal [Shen et al., 2009; Shen et al., 201a; Shen et al., 2010b; Kolodkin et al., 2013] mucosae have been performed previously. These studies have contributed immensely to working models of HIV-1 transmission using tissue explants, not only from humans but other animals as well. Their usefulness in studying T/F selection during transmission events cannot be understated, and human colonic explants were used to simulate such transmission events *ex vivo* for the current study. Via an approach adapted from previously published reports [Trifonova et al., 2018; Shen et al., 2010a; Shen et al., 2010b], mucosal explants were fitted to an apparatus as represented in **FIGURE 4**. The explants were challenged with an inoculum containing 10-ng equivalents of p24 of 92US657 HIV-1 clinical isolate. After 2 hours, the subnatant fluid was sampled and treated with vRNA lysis buffer; vRNA was isolated using Quick-RNA Viral Kit (by Zymo Research, catalog #R1035). To quantify HIV*gag* copy numbers of HIV-1 that transcytosed from the apical surface of the explants to the basolateral subnatant, RT-qPCR was performed on vRNA isolates as described in a later section (see "Reverse Transcription-Quantitative PCR"). The inoculum was also analyzed by RT-qPCR for its *HIVgag* copy number in order to determine the percentage of transcytosis during the given time period. The inoculum

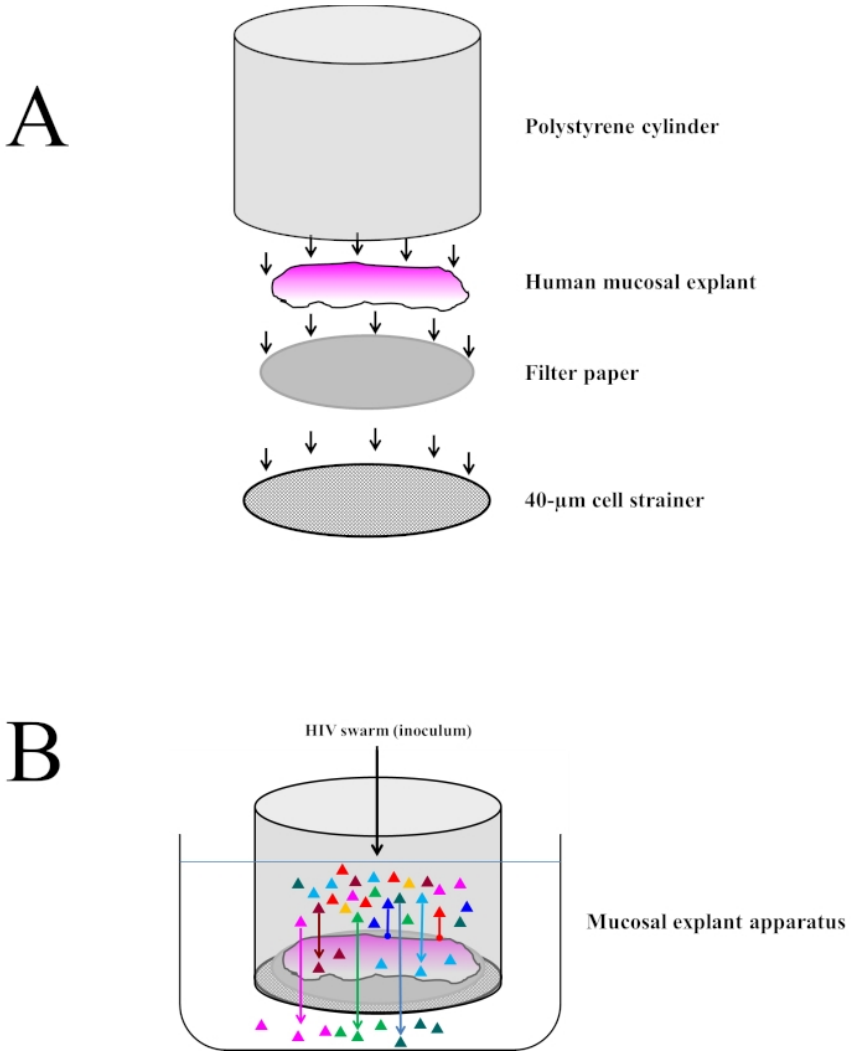


Figure 4: Experimental Setup for the *ex vivo* Mucosal Tissue Model: Adapted from Shen et al. [2010b], this explant model is utilized here to capture HIV variants from an infectious inoculum that may be selected by immunological and physiological factors present in the mucosa at the time of infection. (A) The apparatus is constructed by fixing sterile filter paper to a 40-µm cell strainer using surgical glue. Then a rectal mucosa explanted from a human donor would be fixed on top of the filter paper using surgical glue on the periphery, ensuring that excess glue does not occlude the center of the tissue on either the basolateral or apical sides. Finally, a plastic cylinder, such as a cut portion of a cylindrical cuvette, would be fixed to the cell strainer with surgical glue in a manner that it encloses the tissue without leaving any openings for virus to pass around the tissue. The cylinder acts as a reservoir for the infectious inoculum, while the apparatus is allowed to sit in an appropriate sized well, such as a 6-well or 12-well culture plate, and filled with culture media to act as a subnatant. (B) The reservoir would then be filled with an appropriate titer of an HIV-1 swarm and allowed to pass through the mucosal tissues and into the subnatant. Virus can then be sampled from the subnatant and the mucosa, if desired, to identify selected variants from the inoculum.

was sampled from a separate well within which only contained viruses from the inoculum in the exact amount that challenged the colonic explants.

In vivo Penile SIV Inoculation of *Rhesus macaques*

The *in vitro* colorectal and cervical systems employed by Noel et al. [2017] and Collins et al. [2000], respectively, expanded the transcytosis model to factor the roles of mucosal myeloid cells in HIV transmission. Following from this, the mucosal explant model is a further improvement from this experimental approach in that it utilizes the entire mucosa, not merely a coculture of individual cell types, to study HIV-1 transmission in a more clinically relevant context. Transmission is a complex process that cannot be understood in full by examination of transcytosis through a transwell system alone. Each of these three model systems, however, contributes to that understanding in specific ways for which every other one does not. This is to say: Each model system is, by themselves, necessary but insufficient to study transmission in its entirety.

Transcytosis through a homogenous monolayer allows researchers to study only the role played by epithelial cells in modulating transmission selectivity, which is likely obfuscated by complex interactions with immune cells, cytokines, and antibodies, whether from *in vitro* coculture systems, *ex vivo* tissues, or *in vivo* animals. This is one of the most important advantages of using transwell inserts *in vitro* to study HIV-1 transmission. The disadvantage in isolating this factor from all other factors normally associated with mucosal tissues (i.e. immune cells, tissue architecture, cytokines, antibodies, complement, etc.) is its inability to garner a more

complex view of HIV-1 transmission beyond just the epithelium. Accounting for these limitations requires deliberate manipulation of the experimental system by introducing additional cells and mucosal components to the model, hence the greater power intrinsic to those models employed by Noel et al. [2017] and Collins et al. [2000].

The mucosal explant model, on the other hand, provides greater avenues of study that overcome the limitations of *in vitro* transcytosis assays. The discoveries made using this experimental approach over the last two decades greatly enhanced our understanding HIV transmission. Despite the difficulty of isolating the specific roles played by each molecular and cellular component of mucosae, their versatility of use, ease of maintenance in culture, and tissue similarity to common routes of exposure make it a reliable experimental model for virologists interested in HIV transmission. Furthermore, *ex vivo* studies using tissue explants enable researchers to isolate the roles played by all of the mucosal components in concert (i.e. cells, antibodies, complement, etc.) at the site of transmission, and to investigate how those components interact with T/F viruses to modulate their selective transmission. The advantage of such a system isolated from the context of the larger organism is also its drawback: Non-mucosal factors that influence the genetic bottleneck would be absent in these studies, and their roles couldn't be investigated, even if immunological and physiological factors of the mucosa influence T/F selection across organ systems distal to the mucosal site of transmission. There is certainly evidence that bypassing the mucosal barrier significantly increases the number of T/F variants in an infected animal [Keele et al., 2009], the fact of which suggests an important role played by the epithelium in selection, but this alone cannot be the only factor. Dissemination of transmitted variants from the mucosa to lymph nodes is a necessary step in the establishment of

productive infection [Spira et al., 1996; Masurier et al., 1998; Shen et al., 2010a]. *Ex vivo* studies using mucosae are intrinsically disadvantaged in that they cannot be used to study the effect of dissemination on transmission selectivity, and knowledge of this process is still limited. It would be crucial, then, to employ *in vivo* animal models to expand the investigation of genetic bottlenecks beyond the mucosa, beyond select immune cells of the mucosa, and beyond the epithelial barrier external to the mucosa.

In order to circumvent the failures encountered in the two previous experimental approaches and to test Hypothesis #1 in the context of systemic viral spread of T/Fs in an organism, I used sera from a prior animal study designed to replicate the findings of Gupta et al. [2013]. Gupta et al. [2013] reported that HIV-1 transcytosis across a model epithelium was enhanced in the presence of non-neutralizing concentrations of antibody and low pH; they also found it enhanced transcytosis of T/F viruses, particularly. Gupta et al. [2013] attributed the mechanism of this enhancement to engagement between the neonatal Fc receptor (FcRN) on epithelial cells to the Fc domain of anti-HIV IgG. Our lab (in collaboration with Dr. Christopher Miller from University of California, Davis) performed a study in 2014 using *R. macaques* to test this hypothesis *in vivo*. Adapted from the methods in a previously reported study on SIV penile transmission [Qureshi et al., 2011], animals were infected via the penile route of exposure with an inoculum of SIVmac251 (sourced from SIV stocks generated by Del Prete et al. [2013]) that were pre-incubated in anti-SIV IgG or control IgG at either low pH (pH 6.0) or neutral pH (pH 7.4). The animals were challenged weekly with increasing doses of SIVmac251 from 1,000 $TCID_{50}$ up to 100,000 $TCID_{50}$ until infection was achieved. Plasma was collected from animals biweekly until viral setpoint was achieved; vRNA and p24 titers from each animal were isolated

and analyzed by RT-qPCR and ELISA, respectively. This protocol is outlined in **FIGURE 5**. Even though antibody opsonization and low pH have since been shown not to enhance transcytosis of virus across the penile epithelium in the current study and elsewhere in literature [Gonzalez and Sagar, 2016], it remains unclear whether this combination influences transmission selectivity. Thus, my intention for using these samples was not strictly to test whether antibody and/or low pH were associated with protection or enhancement of transcytosis through mucosal tissues: Rather my intention was to determine whether bottlenecks associated with such transmission exhibited observable transmission selectivity favoring T/F variants over others in a viral swarm and whether an immunologically important molecule like IgG could modulate that selectivity.

Sera at peak virema were treated with vRNA lysis buffer and at least 20,000 copies of SIV vRNA were isolated from the sera of each tested animal using the Quick-RNA™ Viral Kit (by Zymo Research, catalog #R1035) according to the manufacturer's protocol. vRNA was then reverse-transcribed to cDNA using SuperScript™ III First-Strand Synthesis System (by ThermoFisher Scientific, catalog #18080051) according to the manufacturer's protocol. vRNA strands from cDNA-vRNA duplexes were digested by adding 0.5 µL/sample of 2 U/µL RNaseH (by ThermoFisher Scientific, catalog #18021014) and incubated at 37°C for 1 hour. Then single genome amplification was performed as described in a later section (see "Single Genome Amplification by Nested PCR).

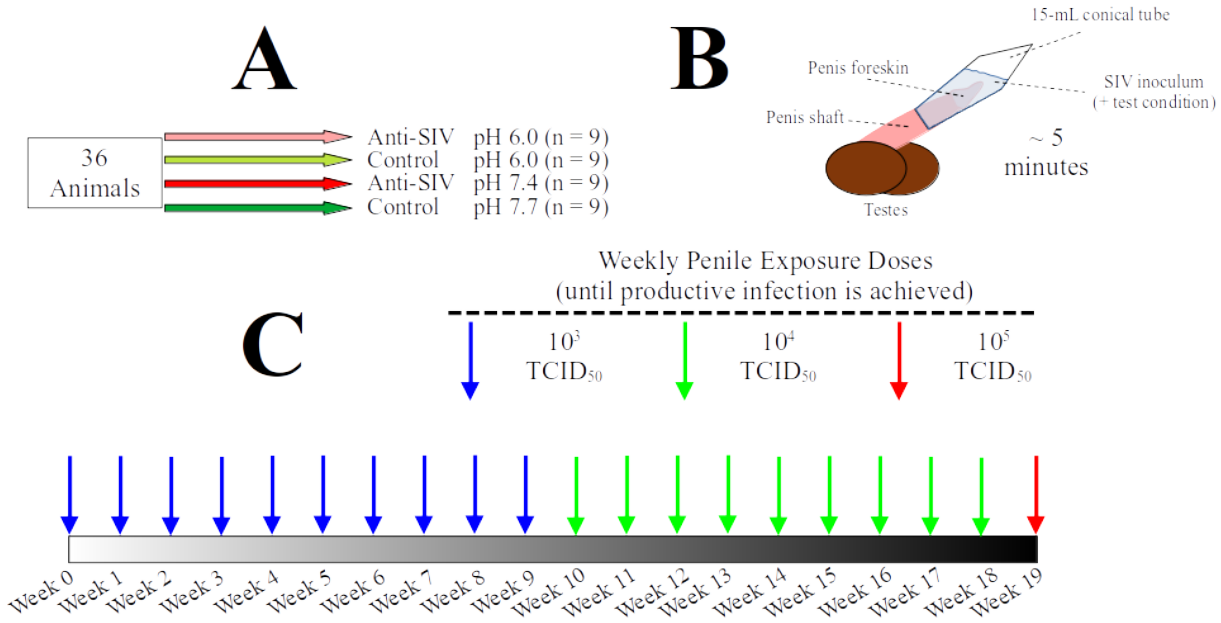


Figure 5: Experimental Setup and Schedule for *in vivo* Penile Inoculations: (A) Thirty-six *R. macaques* were divided equally into 4 different experimental groups. Group 1 animals were challenged with anti-SIV-coated SIVmac251 viruses at pH 6.0, group 2 animals were challenged with non-opsonized virus at pH 6.0, group 3 animals were challenged with anti-SIV-coated viruses at pH 7.4, and group 4 animals were challenged with non-opsonized virus at pH 7.7 (taken from historical controls for IACUC compliance). (B) Penile inoculations occurred by fitting a container, such as a 15-mL conical tube, containing the inoculum and one of the four test conditions over the penis shaft and exposing the retracted foreskin to the infectious medium for 5 minutes. (C) These inoculations occurred weekly for 20 weeks, with a dose of 1,000 $TCID_{50}$ for the first 10 weeks, 10,000 $TCID_{50}$ the next 10, and 100,000 $TCID_{50}$ on the final week until productive infection was achieved.

Testing Hypothesis #2

In vitro Infection of T-Lymphocytes Directly or through Virological Synapses with DCs

Transmission selectivity of HIV-1 T/Fs likely follows a complex, multifaceted process that starts with penetration of mucosal barriers to the productive infection of $CD4^+$ T-lymphocytes, either directly or indirectly through *trans*-infection. Prior studies have shown that T/Fs are only slightly more infectious toward vulnerable target cells, like activated $CD4^+$ T-

lymphocytes [Ma et al., 2009; Ochsenbauer et al., 2012; Parrish et al., 2012; Stieh et al., 2016], although such minute differences in infectivity aren't likely clinically relevant. In any case, preference for direct infection is a particular characteristic of clade B T/Fs that tend to contain fewer N-glycans than their chronic counterparts [Derdeyn et al., 2004; Chohan et al., 2005; Frost et al., 2005; Li et al., 2006; Liu et al., 2008; Gnanakaran et al., 2011]. T/Fs from clade C, on the other hand, are more inclined toward *trans*-infection than direct infection due to a higher concentration of N-glycans on gp120 variable loops, which enhances binding to carbohydrate receptors on myeloid cells [Shen et al., 2010a; Parrish et al., 2013; Shen et al., 2014; Trifonova et al., 2018; Bertram et al., 2019; Perez-Zsolt et al., 2019]. This is not to say, however, that clade-B T/Fs are incapable of *trans*-infection or that *trans*-infection does not play a role in the bottleneck for under-glycosylated strains. In fact, Parrish et al. [2013] showed that T/Fs have exhibited preference toward monocyte-adherence and *trans*-infection than their chronic counterparts, whether from clade B or calde C. Whatever the characteristic differences are between transmitted and non-transmitted viruses or between clades B and C viruses, selective pressures that propagate the genetic bottleneck of HIV transmission depends in large part on the phenotype of the selected T/F strain and that virus's interaction with various cell types in the mucosa and beyond. The second hypothesis of this study thusly focuses on such interactions and that *in vitro* modeling could replicate transmission selectivity as it has been known to occur in clinical studies. The power and innovation of this method is that it does not deviate in procedure from peer-reviewed and accepted infection assays as described in literature [Geijtenbeek et al., 2000; Ceballos et al., 2009; Stax et al., 2009; Balandya et al., 2010; Parrish et al., 2012; Shen et al., 2014; Perez-Zsolt et al., 2019; Nijmeijer et al., 2021], with the only exception being that

phylogenetic analysis was used to characterize potential T/F strains, if they were captured by the assay.

In order to test Hypothesis #2, I sequenced and phylogenetically analyzed output virus from infection assays using $CD4^+$ T-lymphocytes, either in coculture with or without PBMC-derived DCs. First, PBMCs were isolated from blood kindly donated to the lab by the nursing station on the first floor of Hewitt Hall, University of California, Irvine. PBMCs were isolated using SepMate™ PBMC Isolation Tubes and Lymphoprep™ (by StemCell™ Technologies, catalog #85450 and #07861, respectively) according to manufacturer's recommended protocols. Unstimulated $CD4^+$ T-lymphocytes were isolated by negative selection using either the $CD4^+$ T Cell Isolation Kit (by Miltenyi Biotec, catalog #130-096-533) or the EasySep™ Human $CD4^+$ T Cell Isolation Kit (by StemCell™ Technologies, catalog #17952). DCs were acquired first by isolating monocytes through positive selection using either CD14 MicroBeads (by Miltenyi Biotec, catalog #130-050-201) or the EasySep™ Human CD14 Positive Selection Kit II (by StemCell™ Technologies, catalog #17858), then immature DCs were produced by incubating isolated monocytes in 400 U/mL of interleukin-4 (IL4) and 1,000 U/mL of GM-CSF at 37°C for six days.

For direct infection, $CD4^+$ T-lymphocytes were either left unstimulated or stimulated with 10 U/mL of interleukin-2 (IL2) and 5 μg/mL phytohemagglutinin (PHA). Stimulated and unstimulated cells $CD4^+$ T-lymphocytes were maintained in 96-well culture plates with 200 μL of complete RPMI medium per well at 37°C. The T-lymphocytes were challenged with an inoculum swarm of clade-B HIV-1 clinical isolates: 92US712, 92HT593, 92US657, 92US660, and JRCSF (obtained from the NIH AIDS Reagent Program, catalog #ARP-3299, #ARP-2051,

#ARP-2053, #ARP-1722, and #ARP-394, respectively) at viral titers equivalent to 1 to 10 ng of p24. For *trans*-infection, DCs were incubated with the inoculum of the same clinical isolates at the same viral titers for 3 hours at 37°C. Then unbound virus was washed 5 times by centrifugation (1,000 rpm for 5 minutes), gentle aspiration of the supernatant, and resuspension in warm PBS, pH 7.4. On the final wash, the pelleted DCs with adhered virus were resuspended in complete RPMI medium and cocultured with stimulated or unstimulated T-lymphocytes. A conceptual representation of these modes of infection with regard to DCs and T-lymphocytes and how they would be studied to investigate infection selectivity *in vitro* are shown in **FIGURE 6**.

After 2 hours post infection, unbound virus remaining in the supernatant were removed by centriguation (1,000 rpm for 5 minutes) and gentle aspiration, then replaced with an equal amount of complete RPMI. Samples of the supernatant were collected at 3 days, 5 days, 7 days, and, in some cases, 10 days post infection. Viruses were lysed in vRNA lysis buffer, and vRNA was isolated using the Quick-RNA™ Viral Kit (by Zymo Research, catalog #R1035) according to the manufacturer's protocol. Viral titers in each of the experimental conditions were quantified by measuring the *HIVgag* copy number through RT-qPCR as described in a later section (see "Reverse Transcription-Quantitative PCR"). Downstream PCR of HIV-1 samples are described in a later section (see "Bulk Amplification by Nested PCR").

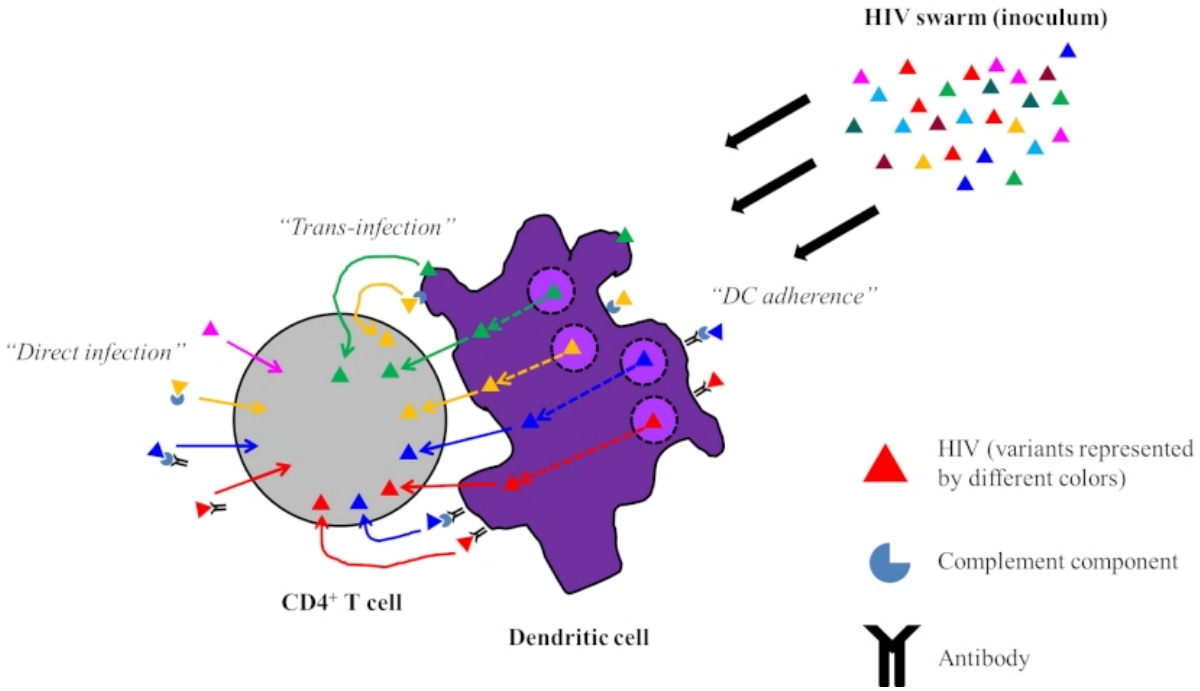


Figure 6: Modes of Infection and their Propagated Selection Pressures under Investigation by the *in vitro* Cell Culture Model: The two major modes of infection being tested in these approaches are *trans*-infection (where a monocyte like a dendritic cell transfers virus to a target T-lymphocyte) and direct infection (where a T-lymphocyte is infected by a virus directly without the involvement of an autologous transmitter like a dendritic cell). An additional kind of interaction that normally occurs between dendritic cells and HIV-1 is "DC adherence" (where virus binds to cell surface receptors like lectins), and this can play important roles in not only *trans*-infection but persistence of a temporary reservoir as dendritic cells disseminate the virus during acute infection. Each infection mode and cellular interaction is likely mediated, interrupted, or influenced in some way by immunologically important molecules, like complement and antibody. Thus an additional aim of this investigation would be to determine if and how antibodies and/or complement would be capable of influencing the selection pressures favoring (or disfavoring) of particular variants from a diverse viral swarm during infection and cellular adherence. Interferon priming is not shown in this diagram as that is a specific treatment of T-lymphocytes before HIV-1 inoculation; the purpose here is to generalize physical interactions between the virus and cell surfaces that may involve interfacing with other molecules.

Testing Hypothesis #3

The following experiments are modified versions of experiments carried out as previously described. And they are concerned more specifically with how certain immunologically important compounds (i.e. antibodies, complement, and cytokines) may modulate the selective pressures observed in those experiments. In fact, the experiments as described above serve as controls for the testing of the compounds listed below. Therefore, the reader may be referenced to procedures above as they read through the modifications described below.

Testing the Effect of Antibodies on the Selection of Variants during *in vitro* Infection Assays

Not only are antibodies crucial in the immune defense of a host from an infectious viral particle like HIV, they may also be crucial in selecting against (or for) certain viral variants. Of the culmination of antibodies from the donor that bind and possibly neutralize certain variants, transmission and infection may be more permissive to unopsonized virus or virus opsonized by non-neutralizing antibodies. For exampe, Shen et al. [2010b] showed that gp41-binding antibodies preclude cell-free transmission of HIV through the colorectal mucosa. There is little information on how such antibodies may facilitate cell-associated transmission or *trans*-infection of $CD4^+$ T-lymphocytes from DCs. Another paper suggested a still unconfirmed hypothesis that non-neutralizing antibodies could facilitate the transmission of T/Fs through mucosae and infect $CD4^+$ T-lymphocytes, either directly or by transfer from DCs [Gupta et al., 2013]. Therefore, infection assays using DCs and $CD4^+$ T-lymphocytes was modified to test for the effect of antibodies in modulating the genetic bottleneck and selective pressures that propagate it. For this experiment, four separate conditions were tested: viral opsonization with a low concentration of

anti-HIV IgG (HIVIG), a commercially available non-neutralizing polyclonal antibody against gp120, high concentration of HIVIG, high and low concentrations of control IgG (IVIG), or no antibody treatment at all. Separate aliquots of 92US657 were incubated in the presence of 200 μg/mL HIVIG, 22 μg/mL HIVIG, 200 μg/mL IVIG, 22 μg/mL IVIG, or absent of any antibody for 1 hour at 37°C. Then 10-ng p24 equivalents of opsonized or unopsonized virus were mixed with immature DCs for about 3 hours at 37°C, washed, and cocultured with stimulated $CD4^+$ T-lymphocytes using the same procedure as described above (see "*In vitro* Infection of T-Lymphocytes Directly or through Virological Synapses with DCs").

Testing the Effect of Antibodies on the Selection of Variants during *in vivo* Animal studies

Given failures encountered during transcytosis assays, I took on samples from an animal study performed in collaboration with Dr. Miller from University of California, Davis. This study involved penile inoculation of *R. macaques* with viral swarm of SIVmac251, incubated in the presence of non-neutralizing polyclonal anti-SIV or an IgG control. Antibodies were originally tested in transcytosis to examine transmission modulation by antibodies and changes in pH, per the hypothesis posited by Gupta et al. [2013]. In lieu of a working transcytosis model, *in vivo* penile inoculation was considered an appropriate substitution, as it not only infected animals through mucosal transmission but also enabled a more physiologically and clinically representative antibody-mediated modulation of the genetic bottleneck.

Using a method adapted form Qureshi et al. [2012], the current study utilized a method of fixing a 50-mL conical flask containing test mixtures described previously (see "*In vivo* Penile SIV Inoculation of *Rhesus macaques*") to the penis of 4 groups of 9 animals. Plasma was sampled from each animal with continual inoculation through the penile route at increasing viral

titers until productive infection was established (as described). Each sampled plasma was analyzed by RT-qPCR for SIV vRNA content. Viruses from plasma samples at peak viremia were lysed with vRNA lysis buffer, and vRNA was isolated using Quick-RNA Viral Kit (by Zymo Research, catalog #R1035). Twenty thousand to three hundred, twenty thousand vRNA copies were reverse-transcribed to cDNA using the SuperScript™ III First-Strand Synthesis System (by ThermoFisher Scientific, catalog #18080051) according to the manufacturer's protocol in order to optimize SGA-optimal dilution ratios (see section "Single Genome Amplification (SGA) by Nested PCR").

Testing the Effect of Complement on the Selection of Variants during *in vitro* Infection Assays

As discussed in an earlier section of chapter 1 (see "Antibodies and Complement" under "The Genetic Bottleneck and Transmitted/Founder (TF) Variants"), complement opsonization of virus plays a complex multifaceted role in HIV transmission (reviewed by Yu et al. [2010]). Seminal complement potently enhances transcytosis through [Hussain et al., 1995; Kinlock et al., 2014; Shen et al., 2014; Day et al., 2022] and possibly, though unlikely, infection epithelial cells [Bouhlal et al., 2002]. Complement is also capable of enhancing viral infection of undifferentiated monocytes through CR1 and CR3 [Thieblemont et al., 1993a], LCs through CR4 and CR5 [Nijmeijer et al., 2021], DCs through CR3 [Bouhlal et al., 2007], $CD4^+$ T-lymphocytes through CR2 [Boyer et al., 1991; June et al., 1991; Willey et al., 2011], *trans*-infection [Bouhlal et al., 2007; Forthal et al., 2012; Nijmeijer et al., 2021]. In sum, the interactions between HIV and complement enhance its viral infectivity in target cells and in penetration of mucosal barriers. However, despite current knowledge on CDE, there is very little known about how

complement may modulate transmission selectivity of T/F viruses beyond conjecture. Hypothesis #3 is intended to elucidate some of those unknowns, and the *trans*-infection assay is an apt procedure for testing it, as this very experiment has been performed previously but without downstream phylogenetic analyses or using laboratory HIV-1 strains.

Complement was isolated and pooled from blood plasma of three different HIV-negative subjects kindly donated to us by nurses in the nursing station in the first floor of Hewitt Hall. An aliquot of 92US657 was incubated in complete RPMI media with fetal bovine serum replaced with activated or heat-inactivated complement (control) for 1 hour at 37°C. Then direct and *trans*-infection assays were performed as described (see "*In vitro* Infection of T-Lymphocytes Directly or through Virological Synapses with DCs"). Each culture condition was challenged with viral inoculum titers of 0.1-ng equivalents of p24 per well. Supernatants were sampled as described (see "*In vitro* Infection of T-Lymphocytes Directly or through Virological Synapses with DCs") and processed for PCR amplification (see "Bulk Amplification by Nested PCR").

Testing the Effect of Interferon-α2 and Stimulation Status on the Selection of Variants during *in vitro* Infection Assays

Based on previous reports, T/Fs are generally more resistant to interferon-mediated antiviral responses than their non-transmitted counterparts (reviewed by Monaco et al. [2017], Soper et al. [2018], and Nijmeijer and Geijtenbeek [2019]). Such resistance is likely due to the enhanced RC of T/F strains, even among T/Fs and CCs with comparable infectivities [Sugrue et al., 2022]. Interferons may select for high-RC strains over low-RC strains during acute infection, and that such selective pressures may be a leading factor propagating the genetic bottleneck.

This is not to say that it is the only factor involved in the genetic bottleneck (or even that there isn't a case where interferon signaling never played a role favoring T/Fs in a bottleneck); it is certainly possible that other factors could have competed in the modulation of a net selective pressure toward one strain over all others. However, the phenomena guiding such selection, especially in regard to interferons, are vaguely understood. Elucidating these mysteries is crucial to understanding the bottleneck, identifying potential T/Fs, and developing more efficacious vaccines and therapies. In fact, a study published in 2017 showed that Ifnα-signaling enhances clearance of HIV-infected cells through ADCC [Tomescu et al., 2017], suggesting interferons as being a potential invaluable tool in the treatment of HIV infection, regardless of how it factors in the genetic bottleneck. Thus, *in vitro* experiments as outlined previously (see "*In vitro* Infection of T-Lymphocytes Directly or through Virological Synapses with DCs") would be repeated with the additional parameter of treating primary $CD4^+$ T-lymphocytes with Ifn-α2 then infecting them with HIV-1 swarms from various clinical isolates either directly or in *trans*.

Another probable factor of selection is the stimulation status of cells targeted by HIV. The stimulation of $CD4^+$ T-lymphocytes, an occurrence that normally occurs in response to cancer [Wang et al., 2023] or infection [Wik and Skålhegg, 2022], results in a higher representation of cell-surface markers and receptors, like CD4, CCR5, and CXCR4 [Hong et al., 1998; Pedersen et al., 2021]. The increase in cell-surface receptors for HIV cellular entry [Hong et al., 1998; Pedersen et al., 2021], cellular proliferation [Xing et al., 2010; Au-Yeung et al., 2014; Pedersen et al., 2021], and shifts in metabolic profile [Frauwirth et al., 2002; van Grevenynghe et al., 2008; MacIver et al., 2013; Man and Kallies, 2015; Menk et al., 2018; Gramatica et al., 2021; Wik and Skålhegg, 2022] enhance cell-free infection by HIV. This would

suggest that stimulated T-lymphocytes may be more permissive to clade-B T/F variants with a lower gp120 N-glycosylation signature than chronic strains during early acute stages. On the other hand, unstimulated $CD4^+$ T-lymphocytes can still be infected in *trans* by virus-adhered DCs more efficiently than by cell-free infection [Weissman et al., 1995], which could suggest an alternative mode of selection that could favor more N-glycosylated clade-C T/F variants or chronic variants of either clades B or C. Additionally, most *in vitro* infection assays involve infection of T-lymphocytes under stimulatory conditions in order to enhance viral output for downstream assays. This complicates *in vitro* T/F research: Stimulation in this setting artificially alters the outcome of selection and produces dubious results (evidence of which will be shown in Chapters 5 and 6). One of the most telling indications that such selection pressures exist along the stimulation axis is from a study published in 2013 indicating that resting $CD4^+$ T-lymphocytes were productively infected with a particular T/F that was phylogenetically distinct from T/Fs associated with persistently activated $CD4^+$ T-lymphocytes [Bacchus et al., 2013]. Beyond this, not much else is known about how stimulation affects the net selective pressure for one type of virus over any others. This is not to imply that stimulation itself does not modulate infection; however, HIV transmission does not always occur in the presence of persistent inflammation (its role in chronic infection is well studied, however, and reviewed by Nasi et al. [2014]), and the preponderance of studies utilizing stimulated T-lymphocytes obfuscates any understanding on how productive infection by T/F variants may be influenced in the absence of stimulation.

To overcome this methodological bias, $CD4^+$ T-lymphocytes isolated from PBMCs were divided into three groups: Unstimulated (U or Unstim), Post-stimulated (P or Post-stim), and

Stimulated (S or Stim). The purpose of such divisions was to test how stimulation status of infected T-lymphocytes and the timing of stimulation relative to infection modulate selection. The cells in S group were simulated with IL2 and PHA prior to infection to replicate immunological conditions wherein pre-existing inflammation (a consequence of a concurrent STIs, for example) was present at the site of transmission. Given that T-lymphocyte stimulation is a common practice of *in vitro* infection assays, the procedure outlined for this group is a standard procedure generally accepted by HIV researchers. The cells in P group were stimulated with IL2 and PHA posterior to infection. The experimental conditions here were intended to replicate immunological conditions of initial HIV infection during a transmission event where $CD4^+$ T-lymphocytes do not become stimulated until after productive infection has occurred. The cells in U group were not stimulated with either IL2 or PHA; this acts as a control for any kind of stimulation-mediated selection.

PBMCs were isolated from blood plasma acquired from the nursing station at Hewitt Hall. From those PBMCs, $CD4^+$ T-lymphocytes and DCs were isolated using previously described protocols (see "*In vitro* Infection of T-Lymphocytes Directly or through Virological Synapses with DCs"). The $CD4^+$ T-lymphocytes were divided into the three groups (U, P, S) and described in the previous paragraph. Half of the T-lymphocytes in each group was treated in the presence of 1,000 U/mL Ifnα2 and the other half with a vehicle control at 37°C for 24 hours. During the time of interferon conditioning, each Ifn-treated and control halves of the S group were also treated with 10 U/mL IL2 and 5 μg/mL PHA. After treatment conditioning, the $CD4^+$ T-lymphocytes were washed five times in PBS in the same manner described previously (see "*In vitro* Infection of T-Lymphocytes Directly or through Virological Synapses with DCs") to

remove excess cytokines then challenged with inocula sampled from five clinical isolates (92US712, 92HT593, 92US657, 92US660, and JRCSF) in swarms containing 0.16 – 0.20 ng equivalents of p24 per culture well. For *trans*-infection experiments, the viruses were adhered to DCs and cocultured with T-lymphocytes as described previously (see "*In vitro* Infection of T-Lymphocytes Directly or through Virological Synapses with DCs"). After 2 hours post challenge, cell-free virus in the supernatant were washed from the wells. On the next day, P group cells were challenged in the same manner as S group cells were. U group cells received no stimulation at all. Supernatants were sampled and replaced with fresh complete RPMI on days 3, 5, 7, and 10. A visual set up of these experiments is shown in **FIGURE 7**.

Viruses from each supernatant sample were lysed in vRNA lysis buffer, and vRNA was isolated using the Quick-RNA™ Viral Kit (by Zymo Research, catalog #R1035) according to the manufacturer's protocol. Viral titers in each of the experimental conditions were quantified by measuring the *HIVgag* copy number through RT-qPCR as described in a later section (see "Reverse Transcription-Quantitative PCR"). Downstream PCR of HIV-1 samples are described in a later section (see "Bulk Amplification by Nested PCR").

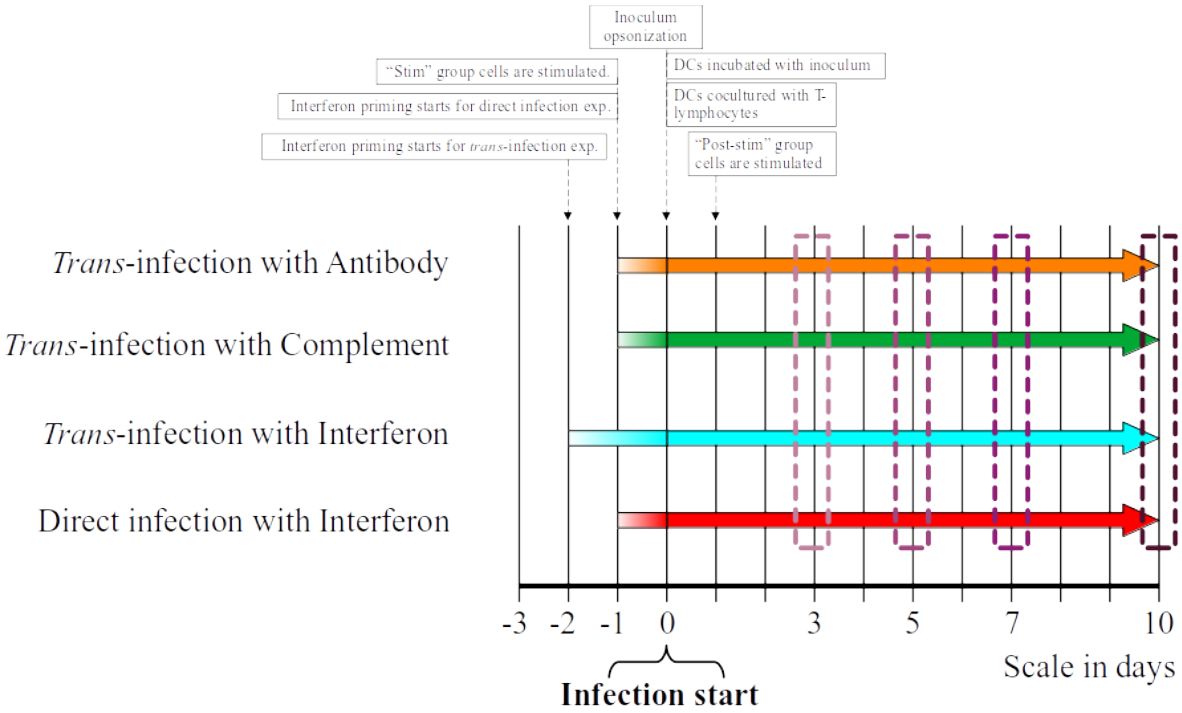


Figure 7: Experimental timeline for *in vitro* infection assays: Infection of T-lymphocytes will be tested in *trans*- and direct infections. For *trans*-infection, HIV would be coated with or without antibody or complement then incubated with dendritic cells for several hours prior to thorough washing and coculturing with unstimulated, post-stimulated, or stimulated T-lymphocytes. For direct infection, the coated HIV swarm would be placed in a monoculture of T-lymphocytes; no other cells would be cocultured. For interferon treatment, T-lymphocytes would be treated with Ifnα2 in order to activate interferon-stimulated genes at least one day prior to stimulation for cells intended to be stimulated, otherwise they were left in media containing interferon in order to maintain its primed status. They would then be cocultured with virus-bound dendritic cells one day after stimulation for *trans*-infection; for direct infection, they would be challenged with virus directly. For directly infected post-stim cells, stimulation occurs one day after viral challenge. Supernatant from each infection was sampled on 3, 6, 7, and 10 days post infection (dpi). After sampling on 10 dpi, the cells were lysed in cell lysis buffer and stored in -80°C for any potential downstream assays.

Bulk Amplification by Nested PCR

For downstream PCR, cDNA was synthesized from vRNA using SuperScript™ III First-Strand Synthesis System (by ThermoFisher Scientific, catalog #18080051) according to the manufacturer's protocol. vRNA strands from cDNA-vRNA duplexes were digested by adding

0.5 µL/sample of 2 U/µL RNaseH (by ThermoFisher Scientific, catalog #18021014) and incubated at 37°C for 1 hour. Nested bulk PCR methods here were adapted from previously published sources [van Maarseveen et al., 2006; Locateli et al., 2008; Fischer et al., 2010; Zarei et al., 2016; Kafando et al., 2017]. During first round PCR, the *HIVenv* gene was amplified in 20-µL reactions using 1 µL – 5 µL of sample cDNA as template and in the presence of 0.4 µM of forward and reverse primers (ENV_F and ENV_R) over 35 cycles using the PCRBIO HiFi *Pfu* Polymerase according to the cycling parameters recommended by the manufacturer (by PCR Biosystems, cat #PB10.41-20). For second round PCR, 2.5 µL of each first-round PCR product are added to each of four different 50-µL reactions and amplified over 45 cycles at the same primer concentrations utilized in the first round of PCR and using cycling parameters recommended by the manufacturer of the PCR reagents. For each sample, PCR amplifications in round 2 were performed over four distinct regions of *gp120*: (1) the signal peptide and the upstream portion of conserved domain 1 (SP), (2) variable loops 1 and 2 (V1V2), (3) variable loop 3 (V3), (4) variable loop 4, conserved domain 4, and variable loop 5 (V4V5). Target bands from each respective reaction were analyzed by agarose gel electrophoresis, extracted, and purified using the Zymoclean Gel DNA Recovery Kit (by Zymo Research, cat #D4008). Purified amplicons were submitted to the GHTF for sequencing through the MiSeq™ System by Illumina®. Sequences were processed further according to procedures outlined in a later section ("Phylogenetic Analysis and Statistics"). Primer sequences are listed in **TABLE 1**, PCR schema are shown in **TABLE 2**, and PCR-amplified regions are highlighted in **FIGURE 8**.

Table 1: Primers for HIV PCR Amplification with 5' Adapters for MiSeq Sequencing Platform

PCR Round	Primer Name	Sequence (5' → 3')*
Round 1	ENV_F	GTTTCTTTTAGGCATCTCCTATGGCAGGAAGAAG
	ENV_R	GTTTCTTCCAGTCCCCCCTTTTCTTTTAAAAAG
Round 2	SP_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGAAAGAGCAGAAGACAGTGG
	SP_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTTCTTGTGGGTTGGGGT
Round 2	V1V2_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACAGACCCCAACCCACAA
	V1V2_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAATCGCAAAACCAGCCG
Round 2	V3_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACGGACAATGCTAAAACCA
	V3_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTTCTGGGTCCCCTCCT
Round 2	V4V5_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACCCAGAAATTGTAACGCAC
	V4V5_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTTCTCTCTGCACCACTCTT
Round 2	712_V4V5_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGACCCAGAAATTATGACTCTC
	712_V4V5_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTTCTCTCTGCACCACTCTC
qRT-PCR	SK462	AGTTGGAGGACATCAAGCAGCCATGCAAAT
	SK431	TGCTATGTCAGTTCCCCTTGGTTCTCT

* Underlined nucleotides are Illumina Nextera adapters

Table 2: PCR Cycling Parameters for HIV-1 Bulk Amplifications

HIV-1 qRT-PCR

	Step	*Temperature*	*Time*
	1	50°C	30 min
	2	95°C	15 min
45 Cycles	3	94°C	15 sec
	4	55°C	30 sec
	5	72°C *	30 sec

HIV-1 Round 1 PCR

	Step	*Temperature*	*Time*
	1	95°C	1 min
35 Cycles	2	95°C	15 sec
	3	58°C	15 sec
	4	72°C	94 sec
	5	72°C	15 min
	6	4°C	HOLD

HIV-1 Round 2 PCR

	Step	*Temperature*	*Time*
	1	95°C	1 min
2 Cycles	2	95°C	15 sec
	3	55°C **	15 sec
	4	72°C	9 sec or 15 sec ***
43 Cycles	5	95°C	15 sec
	6	68°C	15 sec
	7	72°C	9 sec or 15 sec ***
	8	72°C	15 min
	9	4°C	HOLD

* Fluorescence measurement for qRT-PCR at this step.

** 48°C to be used for 92US712-specific primers. 52°C is the default annealing temperature for all other V4V5 primers.

*** 9-second extension time for SP and V3 PCR. 15-second extension time for V1V2 and V4V5 PCR.

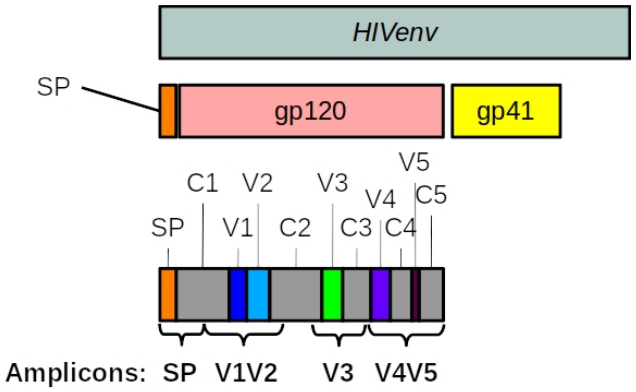


Figure 8: Variable Loop Regions of *HIVenv* Amplified by PCR: Regions of HIV gp120 would be PCR-amplified in two steps. First step is the bulk amplification of full-length *HIVenv.* The next step is the bulk amplification of SP, V1V2, V3, and V4V5 regions, indicated by the brackets in the diagram.

Single Genome Amplification (SGA) by Nested PCR

A common problem often associated with bulk PCR is the artificial introduction of substitutions into daughter strands. PCR amplification can result in 1 error per 1,000 to 10,000 nucleotides, even using high-fidelity PCR reagents on the market [Zanini et al., 2017]. This raises some concerns as to the accuracy of sequencing reads from shorter regions (like SP, V1V2, V3, and V4V5) that were subsequently PCR-amplified from a larger template like the *HIVenv*. This is especially the case during bulk PCR amplification methods as described in the previous section, and using DNA templates concentrated in at least 4,000 molecules per reaction. These errors would have to be handled and corrected during bioinformatic analysis, usually through consensus matching or software that eliminates regions with an average per nucleotide Phred quality score < 30 [Zanini et al., 2017].

Another workaround for this problem is by using single genome amplification (SGA). SGA is a PCR procedure that amplifies one single unique template in two PCR rounds, and such a procedure has been instrumental in the discovery and analysis of T/F viruses in the last fifteen years [Keele et al., 2008; Salazar-Gonzalez et al., 2008; Keele et al., 2009; Salazar-Gonzalez et

al., 2009; Stone et al., 2010; Gnanakaran et al., 2011; Willen et al., 2011; Bar et al., 2012; Wu et al., 2012; Bull et al., 2013; Parrish et al., 2013; Tsai et al., 2014; Park et al., 2016; Tully et al., 2016; Iyer et al., 2017; Yuan et al., 2017]. It involves one single round of PCR on serially diluted cDNA samples from a stock such that the dilution that produces a reaction-positive rate of 20% to 30% by the end of a second round of PCR would contain sufficient template to amplify a single unique strain or genome in the swarm from which it came [Keele, 2010]. This method is particularly sensitive as it is able to isolate a single template from a mixture of diverse templates that differ even by a few nucleotides out of thousands [Keele et al., 2008]. The proporition of reaction-positive amplifications of a particular variant is nearly equal in proportion of that variant in the original swarm, making sequences acquire by SGA numerically representative. It is also useful in limiting PCR-induced errors, like substitutions or recombinations, by diluting the starting template to a single amplifiable copy rather than thousands to millions of amplifiable copies. One drawback of this method is the amount of trial-and-error involved in identifying an SGA-optimal dilution; the researcher often starts blind when experimentally determining such dilution, and this dilution is impacted heavily by the PCR reagents used. For example, two different dNTP aliquots, one being more degraded than the other, could produce two vastly different reaction-positive rates, even at the same template dilution and with all other PCR reagents being the unchanged. The mathematical model developed by Butler et al. [2009] provided an invaluable tool for researchers to accelerate identification of SGA-optimal dilutions. There is, however, no other implementation that could reverse the damage to PCR reagents caused by the inescapable reality of the Second Law of Thermodynamics (aka continual freeze-thawing).

cDNA was synthesized from vRNA samples using SuperScript™ III First-Strand Synthesis System (by ThermoFisher Scientific, catalog #18080051) according to the manufacturer's protocol. vRNA strands from cDNA-vRNA duplexes were digested by adding 0.5 µL/sample of 2 U/µL RNaseH (by ThermoFisher Scientific, catalog #18021014) and incubated at 37°C for 1 hour. For the SGA protocol adapted from previously published methods [Keele et al., 2008; Salazar et al., 2008], cDNA was diluted serially from 10^{-1} to $10^{-10,000}$ in steps of 1/10. One microliter of each serial dilution was amplified first by one round of PCR using 0.4 µM of SIV-specific primers (SIVmacTatF1 and SIVmacEnvR1) to amplify the *SIVenv* gene in 20 cycles with the PCRBIO HiFi *Pfu* Polymerase (by PCR Biosystems, cat #PB10.41-20) according to the cycling parameters recommended by the manufacturer. The second PCR round was performed on 1 µL of the first-round PCR product (generated from each serially diluted template) using 0.4 µM of round 2 forward and reverse primers with 5' M13 sequencing adapters (SIVmacTatF2-M13 and SIVmacEnvR2-M13) over 45 cycles according to the manufacturer's recommendation. PCR products from the second round were analyzed by electrophoresis through 1% (w/v) agarose gel to determine the dilution ratio that produced a 20% to 30% reaction-positive rate. The nested PCR protocol as described above was repeated for that dilution until a total of 30 to 40 positive reactions were achieved. Target bands from each reaction were analyzed by agarose gel electrophoresis, extracted, and purified using the Zymoclean Gel DNA Recovery Kit (by Zymo Research, cat #D4008). Purified amplicons were submitted to the GHTF for sequencing through the SMRTLink™ Sequencing Protocol by Pacific Biosciences. Sequences were processed further according to procedures outlined in

"Phylogenetic Analysis and Statistics." Primer sequences are listed in **TABLE 3**, PCR schema are shown in **TABLE 4**, and PCR-amplified regions are highlighted in **FIGURE 9**.

Table 3: Primers for SIV SGA with 5' Adapters for SMRTLink Sequencing Platform

PCR Round	Primer Name	Sequence (5' → 3')*
Round 1	SIVmacTatF1	CCTCCCCCTCCAGGACTAGC
	SIVmacEnvR1	TGTAATAAATCCCTTCCAGTCCCCCC
Round 2	M13-SIVmacTatF2	GTAAAACGACGGCCAGTTATAATAGACATGGAGACACC
	M13-SIVmacEnvR2	CAGGAAACAGCTATGACATGAGACATRTCTATTGCCAATTTGTA

* Underlined nucleotides are M13 adapters (forward and reverse) for Pacific Biosciences Sequencing

Reverse Transcription-Quantitative PCR (RT-qPCR)

RT-qPCR was performed using the qPCRBIO SyGreen 1-Step Detect Lo-ROX system (by PCR Biosystems, cat #PB25.11). Five microliters of a vRNA sample (isolated according to procedures outlined previously) was diluted with 20 µL of master mix according to the manufacturer's recommended protocol. PCR cycling parameters are outlined in **TABLE 2** and primers used for each RT-qPCR assay are listed in **TABLE 1**.

Table 4:PCR Cycling Parameters for SIV Single-Genome Amplification

SIV SGA Round 1 PCR

	Step	*Temperature*	*Time*
	1	95°C	1 min
20 Cycles or 35 Cycles*	2	95°C	15 sec
	3	55°C	15 sec
	4	72°C	94 sec
	5	72°C	15 min
	6	4°C	HOLD

SIV SGA Round 2 PCR

	Step	*Temperature*	*Time*
	1	95°C	1 min
2 Cycles	2	95°C	15 sec
	3	55°C	15 sec
	4	72°C	94 sec
43 Cycles	5	95°C	15 sec
	6	65°C	15 sec
	7	72°C	94 sec
	8	72°C	15 min
	9	4°C	HOLD

* 20 cycles for > 4,000 vRNA copies
35 cycles for < 4,000 vRNA copies

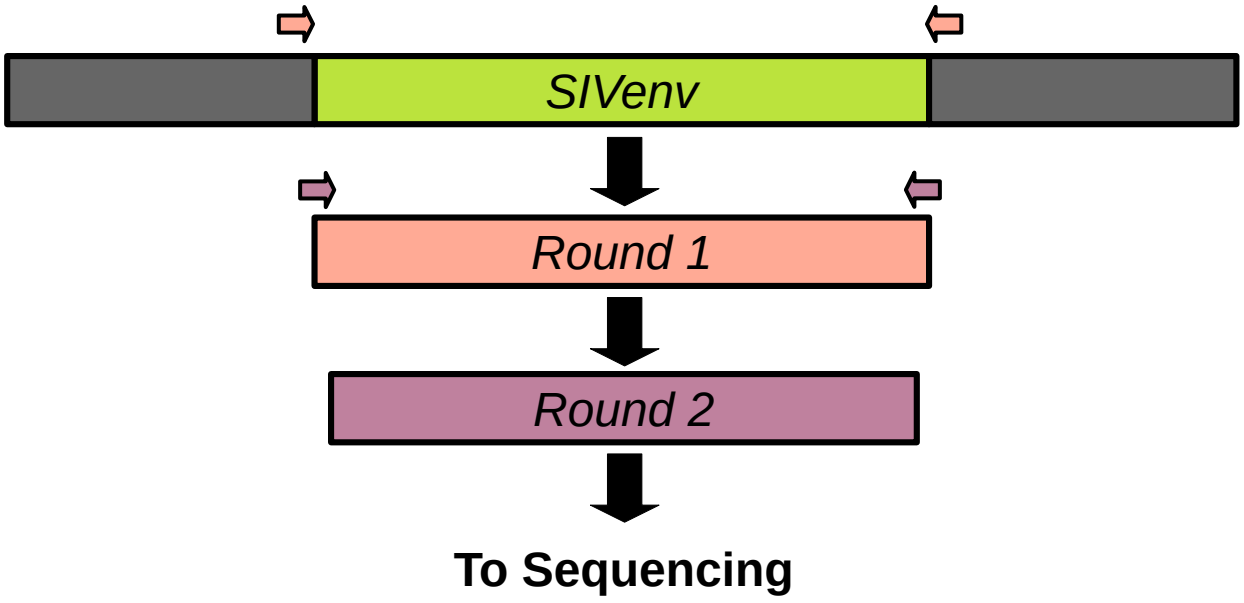


Figure 9: SIV Single-Genome Amplification Scheme: After reverse transcription of vRNA isolates from plasma samples, the *SIVenv* gene is PCR-amplified in the first round using primers depicted here in salmon-colored arrows. The PCR product from the first round is used as template in the second round using primers depicted in plum-colored arrows. The final product after second-round PCR would then be sequenced using the SMRT-Link™ sequencing platform by Pacific Biosciences.

Phylogenetic Analysis and Statistics

HIVenv sequences for T/F viruses from patients infected with HIV-1 subtypes B (AD17, CH040, CH058, CH077, CH106, CH470, CH607, REJO, RHPA, SUMA, THRO, TRJO, WITO) [Keele et al., 2008; Li et al., [2010]; Ochsenbauer et al., 2012; Liu et al., 2013] and C (703010131, 92BR025, 98IN012, ETH2220, CH067, CH185, ZM180M, ZM184F, ZM235F, ZM246F, ZM247F, ZM249M) [Salazar et al., 2008; Ping et al., 2013; Asmal et al., 2015; Rademeyer et al., 2016] were obtained from GenBank (accession numbers are shown in **APPENDIX 1** and a phylogenetic tree of the Env protein for these T/Fs are shown in **APPENDIX 2**).

HIV-1 paired-end amplicon sequences in .fastq format from the MiSeq™ system (Illumina®) were merged based on QC scores using FastP (v0.20.1). Quality control assessment before and after merging was performed using FastQC (v0.11.9). Sequence reads in .sam format from SMRTLink™ platform (Pacific Biosciences) were collapsed by consensus using the seqinr package in R (v4.1.2). Sequence alignments in .fasta format were performed using the ClustalX algorithm via the mafft alignment software (v7.490) and CLC Genomics Workbench (v19). Phylogenetic trees were constructed using FastTree (v2.1.11), MEGA-11 (v11.0.13), and CLC Genomics Workbench (v19). Phylogenetic trees, metadata, sliding window analysis, and statistical tests were performed using base R, ape, apTreeshape, Biostrings, rbiom, seqinr, dplyr, unikn, and stringr packages in R (v4.1.2). Blast alignments for sequence comparisons and searches were performed with blastp (v2.12.0+) [Altschul et al., 1997].

Chapter 3: Findings and Speculations—Transcytosis

Transcytosis through a Model Epithelium

In vitro and *Ex vivo* Models of HIV-1 Transcytosis

The *in vitro* transcytosis assay was performed as shown in **FIGURE 3**. In eleven separate experiments, a challenge dose of 92US657 equivalent to 4 ng of p24 was added to the apical side of an endometrial (HEC1A) model epithelium grown on a transwell insert to a TEER between 12 ± 2 $\Omega \cdot cm^2$ and 38 ± 4 $\Omega \cdot cm^2$. Transcytosis was allowed to occur between 6 and 12 hours across these experiments. Despite similar TEERs, challenge titers, and inoculation times reported in literature using the same cell line [Bomsel et al., 1998; Hocini et al., 2001; Saïdi et al., 2007; Kinlock et al., 2014], *HIVgag* vRNA was not detected in the basolateral chamber by RT-qPCR in any of the experiments (data not shown). During further optimization experiments wherein transcytosis was tested with varying titer doses, varying TEERs, and varying inoculation times, similar null results were found (data not shown).

In one optimization trial, however, artificial levels of transcytosis were found over an inoculation period of 35 hours with a challenge viral titer equivalent to 4 ng of p24 on the apical side of cells grown after 4, 6, 8, 10 and 12 days (**FIGURE 10**). Transcytosis levels across an intact epithelium exhibiting TEERs between 15 $\Omega \cdot cm^2$ and 40 $\Omega \cdot cm^2$ *in vitro* resulted in a range of 0.05% to 2% of the challenge dose in previously published studies [Bomsel et al., 1998; Hocini et al., 2001; Saïdi et al., 2007; Kinlock et al., 2014]. Based on data from the optimization experiment, however, to achieve this range in transcytosis efficiency with 4 ng-equivalents of p24, intact model epithelia had to be grown on transwell inserts with a pore size of 3 μm over an inoculation time of 35 hours (**FIGURE 11**). Repetition of this assay with smaller pore sizes

resulted in much lower transcytosis efficiencies (data not shown), contrary to the findings of Bomsel [1997]. Typical inoculation times as reported in literature ranged from 1 hour [Saïdi et al., 2007] to 3 hours [Hocini et al., 2001], with 12 hours being the longest reported time [Gupta et al., 2013], versus the 35 hours necessary to achieve similar levels of transcytosis compared to those reports. Two studies stand out as reporting widely different transcytosis efficiencies as observed elsewhere in literature [Bomsel, 1997; Day et al., 2022], though these studies tested alternative mechanisms of HIV-1 transmission that cannot be compared to cell-free viral transcytosis.

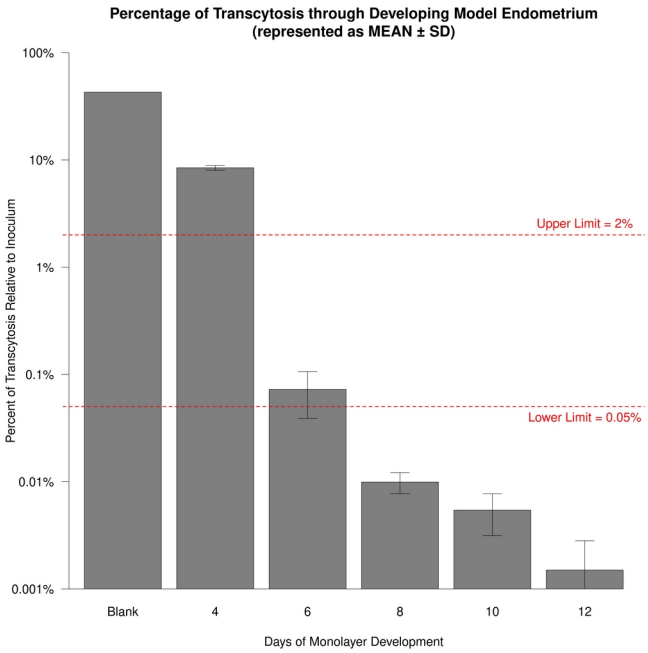


Figure 10: Percentage of Transcytosis through a developing Epithelial Layer of HEC1A Cells: *In vitro* transcytosis was performed through a model endometrium layer using HEC1A cells formed over 4, 6, 8, 10, and 12 days. Successful transcytosis is defined as the passage of virus 0.05% to 2% of virus from the apical side to the basolateral side of the trans-well insert, based on published data.

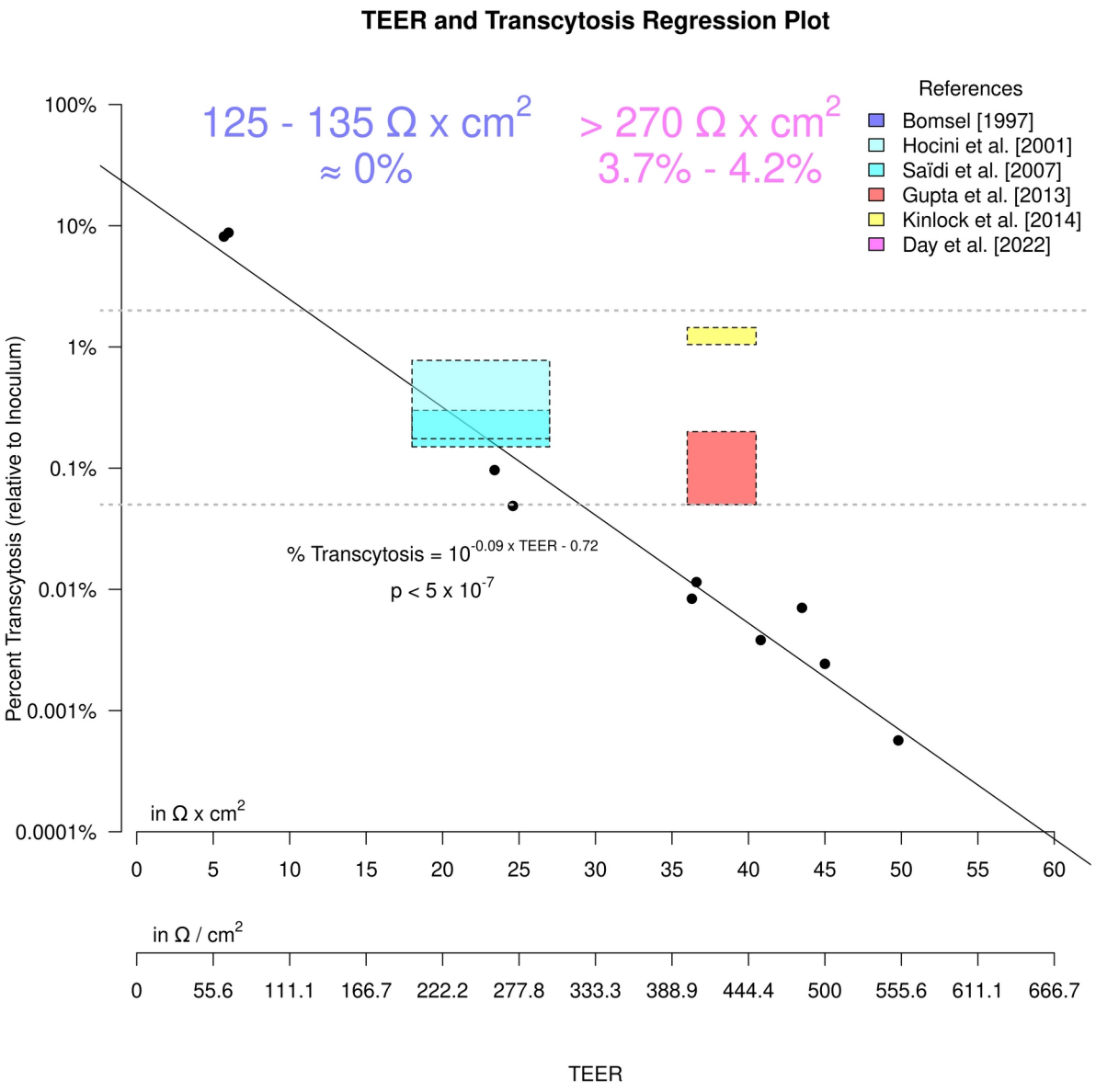


Figure 11: Regression Plot of Transcytosis Efficiency with Trans-epithelial electrical resistance (TEER): Transcytosis was performed across trans-well inserts with 3-µm pore sizes and a model epithelium developed to varying TEERs (shown here in two different units) over the period of 35 hours. Also shown here are ranges of TEERs crossed with percentages of transcytosis for 6 different publications on HIV transcytosis. Also shown is the range of successful transcytosis efficiencies from 0.05% to 2 % by the dotted line. Transcytosis efficiencies reported by Bomsel [1997] and Day et al. [2022] are outliers here for the reason that they tested cell-assisted and complement-mediated transfer of virus across an extremely tight epithelium, respectively. Test of regression: t-statistic.

While a significant regression trend was observed between TEERs and transcytosis efficiency (t-statistic: $p < 5 \times 10^{-7}$) (**FIGURE 11**), there remains considerable doubt as to the relevance of these data. Transcytosis here is likely an artefact of atypically long inoculation times beyond which is necessary for HIV-1 to infect an individual through mucosal transcytosis or cell-mediated transmission. It is unknown whether the tested conditions here are physiologically analogous to mucosal barriers in living persons, even if they resulted in a bottleneck. Therefore viruses that were artificially transcytosed in this manner were not processed for sequencing and phylogenetic analysis as results may be influenced considerably by confounding factors.

In order to determine whether transcytosis failure was due to instrument error, the optimization experiment was repeated using inserts paired by TEER. One of these pairs was challenged with virus titered at 4 ng p24 while the other was challenged with Lucifer Yellow, a fluorescent dye used to measure the intactness of a model epithelium. Passage of Lucifer Yellow from the apical side to the basolateral side of a model epithelium is a well-accepted test for non-transcytotic epithelial leakage *in vitro* given that such passage does not involve endocytosis and exocytosis [Odriscoll et al., 1991]. Leakage, as such, would be a proxy for a non-intact cellular monolayer, which could artificially impact the transcytosis efficiency and selectivity. In this experiment, monolayers exhibited 2.5% ± 0.3% leakage, comparable to those reported in literature (1.5% by Hocini and Bomsel [1999] and 2.7% by Day et al. [2022]). One outlier in comparisons of other studies to this one is in data reported by Bomsel [1997]. The author reported a leakage of 0.24% ± 0.04% using radiolabeled Inulin as a reporter, but this is likely due to the formation of a super-intact epithelium designed to negate cell-free transcytosis (which

explains the ~ 0% transcytosis rate found in that study). According to principal component analysis shown in **FIGURE 12**, there is a strong negative correlation between transcytosis efficiency and TEER, as expected despite transcytosis failures encountered in these trials. However, the data indicated that leakage (the percentage of Lucifer Yellow that pass through junctions between cells in a monolayer) was not related to the measured TEER in the assayed monolayers, as suggested by the pendicular relationship between these two datasets in the analysis. A further note of interest is the negative correlation between the leakage and the transcytosis assays, which was unexpected. It suggests that more efficient transcytosis was associated with a more intact membrane. Though the failures of the transcytosis assays are still under investigation, these data may indicate that transcytosis depends on penetration through cells rather than bypassing the epithelium entirely.

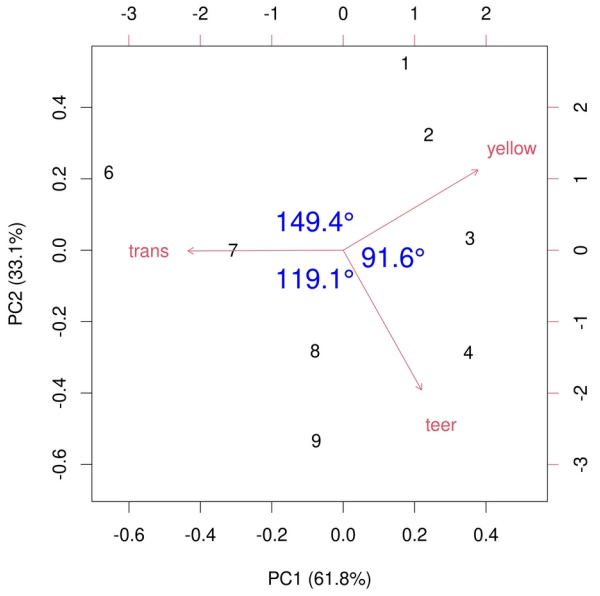


Figure 12: Principal Component Analysis of Transcytosis Assay with TEER and Monolayer Leakage Data: Data was scaled to a 0 – 1 reference, then analyzed through principal component analysis to determine correlates for reporters for membrane leakiness and transcytosis efficiencies.

These investigations included an *ex vivo* study using a colonic mucosal explant that unfortunately also did not result in observable transcytosis (data not shown) but this could be due to improper assembly of the excised tissue in the transwell. The *ex vivo* and *in vitro* studies have since not been repeated in order to focus on a demonstrably working alternative model of transcytosis.

In vivo Model of SIV Transcytosis

A "Living" Transcytosis Assay

Given failures in previous experiments, there was no direct way to test Hypothesis #1 or simulate transmission across mucosal barriers *in vitro* or *ex vivo*. There was, however, still a means of investigating how transcytosis results in restrictive bottleneck in favor of T/F quasispecies and whether immunologically important molecules influence that selection. Previously published reports on SIV/SHIV infection assays *in vivo* showed that a genetic bottleneck, infection of an animal by a single T/F variant, is a consistent outcome of transmission across rectal [Keele et al., 2009; Liu et al., 2010; Yuan et al., 2017; Chen et al., 2018], vaginal [Stone et al., 2010], and penile [Ma et al., 2011; Qureshi et al., 2012] mucosae. This suggests that interaction between the host mucosa and the viruses penetrating it, whether through cell-free transcytosis or cell-assisted transmission, could be an important selective factor for T/F strains. This is especially the case when bypassing mucosal barriers resulted in productive infection by numerous, phylogenetically disparate T/Fs [Keele et al., 2009]. The greater homogeneity of quasispecies infecting animals via the intrarectal route compared to the intravenous route strongly suggests that there exists some factor(s) within or among the rectal mucosa crucially

important for influencing and restricting the transmission of select viral strains from a diverse viral swarm [Keele et al., 2009].

Our lab participated in a penile inoculation study using *R. macaques* in collaboration with Dr. Christopher Miller at the University of California, Davis in order to replicate the results of Gupta et al. [2013] *in vivo*. While the study itself was not crucial to the testing of the current hypotheses, plasma collected from infected animals could be used to determine how virus-specific IgG antibodies and changes in pH (both components of genital fluids) modulate the selection of T/Fs through an intact mucosal barrier. Even though antibody opsonization of an SIVmac251 swarm reduced the infectiousness of viruses and resulted in lower viremia in contrast to the researchers' predictions (data not shown), it was not known whether infection by antibody-opsonized virus resulted in selection of a particular phenotype of T/Fs compared to non-opsonized virus.

The phylogenetic relationship of sequenced variants from each productively infected animal is represented in **FIGURE 13**, showing also the reference sequence for the viral stock (SIVmac251.CM.p3a6; accession KC522220) used in the study [Del Prete et al., 2013]. The same phylogenetic tree depicting the relationship among all the inoculated animals are also depicted according to their treatment groups (**FIGURE 14**). The viral variants sequenced from each animal clustered with particular lineages in the inoculum, which demonstrates that infections largely originated from the inoculum itself and not from another source. None of the sequences exhibited significant APOBEC-3G hypermutation signatures according to the hiv.lanl.gov hypermut tool [Rose and Korber, 2000] (data not shown). One sequence (CV_56) did return a p-value nearly equal to 0.05 from the analysis, however this was not removed from

the pool as it was part of the inoculum and its hypermutated sites could have contributed to bottleneck restriction.

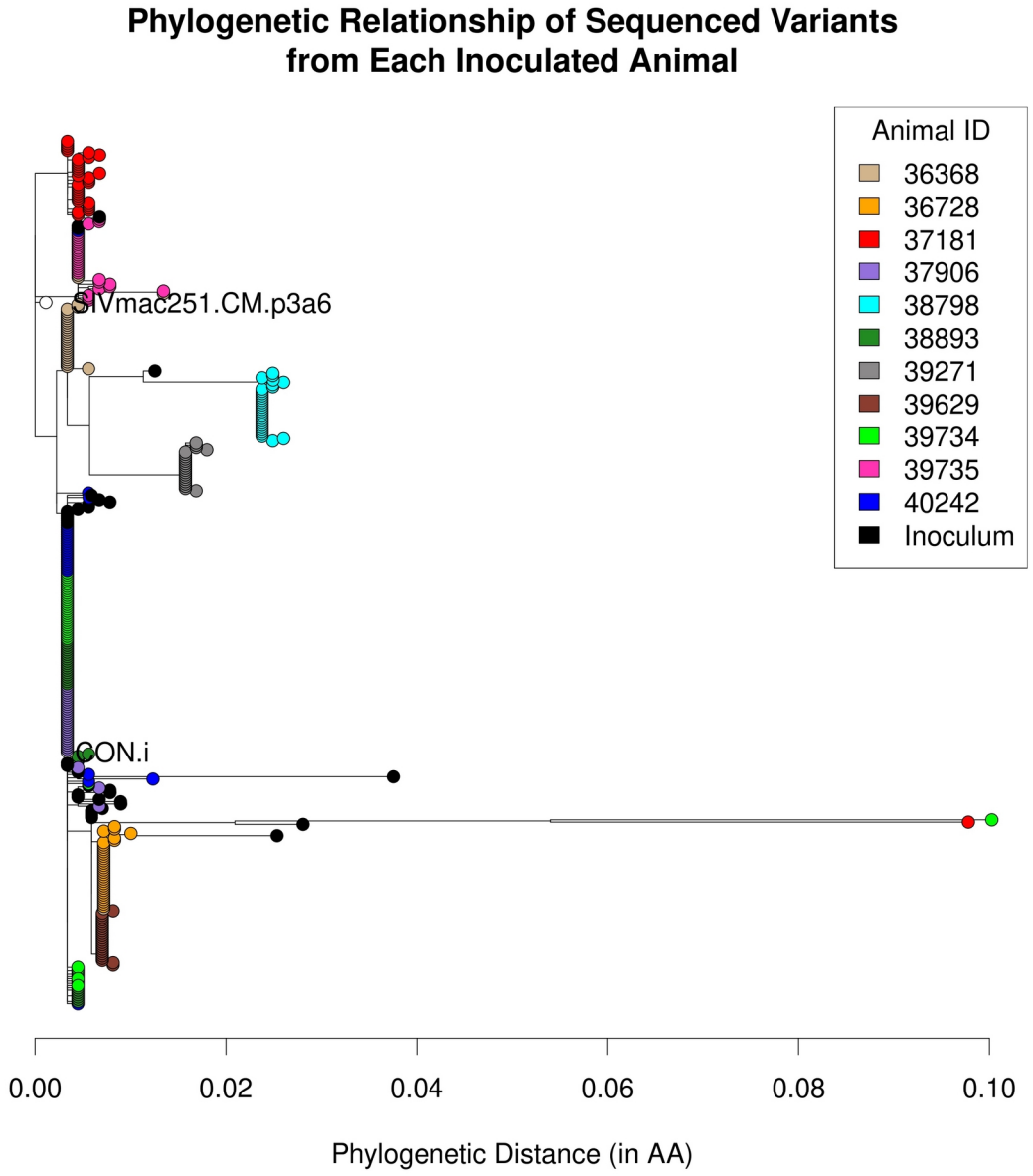


Figure 13: Phylogenetic Tree of SGA amplicons from each Animal (tips labeled by animal): Viruses were lysed and viral RNA was isolated from plasma samples of infected animals during peak viremia. After reverse-transcription of viral genomes, the *SIVenv* gene was PCR-amplified through SGA and sequenced using the SMRTLink™ platform by Pacific Biosciences. Reads were collapsed to their respective consensus and aligned by mafft and displayed in this Maximum Likelihood tree by FastTree. Viral sequences are labeled according to animal ID.

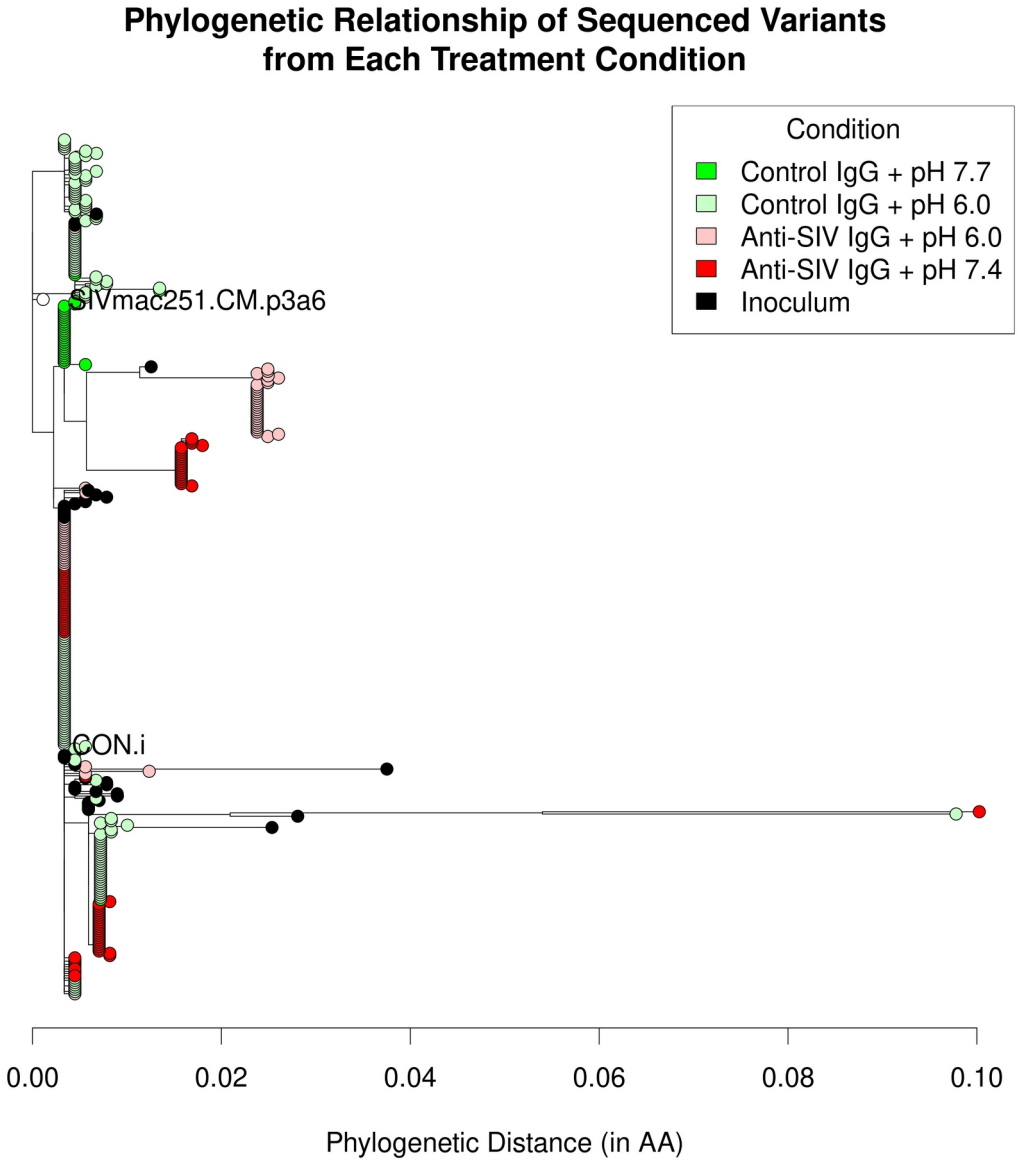


Figure 14: Phylogenetic Tree of SGA amplicons from each Animal (tips labeled by treatment): Viruses were lysed and viral RNA was isolated from plasma samples of infected animals during peak viremia. After reverse-transcription of viral genomes, the *SIVenv* gene was PCR-amplified through SGA and sequenced using the SMRTLink™ platform by Pacific Biosciences. Reads were collapsed to their respective consensus and aligned by mafft and displayed in this Maximum Likelihood tree by FastTree. Viral sequences are labeled according to treatment condition.

Weak selection pressures are inferred by stochastic transmission and by the animal-specific (rather than treatment-specific) nature of T/F quasispecies

All eleven animals were predominantly infected by a single T/F virus (the phylogenetic relationship of which is shown in **FIGURE 15**), although an additional one to four lineages were represented by a minority of variants that were not phylogenetically related to major T/Fs in 5 animals (36368, 37906, 39271, 3883, and 40242). These variants could either be established T/Fs that co-transmitted with the main T/F in each animal or they are short-lived or transient variants sampled at a time before full bottleneck restriction had been achieved. Yuan et al. [2017] have detected minor, short-lived, and likely defective variants from experimentally infected animals, and the possibly transient variants found in this study may be a replication of the finding from the 2017 study. The poorly understood phenomenon of quasispecies transience, wherein a minor co-transmitted viral variant persisted for only a short time before becoming phased out, had been observed in a prior study involving plasma and genital fluids from female subjects recently infected with HIV [Piantodosi et al., 2019]. In the 2019 study, the percentage of single nucleotide variants (individual viral variants that differed from the major T/F by a single nucleotide change) in the plasma dropped during the acute infection period from 7 dpi to 11 dpi, and the highest measured percentage of any particular variant was 6%. Another study prior to this has reported instances in clinical HIV-1 infection of minor variants among multiple T/F viruses within a single subject disappearing after 20 days of infection [Kijak et al., 2017]. A minority of sequenced variants from some of the singly infected animals in the current study do cluster with different lineages in the inoculum at the DNA level other than the majority strain in the host (presumed to be the lineage established by the main T/F virus in that animal) (data not

shown), suggesting further that these underrepresented viruses may have transmitted non-selectively regardless of their contribution to the viremic load.

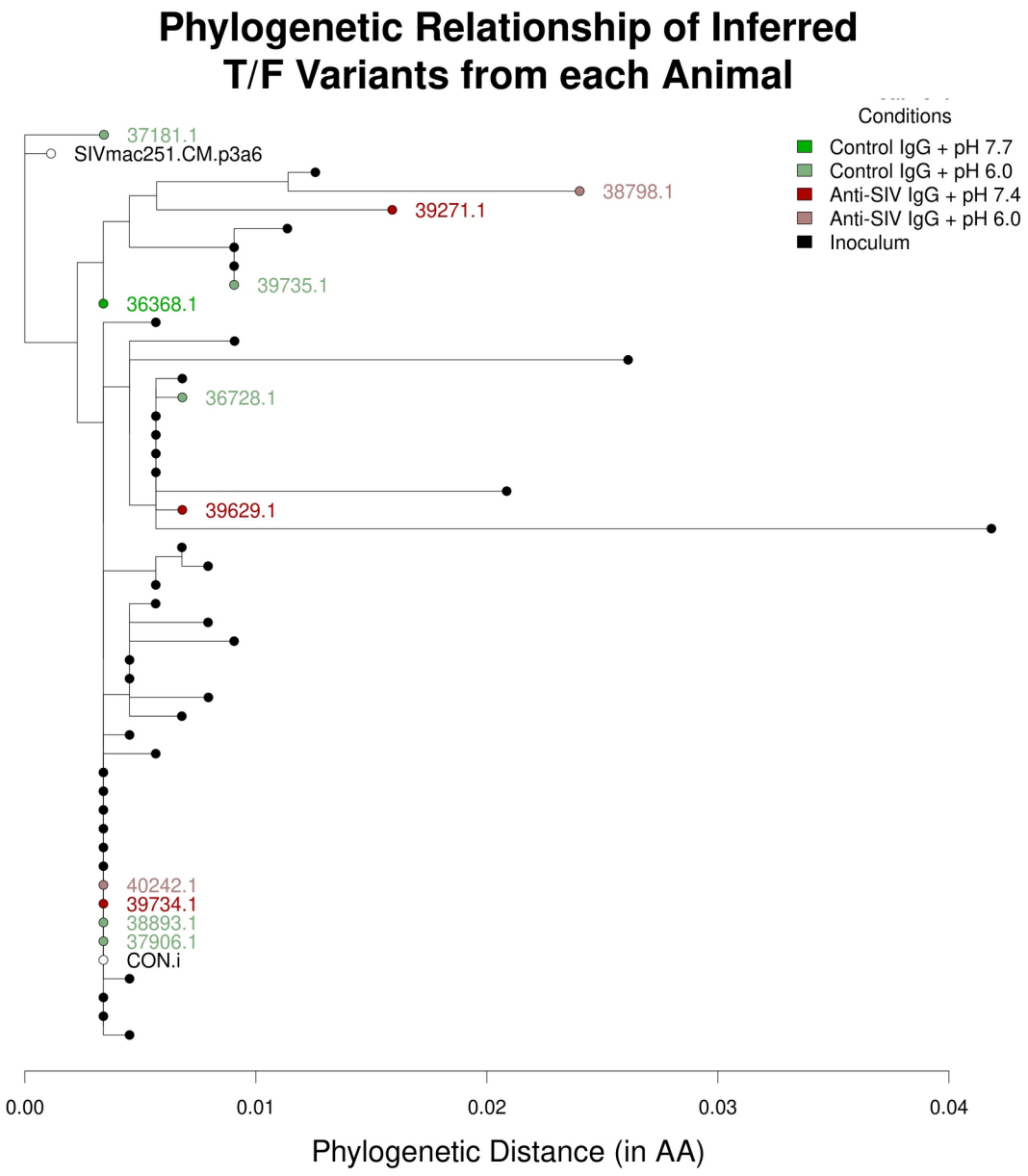


Figure 15: Phylogenetic Tree of T/F viruses inferred from each inoculated animal (tips labeled by treatment): Sequences for T/F viruses were inferred by consensus and aligned by mafft, then displayed in this Maximum Likelihood tree by FastTree. Viral sequences are labeled according to treatment condition.

Of the 5 animals, animal 40242 contained viruses from the highest number of disparate lineages (**FIGURE 16A, APPENDIX 3**) with 36368 containing the second highest (**FIGURE 16B**). The presentation of minor variants extraneous to the major T/F lineage is not likely the result of random mutation, as they originated from the inoculum itself, evident by co-clustering of minor variants with discrete variants found in the inoculum . One of the variants from animal 40242, as well as another from 36368, cluster with the T/F variant infecting 39735, which casts some doubt as to whether the diversity of variants in these four animals was due to infection by multiple selected T/F variants or infection by less fit, transient variants. In fact, the genetic diversity among viral variants from animal 37181 was among the highest observed in this study, yet the tree depicting the phylogenetic relationship among these diverse viruses indicate they originated from a single T/F lineage (**FIGURE 16C**), in as much as variants from animal 38798, which exhibited far lower diversity, also originated from a single T/F virus (**FIGURE 16D**). The percentages of sequenced variants that did not coalesce with the major T/F in each animal was around 13.8% or lower (**FIGURE 17**), which was lower than percentages found in studies involving rectal [Keele et al., 2009] and vaginal [Stone et al., 2010] inoculation but still higher than those found in studies involving penile inoculation [Ma et al., 2011]. It is still unknown whether these are transient quasispecies or additional productively replicating T/Fs. Their presence in the animals may not have been the result of selective transmission given their low representation in the viremic pool of the particular animals in question and thus were eliminated from further analysis to avoid a confounder effect.

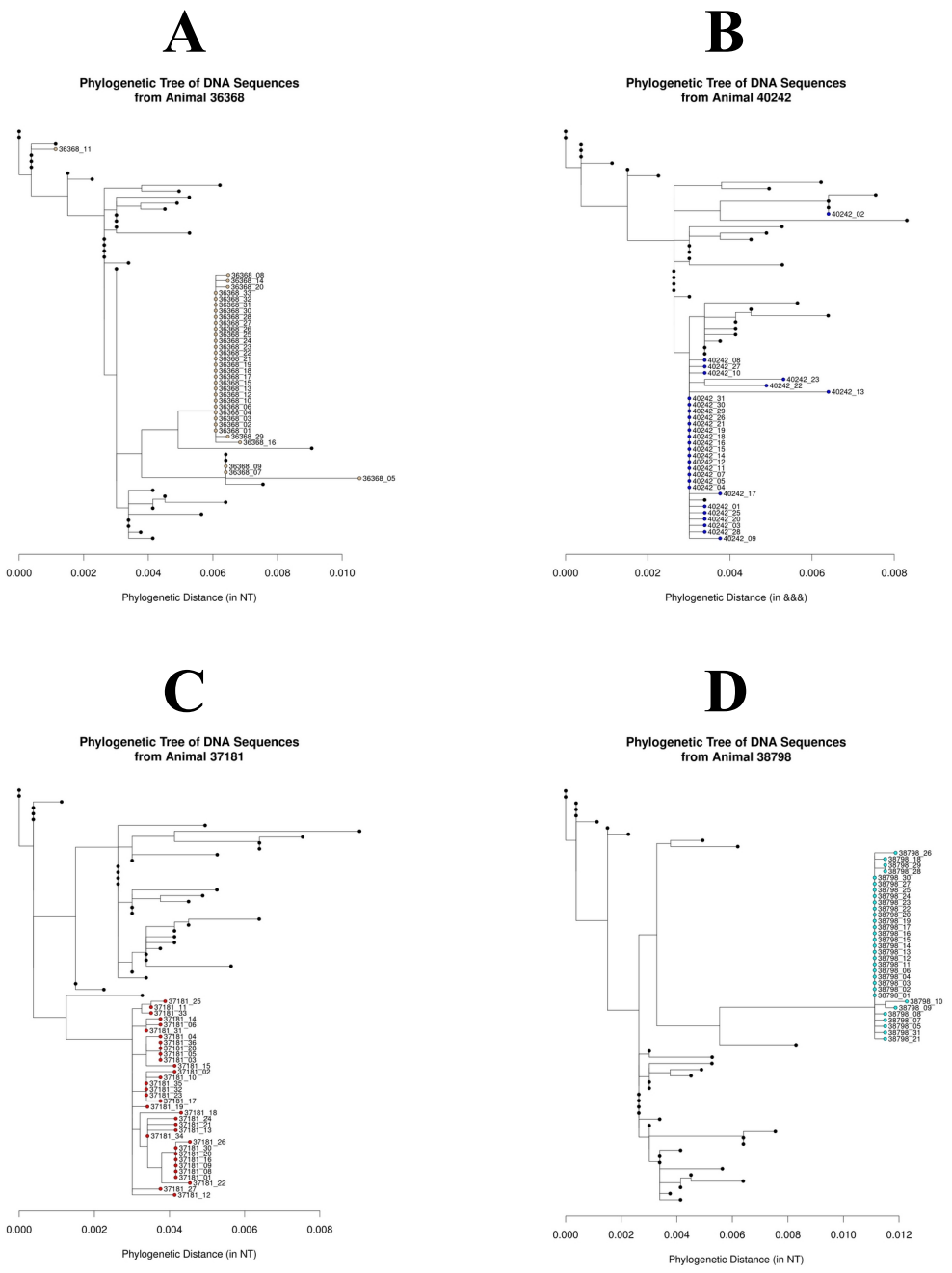


Figure 16: Phylogenetic Trees of four animals with and without suspect transient strains: Maximum likelihood trees for *SIVenv* DNA sequences (aligned by mafft and constructed by FastTree) are shown for animals showing evidence of several suspect transient quasispecies (A and B), an animal showing evidence of infection by a single T/F with high genetic divergence (C), and an animal showing evidence of infection by a single T/F with low genetic divergence (D).

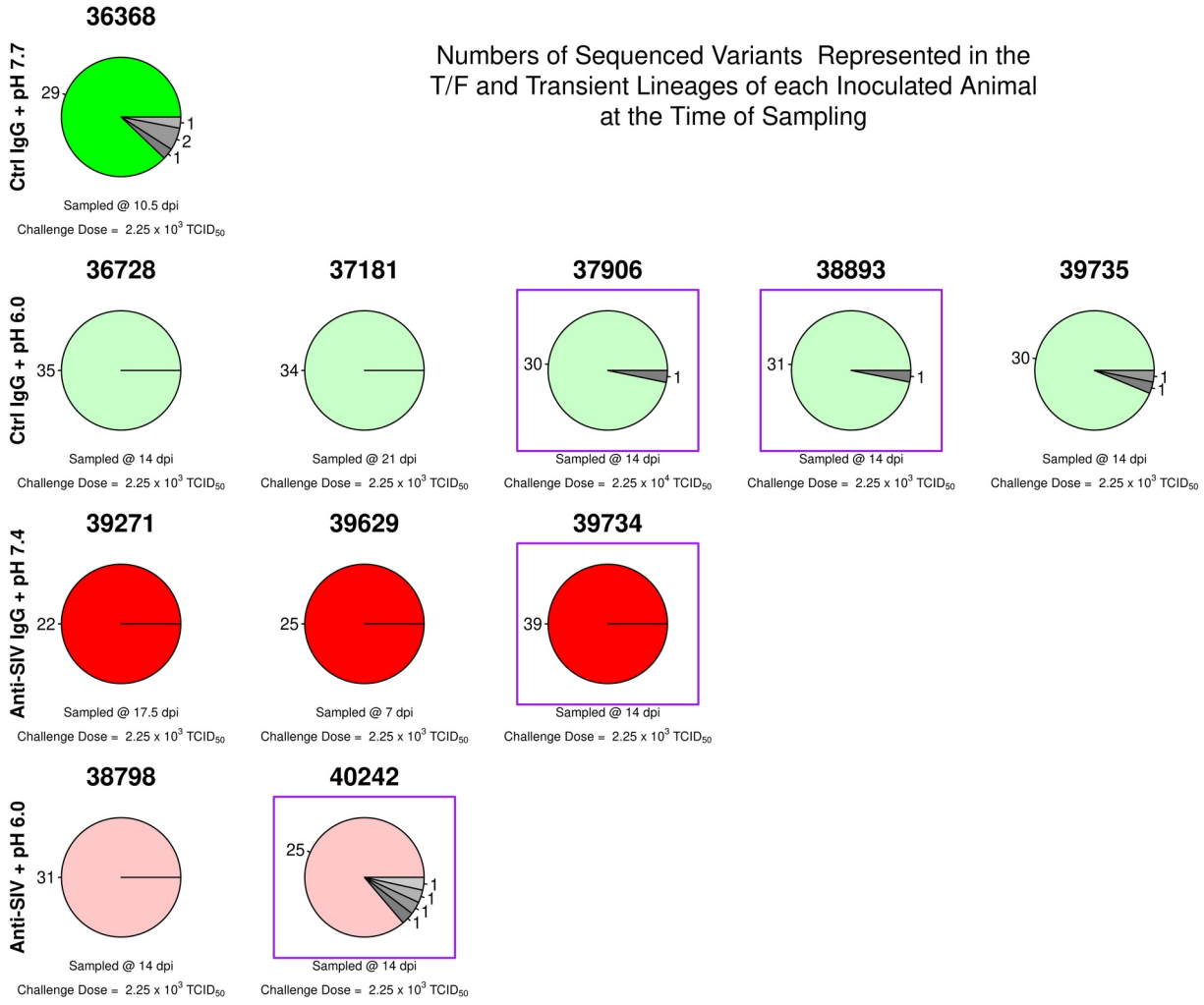


Figure 17: Numbers of variants represented in T/F and transient lineages of each inoculated animal. Pie charts represent the distribution of the numbers of sequence reads in discrete lineages for each animal. Colors indicate treatment conditions (shown to the left). Pie charts in purple squares indicate animals that were stochastically infected by the predominant strain from the inoculum (CON.i). Data below each pie chart shows the sample date (in days post infection) for the sequenced SGA products as well as the challenge dose (in $TCID_{50}$) necessary to achieve productive infection through the penile route.

The consensus sequences for the variants infecting each animal (inferred to be the selected T/F in each animal) are unique and specific to seven of the animals, which corroborates the finding by Yuan et al. [2017] transmission of T/F variants is animal-specific. However, four

of the animals (37906, 38893, 39734, and 40242) share identical T/Fs that cluster with the majority strain in the inoculum (the inoculum consensus, labeled as "CON.i" in some phylogenetic trees) (**FIGURE 15** and **FIGURE 18**). These animals could have been infected through a common selection pattern, but this is unlikely as co-clustering of these strains with the inoculum consensus implies stochastic transmission rather than selective transmission. The co-clustering of T/F strains with the inoculum consensus confounds downstream analyses on selected polymorphisms in that they are compared to animals infected by viruses from minor lineages. In order to characterize any antibody-influenced selection pattern, animals infected stochastically were removed from further analysis unless specified. It should be noted, however, that antibody- or control-mediated selection pressure may not have been strong enough to favor specific T/Fs during transmission, hence the stochastic transmission of a predominant inoculum strain rather than a minor one in some of the animals. Therefore, elimination of these outlier sequences enables characterization of the selection directionality (i.e. whether a particular polymorphism generally favored or disfavored by antibody treatment) rather than the selection strength itself (i.e. the degree of bias toward a particular polymorphism favored or disfavored by antibody treatment).

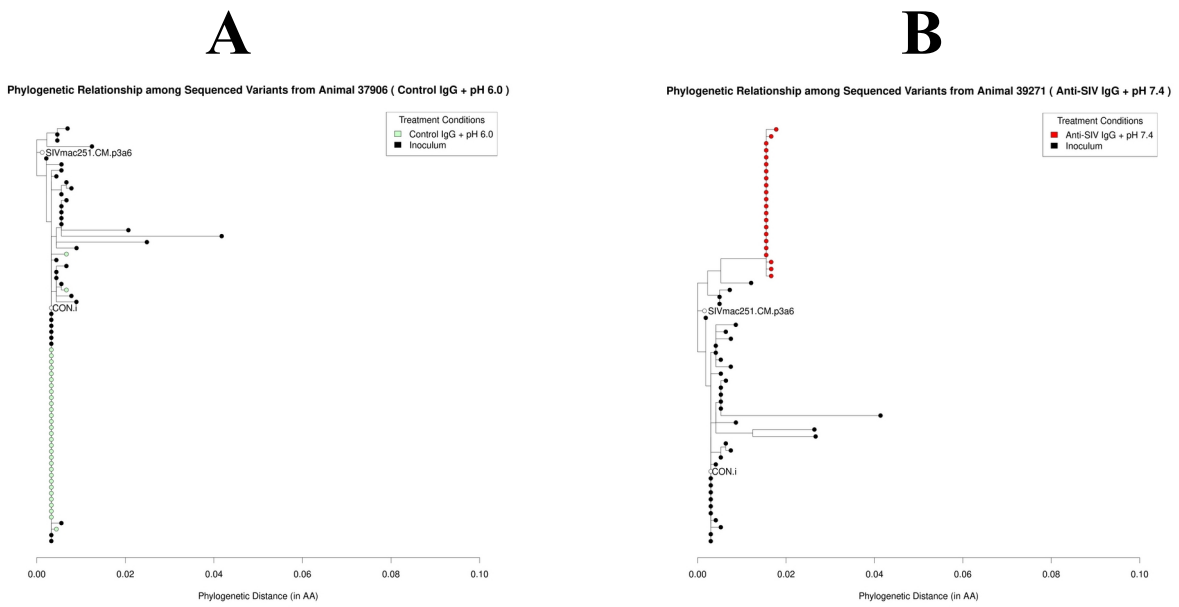


Figure 18: Sample Phylogenetic Trees of stochastic and selective transmission. Maximum-likelihood trees were constructed with FastTree and shown here for an example of an animal infected stochastically (A) and selectively (B) from a diverse viral swarm.

Identification of Potentially Selected Polymorphisms

Regarding the question of antibody-mediated selection, there does not appear to be an obvious pattern of selection merely on inspection of the phylogenetic trees representing relationships among output viruses from animals infected with opsonized or non-opsonized virus (**FIGURE 19A**). Furthermore, pH of the inoculating medium in combination with or without SIV-specific IgG medium does not appear to foment observable patterns in selection either (**FIGURE 19B**). There was also no obvious pattern of treatment-specific representation among the variants from each animal or the T/Fs infecting them specifically (**FIGURE 14**), which may indicate that either selection does not exist or, if it does, may have been too weak to be captured

by the *in vivo* model. This is a likely conclusion given that nearly 40% of the animals were infected stochastically.

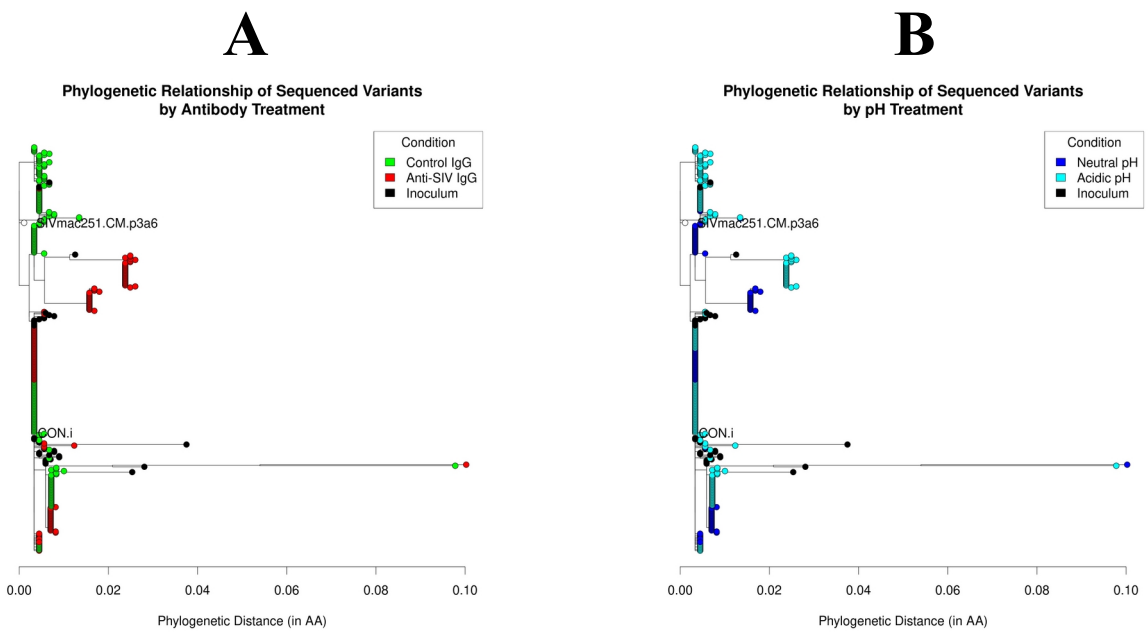


Figure 19: Global phylogenetic trees by antibody and pH conditions separately. Maximum-likelihood trees were constructed with FastTree for all sequenced SGA products. Tips are colored according to antibody treatment (A) and pH treatment (B).

Next, sliding window analysis was performed for all SIV ENV protein sequences from each animal and inoculum residue by residue (after removal of transient and stochastic quasispecies), comparing the distribution of polymorphisms at each site across the control (CTRL), antibody-treated (ANTI), and inoculum (INOC) variants. The threshold cutoff for potential selected polymorphisms at any site along the sliding window was arbitrarily set to 90% identity with the inoculum consensus. By this point, statistical tests were performed to ascertain whether a selection pattern exists at the given amino acid site for a given treatment condition. This analysis revealed thirty-five polymorphisms that were favored (or disfavored) to some

extent during transmission (**FIGURE 20**). Twenty-three of these polymorphisms had higher representation among sequences from antibody-opsonized viruses, ten had higher representation among non-opsonized viruses, and two had higher representation among viruses in the inoculum. **FIGURE 21** shows the p-values (with and without Bonferroni correction [Simes, 1986]) for Fisher's exact test comparisons between the antibody, control, and inoculum groups as well as the experimental group in which the polymorphism had the highest representation. The antibody-treated group generally favored mutations in the gp120 domain, particularly in variable loop 4 (V4), and in the gp41 domain from amino acid positions 697 to 745. However, polymorphisms from positions 763 to 863 were more favored by the control than by antibody treatment. These polymorphisms may not be relevant for infection and replication as they lie in the intergenic region between gp41 and Nef.

Not all polymorphisms were equally represented within treatment groups, let alone among the animals. The A567T polymorphism had much higher representation in the control group than the inoculum or the antibody group. However, all variants with this polymorphism clustered to a single animal (37181), suggesting it may not be as important for selection despite a significant difference in the distribution of this polymorphism in the control group compared to the antibody-group ($p < 2 \times 10^{-5}$, Fisher's exact test with Bonferroni correction). Conversely, the T131_T133del mutation was favored in the absence of antibody, but it was present in only two animals (animals 39629 from the antibody group and 36368 from the control group) and its distribution in the control was not significantly different from that in the antibody group. A similar effect was observed among many of the selected polymorphisms identified by sliding

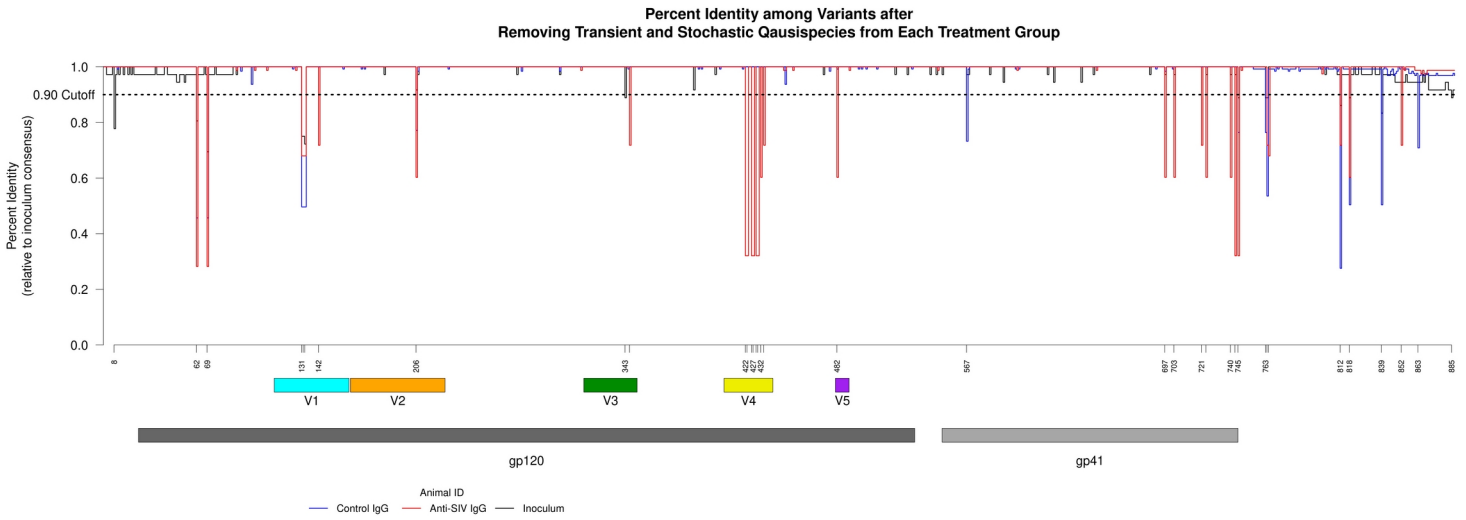


FIGURE 20: Sliding window analysis of amino acid identity (relative to inoculum consensus) of sequenced variants by antibody treatment. Sequence reads that were the result of suspect transient quasispecies (i.e. variants with extremely low representation in a particular animal plasma sample and were not part of the main infecting T/F lineage for that animal) were removed; the entire sequence set from animals infected stochastically were also removed in order to more precisely gauge the outcome of weak selection pressures for the observed bottlenecks on polymorphism selection. The above plot shows a visual depiction of the sliding window analysis used to identify potentially selected polymorphisms by changes in the percent representation of the inoculum consensus amino acid identity at each position along a protein alignment. 90% identity with the inoculum consensus was chosen as an arbitrary cutoff to perform Fisher's exact tests of the distribution of polymorphisms at each position by treatment condition.

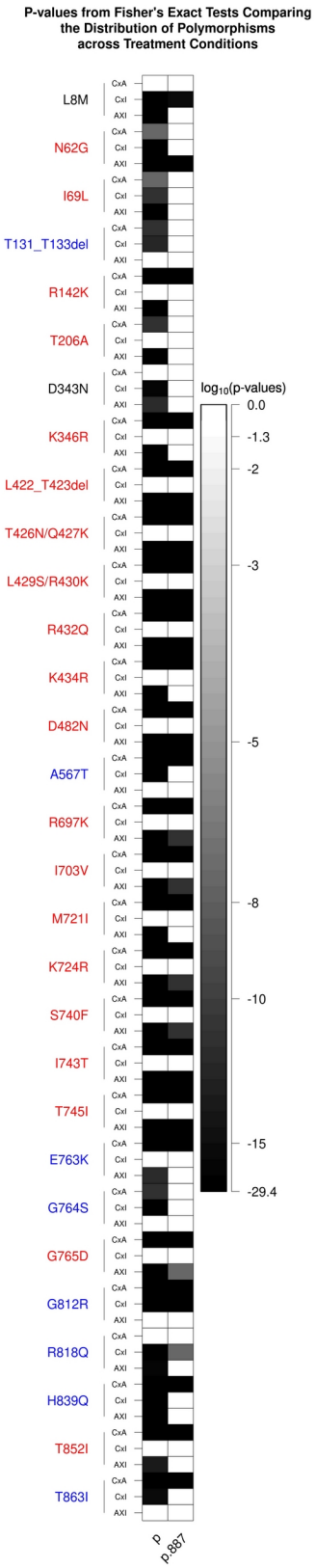


Figure 21: Fisher's exact test results (with and without Bonferroni correction) for potentially selected polymorphisms. Potentially selected polymorphisms at amino acid positions where the inoculum consensus was represented at less than 90% were evaluated for selection using Fisher's exact test, with and without Bonferroni correction. Polymorphisms in red were selected in animals infected with antibody-coated inoculum while those in blue were selected in animals with naked inoculum. The letter code to the right of each polymorphisms represent the type of comparisons being made: CxA (polymorphism distributions between the control and antibody treatments), CxI (polymorphism distributions between the control and inoculum), and AxI (polymorphism distributions between the antibody treatments and inoculum). P-values for Fisher's exact tests are shown in two columns (without Bonferroni correction in the first column and with Bonferroni correction in the second) as shades of white to black along a sliding scale represented to the right.

window analysis, and these could be non-selected, linked mutations. Such mutations happen to have very high representation in a single animal but are not present among any other animal in a given experimental group, raising doubts as to their relevance in antibody-mediated selection. Any apparent dominance of particular polymorphisms in any of the groups could be artificially influenced merely by their genetic linkage to selected T/F polymorphisms.

Refinement of sliding window analysis toward T/F sequences rather than bulk sequences reduces the bias of overrepresented polymorphisms in the search for truly selected polymorphisms

Per the likelihood of linkage artefacts, sliding window analysis was repeated using only the consensus T/F sequences infecting each animal. Performing the analysis in this manner (comparing individual T/Fs rather than bulk viral variants from each animal) effectively under-represents the distribution of polymorphisms in each treatment group, but the statistical advantage here is that identification of truly selected polymorphisms would be less biased toward non-selected, genetically linked polymorphisms. Stochastic T/Fs were re-included in this analysis in order to factor the strength of selection toward polymorphisms among all discovered T/Fs and to remain conservative in identifications. Sliding window analysis, shown in **FIGURE 22**, revealed largely the same polymorphism hits as the bulk analysis did, except that percent identities at each of the candidate sites relative to the inoculum consensus was generally higher than those in the bulk analysis. However, Fisher's exact test showed statistically significant differences in the representation of only three polymorphisms without applying Bonferroni corrections. The first polymorphism was L422_T423del/T426N/Q427K/L429S/R430K

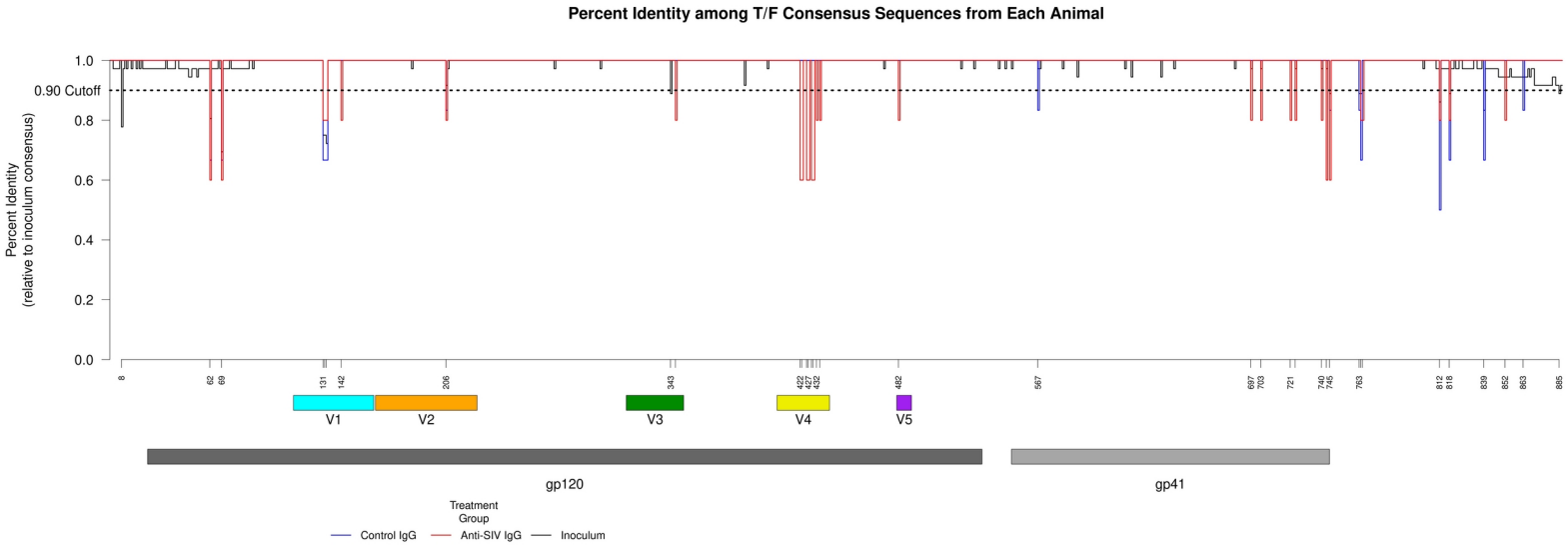


Figure 22: Sliding window analysis of amino acid identity (relative to inoculum consensus) of inferred T/F sequences. Analysis performed and depicted here as described in **Figure 20**. Here, main difference was that comparisons were made among inferred T/F quasispecies, including the T/Fs of stochastically infected animals. The purpose of the sliding window analysis of T/F sequences was to identify potentially selected polymorphisms specific to the major T/Fs without the confounder effect of random mutations due to a star-like phylogeny.

(NKSK), which was diverged from the consensus of L422/T423/T426/Q427/L429/R430 (LTTQLR), and had higher representation in the group inoculated with antibody-opsonized virus compared to the inoculum ($p = 0.01$, Fisher's exact test without Bonferroni correction). This region in question is part of V4 of SIV gp120. A prior study has shown that ablation of this variable loop impairs HIV-1 infectivity and viral assembly and can potentially expose gp41 for antibody-mediated fusion inhibition [Yuan et al., 2013]. While not much else is known regarding V4, it can be surmised that it plays an important role in shielding gp41 from potentially neutralizing antibodies.

The second was I743T and it also had higher representation in the group inoculated with antibody-opsonized virus compared to inoculum ($p = 0.03$, Fisher's exact test without Bonferroni correction), and it is located in the C-terminal end of gp41. The C-terminal (cytoplasmic tail) of gp41 is important for gp120 viral membrane and matrix (MA) protein anchor [Freed and Martin, 1996; Turner and Summers, 1999]. It is unclear how this particular amino acid (I743) or the antibody-selected mutation (I743T) influences the viral spike-MA anchor, but neither polymorphism appear to impede viral infectivity during the acute phase of infection.

The third polymorphism was G812R, which had higher representation in the control group compared to the inoculum although such difference was nearly significant ($p = 0.05$, Fisher's exact test without Bonferroni correction). It is located in the region between gp41 and Nef, and it is not likely incorporated into the mature virion. Whatever role it may play in infectivity, replicability, or other functions of SIV/HIV is not currently known.

Statistical measures of selection reveal NKSK was positively selected by antibody-opsonization

To determine whether the observed distributions of polymorphisms under particular treatment conditions were due to selection or chance, Ka/Ks (also known as dN/dS) ratios were calculated for genetic changes between treatment groups. The Ka/Ks ratio is a measure of the rate of non-synonymous mutations (Ka) over the rate of synonymous mutations (Ks) [Li et al., 1985; Nei and Gojobori, 1986; Li, 1993; Pamilo and Bianchi, 1993; Comeron, 1995; Ina, 1995], and it is an invaluable tool in determining modes of selection, whether positive or negative, influencing the evolution of genes and organisms [McDonald and Kreitman, 1991; Pamilo and Bianchi, 1993; Muse and Guat, 1994; Wang et al., 2011]. A Ka/Ks ratio > 1 results from a higher nonsynonymous substitution rate than a synonymous substitution rate, and is an indication of positive selection, whereas a Ka/Ks ratio < 1 results from a lower nonsynonymous substitution rate than a synonymous substitution rate and would indicate negative (or purifying) selection. Positive selection, in this sense, means that mutations for amino acid changes were favored under a particular environmental change (i.e. antibody-opsonization of SIV) and that those mutations likely enhanced the fitness of the infectious virus. On the other hand, negative selection occurs when amino acid changes result in lower fitness or a general detriment to viral infectivity and replication during infection under a particular treatment condition, thus such selection is considered purifying because it removes potentially harmful mutations from the progeny of a productively infectious virus. Whenever Ka/Ks = 1, there is no selection pressure influencing the accumulation of nonsynonymous mutations, and that any such accumulation are likely due to chance rather than selection.

A heatmap showing the Ka/Ks ratios for the entire *SIVenv* amplicon is displayed in **FIGURE 23** with Ka/Ks comparisons between treatment groups. Sequences from animals 37906, 38893, 39734, and 40242 were removed from this analysis in order to evaluate modes of selection without the outlier effect of stochastic transmission. SGA sequences are arranged in rows by hierarchical clustering of Ka/As ratios and in columns by treatment groups; the alignment between hierarchy and treatment would indicate correlation between treatment condition and a particular mode of selection. As expected for the *SIVenv* gene, mutations are generally deleterious as indicated by Ka/Ks ratios < 1 for the entire gene in the control, anti-SIV, and inoculum groups, and this is consistent with reports in literature that most mutations impair viral fitness especially during early acute infection [Zanini et al., 2015].

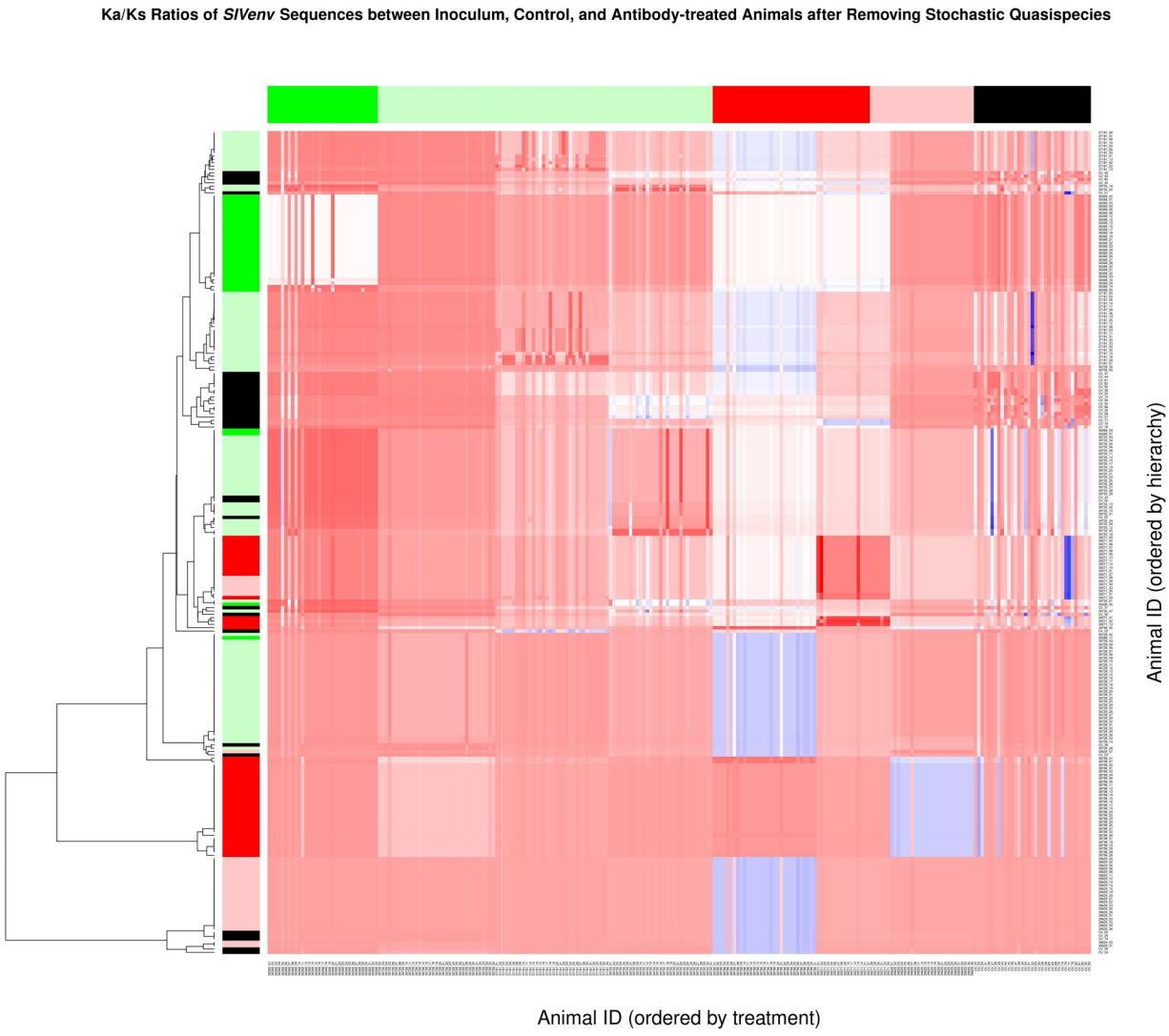
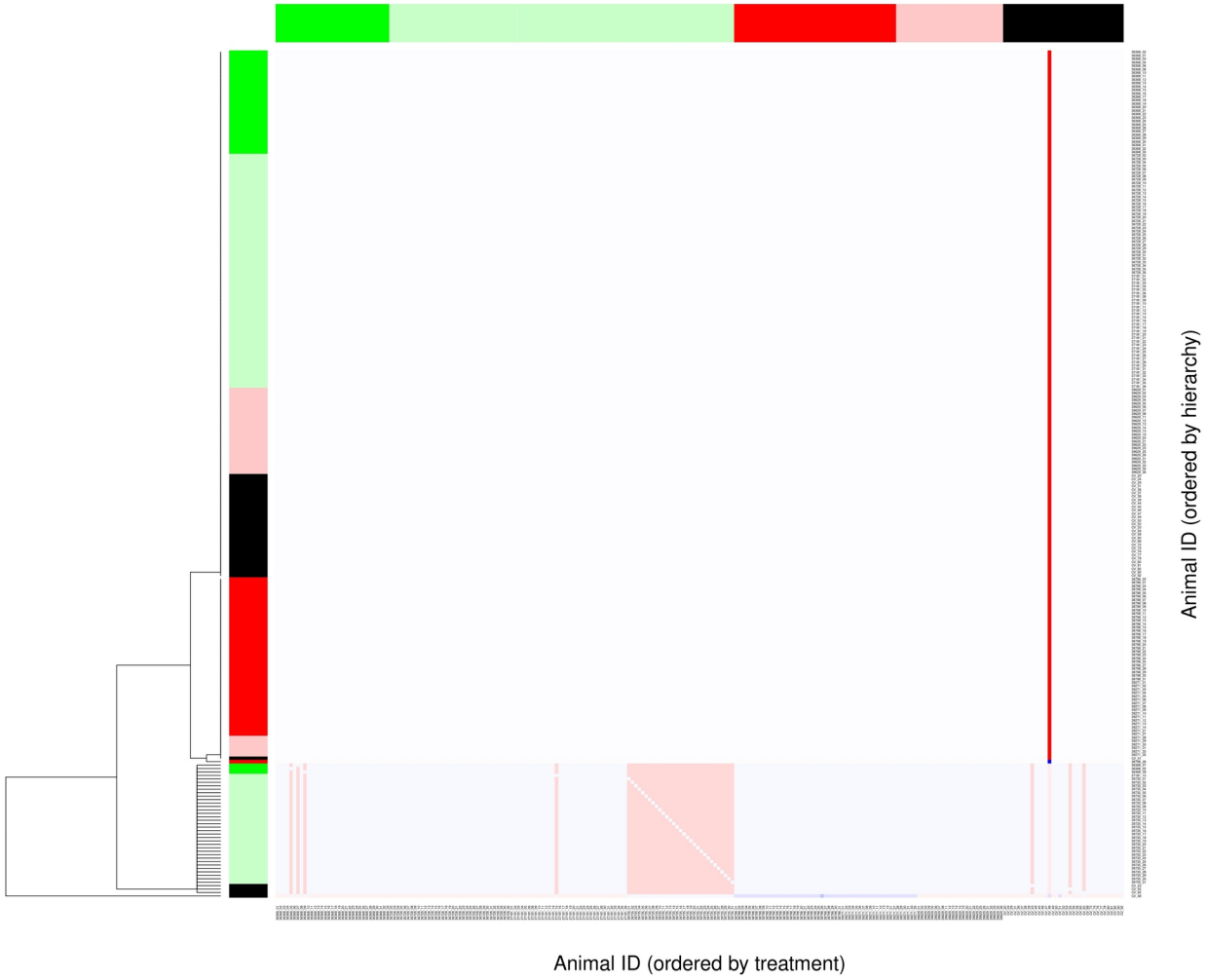


Figure 23: Heatmap of Ka/Ks ratios calculated for *SIVenv* gene between SGA-sequenced variants. Ka/Ks ratios were calculated for the entire *SIVenv* gene among the various treatment conditions, depicted on the top horizontal and left vertical panels: red is for antibody-opsonization in neutral pH, light red is for antibody-opsonization in acidic pH, green is for control in neutral pH, light green is for control in acidic pH, and black is for inoculum. Ka/Ks ratios between each sequence read are organized by hierarchical clustering (on the left panel) and by treatment (on the top panel). Ka/Ks ratios for each comparison that is greater than 1 is shown in blue, less than 1 in red, and equal to 1 in white.

However, there is some correlation between SGA sequences from animals infected with antibody-opsonized virus and a Ka/Ks ratio > 1, although the presence of mostly deterious mutations obfuscates the discovery of mutations that may enhance fitness under certain conditions. To investigate this further, Ka/Ks ratios were calculated for the particular polymorphisms (NKSK, I743T, and G812R) that had higher representation among the control and antibody-opsonized groups. The heatmaps of Ka/Ks ratios for I743T (**FIGURE 24A**) and G812R (**FIGURE 24B**) do not reveal any significant patterns of selection between the treatment groups, suggesting that these mutations are not under selection pressure. However, the heatmap for NKSK (**FIGURE 24C**) shows a strong correlation between treatment groups and Ka/Ks ratio across all comparisons. The Ka/Ks ratios between the antibody group and the inoculum were significantly different from those between the control group and the inoculum (**FIGURE 25**), suggesting that antibody-opsonization of SIV establishes selection pressure favoring the NKSK polymorphism. Although the Ka/Ks ratios presented here suggests that antibody-opsonization of virus fomented positive selection pressure favoring the NKSK polymorphism, there is still insufficient information to determine whether such pressure was conducive to sufficient demarcation between the antibody-opsonized group and the control group, including the inoculum.

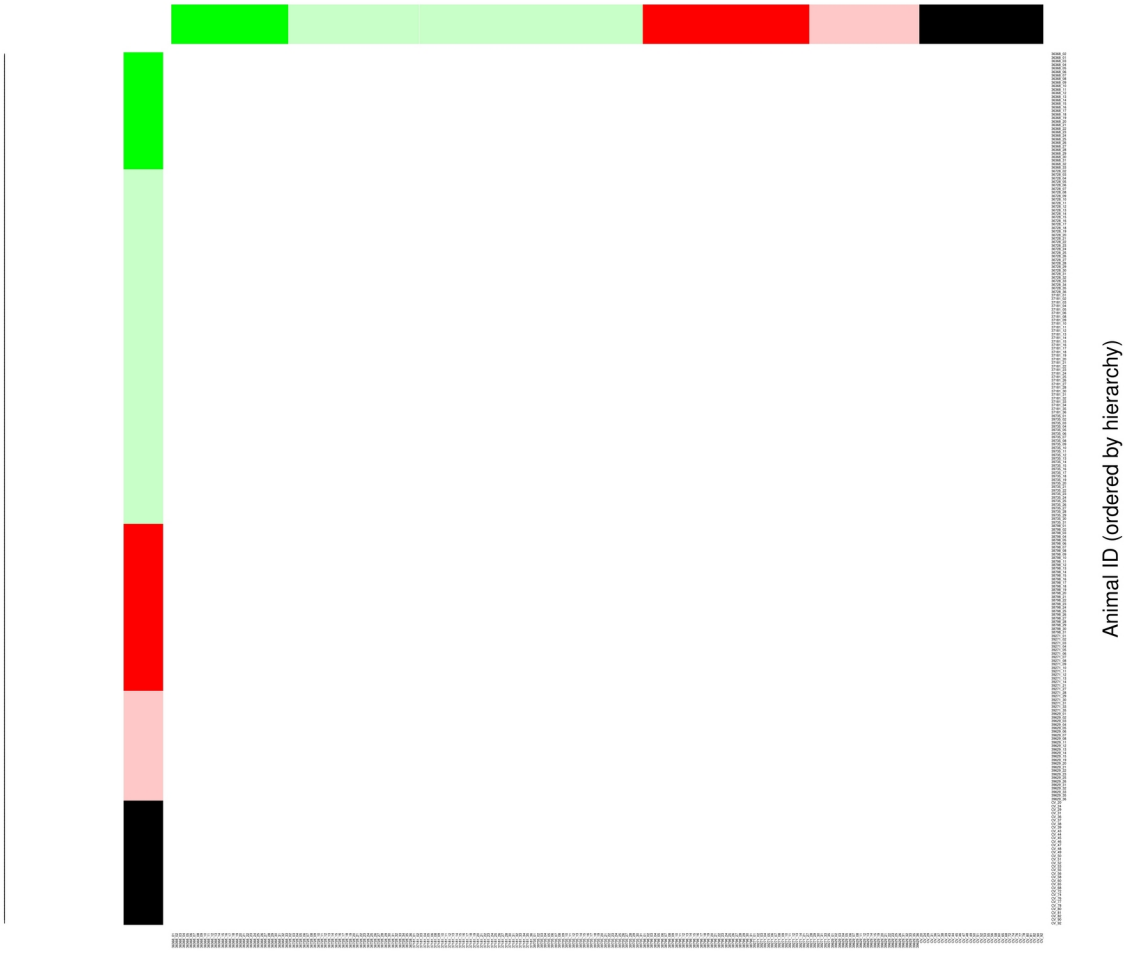
A

Ka/Ks Ratios of QQIHTQQ and QQTHIQQ Sequences between Inoculum, Control, and Antibody-treated Animals after Removing Stochastic Quasispecies



B

Ka/Ks Ratios of ILQGLSA and ILQRLSA Sequences between Inoculum, Control, and Antibody-treated Animals after Removing Stochastic Quasispecies



C

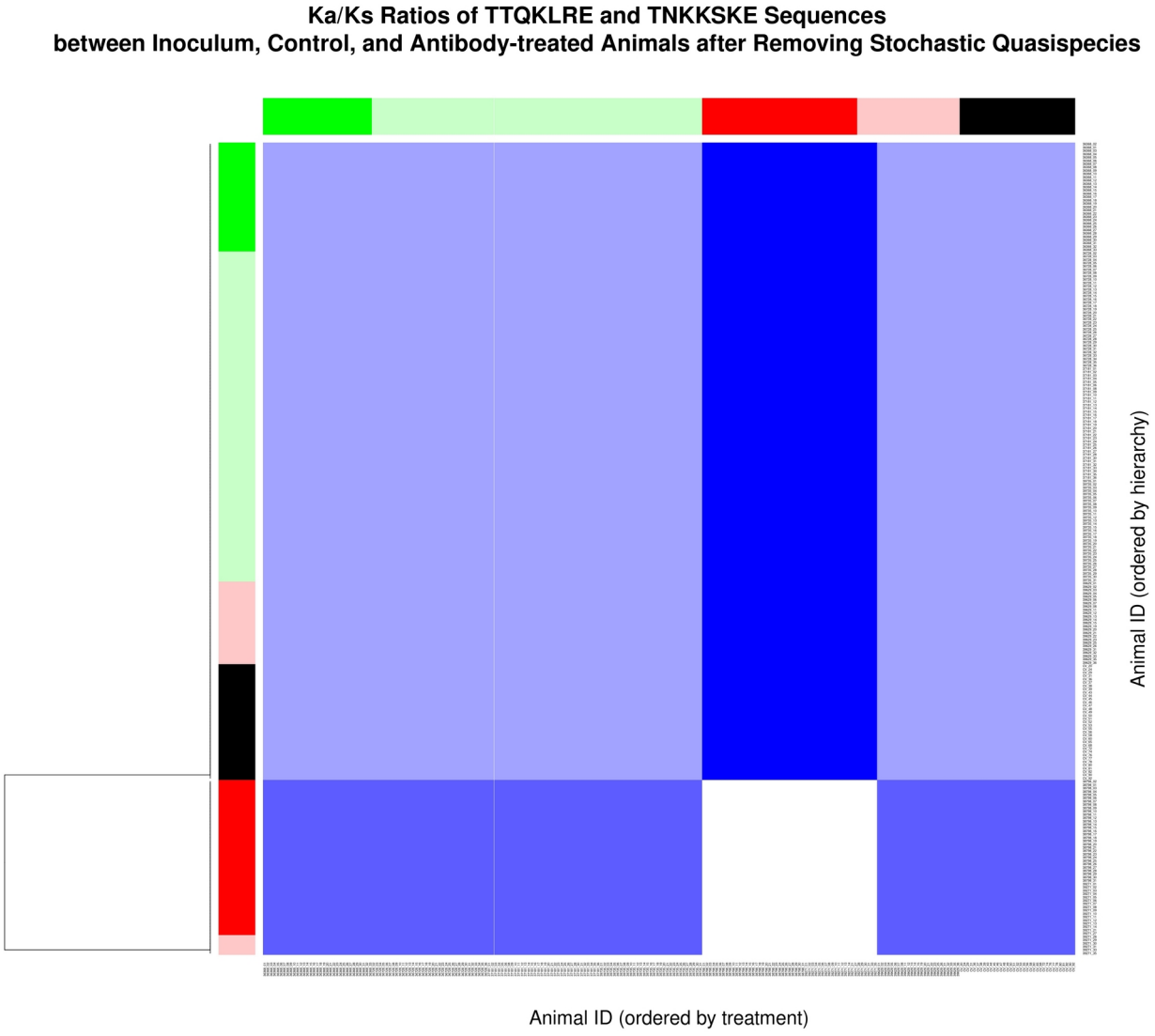


Figure 24: Heatmap of Ka/Ks ratios calculated for I743T (A), G812R (B), and NKSK (C) selected polymorphisms between SGA-sequenced variants (this and two previous pages). Ka/Ks ratios were calculated for a span of 7 amino acid containing the polymorphisms I743T (A), G812R (B), and NKSK (C) among the various treatment conditions, depicted on the top horizontal and left vertical panels: red is for antibody-opsonization in neutral pH, light red is for antibody-opsonization in acidic pH, green is for control in neutral pH, light green is for control in acidic pH, and black is for inoculum. Ka/Ks ratios between each sequence are organized by hierarchical clustering (on the left panel) and by treatment (on the top panel). Ka/Ks ratios for each comparison that is greater than 1 is shown in blue, less than 1 in red, and equal to 1 in white.

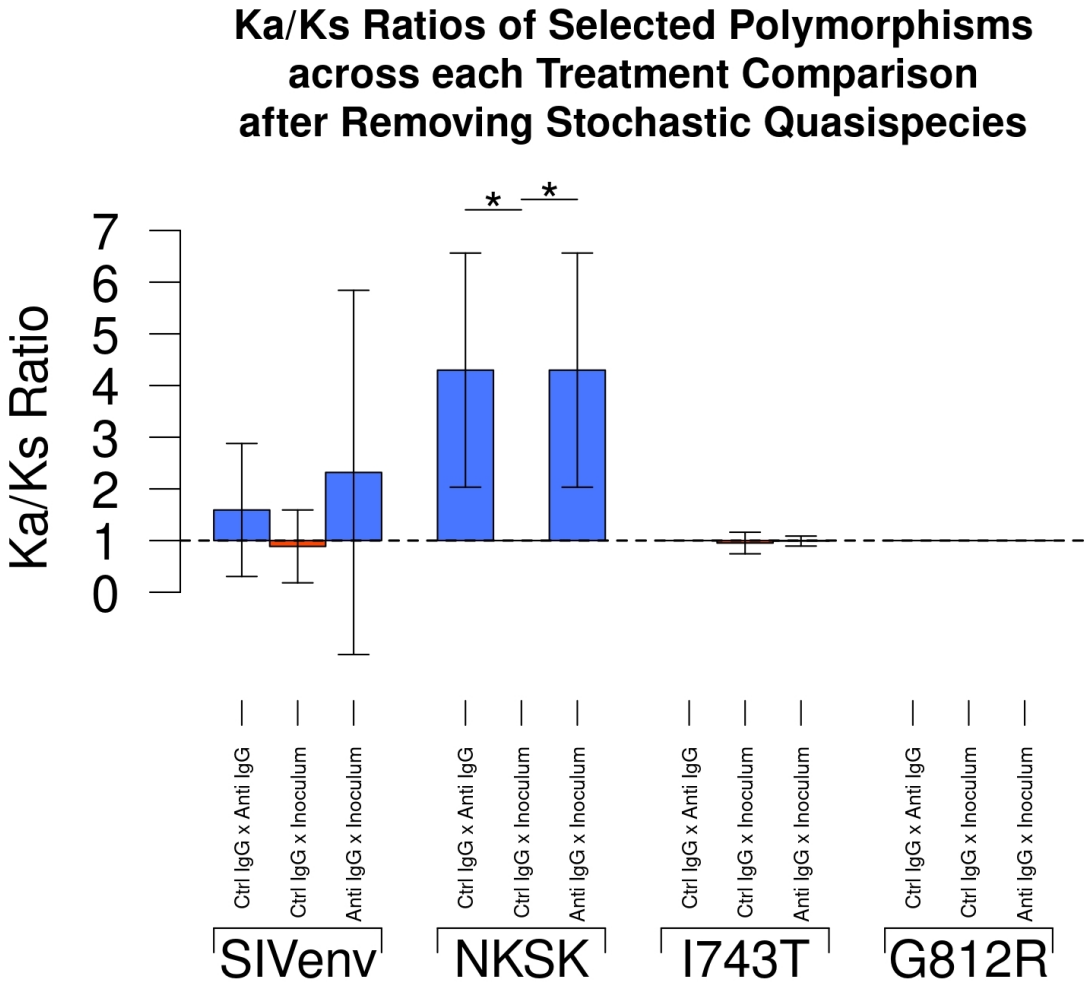


Figure 25: Mean Ka/Ks Ratios of each selected polymorphism between different comparisons. Ka/Ks ratios for each comparison and each polymorphism are displayed as mean ± standard deviation. * $p < 0.05$; Welch's t-test.

One additional test of selective adaptation/pressure that may answer aforementioned question is the McDonald-Kreitman test (MKT) [McDonald and Kreitman, 1991]. This is a variation of the Ka/Ks calculation for selection in the sense that it compares Ka/Ks ratios for fixed and polymorphic mutations in order to determine whether genetic variations within a particular species or group drove its evolution by comparing it to genetic variations between the species and an outgroup. In this regard, an "in" group represents any experimental group or species (for example, viral output of antibody-opsonized T/Fs) for which one wishes to determine whether a selection pressure played a role in its diversification from an "out" group (for example, the viruses of the infectious inoculum). This is calculated by dividing the nonsynonymous mutation rate of an "in" group by the synonymous mutation rate of an "out" group. The ratio here is a measurement of the selection pressure (whether positive or negative) that led to diversification between the two groups, and it also represents a proxy for the distribution of mutations that have become fixed between the "in" group and the "out" group, essentially demarcating the genetic differences between the two groups.

The Ka/Ks ratio, as described above, only refers to the genetic diversity between two groups, and it is a measurement of selective pressures that established the "fixed" mutations that demarcate one group from another. This generally only informs the researcher whether selection pressures were sufficient to demarcate one group from another, but it is not without statistical biases. The Fixed Ka/Ks ratio, on its own, lacks information on genetic diversity internal to a particular group, and this is a crucial point of contention while studying drivers of selection and evolution: The observed demarcation in mutation signatures between two groups could be an artefact merely of high genetic diversity *within* a group. Thus, the Fixed Ka/Ks ratio must be

compared with a Polymorphic Ka/Ks ratio. This ratio is calculated in the same way as the Fixed Ka/Ks ratio is calculated, except that the "out" group is the same as the "in" group. It is a measure of selection pressure within a population or group and represents a baseline for internal genetic diversity. If the Fixed Ka/Ks ratio is greater than the Polymorphic Ka/Ks ratio, then the observed differences between the "in" and "out" group are not likely influenced by the genetic diversity within the "in" group and therefore positive selection pressure contributed significantly to diversification between the "in" and "out" groups. If, however, the Fixed Ka/Ks ratio is lower than the Polymorphic Ka/Ks ratio, then the observed differences between the "in" and "out" groups could have been influenced by genetic diversity and that the demarcations between the two may not be fixed. It follows from this interpretation that an "in" group and "out" group are not as diversified from one another and that negative (or purifying) selection pressures maintain their similarities by removing potentially deleterious mutations from the genetic pool. Normalizing the Polymorphic Ka/Ks ratio to the Fixed ratio (done by dividing the former by the latter) computes the neutrality index (NI) [Meiklejohn et al., 2007; Stoletzki and Eyre-Walker, 2011], which represents a less biased assessment of selection pressure within and between populations: positive selection is indicated by $NI < 1$ while negative selection by $NI > 1$.

MKT analysis of the entire *SIVenv* and the candidate polymorphisms NKSK, I743T, and G812R are shown in **FIGURE 26**. As shown previously in the heatmap depicted in **FIGURE 23**, mutations across the envelop gene are generally inimical to replication fitness and selected out of the genetic pool, and as stated previously the background of deteterious mutations obfuscates the identification of particular regions of the *env* gene under positive selection. Indeed, MKT analysis here corroborates the initial conclusions garnered from the Ka/Ks

heatmaps (**FIGURE 24**) and graphs (**FIGURE 25**). Positive selection favors the NKSK polymorphism during transmission of antibody-opsonized virus compared to the control, as indicated by an NI = 0.17. It was curious to find that NI = 1.00 in all other comparisons for this polymorphism; the reasoning here could be that the nucleotide substitution rates are equal in every group but that antibody-opsonization contributed to an enhancement of selection pressure for certain amino acid changes, hence why the Ka/Ks ratios are much higher among comparisons wherein the antibody group was included. For the G812R polymorphism, there is no difference in the Fixed Ka/Ks ratios and the Polymorphic Ka/Ks ratios across all comparisons (NI = 1.00). However, the ratio and neutrality index calculations for the I743T polymorphism yielded extreme values, with Ka/Ks ratios ranging from nearly 0 to 22.37 and neutralization indices as low as 0.04. The reason for such extreme values was due to the reduction in sample sizes as a result of removing sequences from stochastically transmitted quasispecies. The remaining sequence regions that flank the I743T region exhibited very low sequence diversity, and as a result led to Ka and Ks rates of nearly 0 or undefined. Calculating the Ka/Ks ratios among this reduced data set led to Ka/Ks ratios that were either very small or very large. Given that I743T likely isn't under any kind of selection pressure, in as much as G812R was not, I743T was rejected as candidate polymorphism for T/F selection and was only included in further anlysis as an internal control of sorts for a polymorphism that only artificially appeared to be selected.

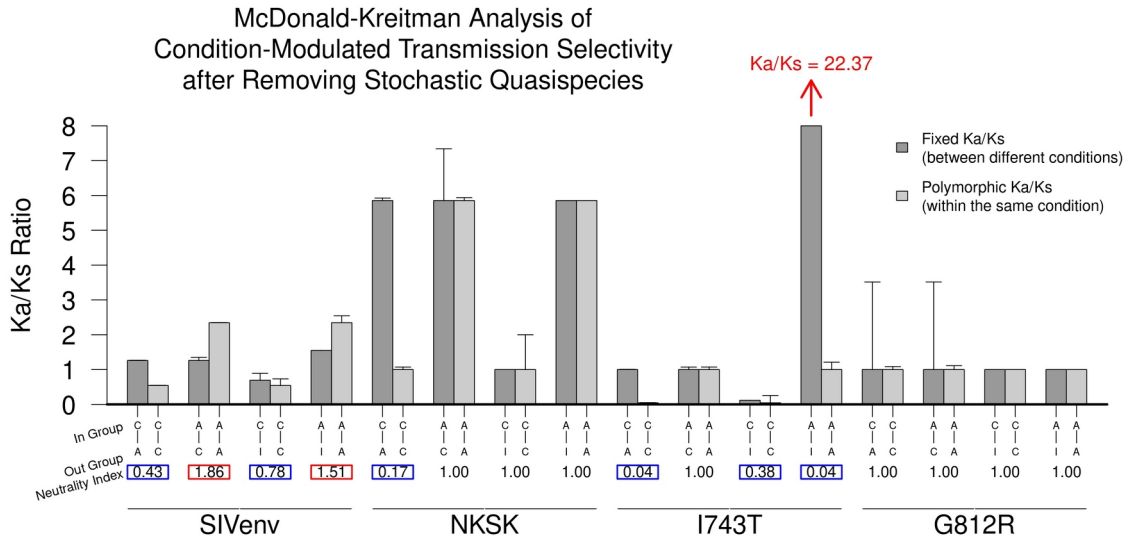


Figure 26: McDonald-Kreitman (MKT) analysis of genetic divergence within and between experimental groups and the determination of positive and negative selection pressures. Ka/Ks ratios were calculated as comparisons between particular "In groups" and "out groups". Groups are designated by a letter: "C" is for control, "A" is for antibody treatment, and "I" is for inoculum. The neutrality index [Meiklejohn et al., 2007; Stoletzki and Eyre-Walker, 2011] is calculated as dividing the "polymorphic" Ka/Ks ratio within the "in" group by the "fixed" Ka/Ks ratio between the "in" and "out" groups. NI < 1 indicates that positive selection led to fixation of the polymorphism within a condition (indicated by a blue box). NI > 1 indicates that negative selection disfavors the particular polymorphism and purifies it from a condition (indicated by a red box). NI = 1 indicates that a particular polymorphism is neutral, neither favored nor disfavored by selection.

BLAST searches for similar polymorphisms and common mechanisms of selection in literature

In order to investigate the phenotypic functions of these polymorphisms (NKSK especially) further and determine what roles they may play in selection, replication fitness, or viral infectivity, all six polymorphisms (LTTQLR, NKSK, I743, I743T, G812, and G812R) were submitted as protein BLAST queries in order to search for published reports in the literature that may have previously studied these polymorphisms (or at least the phenotypes and functions of the given regions). Due to the propensity of BLAST's algorithm in reporting the same accession

numbers for two different sequences, the complete list of accession numbers from each BLAST query were reduced and refined to deduplicate subject accession numbers from each queried sequence, identify accession numbers unique to a particular sequence, and search publications for those accession numbers with the lowest E-value and for those that are unique to a particular sequence (**APPENDICES 4 – 15**). **TABLE 5** shows the overall summaries of BLAST queries, results, accession numbers, and reduction thereof, as well as summaries of related publications unique to each polymorphism or with low E-values. Many of the Genbank entries uncovered here are likely nonspecific hits due to sequence similarities with SIV isolates, clones, vaccines, or constructs used in those studies rather than the polymorphisms themselves [Chakrabarti et al., 1987; Franchini et al., 1987; Hirsch et al., 1987; Unger et al., 1992; Bobadilla et al., 2013; Del Prete et al., 2013; Kulkarni et al., 2013]

Interestingly, a common theme of some of the papers that were returned from the protein BLAST search was concerned with mutations associated with SIV neurovirulence [Sharma et al., 1992; Anderson et al., 1993; Flaherty et al., 1997; Chen et al., 2006]. Another paper, found to be unique to NKSK, was on the phenotype-specific compartmentalization of SIV in the male genital tract (MGT) of infected animals [Tortorec et al., 2008], which is pertinent to the current study in the sense that these animals were infected through the male genital tract. One paper [Unger et al., 1992] reporting on the function of *SIVnef* showed that Nef was not necessary for viral infectivity and replication fitness; strangely, this hit was determined to be unique to the LTTQLR polymorphism in V4, and Unger et al. [1992] did not strictly study the *SIVenv* gene or deposited

TABLE 5: BLAST search results for selected polymorphisms among T/F and inoculum consensus sequences

Polymorphism	Abbreviation	Strain	Animals Infected with Polymorphism Variant	Group	*Directionality of Selection Pressure p-value*¶	SIV gp160 Blast Query‡	Protein Domain	# of Accession Hits: Raw	# of Accession Hits: After Deduplication	# of Accession Hits: Unique to Polymorphism	E-value cutoff⁑	References for hits lower than E-value Cutoff	References Unique to Blast Query	Major Finding of Reference
L422_T423del/T426N/Q427K/L429S/R430K	NKSK*	Selected Polymorphism	36798 39271	Anti-SIV, pH 6.0 Anti-SIV, pH 7.4	Favored by Anti-SIV CA: 0.18 CI: 1.00 AI: 0.01	WFLNWVEDRN–LTNKKSKE	gp120 V4	179	100	2	10^{-40}	Franchini et al. [1987] Flaherty et al. [1997] Chen et al. [2006] Tortorec et al. [2008]	SIVmac is more closely related to HIV-2 than to HIV-1 Neurovirulence of macrophage-tropic viruses depend on functional gp41 and nef Compartmentalization of SIV in different brain regions is phenotype-specific Compartmentalization of SIV in male genital tract is phenotype-specific
L422/T423/T426/Q427/L429/R430	LTTQLR	Inoculum Consensus	36368 36728 37181 37906 38893 39629 39734 39735	Control, pH 7.7 Control, pH 6.0 Control, pH 6.0 Control, pH 6.0 Control, pH 6.0 Anti-SIV, pH 7.4 Anti-SIV, pH 7.4 Control, pH 6.0		WFLNWVEDRNLTLTTQKLRE		430	92	10	10^{-9}	Chakrabarti et al. [1987] Franchini et al. [1987] Hirsch et al. [1987] Unger et al. [1992] Anderson et al. [1993] Chen et al. [2006] Bobadilla et al. [2013] Kulkami et al. [2013]	SIVmac is phenotypically related to HIV-1 and HIV-2 SIVmac is more closely related to HIV-2 than to HIV-1 Premature truncation of SIV and HIV gp160 does not impair infectivity SIVnef gene is dispensable for viral infectivity and replication. Polymorphisms associated with SIV neurovirulence Compartmentalization of SIV in different brain regions is phenotype-specific Development of lentiviral vectors using vpx Intradermal/intramuscular SIVenv DNA delivery produced robust ADCC and bnAb responses in vaccinated animals
I743T	I743T*	Selected Polymorphism	36798 39271	Anti-SIV, pH 6.0 Anti-SIV, pH 7.4	Favored by Anti-SIV CA: 0.18 CI: 1.00 AI: 0.03	YSQQTHTQQ	gp41	86	86	4	10^{-5}	Kulkami et al. [2013]	Flaherty et al. [1997] Chen et al. [2006] Intradermal/intramuscular SIVenv DNA delivery produced robust ADCC and bnAb responses in vaccinated animals	Neurovirulence of macrophage-tropic viruses depend on functional gp41 and nef Compartmentalization of SIV in male genital tract is phenotype-specific
I743	I743	Inoculum Consensus	36368 36728 37181 37906 38893 39629 39734 39735	Control, pH 7.7 Control, pH 6.0 Control, pH 6.0 Control, pH 6.0 Control, pH 6.0 Anti-SIV, pH 7.4 Anti-SIV, pH 7.4 Control, pH 6.0		YSQQIHTQQ		82	82	0	10^{-5}	Kulkami et al. [2013]	Intradermal/intramuscular SIVenv DNA delivery produced robust ADCC and bnAb responses in vaccinated animals	
G812R	G812R	Selected Polymorphism	36368 37181 39271 39735	Control, pH 7.7 Control, pH 6.0 Anti-SIV, pH 7.4 Control, pH 6.0	Favored by Control CA: 0.55 CI: 0.05 AI: 0.51	VYQILQPILQRLSATL	Intergenic Region	228	93	5	10^{-10}	Chakrabarti et al. [1987] Hirsch et al. [1987] Anderson et al. [1993] Chen et al. [2006]	SIVmac is phenotypically related to HIV-1 and HIV-2 Premature truncation of SIV and HIV gp160 does not impair infectivity Polymorphisms associated with SIV neurovirulence Compartmentalization of SIV in different brain regions is phenotype-specific	
G812	G812	Inoculum Consensus	36798 36728 37906 38893 39629 39734	Anti-SIV, pH 6.0 Control, pH 6.0 Control, pH 6.0 Control pH 6.0 Anti-SIV, pH 7.4 Anti-SIV, pH 7.4		VYQILQPILQGLSATL		318	100	12	10^{-5}	Anderson et al. [1993] Chen et al. [2006] Tortorec et al. [2008]	Polymorphisms associated with SIV neurovirulence Compartmentalization of SIV in different brain regions is phenotype-specific Compartmentalization of SIV in male genital tract is phenotype-specific	

* = linked mutations
‡ = Amino acid sequences for BLASTP queries are based on refseq AGG38452.1
¶ = Two-tailed Fisher's Exact Test, Bonferroni-corrected
CA = Fisher's exact test comparing distribution of polymorphisms between Control and Anti-SIV groups
CI = Fisher's exact test comparing distribution of polymorphisms between Control group and Inoculum
AI = Fisher's exact test comparing distribution of polymorphisms between Anti-SIV group and Inoculum
⁑ = E-value cutoffs determined based on lowest E-values of each BLAST alignment (refer to APPENDIX XXX)

SIVenv gene sequences to Genbank. It could be a nonspecific hit because Unger et al. [1992] incorporated portions of *SIVenv* in their cloned constructs of *SIVnef*; Env was not a focus of research in this paper.

Mechanisms controlling neurovirulence does not explain antibody-mediated selection

Viral neuroinvasion and neurovirulence are well-documented phenomena of SIV [Sharma et al., 1992; Anderson et al., 1993; Flaherty et al., 1997; Chen et al., 2006] and HIV [Dohgu et al., 2011; Gonzalez-Perez et al., 2012; Gonzalez et al., 2017; Brese et al., 2018; Cantres-Rosario et al., 2019] infections. Anderson et al. [1993] and Flaherty et al. [1997] studied neurovirulence using SIVmac239 as a model for neuroinvasion in *R. macaques*. The latter group of researchers established lines of infectious clones containing mutations discovered in animals with clear evidence of viral neuroinvasion by the former group of researchers. SIVmac239.R71 [Anderson et al., 1993] and SIVmac239.17E-Fr [Flaherty et al., 1997] are neurovirulent strains that happen to also be macrophage-tropic (which is an important feature of neurovirulence as viral adherence to macrophages and microglia is central to how viruses bypass or penetrate the blood brain barrier). SIVmac239.17E-Cl, on the other hand, is macrophage tropic but yields no demonstrable evidence of neurovirulence [Flaherty et al., 1997]. Phylogenetic tree analysis shows that these clones is highly diverged from T/Fs and inoculum sequences used in the current studies, but this is likely due to differences in clinical isolates (namely SIVmac239 used by prior studies and SIVmac251 used in the current study) (**FIGURE 27**). Sliding window analysis does show polymorphisms shared by neurovirulent strains and T/Fs in this study (as well as macrophage tropic strains) (**TABLE 6**); however, none of the selected polymorphisms, not even

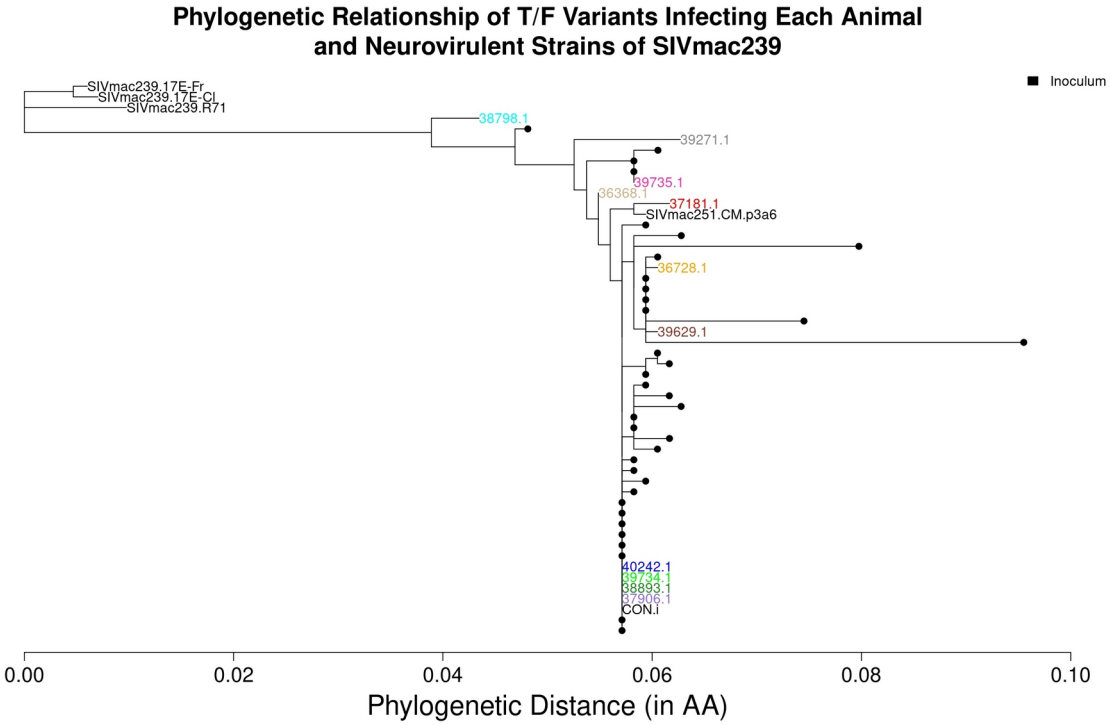
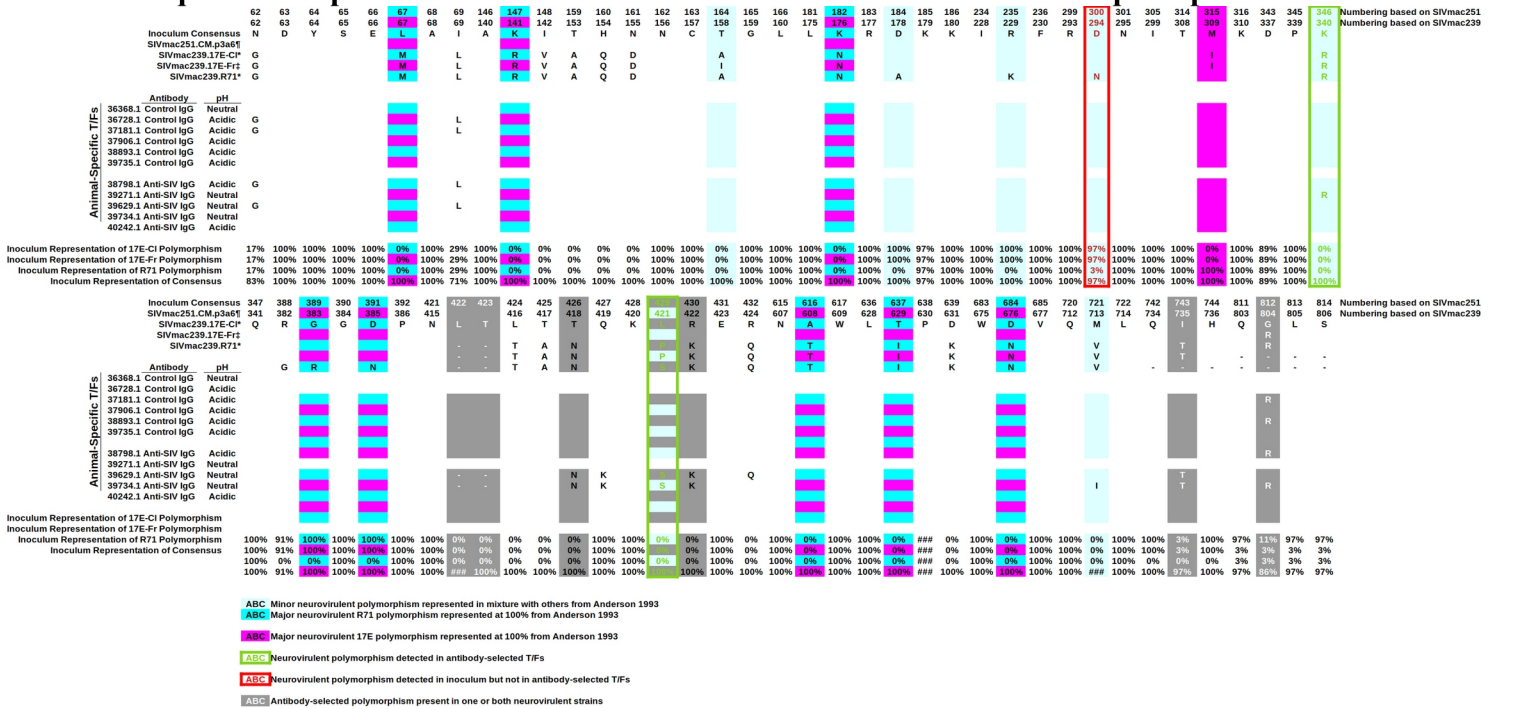


Figure 27: Phylogenetic tree depicting the relationship between inferred T/F quasispecies and neurovirulent strains from Anderson et al. [1993] and Flaherty et al. [1997]. Maximum-likelihood trees developed for inferred T/F sequences and their relationship to inoculum variants and macrophage-tropic neurovirulent SIV strains (SIVmac239.R71 [Anderson et al., 1993] and SIVmac239.17E-Fr [Flaherty et al., 1997]) as well as a macrophage-tropic non-neurovirulent SIV strain (SIVmac239.17E-Cl [Flaherty et al., 1997]).

NKSK, from antibody-selected T/Fs were shared by either the R71, 17E-Cl, or 17E-Fr strains, nor were neurovirulence-important polymorphisms present in the T/Fs of the current studies. Similar observations were made in comparisons between T/Fs of the current study and central nervous system (CNS) compartmentalized SIV strains reported by Chen et al. [2006]. Neither sliding window analysis (data not shown) or phylogenetic tree inference (**FIGURE 28**) showed relationships between T/Fs and neurovirulent strains of two different clinical isolates of SIV. In light of these analyses, it was concluded that mechanisms of neurovirulence and CNS compartmentalization were not likely related to antibody-mediated selection pressures in the current study.

Table 6: Sequence comparisons between neurovirulent strains and inferred T/F quasispecies



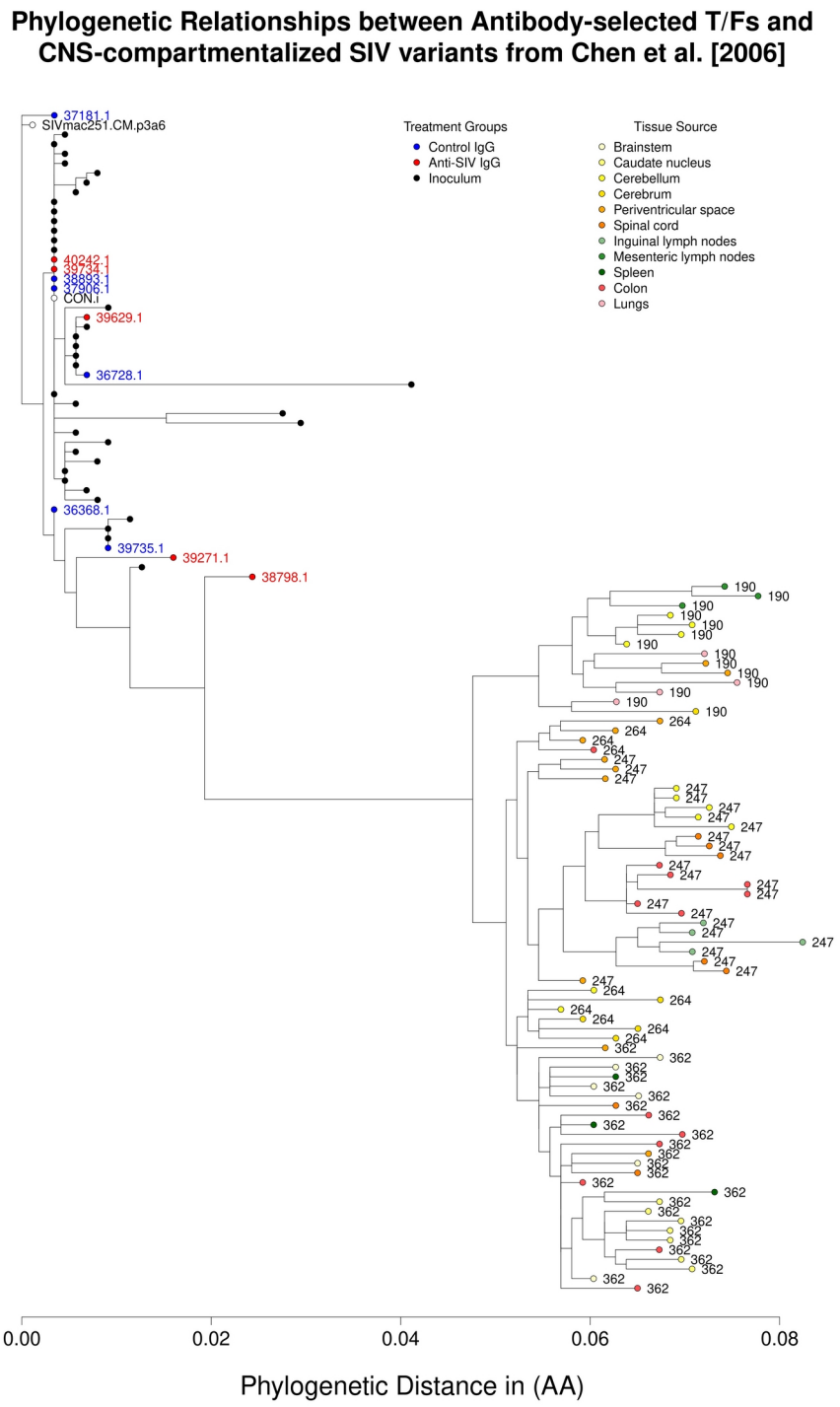


Figure 28: Phylogenetic tree depicting the relationship between inferred T/F quasispecies and CNS-compartmentalized strains from Chen et al. [2006]. Maximum-likelihood trees developed for inferred T/F sequences and their relationship to inoculum variants and CNS-compartmentalized SIVmac239 variants from Chen et al. [2006]. Numbers next to tips indicate animal IDs.

Male genital tract compartmentalization does not explain antibody-mediated selection through the penile route

Regarding the question of tissue compartmentalization, it was shown that antibody-selected T/Fs did not share the same polymorphisms as they did with strains of SIVmac239. This is not to say that viruses from the current study are not themselves neurovirulent or that they would not contribute to systemic infection of the CNS during chronic infection. As of this analysis at this time, mechanisms of viral neurovirulence are not likely related to any of the selected polymorphisms of the T/Fs of the current study. However, some of the hits from the BLAST came from a paper on strain-specific tissue compartmentalization of the MGT [Tortorec et al., 2008]. In this study, cynomolgus macaques were infected intravenously with a viral swarm of SIVmac251, then tissues and plasma were analyzed for SIV infection and compartmentalization. Even through the intravenous route, the researchers found that the virus took residence in tissues of the MGT in a phenotype-specific manner as early as the acute phase and such populations persisted through the chronic phase. While the animals used by Tortorec et al. [2008] were not infected in the same manner as those used in the current study, the migration of certain strains to the MGT could be related to mechanisms of penile inoculation, and that selective pressures guiding both processes may be similar. To investigate this further, phylogenetic analysis was performed to compare the polymorphisms of the current T/Fs with those of the MGT-resident viruses. Since Tortorec et al. [2008] focused their analyses on V1V2, comparisons were restricted to the V1V2 regions of the T/Fs and inocula discovered by the present study. It should be noted that the V1V2 hits from the BLAST results could have been nonspecific, as the queries used were from V4 and gp41. Nevertheless, analysis was performed

in case similarities could be found between MGT-related viruses and antibody-selected T/Fs through the penile route.

A hierarchical tree depicting the phylogenetic relationship among MGT-compartmentalized strains, antibody-selected and control T/Fs is shown in **FIGURE 29**. The tree did not reveal co-clustering of viruses from either treatment group or inoculum and viruses from the report by Tortorec et al. [2008], even though they both originated from a similar clinical isolate (albeit not the same exact stock). The tree does show, however, tissue-specific clustering of MGT-related viruses, indicating that this analytical approach was sensitive enough to detect similarities in viruses from the standpoint of V1V2 sequences, which are normally subject to high mutation rates. Sliding window analysis indicates strong similarities between MGT-resident viruses and antibody-selected T/Fs across the V1V2 region, but not in any site-specific manner (data not shown). In fact, it would appear that the similarities between the viruses from both studies are due in large part to the inoculum also exhibiting a high degree of similarity with MGT-resident viruses; this would suggest that such similarities are not due to a common mechanism of MGT compartmentalization and antibody-selected transmission of particular T/Fs.

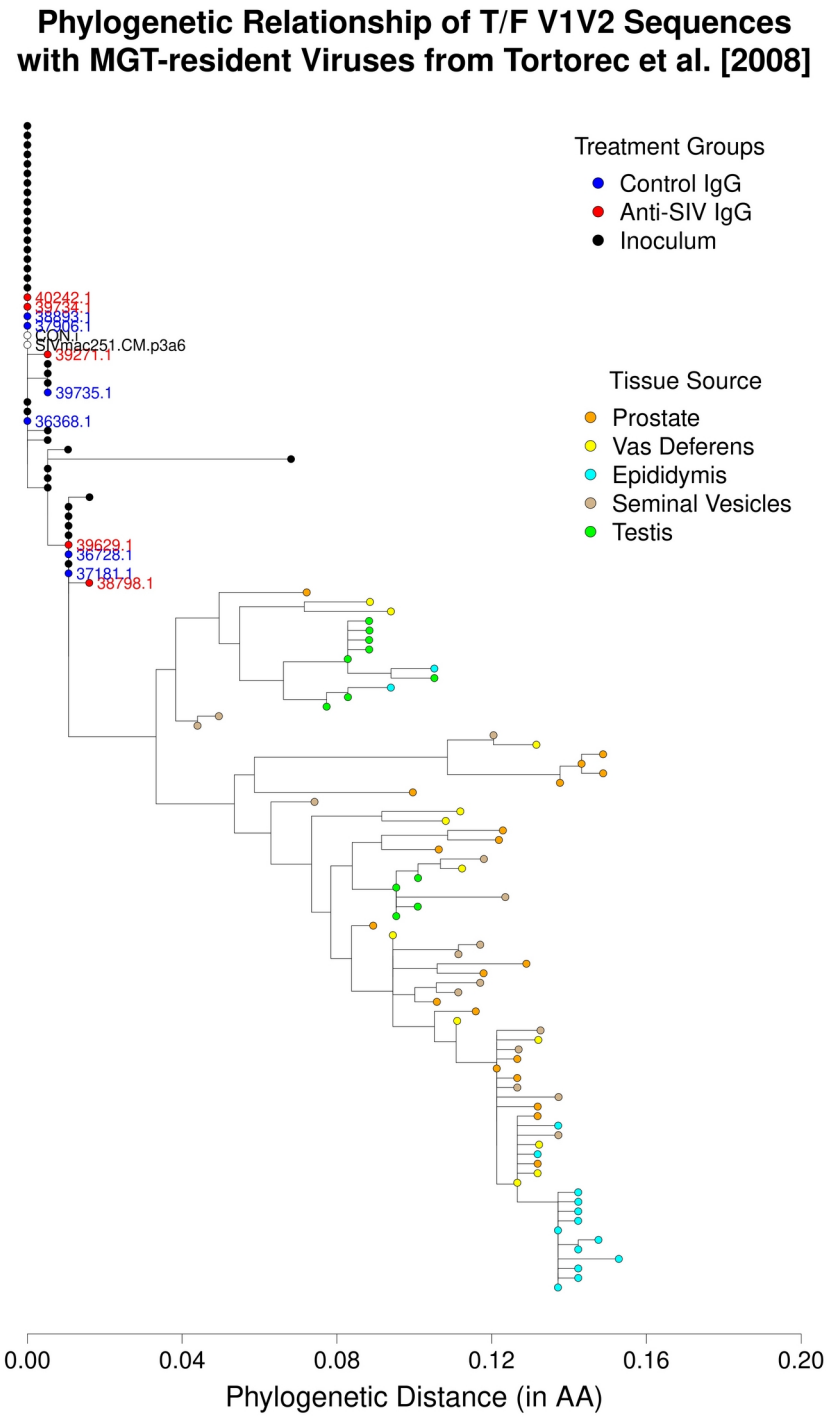


Figure 29: Phylogenetic tree depicting the relationship between inferred T/F quasispecies and MGT-compartmentalized strains from Tortorec et al. [2008]. Maximum-likelihood trees developed for inferred T/F sequences and their relationship to inoculum variants and MGT-compartmentalized SIVmac239 variants from Tortorec et al. [2008].

A potentially novel polymorphism favored by a potentially unknown mechanism of positive selection

In searching for a mechanism of antibody-mediated selection, it was clear that neuroinvasion and MGT-compartmentalization were not related to the mechanism of selection under the current study, although this does not discount the possibility of MGT- and CNS-compartmentalization of T/F progenies selected through the penile route. There is an apparent lack of sources in literature on the antibody-selected NKSK polymorphism, let alone papers on possible mechanistic explanations for such selection. Therefore, it is a conclusion of this paper and a partial answer to Hypotheses #1 and #3 that there exists some form of selection pressure, however weak it may be, in the penile mucosa that selected for this polymorphism during transmission of an antibody-opsonized virus, and that such selection pressure may be influenced by other antibodies of different neutralizing capacities at different titers as well.

Chapter 4: Findings and Speculations—Antibodies

Integrating Two Accepted Experimental Approaches in a Novel Way

The goals of this research undertaking was, described in the broadest of terms, to simulate clinical infection *in vitro* and to use that model to study T/F viruses with greater replicability. A working model of such design and output would become an invaluable tool to other researchers interested in developing cures and vaccines not only against HIV-1 but other viruses as well. This new model is a synthesis of two commonly employed approaches: (1) tissue culture experiments involving *in vitro* infection of primary isolations of PBMCs or their subsets (i.e. DCs, T-lymphocytes, macrophages, etc.) and (2) phylogenetic analysis of viruses from infected tissues or cells. The former approach has often been utilized in research to study interactions between virus and immortalized cell lines in controlled laboratory settings [Dalgleish et al., 1984; Kozlowski et al., 1995; Hong et al., 2002; Yuan et al., 2013; Sarzotti-Kelsoe et al., 2014; Ikeda et al., 2019; Pierini et al., 2021] while the latter approach has been limited to studying the phylogenies and phenotypes specific to T/F and chronic variants in clinical settings [Frost et al., 2005; Keele et al., 2008; Salazar-Gonzalez et al., 2008; Mild et al., 2010; Bull et al., 2013; Yongshong et al., 2015; Affi et al., 2022]. When the former is used in a clinical setting, *ex vivo* explants may be used or PBMCs from HIV-naive or -infected individuals are studied *in vitro*, but rarely would this be done to discover novel HIV T/Fs [Hussain et al., 1995; Shepard et al., 2000; Kawamura et al., 2008; Stacey et al., 2009; Jenabian et al., 2010; Shen et al., 2014; Bertram et al., 2019; Pedersen et al., 2021]. When the latter approach is used in a controlled experiment, it is usually done with infected animals rather than tissue culture of isolated human PBMCs, limiting the translational benefits of novel T/F discoveries from animals

to humans [Keele et al., 2009; Shen et al., 2010c; Stone et al., 2010; Salgado et al., 2014; Stieh et al., 2016; Yuan et al., 2017]. In the few times where T/Fs are studied *in vitro*, features of T/F and cell interactions are studied in the reverse direction [Liu et al., 2013; Song et al., 2016; Iyer et al., 2017; Schwartz et al., 2018; Ashokkumar et al., 2020; Day et al., 2022; Sugrue et al., 2022], rather than discovered in forward.

The current model was designed to combine both of these approaches to accelerate discovery of new T/Fs but in a laboratory setting rather than a clinical one. Its is its versatility in testing any permutation of physiological or immunological conditions that may modulate HIV-1 transmission selectivity, which could provide researchers new insights into yet undiscovered T/F-specific polymorphisms and host cell interactions and selection pressures that favor certain quasispecies over others during HIV-1 transmission. These data could be useful for rational and empirical vaccine design in the future. Thus, the experimental results presented in the following pages are the output of this model in action, and its ability to capture or simulate a clinical-like infection (i.e. selective infection by rare variants from an inoculum, genetic bottleneck, selective pressures favoring certain polymorphisms, etc.) will be evaluated.

The Effect of Antibody in Variant Selection during *Trans*-Infection

Using the model to test antibody-opsonization as a potential selection pressure for HIV-1 transmission

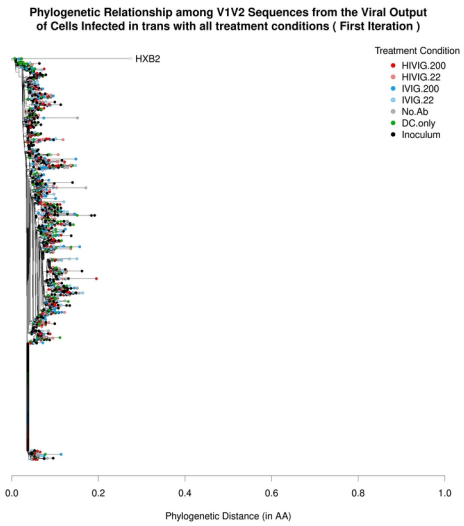
In followup from the previous study using antibodies as a selection modulator of T/F viruses, the current approach involved using polyclonal antibodies (HIVIG) against HIV-1 gp160 to determine how they influence the infection selectivity of HIV-1 clinical isolate 92US657

during *trans*-infection. First, a sample viral inoculum was incubated in the presence of 200 μg/mL of HIVIG or an isotype control (IVIG). Then primary isolated DCs were incubated with opsonized viruses to allow sticking and phagocytosis of viruses. Then those DCs were cocultured with primary isolated CD^+ T-lymphocytes to allow transfer of virus, either through extracellular transfer or immunological synapsing, These transfer experiments were also performed with a lower antibody concentration (22 μg/mL) to capture a dose-dependent response in viral selection, if one exists. It also included a "no antibody" control to ensure non-specific antibody interactions didn't artificially influence selection and a "DC only" control to delineate viral sticking phenotypes from viral transfer phenotypes. **FIGURE 6** presents a simplified view of this experimental setup.

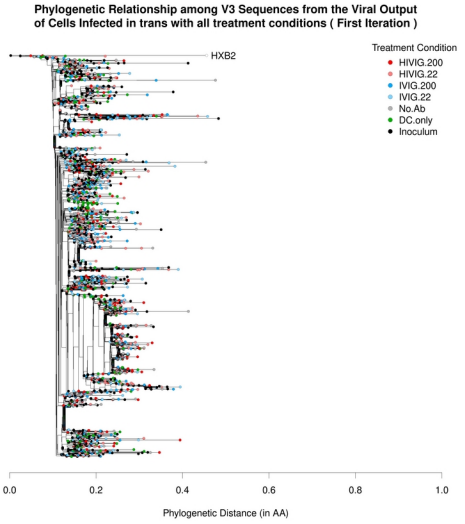
Cells were challenged with a virus dose containing 10 ng of p24 equivalents, and this resulted in extremely high copy numbers that could not be accurately quantified through qRT-PCR (data not shown). Best estimates place the average *HIVgag* copy number in each of the conditions at > 1,500,000 cp/μL. Given the high number of viral sequences and sequence features (i.e. unique sequences), phylogenetic analyses were performed in performed in 5 iterations with a random sample of 500 sequences from each treatment condition per iteration and 1,000 sequences from the inoculum per iteration. All trees showing phylogenetic relationships among viral sequences for each iteration performed hitherto indicated similar clustering patterns of randomly sampled sequences (data not shown); thus, there was no evidence of sampling bias during each iteration. Therefore, the first iteration was used as a representation of all iterations.

Phylogenetic trees showing relationships among sampled viral sequences indicate no discernible bottleneck resulting from any of the treatment conditions regardless of amplicon or iteration (**FIGURE 30**). The numbers of sequences represented in each observed lineage does not change with changes in antibody treatment (**FIGURE 31A**) or concentrations (**FIGURE 31B**). There were also no apparent differences in representation of sequences at different nodes/lineages when comparing the IVIG control with the no antibody control (**FIGURE 31C**), suggesting that non-specific antibody interactions do not influence infection selectivity. However, on cursory examination of the trees, it is unlikely that the experimental application of this model as designed was insensitive to detect infection selectivity, let alone a genetic bottleneck.

A



B



C

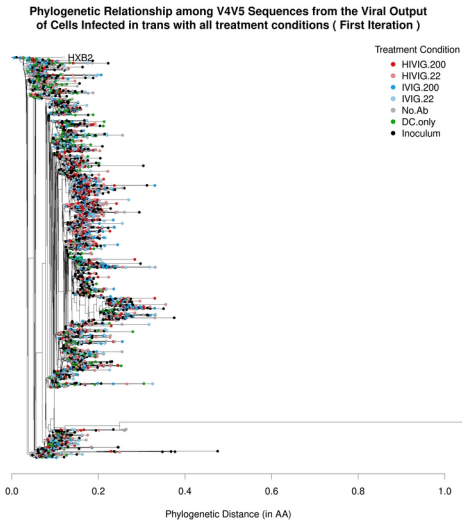
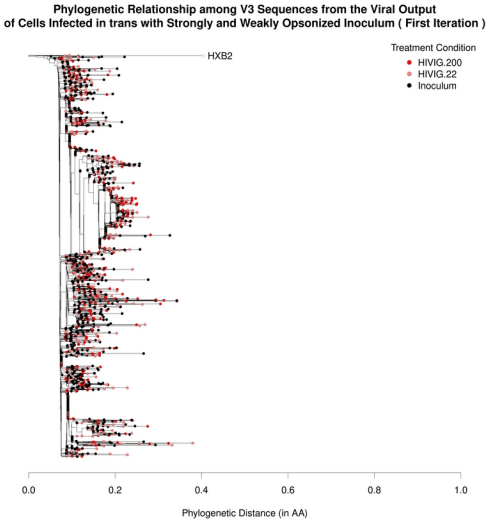


Figure 30: Phylogenetic trees of V1V2, V3, and V4V5 amplicons for each experimental condition. Maximum likelihood trees were developed for each *in vitro* infection experiment based on V1V2 (A), V3 (B), and V4V5 (C) alignments. Tips are colored by treatment condition and IgG concentration. HIVIG refers to polyclonal (non-neutralizing) IgG anti-HIV while IVIG refers to the polyclonal isotype control. HIVIG.200 refers to opsonization by 200 µg/mL HIVIG. HIVIG.22 refers to opsonization by 22 µg/mL HIVIG. IVIG.200 refers to opsonization by 200 µg/mL IVIG. IVIG.22 refers to opsonization by 22 µg/mL IVIG. No.Ab refers to the "No antibody" control. "DC.only" refers to the "DC only" control, without coculture with T-lymphocytes. Inoculum refers to the starting challenge sample.

A

B



C

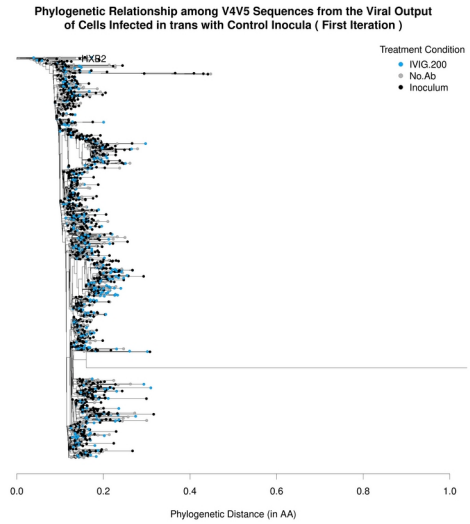


Figure 31: Representative phylogenetic trees of sequences from HIVIG and IVIG groups (A), low and high concentrations of HIVIG (B), and both IVIG and "No Antibody" controls (C). Maximum-likelihood trees were constructed for output viruses from cells infected with opsonized virus and non-opsonized virus (A). The tree in A is shown for the V1V2 alignment as an example. The same type of tree was generated to show whether changes in antibody doses led to changes in antibody-mediated selection (B). The tree in B is shown for the V3 alignment as an example. The final tree here was generated for the comparison between an IVIG treatment and the "No antibody" group as a control for non-specific antibody interactions (C). The tree in C is shown for V4V5 as an example.

"Sticky" viruses that do not transfer are a unique test of this model system

One purpose of the "DC only" and "no antibody" controls is to determine whether a bottleneck of transferring variants could be detected from the pool of variants that stick to the surface of DCs or that reside in endosomes prior to synpatic transfer. This can be detected by phylogenetic comparisons of variants from both control groups, and at least one (at most two) levels of bottlenecks may be observed in the same way it could be detected in the *in vivo* model: the first level would be a restrictive bottleneck from the inoculum to the "DC only" group and the second would be a restrictive bottleneck from the "DC only" group to the "no antibody" group. An example phylogenetic tree depicting theoretical first and second level bottlenecks from a sample inoculum is shown in **FIGURE 32**.

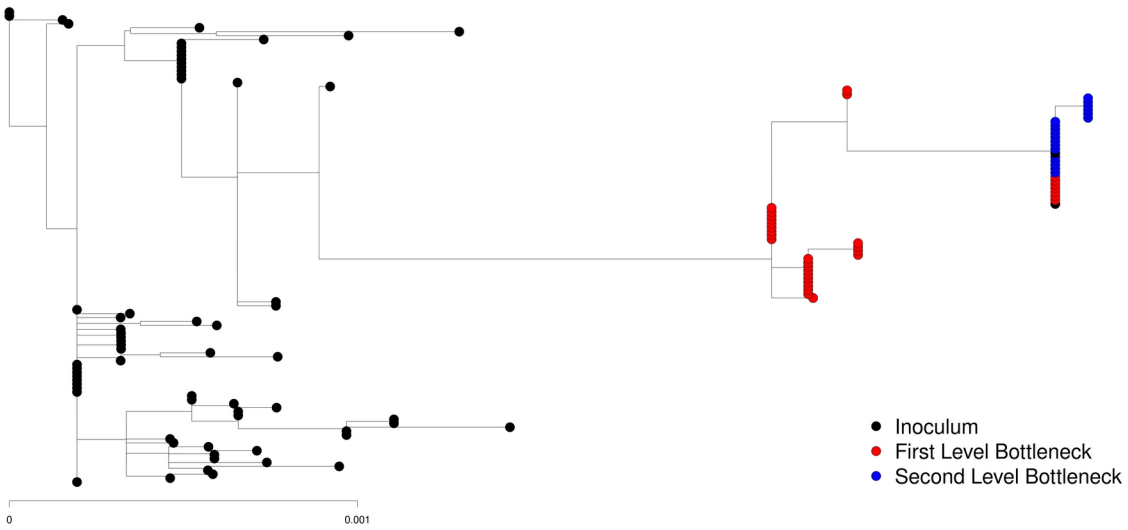


Figure 32: An example phylogenetic tree depicting a two-leveled bottleneck. This maximum-likelihood tree was constructed in order to demonstrate the concept of a two-level bottleneck. During the viral challenge of cells (i.e. dendritic cells) with a particular viral inoculum, the output virus from this challenge may cluster to a single lineage, as shown in the tree above by red tips. Variants in this major lineage may bind to or infect the challenged cells, and the variants within this lineage may be divergent from one another by a star-like phylogeny, producing minor variants within the first bottleneck. After coculturing the challenged cells with a different type of target cell (i.e. T-lymphocytes), viruses may selectively transfer from one of the minor lineages of the first bottleneck, leading to second bottleneck that was selected or restricted from variants bound to the challenged cells. This theoretical selection pattern was searched throughout data sets in order to determine if *trans*-infection T-lymphocytes is selective not just from an inoculum but from a smaller subset of viruses bound to the surface of dendritic cells.

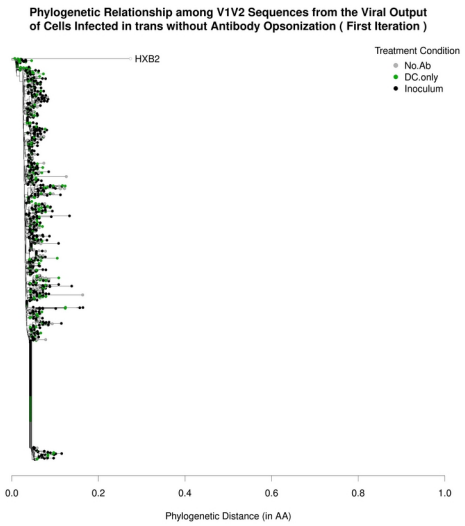
No typical bottlenecks were observed in these comparisons (**FIGURE 33A**). However, two unexpected populations of viruses were detected that did not follow expected selection patterns, from the standpoint of V3 (**FIGURE 33B**) and V4V5 (**FIGURE 33C**) amplicons. The first population of variants cluster to a particular lineage of the "DC only" group that does not appear in the "no antibody" group (or if they do, they are represented as a scant minority). Its selection pattern appears to be bottleneck-like as individual variants from this lineage originated from one or two non-majority variants in the inoculum. These viruses can be functionally described as viruses that adhere to DCs (either on the surface or in endosomes) but are not readily or actively transferred to susceptible T-lymphocytes; while this conjecture hasn't been demonstrated by a separate laboratory experiment it can be inferred from the tree itself. The viruses in this lineage are denoted by green bracket in phylogenetic trees.

The second population does not appear to be a bottleneck-like restriction of variants, as each node occupied by a variant in this lineage is co-occupied by a member from the inoculum, suggesting independent infections by multiple variants likely in a stochastic manner. These variants cluster to multiple lineages (denoted by a gray bracket) in the "no antibody" group that were only scantly present (if at all) in the "DC only" group. This suggests that there existed a population of viruses from these transfer experiments that circumvented the *trans*-infection route by directly infecting T-lymphocytes. The presence of these viruses indicates that DCs challenged with the inoculum were insufficiently washed prior to coculture, raising doubts as to whether sequenced variants are really the result of *trans*-infection. While these data do suggest the presence of viruses in the challenge stock that preferentially infect T-lymphocytes directly rather than doing so in *trans*, selection of particular polymorphisms in this study may not be attributed

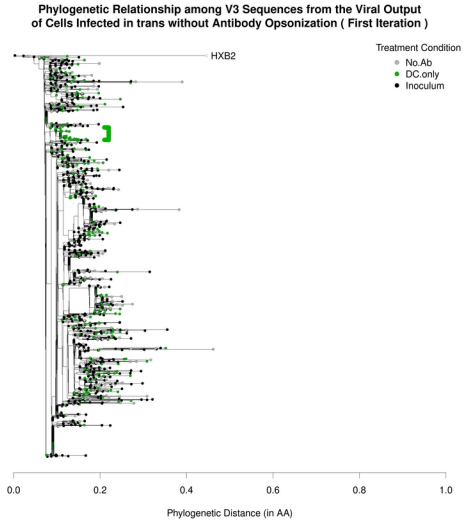
to a *trans*-infection mechanism. Due to this possibility, an exercise of caution is necessary in interpreting these results.

The subpopulation of variants from the "DC only" group that adhere to DCs without being involved in *trans*-infection of T-lymphocytes was of particular interest because infection of and adherence to monocytes have been studied previously [Weissman et al., 1995; Geijtenbeek et al., 2000; Hong et al., 2002; Kwon et al., 2002; Binley et al., 2006; Nabatov et al., 2006; Shan et al., 2007; van Montfort et al., 2008; Shen et al., 2010a; Chen et al., 2013; Bertram et al., 2019; Jarvis et al., 2019; Perez-Zsolt et al., 2019; Koh et al., 2020], and they present an opportunity to evaluate the predictive power of this model. Cellular adherence of HIV-1, particularly in binding to surface receptors on DCs, depends on length, net charge, and levels of N-glycosylation of variable loops [Nabatov et al., 2006]. For this reason, the aforementioned characteristics of variable loops V3, V4, and V5 were compared among the viruses in the two controls and the inoculum in order to determine whether selection favors particular phenotypes.

A



B



C

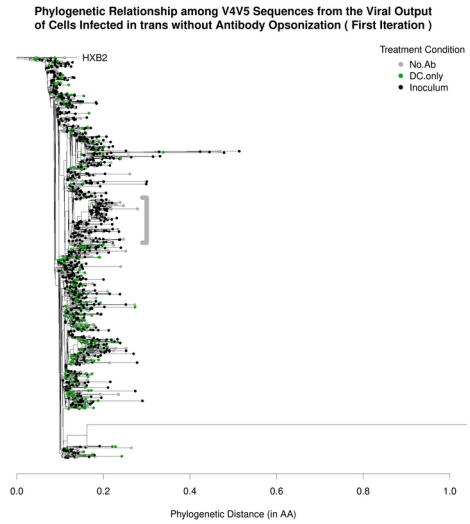


Figure 33: Phylogenetic trees of sequenced V1V2, V3, and V4V5 variants from the "No antibody" group and the "DC only" group. Maximum-likelihood trees were developed showing the phylogenetic relationships of variants from the "No antibody" group, "DC only" group, and the inoculum. These trees were based on V1V2 (A), V3 (B), and V4V5 (C) alignments. The green bracket in B indicates a viral lineage preferentially infecting dendritic cells. The gray bracket in C shows the presence of viruses that infect T-lymphocytes preferentially without interacting with dendritic cells (an indication of insufficiently washed dendritic cells).

V4 lengthening and N-glycosylation are favored by the selection pressures inherent to the "DC only" group

The polypeptide lengths of V3, V4, and V5 were not significantly different among the treatment groups and inoculum (**FIGURE 34A**), although there was a slight, albeit insignificant trend toward a longer V4 among viruses in the "DC only" group compared to both the "No antibody" group and the inoculum stock, which agrees with a previous report showing high high amino acid diversity and length in V4 [Hoffman et al., 2002]. Closer examination of the polymorphisms that contributed to this longer variable loop revealed two major variants with a net increase in V4 by approximately fourteen (F391_N392insWK/S393_T394insMVEQMHEDIIS/W395_G404insQSLKPCVKLTPLCVTL/ G404_K405insNYVKD) and six amino acids (W395_G404insNGTWYW), respectively. Variants with these insertions were generally favored for DC sticking/infection ($p < 2 \times 10^{-14}$, Fisher's exact test with Bonferroni correction). It should be noted that the variants with a 14-AA longer V4 were linked to deletions of 20 residues in V4 and 6 and 5 residues from gp120 domains C3 and C4, respectively. In sum, the selected insertions in V4 only increases the length of the entire V4V5 amplicon (containing gp120 domains C3V4C4V5C5) by approximately three amino acids. The insertions here might be important for selective adherence/infection of DCs without compromising the fitness of the virus in question due to extreme changes in gp120 length. In fact, the variant with a shorter V4 insertion (W395_G404insNGTWYW) was not linked to upstream or downstream deletions, which could be a further indication that there exists a constraint on gp120 length. Indeed, lengthening of the variable loops have previously been shown to impair viral infectivity and replicative fitness, despite being selected as a means of

immune evasion during chronic and late acute infection [Ganusov et al., 2011; Gnanakaran et al., 2011; Ochsenbauer et al., 2012; Hamoudi et al., 2013; Gonzalez et al., 2015; Lamers et al., 2019].

On the question of ionic charge in the variable loops, analysis showed that V3 generally had a more net positive charge and V4 and V5 both had a more net negative charge (**FIGURE 34B**), although less than 25% of V4 and V5 variants were cationic. These results are in agreement with another study that examined net ionic charge characters of HIV-1 gp120 variable loops among patients with severe infections [Seclén et al., 2011]. The distribution of net charges in V4 and V5 were statistically more constrained among the "DC only" group compared to the "No antibody" control and the inoculum (V4: $p < 5 \times 10^{-9}$ between "DC only" and "No antibody" and $p < 5 \times 10^{-5}$ between "DC only" and inoculum; V5: $p < 0.01$ between "DC only" and "No antibody" and $p < 0.05$ between "DC only" and inoculum; Welch's two-tailed t-test). There may be correlation (not tested here) between a constraint on ionic charge character of V4 with its lengthening; Lamers et al. [2019] noted a similar occurrence in the V1 loop, wherein a longer V1 tended to be more ionically constrained (i.e. insertions of cationic amino acids were balanced with insertions or anionic amino acids). Although such correlation has yet to be shown for V4 or V5 loops, there may exist selection pressure that preserves virus infectivity of and/or adherence to DCs by maintaining ionic charge character with increases in V4 length. In fact, mutations in one variable loop often correlate with mutations in a different loop in order to maintain viral fitness during *trans*-infection [Nabatov et al., 2006], thus effects measured here for V4 may have broader impacts on other variable loops that cannot be captured given the current scope of this project.

Finally, and perhaps more crucial to DC adherence, the N-glycosylation character of V3, V4, and V5 loops among viruses in the two control groups and the inoculum was examined (**FIGURE 34C**). More than half of the viruses in the "DC only" group contained six or more potential N-glycosylation sites in V4. Viruses with that many N-glycosylation sites in the "no antibody" group and the inoculum made only 25% of the population of viruses in those groups. In fact, the majority of viruses in the latter groups contained fewer potential N-glycosylation sites. This suggests that in the absence of $CD4^+$ T-lymphocytes, there may be selection pressure favoring viruses with a higher N-glycosylation signature in order to infect and persist among a monocyte-only population. Mannosylation and complex glycan shielding of gp120, particularly in V1 and V3, contribute to a higher binding affinity to DC-SIGN, siglec-1, and other lectins on monocytes [Becker et al., 2004; Nabatov et al., 2006; Izquierdo-Useros et al., 2014; Bertram et al., 2019]; the benefits of this kind of post-translational modification of the HIV viral spike is likely not limited to just evading antibody recognition but may include enhancement of monocyte infection and lymphocyte *trans*-infection. This is a likely explanation for the observed bottleneck in the "DC only" treatment group. It should be noted that Nabatov et al. [2006] found evidence that DC infection and adherence is slightly more associated with N-glycosylation of V1 and V3, although this wasn't found to be necessary to promote infection of DCs and that any number of N-glycosylated residues on any variable loop is generally sufficient to enhance infection of these cell types. In the case of the current study, however, N-glycans of V4 could be particularly more permissive to DC infection/adherence than the N-glycans of V3 or V5.

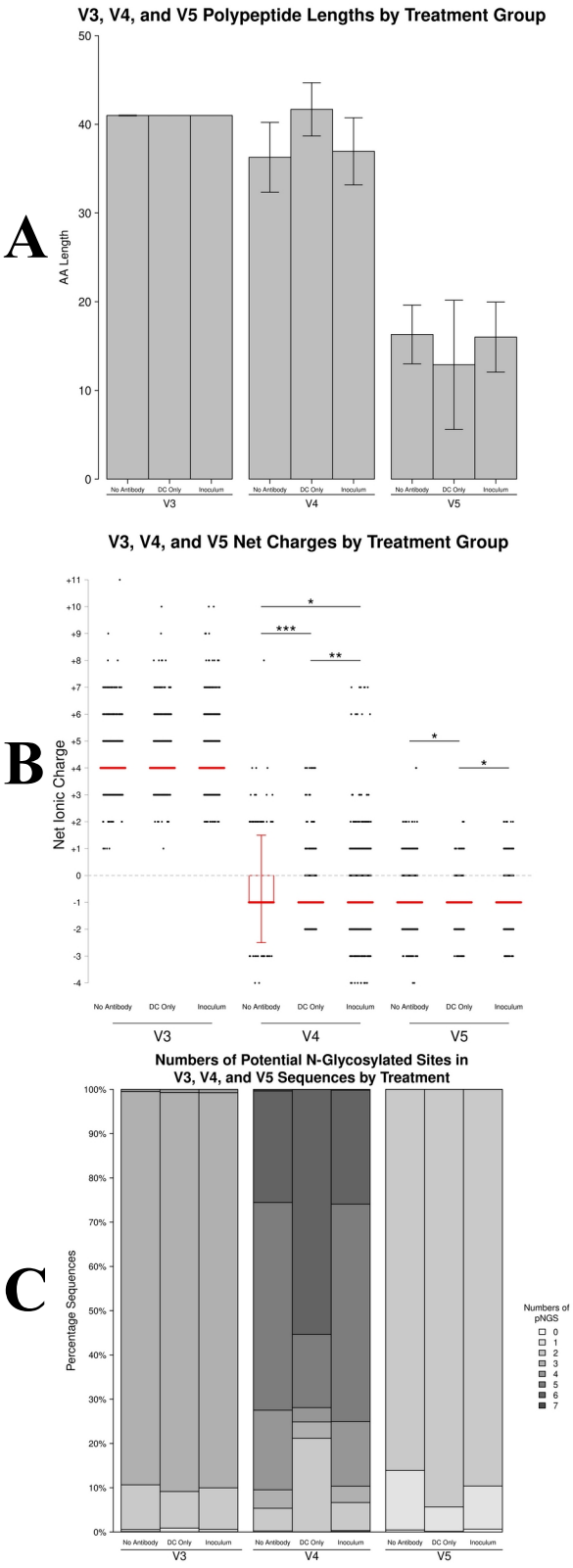


Figure 34: Functional characteristics of V3, V4, and V5 loops among variants of the "No antibody", "DC only," and "Inoculum" groups. (A) the polypeptide lengths of V3, V4, and V5 loops were counted for each infection condition ("No antibody," "DC only," and inoculum) and shown here as mean ± standard deviation. (B) The net charges of each variable loop was also analyzed and shown here as a box plot with the median (red line) and interquartile ranges (in the red box). Error bars indicate the minimum and maximum of each quartile analyses. Outliers beyond the minimum and maximum of each analyses are displayed as black points. (C) The numbers of potential N-glycosylated sites were enumerated for each read in each infection condition for the V3, V4, and V5 loops and displayed as percentages of each set of sequences.

* $p < 0.05$, Welch's t-test
** $p < 0.0005$, Welch's t-test
*** $p < 0.000005$, Welch's t-test

Co-receptor tropism is insufficient to explain mechanism of selection toward DC infection

Sliding window analysis showed selected polymorphisms with 6 to 14 amino acid insertions into the V4 loop that were favored by DCs in the absence of T-lypmhocytes. However, the bottleneck that was observed in the "DC only" group was associated with the V3 amplicon, not V4V5. Furthermore, among the five iterations performed for this data set, a questionable "DC only" bottleneck in the V4V5 amplicon was observed only for one iteration (as opposed to an obvious bottleneck in all five iterations for the V3 amplicon) (data not shown). This does not strictly mean that selection for the V4 phenotypes, as observed, were artificial (it also does not mean that the apparent distributions of polymorphisms at any genetic region in these experimental groups could not have arisen by linkage disequilibrium with other undetected polymorphisms); rather, the selected phenotypes of V3 were more directly associated with the "DC only" bottleneck is of a smaller scope than what is measurable in broader characters, such as amino acid length, ionic charge, or levels of N-glycosylation. Selection pressure acting on and near V3 is likely more refined and centered on single point mutations rather than an entire variable loop.

In speculation on the mechanism of selection, it is important to take the function of the V3 stem loop into consideration. The V3 and V4 domains contain binding sites to co-receptors that are necessary for the infection machinery [Hoffman et al., 1998; Hoffman et al., 2002; Biscone et al., 2006; Tamamis and Floudas, 2013; Tamamis and Floudas, 2014], but polymorphisms in the V3 region directly correlate with the R5- and X4-tropism of HIV-1 [Cardozo et al., 2007]. Generally, co-receptor affinity is determined by the "11/24/25 rule" formulated by Cardozo et al. [2007] through a process of induction: X4-tropic viruses tend to

carry positive charges at positions 11, 24, and 25 of the V3 region (identical to positions 306, 321, and 322 of gp120) while R5-tropic viruses carry neutral or negative charges at those positions. Based on sliding window analysis with this insight in mind, over 99% of the viruses in the inoculum were R5-tropic, amounting to a ratio of 1 X4 virus for every 110 R5 viruses, which is typical considering the clinical isolate used here (92US657). There was no evidence of selective sway from this major phenotype in any of the treatment conditions ($p > 0.2$, Fisher's exact test without Bonferroni correction). Given a previous report suggesting DCs favor X4-tropic HIV-1 rather than the R5 variant in *trans*-infection of T-lymphocytes [van Montfort et al., 2008], it is tempting to ascribe the selection mechanism in preference of non-transferring, DC-adhered viruses to R5 tropism. This potentially new hypothesis is further predicated on a previous finding that the amino acid identity at this position modulates binding affinities in favor of CCR5 [Tamamis and Floudas, 2014] over binding to CXCR4 [Tamamis and Floudas, 2013]. However, the current experiment does not offer a reliable test of this particular hypothesis because the vast majority of viruses in the inoculum were R5-tropic, and the quantity of these viruses likely overcame any selection pressure favoring the X4 variant, if such pressure existed. In fact, the X4-tropism phenotype was represented in the "DC only" group at 66% of what was represented in the original inoculum, suggesting that a slight, albeit non-significant, purifying selection pressure may have filtered the X4 phenotype or any linked deleterious trait from the population of infectious viruses in the inoculum (data not shown). In the absence of followup experiments, the observed bottleneck in V3 among the "DC only" viruses isn't likely due to co-receptor affinity.

Two selected polymorphisms of V3 and C3 are favored by the selection pressures inherent to the "DC only" group

Infection mechanism aside, sliding window analysis of V3 variants in the "DC only", "No antibody", and inoculum viral populations did show patterns of significant selection among three polymorphisms at two different amino acid sites: 317 and 337. The "DC only" group favored an L317I substitution (DC only versus No antibody, $p < 10^{-8}$, Fisher's exact test with Bonferroni correction; DC only versus inoculum, $p < 10^{-14}$, Fisher's exact test with Bonferroni correction) and D337E (DC only versus No antibody, $p < 10^{-216}$, Fisher's exact test with Bonferroni correction; DC only versus inoculum, $p < 10^{-13}$, Fisher's exact test with Bonferroni correction), and it disfavored the L317F substitution (DC only versus No antibody, $p < 10^{-195}$, Fisher's exact test with Bonferroni correction; DC only versus inoculum, $p < 10^{-197}$, Fisher's exact test with Bonferroni correction). Although the amino acid position 337 resides in the C3 region of gp120, it is located very near V3 that it may play a role in V3-related functions. Distribution of polymorphisms at these two sites in the "No antibody" group is not significantly different from that in the inoculum ($p > 0.16$, Fisher's exact test with Bonferroni correction), suggesting that the occurrence of variants in the "No antibody" group from the challenge inoculum is the result of stochastic rather than selective infection. A graphical representation of the percentages of variants with particular polymorphisms at these two positions is shown in **FIGURE 35A**.

Interestingly, the L317I and D337E polymorphisms are linked in only a minority of the inoculum (< 0.4%), which should be represented in the "No antibody" and "DC only" groups at around the same percentage should there be no selection acting on this particular linkage. This

was only true in the "No antibody" group ($p = 0.18$, Fisher's exact test without Bonferroni correction), but linkage of these polymorphisms were represented in the "DC only" at a substantially higher percentage (~ 30%) (DC only versus inoculum, $p < 10^{-251}$, Fisher's exact test with Bonferroni correction; DC only versus No antibody, $p < 10^{-140}$, Fisher's exact test with Bonferroni correction) (**FIGURE 35B**). The substitutions for an isoleucine at position 317 and a glutamate at position 337 unlinked from one another occurred at 84.5% lower representation ($p < 10^{-77}$, Fisher's exact test with Bonferroni correction) and 335% higher representation, respectively, in the "DC only" group compared with the inoculum ($p < 10^{-7}$, Fisher's exact test with Bonferroni correction). This implies that the L317I mutation greatly impairs viral fitness while the D337E mutation slightly enhances it. When both substitutions are linked, however, not only does the D337E substitution appear to nullify the viral fitness loss incurred by L317I, there may also be a synergistic enhancement in viral fitness as evidenced by the significant increase in representation of the linked mutations 11.8 fold and 10.4 fold of the unlinked mutations (L317I and D337E, respectively) in the "DC only" group. Furthermore, the enhancement of percentages of the linked polymorphisms (or some related genetic signature hitherto undiscovered) may be the result of direct infection of DCs rather than strict adherence to surface lectins. This possibility may be seriously considered in the experimental condition where vulnerable T-lymphocytes were not present, and thus in a competition between infectious variants, the viruses with the highest fitness may be the types of viruses that preferentially infect monocytes. It could also explain why the variants that were strongly selected in the "DC only" group were not selected in the "No antibody" group where the presence of T-lymphocytes could have shifted selection pressures in such a way that viruses adapted to *trans*-infection would out-compete DC-

infecting viruses. While it is difficult to ascertain whether the enhancement of fitness was strictly due to the linkages of these particular polymorphisms or linkages to polymorphisms in other variable loops, the changes in the representation of polymorphisms at positions 317 and 337 from the inoculum to the "DC only" group are suggestive of some form of selection pressure acting on these positions that may be relevant for viral infection of DCs.

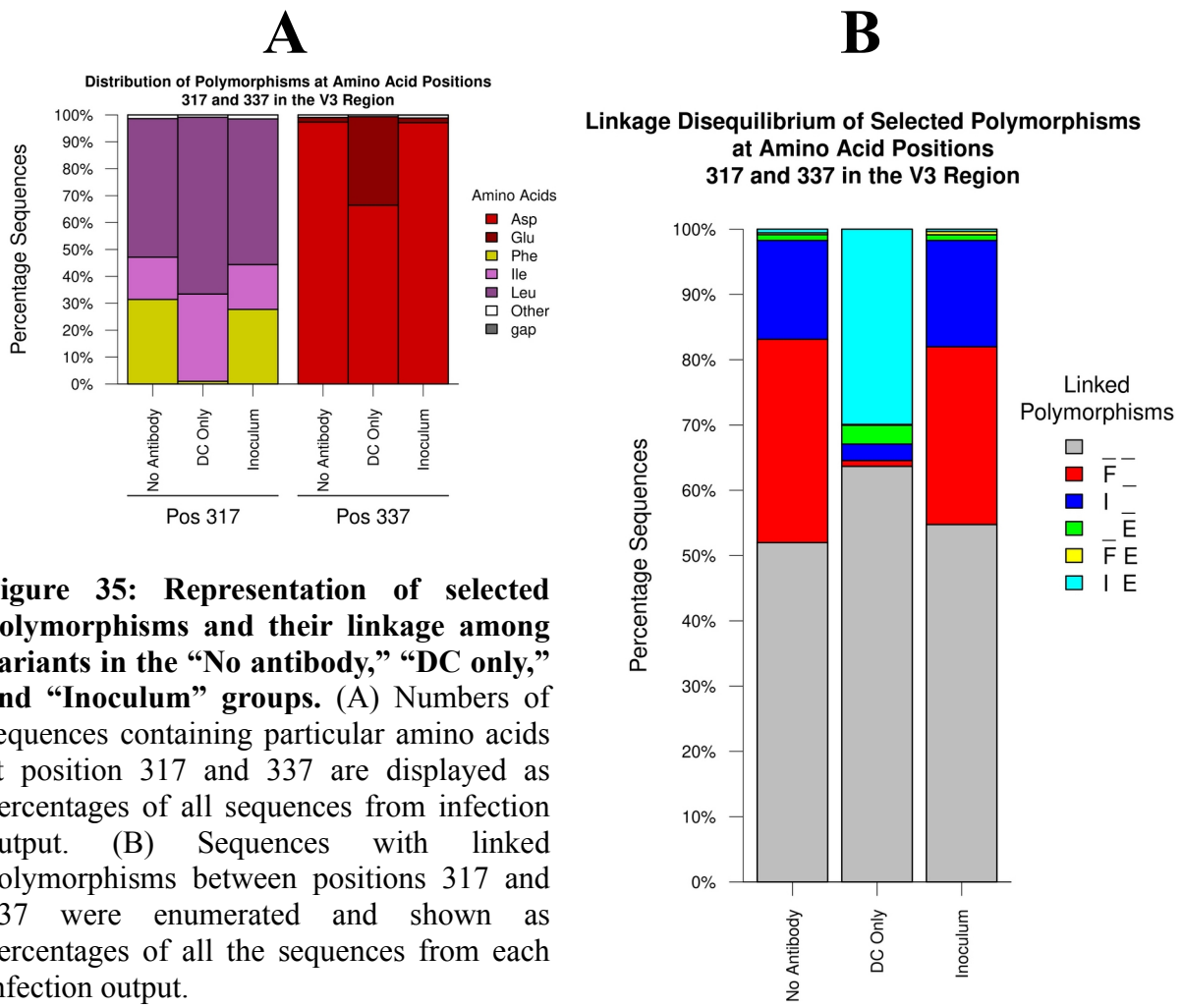
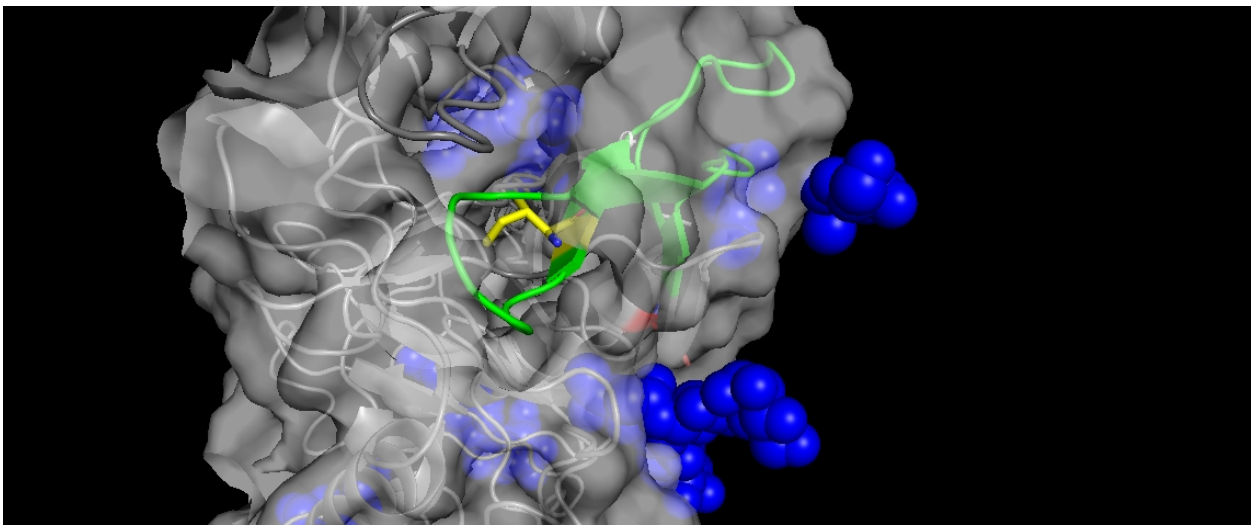


Figure 35: Representation of selected polymorphisms and their linkage among variants in the "No antibody," "DC only," and "Inoculum" groups. (A) Numbers of sequences containing particular amino acids at position 317 and 337 are displayed as percentages of all sequences from infection output. (B) Sequences with linked polymorphisms between positions 317 and 337 were enumerated and shown as percentages of all the sequences from each infection output.

The amino acid position 317 is variable among clades A, B, and C viruses [Huang et al., 2005], and it tends to fluctuate between hydrophobic amino acids like phenylalanine, tyrosine,

leucine, and isoleucine (although gaps in this position have also been reported) [HIV Sequence Compendium, 2002; HIV Sequence Compendium, 2021]. Based on its placement in a hyrodphobic pocket of gp120, the hydrophobicity of amino acids at this position may be intended to stabilize the gp120 tertiary structure with either aliphatic or aromatic amino acids depending on the chemical properties of proximal amino acid side chains (**FIGURE 36A**). The 337 position residing in the conserved region C3, on the other hand, tends to be far more variable with neutral and charged amino acids as well as gaps, aspartate and glutamate having the highest representation among reported sequences [HIV Sequence Compendium, 2002; HIV Sequence Compendium, 2021]. The placement of this amino acid position faces away from the gp120 core (**FIGURE 36B**). Its spatial orientation and predominance toward negatively charged residues may be suggestive of roles in forming salt bridges with surface proteins on monocytes; such an interaction may play an important role in the initial infection of DCs, though not exclusively, contributing to its spread during early infection. The mechanism of DC infection from the standpoint of these mutations, however, remains unknown. Therefore, the literature was searched for a probable mechanistic explanation.

A



B

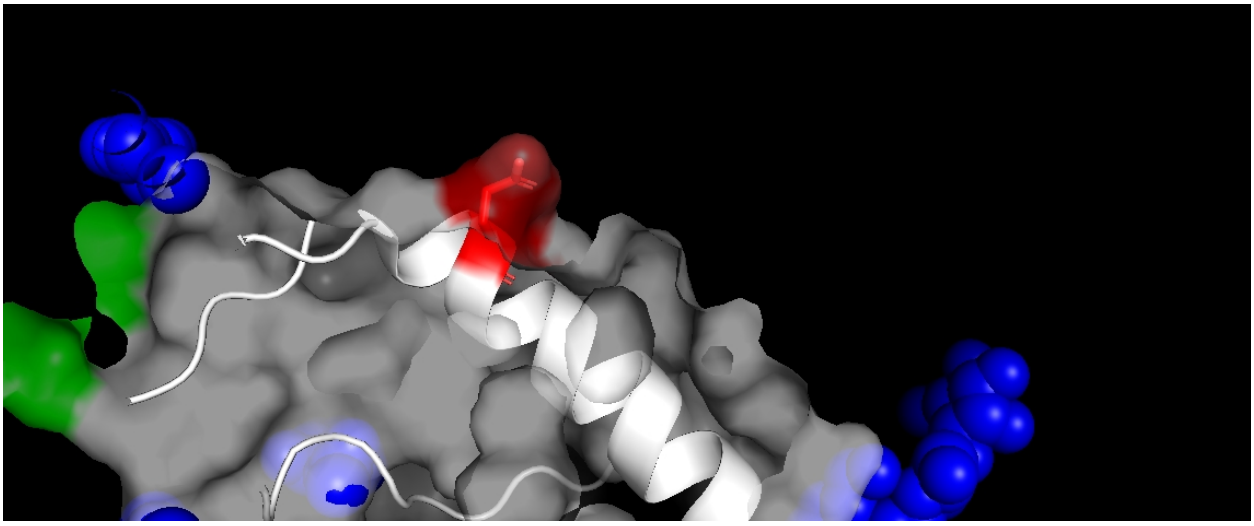


Figure 36: Spatial orientation of polymorphisms I317 and E337 on gp120. Space-filling cartoons of the amino acid polymorphisms I at position 317 (A) and E at position 337 (B). The V3 loop is shown in green, while the conserved region C3 is shown in white. The amino acid I is shown in yellow in A while the amino acid E is shown in red in B. N-glycans are shown in blue. *Sourced from rcsb.org: ref # 7SC5.* [Yang et al., 2022]

Protein-BLAST searches for the V3 consensus sequence (from the inoculum) with the 317 and 337 positions substituted with an I and E, respectively, returned very few accession results wherein both mutations were linked. The only GenBank entries found (related to GenBank accession AMD05787) with the linked mutations were from an unpublished report under the title "Broad activation of latent HIV-1 *in vivo*" by Barton et al. An E at the 337 position was associated with another accession number (QCF44990) regarding a similar topic: an unpublished report under the title "Tracking the origin of latent infection in T-cell subsets in HIV-infected individuals on antiretroviral therapy through envelope sequence and function" by Roche et al. One interesting query result (AFV65611) from this search pointed to a 2012 publication by Poon et al. [2012] wherein a *de novo* I and V mutations at position 317 correlated with the emergence of CXCR4 tropism among variants in a chronically infected human. While this substitution may have given rise to co-receptor switching in one person, the same mutation was insufficient to do the same in others [Poon et al., 2012]. Furthermore, an E at position 337 was present in submitted sequences by these authors, thus it is unlikely that this residue was responsible for co-receptor switching, let alone a possible mechanistic explanation for DC infection in the current study. Another entry (AIX95706) returned by the BLAST search referred to a study published by Sturdevant et al. [2015]. The authors traced the phylogenetic relationship of viruses replicating in the CNS, cerebral spinal fluid (CSF), and blood plasma compartments of 33 infected patients. This study was considered for a possible mechanism to explain the DC-infecting phenotype of E337 because microglial infection and brain compartmentalization was associated with mutations in the V3 region of macrophage-tropic viruses [Rossi et al., 2008]. A glutamate at position 337 was associated with viral variants equilibrated among the three

compartments in one patient (9018) from the 2015 study; however, the 317 position was identified with a phenylalanine, not an isoleucine. Complicating the connection further, the E337 residue wasn't found in variants among other patients exhibiting similar compartment equilibration, complicating the link between this polymorphism and infection of monocytes.

Despite an exhaustive protein-BLAST search of V3 sequences, it remains unclear how the I317 and E337 polymorphisms confer the DC-infecting phenotype, if at all, to the variants in the current study. It also remains unclear whether interactions with DC surface proteins contribute to the selection pressures observed here, whether or not those pressures are acting on these polymorphisms or some other unknown linked polymorphism. It is clear, however, that these particular variants were selected as to increase their representation in the progeny of DC-infecting viruses nearly 75 fold from their numbers in the inoculum. This may be a strong indication of transmission selectivity captured by the *in vitro* model that would provide researchers further insights into the mechanisms that contribute to the genetic bottleneck.

Chapter 5: Findings and Speculations—Complement

The Effect of Complement in Variant Selection during Infection

Using the model to test complement-opsonization as a potential selection pressure for HIV-1 transmission

The literature on HIV-1 infection is replete with examples of ADE [Robinson et al., 1987; Takeda et al., 1988; Homsy et al., 1989; Kozlowski et al., 1995; Trischmann et al., 1995], a phenomenon where antibodies at non-neutralizing levels or efficacy enhance the infectivity of opsonized virus through FcγR engagement [Homsy et al., 1989; Trischmann et al., 1995; Forthal et al., 2012] or indirectly by strengthening the interaction between V3 and CCR5 [Guillon et al., 2002]. The enhancement of infectivity by antibody-opsonization was the empirical basis for the use of antibodies in testing Hypothesis #3 in conjunction with Hypotheses #1 and #2: that the antibody mediation between virus and host cells may enhance infectivity of particular variants over others through mucosal penetration and/or *trans*-infection. ADE is not the only measurable phenomenon of infection enhancement by host factors; the immunological complement system has been reported to do the same in antibody-dependent [Robinson et al., 1989; Robinson et al., 1990; Tremblay et al., 1990; Willey et al., 2011] and -independent manners [Montifiori et al., 1989; Ebenbichler et al., 1991; Bouhlal et al., 2007; Nijmeijer et al., 2021]. It was on this basis that a second test of Hypotheses #2 and #3 was performed: complement opsonization of virus may select for particular viral variants over others during direct and *trans*-infection.

This experiment was performed using the same methodology employed for the test of antibody-mediated selection in the prior experiment, except with modifications to improve on its design and implementation. To compensate for the extremely high viral output that likely

overcame the weak antibody-propagated selection pressures, the challenge dose of the 92US657 utilized in this experiment was diluted from 10 ng to 0.1 ng per 100,000 target cells. This experiment also tested direct infection of primary T-lymphocytes as well as *trans*-infection under stimulated and unstimulated conditions. Where the challenge inoculum was incubated in the presence of antibodies at two particular concentrations, the inoculum here was incubated in media with human complement from three HIV-negative donors. This incubation allowed opsonization by complement, and for the control human complement was heat-inactivated to denature proteins that may interact with virus. Opsonized virus was incubated and allowed to adhere or be phagocytosed by DCs. Those DCs were then cocultured with stimulated or unstimulated T-lymphocytes, or opsonized virus was applied directly to stimulated T-lymphocytes. The purpose here was to test how complement-opsonization influenced infection selectivity of HIV-1 during *trans*-infection and direct infection.

Complement opsonization of challenge virus resulted in slightly lower infection (though not significantly) than did the non-opsonized control during *trans*-infection of unstimulated T-lymphocytes (**FIGURE 37**) while complement opsonization did not lead to changes in either direct or *trans*-infection of stimulated cells. The non-significant trend toward lower infectivity of the opsonized virus compared to the "naked" virus during *trans*-infection of unstimulated cells suggests that complement-dependent enhancement of infectivity did not occur, which has been observed previously in other studies with a similar experimental design [Bouhlal et al., 2007; Willey et al., 2011; Nijmeijer et al., 2021].

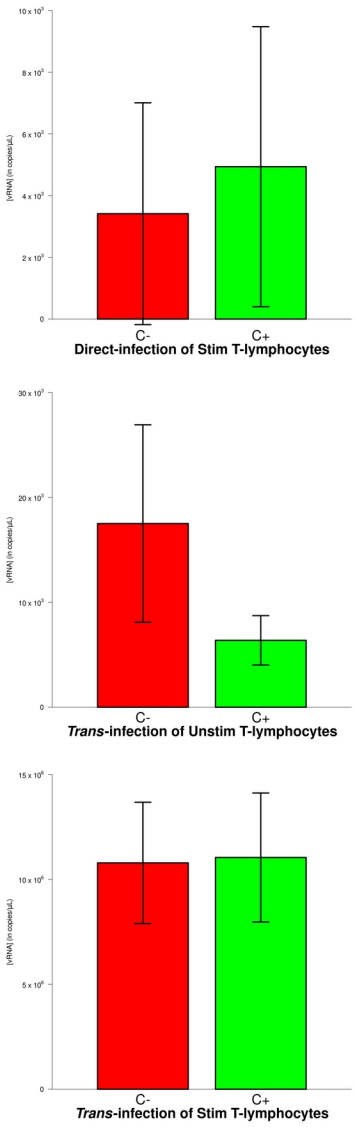


Figure 37: *HIVgag* vRNA copy numbers of viral output from directly and *trans*-infected with or without complement opsonization of virus (assessed by RT-qPCR). vRNA isolates from *in vitro* infection experiments with complement-opsonized ("C+", green) and non-opsonized ("C-", red) viral inocula were analyzed for *HIVgag* copy numbers. (top) Direct infection of stimulated T-lymphocytes, (middle) *trans*-infection of unstimulated T-lymphocytes, and (bottom) *trans*-infection of stimulated T-lymphocytes. Data are represented as mean ± standard deviations (n = 3).

However, based on phylogenetic trees showing the relationship of viral variants in each of the infection conditions, there appears to be some patterns of selection among treatment conditions. There is strong bottleneck-like complement-mediated selection on the basis of *trans*-infection of unstimulated T-lymphocytes among V3 (**FIGURE 38A**) and V4V5 amplicons (**FIGURE 38C**). Paradoxically, this selection appears to be attenuated with stimulation of T-lymphocytes during

trans-infection, as detected in the V3 (**FIGURE 38B**) and V4V5 amplicons (**FIGURE 38D**); this attenuation of selection pressure is evidenced by a lower representation of treatment conditions at two phylogenetic lineages. A stronger bottleneck was observed among directly infected stimulated T-lymphocytes in the absence of complement opsonization detected in V1V2 (**APPENDIX 16**) and V3 (**APPENDIX 17**) but not V4V5 (**APPENDIX 18**) amplicons; however, this is likely an artefact viral variants adapted to multiple laboratory passages grown

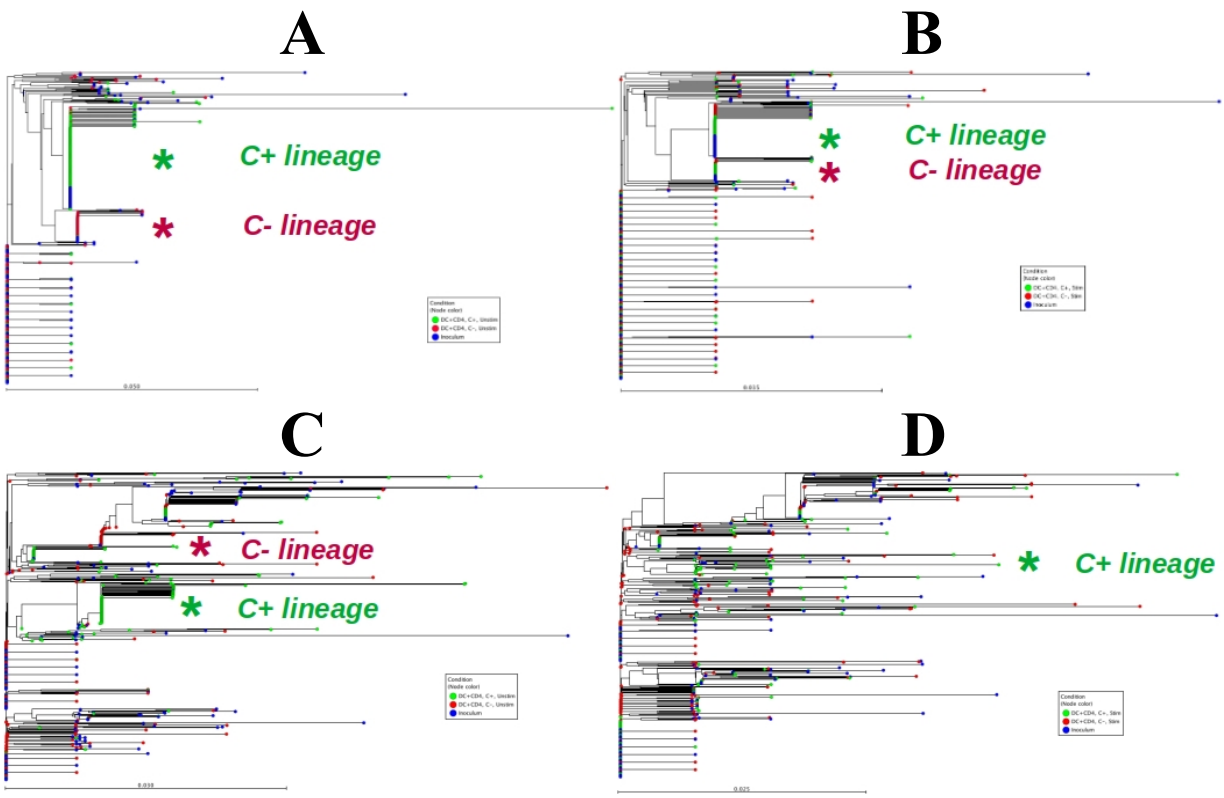


Figure 38: Phylogenetic trees of viral output from trans-infection of stimulated and unstimulated T-lymphocytes by inoculum opsonized with complement or without complement. Neighbor-joining trees were constructed using QIAGEN CLC Genomics Workbench for V3 (A and B) and V4V5 (C and D) alignments. Viral outputs represented in each tree are from *trans*-infection of unstimulated T-lymphocytes (A and C) and stimulated T-lymphocytes (B and D). Trees depict the relationship of 500 randomly sampled variants from each infection condition in replicate C. Phylogenetic trees are representative of trees from the other two replicates.

under the same conditions. This artefact may not be relevant to understanding complement-mediated selection, but its applicability in the accuracy of this model will be discussed.

Complement-mediated selection is modulated by stimulating T-lymphocytes prior to *trans*-infection

As described earlier, *trans*-infection of unstimulated T-lymphocytes revealed two lineages from the inoculum that were selected along the complement axis (i.e. with or without complement), which were labeled as the C+ and C- lineages, respectively. Based on the phylogenetic relationships among the V3 region of HIV-1 variants, viruses in one lineage were favored by complement opsonization while the other lineage were favored in the absence of opsonization ($p < 10^{-42}$, Fisher's exact test without Bonferroni correction). The variants from the inoculum are represented at the complement-selected and non-complement-selected lineages at 23% and 18%, respectively, which is significantly lower than the majority variants at each lineages ($p < 10^{-6}$ and $p < 10^{-15}$, respectively, Fisher's exact test without Bonferroni correction). The same pattern of selection was found in the V4V5 regions as well (C+ vs C-: $p < 10^{-27}$; C+ vs I: $p < 0.05$; C- vs I: $p < 10^{-3}$; Fisher's exact test, without Bonferroni correction). This suggests there existed a strong selection pressure that favored particular HIV-1 quasispecies from the inoculum on the basis of whether or not it was opsonized by human complement at both V3 and V4V5.

The selection pressure here was observed among unstimulated cells *trans*-infected by DCs. Although complement-mediated selection under these conditions did not strictly result in a bottleneck, adjusting the experimental parameters (i.e. stimulation status of T-lymphocytes,

challenge dosage) could have modulated the selection pressure in such a way as to result in one. To determine whether stimulation status affected complement-mediated selection toward or away from either of these lineages, the above analysis was repeated by examination of the phylogenetic trees of complement- and control-opsonized virus infecting stimulated T-lymphocytes through *trans*-differentiation over the V3 (**FIGURE 38B**) and V4V5 regions (**FIGURE 38D**). Prior stimulation of T-lymphocytes resulted in 19% reduction and 72% reduction in the representation of the complement-selected variants at the C+ lineage and control variants at the C- lineage, respectively, in the V3 region. These reductions were found to be statistically significant by Fisher's exact test without Bonferroni correction ($p < 10^{-4}$ and $p < 10^{-10}$). With regard to the V4V5 region on the other hand, stimulation did not result in a reduction of complement-opsonized viruses at the C+ lineage ($p = 1$, Fisher's exact test without Bonferroni correction), but it did lead to a complete disappearance of the control viruses at the C- lineage despite a non-significant p value ($p = 1$, Fisher's exact test without Bonferroni correction). These changes can be better visualized through the pie diagrams in **FIGURE 39**. It should be noted that the statistics measuring the distribution of virus treatments and stimulation status of target cells at the C+ and C- lineages may be influenced by an as yet unknown confounding factor. It likely may have to do with the paucity of V4V5 sequences from either virus groups represented in the C+ and C- lineage during T-lymphocyte stimulation. It may be that the selection pressure on V4V5 was weaker than the pressure on V3, such that stimulating T-lymphocytes sufficiently overcame those pressures (not just modulated it) and allow infection by viruses with more diverse V4V5 sequences. In this regard, stimulation of T-lymphocytes may be an immunological feature that weakens selection pressures as stimulation is associated with a

higher surface expression of common HIV-binding co-receptors, rendering target cells more vulnerable to infection [Qin et al., 2005].

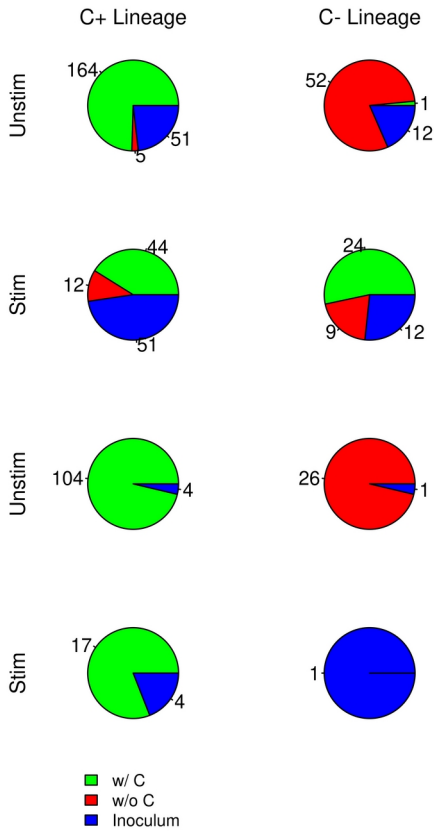


Figure 39: Representation of complement- and non-complement-selected variants in the C+ and C- lineages with or without T-lymphocyte stimulation. Variants in each lineage (C+ and C-) were enumerated according to the infection conditions with which they were associated: with complement (green) and without complement (red). Numbers of condition-associated variants are listed next to their respective portions of the pie charts. Stimulation status of target T-lymphocytes are shown on the left, and the top four pie charts are for lineages in the V3 region and the bottom four are for those in the V4V5 region.

Differential selection patterns with and complement favor polymorphisms in V3 at position 317 and gaps in V4V5

Opsonization of virus by complement and lack thereof led to a differential pattern in selection pressures where opsonization favored variants from one lineage (C+) and non-opsonization favored variants from another (C-). Stimulation of T-lymphocytes weakened those selection pressures, allowing variants from either lineage having a higher probability of infecting target cells under the different conditions. Sliding window analysis of sequences revealed a

single amino acid site in the V3 region of significant diversity, the identity of which is modulated significantly with T-lymphocyte stimulation (**FIGURE 40**). The inoculum consensus at this site is a phenylalanine (F317) whereas the polymorphisms under selection pressure are substitutions to a leucine (F317L) or isoleucine (F317I), very much the same polymorphisms selected in the DC-sticking experiments of the previous study (**FIGURE 35A**). Compared to the inoculum, the consensus amino acid identity at position 317 dropped significantly from 61% to around 21% during *trans*-infection of unstimulated T-lymphocytes with complement-opsonized virus ($p < 10^{-294}$; Fisher's exact test with Bonferroni correction). Stimulation modulated this selection pressure by increasing consensus representation from 21% to 78% ($p \approx 0$; Fisher's exact test without Bonferroni correction). A similar trend in stimulation-dependent modulation of selection pressure was found in the non-opsonized virus condition by an increase in consensus representation from 63% to 85% ($p < 10^{-82}$; Fisher's exact test with Bonferroni corection).

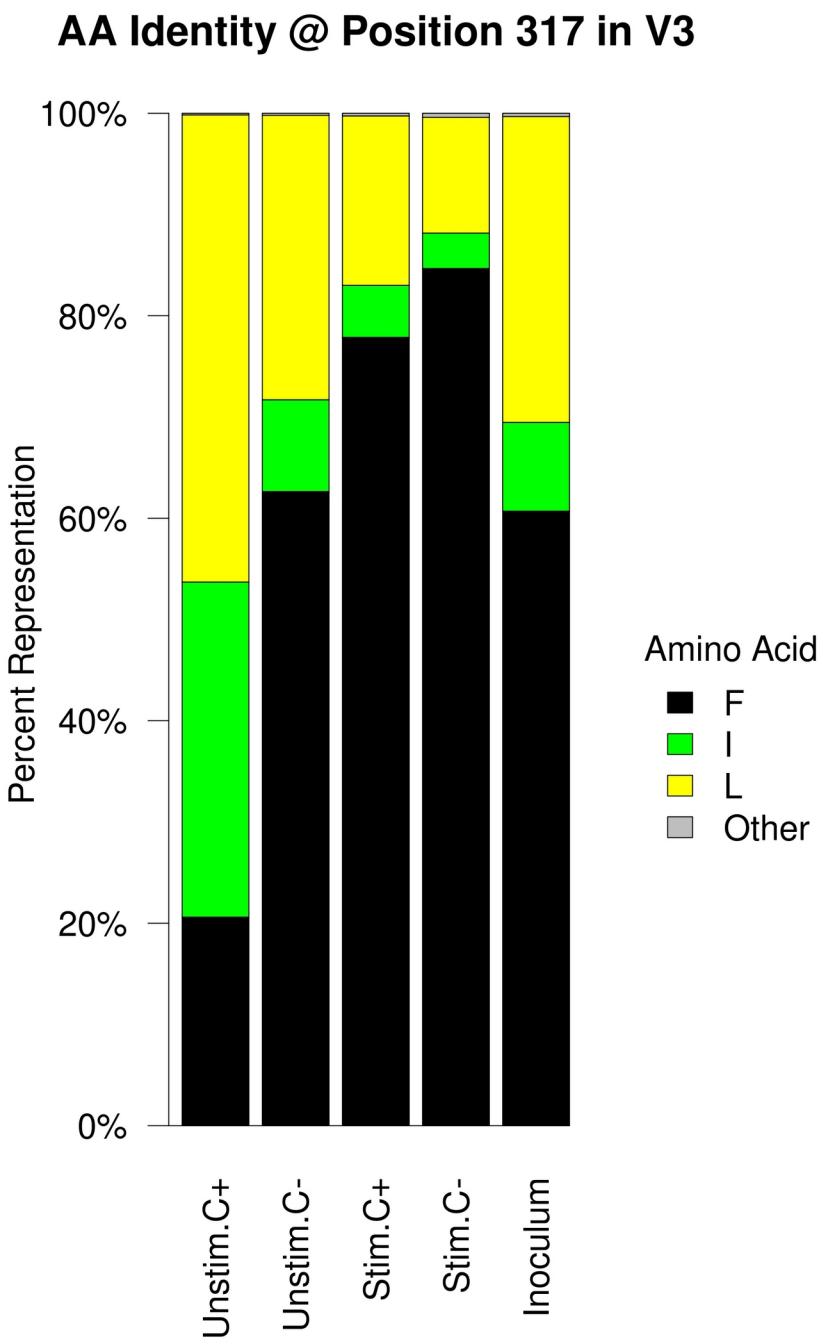


Figure 40: Amino acid representation at position 317 in V3 under various infection conditions. The number of variants in each *trans*-infection condition (listed in the X axis) were enumerated based on their amino acid identity at position 317 and displayed as percentages of 3,000 randomly sampled reads (6,000 for the inoculum).

The amino acids, leucine and isoleucine, at position 317 were represented in the inoculum at 30% and 9%, respectively. These amino acid identities represent minor variants in the inoculum as the majority (or consensus) amino acid identity was a phenylalanine (at 61%). The absence of any selection pressure would have resulted in a similar distribution of these amino acid identities at position 317 in any given infection condition (i.e. *trans*-infection of unstimulated cells with complement-opsonized virus) due to stochastic infection. However, the representation of leucine and isoleucine during infection of unstimulated T-lymphocytes with complement-opsonized virus increased by 16% and 24%, respectively, compared to the inoculum ($p < 10^{-46}$ and $p < 10^{-172}$, respectively; Fisher's exact test with Bonferroni correction). There does appear to be a preponderance for an isoluecine at this position with complement opsonization of virus given that its representation incurred the highest increase when compared to the increased incurred by leucine ($p < 10^{-43}$; Fisher's exact test with Bonferroni correction). This suggests that under these conditions the F317I substitution (and any genetically linked mutation) was favored during *trans*-infection of unstimulated T-lymphocytes by complement-opsonized virus, which happens to a similar mutation found in the direct infection of DCs with the L317I mutation in the previous study (**FIGURE 35A**). Unlike that study, however, the mutation to an isoleucine at this position was not linked to a mutation to a glutamate at position 337 as was previously shown to be significantly favored during direct infection of DCs (**FIGURE 35B**). It wasn't known whether the favoritism for isoleucine in the prior study was the result of positive selection or purifying selection against delterious mutations (i.e. substitution to phenylalanine in a hydrophobic core comprised of mostly aliphatic amino acids); in the case of *trans*-infection of

unstimulated target cells with complement opsonized virus, it is more likely (although still not known) that isoleucine was selected positively rather than indirectly through purifying selection.

Surprisingly, the representations of the conserved and mutant amino acid identities at this position were unchanged in the *trans*-infection of unstimulated T-lymphocytes with naked virus ($p > 0.05$; Fisher's exact test without Bonferroni correction). This was unexpected because the lack of complement opsonization favored HIV variants in the C- lineage; selection pressure does exist but it may not be focused on the 317 position. As of this time, amino acid sites in V3 other than position 317 with the consensus amino acid identity are represented at > 98%, which complicates further analysis on the polymorphisms selected in the transfer of non-opsonized virus during *trans*-infection of unstimulated cells. All in all, these data suggest that under conditions where unstimulated T-lymphocytes were infected in *trans*, stimulation of T-lymphocytes weakens selection pressure to such an extent that enables stochastic infection of viral variants, even if the mode of selection isn't precisely known.

The identification of a single selected polymorphism in V3 was relatively straightforward: (1) it was the only amino acid position that had underwent significant divergence from the consensus and (2) stimulation of T-lymphocytes reduced the strength of selection as to allow infection by viral variants in the inoculum in a stochastic, random manner. However, V4 and V5 regions presented a greater challenge to identifying selected polymorphisms, particularly due to their propensity toward indel mutations [Lamers et al., 2019; Palmer et al., 2019]. To accelerate the unbiased discovery of selected polymorphisms in these regions, a simple mathematical expression was used based on how selection modulation (i.e. stimulation of T-lymphocytes) led to a change in the percentage of inoculum representation at

each lineage with and without T-lymphocyte stimulation. This is predicated on the fact that a constant sample of inoculum sequences were used in all comparisons and phylogenetic trees, making their representation in the given lineages constant in the infections of both stimulated and unstimulated target cells. This was calculated by the difference between the percentage of inoculum reads at a given lineage during infection of stimulated cells and the percentage of inoculum reads at the same lineage during infection of unstimulated cells. For the C+ and C- lineages, applying this calculation shows a 15.3% and 96.3% stimulation-dependent increase in the representation of incoulum reads at each lineage, respectively. It is likely that the representation of conserved amino acid identities at sites under complement- and non-complement-mediated selection pressures may undergo an increase in representation at the same or higher levels as calculated for inoculum representation at each lineage. Therefore selected polymorphisms can be identified indirectly by performing sliding window analysis to search for positions along an alignment where the stimulation-dependent increase in consensus amino acid is equal to or greater than the stimulation-dependent increase of the percentage of inoculum sequences in each of the given lineages. **FIGURE 41** shows a summary of this sliding window analysis.

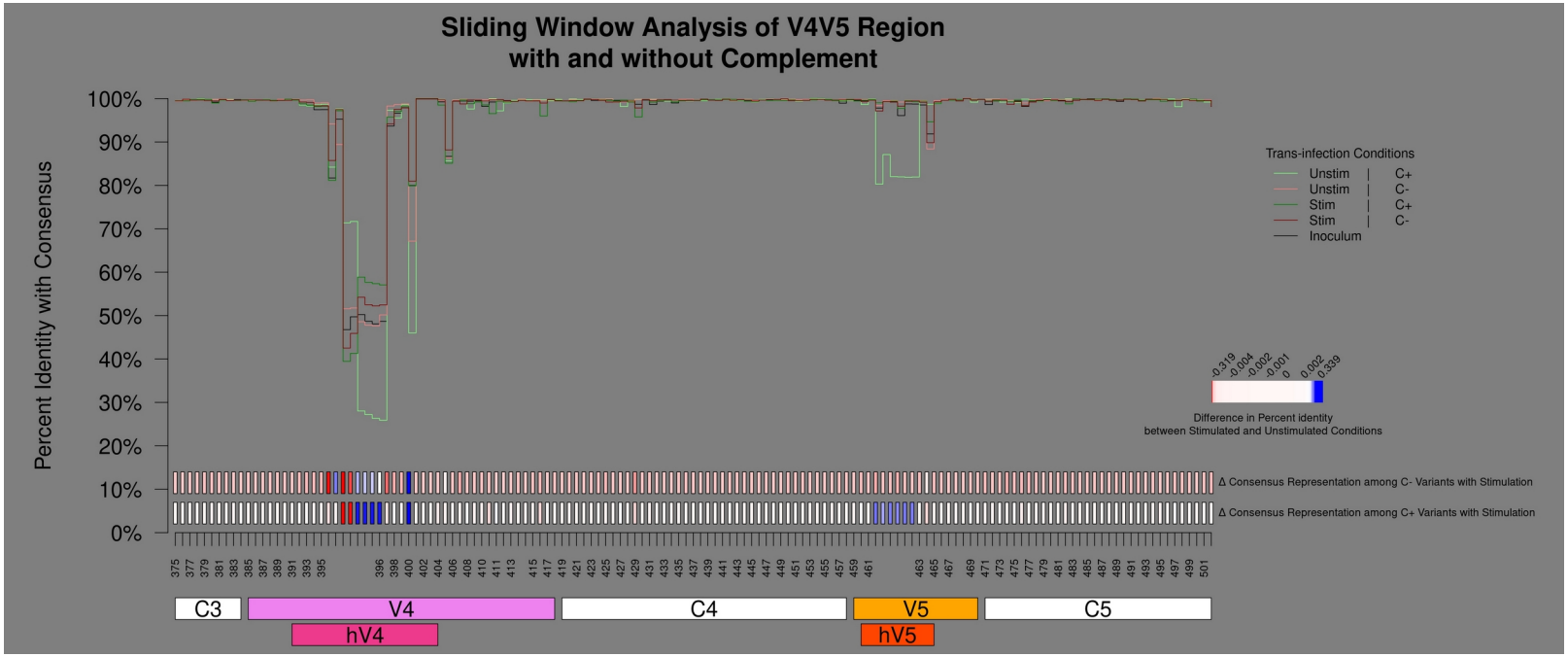


Figure 41: Sliding window analysis of V4V5 regions among infection conditions with and without complement. A visual depiction of sliding window analysis of aligned sequences from the output viruses of each infection condition (*trans*-infection of unstimulated and stimulated T-lymphocytes by complement-opsonized and non-opsonized viral inocula). Data here represents the percentage of inoculum consensus amino acid identities at each position along the V4V5 amplicon. Differences in the consensus representation between unstimulated and stimulated *trans*-infections in C- and C+ conditions is shown for the purpose of identifying potentially selected polymorphism that are specific to the C- and C+ lineages. HIV gp120 domains are shown to the right of the x axis. Amino acid positions are numbered with respect to HXB2 gp120 protein (*GenBank accession # AAB50262.1)*.

In the absence of complement, the major selected polymorphism was W395_W396ins(D/Y) in the V4 loop during *trans*-infection of unstimulated T-lymphocytes compared to its representation in the inoculum ($p < 10^{-65}$; Fisher's exact test with Bonferroni correction). The "D/Y" notation here indicates the insertion of either an aspartate or tyrosine. Stimulation of T-lymphocytes resulted in a 92% reduction of the representation of this polymorphism in the viral output of infected cells ($p < 10^{-65}$; Fisher's exact test with Bonferroni correction). In fact, the representation of W395_W396ins(D/Y) in the stimulation group without complement is equal to that in the inoculum, suggesting the complete absence of selective pressure for this phenotype under stimulatory conditions ($p = 1$, Fisher's exact test without Bonferroni correction). It could explain how V4V5 variants selected in the C- lineage disappeared with stimulation of T-lymphocytes during *trans*-infection. It is important to state, however, that given the highly divergent nature of the V4 loop, the selection of either of these insertions may depend more on negative selection against other genetically linked mutations; this has yet to be confirmed.

While the insertion of either an aspartate or tyrosine in this position was present at higher representation during *trans*-infection with naked virus, there was a 3.3-fold bias toward the insertion of an aspartate over a tyrosine ($p < 10^{-58}$; Fisher's exact test with Bonferroni correction) (**FIGURE 42A**). In the inoculum, the asparate at this position was represented at 0.78-fold that of the representation of the tyrosine at the same position, although the difference in representations between both amino acid identities was not significant in the inoculum ($p > 0.3$; Fisher's exact test without Bonferroni correction). The distribution of these two insertions were also not significantly different in comparisons between the stimulation condition and the

inoculum ($p > 0.1$; Fisher's exact test with Bonferroni correction), which suggests that if selection pressure existed in the stimulation condition, it was greatly attenuated compared to selection pressure during the *trans*-infection of unstimulated T-lymphocytes by naked virus.

While the distribution of the insertions of these single amino acids may account for the disappearance of C- lineage variants from the "stimulation" phylogenetic tree (**FIGURE 38D**), selection pressures favoring these insertions may be modulated by yet an additional genetic factor: namely, the insertion of a subsequent arginine. Selected polymorphisms W395_W396insD and W395_W396insY were represented during *trans*-infection of unstimulated cells with naked virus at 12-fold and and 2.9-fold, respectively, of their representations in the inoculum ($p < 10^{-64}$ and $p < 10^{-5}$, respectively; Fisher's exact test with Bonferroni correction) (**FIGURE 42B**). As shown earlier, representation of these particular polymorphisms in the stimulated condition were scant compared to the unstimulated condition, and they were nearly equal to that in the inoculum. On the other hand, the insertion of a subsequent arginine to the aspartate at this position (W395_W396insDR) was nearly unchanged between the unstimulated and inoculum groups ($p > 0.6$; Fisher's exact test without Bonferroni correction) and between the unstimulated and stimulated groups ($p > 0.9$; Fisher's exact test with Bonferroni correction). The representation of variants with an arginine subsequent to the insertion of a tyrosine (W395_W396insYR) was also unchanged between the inoculum and the stimulation group ($p > 0.9$; Fisher's exact test with Bonferroni correction). However, this mutant was completely absent from the unstimulated group, which could have either been due to a lower replication fitness imparted by the tyrosine for which the arginine could not compensate or due to low probability of being transferred through stochastic infection given its low representation in

the inoculum (2%). The data here suggest quite strongly that there exists a selection pressure weakly favoring the insertion of a single aspartate between W395 and W396 among non-opsonized virus and that this selection pressure weakens significantly during stimulation of T-lymphocytes or even among variants with a subsequent arginine.

In the presence of complement, several selected polymorphisms were identified: four in V4 and one in V5. Regarding V4, the opsonization of virus by complement selected variants

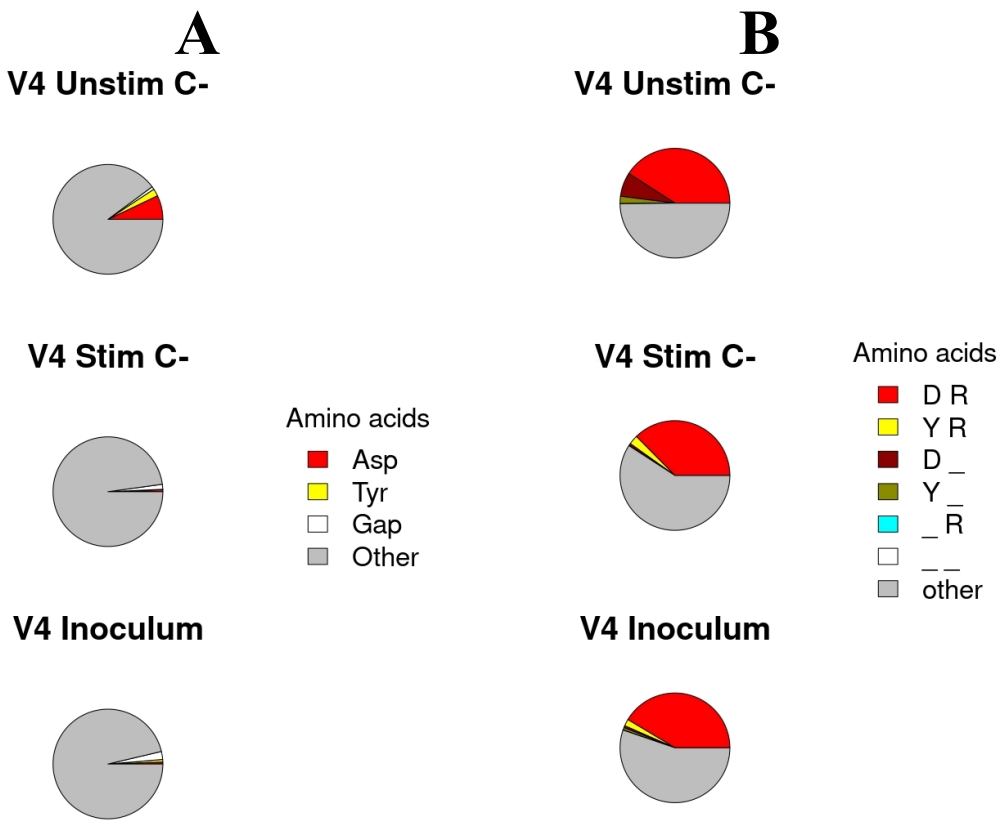


Figure 42: Distribution of viral variants in each group with single (A) or double (B) amino acid insertions between W395 and W396 in the V4 loop. (A) Variants in each infection condition were enumerated based on the single amino acid identity being inserted between W395 and W396. Data is depicted as a percentage of 3,000 randomly sampled sequences (6,000 for the inoculum) for each *trans*-infection condition. (B) Variants were enumerated based on double insertions between W395 and W396 and shown as a percentage of sampled sequences from the *trans*-infection conditions and inoculum.

with W395_W396insNGTWY, which was present in about 1% of opsonized viral variants infecting the unstimulated T-lymphocytes, a 4.1-fold increase in representation from the inoculum and likely an indication of positive selection ($p < 0.005$; Fisher's exact test with Bonferroni correction) (**FIGURE 43**). Another was W395_W396insTWY, though this was found to be negatively selected as its representation in the *trans*-infection, unstimulated group was reduced by 0.1-fold compared to the inoculum ($p < 10^{-5}$; Fisher's exact test with Bonferroni correction). Interestingly, a third polymorphism, W395_W396insXXNG (where X represents any amino acid identity), was also shown to be negatively selected in the same infection condition with a 0.23-fold decrease compared to the inoculum ($p < 0.001$; Fisher's exact test with Bonferroni correction).

A pattern was realized in these analyses such that the insertion of an N-glycosylated site (W395_W396ins(N!PT/N!PTX/N!PTXX), where !P symbolizes any amino acid <u>*not*</u> a proline) may be the driver of positive selection of viral variants during the *trans*-infection of unstimulated T-lymphocytes with complement-opsonized virus. However, there was no significant difference in the representations of these polymorphic potential N-glycosylated sites among the different groups ($p > 0.1$, Fisher's exact test without Bonferroni correction). In fact, N-glycosylated polymorphisms with one or two amino acid insertions N-terminal to an inserted N-glycosylated site (W395_W396ins(X/XX)N!PT) were negatively selected in the unstimulated cells infected with opsonized virus and postively selected in the stimulated cells infected with opsonized virus compared to the inoculum, as shown in **FIGURE 43** ($p < 10^{-77}$ and $p < 10^{-15}$, respectively; Fisher's exact test with Bonferroni correction). These two inserted amino acids happen to have the same genotypic character selected in the absence of complement: W395_W396ins(DR/YR).

And the presence of the arginine was not found to be influenced by any selective pressure propagated during *trans*-infection of stimulated or unstimulated cells by non-opsonized virus. It was suspected that the presence of an R N-terminal to a potential N-glycosylated site may be a modulator of infection selectivity during *trans*-infection, depending on whether target cells are stimulated or not. However, this turned out not to be the case as the distributions of the W395_W396insRN!P(T/S/C) and W395_W396insN!P(T/S/C) were unchanged across the various groups ($p > 0.1$; Fisher's exact test with Bonferroni correction). These results raise serious questions about whether insertions between W395 and W396 are important for selection of variants during *trans*-infection of T-lymphocytes, stimulated or unstimulated, by complement-opsonized virus. However, it was not the only potentially selected polymorphism that could be responsible for the differential selection patterns observed in the C+ lineage with and without stimulation of T-lymphocytes.

The fourth polymorphism potentially selected in the V4 loop by complement opsonization was W400R, which was represented in the unstimulated and stimulated groups at 53.7% and 19.7%, respectively. From the inoculum, this mutation incurred a nearly three-fold increase in the unstimulated group from the inoculum, whereas its representation in the stimulation group was unchanged ($p < 10^{-241}$ and $p > 0.3$, respectively; Fisher's exact test without Bonferroni correction). The substitution of an aromatic amino acid for a positively charged one may be a driver (likely one of several) for positive selection of variants in the C+ lineage during *trans*-infection of unstimulated T-lymphocytes with a complement-opsonized viral swarm, especially since selection pressure for this particular phenotype disappears after stimulation. It should be noted that this same mutation was also selected in the absence of complement, where

its representation in the unstimulated T-lymphocytes without complement was 32.5%; stimulation of T-lymphocytes also reduced its representation to 18.3%. The significantly higher representation of this mutation in the unstimulated, non-opsonized group is an indication of positive selection ($p < 10^{-44}$; Fisher's exact test without Bonferroni correction), and that this pressure was attenuated with stimulation of T-lymphocytes. While it may certainly play a role in selection in the absence of complement, the representation of this mutation in the complement-opsonized group was significantly higher than its representation in the non-opsonized group during *trans*-infection of T-lymphocytes ($p < 10^{-59}$; Fisher's exact test with Bonferroni correction). This implies that selection pressure acting on this polymorphism is likely stronger in the presence of complement than in its absence.

The last potentially selected polymorphism is in the V5 loop where four to five amino acids found to be inserted between N462 and N463 at nearly 100% in both the stimulated group and the inoculum were deleted in the unstimulated group infected in the presence of complement. Because the consensus sequence between N462 and N463 cannot be assigned numbers with reference to HXB2 and in order to simplify the nomenclature for mutations here, these sequences were assigned new numbers as follows: N462, K462.1, T462.2, G462.3, T462.4, E462.5, and N463. The polymorphism in question would be K462.1_T462.4del. These deletions were present in the inoculum at only 0.0167%, but it was present in the stimulated and unstimulated groups at 0% and 12.5%, respectively. This was not the only deletion genotype detected; another deletion of equal length, T462.2_E462.5del, was present in the unstimulated group at nearly 5% (which was undetected in the inoculum at the given sample size of 6,000 reads out of a total of 148,703 reads, after data cleanup). The representation of the

K462.1_T462.4del and T462.2_E462.5del were significantly higher in the unstimulated group than in the inoculum, suggesting the existence of positive selection pressure that favored these deletions during *trans*-infection of T-lymphocytes by complement-opsonized virus ($p < 10^{-181}$ and $p < 10^{-68}$, respectively; Fisher's exact test with Bonferroni correction). The absence of these deletions in the stimulated group also suggests that stimulation of T-lymphocytes modulates selection pressures for these particular genotypes ($p < 10^{-116}$ and $p < 10^{-42}$, respectively; Fisher's exact test with Bonferroni correction). In fact, the distribution of both deletion genotypes were not significantly different between the stimulated group and the inoculum, suggesting that stimulation of T-lymphocytes may have completely nullified selection pressures favoring either of these genotypes under complement opsonization ($p = 1$, Fisher's exact test without Bonferroni correction).

Two other mutations were found to be significantly selected under these conditions: N462D and E462.5K. The latter was found to be genetically linked to the K462.1_T462.4del, and so whatever role it plays in selection, if any at all, it would likely be related to the K462.1_T462.4del polymorphism. Unlinked E462.4K was not selected under any condition as its distribution in both the stimulated and unstimulated conditions were unchanged relative to the inoculum ($p > 0.1$ and $p > 0.3$, respectively; Fisher's exact test without Bonferroni correction). The former mutation here, N432D, was not linked to either deletion genotype appears to be selected positively independent from the aforementioned deletions. It was present in the unstimulated group at 19.2%, in the stimulated group at 0.63%, and in the inoculum at 1.65%. The distribution of this substitution among the various treatment groups imply that it is positively selected in *trans*-infection of unstimulated cells and negatively selected in *trans*-infection of

stimulated cells ($p < 10^{-185}$ and $p < 0.01$, respectively; Fisher's exact test with Bonferroni correction).

The differentiation in selection pressure between unstimulated and stimulated cells for both of these polymorphisms (K462.1_T462.4del/E462.5K and N462D) may indicate that complement opsonization selects for both genotypes in the *trans*-infection of unstimulated T-lymphocytes and against them in stimulated T-lymphocytes. Though not yet confirmed through follow up experimentation, the mutations in V5 here may be responsible for the differential selection pressures between the C+ lineage with and without stimulation of T-lymphocytes. Since these genotypes were found not to be genetically linked, it is unlikely that complement is acting on variants with either genotype in exactly the same way. It is clear, however, that whether or not selection is acting on these particular genotypes, their apparent distributions in among different experimental conditions is strong evidence that there exists selection pressure for particular viral genotypes that can be modulated through complement opsonization of a viral swarm and stimulation of T-lymphocytes prior to *trans*-infection. Furthermore, it is a demonstration of the capacity of this experimental model to detect selection pressures that may be relevant to discovery of T/F strains a clinical setting, especially since viruses passed from the donor to the recipient are often osponized by complement in the semen [Bouhlal et al., 2002] and T-lymphocytes are rarely stimulated in virally invaded mucosae provided that there isn't a concurrent infection.

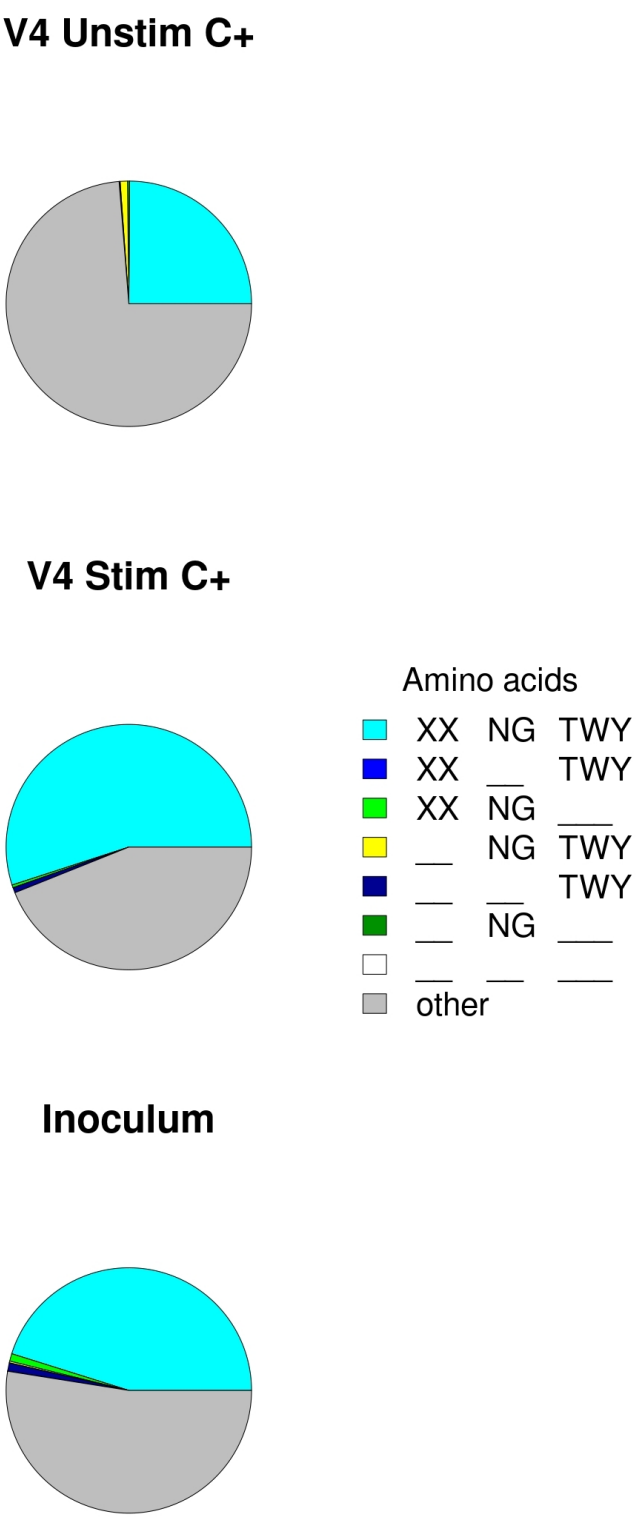


Figure 43: Distribution of viral variants in each group infected with opsonized virus with insertions between W395 and W396. Variants with more complex amino acid insertions between W395 and W396 of the V4 loop were enumerated per *trans*-infection condition. Data is depicted here in pie chart format as percentages of each polymorphism among 3,000 randomly sampled sequences from the *trans*-infection of unstimulated T-lymphocytes with opsonized virus (top) and from the *trans*-infection of stimulated T-lymphocytes with opsonized virus (middle). Percentages are for the inoculum (bottom) are based on randomly sampled 6,000 sequences. "X" refers to any amino acid identity, and "_" refers to gaps.

Selection toward complement-opsonized and control viruses during *trans*-differentiation does not occur at the level of DC adherence or phagocytosis

Strong selection pressures favoring control or opsonized virus were found during *trans*-infection of unstimulated T-lymphocytes. The origin of such selection can be propagated at any point during *trans*-infection, from the opsonization of the virus itself, to the adherence of the opsonized virus to DCs, and finally during the cell-to-cell viral transfer. In order to investigate how viral adherence to DCs may modulate or propagate this selection pressure, qPCR and phylogenetic analysis was performed on DCs during a "stickiness" assay, wherein a viral inoculum stock, either complement-associated or control, was incubated with DCs for 3 hours, washed in PBS, and then lysed to ascertain not only the amount of viruses that successfully adhered to the surface membrane of DCs but also those that were engulfed into endosomes.

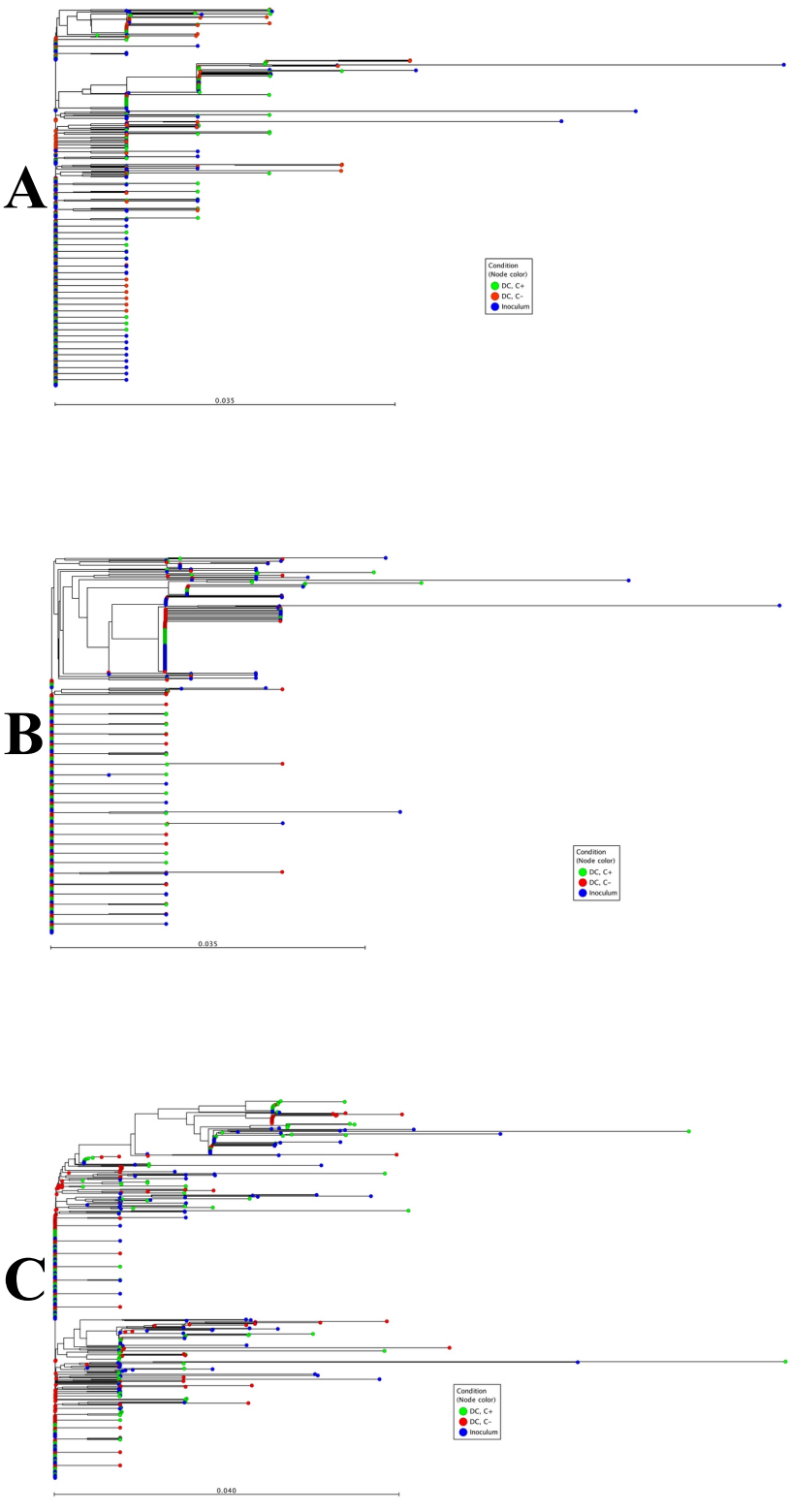


Figure 44: Phylogenetic tree of V1V2 (A), V3 (B), and V4V5 (C) amplicons of DC-adhered viruses with and without complement. Neighbor-joining trees were constructed for 500 randomly sampled sequences from the DC-sticking assay with complement-opsonized virus (green), non-opsonized virus (red), and the inoculum (blue). Trees are based on V1V2 (A), V3 (B), and V4V5 (C) alignments.

The phylogenetic tree of DC-adhered viruses opsonized or not opsonized by human complement is shown in **FIGURE 44A** for V1V2, **FIGURE 44B** for V3, and in **FIGURE 44C** for V4V5.. There is no obvious bottleneck or patterns of selection observed in any of these trees, which suggests that DC adherence is unbiased and that complement-mediated selection occurs during the viral transfer process. To determine quantitative effect of DC adherence with and without complement opsonization, three different challenge doses were tested during a DC sticking assay: 0.050 ng-equivalents, 0.075 ng-equilvants, and 0.100 ng-equivalents of HIV-1 p24 with or without human complement opsonization. As expected, there was a direct dose-dependent response in the mean *HIVgag* copy number adhered to each DC to increasing challenge doses (**FIGURE 45**). At the lowest tested dose (0.050 ng of p24-equivalents), there was a trend (although not statistically significant) trend toward higher adherence to DCs among complement-opsonized virus compared to the non-opsonized control. This complement-dependent increase waned with higher challenge doses, as it appears that the non-opsonized virus adhered to DCs at higher level than did the complement opsonized virus. In fact, there was a statistically significant increase in the amount of virus adhered to DCs among control viruses compared to the complement-opsonized challenge ($p < 0.01$, Welch's T-test). This suggests that complement enhances virus-DC association at low viral loads impedes maximal binding of virus to DCs at higher viral loads.

<u>Complement-directed selection and DC adherence could be explained by two opposing forces</u>

Taken altogether, these data suggest that two different viral lineages were under some form of selective pressure along the complement axis: viruses in one lineage were favored in the

presence of complement whereas the other was favored in its absence. Additionally, that pressure is likely propagated at some point in the infection process before transfer of viral particles and after DC adherence. This is predicated on the finding that virus adheres non-specifically to DCs prior to *trans*-infection. Interestingly, complement appears to enhance binding at low viral titers but that enhancement wanes at higher titers. In speculating on possible mechanisms for these observations, it was important to consider virus-cell interactions independent of complement opsonization. One possible explanation for this could be that the binding affinity between complement-opsonized virus and complement receptors (i.e. CR3) is higher than the binding affinity between "naked" virus and lectins (i.e. DC-SIGN or Siglec-1) on DCs, and to the further detriment of the virus, complement binding may also have impaired N-glycan binding to surface lectins. This hypothesis is supported by an earlier study by Bouhlal et al. [2007] in which blocking CR3 reduced, though not significantly, DC-adherence of complement-opsonized R5-tropic virus to such an extent that the "naked" virus exhibited greater adherence to CR3-blocked DCs by comparison (either through CCR5 or other receptors). However, the speculation that complement opsonization of virus blocks viral adherence to lectins is contradicted by the same study wherein blockage of DC-SIGN attenuated complement-dependent enhancement of DC adherence. Bouhlal et al. [2007] found evidence that DC-SIGN participation in viral sticking (and transfer) is a complement-dependent process. It should be noted that the the 2007 study methodologically differed from the current one in that Bouhlal et al. [2007] used double the challenge dose used in the current study and output virus from infected DCs were sampled 6 days post infection rather than 3 hours in the current study. The output viruses sampled and studied in the 2007 study reflected the progeny of DC-infecting

viruses whereas the output in the current study represents mainly adherent virus, regardless of their infectiousness. The discrepancies in methods between this and the previously published report complicates any comparison between them.

These findings also suggest that there is an upper limit to complement-dependent enhancement of DC binding, and this upper limit may depend on the viral titers of an infectious swarm. Below this limit, complement acts as a stronger bridge between opsonized virus and monocytic receptors, enabling enhancement of DC adherence; above this limit, all monocytic receptors (whether or not DC-SIGN is also involved in CDE) on the surface of DCs may

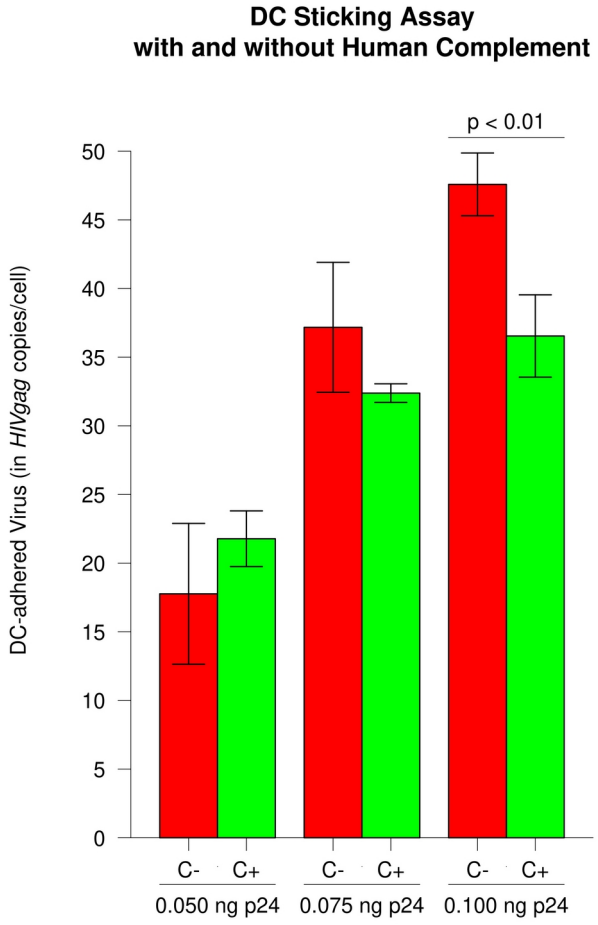


Figure 45: HIV-DC sticking assay with varying titers of complement-opsonized and non-opsonized virus. Viral isolates from lysed dendritic cells of the DC stick assay, challenged with 0.050, 0.075, and 0.100 ng-equivalents of p24 were measured for *HIVgag* vRNA content and normalized to the amount of dendritic cells added to each tissue culture well (~ 50,000 cells). DC sticking assay were performed with complement-opsonized (green) and non-opsonized (red) viral inocula. Data is presented as mean ± standard deviation (n = 3). p value is the result of Welch's t-test.

be engaged and occupied, thus limiting how many more "bridges" could be formed between DCs and virus through cell surface receptors. If the findings here are repeatable beyond just the experiments performed here and that such an upper limit exists, then viral titers beyond this upper limit would contribute to lower infectivity in *trans*, regardless of whether it would contribute to changes in selective pressures favoring viral variants in the "C+" lineage.

The aforementioned explanation is predicated on known examples of CDE [Bouhlal et al., 2007; Willey et al., 2011; Nijmeijer et al., 2021] and the ability of virus to bind to DC-SIGN and other lectins through complement-dependent [Montifiori et al., 1989; Bouhlal et al., 2007] and -independent [Kwon et al., 2002; de Witte et al., 2007; Stax et al., 2009; Shen et al., 2010a; Nijmeijer et al., 2021] means. However, another possible explanation is predicated on an immunological occurrence with an opposite effect: the intended anti-viral function of complement. The complement cascade system is an important immunologic feature known for its anti-pathogenic properties, including endosomolysis (reviewed by Brown [1991], Yu et al. [2010], Sarma and Ward [2011], and Posch et al. [2020]). The adherence of opsonized HIV to DCs may simply be a consequence of DCs interacting with targeted viruses through CR3-related signaling, resulting in phagocytosis [Brown, 1991], macropinocytosis [Mercer and Helenius, 2009] and endosomolysis of the virus. While this function is not dissimilar to CDE (in very much the same way that ADE occurs initially through the intentionally microbicidal process of phagocytosis prior to transfer to the enhancement of *trans*-infection), there is a higher likelihood of viral lysis in the DC sticking assay because it lacks any T-lymphocytes to infect in *trans*. The lack of such target cells in this case ultimately reduced the number of pathways that are opportune to the virus to maintain its survival and replicability, leading ultimately to a higher

probability of killing endocytosed viruses. Therefore, the observed enhancement of DC adherence at higher viral titers could simply be a reflection of incomplete culling of opsonized virus and maximized CR3. At higher titers, all of the CR3s of DCs may be engaged with opsonized virus, making it a rate-limiting step in monocytic anti-viral responses.

"To kill or not to kill": A Sliding Scale

It is particularly important to note that the biochemical features (i.e. complement bridging between virus and DCs, possible competition between complement receptors and lectins for viral association, etc.) of both possible explanations above are largely the same but that they differ mainly in outcome of the viral infectivity. The immunologic purpose of the complement cascade system is the neutralization and destruction of potentially inimical foreign particles (reviewed by Sarma and Ward [2011]. This is not always the case with HIV, as literature on the topic of CDE demonstrates the ability of the virus to "hijack" the complement system in order to enhance its infectivity [Thieblemont et al., 1993a; Larsson, 2005; Bouhlal et al., 2007; Willey et al., 2011; Tremblay et al., 1990; Nijmeijer et al., 2021; Day et al., 2022]. The outcome of CR3-related signaling in the context of this model makes the difference in whether DCs are behaving in accordance with or against its immunologic purpose: "to kill" opsonized virus or "not to kill" opsonized virus.

These two outcomes can be said to be in competition with one another along the same pathway, and complement regulators and viral variants by virtue of their particular phenotypes and functions could sway this competition in favor of pathogenic survival. For simplicity and by extension of the discussion on the dichotomous nature of the complement system by Yu et al.

[2010], one may imagine a sliding scale where on one end of this scale is the complete destruction of viruses and an inability of DCs to infect autologous T-lymphocytes in *trans*, and on the other the complete failure of DCs to destroy *any* virus, leading ultimately to unfettered *trans*-infection, limited only by the replicative fitness of the viral variants themselves. Most cases of CDE (and ADE for that matter) likely fall somewhere in the middle along this sliding scale or gradient; the tipping of the scales toward one end or another would likely be influenced by factors (i.e. combinatorial mixtures of cytokines, viral phenotypes, antibodies, stimulation status of target cells, etc.) present in host cells, the host body, or even biological compounds introduced into an experimental model [Yu et al., 2010].

To determine the outcome of this biochemical battle (for lack of a better term) and to determine where along this theoretical scale viral survival is being favored or disvafored by competing factors, it is important to refer back to **FIGURE 37**. When T-lymphocytes are absent, it is reasonable to expect lower virus survival due to the abolition of *trans*-infection. DCs could transfer virus among each other, but this wouldn't likely improve survival because the selective pressure for viral variants are the same among these cells as they were isolated and differentiated from the same host. Therefore, along this scale, DC-adhered viruses might be tipped closer to the "kill" end of the scale. It must be stressed that this does not mean <u>all</u> viruses under these circumstances would be killed; rather, there probably exists a net force in favor of destroying viruses, even if some viruses survive destruction and infect the DCs to which they were associated. Conversely, with experimental conditions being reversed where virus-bound DCs were co-cultured in the absence of complement opsonization, one might expect the scale being tipped more in favor of "not killing" bound virus. This would be a reasonable prediction given

that (1) complement cascade signaling wouldn't be engaged and (2) there exists a population of target cells in the presence of virus-transferring DCs. Likewise, this alternative does not necessarily mean that no viruses are lysed or neutralized, rather, it is a net force in favor of transferring viruses to autologous cells, even if in the process fit viruses are lysed in a complement-mediated process. In order to ascertain whether complement is being employed here for its intended anti-viral purpose or being hijacked by the virus to enhance transfer, one must compare the level of infection observed during *trans*-infection with either complement-opsonized or control viruses. As shown in **FIGURE 37**, the *trans*-infection of unstimulated T-lymphocytes with complement-opsonized virus was reduced by nearly 64% compared to control. This suggests that complement in this circumstance is tipping the scale closer to "kill" and away from "not kill." However, this was in the case of *trans*-infection of unstimulated T-lymphocytes. The next question is: Does the same pressure exist when T-lymphocytes are stimulated? In this case, the answer would be no. There was no change in level of infection between complement and control during *trans*-infection of stimulated T-lymphocytes. This suggests that stimulation may overcome complement-mediated transfer of virus to target cells, probably due to a higher level of surface CD4 and CCR5 receptors on the surface of those target cells.

The sliding scale and upper limit hypotheses as presented above are only theoretical and meant to explain (or rather, to speculate) on the observations made here in the absence of follow-up experiments. Whatever mechanism explains the difference in DC adherence between complement- and non-opsonized virus at higher viral titers is not yet known, but the most important conclusion drawn from these data is that complement-mediated selection exists and acts on two different lineages in the 92US657 clinical isolate during *trans*-infection, and that it

does not occur at the level of DC adherence. This has implications for the genetic bottleneck (even if a true bottleneck wasn't found in this particular study) and transmission selectivity of T/Fs in clinical cases due to the propensity of viruses to become opsonized by complement and antibodies during transmission. This model may become a useful tool to researchers interested in how complement may be a selective factor for certain T/Fs and whether polymorphisms from complement-selected T/Fs would be particularly informative in the development of an effective HIV-1 vaccine.

Chapter 6: Findings and Speculations—Interferon

The Effect of Interferon in Variant Selection during Infection

Interferons: A Stronger Test

A common phenotype of known T/F strains is resistance to interferon-mediated responses to viral infection. Previous reports have shown that T/Fs are generally resistant to Type-1 [Fenton-May et al., 2013; Parrish et al., 2013; Iyer et al., 2017; Ashokkumar et al., 2020; Gondim et al., 2021; Sugrue et al., 2022] and Type-2 [Rihn et al., 2017] interferons (reviewed extensively by Utay and Douek [2016]). Despite the selection pressures detected in previous model applications using antibody and complement, there wasn't a consistent pattern of selection. Furthermore, there is a scarcity of research regarding antibody- and complement-mediated selection in published literature, making it difficult to confirm applicability of this model to clinical infection. However, given extensive research regarding the characteristic of interferon resistance among reported T/Fs, utilizing Type-1 interferons (namely Ifnα2) in these experiments might be a stronger test of the model than previous tests.

The experiment was done in exactly the same way the complement-involved experiment was performed with the exception that (1) it included an additional group of directly infected unstimulated T-lymphocytes, (2) the 92US712 clinical isolate was used instead of the 92US657, and (3) only the V1V2 region was PCR-amplified for phylogenetic analysis because it was the only successful amplification in this experiment.

A strong bottleneck was found with Ifnα2-treated, unstimulated T-lypmhocytes

Based on phylogenetic trees, there were no observed bottlenecks in the viral output of *trans*-infected stimulated (**FIGURE 46A**) or unstimulated (**FIGURE 46B**) T-lymphocytes with or without Ifnα2 priming. The same observation was made among viruses from the direct infection of stimulated T-lymphocytes with or without Ifnα2 (**FIGURE 46C**). However, direct infection of unstimulated T-lymphocytes primed with Ifnα2 resulted in a substantial bottleneck reminiscent of clinical infection (**FIGURE 46D**); the lineage related to the observed bottleneck here was specific to Ifnα2 treatment, although there was selection associated with the absence of Ifnα2 treatment, but not strongly enough to produce a similar bottleneck. Given data in the literature showing that T/Fs tend to be interferon-resistant and that the bottleneck observed here, it was conjectured that the mechanism involved in T/F selection during the genetic bottleneck in HIV-1 transmission may have been captured by this experimental model, and thus further investigation was warranted.

Each experimental replicate resulted in a bottleneck of phylogenetically distinct variants

Direct infection of interferon-primed, unstimulated T-lymphocytes resulted in a restrictive bottleneck of variants from an HIV-1 swarm, and thus the next aim was to determine how phylogenetically similar these selected variants were to T/F strains identified in the literature. Since these experiments were performed in three experimental replicates, phylogenetic trees were constructed depicting the relationship between thirteen known clade-B T/F strains and a larger sample size (500 reads) of each infection group in each replicate: repA (**FIGURE 47A**), repB (**FIGURE 47B**), and repC (**FIGURE 47C**). A notable pattern from these trees was that the

bottlenecks observed in each replicate appear to originate from entirely different lineages in the inoculum, implying different strains were selected in each infection. T/Fs were also shown to be phylogenetically distinct from the experimental strains assayed in these experiments. Consensus sequences of the interferon-selected strains in each replicate were aligned with T/F strains and their relationships are depicted in a tree to better visualize the phylogenetic distinctiveness of not only the T/Fs but between the replicate consensus sequences themselves (**FIGURE 48**). repA, repB, and repC consensus sequences, as closely related to one another as they are relative to T/F variants, are neither identical nor belong to the same lineages in the 92US712 inoculum (the consensus for which is symbolized by the "CON.i" tip). The selection of different strains under the same condition may be analogous to the report by Yuan et al. [2017] showing that T/Fs are animal-specific and that phylogenetically different T/Fs with similar founder-specific signatures (i.e. amino acid motifs commonly associated with T/Fs) can be selected under the same mode of transmission. Another notable pattern here is that the divergence between the T/Fs and the replicate consensus sequences are likely the result the phylogenetically disparate origin of the 92US712 isolate compared to each T/F strains, as evident by the distances between the T/Fs and CON.i. In fact, each the 92US712 inoculum appears as phylogenetically distinct from each of the T/Fs as each T/F are from one another. Therefore, if any genotypic or functional similarity exists between the distinct viruses selected in each replicate, it would be identified through sliding window analysis and searching for founder-specific signatures as Yuan et al. [2017] did in their research.

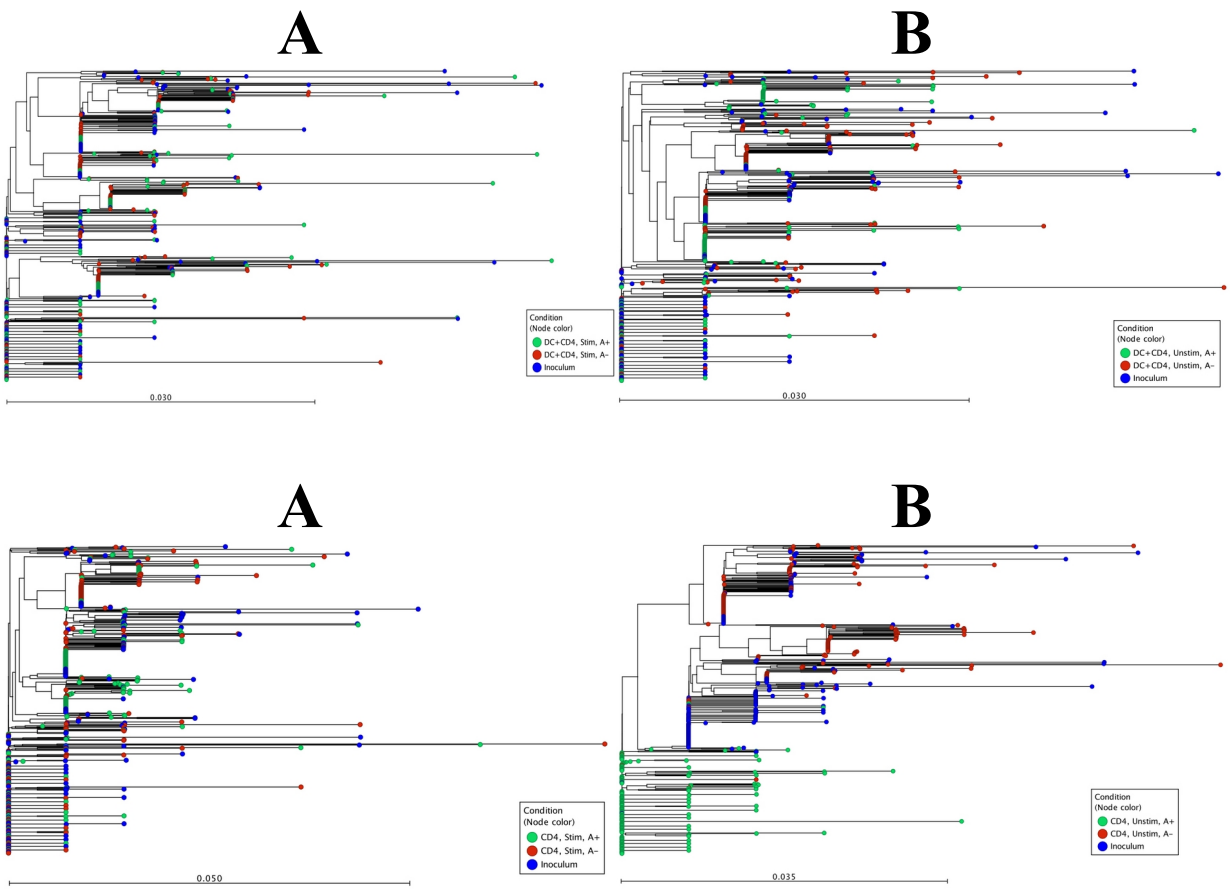


Figure 46: Phylogenetic trees of T-lymphocytes infected with and without Ifnα2 during direct and *trans*-infection. Neighbor-joining trees were constructed based on V1V2 alignments for the output virus from *trans*-infected simulated T-lymphocytes (A), *trans*-infected unstimulated T-lymphocytes (B), directly infected stimulated T-lymphocytes (C), and directly infected unstimulated T-lymphocytes (D). Green tips refer to output virus from interferon priming, red tips refer to output virus from the control, and blue tips refer to the inoculum. Tress were constructed with 200 randomly sampled reads from each infection condition in replicate repC.

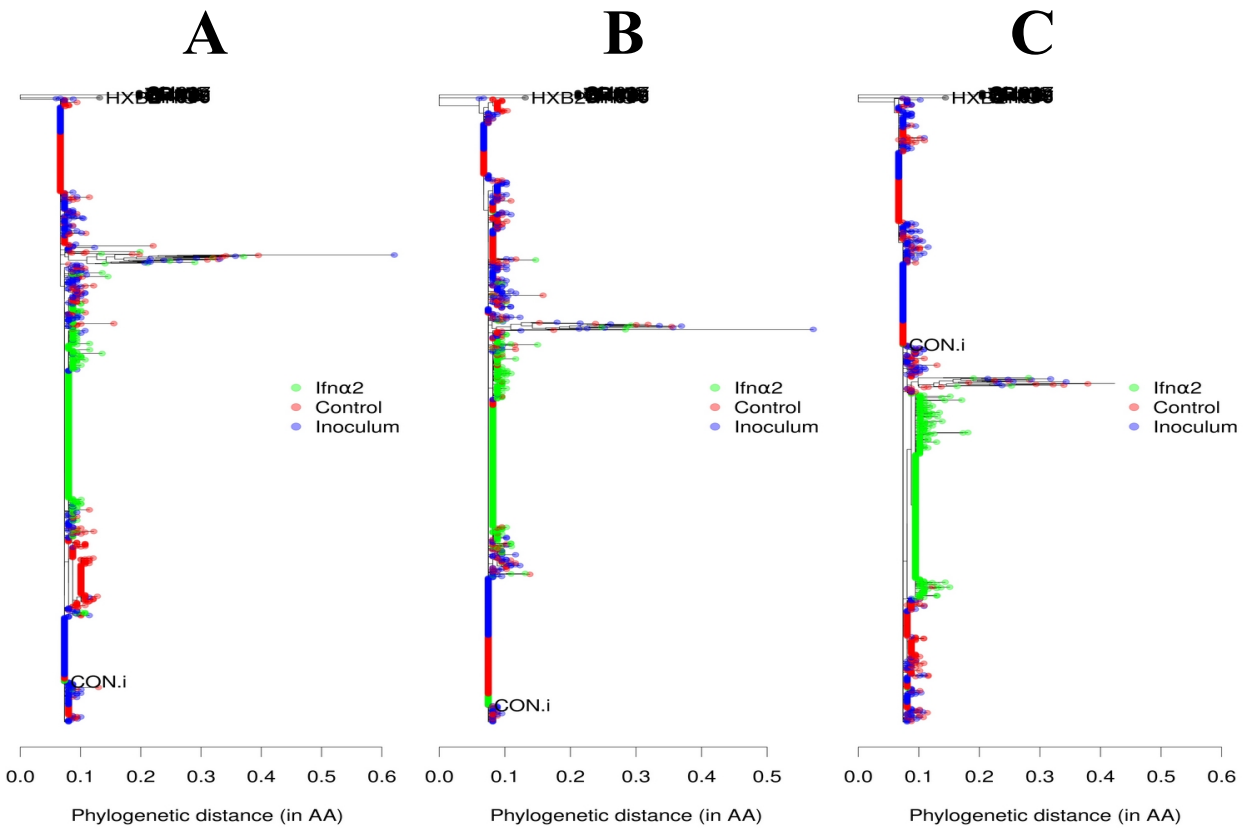


Figure 47: Phylogenetic tree of variants from directly infected unstimulated CD4$^+$ T-lymphocytes with or without prior Ifnα2 treatment by three different replicates. Maximum-likelihood trees were generated from aligned V1V2 amplicons of 500 randomly sampled sequences from each condition in directly infected unstimulated cells of replicates repA (A), repB (B), and repC (C).

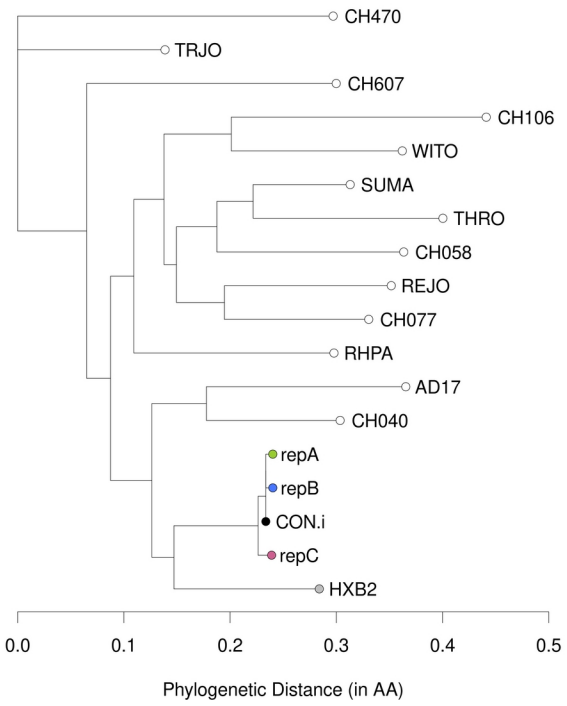


Figure 48: Phylogenetic tree of consensus sequences of interferon-selected viruses from each replicate. Maximum-likelihood trees were constructed based on consensus sequences from each replicate, shown as a green tip (repA), blue tip (repB), rep tip (repC), and black (inoculum consensus). Thirteen clade-B T/Fs are also shown in this tree.

Selected polymorphisms of each bottleneck arose through positive and negative selection, not stochastic infection

TABLE 7 shows a list of amino acid substitutions in this data set that were associated in the bottlenecks of each replicate as well as the direciton of selectivity (i.e. negative/purifying or positive).

In this table, "negative" indicates the existence of selective pressure conserving the consensus amino acid among interferon-selected variants while "positive" indicates selective pressure favoring the substitution. The former can be the result of either, but not strictly only, the higher replication capacity imparted by the consensus amino acid (or a linked polymorphism) or the lower replication capacity by the substitution amino acid (or a linked polymorphism). These are polymorphisms are not likely *de novo* mutations that occur after infection because these polymorphisms exist among viral variants in the inoculum. The importance of these genotypes is

that they either establish or are linked to undetected polymoprhisms in the viral genome that establish higher replication fitness in an infected cell primed with an anti-viral response prompted by Ifnα2.

TABLE 7: Interferon-selected polymorphisms (and direction of selectivity)

mutants	V84A	M147I	N160K	A163V	N167D	S190R	V208A
repA		negative	positive		negative		negative
repB		negative			negative	positive	
repC	positive	negative	negative	positive	positive		

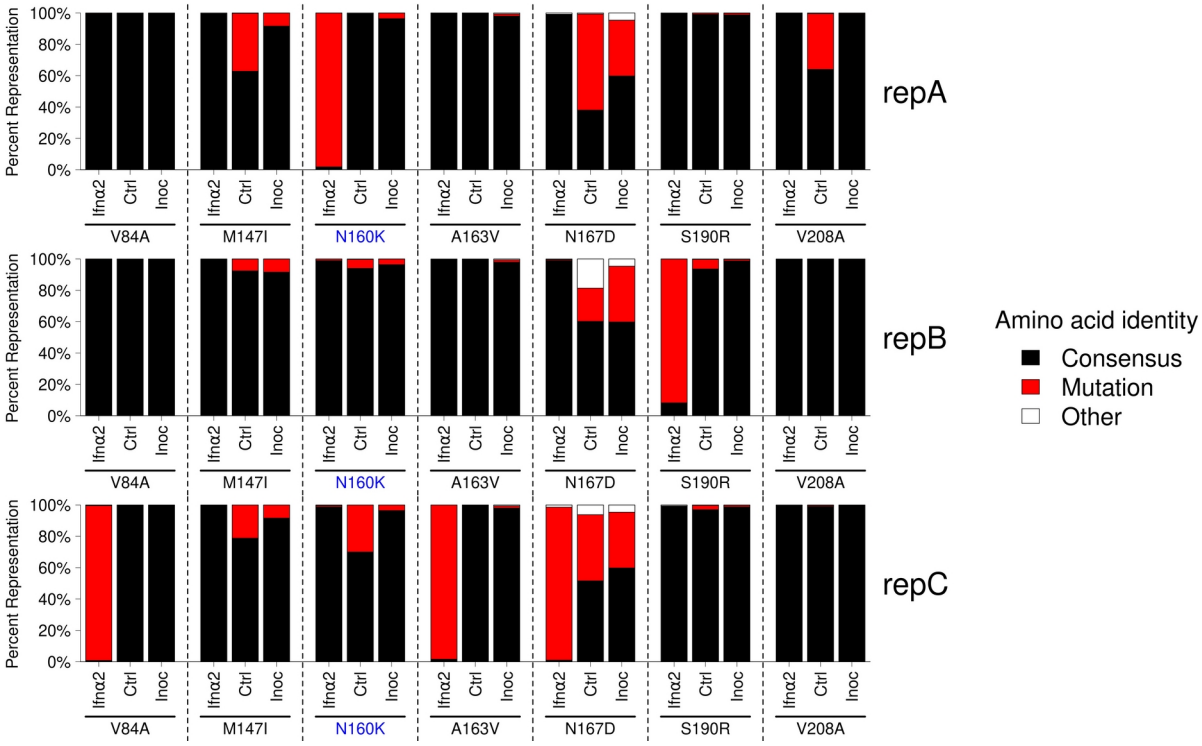


Figure 49: Percent representation of each selected polymorphism (consensus and mutation) in each replicate. Variants in each replicate (repA, repB, and repC) were enumerated on whether they contained a "mutation" amino acid identity or "consensus" amino acid identity at each position where selected polymorphisms were identified. Data is visualized as a percentage of 500 randomly sampled variants from each condition of each replicate.

The polymorphisms identified among interferon-selected variants in the three replicates were V84A, M147I, N160K, A163V, N167D, S190R, and V208A, and the distribution of each of these polymorphisms per replicate per experimental condition are shown in **FIGURE 49**. However, these polymorphisms weren't selected in any particular replicate to the same degree or even the same directionality as in other replicates. For example, selected polymorphisms associated with repA variants were M147I, N160K, N167D, and V208A. M147I, N167D, and V208A were negatively selected during the direct infection of interferon-primed unstimulated T-lymphocytes compared to the control ($p < 10^{-62}$, $p < 10^{-118}$, $p < 10^{-60}$, respectively; Fisher's exact test with Bonferroni correction). N160K was selected positively compared to the control ($p < 10^{-277}$; Fisher's exact test with Bonferroni correction). For repB on the other hand, only three were detected: M147I, N167D, and S190R with M147I and N167D being shared with repA. Like repA, M147I and N167D was negatively selected in interferon-primed cells of repB compared to those in control cells ($p < 10^{-6}$ and $p < 10^{-33}$, respectively; Fisher's exact test with Bonferroni correction). The S190R mutation was positively selected in interferon-primed cells compared to control cells ($p < 10^{-187}$; Fisher's exact test with Bonferroni correction), which was not present at all in either the interferon-primed or control cells of repA. Finally, repC-associated selected polymorphisms were V84A, M147I, N160K, A163V, and N167D, where the first, fourth, and fifth polymorphisms were positively selected compared to the control ($p < 10^{-281}$, $p < 10^{-279}$, $p < 10^{-91}$, respectively; Fisher's exact test with Bonferroni correction) and the second and third were negatively selected compared to the control ($p < 10^{-30}$ and $p < 10^{-41}$, respectively; Fisher's exact test with Bonferroni correction). Of the three replicates, V208A was specific to repA, S190R was specific to repB, and V84A and A164V were specific to repC. Although N160K was shared

between repA and repC, the direction of selection was not, and likewise for N167D. Only repA and repB shared the same directionality of selection for N167D in interferon-primed cells. The only polymorphism that was shared to some degree by all three replicates and in the same manner was M147I: this was negatively selected in all replicates though the degree of selection was not significant in repB when compared with the control ($p < 10^{-62}$ for repA, $p < 10^{-6}$ for repB, and $p < 10^{-30}$ for repC; Fisher's exact test without Bonferroni correction). The lack of uniformity in selected polymorphisms and directionality of selection may not necessarily be a weakness of the *in vitro* model here, and it certainly doesn't present an impediment to further investigation. But it does cast a modicum of doubt over any conclusions that can be made, especially in the absence of any knowledge on how linked mutations in other variable loops might also contribute to selection of variants during direct infection. That being the case, a clear bottleneck was still found in all replicates, and these bottlenecks appear as a cosnequence of selection pressures. Just because some amino acid residues may be selected in one replicate and not in the other doesn't mean that selection isn't acting on some other residue common among all of the replicates, even if it hasn't yet been discovered.

The percent representations of the consensus amino acid for negatively selected polymorphisms (such as M147I and V208A) are typically similar and not always significantly different from those in the inoculum, which raises questions whether they are really the result of negative selection or the result of stochastic infection. This may certainly be the case in a background of positively selected polymormphisms like N160K for repA. In fact, in each of the three replicates there was at least one polymorphism that was selected postively and quite strongly compared to their representation in the inoculum. For repA, positive selection led to an

enhancement N160K's representation in interferon-primed cells 27.3-fold compared to the inoculum ($p < 10^{-245}$; Fisher's exact test with Bonferroni correction). For repB, positive selection with the S190R polymorphism was increased nearly 92-fold compared to the inoculum ($p < 10^{-222}$; Fisher's exact test with Bonferroni correction). For repC, the representation of variants with V84A, A163V, and N167D was clonally expanded 70.3-fold compared to the inoculum ($p < 10^{-259}$; Fisher's exact test with Bonferroni correction), but it may be much higher than that because the V84A mutant was not detected in the inoculum and was more likely to have been expanded from that population ($p < 10^{-283}$; Fisher's exact test with Bonferroni correction). This analysis indicates that positive selection may be a more potent driver of selection than purifying/negative selection. This doesn't negate any role negative selection may play in the establishment of these genetic bottlenecks; *de novo* mutations of positively selected strains may impair the virus's fitness during infection, and thus those mutations would likely be purified out of a viral progeny, ensuring clonal expansion of a founder variant. Furthermore, abundances of consensus residues in these variants is not likely the result of stochastic infection because there would be an equal proportion of other non-selected polymorphisms between the interferon-primed and inoculum groups, like what can be seen in the control where if selection does exist it is greatly attenuated (**FIGURE 49**). Therefore, stochastic infection is not the likely cause of the high proportion of selected consensus polymorphisms in the interferon-treated group, and their persistence may be due to negative selection against deleterious mutations.

Variants of each bottleneck do not share the same polymorphisms as documented T/Fs: Are they novel T/Fs?

Two of these polymorphisms stand out as being of particular interest: M147I and N160K. The selected polymorphism M147I was of interest here because the mutation from methionine to isoleucine was negatively selected, suggesting the existence of selection pressure acting to conserve the methionine in replicating viruses. The particular mutation was also found among variants with diversified amino acid compositions and polymorphisms, suggesting further that selection may be strong enough to select for the methionine at this position over its mutation in a background of many other mutations in other positions. And it likely did not arise *de novo* as the interferon-selected variants with the consensus M147 amino acid character were present at very low percentages in the inoculum (**FIGURE 46D** and **FIGURE 47**). Therefore, its selection here among disparate, scarce variants in the inoculum may indicate that the methionine at position is important for its replicability in cells made hostile to viral replication (though this conjecture has not yet been confirmed).

The selected polymorphism N160K was of interest because the substitution of an asparagine in an N-glycosylated site to a lysine suggests a selection pressure that disfavors N-glycosylation. More importantly, clade-B T/Fs tend to be selected for their lower N-glycosylation profile compared to non-transmitted variants [Liu et al., 2008; Salazar-Gonzalez et al., 2008; Curlin et al., 2010; Gnankaran et al., 2011]. The viruses in the 92US712 clinical isolate is clade-B, and T/Fs present in this isolate would likely be under-glycosylated and interferon-resistant. The representation of this particular variant in interferon-primed cells was greatly increased 27.3-fold of its representation in the inoculum ($p < 10^{-245}$; Fisher's exact test

with Bonferroni correction), in very much the same way that most T/Fs are selected from a rare variant among a viral swarm from an HIV-infected donor [Keele et al., 2008; Salazar-Gonzalez et al., 2008; Frange et al., 2013; Fisher et al., 2010]. Thus, it is possible (though not yet confirmed) that the virus establishing a productive and highly replicative infection in interferon-primed cells in repA may be a transmitted/founder virus. There is some doubt as to the validity of this conclusion because the N160K polymorphism wasn't selected in repB and even negatively selected in repC. However, there is no data on the net abundance of potential N-glycosylated sites in the rest of gp120 variable loops, so the question remains open as to whether this site is important for selection.

Given the possibility that these selected variants may represent T/Fs, discovered or yet to be discovered, the next step was to examine whether selected polymorphisms of variants from each bottleneck are shared by known T/F quasispecies. **TABLE 8** lists known 13 different T/F strains aligned with consensus sequences (repA, repB, repC, and CON.i) and numbered according to alignment of residue positions with HXB2. Among the thirteen T/F quasispecies analyzed phylogenetically, all of them contained one to several sequence identities with bottlenecked variants repA, repB, and repC. For polymorphism N160K, the consensus asparagine being selected in repC while the mutant lysine was selected in repA, one of elevent T/Fs contained the mutant at the same position while the remaining contained the consensus identity. Conversely for polymorphism N167D, where the consensus asparagine was selected in repA and repB and the mutant aspartate in repC, eleven of the T/Fs contained the mutant at this position while only one contained the consensus identity. Most of these T/Fs exhibited high

Table 8: Shared polymorphisms between discovered T/Fs and bottlenecked variants

HXB2 Position	83	84	85	146	147	148	159	160	161	162	163	164	165	166	167	168	189	190	191	207	208	209
HXB2	E	V	V	R	M	I	F	N	I	S	T	S	I	R	G	K	T	S	Y	K	V	S
CON.i	E	V	V	G	M	M	F	N	V	T	A	S	I	R	N	K	T	S	Y	K	V	S
repA	E	V	V	G	M	M	F	K	V	T	A	S	I	R	N	K	T	S	Y	K	V	S
repB	E	V	V	G	M	M	F	N	V	T	A	S	I	R	N	K	T	R	Y	K	V	S
repC	E	A	V	G	M	M	F	N	V	T	V	S	I	R	D	K	T	S	Y	K	V	S
WITO	E	V	V	A	N	V	F	N	T	T	T	V	I	R	D	K	T	S	Y	K	V	S
CH040	E	V	E	G	E	M	F	K	I	T	T	D	I	K	D	R	T	R	Y	K	V	S
CH058	E	I	V	-	-	S	F	N	I	P	T	S	M	Q	D	K	S	T	Y	K	V	S
CH077	E	V	E	S	S	S	F	N	I	T	T	S	I	R	D	K	S	K	Y	K	V	S
CH106	E	V	V	G	Q	L	F	N	T	T	T	S	V	Q	D	K	S	S	Y	K	V	S
THRO	E	V	V	K	T	A	F	N	I	T	T	N	V	R	D	K	T	S	Y	K	I	T
RHPA	E	V	V	-	-	N	F	N	V	T	S	G	I	R	D	K	T	S	Y	K	I	S
REJO	E	V	E	S	S	L	F	N	I	T	T	T	P	R	D	K	I	S	Y	K	V	T
CH470	E	V	K	T	I	S	F	N	V	T	K	S	V	G	N	K	T	S	Y	K	I	S
TRJO	E	L	V	E	E	T	F	N	I	T	T	A	T	G	D	K	T	S	Y	K	I	S
SUMA	E	V	V	A	T	S	F	N	V	T	T	N	M	R	D	K	N	S	Y	K	V	S
AD17	E	V	E	-	-	V	F	N	V	T	T	E	V	R	D	K	T	S	Y	K	V	T
CH607	E	V	R	S	L	K	F	N	I	S	T	N	V	Q	G	E	D	T	Y	K	V	S
Consensus Identity		V			M			N			A				N			S			V	
Mutant Identity		A			I			K			V				D			R			A	

conservation of a valine at positions 84 and 208 as did the experimental replicates, however this is likely because mutations in these two positions might generally impair viral replication and thus would be selected negatively. The only exception to this conjecture would be in repC where the V84A polymorphism was selected positively, and this could be either because an alanine is structurally more similar to valine than a leucine or isoleucine (as seen in the minority of T/Fs here) or whatever negative effect these amino acids have to the replication capacity of the bottlenecked virus might be overcome or rescued by linked polymorphisms with potentially more beneficial contributions to replication capacity in interferon-primed cells. S190R, which was positively selected in repB, only showed up in T/F strain: CH040. Nine T/Fs contained the consensus amino acid serine. At this position, it would appear that either an amino acid with a small hydrophilic side chain (serine or threonine) or an amino acid with a long positively charged side chain (lysine or arginine) would be present, very much like the replicates where either an consensus serine or mutant arginine was present. More interesting is the fact that CH040 contained a lysine at position 160 and arginine at position 190, although the arginine lysine and

arginine at these positions were unlinked in repA and repB, respectively. It could be that the identity of amino acids selected at two different positions might be important and informative to T/F selection in clinical infection where both positions might be important for interferon resistance, if at all. Amino acid identity at positions 160, 167, and 190 might be important for T/F selection in much the same way they were important for selection under these test conditions. The experimental model here may have captured T/F phenotypes, although this is far from certain in the absence of followup experiments and information on linkage with mutations in other variable loops.

Two positions stand out as being highly divergent from amino acid identities discovered in this experimental model: position 147 and 163. The M147I polymorphism, negatively selected in all replicates, was found in only one T/F: CH470. This position is highly conserved in the inoculum at approximately 91.6% with the remaining 8.4% occupied by an isoleucine at this position. It is unclear whether or not this position is in fact important for T/F selection as were positions 160, 167, and 190, and even if a methionine might be present in other T/F quasispecies, it would still be unclear whether a methionine or an isoleucine would be important for interferon-mediated selection of infectious variants. In fact, examination of amino acid identities at position 147 shows that it fluctuates greatly among molecules with hydrophilic, hydrophobic, charged side chains, not to mention that gaps were detected in three T/Fs (CH058, RHPA, and AD17). Furthermore, position 147 is located in or near a region of hypervariability, vulnerable to indels that are often adapted to evade immune responses during chronic infection [Liu et al., 2008; Wood et al., 2009; Kafando et al., 2019]. The identity of amino acids here and proximal amino acids might not be as important as the length of the variable loop for the purpose

of immune evasion and conservation of CD4 binding affinity. Therefore, its relevance to virus selection, particularly T/F selection, is unclear and it is too early, without further testing, to comment on its role in interferon resistance.

While the methionine at 147, conserved in all three replicates, might not be important for T/F selection, the polymorphism A163V may be interpreted differently, however. This was positively selected in repC, did not show up in any of these T/F viruses. It might be present in others, but among the 13 T/Fs examined here the amino acid identities tended to either be positively charged or hydrophilic, while consensus alanine and mutant valine at this position are small hydrophobic molecules. This is distinguishable from position 190 where amino acids shared similar functional side chains despite fluctuating considerably in identity. The differences in chemical properties among amino acid identities in this position discovered through this *in vitro* model and the identities in 13 documented and well-studied T/Fs may stand out as a limitation of this experimental model to capture T/F-associated polymorphisms. However the 92US712 inoculum is approximately 98% conserved for an alanine at this position with the mutant valine making up 1.4% of amino acid identities. Threonine, which was contained in the majority of these T/Fs (11 out of 13), was present at this position in 0.4% of variants in the inoculum. And even rarer variants have been selected in this assay. Thus, the discovery of T/F-associated polymorphisms may only be as limited as the number of times it is repeated; the lack of any discovery here might not be an intrinsic limitation of the model itself (will be discussed further in the final chapter).

Interferon and Direct Infection: A Followup Experiment

Direct infection of T-lymphocytes using multiple clinical isolates and new stimulation conditions

Given the promising results of the last experiment, it was repeated this time with new experimental parameters to expand our search for T/F-associated phenotypes and polymorphisms. Firstly, the last experiment only examined V1V2 sequences; the next experiment examined sequenced regions of SP (signal peptide), V1V2, V3, and V4V5. Secondly, the last experiment only used a sample inoculum of 92US712. The next experiment will test selection from three new isolates (92HT593, 92US660, and JRCSF) in addition to two previously tested (92US657 and 92US712). And lastly, the next experiment was focused on direct infection of cells with different stimulation statuses: unstimulated (which was not stimulated with IL-2 or PHA), stimulated (which was stimulated with IL-2 and PHA prior to infection), and post-stimulated (which was stimulated with IL-2 and PHA posterior to infection). The purpose of the post-stimulated cells was to simulate what most likely occurs clinically within the first several hours to days of acute infection when $CD4^+$ T-lymphocytes are initially at a resting state when exposed to virus until epigenetically activated to a pro-inflammatory state in response to the virus, either because they have become infected or were activated by nearby monocytes. This experimental condition has not been tested by others in the field, and it may represent a clinically relevant testing condition, especially in these experiments where the goal is to replicate a clinical-like infection *in vitro*. These experiments were performed by directly infecting a monoculture of $CD4^+$ T-lymphocytes isolated from donor PBMCs, rather that doing so by *trans*-infection in co-culture with dendritic cells. Where *trans*-infection did not produce a bottleneck in any condition involving interferon-primed cells, direct infection did.

In processing sequence data, there was a large degree of sequence variability unexpected for HIV-1 sequences, even those that are APOBEC-hypermutated. In fact, examination of sequenced reads for hypermutation signatures (employed through the Hypermut tool on hiv.lanl.gov) showed that there was no significant differences in the rate of hypermutation signatures and the baseline G-to-A mutation rate (**APPENDIX 19**). The only clinical isolate wherein the infection output appeared to exhibit a higher hypermutation signature than what was predicted by the baseline mutation rate was 92US657, and this only occurred in the V3 region; however, most of the PCR amplifications failed for this isolate and the data shown in **APPENDIX 19** may have been the based on a very small subset of sequences (i.e. 1 to 3 reads). In any case, viral sequences in many of the infection conditions diverged significantly from inoculum at greater than 15% (data not shown), and this was particularly the case among variants from infections using 92US660 and 92US712 clinical isolates. It wasn't clear whether highly mutated reads from these experiments were real variants or artefacts of PCR-induced errors. Denoising attempts were made using qiime2 (v2021.8) [Bolyen et al., 2019], however these attempts tended to remove too many reads from each data set, more than is expected if they were the result of PCR errors (data not shown). They generally produced phylogenetic trees with (1) such low sequence diversities that some variants could not be traced to any individual node in the inoculum due to the removal of inoculum variants that may have been real and (2) with greatly exaggerated bottlenecks that overestimates the selection strength that could have produced those bottlenecks (data not shown). This is not to say that those bottlenecks weren't real, but rather that the program may used overly strict parameters that removed sequences highly divergent from clustered bottlenecks. For this reason, attempts at denoising these samples was done in lieu

of qiime2 by (1) removing potentially contaminating sequences (i.e. a 92US593 sequence in a 92US712 sample), (2) removing highly divergent sequences (any sequence where its total percent identity was less than 85% with the consensus inoculum; this filter was not applied to the inoculum itself), and (3) removing reads with polypeptide lengths less than the HXB2 amplicon length minus 50 aa and greater than the HXB2 amplicon length plus 50 aa. This denoising attempt has the advantage of removing highly divergent sequences that may have been the result of APOBEC-3G hypermutation and PCR errors without removing real sequences due to overly stringent filter parameters. While it may not have removed all sequences with PCR errors, it wouldn't overemphasize bottlenecks in comparison to non-bottlenecked sequences should they arise.

Cells primed with interferon-alpha 2 were selectively infected by fewer, clonally expanded variants

A global phylogenetic tree of each amplicon (SP, V1V2, V3, and V4V5) is shown in **FIGURE 50**. These trees show the phylogenetic relationship of all of the viruses that produced amplicons under the various infection conditions (with and without Ifnα2), although stimulation statuses are not depicted here in order to show only the general effect of Ifnα2 on infection selectivity. These trees also show tips the size of which represent the numbers of each unique read called "features." The size is represented in log scale and the larger the circle the greater the number of reads for a particular feature. For example, a feature with only 6 reads would be represented at a tip with a smaller circle than a feature represented at a tip with 6,000 reads.

Because of the log scale, however, there is no difference in size between tips that represent two different features represented in the data set by 1 or 10 reads, respectively.

Not every virus or condition amplified successfully by PCR among the various regions. PCR amplification of 92US657 produced too few reads to be depicted in phylogenetic trees, so it was excluded from further analyses (except for analysis of APOBEC-3G hypermutation signatures). For the viruses where amplifications worked, bottlenecks and selection patterns can be detected by fewer and larger circles of a particular condition. This is because a bottleneck or selection toward particular lineages is defined by the clonal expansion of a particular lineage; selection here would entail a restriction of infection by many lineages, if not most of all, except for variants in a particular lineage favored by a particular condition. On inspection of the trees in **FIGURE 50**, Ifnα2 treatment resulted in variants that can be represented by fewer, larger tips, which indicates clonal expansion. The control also produced fewer, larger circles in comparison to the inoculum, but not to the same degree as Ifnα2 treatment did.

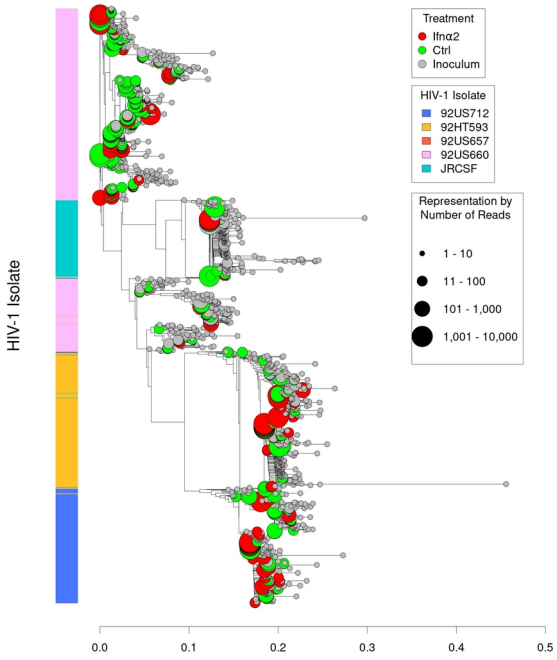
Genetic bottlenecks were strongly associated with unstimulated and post-stimulated T-lymphocytes

A major limitation of these trees, however, is that it only shows a global representation of viruses on the basis of whether cells were treated with or without Ifnα2 and how that could produce selection. While it does show selection exists, it cannot delineate between the different stimulation statuses that may have played roles in those bottlenecks and selection pressures. To visualize these bottlenecks more easily, phylogenetic trees were constructed for each amplicon of each infectious inoculum delineated by stimulation status and interferon treatment where

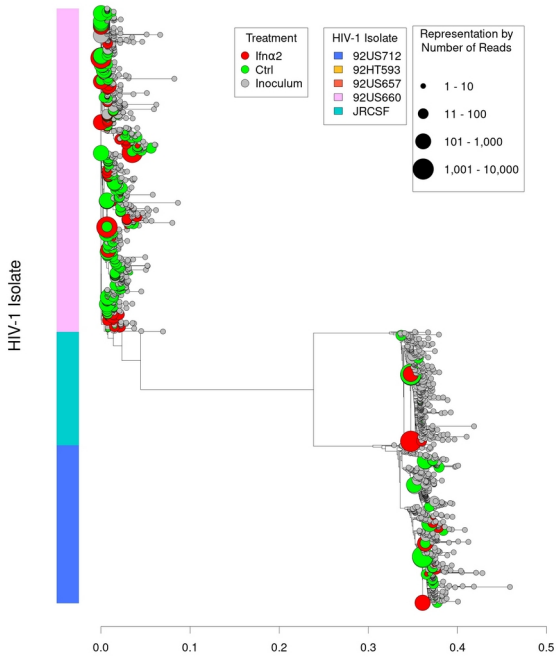
numbers of each feature are represented by the size of tips. These trees also show relationships of sequenced reads to thirteen clade-B T/Fs.

FIGURE 51 and **FIGURE 52** show phylogenetic trees that contained bottlenecks in directly infected interferon-primed unstimulated and post-stimulated cells, respectively. Among infection of unstimulated cells, bottlenecks were observed in the SP region of interferon-selected variants of 92HT593 (**FIGURE 51A**), 92US712 (**FIGURE 51B**), and JRCSF (**FIGURE 51C**) inocula and in the V3 region of interferon-selected variants of 92US712 (**FIGURE 51D**). Among post-stimulated cells, bottlenecks were found in V1V2 (**FIGURE 52A**) and V4V5 (**FIGURE 52C**) of 92US712 inocula and V3 (**FIGURE 52B**) of JRCSF inocula. The SP bottleneck in unstimulated cells primed with Ifnα2 was the most interesting because two closely related features (part of a single lineage) clustered with the same lineage from which CH106 evolved (**FIGURE 51C**). The major variant selected in this bottleneck may share selected polymorphisms with the CH106 SP region, and this similarity may indicate a mechanism of selection. These experiments didn't isolate a bottleneck in the V1V2 region of interferon-primed unstimulated cells infected with the 92US712, which is in contradiction of the prior experiments that showed not only was this condition associated with a bottleneck but they contained

A


Phylogenetic Representation of Unique Reads from SP Amplification of Viruses in Clinical Isolates and Treatment Conditions
HIV-1 Isolate
Treatment
Ifnα2
Ctrl
Inoculum
HIV-1 Isolate
92US712
92HT593
92US657
92US660
JRCSF
Representation by Number of Reads
1 - 10
11 - 100
101 - 1,000
1,001 - 10,000
0.0
0.1
0.2
0.3
0.4
0.5

B


Phylogenetic Representation of Unique Reads from V1V2 Amplification of Viruses in Clinical Isolates and Treatment Conditions
HIV-1 Isolate
Treatment
Ifnα2
Ctrl
Inoculum
HIV-1 Isolate
92US712
92HT593
92US657
92US660
JRCSF
Representation by Number of Reads
1 - 10
11 - 100
101 - 1,000
1,001 - 10,000
0.0
0.1
0.2
0.3
0.4
0.5

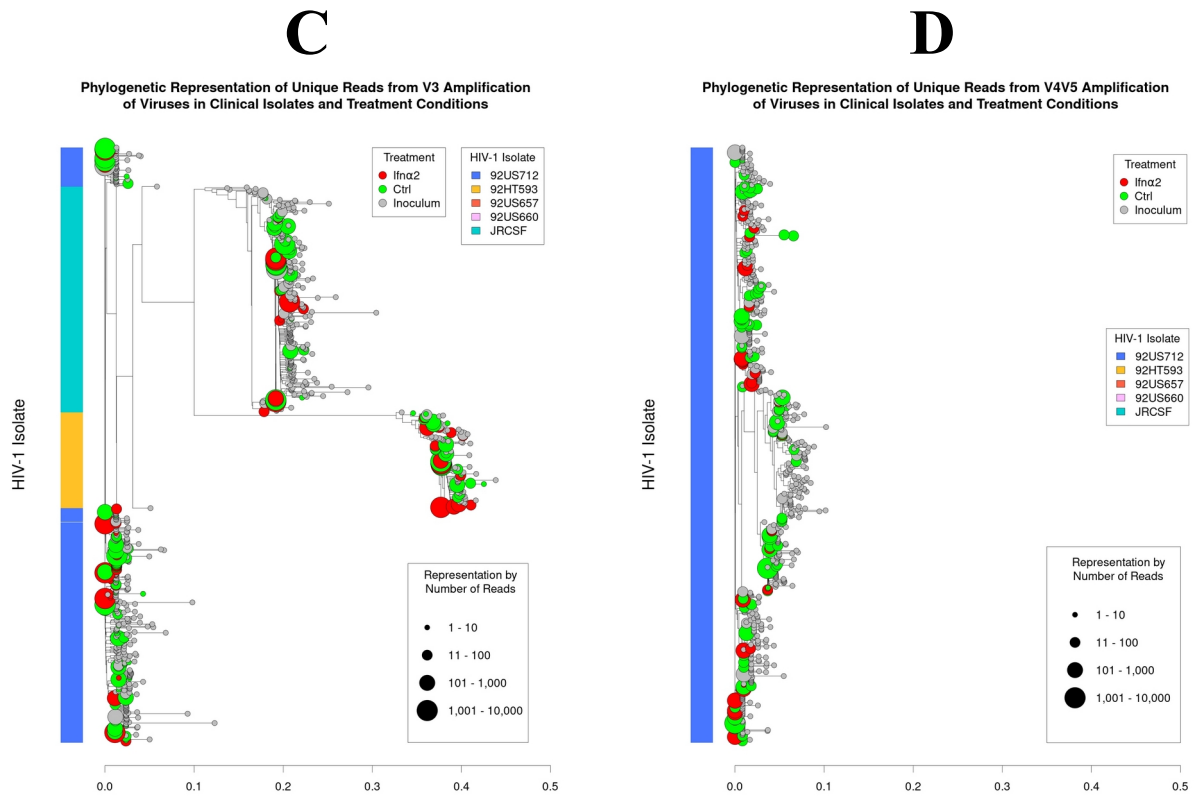


Figure 50 (this page and previous page): Global Phylogenetic trees for each amplicon (SP, V1V2, V3, and V4V5). Maximum-likelihood trees were constructed for the global alignments of SP (A), V1V2 (B), V3 (C), and V4V5 (D) amplicons of each infection condition. Tips are labeled according to their treatment conditions (red for interferon priming, green for control, and gray for inoculum) and their representation by number of reads (size of the tips). The y-axis of each tree is also colored according to the clinical isolate from which each tip originated. The x-axis of each tree shows the phylogenetic distance in amino acid changes per site.

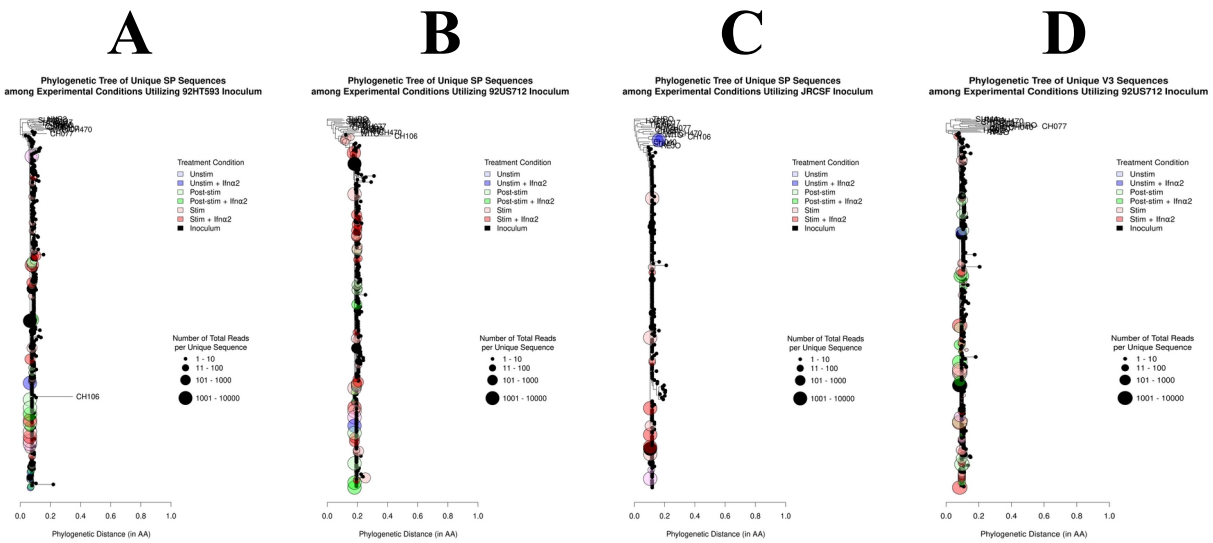


Figure 51: Phylogenetic trees showing bottlenecks associated with direct infection of interferon-primed unstimulated cells. Maximum-likelihood trees for genetic bottlenecks in the aligned SP (A, B, and C) and V3 (D) amplicons of interferon-primed, unstimulated T-lymphocytes infected with sampled inocula of 92HT593 (A), 92US712 (B and D), and JRCSF (C). Tips are colored according to treatment conditions and sized according to numbers of unique sequences in each condition.

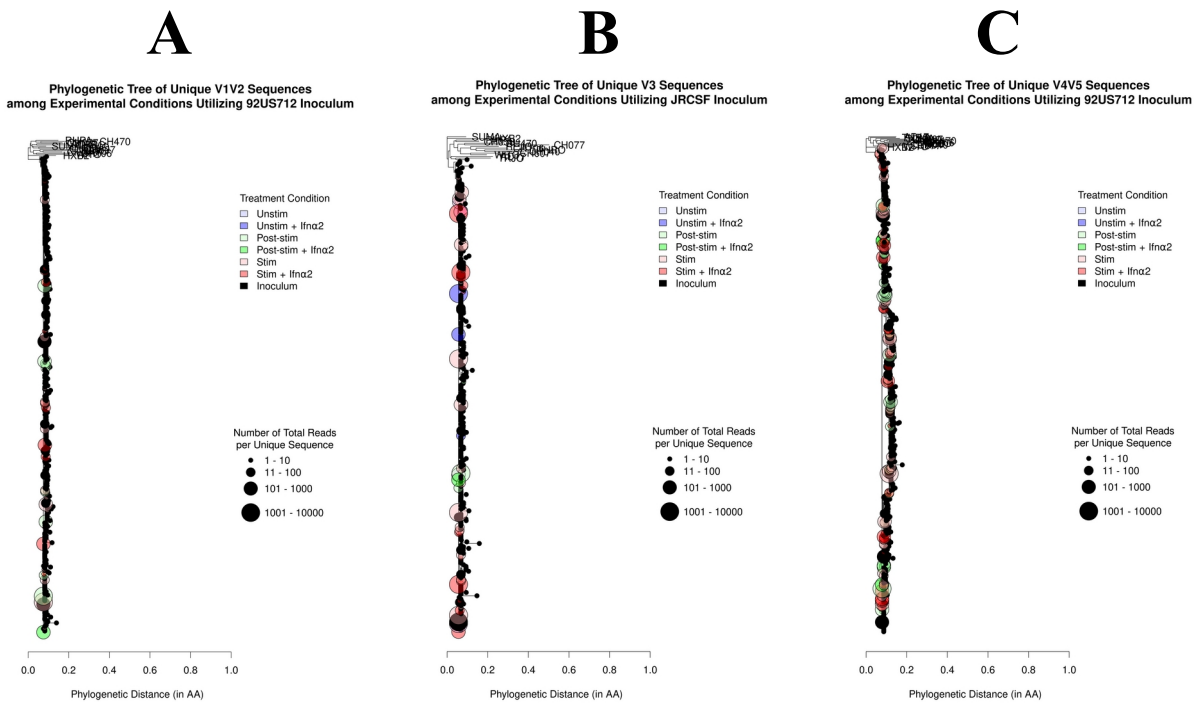


Figure 52: Phylogenetic trees showing bottlenecks associated with direct infection of interferon-primed post-stimulated cells. Maximum-likelihood trees for genetic bottlenecks in the aligned V1V2 (A), V3 (B), and V4V5 (C) amplicons of interferon-primed, post-stimulated T-lymphocytes infected with sampled inocula of 92US712 (A and C) and JRCSF (B). Tips are colored according to treatment conditions and sized according to numbers of unique sequences in each condition.

polymorphisms similar to known T/Fs. However, the PCR amplification of the V1V2 regions in these experiments failed, likely due to extremely low copy numbers. And post-stimulated T-lymphocytes infected with the same clinical isolate produced an interferon-associated bottleneck in V3 (**FIGURE 52B**), so sequences found in these experiments may provide valuable insight into the V3 region and how selection might favor selected polymorphisms here as V1V2 polymorphisms did in prior experiments. While there may have been selection pressures favoring variants during infection of stimulated cells, it was never associated with a bottleneck as did the other two stimulation statuses.

Viral lineages selected by Ifnα2 can be unique or shared among replicates

As shown in **FIGURE 48** from the prior experiment, genetic bottlenecks were produced in unstimulated cells primed with Ifnα2. Those bottlenecks were replicate-specific, analogous to the way T/Fs can be animal-specific [Yuan et al., 2017]. To determine if those findings were reproducible, phylogenetic trees were redrawn to show the relationships of output viruses from each replicate in the current experiment. In this analysis, not every replicate produced a bottleneck, but it was clear that each replicate produced viral variants that clonally expanded from one or several lineages. **FIGURE 53A** shows an example of an infection of post-stimulated $CD4^+$ T-lymphocytes that produced a single replicate-specific interferon-associated bottleneck where all of the output virus from the infection condition in one replicate (rep3) clustered to a single lineage. In this phylogenetic tree, a different replicate (rep2) revealed multiple clonally expanded interferon-associated viruses from multiple lineages, but they were still specific to the particular replicate. Conversely, **FIGURE 53B** shows a phylogenetic tree of viruses from an infection of interferon-primed unstimulated cells where all three replicates share a common viral lineage. Aside from a few unshared lineages that were clonally expanded in rep1, the major lineage shared by each replicate also happened to be closely related to the major consensus lineage in the inoculum, which may suggest stochastic infection rather than selective infection. Examples of replicate-specific and replicate-shared lineages were found in other phylogenetic trees (data not shown), and this led to a question, albeit counterintuitively, of whether the overlap of viral lineages across replicates could be explained by a loss of selection or overcoming selective pressures.

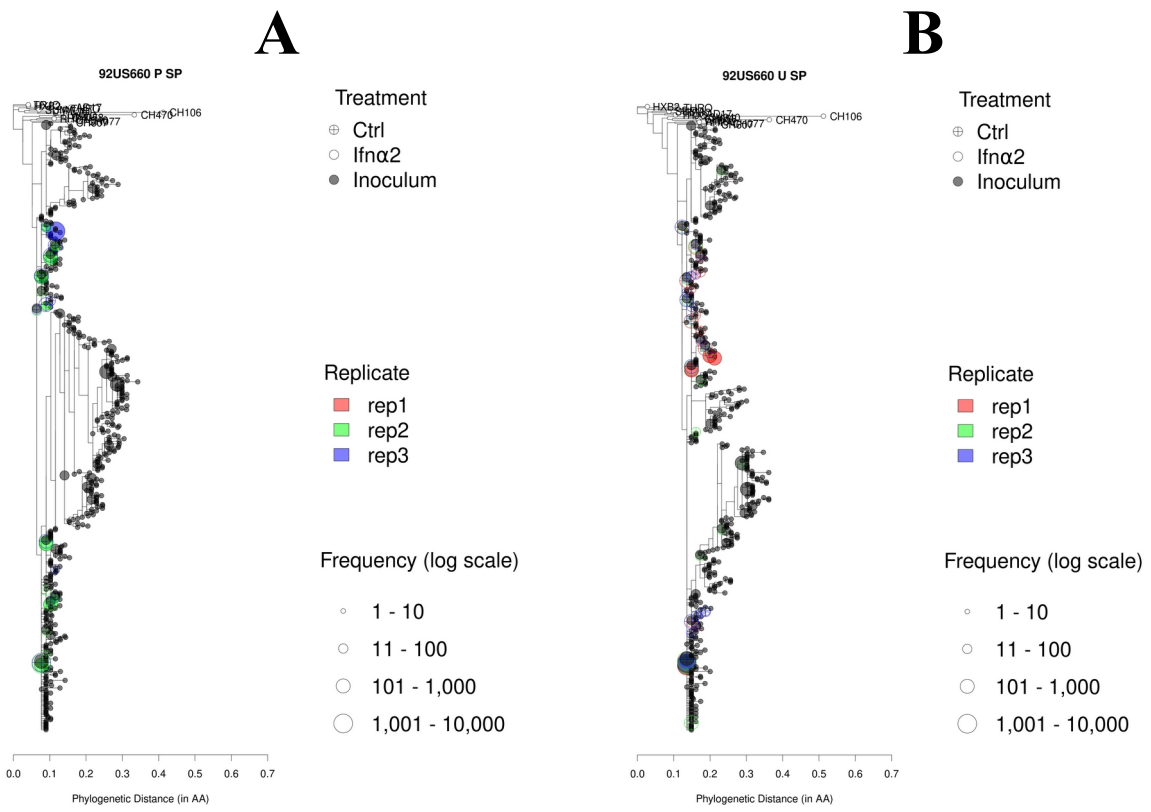


Figure 53: Phylogenetic trees showing clonal expansion at viral lineages specific to (A) and shared among (B) experimental replicates. Maximum-likelihood trees were constructed for the SP amplicons of each replicate from T-lymphocytes infected. These example bottlenecks that are unique/specific (A) to a particular replicate or shared (B) among at least two replicates. Tips are colored according to the replicate from which they originated and the size of tips represent the numbers of unique sequences from each replicate. Controls are labeled by empty circles with a cross inside it, and Ifnα2-associated variants are labeled by tips filled in with the color indicating the replicate number. Tips colored in dark gray indicate variants from the inoculum.

A pattern observed in previous experiments reported here is that stimulation of T-lymphocytes appear overcome selection pressures that would favor certain variants of HIV-1 over others from an infectious inoculum; this generally results in stochastic infection rather than selective infection. Provided that this is likely the case with output virus from stimulated T-lymphocytes in the current experiment, the distribution of unique and shared viral lineages

among each of three replicates and each of 5 clinical isolates were compared among output viruses from unstimulated, post-stimulated, and stimulated cells (acting as a control for loss of selectivity). The results of this analysis for SP, V1V2, V3, and V4V5 amplicons are shown in **FIGURES 54A** to **54D**, respectively. There were no significant differences in the number of

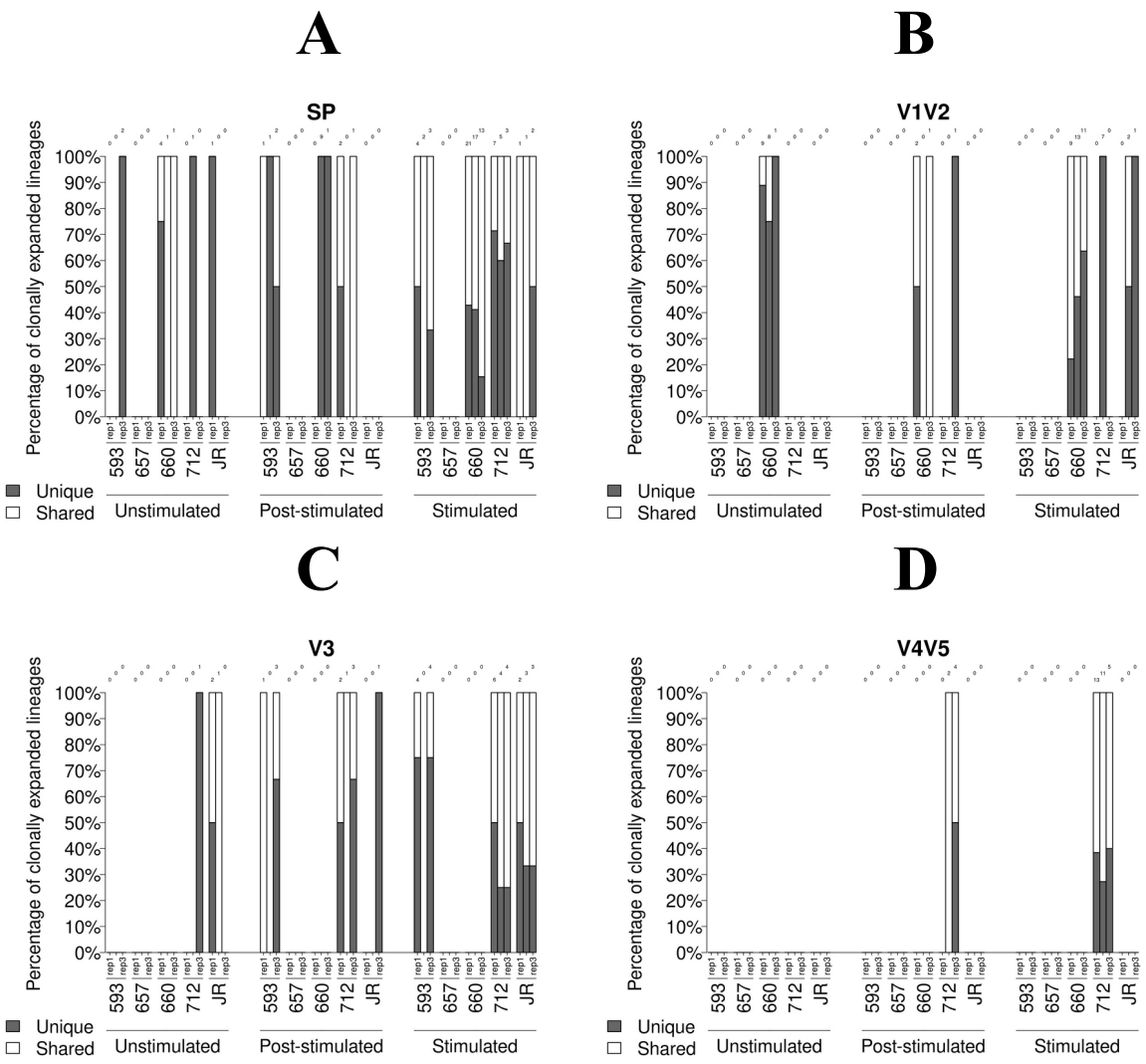


Figure 54: Numbers of lineages in each infection condition that were specific (unique) to a replicate or shared among replicates. Clonally expanded lineages in directly infected, interferon-primed T-lymphocytes for SP (A), V1V2 (B), V3 (C), and V4V5 (D) amplicons were enumerated. Clonal expansions that are specific/unique and shared among particular replicates are shown in dark gray and white, respectively, as percentages of the total number of clonally expanded lineages (which are indicated at the top of each bar).

shared lineages between unstimulated or post-stimulated infection conditions and the infection of stimulated cells ($p > 0.2$; Fisher's exact test without Bonferroni correction). Although it was apparent on examination of these analyses that the number of bottlenecks or clonal expansions increased with stimulation of T-lymphocytes (**FIGURE 55**), as expected given that greater permissibility to the inoculum would overcame whatever selection pressures were present in unstimulated and post-stimulated cells; the differences measured here, however, were not statistically significant ($p > 0.1$; Welch's t-test).

Caution is warranted in interpreting these results for two reasons. Firstly, PCR amplification was not always successful, either because of low viral yields (data not shown) or because of low primer-binding affinity. The PCR failures of one or two replicates but not of a third would lead to an under-representation of the shared lineages, making it appear that there were more replicate-specific lineages in a particular condition than there may actually have been. Secondly, there was a large degree of genetic variability in the datasets, some of which could have been due to PCR errors and/or APOBEC hypermutation. Despite data cleanup to lower the genetic variability below 15%, this likely did not remove all genetic artefacts from the data sets. Ultimately, this implies that the percentage of viral lineages shared among replicates may be far lower than what was actually measured because PCR and host cell-induced artefacts would be unique to a particular replicate but they didn't originate from the inoculum. For the problems inherent in this analytic approach, the results are mostly inconclusive other than the finding that stimulation of T-lymphocytes tends to allow infection by viruses from a greater number of viral lineages, although not to a statistically significant degree as shown here. Stimulation of T-

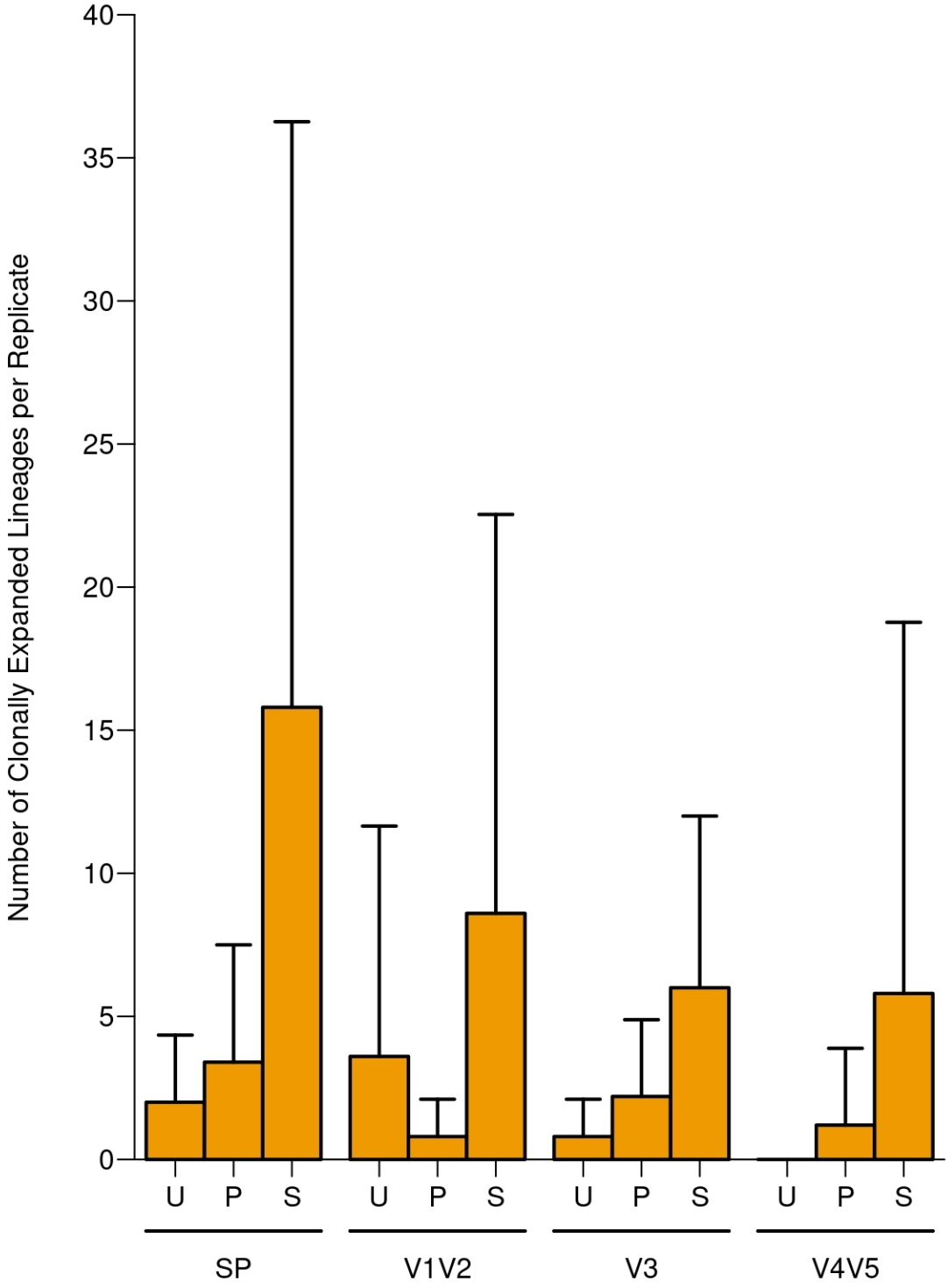


Figure 55: Total number of clonally expanded lineages per replicate. The numbers of clonally expanded lineages in each replicate of interferon-primed cells directly infected with each sampled inocula were counted and displayed here as mean ± standard deviation.

lymphocytes could be a hallmark of selection modulation that has been observed in prior experiments, and further experimentation would be required to determine the etiologic source of such modulation from the standpoint of metabolic and other biochemical changes due to target cell stimulation.

Final comment on the progress analyzing data from this experiment

Further analysis of this dataset was discontinued due to issues inherent with abnormal variation in the data set that may have been the result of PCR errors or some other unforeseen phenomena; in fact, PCR artefacts induced by reading errors and recombinations tend to be a common problem in HIV phylogenetics research [Zanini et al., 2017]. For this reason, only the 3-dpi sample was examined here, even though there were samples collected at 5 dpi and 7 dpi. As of this writing, the sequencing for the 5-dpi sample has been finished but not yet processed. To account for PCR errors, however, future PCR amplifications will utilize barcodes to act as unique molecular identifiers (UMIs) in order to demarcate real nucleotide identities and PCR errors in sequencing reads [Zhou et al., 2021; Zhou et al., 2022]. This approach enables the correction of artefacts introduced through amplification and sequencing and would greatly improve the quality of the data compared to what is shown here.

Chapter 7: Discussion and Model Evaluation

Summary of studies

Presented here are the results of experimental tests for three hypotheses: that transcytosis through an epithelium results in selection of T/F strains (Hypothesis #1), that *trans*-infection results in in selection of T/F strains (Hypothesis #2), and that immunologically important molecules, like antibodies, complement, and interferons, results in selection (or modulation of existing selection patterns) of T/F strains at the level of transcytosis or *trans*-infection. In general, the experiments here revealed several conditions that not only resulted in selection of uncommon variants from a viral swarm, some even resulted in genetic bottlenecks reminiscent of genetic bottlenecks that normally occur during acute infection.

Transcytosis

In vitro transcytosis could not be performed due to a fundamental problem with this experimental approach, and it would require additional work to resolve those issues; such work is beyond the scope of the projects presented here. The workaround for this experimental failure was to use plasma samples from *in vivo* animal studies where animals were challenged through the penile route by exposing foreskin tissue to a viral SIVmac251 swarm and allowing viral variants to penetrate the epithelium by transcytosis or being drawn in by tissue-resident myeloid cells. Viruses were opsonized by anti-SIV antibodies (or non-opsonized in controls) in mixtures containing neutral or low pH. Originally, these experiments tested an hypothesis viral transcytosis of SIV would be enhanced by antibodies and low pH through the FcRN on mucosal tissues [Gupta et al., 2014]. Although enhancement was not found (data not shown), there was

selection for variants with particular polymorphisms. Namely, a set of linked selected polymorphisms were found to have the most significant correlation with antibody opsonization: T426N/Q427K/L429S/R430K (NKSK). These polymorphisms were found in the V4 loop of SIV and literature and BLAST searches did not reveal any insight into possible mechanisms that could explain the selection pressures propagated by antibody and downstream interactions with tissues and other cells. Despite not identifying the etiologic source of such selection pattern, the fact that an uncommon phenotype (NKSK) was clonally expanded in some (though not all) animals infected with antibody-opsonized virus was a strong indication of a selective pressure favoring a particular variant during transcytosis (at the earliest) and viral dissemination (at the latest). Furthermore, the distribution of this polymorphism was unchanged between the control and the inoculum, suggesting that clonal expansion of NKSK-containing variants was more associated with antibody opsonization than without it.

This polymorphism was identified by analyzing the Ka/Ks ratios of sequence reads from each animal according to their treatment conditions [Li et al., 1985; Nei and Gojobori, 1986; Li, 1993] as well as ratios for polymorphisms that are fixed between two different groups [McDonald and Kreitman, 1991]. However, it should be noted that this analysis is not without problems. Significant changes in the distribution of particular amino acids with similar functional groups, such as charged or polar amino acids, may appear to result from selection pressure, even in cases where selection was non-existent [Xia and Kumar, 2006]. For this reason, care must be taken to examine whether observed changes in amino acid identities, particularly between amino acids of similar chemical properties, are truly the result of selection or random non-deleterious mutations. Such validation has yet to be performed, but it is planned

for the future. Regardless of whether the NKSK polymorphism is truly selected positively, it would require more work to determine whether this selection pattern is a reproducible result of antibody coating and, if so, how it immunologically and physiologically establish that selection pattern.

In vitro infection of cells with antibody-opsonized virus

Phylogenetic analysis of T/F strains and other selected variants from the standpoint of infecting animals with a viral swarm under different conditions is not an original experimental approach; it has been done many times prior by other labs in the field [Chen et al., 2006; Tortorec et al., 2008; Keele et al., 2009; Liu et al., 2010; Ma et al., 2011; Tsai et al., 2014; Yuan et al., 2017]. Thus, *in vitro* challenges of healthy donor cells with HIV-1 clinical isolates was the first real test of this novel approach. In testing Hypotheses #2 and #3, a sample of clinical isolate 92US657 was coated with antibody HIVIG or left uncoated in the presence of non-binding IVIG as a control were incubated with dendritic cells to saturate its surface cell receptors with bound virus, either through antibody-dependent or -independent means. Then those dendritic cells were cocultured, after thorough washing, with vulnerable target T-lymphocytes to mediate viral transfer. Unfortunately, the selective transfer of virus to stimulated T-lymphocytes were observed in an antibody-dependent or -independent manner. Even if a selection pattern were found, the presence of non-adhering viral variants in the "No antibody" group (which was absent in the "DC only" group) indicates that dendritic cells were not sufficiently washed prior to coculture and casts doubt as to whether such selection would really have been the result of *trans*-infection.

However, the identification of a minor lineage that was clonally expanded in the "DC only" group indicated a group of variants from a particular lineage that was favored for direct infection of these monocytes. These variants were distinguishable from others as they contained an L317I/D337E linked polymorphism in V3 and C3, respectively. The role these mutations play in selecting DC-preferring variants is not known, and they likely only played a role in the absence of T-lymphocytes where selection pressures could have favored less competitive variants like L317I/D337E-habording strains. While the etiologic source of such selection pattern is still unknown, the 75-fold expansion of variants with this particular phenotype in DC cells from the inoculum is strong evidence that this approach successfully captured an instance where an as-yet unidentified selection pressure favored a particular phenotype over others. Even if a bottleneck never occurred and even if these variants were not T/F strains, there isn't reason sufficient to discount the possibility that selection pressures associated with dendritic cells couldn't also play an important role in T/F selection. Further work would be required before such a role could be uncovered.

In vitro infection of cells with complement-opsonized virus

The next test of this experimental model was to repeat the prior experiments with complement opsonization instead of antibody opsonization of the virus with methodological corrections for issues encountered in the last study. Naked virus and complement-opsonized virus led to selective *trans*-infection of variants from two different lineages among V3 and V4V5 variants. One lineage was termed "C+" to indicate that selection favored these variants when opsonized by complement, while the other lineage was termed "C-" to indicate that these

variants were favored when naked. Stimulation of target cells modulated the strength of selection as evident by the increased representation of previously non-selected variants in each lineage during the *trans*-infection of T-lymphocytes with and without complement-opsonization. In V3, complement opsonization selected for an isoleucine at position 317 and against a phenylalanine at this position during *trans*-infection of unstimulated T-lymphocytes; stimulation altered the selection pattern so as to allow infection by many more variants containing the F317 polymorphism and far fewer variants containing the I317 polymorphism, which suggests that selection for variants in the "C+" variants could depend on the amino acid identity at this position. However, the distribution of mutations at position 317 appears unchanged between *trans*-infection with naked virus and the inoculum, and although stimulation does shift the variant distribution in favor of F317 selection of variants in the "C-" lineage during *trans*-infection might be due to polymorphisms in a different region.

For the V4 loop, *trans*-infection of unstimulated cells with and without opsonization led to differential selection patterns when cells were stimulated during *trans*-infection, and this enabled the identification of polymorphisms and mutation signatures specifically selected in the "C-" lineage. Variants from this lineage were defined by the V4 polymorphisms W395_W396ins(D/Y) (particularly in its hypervariable region), and stimulation largely disfavored those mutations in favor of the W395_W396ins(YR) mutation. It would appear that a subsequent arginine protected the N-terminal amino acid from purifying selection. While it is unclear whether these selected variants are genetic features of T/Fs or if their selection is one piece of larger network of alternative selection pressures favoring T/Fs, their clonal expansions from a lineage lowly represented in the inoculum is evidence that this model captured an

instance of selection and that further investigation of this could reveal greater insight into how complement opsonization may influence selection pressures in favor of T/F viruses during a clinical bottleneck.

In addition to complement-mediated selection, complement-dependent enhancement (not selection per se) was observed in the direct infection of stimulated T-lymphocytes, while it was absent in the *trans*-infection of stimulated T-lymphocytes. In fact, complement may have even been protective in the infection of unstimulated cells through DC-dependent transfer. Furthermore, complement-mediated binding to DC surface receptors likely has a saturation endpoint, beyond which the addition of more viruses does not enhance binding further. This could explain why complement opsonization can be protective in the *trans*-infection of unstimulated T-lymphocytes: coating a virus with complement limits how many viruses could be presented to T-lymphocytes either from surface receptors or synaptic transfer of viruses trapped in endosomal compartments. Stimulation of T-lymphocytes likely overcomes this merely on the basis that it increases CD4 and CCR5 expression on the surface of T-lymphocytes. Complement is also anti-pathogenic as it stimulates and directs immune responses to destroy coated pathogens; thus, selection of variants by complement opsonization likely depends on a sliding scale or "balancing act" where the selection pressures observed here can only be achieved when the level of activated complement is below the "kill" levels" but above the "selection threshold".

In vitro infection of cells initially primed with Type-1 interferon

In a final test of this model (and perhaps the most important one yet) is to prime T-lymphocytes with Ifnα2 prior to direct and *trans*-infection to see whether highly fit T/F-like

variants could be selected in cells made hostile to HIV replication, and resistance to Type-1 interferons is a common characteristic of T/F viruses [Lu et al., 2011; Fenton-May et al., 2013; Kane et al., 2013; Parrish et al., 2013; Iyer et al., 2017; Ashokkumar et al., 2020; Gondim et al., 2021; Sugrue et al., 2022]. In the first attempt of this experiment, T-lymphocytes were primed with Ifnα2 or with a vehicle control, then half was stimulated and the other half left unstimulated. After this, they were either cocultured with dendritic cells previously challenged with virus or they were directly challenged by the viral inoculum itself, this time 92US712. Only the V1V2 region was analyzed phylogenetically due to PCR failures in V3 and V4V5 with poorly designed primers. While phylogenetic analysis did not show bottlenecks in any of the *trans*-infection experiments and some selection pressures (albeit weak) in the direct infection of stimulated cells, strong replicate-specific genetic bottlenecks were observed during the direct infection of T-lymphocytes. These bottlenecks clustered to different lineages in the viral inoculum with different selected polymorphisms with only one shared by all three replicates: M147. They also share, to some degree, polymorphisms among different T/Fs but inferred strains selected in each replicate are as disparate from one another and from the T/Fs as the T/Fs are themselves, which altogether could be a recapitulation of the finding that T/Fs tend to be animal specific and do not always share the exact same mutation profile [Yuan et al., 2017]. In a repetition of this experiment using multiple clinical isolates (92US712, 92HT593, 92US657, 92US660, and JRCSF), patterns of highly selective infections and genetic bottlenecks were observed in SP, V1V2, V3, and V4V5 in both replicate-specific and replicate-shared manners. The number of these variants infecting T-lymphocytes increased, although not significantly, with stimulation of T-lymphocytes. The results here indicate that the interferon-priming of cells *in*

vitro established some form of selection pressure that favored uncommon variants from the inoculum, and that such pressures were attenuated by stimulation of target cells. Some of the selected polymorphisms of those variants, particularly in V1V2, were shared with some T/F viruses. In fact, the sequence of the SP region of JRCSF in the viral output of directly infected unstimulated cells clustered phylogenetically with T/Fs, a strong indication that not only did this *in vitro* model may have recapitulated and captured the selection conditions favoring T/F viruses during clinical infection but it likely captured at least one variant that phenotypically resembles a T/F virus. An analysis of the selected polymorphisms in the latter half of this particular study was not performed due to issues with sequence variation that cannot be accounted for as of yet, although those issues have been addressed and efforts to correct them are underway.

The Identification of Potential, but unconfirmed, T/F quasispecies

The prevailing question that is central to these studies is, as predicated in the three main hypotheses: "Are these selected variants novel, yet-to-be discovered T/Fs?" Answering this question is as complex as the analyses undertaken to uncover them, and whatever means it would take to address them likely wouldn't lead to any straightforward answers. Almost certainly the viruses found in the animal infection studies are T/F viruses, and they may even be novel T/F quasispecies that haven't yet been discovered. These viruses were likely selected in an antibody-dependent manner, and the sequences of those selected variants may provide valuable insights into how antibody might select for those particular variants and the polymorphisms of those variants could provide bioinformaticians a rational basis for finding epitopes that could be useful in vaccine design. However, in the path from infection through transcytosis to virus

dissemination after transcytosis, genetic selection largely occurs as a series of filters that start with a viral swarm and end with a single, clonally expanded T/F virus in a host body [Mosa, 2021]. These filters are virus-dependent and host-dependent, they can depend on the epithelial tissues at the site of transmission, and they can depend on immunological conditions (i.e. concurrent STIs) present at the time of transmission. The genetic bottleneck can have a selective component or a stochastic component, both co-occurring or occuring exclusively of one another, and they likely play important roles in shaping such bottlenecks. Mosa [2021] reviews how the genetic bottleneck is likely the result of a complex interplay of multiple factors and filters (whether propagated as a function of the virus or the host) that restrict the number of transmitted viruses from a viral swarm to a recipient, and that understanding the genetic bottleneck has been an immense challenge to researchers, especially from the standpoint of animal models, due to the impracticality of controlling every possible filter, especially those for which knowledge is limited. The culmination of these filters, whether known or unknown, act on viral swarms in a Black Box, if you will, where it would be nearly impossible to discern how such selection pressures were propagated after productive infection has already occurred.

The Black Box: The Central Problem of T/F Research in Animal Models

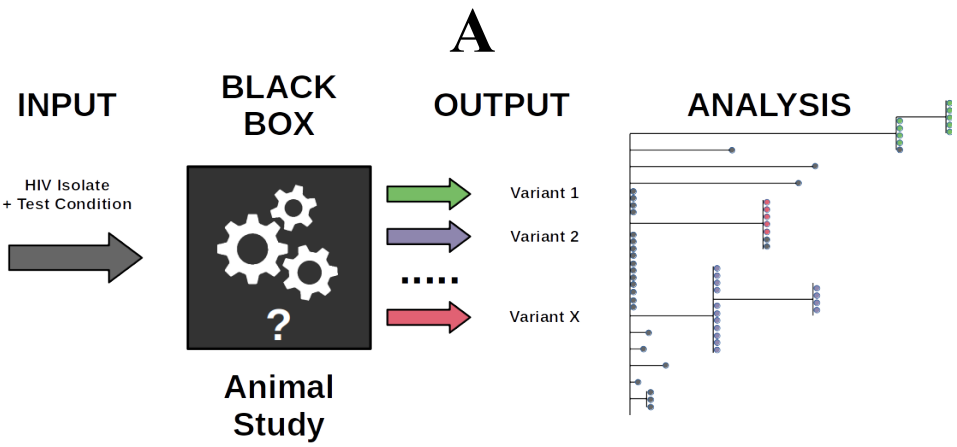
This recapitulates the central problem in T/F research from the beginning when the phenomenon of the genetic bottleneck was initially discovered in 2008. All information on T/Fs were found in reverse, where T/F quasispecies were inferred from phylogenetic analysis of tissue and blood samples from HIV-positive patients, and then laboratory testing of viral phenotypes led to *a priori* conjectures and speculations about how that particular variant was selected during

acute infection [Keele et al., 2008; Salazar-Gonzalez et al., 2008; Salazar-Gonzalez et al., 2009; Gnanakaran et al., 2011; Ochsenbauer et al., 2012; Parrish et al., 2013]. The problem of course is that there is no measurable probability assigned to or an empirical means of determining how likely those characteristics arose as a result of any particular selection pressure that may or may not have existed in tested patients; the lack of such insights presents serious problems to understanding and identifying T/Fs as well as the immunological and physiological conditions that gave rise to them.

Animal studies have afforded scientists a little more power in discerning how T/Fs can be selected from experimentally propagated selection pressures, but there is still a Black Box associated with every single animal study, as conceptualized in **FIGURE 56A**: an incomplete control of the an entire host body's immunological and physiological properties to fine tune selection pressures at cellular or molecular levels to see precisely how they select for those T/Fs. Worse yet, identifying novel T/Fs from an inoculum is still virtually impossible [Frange et al., 2013] even if it were possible to control for every permutation of conditions present during the infection of an animal. The *in vivo* transcytosis model using SIV, either opsonized or non-opsonized by antibody, in this study is no exception. The Black Box exists still, and it cannot be overcome in a model of limited manipulations. But what about a model of limitless possibilities in not only of manipulations of single factors but in permutations of multiple factors? This is where the *in vitro* model may become the next important tool in T/F research.

An inherent strength of the *in vitro* model is the ability to test any factor and combination of factors in any particular permutation the researcher desires with greater experimental control, as conceptualized in **FIGURE 56B**. It is a potentially cost-effective means with high statistical

power to study and identify novel T/Fs; it uses two well-accepted to approaches to studying HIV/SIV infection: (1) establishing productive infection in a tissue culture dish using donor tissue and cells and (2) phylogenetic analysis of output virus to determine whether test conditions selected particular variants from a viral inoculum. Downstream bioinformatics research could lead to novel discoveries, not only of new T/Fs but also of new conditions that may have influenced T/F selection in previous clinical cases that have as-yet been incompletely understood. This was the empirical aim of the studies presented here; so the prevailing question remains: "Are these selected variants novel, yet-to-be discovered T/Fs?"


A
INPUT
BLACK BOX
OUTPUT
ANALYSIS
HIV Isolate
+ Test Condition
?
Variant 1
Variant 2
.....
Variant X
Animal Study

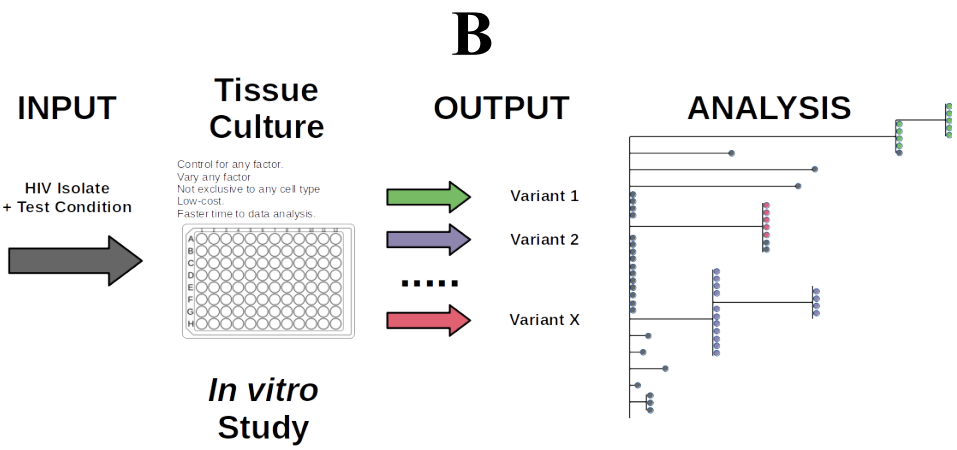

B
INPUT
Tissue Culture
OUTPUT
ANALYSIS
HIV Isolate
+ Test Condition
Control for any factor.
Vary any factor
Not exclusive to any cell type
Low-cost.
Faster time to data analysis.
Variant 1
Variant 2
.....
Variant X
In vitro Study

Figure 56 (previous page): Conceptual comparisons between animal studies and *in vitro* experiments to study the phylogenetics of HIV transmission: (A) Animal model systems are often used to study genetic bottlenecks in HIV/SIV transmission in the forward direction (i.e. challenging an animal with an infectious inoculum with some test condition and measure/characterize viral variants that are produced as a result of selection). The animal model is an experimentally sound approach to studying the genetic bottleneck, and it has provided researchers with important information on how selection pressures influence or rather shape the genetic bottleneck from a viral swarm. However, tracing the path from inoculum challenges to viral output is not so straightforward, as other selection pressures may exist in the organism that are beyond those directly under the researcher's control and that they may influence and shape those bottlenecks and output variants in ways that cannot be predicted or accounted for by the researcher. The animal, in a sense, represents a Black Box comprised of an internal mechanism largely unknown to the researcher, and therefore couldn't cannot be fully understood or controlled or even used to correlate the "INPUT" to the "OUTPUT." (B) *In vitro* model systems, on the other hand, are used quite often to study the molecular functions and characteristics of not just HIV-1 in general, but also of strains in particular like T/F quasispecies or clinical isolates. However, very few researchers have used them to model the genetic bottleneck in a manner resembling to what was attempted in the current studies. It is a versatile means of acquiring important information in the fastest turn-around time, offering the researcher greater control over variables and factors, and it can provide scientific insight on nearly equal grounds as animal models can. Although the researcher may not be able to control every possible factor or variable in cell culture, it still provides significant more control than the animal model; where genetic bottlenecks in the organism would be shaped potentially by all of the cellular and molecular components involved in propagating those selection pressures, *in vitro* model systems like those shown in (B) can be "dissected" and broken down into component parts in order to investigate how each component contributes to the genetic bottleneck specifically and in isolation from other components. Data acquired from the *in vitro* model wouldn't necessarily be any more accurate than data from the animal studies, it merely makes it easier to correlate the parameters of an *in vitro* experiment to its results than it is to correlate parameters in an animal Black Box to the results.

<u>These are not the T/Fs you're looking for, but they might be related</u>

The shortest, most accurate answer I can give is: "Maybe but I do not know." This answer should be insufficient for most researchers, but given the scope of this study, the aim was not so much to identify T/Fs from an inoculum under different conditions (although attempts were made regardless) but rather to determine whether it was feasible to select and clonally

amplify uncommon strains from an inoculum under different conditions. This was definitely found to be the case, and in some cases those clonally expanded variants shared selected polymorphisms with at least one of thirteen documented clade-B T/Fs. The identification of T/Fs from an inoculum was a secondary aim. In following this aim to a natural conclusion, phylogenetic analysis of the SP region of Ifnα2-primed unstimulated T-lymphocytes with a sample inoculum of JRCSF returned a bottleneck that clustered among T/F lineages more strongly than other inocula did. If this research design were capable of identifying T/Fs from an inoculum given some experimentally propagated selection pressure, T/F discovery may be improbable (though not impossible) given the low likelihood of selecting or predicting particular T/F variants from an inoculum [Frange et al., 2013] and the improbability that the exact same variant selected in one replicate would show up in a second or third replicate under the exact same condition [Yuan et al., 2017]. It isn't likely that the strains discovered here are bona fide T/F variants, but they don't have to be for this model to be instrumental in discovering viral epitopes conducive to a working vaccine. If desired, T/Fs could still potentially be discovered with fine tuning and perfecting the model system in successive implementations after improvement from failures encountered in previous implementations; this will be discussed in greater detail later on. For now, the important conclusions to draw from these results in response to the prevailing question are that (1) it is possible to test any combination of host-derived factors in the propagation of selection pressures that would selectively favor (or disfavor) certain variants from an infectious inoculum and that (2) those variants, while not strictly T/F quasispecies by any means, contain selected polymorphisms. that can be directly correlated to the selection pressures that established them in the first place. This model is already an

advantage over animal studies on the first conclusion in that it is nearly impossible to control for minor changes in individual factors that could modulate selection pressures propagated by major test conditions in an animal model and on the second conclusion in that it is nearly impossible to correlate selected polymorphism back to all possible combination of factors, however minutely that may have changed, in a Black Box at the time of infection.

Does this *in vitro* approach really simulate a clinical infection?

Another question that may be posed here is, provided that the experimental results returned T/F or T/F-like phenotypes, if at all, and provided that they were successful in showing selected polymorphisms under particular experimental conditions, how closely do these experiments simulate a clinical infection? This question is much more straightforward to answer than the question asked in the previous section; that is, "Not at all! And that's probably another strength of the model, rather than its weakness." It is true, this model and the hypotheses formed alongside were designed designed to replicate or simulate a clinical-like infection *in vitro*. But it is a prohibitively costly and methodologically complicated venture to simulate all of a host body's immune system and tissues in any single well of a tissue culture dish or plate. As we performed experiments with this model in practice, it became apparent that it may provide more valuable insight to focus on one or two factors at a time instead of having to reductively simulate all possible virus-host interactions inside of a single dish. It is certainly true that T/F selection likely occurs as a result of all these interactions working in tandem, with some contributing more than others depending on selection pressures at play, but isolating one factor at a time provides

greater success in determining how the isolated selection pressure played a role in the clonal expansion of particular variants over all others from an inoculum.

The complement opsonization experiments is a reasonable example of why isolating individual factors is more fruitful than testing a combination of factors *in vitro* simultaneously. It enabled the discovery of two different lineages that were differentially selected on the basis of whether virus was opsonized or naked. The addition of an additional factor, such as antibodies, let alone many more factors, may have obfuscated discovery of those complement-related lineages if the selection pressure for antibody-opsonized virus was stronger toward an entirely different lineage than the "C+" lineage or the "C-" lineage. This is not to say that such a discovery wouldn't have been useful though (discussed in more detail in a later section); the major point to be realized here is that testing individual factors rather than attempting to simulate a clinical infection in a "petri dish" is probably more valuable to epitope and polymorphism discovery than trying to uncover every possible factor of a Black Box and controlling them blindly in tissue culture.

The closest approximation to simulating clinical infections *in vitro* was by using healthy donor PBMCs in these experiments, instead of immortalized cell lines like SupT1 lymphocytes or THP-1 macrophages. Lab-grown cell lines are not an improper tool to use in certain experimental approaches, contingent on the empirical goals of the researcher, and in fact they have been used to garner important information on how HIV counteracts anti-viral responses in myeloid cells [Ikeda et al., 2019] and force cell cycle progressions in order to promote proviral integration and expression [Evans et al., 2018]. They have also been used to cultivate virus to high titers for cryoEM [Briggs et al., 2004; Moscoso et al., 2014; Gorman et al., 2022], cryoET

[Zhu et al., 2008; Earl et al., 2013], and X-ray diffraction studies [Huang et al., 2005]. Sarzotti-Kelsoe et al. [2014] used a $CD4^+$ T-lymphoblastoid line (A3R5) to improve on existing neutralization antibody assays for HIV-1, which has become an invaluable research tools for immunologists. However, macrophage lines like THP-1 and U937 do not respond to anti- and pro-inflammatory stimuli as strongly as PBMC-isolated macrophages do [Nascimento et al., 2022], making them less suitable to simulate any kind of clinical-like infection *in vitro* than using cells derived from a healthy donor. In every experiment performed, clinical isolates were used instead of lab strains of HIV. As with the case of immortalized macrophages in the lab, lab strains can be useful when used to test for specific virus-specific functions, like understanding how HIV-1, regardless of co-receptor tropism, adhere strongly to DC surface lectins [Geijtenbeek et al., 2000; Balzarini et al., 2004; Bouhlal et al., 2007] and to complement receptors [Bouhlal et al., 2007]. However, lab strains tend to be abnormally infectious toward cells that do not normally become infected during acute infection. Bouhlal et al. [2002] achieved infection of an immortalized epithelial cell line using different complement-opsonized HIV isolates while Chenine et al. [1998] conversely achieved a similar result in normal primary intestinal explants using laboratory strains of HIV-1. These authors suggested that complement opsonization [Bouhlal et al., 2002] or a syncytial-inducing phenotype of HIV-1 [Chenine et al., 1998] could mediate transcytosis in this way although no other report has yet to be found showing a clinical infection of normal epithelial cells was possible. For reasons such as these, laboratory strains like $HIV\text{-}1_{NDK}$, which was originally isolated from an infected patient in Zaïre [Spire et al., 1989], no longer resemble the wild-type features associated with clinical isolates of HIV-1, and using them for *in vitro* infection experiments with the aim of studying a clinical-like

genetic bottleneck would be counterintuitive. In fact, care should be taken to ensure that clinical isolates utilized in experiments such as these are not overpassaged to the point that they lose wild-type traits typical of HIV-1 during a clinical infection.

Another way in which these experiments attempt to simulate a clinical infection *in vitro* without burdening the system with an overly complex array of factors, tissues, and cells is the introduction of post-stimulated cells. During an acute infection, T-lymphocytes are rarely present at the time of initial infection. Although reports have previously shown that certain subsets of $CD4^+$ T-lymphocytes, like Th17, are present and become the first cells to become infected [Stieh et al., 2016], other reports have shown macrophages and dendritic cells to become the initial carriers of HIV-1, if not directly infected by them [Hu et al., 2000; Arthos et al., 2008; Koh et al., 2020]. In the case of the study by Stieh et al. [2016], animals infected in that study exhibited inflammation at the site of viral exposure, which could have artificially stimulated T-lymphocytes into migrating to the site of inflammation and becoming infected in the process. In any case, rarely T-lymphocytes would be stimulated unless activated by the macrophages and dendritic cells in response to an infection, as has been shown previously [Haaland et al., 2009]. Therefore, the purpose of the post-stimulated group was to infect T-lymphocytes with a clinical isolate of HIV-1 as if for the first time, and over the time frame of 24 hours they become stimulated in response to a growing viral infection. No other examples in literature have been found showcasing an experimental procedure such as this, which may be an indication that the post-stimulated model is an original attempt to create a clinical simulation of *in vitro* infection.

As stated already, these experiments are not at all true simulations of a clinical infection. They likely fall hopelessly short of it. However, the relevance of these results to clinical

infections are not so dependent on how closely an entire body can be replicated in a dish but rather how closely do the cells, viruses, and factors used in these experiments resemble the normal source. Are primary cultured cells being used, or are they immortalized cancer cell lines with severely altered morphologies, expression profiles, and karyotypes? Are clinical isolates of HIV-1 with limited passages being used, or are they laboratory strains passaged many times over years and expanded artificially to adapt under specific growth and infection conditions? Lastly, are cells being primed and treated in these experiments in a fashion that would resemble how they might behave or respond during a real clinical infection (like post-stimulation to simulate a gradual response to infection or stimulation to resemble an immune response to a concurrent infection [Haaland et al., 2009]), or are they always stimulated with the intention of enhancing viral output with little regard to how it may alter the results of such experiments and mislead researchers to the wrong conclusions? A simpler approach to study this phenomenon might provide more accurate insight than any Black Box or endlessly complicated *in vitro* experiments.

Limitations of this Model

A lot of effort has been made to show how this experimental model could recapitulate selection pressures that favor (or disfavor) particular variants, whether or not they share similarities with known T/F variants. However, there has yet to be a discussion of the limitations of this model. The uncertainty of whether these identified variants and polymorphisms are clinically relevant to T/Fs was implied previously. But this uncertainty is more likely due to sampling and chance than any real fault of the model itself. For example, direct infection of unstimulated, interferon-primed T-lymphocytes with the JRCSF inoculum showed a bottleneck

in the SP region that clustered with known T/F viruses. A single bottleneck was also found, as predicted, in the V3 region under the exact same replicate of the same infection condition. However, the V3 region in this bottleneck did not cluster with the V3 regions of the thirteen clade-B T/F viruses, like what was found in the SP region. This is probably because the SP region is more important for interferon-mediated selection of variants than the V3 region is, especially since the bottleneck in V3 clustered with the majority strain in the inoculum. The same, however, was not found in other infection conditions. Direct infection of post-stimulated, interferon-primed cells with the 92HT593 clinical isolate resulted in bottlenecks in the SP region for three replicates and in the V3 region for two replicates. However, these bottlenecks did not cluster with T/F variants; in fact, the bottlenecks of every PCR-amplifiable replicate of post-stimulated cells infected with 92HT593 clustered to the majority strain in the inoculum, suggesting stochastic infection rather than selection (or at the very least: purifying selection).

This doesn't necessarily falsify the hypothesis that interferon-mediated selection doesn't exist, or that the lack of consistency in variant capture in two different settings devalues the model's relevancy to studying genetic bottlenecks. Rather it could indicate that selection pressures can be so weak that even in the presence of interferons where T-lymphocytes naturally become inimical to HIV replication, breakthrough infections can still occur by virtue of numbers of particular variants rather than selection of more fit but rarer variants. Conversely, selection pressures can be so strong that infection is hardly ever achieved, even by the more fit T/F-like variants that may exist in clinical isolates. By induction, it can be surmised that this model is limited in its capacity to detect variants by the very nature of the selection pressures being tested: A factor being tested in one condition with one HIV-1 clinical isolate inoculating one set of cells

isolated from one particular donor may produce a wildly different result than the same factor being used in the same condition with the same HIV-1 clinical isolate inoculating the same set of cells isolated from a completely different donor. The same can also be true for experimental replicates: Cells in one replicate might respond to a factor differently than cells in a different replicate, even with all things being equal. This could be propagated by stochastic fluctuations in gene expression profiles between two different replicates, and these could be sufficient to produce unexpected responses to HIV infection. While these do not necessarily render the whole experimental model useless as implemented here, they do present real challenges in the replicability and relevancy of this model to clinical bottlenecks. And these challenges (or rather limitations) are inextricably linked to the way in which an experiment is designed, from the choice of clinical isolate to the isolation procedures to acquire test cells to be used in said experiments. Thus, there exist limitations that are either intrinsic or extrinsic to this experimental model, and an understanding of both is necessary for careful interpretation of results per the former and proper experimental and analytic approaches to account for issues encountered by the latter.

Intrinsic limitations of the model

In speaking of limitations that are "intrinsic" to this model system, it is important clarify what is meant by "intrinsic." "Intrinsic" here means a feature engineered into the model system. This could be in the choice of viral inoculum (i.e. JRCSF or SIVmac251). It could be the cell cultures themselves used as a target of infection (i.e. monocyte-derived macrophages or Langerhans cells expressing high levels of CD4 and CCR5). It could be the test condition

applied with viral inoculation (i.e. interferon-priming or interferon-priming with KD of several interferon-stimulated genes). Or it could be as simple as centrifuge-pelleting cells prior to inoculation or not. Each methodological choice can have an impact on the output data of that particular experiment merely by virtue of altering the selection pressures in a particular infection experiment by those components or designs.

This point can be illustrated better with an hypothetical example. Let's say a researcher wishes to understand how selection pressures in the brain could transfer select variants from the JRCSF clinical isolate (which was itself isolated from the cerebral spinal fluid of an AIDS patient with HIV-associated dementia [Koyanagi et al., 1987]). They set up a transcytosis experiment using a brain organoid as the tissue and challenge it with a sample inoculum of JRCSF. They analyze their data and find that a variants A, B, and C were found replicating in the organoid. They may repeat this experiment a second time and might find entirely new variants replicating in the tissues: variants X, Y, and Z. Future replicates might not ever produce the exact same variants as discovered previously but they might share some combination of previously discovered variants: like a replicate with variants A, X, and Z or another replicate with variants C, Y, and X. The researcher may be inclined to think that selection pressures do not exist over neurovirulent strains and that viral across the blood-brain barrier occurs sthochastically rather than selectrively, but this conclusion may only be true in the manner that it was tested under those particular conditions. The quality of results are only as limited as the experiment is designed, and fine-tuning the experimental parameters to test smaller, more specific components (such as endothelial cells or microglial cells) of the organoid tissue might produce entirely different but more reproducible results. Maybe doing so would produce varied

results where at least variants A and B was consistently found in every replicate with the other four being minor outliers. And in these two experiments, the limitations intrinsic to both are so different that they influence the results in vastly different ways.

This issue was encountered in the direct infection studies where T-lymphocytes were primed with Ifnα2, and some replicates clustered to an uncommon variant from the inoculum while another clustered to the most common variant. The number of selected variants were more numerous in stimulated T-lymphocytes and control cells (not treated with Ifnα2) compared to unstimulated T-lymphocytes primed with Ifnα2. This was likely due to a loss of selection pressures during T-lymphocyte stimulation, and it could explain why concurrent infections often result in the transmission of multiple T/F variants in a more stochastic manner [Haaland et al., 2009]. More importantly, clinical infection by multiple T/F variants often resulted in the clonal expansion of viruses with lower replicative capacity [Macharia et al., 2020], which can mean that additional variants captured in this model by a loss of selection pressure due to T-lymphocyte stimulation could have also had lower replicative fitness, and the researcher would not know unless they specifically tested the replication capacity of those output variants They also likely wouldn't have known that a lower replication capacity was the result of stimulation or some other uncontrolled minute changes in the model system. The selection (or rather absence of selection) that led to accumulation of these (probably) less fit viruses was entirely the result of the way these experiments were designed and performed. It follows from this that any experimental iteration of these approaches, depending on the parameters employed in concert, ultimately lead to outcomes that are the direct result of not only known parameters but unknown parameters intrinsically part of the experiment itself. In a sense, the intrinsic limitations are

essentially the Black Box of any *in vitro* model system employed here to study the HIV-1 genetic bottleneck. Overcoming these Black Boxes may not always be possible in every experiment. Indeed, every new *in vitro* experiment is bound to have some immeasurable uncertainty or inaccuracy linked to its design. Unlike the Black Box of animal studies, *in vitro* Black Boxes are more easily disassembled into constituent parts and tested in isolation of one another in experiments with ever decreasing sizes of those Black Boxes—or rather fewer and fewer intrinsic limitations. Breaking down the system into yet smaller components and studying how each component contributes to the larger whole would enable the researcher to have greater power in interpreting results and understanding the system being studied.

Extrinsic limitations of the model

As discussed already, an intrinsic limitation is any feature engineered into an experiment that impacts the results of those experiments in ways that the researcher may not be aware. Intrinsic limitations are inescapable in every *in vitro* experiment, but experiments in sequence with gradual modifications to deal with experimental limitations of the prior trial would lead to more reproducible results and rationally sound conclusions. On the other hand, there are limitations in a particular experiment (or rather how data is collected from the experiment) that can influence those results in equally subtle ways. These are extrinsic limitations are factors that are outside of or not directly part of an experiment but can still produce uncertainty in the results. This would be more commonly encountered as the way in which data is collected; it doesn't influence or impact the true outcome of an experiment but it definitely can still influence the interpretation of experimental results.

An extrinsic limitation can be visualized by reference to the prior example of a researcher studying cerebral selection pressures favoring neurovirulent strains from an infectious inoculum. Let's say the researcher exposed two different cell types from the brain (endothelial cells and microglial cells) using a sample inoculum of JRCSF. Next, the researcher wants to PCR-amplify, sequence, and phylogenetically analyze variants preferentially replicating or penetrating both cell types. They use primers for Illumina sequencing and they get sequence data on *HIVpol* and *HIVenv*. The researcher finds variants specific to both *HIVpol* and *HIVenv* that are more highly represented in microglial cells than in endothelial cells, and the representation of those variants or mutation signatures are 100x enhanced whatever their levels were in the original inoculum. This is an important discovery, but the researcher doesn't know which polymorphisms in *HIVpol* are linked to which in *HIVenv*. Even worse, the researcher doesn't have a reliable means of knowing which polymorphisms are the result of PCR errors or true genetic diversity. The inability to know whether a particular polymorphism A in *HIVpol* is genetically linked to polymorphism B in *HIVenv* is not a problem intrinsic to the experiment itself; this is a problem of data validation that would require some other resolution beyond or "outside" the experiment itself. After careful consideration and learning about how other scientists dealt with these issues, they decide to perform single genome amplification (SGA) using primers specific that contain unique molecular identifiers (UMIs) in order to distinguish between real mutations and PCR-/sequencing-induced errors. In this way, the extrinsic limitation that influenced the results was in the primer design, not so much the experimental design, and that it can be overcome by designing new primers to better validate the data acquired from this experiment. The experiment

itself likely didn't have to be repeated to remove or at least improve on the errors that resulted from extrinsic issues.

A similar issue was encountered in the current study. In fact, the possibility of genetic disequilibrium was constantly found throughout every result in every study (minus the *in vivo* animal work). Variants that were particularly infectious toward dendritic cells contained two highly represented selected polymorphisms L317I and D337E of V3 amplicons. In fact, viral variants with these polymorphisms were represented in the inoculum and the "no antibody" control (that is, coculture between T-lymphocytes and dendritic cells) at nearly equal percentages, suggesting quite strongly that some genetic aspect of these variants were selected and preferred for direct infection of DCs. But it is uncertain whether the cause was one or both of these polymorphisms or some other polymorphism linked elsewhere in the HIV genome that had been part of some other unsequenced, non-PCR-amplified region of the genome, like *HIVpol* or *HIVnef*. Even though V4 and V5 polymorphisms were also found associated with DC-infecting variants, the separate amplifications of both V3 and V4V5 regions disconnected the genetic linkages between V3 polymorphisms in one variant from those in V4V5 of the same variant. Another example of this problem was in the *trans*-infection of unstimulated T-lymphocytes with complement-opsonized virus. Polymorphisms were found in V3, V4, and V5 loops, but it wasn't clear whether selection pressures favoring those variants were acting on precisely those polymorphisms. They might very well be acting on a different part of the envelope spike protein; complement has been shown to interact specifically with certain regions of gp41 [Ebenbichler et al., 1991], but this region of the *HIVenv* gene was not sequenced and it isn't known whether selection pressures propagated by complement opsonization might have

favored mutations in gp41 more strongly than mutations in V3, V4, and V5 of gp120. For these reasons, it became impossible to determine whether polymorphisms were linked and therefore exceedingly more difficult to determine correlates of selection by reference to one or even several polymorphisms across multiple regions separately amplified.

Another issue encountered that turned out to be an extrinsic limitation is the high genetic variability among the experimental outcomes of the last interferon experiment, particularly in 92US660 and 92US712. It became clear that this high genetic variability wasn't the result of APOBEC-driven hypermutation; had it been, the intrinsic limitation here could easily be overcome by correcting mutations in post or removing sequences with hypermutation signatures. But that wasn't the case: It wasn't very clear whether these mutations were the result of PCR errors, and therein lies the intrinsic limitation. The uncertainty of knowing the source of these mutations complicated the interpretation of these results and the search for selected polymorphisms. This limitation can be overcome using a primers with specific barcodes (or UMIs) that would allow one to track sequence reads with a particular barcode and build consensus reads based on those barcodes [Zhou et al., 2021; Zhou et al., 2022]. This would collapse all reads down to a single consensus for each tagged genetic variant, thus removing PCR artefacts from the dataset. In doing so, such an intrinsic limitation could be overcome and analysis can finally be performed to the next logical conclusions.

A convenient rule of thumb

A good rule of thumb to discern "intrinsic" from "extrinsic" limitations is this: if one must repeat an entire experiment with parameters in order to correct a problem or limitations

encountered in the last time it was performed, then one is dealing with an "intrinsic limitation." If one must change some aspect of how data is collected and analyzed from an experiment without necessarily repeating it with new parameters in order to improve data quality or interpretation, then one is dealing with an "extrinsic limitation."

Practice makes perfect: Why this Model is Relevant

Future Directions

Despite the discoveries made in these experiments (such as polymorphisms that were positively selected in particular infection conditions), there is still much to be learned about the genetic bottleneck and how particular physiological and immunological factors may be responsible for it. It isn't known whether selected polymorphisms were themselves acted upon by selective pressures or if they only appear that way by linkage disequilibrium to unsequenced or uncaptured polymorphisms. It also isn't known whether additional cellular properties may have been responsible for the observed selection pressures in these studies, such as minor fluctuations in gene expression in particular cells that may have tipped the balance of selection in favor of one minor variant over another one. Indeed, despite evidence of strong selection, many variants appear to become predominant because of their high representation in the inoculum (implying infection bias toward a particular variant by non-selective means). In the animal studies, for example, the majority of animals were infected stochastically rather than selectively; the consensus T/F infecting many animals were identical to the inoculum consensus CON.i, even among antibody-treated animals. Removal of stochastically transmitted variants enabled the identification of antibody-selected variants and the selected polymorphisms NKSK. It remains

unclear whether this was truly selected by antibody opsonization and that it was subject to weak selection pressure, but followup experiments would be necessary to confirm whether selection of this polymorphism is reproducible. The same is true for all the other results acquired from the various experiments performed with the *in vitro* model system: Selected variants and polymorphisms must be confirmed by testing their selection in similar conditions.

Competition assays would be a good experimental approach here. This would be done by developing an infectious molecular clone for a selected variant or containing a selected polymorphism under a particular condition and serially diluted purified virions with an infectious inoculum. Then repeating the infection assay in multiple replicates to determine if the selected variant or polymorphism is clonally expanded among output viruses. A rescue experiment can be performed where a selected polymorphism would be cloned into a non-selected variant to determine if the polymorphism would sufficiently increase the fitness of non-selected viruses under a particular infection condition (i.e. complement opsonization or interferon priming).

Another important question that has arisen from these studies is how interferon priming establishes the selection pressures favoring interferon-resistant variants? There are clues in literature that generally show that interferon-resistant T/Fs have higher replication capacity than non-transmitted variants in the presence of interferon [Fenton-May et al., 2013; Iyer et al., 2017; Ashokkumar et al., 2020; Gondim et al., 2021; Sugrue et al., 2022]. Other reports in literature show that HIV-1 expresses anti-restriction factors that exhibit particular resistance against interferon-stimulated genes, like TRIM5 [Stremlau et al., 2004; Grutter et al., 2012], SAMHD1 [Goldstone et al., 2011; Laguette et al., 2011], APOBEC-3G [Wang et al., 2008; Ikeda et al., 2019], GBP5 [Krapp et al., 2016], IFITM [Lu et al., 2011], Tethrin [Neil et al., 2008], and MX2

[Kane et al., 2013]. Since viral samples from 5 dpi, 7 dpi, and 10 dpi are currently on hand, it wouldn't be unfeasible to design new primers to PCR-amplify, sequence, and phylogenetically analyze interferon-selected variants in the various anti-restriction factors expressed by HIV-1, like Vpr, Vif, Vpu, and Nef (the complex molecular interactions between restriction factors and HIV-1 genes are reviewed by Ramdas et al. [2020] and Coiras et al. [2010]). Knowing whether polymorphisms are selected in any of these non-envelope genes might provide better insights into the mechanism of interferon-resistance not only found in these experiments but among T/F viruses as well. Additional rescue experiments could be performed to determine whether those polymorphisms cloned into non-selected variants increase the replication fitness of those viruses in the presence of Ifnα2.

Finding T/F-like variants from a diverse swarm was an aim of these experiments. In performing these experiments, variants have been found that not only share polymorphisms with one or several of 13 known clade-B T/F viruses but some even cluster with those T/Fs on a phylogenetic tree. This was particularly the case in interferon experiments. However, these instances were far and few between, and most bottlenecks found were probably due to stochastic infection rather than selective infection. This represents an intrinsic limitation, where lack of T/F-like phenotypes found in these experiments may have been due to their absence from the clinical isolates. These isolates were acquired from patient samples early during acute infection, so they theoretically should still contain some trace, however non-abundant, in those clinical isolates. The amount of T/Fs still present in the clinical isolates used here are not known, in fact. It is possible that over successive passages of viruses for each clinical isolate, the amount of clinical T/Fs may have dwindled or accumulated mutations to such a degree that they no longer

resemble the T/Fs from which they evolved. Thus, a final experiment to be performed here (and probably more importantly, a final test of this experimental model) would be to perform a competition assay where an infectious molecular clone of T/F viruses, such as those developed by Ochsenbauer et al. [2012] and Parrish et al. [2013], could be expressed and serially diluted in any number of clinical isolates and tested in competition assays to observe whether these T/Fs can be selected and clonally expanded under particular infection conditions. A confirmation of T/F expansion during competition assays wouldn't necessarily "prove" the strength of this model; in fact, T/Fs by their phenotypic properties exhibit higher replicative fitness in the presence of interferon than their non-transmitted counterparts [Sugrue et al., 2022]. It would however show that the model is sensitive enough to capture known interferon-resistant strains from a viral swarm, and this would be important since the goal here is to model the genetic bottleneck *in vitro*. Therefore, the last set of experiments proposed here would be to perform competition assays with variants in a clinical isolate containing a serial dilution of T/F quasispecies.

Improving the Search for a Vaccine

Reviewed and discussed to considerable detail in Chapter 1 under the section "Vaccine Trials and their Failures," there has been considerable work toward the development of HIV vaccines. Every vaccine trial has failed, except for one: the RV144 trial from 2003 – 2009, which resulted in no greater than 31% efficacy [Rerks-Ngarm et al., 2009]. The literature is replete with clues for why developing an effective vaccine has been so difficult, with some reports suggesting ADE as a reason [Robinson et al., 1987; Forthal et al., 2012], some suggesting

CDE was another reason [Dolin et al., 1991; Thieblemont et al., 1993b], and others suggesting broad phenotypic differences between vaccine epitopes and wild-type strains as possible reasons [Kijak et al., 2013; Karnasuta et al., 2017]. Very little is known about whether using epitopes from T/Fs would improve vaccine designs; in fact, there is very little information found in literature on the subject of how effective such a vaccine would be compared to those that failed. Given that T/Fs are selected in severe genetic bottlenecks during acute infection by some selective pressure as yet poorly understood and that it is precisely those T/Fs (or some common epitopes that distinguish them from non-transmitted variants) for which a successful vaccine must be based on, then it follows that T/F research is absolutely crucial to vaccine development and deployment. But before researchers can take on such ventures, the way vaccines are developed from immunological data must be discussed. The success of a selected vaccine epitope is contingent upon the accuracy and statistical power of the approach to developing vaccines.

There are two major approaches to vaccine design: rational and empirical. The rational design often involves identifying an array of broadly neutralizing antibodies from an HIV-infected patient that show promise in neutralizing a panel of viruses, including T/F strains (reviewed by Rueckert et al. [2012], Haaren et al. [2017], and Sharma et al. [2019]). The central argument in defense of this approach is that if a broadly neutralizing antibody (bnAb)was already developed in Nature, and that it led to a demonstrable suppression of viremia in infected patients, then perhaps those antibodies could be used to form immune complexes with viral spike proteins, viral epitopes, or live-attenuated viruses and enable training of B cells to produce those very same bnAbs and immunize the vaccinated individual. The design can be improved upon in

iteration from animal studies or *in vitro* experiments, but the important distinguishing point here is that rational design depends on bnAbs already developed and acquired from an infected individual and using that to inform the design of a successful vaccine. This is not a bad approach; in fact, research using the rational approach has produced some promising results in HIV vaccinology [Sanders et al., 2015; Gach et al., 2019]. However, it may not be the best approach to vaccine design.

The inherent problem with the rational design is that it relies too much on the forward direction, and it oversimplifies antibody functionality, reductively assigning "yes or no" answers to what are otherwise incredibly complex questions. For example, given that bnAbs effectively suppressed viremia in some patients, how does it work? Which effector functions are these bnAbs enhancing? What titer must they be expressed in an individual to effectively foment that response? Would these work against all viruses or just the ones that the sampled individual happened to be infected with? Answers to these questions may not be available, but the rational design relies mainly on information that *is* available, and it may not be enough. Furthermore, the rational design is not typically suited to dealing with what Regenmortel [2018] calls "The Inverse Problem," where an experiment produces unexpected deviations from a logically sound causal chain of events. On the assumption that a rationally designed vaccine *should* be effective, the researcher may encounter cases where it is *not* effective at all, and they have to find the reason for the experimental failure or determine that it is was the result of a statistical outlier and recommend that it was due to an experimental aberration rather than a real, measurable biological phenomenon.

In contrast to the rational design, the empirical design is a longer, usually more arduous approach to developing a vaccine than the rational design. Where the rational design lets Nature do all the work in the forward direction, so to speak, and researchers capitalize on the presumed success of those antibodies, the empirical design requires that scientists search for working vaccine epitopes almost from scratch and perform experiments, both *in vitro* or *in vivo*, in sequential order to improve on currently existing designs (reviewed by Rueckert et al. [2012] and Regenmortel [2018]). There is nothing atypical about this approach; it is the hallmark of the scientific method, and the empirical approach to vaccine development is nothing more than a reinvitation of the scientific method to study epitopes, determine what formulations produce promising results, explain observed phenomena (including deviations), and then repeat experiments with modified parameters in order to optimize vaccine designs [Rueckert et al., 2012; Regenmortel, 2018; Sharma et al., 2019].

One way in which this experimental model can be beneficial to vaccine development would be its use as an empirical tool to test the efficacy of particular epitopes and then optimize the design of those epitopes and immune complexes formed from them to produce more robust effector responses against HIV-1, particularly against T/F strains. It also wouldn't be exclusive to the empirical design; it can be of use to researchers using the rational design and assist them in solving "Inverse Problems" warned by Regenmortel [2018]. bnAbs can be tested in this *in vitro* model in order to evaluate vaccine efficacy, and its low-cost, capacity to use clinical isolates and primary tissue cultures, and propensity to high repetition of experiments would make it an invaluable tool for researchers to not only find a new vaccine but fully study its efficacy and any observed deviations that wasn't anticipated in the rational design.

A new analytic approach with a scope that goes beyond HIV

As stated previously, the inherent power of this model system is in its well-validated experimental implementation. This is to say that *in vitro* culture systems have been used extensively in scientific experiments for decades, and its implementation here does not deviate from how these experiments are normally performed. The novel aspect here is that it attempts to model an otherwise complex virological phenomenon known as the genetic bottleneck *in vitro* where it has never been attempted before (as of December 2023 of writing this section). Indeed, it has produced bottlenecks as well as some deviations that couldn't be accounted for with the available data, but future iterations of this model system can be improved with modified parameters, new clinical isolates, testing new infection conditions, primers optimized to specific strains and genomic regions, and utilizing any donor tissue or cells so desired for these experiments. Its greatest strength is its versatility. There is nothing particularly special or exclusive about this model that it cannot be performed by any modestly equipped laboratory with access to even the most basic of reagents. There is also nothing to prevent researchers from applying this approach to model genetic bottlenecks for other viruses *in vitro*, such as hepatitis C virus (HCV) or influenza. In fact, there is evidence of a genetic bottleneck during HCV transmission [da Silva et al., 2017]. Thus, this experimental model system, designed initially as a means to study the genetic bottleneck of HIV transmission and the selective pressures that propagate it, is freely given to scientists of the world in the search for cures and vaccines against not only virus but countless other pathogens.

REFERENCES

Abrahams, M. et al. (2009) Quantitating the Multiplicity of Infection with Human Immunodeficiency Virus Type 1 Subtype C Reveals a Non-Poisson Distribution of Transmitted Variants. **83**(8):3556 – 3567.

Adeniji, O. et al. (2021) Siglec-9 defines and restrains a natural killer subpopulation highly cytotoxic to HIV-infected cells. *PLoS Pathogens*. **17**(11):e1010034.

Affi, R. et al. (2022) Plasma sVCAM-1, antiretroviral therapy and mortality in HIV-1-infected West African adults. *HIV Medicine*. **23**:717 – 726.

Aldunate, M. et al. (2013) Vaginal concentrations of lactic acid potently inactivate HIV. *Journal of Antimicrobial Chemotherapy*. **68**:2015 – 2025.

Alexander, M. et al. (2010) Donor and Recipient Envs from Heterosexual Human Immunodeficiency Virus Subtype C Transmission Pairs Require High Receptor Levels for Entry. *Journal of Virology*. **84**(8):4100 – 4104.

Altschul, S. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. **25**(17):3389 – 3402.

Anderson, M. et al. (1993) Analysis of Envelope Changes Acquired by SIVmac239 during Neuroadaption in *Rhesus macaques*. *Virology.* **195**(2):616 – 626.

Anderson, D. et al. (2011) HIV Infection and Immune Defense of the Penis. *American Journal of Reproductive Immunology*. **65**:220 – 229.

Andersson, G. et al. (2020) Stigma reduction interventions in people living with HIV to improve health-related quality of life. *Lancet HIV*. **7**:e129 – e140.

Andrews, A. et al. (2018) Analysis of HIV-1 envelope evolution suggests antibody-mediated selection of common epitopes among Chinese former plasma donors from a narrow-source outbreak. *Scientific Reports*. **8**:5743.

Anwar, A. et al. (2022) Impact of Caveolin-Mediated Endocytosis on the Trafficking of HIV within the Colonic Barrier. *Journal of Virology*. **96**(7):e0020222.

Arhel, N. (2010) Revisiting HIV-1 uncoating. *Retrovirology*. **7**:96.

Arthos, J. et al. (2008) HIV-1 envelope protein binds to and signals through integrin α4β7, the gut mucosal homing receptor for peripheral T cells. *Nature Immunology*. **9**(3):301 – 309.

Ashokkumar, M. et al. (2016) Variability in V1V2 and PNGs in Pediatric HIV-1 Viral Variants Transmitted Through Vertical Route. *AIDS Research and Human Retroviruses*. **32**(10):942 – 943.

Ashokkumar, M. et al. (2018) Unique Phenotypic Characteristics of Recently Transmitted HIV-1 Subtype C Envelope Glycoprotein gp120: Use of CXCR6 Coreceptor by Transmitted Founder Viruses. *Journal of Virology*. **92**(9):e00063-18.

Ashokkumar, M. et al. (2020). *In vitro* replicative fitness of early Transmitted founder HIV-1 variants and sensitivity to Interferon alpha. *Scientific Reports*. **10**:2747.

Asmal, M. (2015) Infection of monkeys by simian-human immunodeficiency viruses with transmitted/founder clade C HIV-1 envelopes. *Virology*. **475**:37 – 45.

Atashili, J. et al. (2008) Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. *AIDS*. **22**:1493 – 1501.

Au-Yeung, B. et al. (2014) A sharp T-cell antigen receptor signaling threshold for T-cell proliferation. *PNAS*. **111**(35):E3679 – E3688.

Bacchus, C. et al. (2013) A single HIV-1 Cluster and a Skewed Immune Homeostasis Drive the Early Spread of HIV among Resting $CD4^+$ Cell Subsets within One Month Post-Infection. *PLoS One*. **8**(5):e64219.

Baeten, J. et al. (2012) Antiretroviral Prophylaxis for HIV Prevention in Heterosexual Men and Women. *NEJM*. **367**(5):399 – 410.

Baiyegunhi, O. et al. (2022) CD8 lymphocytes mitigate HIV-1 persistence in lymph node follicular helper T cells during hyperacute-treated infection. *Nature Communications*. **13**:4041.

Balandya, E. et al. (2010) Semen Protects $CD4^+$ Target Cells from HIV Infection by Promotes the Preferential Transmission of R5 Tropic HIV. *Journal of Immunology*. **185**:7596 – 7604.

Balzarini, J. et al. (2004) Mannose-Specific Plant Lectins from the *Amaryllidaceae* Family Qualify as Efficient Microbicides for Prevention of Human Immunodeficiency Virus Infection. *Antimicrobial Agents and Chemotherapy*. **48**(10):3858 – 3870.

Bar, K. et al. (2012) Early Low-Titer Neutralizing Natibodies Impede HIV-1 Replication and Select for Virus Escape. *PLoS Pathogens*. **8**(5):e1002721.

Barré-Sinoussi, F. 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.

Baxter, J. et al. (2023) Inferring the multiplicity of founder variants initiating HIV-1 infection: a systematic review and individual patient data meta-analysis. *Lancet Microbe*. **4**:e102 – e112.

Bbosa, N et al. (2019a) HIV subtype diversity worldwide. *Current Opinion on HIV and AIDS.* **14**(3):153 – 160.

Bbosa, N. et al. (2019b) Phylogeography of HIV-1 suggests that Ugandan fishing communities are a sink for, not a source of, virus from general populations. *Scientific Reports*. **9**:1051.

Becker, Y. (2004) HIV-1 gp120 Binding to Dendritic Cell Receptors Mobilize the Virus to the Lymph Nodes, but the Induced IL-4 Synthesis by $Fc\varepsilon RI^+$ Hematopoietic Cells Damages the Adaptive Immunity – a Review, Hypothesis, and Implications. *Virus Genes*. **26**(1):147 – 165.

Berger, E. et al. (1998) A new classification for HIV-1. *Nature*. **391**:240.

Bertram, K. et al. (2019) Identification of HIV transmitting $CD11C^{+}$ human epidermal dendritic cells. *Nature Communications*. **10**:2759.

Binley, J. et al. (2006) Inhibition of HIV Env binding to cellular receptors by monoclonal antibody 2G12 as probed by Fc-tagged gp120. *Retrovirology*. **3**:39.

Biscone, M. et al. (2006) Functional impact of HIV coreceptor-binding site mutations. *Virology*. **351**:226 – 236.

Bobadilla, S. et al. (2013) Efficient transduction of myeloid cells by an HIV-1-derived lentiviral vector that packages the Vpx accessory protein. *Gene Therapy*. **20**:514 – 520.

Boeras, D. et al. (2011) Role of donor genital tract HIV-1 diversity in the transmission bottleneck. *PNAS*. **108**(46):E1156 – E1163.

Bolyen, E. et al. (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME2. *Nature Biotechnology*. **37**:848 – 857.

Bomsel, M. (1997) Transcytosis of infectious human immunodeficiency virus across a tight human epithelial cell line barrier. *Nature Medicine*. **3**(1):42 – 47.

Bomsel, M et al. (1998) Intracellular Neutralization of HIV Transcytosis across Tight Epithelial Barriers by Anti-HIV Evelope Protein dIgA or IgM. *Immunity*. **9**:277 – 287.

Bost, K. et al. (1988) Individuals infected with HIV possess antibodies against IL-2. *Immunology*. **65**:611 – 615.

Bouhlal, H. et al. (2002) Opsonization of HIV-1 by Semen Complement Enhances Infection of Human Epithelial Cells. *Journal of Immunology*. **169**:3301 – 3306.

Bouhlal, H. et al. (2007) Opsonization of HIV with Complement Enhances Infection of Dendritic Cells and Viral Transfer to CD4 T Cells in a CR3 and DC-SIGN-Dependent Manner. *Journal of Immunology*. **178**:1086 – 1095.

Boyd, M. et al. (1997) Discovery of Cyanovirin-N, a Novel Human Immunodeficiency-Inactivating Protein That Binds Viral Surface Envelope Glycoprotein gp120: Potential Applications to Microbicide Development. *Antimicrobial Agents and Chemotherapy.* **41**(7):1521 – 1530.

Boyer, V. et al. (1991) Complement Mediates Human Immunodeficiency Virus Type 1 Infection of a Human T Cell Line in a CD4- and Antibody-independent Fashion. *Journal of Experimental Medicine*. **173**:1151 – 1158.

Bracq, L. et al. (2018) Mechanisms for Cell-to-Cell Transmission of HIV-1. *Frontiers in Immunology*. **9**:260.

Brese, R et al. (2018) Ultradeep signle-molecule real-time sequencing of HIV envelope reveals complete compartmentalization of highly macrophage-tropic R5 proviral variants in brain and CXCR4-using variants in immune and peripheral tissues. *Journal of Neurovirology*. **24**(4):439 – 453.

Briggs, J. et al. (2004) The stoichiometry of Gag protein in HIV-1. *Nature Structural and Molecular Biology*. **11**(7):672 – 675.

Brown, E. (1991) Complement receptors and phagocytosis. *Current Opinion in Immunology*. **3**:76 – 82.

Brown, T. and Peerapatanapokin, W. (2019) Evolving HIV epidemics: the urgent need to refocus on populations with risk. *Current Opinion on HIV and AIDS*. **14**(5):337 – 353.

Bull, M. et al. (2013) Human Immunodeficiency Viruses Appear Compartmentalized to the Female Genital Tract in Cross-Sectional Analyses but Genital Lineages Do Not Persist Over Time. *Journal of Infectious Diseases*. **207**:1206 – 1215.

Burman, W. et al. (2008) Episodic antiretroviral therapy increases HIV transmission risk compared to continuous therapy: results of a randomized controlled trial. *Journal of Acquired Immune Deficiency Syndrome*. **49**(2):142 – 150.

Burnie, J. and Guzzo, C. (2019) The Incorporation of Host Proteins into the External HIV-1 Envelope. *Viruses*. **11**:85.

Busch, M. (1988) Laboratory Diagnosis of HIV Infection. *Transfusion Medicine Reviews*. **2**(4):250 – 263.

Butler, D. et al. (2009) The efficiency of single genome amplification and sequencing is improved by quantitation and use of a bioinformatics tool. *Journal of Virological Methods*. **162**:280 – 283.

Cantres-Rosario, Y. et al. (2019) HIV Infection Induces Extracellular Cathepsin B Uptake and Damage to Neurons. *Scientific Reports*. **9**:8006.

Cardozo, T. et al. (2007) Structural Basis for Coreceptor Selectivity by the HIV Type 1 V3 Loop. *AIDS Research and Human Retroviruses*. **23**(3):415 – 426.

Carlson, J. et al. (2014) Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science*. **345**(6193):1254031.

Caskey, M. et al. (2015) Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature*. **522**(7557):487 – 491.

Ceballos, A. et al. (2009) Spermatozoa capture HIV-1 through heparan sulfate and efficiently transmit the virus to dendritic cells. *Journal of Experimental Medicine.* **206**(12):2717 – 2733.

Chakrabarti, L. et al. (1987) Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. *Nature*. **328**(6130):543 – 547.

Chateau, M. et al. (2013) Rectal Transmission of Transmitted/Founder HIV-1 Is Efficiently Prevented by Topical 1% Tenofovir in BLT Humanized Mice. *PLoS One*. **8**(3):e60024.

Chen, M. et al. (2006) Simian immunodeficiency virus envelope compartmentalizes in brain regions independent of neuropathology. *Journal of Neurovirology*. **12**:73 – 89.

Chen, Y. et al. (2013) Binding of HIV-1 gp120 to DC-SIGN Promotes ASK-1-Dependent Activation-Induced Apoptosis of Human Dendritic Cells. *PLoS Pathogens*. **9**(1):e1003100.

Chen, J. et al. (2018) Identification of Unequally Represented Founder Viruses Among Tissues in Very Early SIV Rectal Transmission. *Frontiers in Microbiology*. **9**:557.

Chen, B. (2019) Molecular Mechanism of HIV-1 Entry. *Trends in Microbiology*. **27**(1):878 – 891.

Chenine, A-L. et al. (1998) Primary Intestinal Epithelial Cells Can Be Infected with Laboratory-Adapted Strain HIV Type 1 NDK but Not with Clinical Primary Isolates. *AIDS Research and Human Retroviruses*. **14**(14):1235 – 1238.

Chohan, B. et al. (2005) Selection for Human Immunodeficiency Virus Type 1 Envelope Glycosylation Variants with Shorter V1-V2 Loop Sequences Occurs during Transmission of Certain Genetic Subtypes and May Impact Viral RNA Levels. *Journal of Virology*. **79**(10):6528 – 6531.

Chorin, E. et al. (2014) Peripheral blood mononuclear cells of HIV-infected patients contain CD8 T cells that form conjugates with and kill HIV-infected autologous CD4 T cells. *Immunology*. **144**:412 – 421.

Cicala, C. et al. (2009) The integrin α4β7 forms a complex with cell-surface CD4 and defines a T-cell subset that is highly susceptible to infection by HIV-1. *PNAS.* **106**(49):20877 – 20882.

de Cock, K. et al. (2011) Reflections on 30 Years of AIDS. *Emerging Infectious Diseases*. **17**(6):1044 – 1048.

de Cock, K. et al. (2021) Reflections on 40 Years of AIDS. *Emerging Infectious Diseases.* **27**(6):1553 – 1560.

Cohen, M. et al. (2011) Prevention of HIV-1 Infection with Early Antiretroviral Therapy. *NEJM*. **365**(6):493 – 505.

Cook, P. and MacDonald, A. (2016) Dendritic cells in lung immunopathology. *Seminars in Immunopathology*. **38**(4):449 – 460.

Coiras, M. et al. (2010) Dual Role of Host Cell Factors in HIV-1 Replication: Restriction and Enhancement of the Viral Cycle. *AIDS Review*. **12**:103 – 112.

Collins, K. et al. (2000) Development of an *in vitro* organ culture model to study transmission of HIV-1 in the female genital tract. *Nature Medicine.* **6**(4):475 – 479.

Comeron, J. (1995) A Method for Estimating the Numbers of Synonymous and Nonsynonymous Substitutions per Site. *Journal of Molecular Evolution*. **41**:1152 – 1159.

Cone, R. (2014) Vaginal Microbiota and Sexual Transmitted Infections That May Influence Transmission of Cell-Associated HIV. *Journal of Infectious Diseases*. **210**:S616 – S621.

Curlin, M. et al. (2010) HIV-1 Envelope Subregion Length Variation during Disease Progression. *PLoS Pathogens*. **6**(12):e1001228.

Dalgleish, A. et al. (1984) The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*. **312**(5996):763 – 767.

Davies, O. et al. (2017) Impact of rectal gonorrhoea and chlamydia on HIV viral load in the rectum: potential significance for onward transmission. *International Journal of STD and AIDS*. **28**(10):1034 – 1037.

Day, C. et al. (2022) Complement Receptor 3 Mediates HIV-1 Transcytosis across an Intact Cervical Epithelial Cell Barrier: New Insight into HIV Transmission in Women. *Mbio*. **13**(1):e02177-21.

Deng, H. et al. (1996) Identification of a major co-receptor for primary isolates of HIV-1. *Nature*. **381**(3584):661 – 666.

Derdeyn, C. et al. (2004) Envelope-Constrained Neutralization-Sensitive HIV-1 After Heterosexual Transmission. *Science*. **303**(5666):2019 – 2022.

Devito, C. et al. (2004) Intranasal HIV-1-gp160-DNA/gp41 Peptide Prime-Boost Immunization Regimen in Mice Results in Long-Term HIV-1 Neutralizing Humoral Mucosal and Systemic Immunity. *Journal of Immunology*. **173**(11):7078 – 7089.

Deymier, M. et al. (2015) Heterosexual Transmission of Subtype C HIV-1 Selected Consensus-Like Variants without Increased Replicative Capacity or Interferon-α Resistance. *PLoS Pathogens*. **11**(9):e1005154.

Dinh, M. et al. (2015) Visualization of HIV-1 Interactions with Penile and Foreskin Epithelia: Clues for Female-to-Male HIV Transmission. *PLoS Pathogens*. **11**(3):e1004729.

Doepker, L. et al. (2021) Development of antibody-dependent cell cytotoxicity function in HIV-1 antibodies. *eLife*. **10**:e63444.

Dohgu, S. et al. (2011) Lipopolysaccharide-enhanced transcellular transport of HIV-1 across the blood-brain barrier is mediated by luminal microvessel IL-6 and GM-CSF. *Journal of Neuroinflammation*. **8**:167.

Dolin, R. et al. (1991) The safety and immunogenicity of a human immunodeficiency virus type 1 (HIV-1) recombinant gp160 candidate vaccine in humans. NIAID AIDS Vaccine Clinical Trials Network. *Annals of Internal Medicine*. **144**(5):119 – 127.

Dragic, T. et al. (1996) HIV-1 entry into $CD4^+$ cells is mediated by the chemokine receptor CC-CKR-5. *Nature*. **381**(6584):667 – 673.

D'Souza, G. et al. (2019) The Changing Science of HIV Epidemiology in the United States. *American Journal of Epidemiology*. **188**(12):2061 – 2068.

Dutartre, H. et al. (2016) Cell-Free versus Cell-to-Cell Infection by Human Immunodeficiency Virus Type 1 and Human T-lymphoctropic Virus Type 1: Exploring the Link among Viral Source, Viral Trafficking, and Viral Replication. *Journal of Virology*. **90**(17):7607 – 7617.

Earl, L. et al. (2013) Catching HIV 'in the act' with 3D electron microscopy. *Trends in Microbiology*. **21**(8):397 – 404.

Ebenbichler, C. et al. (1991) Human Immunodeficiency Virus Type 1 Activates the Classical Pathway of Complement by Direct C1 Binding through Specific Sites in the Transmembrane Glycoprotein gp41. *Journal of Experimental Medicine*. **174**:1417 – 1424.

Edo-Matas, D. et al. (2010) Genetic composition of replication competent clonal HIV-1 variants isolated from peripheral blood mononuclear cells (PBMC), HIV-1 proviral DNA from PBMC and HIV-1 RNA in serum in the course of HIV-1 infection. *Virology*. **405**:492 – 504.

Evans, E. et al. (2018) HIV-1 Vif's Capacity To Manipulate the Cell Cycle Is Species Specific. *Journal of Virology*. **92**(7):e02102-17.

Fenton-May, A. et al. (2013) Relative resistance of HIV-1 founder viruses to control by interferon-alpha. *Retrovirology*. **10**:146.

Fiebig, E. et al. (2003) Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS*.**17**:1871 – 1879.

Fischer, W. et al. (2010) Transmission of Single HIV-1 Genomes and Dynamics of Early Immune Escape Revealed by Ultra-Deep Sequencing. *PLoS One*. **5**(8):e12303.

Fischetti, L. et al. (2009) HIV-1 infection of human penile explant tissue and protection by candidate microbicides. *AIDS*. **23**(3):319 – 328.

Fitzgerald, D. et al. (2011) An Ad5-Vectored HIV-1 Vaccine Elicits Cell-mediated Immunity but does not Affect Disease Progression in HIV-1-infected Male Subjects: Results From a Randomized Placebo-Controlled Trial (The Step Study). *Journal of Infectious Diseases*. **203**:765 – 772.

Flaherty, M. et al. (1997) Molecular and Biological Characterization of a Neurovirulent Molecular Clone of Simian Immunodeficiency Virus. *Journal of Virology*. **71**(8):5790 – 5798.

Fong, L. et al. (2002) Productive Infection of Plasmacytoid Dendritic Cells with Human Immunodeficiency Virus Type 1 Is Triggered by CD40 Ligation. *Journal of Virology*. **76**(21):11033 – 11041.

Forthal, D. et al. (2012) Association of Fcγ reeptor IIIa genotype with the rate of HIV infection after gp120 vaccination. *Immunobiology*. **120**(14):2836 – 2842.

Franchini, G. et al. (1987) Sequence of simian immunodeficiency virus and its relationship to the human immunodeficiency viruses. *Nature*. **328**(6130):539 – 543.

Frange, P. et al. (2013) Sexually-Transmitted/Founder HIV-1 Cannot Be Directly Predicted from Plasma or PBMC-Derived Viral Quasispecies in the Transmitting Partner. *PLoS One*. **8**(7):e69144.

Fraser, I et al. (1999) Variations in vagina epithelial surface appearance determined by colposcopic inspection in healthy, sexually active women. *Human Reproduction*. **14**(8):1974 – 1979.

Frauwirth, K. et al. (2002) The CD28 Signaling Pathway Regulates Glucose Metabolism. *Immunity*. **16**:769 – 777.

Freed, E. and Martin, M. (1996) Domains of the Human Immunodeficiency Virus Type 1 Matrix and gp41 Cytoplasmic Tail Required for Envelope Incorporation into Virions. *Journal of Virology*. **70**(1):341 – 351.

Frost, S. et al. (2005) Characterization of Human Immunodeficiency Virus Type 1 (HIV-1) Envelope Variation and Neutralizing Antibody Responses during Transmission of HIV-1 Subtype B. *Journal of Virology*. **79**(10):6523 – 6527.

Gach, J. et al. (2017) Human immunodeficiency virus type-1 (HIV-1) evades antibody-dependent phagocytosis. *PLoS Pathogens*. **13**(12):e1006793.

Gach, J. et al. (2019) Antibody Responses Elicited by Immunization with BG505 Trimer Immune Complexes. *Journal of Virology*. **93**(20):e01188-19.

Gallay, P. et al. (2017) Prevention of vaginal and rectal HIV transmission by antiretroviral combinations in humanized mice. *PLoS One*. **12**(9):e0184303.

Glidden, D. et al. (2019) Mosaic effectiveness: measuring the impact of novel PrEP methods. *Lancet HIV*. **6**:e800-06.

Ganusov, V. et al. (2011) Fitness Costs and Diversity of the Cytotoxic T Lymphocyte (CTL) Response Determine the Rate of CTL Escape during Acute and Chronic Phases of HIV Infection. *Journal of Virology*. **85**(20):10518 – 10528.

Geijtenbeek, T. et al. (2000) DC-SIGN, a Dendritic Cell-Specific HIV-1-Binding Protein that Enhances *trans*-Infection of T Cells. *Cell*. **100**:587 – 597.

Gnanakaran, S. et al. (2011) Recurrent Signature Patterns in HIV-1 B Clade Envelope Glycoproteins Associated with either Early or Chronic Infections. *PLoS Pathogens*. **7**(9):e1002209.

Goff, S. (1992) Genetics of Retroviral Integration. *Annual Reviews of Genetics*. **26**:527 – 544.

Goldstone, D. et al. (2011) HIV-1 restriction factor SAMHD1 is a deoxynucleoside triophosphate triphosphohydrolase. *Nature*. **480**(7377):379 – 382.

Gondim, M. et al. (2021) Heightened resistance to host type 1 interferons characterizes HIV-1 at transmission and after antiretroviral therapy interruption. *Science Translational Medicine*. **13**:eabd8179.

Gonzalez, M. et al. (2015) Conserved Molecular Signatures in gp120 Are Associated with the Genetic Bottleneck during Simian Immunodeficiency Virus (SIV), SIV-Human Immunodeficiency Virus (SHIV), and HIV Type 1 (HIV-1) Transmission. *Journal of Virology*. **89**(7):3619 – 3629.

Gonzalez, O. and Sagar, M. (2016) Antibodies and Acidic Environment Do Not Enhance HIV-1 Transcytosis. *Journal of Infectious Diseases*. **214**:1221 – 1224.

Gonzalez-Perez, M. et al. (2012) Independent evolution of macrophage-tropism and increased charge between HIV-1 R5 envelopes present in brain and immune tissue. *Retrovirology*. **9**:20.

Gonzalez-Perez, M. et al. (2017) Identification of Emerging Macrophage-Trophic HIV-1 R5 Variants in Brain Tissue of AIDS Patients without Severe Neurological Complications. *Journal of Virology*. **91**(20):e00755-17.

Goo, L. et al. (2014) Early development of broadly neutralizing antibodies in HIV-1-infected infants. *Nature Medicine*. **20**(6):655 – 658.

Gorlani, A. and Forthal, D. (2013) Antibody-Dependent Enhancement and the Risk of HIV Infection. *Current HIV Research*. **11**(5):421 – 426.

Gorman, J. et al. (2022) Cryo-EM structures of prefusion SIV envelope trimer. *Nature Structural and Molecular Biology*. **29**(11):1080 – 1091.

Gramatica, A. et al. (2021) Evaluating a New Class of AKT/mTOR Activators for HIV Latency-Reversing Activity *Ex Vivo* and *In Vivo*. *Journal of Virology*. **95**(8):e02393-20.

Grant, R. et al. (2010) Preexposure Chemoprophylaxis for HIV Prevention in Men Who Have Sex with Men. *NEJM*. **363**(27):2587 – 2599.

van Grevenynghe, J. et al. (2008) Transcription factor FOXO3a controls the persistence of memory $CD4^+$ T cells during HIV infection. *Nature Medicine*. **14**(3):266 – 274.

Grütter, M. and Luban, J. (2012) TRIM5 structure, HIV-1 capsid recognition, and innate immune signaling. *Current Opinion in Virology*. **2**(2):142 – 150.

Guillon, C. et al. (2002) Antibody-Mediated Enhancement of Human Immunodeficiency Virus Type 1 Infectivity Is Determined by the Structure of gp120 and Depends on Modulation of the gp120-CCR5 Interaction. *Journal of Virology*. **76**(6):2827 – 2834.

Gupta, S. et al. (2013) The Neonatal Fc Receptor (FcRn) Enhances Human Immunodeficiency Virus Type 1 (HIV-1) Transcytosis across Epithelial Cells. *PLoS Pathogens*. **9**(11):e1003776.

Gupta, S. et al. (2023) Lenacapavir administered every 26 weeks or daily in combination with oral daily antiretroviral therapy for initial treatment of HIV: a randomised, open-label, active-controlled, phase 2 trial. *Lancet HIV.* **10**:e15 – e23.

Haaland, R. et al. (2009) Inflammatory Genital Infections Mitigate a Severe Genetic Bottleneck in Heterosexual Transmission of Subtype A and C HIV-1. *PLoS Pathogens*. **5**(1):e1000274.

Haaren, M. et al. (2017) Natural infection as a blueprint for rational HIV vaccine design. *Human Vaccines and Immunotherapeutics*. **13**(1):229 – 236.

Hamoudi, M. et al. (2013) Genetic diversity of the highly variable V1 region interferes with Human Immunodeficiency Virus type 1 envelope functionality. *Retrovirology*. **10**:114.

Hansen, S. et al. (2019) A live-attenuated RhCMV/SIV vaccine shows long-term efficacy against heterologous SIV challenge. *Science Translational Medicine*. **11**(501):eaaw2607.

Heeney, J. et al. (2006) Transmission of Simian Immunodeficiency Virus SIVcpz and the Evolution of Infection in the Presence and Absence of Concurrent Human Immunodeficiency Virus Type 1 Infection in Chimpanzees. *Journal of Virology*. **80**(14):7208 – 7218.

Hertoghs, N. et al. (2019) Sexually transmitted founder HIV-1 viruses are relatively resistant to Langerhans cell-mediated restriction. *PLoS One*. **14**(12):e0226651.

Heß, R. et al. (2019) Glycosylation of HIV Env Impacts IgG Subtype Responses to Vaccination. *Viruses*. **11**:153.

Hirsch, V. et al. (1987) The Genome Organization of STLV-3 Is Similar to That of the AIDS Virus except for a Truncated Transmembrane Protein. *Cell*. **49**(3):307 – 319.

HIV Sequence Compendium 2021. (2023) Cristian Apetrei, Beatrice Hahn, Andrew Rambaut, Steven Wolinsky, J. Rodney Brister, Brandon Keele, and Christophe Faser, editors. Publisher: Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico. LA-UR-23-22840.

HIV Sequence Compendium 2002. (2003) Kuiken C, Foley B, Freed E, Hahn B, Marx P, McCutchan F, Mellors J, Wolinsky S, and Korber B, editors. Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, LA-UR number 03-3564.

Hocini, H. and Bomsel, M. (1999) Infectious Human Immunodeficiency Virus Can Rapidly Penetrate a Tight Human Epithelial Barrier by Transcytosis in a Process Impaired by Mucosal Immunoglobulins. *Journal of Infectious Diseases*. **179**(Supp 3):S448 – S453.

Hocini, H. et al. (2001) Active and Selective Transcytosis of Cell-Free Human Immunodeficiency Virus through a Tight Polarized Monolayer of Human Endometrial Cells. *Journal of Virology*. **75**(11):5370 – 5374.

Hoffman, T. et al. (1998) HIV type I envelope determinants for use of the CCR2b, CCR3, STRL33, and APJ coreceptors. *PNAS*. **95**(19):11360 – 11365.

Hoffman, N. et al. (2002) Variability in the Human Immunodeficiency Virus Type 1 gp120 Env Protein Linked to Phenotype-Associated Changes in the V3 Loop. *Journal of Virology*. **76**(8):3852 – 3864.

Homsy, J. et al. (1989) The Fc and not CD4 receptor mediates antibody enhancement of HIV infection in human cells. *Science*. **244**(4910):1357 – 1360.

Hong, M. et al. (1998) IL-2 and IFN-γ, but not IL-4 secretion by peripheral blood mononuclear cells (PBMC) are related to $CD4^+$ T cells and clinical status in Brazilian HIV-1-infected subjects. *Revista do Instituto de Medicina Tropical de São Paulo*. **40**(6):351 – 354.

Hong, P. et al. (2002) Human Immunodeficiency Virus Envelope (gp120) Binding to DC-SIGN and Primary Dendritic Cells Is Carbohydrate Dependent but Does Not Involve 2G12 or Cyanovirin Binding Sites: Implications for Structural Analyses of gp120-DC-SIGN Binding. *Journal of Virology*. **76**(24):12855 – 12865.

Hu, J. et al. (2000) Simian Immunodeficiency Virus Rapidly Penetrates the Cervicovaginal Mucosa after Intravaginal Inoculation and Infects Intraepithelial Dendritic Cells. *Journal of Virology*. **74**(13):6087 – 6095.

Huang, C-C. et al. (2005) Structure of a V3-Containing HIV-1 gp120 Core. *Science*. **310**(5750):1025 – 1028.

Huang, L. and Crothers, K. (2009) HIV-associated Opportunistic Pneumonias. *Respirology*. **14**(4):474 – 485.

Huang, Y. et al. (2016a) Diversity of Antiviral IgG Effector Activities Observed in HIV-Infected and Vaccinated Subjects. *Journal of Immunology*. **197**:4603 – 4612.

Huang, X. et al. (2016b) Cytokine cascade and networks among MSM HIV seroconverters: implications for early immunotherapy. *Scientific Reports*. **6**:36234.

Huarte, N. et al. (2016) Functional organization of the HIV lipid envelope. *Scientific Reports*. **6**:34190.

Hussain, L. et al. (1995) Investigation of the complement receptor 3 (CD11b/CD18) in human rectal epithelium. *Clinical Experimental Immunology*. **102**:384 – 388.

Ibata, B. et al. (1997) Migration of Foreign Lymphocytes from the Mouse Vagina into the Cervicovaginal Mucosa and to the Iliac Lymph Nodes. *Biology of Reproduction*. **56**:537 – 543.

Ikeda, T. et al. (2019) HIV1 restriction by endogenous APOBEC3G in the myeloid cell line THP-1. *Journal of General Virology.* **100**(7):1140 – 1152.

The INSIGHT START Study Group (2015) Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection. *NEJM*. **373**(9):795 – 807.

Isaacman-Beck, J. et al. (2009) Heterosexual Transmission of Human Immunodeficiency Virus Type 1 Subtype C: Macrophage Tropism, Alternative Coreceptor Use, and the Molecular Anatomy of CCR5 Utilization. *Journal of Virology*. **83**(16):8208 – 8220.

Iyer, S. et al. (2017) Resistance to type 1 interferons is a major dterminant of HIV-1 transmission fitness. *PNAS*. **114**(4):E590 – E599.

Izquierdo-Useros, N. et al. (2014) HIV-1 Capture and Transmission by Dendritic Cells: The Role of Viral Glycolipids and the Cellular Receptor for Siglec-1. *PLoS Pathogens*. **10**(7):e1004146.

Jain, S. and Rosenthal, K. (2011) The gp41 epitope, QARVLAVERY, is highly conserved and a potent inducer of IgA that neutralizes HIV-1 and inhibits viral transcytosis. *Mucosal Immunology*. **4**(5):539 – 553.

Jarvis, C. et al. (2019) Antigen structure affects cellular routing through DC-SIGN. *PNAS*. **116**(30):14862 – 14867.

Jenabian, M-A. et al. (2010) Differential activity of candidate microbicides against early steps of HIV-1 infection upon complement virus opsonization. *AIDS Research and Therapy*. **7**:16.

Ji, Y. and Lu, H. (2017) "Chapter 10: Malignancies in HIV-Infected and AIDS Patients." *Infectious Agents Associated Cancers: Epdemiology and Molecular Biology*. Cai, Q. et al., editors. Published by Springer Nature Singapore, Pte, Ltd.

Julien, J-P. et al. (2015) Design and structure of two HIV-1 clade C SOSIP.664 trimers that increase the arsenal of native-like Env immunogens. *PNAS*. **112**(38):11947 – 11952.

June, R. et al. (1991) Complement and antibody mediate enhancement of HIV infection by increasing virus binding and provirus formation. *AIDS*. **5**:269 – 274.

Kafando, A. et al. (2017) HIV-1 envelope sequence-based diversity measures for identifying recent infections. *PLoS One*. **12**(12):e0189999.

Kafando, A. et al. (2019) HIV-1 Envelope Glycoprotein Amino Acids Signatures Associated with Clade B Transmitted/Founder and Recent Viruses. *Viruses*. **11**:1012.

Kane, M. et al. (2013) MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature*. **502**(7472):563 – 566.

Kariuki, S. et al. (2017) The HIV-1 transmission bottleneck. *Retrovirology*. **14**:22.

Karnasuta, C. et al. (2017) Comparison of Antibody Responses Induced by RV144, VAX003, and VAX004 Vaccination Regimens. *AIDS Research and Human Retroviruses*. **33**(5):410 – 423.

Kaul, R. et al. (2015) Inflammation and HIV Transmission in Sub-Saharan Africa. *Current HIV/AIDS Reports*. **12**:216 – 222.

Kaul, R. et al. (2022) The Penis, the Vagina and HIV Risk: Key Differences (Aside from the Obvious). *Viruses*. **14**:1164.

Kawamura, T. et al. (2008) Significant Virus Replication in Langerhans Cells following Application of HIV to Abraded Skin: Relevance to Occupational Transmission of HIV. *Journal of Immunology*. **180**(5):3297 – 3304.

Kazazi, F. et al. (1989) Variations in CD4 Expression by Human Monocytes and Macrophages and Their Relationship to Infection with the Human Immunodeficiency Virus. *Journal of General Virology*. **70**(10):2661 – 2672.

Keating, S. et al. (2016) Magnitude and Quality of Cytokine and Chemokine Storm during Acute Infection Distinguish Nonprogressive and Progressive Simian Immunodeficiency Virus Infections of Nonhuman Primates. *Journal of Virology*. **90**(22):10339 – 10350.

Keele, B. et al. (2008) Identification and Characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *PNAS*. **105**(21):7552 – 7557.

Keele, B. et al. (2009) Low-Dose Rectal Inoculation of *Rhesus macaques* by SIVsmE660 or SIVmac251 Recapitulates Human Mucosal Infection by HIV-1. *Journal of Experimental medicine*. **206**(5):1117 – 1134.

Keele, B. et al. (2010) Identifying and Characterizing Recently Transmitted Virus. *Current Opinion on HIV and AIDS.* **5**(4):327 – 334.

Kelley, C. et al. (2017) The Rectal Mucosa and Condomless Receptive Anal Intercourse in HIV Negative MSM: Implications for HIV Transmission and Prevention. *Mucosal Immunology*. **10**(4):996 – 1007.

Kerina, D. et al. (2011) HIV-1 Env gp120 C2V5 Potential N-Linked Glycosylation Site(s) (PNGs) Variations and Amino Acid Length Polymorphisms among Infected Family Members. *Advances in Infectious Diseases*. **1**(1):1 – 13.

Kijak, G. et al. (2013) Molecular Evolution of the HIV-1 Thai Epidemic between the Time of RV144 Immunogen Selection to the Execution of the Vaccine Efficacy Trial. *Journal of Virology*. **87**(13):7265 – 7281.

Kijak, G. et al. (2017) Rare HIV-1 transmitted/founder lineages identified by deep viral sequencing contribute to rapid shifts in dominant quasispecies during acute and early infection. *PLoS Pathogens*. **13**(7):e1006510.

Kinlock, B. et al. (2014) Transcytosis of HIV-1 through Vaginal Epithelial Cells Is Dependent on Trafficking to the Endocytic Recycling Pathway. *PLoS One*. **9**(5):e96760.

Klatzmann, D. et al. (1984) T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature*. **312**(5996):767 – 768.

Klein, K. et al. (2021) Deep Gene Sequence Cluster Analyses of Multi-Virus-Infected Mucosal Tissue Reveal Enhanced Transmission of Acute HIV-1. *Journal of Virology*. **95**(3):e01737-20.

Kogut, M. et al. (2020) Microbiome and pathogen interaction with the immune system. *Poultry Science*. **99**(4):1906 – 1913.

Koh, W. et al. (2020) HIV-Captured DCs Regulate T Cell Migration and Cell-Cell Contact Dynamics to Enhance Viral Spread. *iScience*. **23**:101427.

Kolodkin-Gal, D. et al. (2013) Efficiency of Cell-Free and Cell-Associated Virus in Mucosal Transmission of Human Immunodeficiency Virus Type 1 and Simian Immunodeficiency Virus. *Journal of Virology*. **87**(24):13589 – 13597.

Koup, R. et al. (2010) Priming Immunization with DNA Augments Immunogenicity of Recombinant Adenoviral Vectors for Both HIV-1 Specific Antibody and T-Cell Responses. *PLoS One*. **5**(2):e9015.

Koyanagi, Y. et al. (1987) Dual infection of the central nervous system by AIDS viruses with distinct cellular tropisms. *Science*. **236**(4803):819 – 822.

Kozlowski, P. et al. (1995) High prevalence serum IgA HIV-1 infection-enhancing antibodies in HIV-infected persons: Masking by IgG. *Journal of Immunology.* **154**(11):6163 – 6173.

Krapp, C. et al. (2016) Guanylate Binding Protein (GBP) 5 Is an Interferon-Inducible Inhibitor of HIV-1 Infectivity. *Cell Host and Microbe*.**19**(4):504 – 514.

Kulkarni, V. et al. (2013) Comparison of intradermal and intramuscular delivery followed by *in vivo* electroporation of SIV Env DNA in *macaques*. *Human Vaccines and Immunotherapeutics*. **9**(10):2081 – 2094.

Kwon, D. et al. (2002) DC-SIGN-Mediated Internalization of HIV Is Required for *Trans*-Enhancement of T Cell Infection. *Immunity*. **16**:135 – 144.

Laga, M. et al. (1993) Non-ulcerative sexually transmitted diseases as risk factors for HIV-1 transmission in women: results from a cohort study. *AIDS*. **7**(1):95 – 102.

Laguette, N. et al. (2011) SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature*. **474**(7353):654 – 657.

Lamers, S. et al. (2019) Emerging Patterns in HIV-1 gp120 Variable Domains in Anatomical Tissues in the Absence of a Plasma Viral Load. *AIDS Research and Human Retroviruses*. **35**(6):588 – 596.

Larsson, M. (2005) HIV-1 and the hijacking of dendritic cells: a tug of war. *Springer Seminars in Immunolopathology.* **26**(3):309 – 328.

Letvin, N. et al. (2007) No Evidence for Consistent Virus-Specific Immunity in Simian Immunodeficiency Virus-Exposed, Uninfected *Rhesus* Monkeys. *Journal of Virology*. **81**(22):12368 – 12374.

Li, W-H. et al. (1985) A New Method for Estimating Synonymous and Nonsynonymous Rates of Nucleotide Substitution Considering the Relative Likelihood of Nucleotide and Codon Changes. *Molecular Biology and Evolution*. **2**(2):150 – 174.

Li, W-H. (1993) Unbiased Estimation of the Rates of Synonymous and Nonsynonymous Substitution. *Journal of Molecular Evolution*. **36**:96 – 99.

Li, B. et al. (2006) Evidence for Potent Autologous Neutralizing Antibody Titers and Compact Envelopes in Early Infection with Subtype C Human Immunodeficiency Virus Type 1. *Journal of Virology*. **80**(11):5211 – 5218.

Li, Q. et al. (2009) Glycerol monolaureate prevents mucosal SIV transmission. *Nature*. **458**(7241):1034 – 1038.

Li, H. et al. (2010) High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men. *PLoS Pathogens*. **6**(5):e1000890.

Li, X. et al. (2011) $CD4^+$ T-cell counts and plasma HIV-1 RNA levels beyond 5 years of highly active antiretroviral therapy (HAART). *Journal of Acquired Immune Deficiency Syndromes*. **57**(5):421 – 428.

Liao, H-X. Et al. (2013) Antigenicity and Immunogenicity of Transmitted/Founder, Consensus, and Chronic Envelope Glycoproteins of Human Immunodeficiency Virus Type 1. *Journal of Virology*. **87**(8):4185 – 4201.

Lindbäck, S. et al. (2000) Diagnosis of primary HIV-1 infection and duration of follow-up after HIV exposure. *AIDS*. **14**:2333 – 2339.

Lindenbergh, M. and Stoorvogel, W. (2018) Antigen Presentation by Extracellular Vesicles from Professional Antigen-Presenting Cells. *Annual Review of Immunology*. **36**:435 – 459.

Liu, Y. et al. (2008) Env length and N-linked glycosylation following transmission of Human Immunodeficiency Virus Type 1 subtype B viruses. *Virology*. **374**(2):229 – 233.

Liu, J. et al. (2010) Low-Dose Mucosal Simian Immunodeficiency Virus Infection Restricts Early Replication Kinetics and Transmitted Virus Variants in *Rhesus* Monkeys. *Journal of Virology*. **84**(19):10406 – 10412.

Liu, M. et al. (2013) Vertical T cell immunodominances and epitope entropy determine HIV-1 escape. *Journal of Clinical Investigation.* **123**(1):380 – 393.

Liu, Y. et al. (2020) Broadly neutralizing antibodies for HIV-1: efficacies, challenges and opportunities. *Emerging Microbes and Infections*. **9**(1):194 – 206.

Locateli, D. et al. (2008) An Optimized Nested Polymerase Chain Reaction (PCR) Approach Allows Detection and Characterization of Human Immunodeficiency Virus Type 1 (HIV-1) env and gag Genes From Clinical Samples. *Journal of Clinical Laboratory Analysis*. **22**:106 – 113.

Low, N. et al. (2011) Intravaginal Practices, Bacterial Vaginosis, and HIV Infection in Women: Individual Participant Data Meta-analysis. *PLoS Medicine*. **8**(2):e1000416.

Lu, J. et al. (2011) The IFITM Proteins Inhibit HIV-1 Infection. *Journal Virology*. **85**(5):2126 – 2137.

Lu, C-L. et al. (2016) Enhanced clearance of HIV-1-infected cells by broadly neutralizing antibodies against HIV-1 *in vivo*. *Science*. **352**(6288):1001 – 1004.

Lucar, J. et al. (2018) Sexually Transmitted Infections Among HIV-Infected Individuals in the District of Columnia and Estimated HIV Transmission Risk: Data From the DC Cohort. *Open Forum Infectious Diseases*.**5**(2):ofy017.

Lumngwena, E. et al. (2019) HIV-1 subtype C Envelope function becomes less sensitive to N-glycosylation deletion during disease progression. *BMC Research Notes*. **12**:340.

Lumngwena, E. et al. (2020) Selective transmission of some HIV-1 subtype C variants might depend on Envelope stimulating dendritic cells to secrete IL-10. *PLoS One*. **15**(1):e0227533.

Ma, Z-M. et al. (2009) High Specific Infectivity of Plasma Virus from the Pre-Ramp-Up and Ramp-Up Stages of Acute Simian Immunodeficiency Virus Infection. *Journal of Virology*. **93**(7):3288 – 3297.

Ma, Z-M. et al. (2011) SIVmac251 Is Inefficiently Transmitted to *Rhesus macaques* by Penile Inoculation with a Single *SIVenv* Variant Found in Ramp-up Phase Plasma. *AIDS Research and Human Retroviruses*. **12**:1259 – 1269.

van Maarseveen, N. et al. (2006) A novel real-time PCR assay to determine relative replication capacity for HIV-1 protease variants and/or reverse transcriptase variants. *Journal of Virological Methods*. **133**:185 – 194.

Macher, A. (1988) The Pathology of AIDS. *Public Health reports*. **103**(3):246 – 254.

Macharia, G. et al. (2020) Infection with multiple HIV-1 founder variants is associated with lower viral replicative capacity, faster $CD4^+$ T cell decline and increased immune activation during acute infection. *PLoS Pathogens*. **16**(9):e1008853.

MacIver, N. (2013) Metabolic Regulation of T Lymphocytes. *Annual Review of Immunology*. **31**:259 – 283.

Madlala, P. et al. (2023) Genetic variation of the HIV-1 subtype C transmitted/founder viruses long terminal repeat elements and the impact on transcription activation potential and clinical disease outcomes. *PLoS Pathogens*. **19**(6):e1011194.

Malouli, D. et al. (2021) Cytomegaloviral Determinants of $CD8^+$ T Cell Programming and RhCMV/SIV Vaccine Efficacy. *Science Immunology*. **6**(57):eabg5413.

Man, K. and Kallies, A. (2015) Synchronizing transcriptional control of T cell metabolism and function. *Nature Reviews Immunology*. **15**:574 – 584.

Manches, O. et al. (2014) Dendritic cells in progression and pathology of HIV infection. *Trends in Immunology*. **35**(3):114 – 122.

Martellini, J. et al. (2009) Cationic polypeptides contribute to the anti-HIV-1 activity of human seminal plasma. *The FASEB Journal*. **23**:3609 – 3618.

Masese, L. et al. (2015) Changes in the Contribution of Genital Tract Infections to HIV acquisition among Kenyan High-Risk Women from 1993 to 2012. *AIDS*. **29**(9):1077 – 1085.

Matoba, N. et al. (2004) A mucosally targeted subunit vaccine candidate eliciting HIV-1 transcytosis-blocking Abs. *PNAS*. **101**(37):13584 – 13589.

Matume, N. et al. (2020) HIV-1 subtype C predicted co-receptor tropism in Africa: an individual sequence level meta-analysis. *AIDS Research and Therapy*. **17**:5.

McCormack, S. et al. (2016) Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. *Lancet*. **387**:53 – 60.

McDonald, J. and Kreitman, M. (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*. **351**(6328):652 – 654.

Meiklejohn, C. et al. (2007) Positive and negative selection on the mitochondrial genome. *Trends in Genetics*. **23**(6):259 – 263.

Menk, A. et al. (2018) Early TCR Signaling Induces Rapid Aerobic Glycolysis Enabling Distinct Acute T Cell Effector Functions. *Cell Reports*. **22**(6):1509 – 1521.

Mercer, J. and Helenius, A. (2009) Virus entry by macropinocytosis. *Nature Cell Biology*. **11**(5):510 – 520.

Mild, M. et al. (2010) Differences in molecular evolution between switch (R5 to R5X4/X4-tropic) and non-switch (R5-tropic only) HIV-1 populations during infection. *Infection, Genetics and Evolution.* **10**:356 – 364.

Miller, C. and Hu, J. (1999) T Cell-Tropic Simian Immunodeficiency Virus (SIV) and Simian-Human Immunodeficiency Viruses Are Readily Transmitted by Vaginal Inoculation of *Rhesus macaques*, and Langerhans' Cells of the Female Genetial Tract Are Infected with SIV. *Journal of Infectious Diseases*. **179**(Supp 3):S413 – S417.

Molina, J. et al. (2015) On-Demand Preexposure Prophylaxis in Men at High Risk for HIV-1 Infection. *NEJM*. **373**(23):2237 – 2246.

Monaco, D. et al. (2017) Virus-Host Gene Interactions Define HIV-1 Disease Progression. *Current Topics in Microbiology and Immunology*. **407**:31 – 63.

Montefiori, D. et al. (1989) Antibody-independent, complement-mediated enhancement of HIV-1 infection by mannosidase I and II inhibitors. *Antiviral Research*. **11**:137 – 146.

van Montfort, T. et al. (2008) Dendritic Cells Prefrentially Transfer CXCR5-Using Human ImmunodeficiencyVirus Type 1 Variants to CD4+ T Lymphocytes in *trans*. *Journal of Virology*. **82**(16):7886 – 7896.

Mori, T. and Boyd, M. (2001) Cyanovirin-N, a Potent Human Immunodeficiency Virus-Inactivating Protein, Blocks both CD4-Dependent and CD4-Independent Binding of Soluble gp120 (sgp120) to Target Cells, Inhibits sCD4-Induced Binding of sgp120 to Cell-Associated CXCR4, and Dissociates Bound sgp120 from Target Cells. *Antimicrobial Agents and Chemotherapy*. **45**(3):664 – 672.

Mosa, A. (2021) Implications of viral transmitted/founder (T/F) dynamics on vaccine development. *Human Vaccines and Immunotherapeutics*. **17**(7):2293 – 2297.

Mueller, A. and Strange, P. (2003) The chemokine receptor, CCR5. *International Journal of Biochemistry and Cell Biology*. **36**:35 – 38.

Muema, D. et al. (2020) Association between the cytokine storm, immune cell dynamics, and viral replicative capacity in hyperacute HIV infection. *BMC Medicine*. **18**:81.

Muse, S. and Gaut, B. (1994) A Likelihood Approach for Comparing Synonymous and Nonsynonymous Nucleotide Substitution Rates, with Application to the Chloroplast Genome. *Molecular Biology and Evolution*. **11**(5):715 – 724.

Nabatov, A. et al. (2006) Interaction of HIV-1 with dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin-expressing cells is influenced by gp120 envelope modifications associated with disease progression. *The FEBS Journal*. **273**:4944 – 4958.

Naiman, N. et al. (2019) Antibody-dependent cellular cytotoxicity targeting CD4-inducible epitopes predicts mortality in HIV-infected infants. *EBioMedicine*. **47**:257 – 268.

Nascimento, C. et al. (2022) Comparison of monocytic cell lines U937 and THP-1 as macrophage models for *in vitro* studies. *Biochemistry and Biophysics Reports*. **32**:101383.

Nasi, M. et al. (2014) Persistent inflammation in HIV infection: Established concepts, new perspectives. *Immunology Letters*. **161**:184 – 188.

Nei, M. and Gojobori, T. (1986) Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. *Molecular Biology and Evolution*. **3**(5):418 – 426.

Neil, S. et al. (2008) Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature*. **451**(7177):425 – 430.

Ng'uni, T. et al. (2020) Major Scientific Hurdles in HIV Vaccine Development: Historical Perspective and Future Directions. *Frontiers in Immunology*. **11**:590780.

Nijmeijer, B. and Geijtenbeek, T. (2019) Negative and Positive Selection Pressure During Sexual Transmission of Transmitted Founder HIV-1. *Frontiers in Immunology*. **10**:1599.

Nijmeijer, B. et al. (2021) HIV-1 subverts the complement system in semen to enhance viral transmission. *Mucosal Immunology.* **14**:743 – 750.

Noel, G. et al. (2017) A primary human macrophage-enteroid co-culture model to investigate mucosal gut physiology and host-pathogen interactions. *Scientific Reports*. **7**:45270.

Nunes, R. et al. (2014) Formulation and delivery of anti-HIV rectal microbicides: Advances and challenges. *Journal of Controlled Release.* **194**:278 – 294.

Ochsenbauer, C. et al. (2012) Generation of Transmitted/Founder HIV-1 Infectious Molecular Clones and Characterization of Their Replication Capacity in CD4 T Lymphocytes and Monocyte-Derived Macrophages. *Journal of Virology*. **86**(5):2715 – 2728.

O'Driscoll, D. et al. (1991) Lucifer Yellow fluorescein isothiocyanate uptake by cells of *Morinda citrifolia* in suspension cultures is not confined to the endocytotic pathway. *Journal of Cell Science*. **100**:237 – 241.

Palmer, C. et al. (2016) Glucose Metabolism in T Cells and Monocytes: New Perspectives in HIV Pathogensis. *EBioMedicine*. **6**:31 – 41.

Palmer, J. and Poon, A. (2019) Phylogenetic measures of indel rate variation among the HIV-1 group M subtypes. *Virus Evolution*. **5**(2):vez022.

Pamilo, P. and Bianchi, N. (1993) Evolution of the *Zfx* and *Zfy* Genes: Rates and Interdependence between the Genes. *Molecular Biology and Evolution*. **10**(2):271 – 281.

Park, S. et al. (2016) Molecular clock of HIV-1 envelope genes under early immune selection. *Retrovirology*. **13**:38.

Parker, Z. et al. (2013) Transmitted/Founder and Chronic HIV-1 Envelope Proteins Are Distinguished by Differential Utilization of CCR5. *Journal of Virology*. **87**(5):2401 – 2411.

Parrish, N. et al. (2012) Transmitted/Founder and Chronic Subtype C HIV-1 Use CD4 and CCR5 Receptors with Equal Efficiency and Are Not Inhibited by Blocking the Integrin α4β7. *PLoS Pathogens*. **8**(5):e1002686.

Parrish, N. et al. (2013) Phenotypic properties of transmitted founder HIV-1. *PNAS*. **110**(17):6626 – 6633.

Passmore, J-A. et al. (2016) Genital inflammation, immune activation and risk of sexual HIV acquisition. *Current Opinion on HIV and AIDS*. **11**(2):156 – 162.

Patel, P. et al. (2014) Estimating per-act HIV transmission risk: a systematic review. *AIDS*. **28**:1509 – 1519.

Patterson, B. et al. (2002) Susceptibility to Human Immunodeficiency Virus-1 Infection of Human Foreskin and Cervical Tissue Grown in Explant Culture. *American Journal of Pathology*. **161**(3):867 – 873.

Pau, A. and George, J. (2014) Antiretroviral Therapy: Current Drugs. *Infectious Disease Clinics of North America*. **28**(3):371 – 402.

Pederson, J. et al. (2021) Cell-Extrinsic Priming Increases Permisiveness of $CD4^+$ T Cells to Human Immunodeficiency Virus Infection by Increasing C-C Chemokine Receptor Type 5 Co-receptor Expression and Cellular Activation Status. *Frontiers in Microbiology*. **12**:763030.

Perez-Zsolt, D. et al. (2019) Dendritic Cells From the Cervical Mucosa Capture and Transfer HIV-1 via Siglec-1. *Frontiers in Immunology*. **10**:825.

Piantadosi, A. et al. (2019) Metagenomic Sequencing of HIV-1 in the Blood and Female Genital Tract Reveals Little Quasispecies Diversity during Acute Infection. *Journal of Virology*. **93**(2):e00804-18.

Pierini, V. et al. (2021) SERINC5 Can Enhance Proinflammatory Cytokine Production by Primary Human Myeloid Cells in Response to Challenge with HIV-1 Particles. *Journal of Virology*. **95**(9):e02372-20.

Ping, L-H. et al. (2013) Comparison of Viral Env Proteins from Acute and Chronic Infections with Subtype C Human Immunodeficiency Virus Type 1 Identifies Differences in Glycosylation and CCR5 Utilization and Suggests a New Strategy for Immunogen Design. *Journal of Virology*. **87**(13):7218 – 7233.

Pitisuttithum, P. et al. (2006) Randomized, Double-Blind, Placebo-Controlled Efficacy Trial of a Bivalent Recombinant Glycoprotein 120 HIV-1 Vaccine among Injection Drug Users in Bangkok, Thailand. *Journal of Infectious Diseases*. **194**:1661 – 1671.

Pitisuttithum, P. and Marovich, M. (2020) Prophylactic HIV vaccine: vaccine regimens in clinical trials and potential challenges. *Expert Review of Vaccines*. **19**(2):133 – 142.

Planelles, V. and Barker, E. (2010) Roles of Vpr and Vpx in modulating the virus-host cell relationship. *Molecular Aspects of Medicine*. **31**(5):398 – 406.

Plata, F. et al. (1987) AIDS virus-specific cytotoxic T lymphocytes in lung disorders. *Nature*. **328**(6128):348 – 351.

Pond, S. et al. (2018) HIV-TRACE (TRAnsmission Cluster Engine): a Tool for Large Scale Molecular Epidemiology of HIV-1 and Other Rapidly Evolving Pathogens. *Molecular Biology and Evolution*. **35**(7):1812 – 1819.

Poon, A. et al. (2012) Reconstructing the Dynamics of HIV Evolution within Hosts from Serial Deep Sequence Data. *PLoS Computational Biology*. **8**(11):e1002753.

Popovic, M. et al. (1984) Detection, Isolation, and Continuous Production of Cytopathic Retroviruses (HTLV-III) from Patients with AIDS and Pre-AIDS. *Science*. **224**(4648):497 – 500.

Posch, W. et al. (2020) Role of Complement Receptors (Crs) on DCs in Anti-HIV-1 Immunity. *Frontiers in Immunology*. **11**:572114.

del Prete, G. (2013) Comparative Characterization of Transfection- and Infection-Derived Simian Immunodeficiency Virus Challenge Stocks for *In Vivo* Nonhuman Primate Studies. *Journal of Virology*. **87**(8):4584 – 4595.

Prodger, J. and Kaul, R. (2017) The biology of how circumcision reduces HIV susceptibility: broader implications for the prevention field. *AIDS Research and Therapy*. **14**:49.

Pudney, J. and Song, M. (1994) Electron Microscopic Analysis of HIV-Host Cell Interactions. *Tissue and Cell*. **26**(4):539 – 550.

Pudney, J. et al. (2019) Condylomata Acuminata (Anogenital Warts) Contain Accumulations of HIV-1 Target Cells That May Provide Portals for HIV Transmission. *Journal of Infectious Diseases*. **219**:275 – 283.

Qin, S. et al. (2005) The Chemokine Receptors CXCR3 and CCR5 Mark Subsets of T Cells Associated with Certain Inflammatory Reactions. *Journal of Clinical Investigations.* **101**(4):746 – 754.

Qu, K. et al. (2021) Maturation of the matrix and viral membrane of HIV-1. *Science*. **373**(6555):700 – 704.

Qureshi, H. et al. (2012) Low-Dose Penile SIVmac251 Exposure of *Rhesus macaques* Infected with Adenovirus Type 5 (Ad5) and Then Immunized with a Replication-Defective Ad5-Based SIV *gag/pol/nef* Vaccine Recapitulates the Results of the Phase Iib Step Trial of a Similar HIV-1 Vaccine. *Journal of Virology*. **86**(4):2239 – 2250.

Rademeyer, C. et al. (2016) Features of Recently Transmitted HIV-1 Clade C Viruses that Impact Antibody Recognition: Implications for Active and Passive Immunization. *PLoS Pathogens*. **12**(7):e1005742.

Ramdas, P. et al. (2020) From Entry to Egress: Strategic Exploitation of the Cellular Processes by HIV-1. *Frontiers in Microbiology*. **11**:559792.

Ranganath, N. et al. (2016) Type I interferon respones are impaired in latently HIV infected cells. *Retrovirology*. **13**:66.

Rao, M. et al. (2013) HIV-1 Variable Loop 2 and its Importance in HIV-1 Infection and Vaccine Development. *Current HIV Research*. **11**(5):427 – 438.

Rappocciolo, G. et al. (2019) Efficient HIV-1 *Trans* Infection of $CD4^+$ T Cells Occurs in the Presence of Antiretroviral Therapy. *Open Forum Infectious Diseases*. **6**(7):ofz253.

Ratner, L. et al. (1985) Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature*. **313**(6000):277 – 284.

Regenmortel, M. (2018) Development of a Preventive HIV Vaccine Requires Solving Inverse Problems Which Is Unattainable by Rational Vaccine Design. *Frontiers in Immunology*. **8**:2009.

Rerks-Ngarm, S. et al. (2009) Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand. *NEJM*. **361**(23):2209 – 2220.

The rgp120 HIV Vaccine Study Group. (2005) Placebo-Controlled Phase 3 Trial of a Recombinant Glycoprotein 120 Vaccine to Prevent HIV-1 Infection. *Journal of Infectious Diseases*. **191**:654 – 665.

Rihn, S. et al. (2017) The Envelope Gene of Transmitted HIV-1 Resists a late Interferon Gamma-Induced Block. *Journal of Virology*. **91**(7):e02254-16.

Robb, M. et al. (2016) Prospective Study of Acute HIV-1 Infection in Adults in East Africa and Thailand. *NEJM*. **374**(22):2120 – 2130.

Roberts, L. et al. (2010) Plasma cytokine levels during acute HIV-1 infection predict HIV disease progression. *AIDS*. **24**(6):819 – 831.

Robinson, W. et al. (1987) A Human Immunodeficiency Virus Type 1 (HIV-1) Infection Enhancing Factor in Seropositive Sera. *Biochemical and Biophysical Research Communications*. **149(**2):693 – 699.

Robinson, W. et al. (1989) Complement-mediated, antibody-dependent enhancement of HIV-1 infection *in vitro* is characterized by increased protein and RNA syntheses and infectious virus release. *Journal of Acquired Immune Deficiency Syndromes*. **2**(1):33 – 42.

Robinson, W. et al. (1990) Complement-mediated antibody-dependent enhancement of HIV-1 infection requires CD4 and complement receptors. *Virology*. **175**(2):600 – 604.

Rolle, C-P. et al. (2019) PrEP Implementation and Persistence in a County Health Department Setting in Atlanta, GA. *AIDS and Behavior*. **23**:S296 – S303.

Rose, P. and Korber, B. (2000) Detecting hypermutations in viral sequences with an emphasis on G→A hypermutation. *Bioinformatics*. **16**(4):400 – 401.

Rossi, F. et al. (2008) The V1-V3 region of a brain-derived HIV-1 envelope glycoprotein determines macrophage tropism, low CD4 dependence, increased fusogenicity and altered sensitivity to entry inhibitors. *Retrovirology*. **5**:89.

Rothenberg, R. et al. (1998) Oral transmission of HIV. *AIDS*. **12**:2095 – 2105.

des Roziers, N. et al. (1995) Kinetics of detection of antibodies to HIV-1 and plasma p24 antigens during a severe primary HIV-1 infection. *AIDS*. **9**(5):528 – 529.

Rueckert, C. and Guzmán, C. (2012) Vaccines: From Empirical Development to Rational Design. *PLoS Pathogens*. **8**(11):e1003001.

Saïdi, H. et al. (2007) Differential *in vitro* inhibitory activity against HIV-1 of alpha-(1-3)- and alpha-(1-6)-D-mannose specific plant lectins: Implication for microbicide development. *Journal of Translational Medicine*. **5**:28.

Saïdi, H. et al. (2012) Understanding Factors That Modulate HIV Infection at the Female Genital Tract Mucosae for the Rationale Design of Microbicides. *AIDS Research and Human Retroviruses*. **28**(11):1485 – 1497.

Salazar-Gonzalez, J. et al. (2008) Deciphering Human Immunodeficiency Virus Type 1 Transmission and Early Envelope Diversification by Single-Genome Amplification and Sequencing. *Journal of Virology*. **82**(8):3952 – 3970.

Salazar-Gonzalez, J. et al. (2009) Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *Journal of Experimental Medicine*. **206**(6):1273 – 1289.

Salgado, M. et al. (2014) HLA-B*57 Elite Suppressor and Chronic Progressor HIV-1 Isolates Replicate Vigorously and Cause $CD4^+$ T Cell Depletion in Humanized BLT Mice. *Journal of Virology*. **88**(6):3340 – 3352.

Samji, H. et al. (2013) Closing the Gap: Increases in Life Expectancy among Treated HIV-Positive Individuals in the United States and Canada. *PLoS One*. **8**(12):e81355.

Sanders, R. et al. (2015) HIV-1 neutralizing antibodies induced by native-like envelope trimers. *Science*. **349**(6244):aac4223.

Sanders, R. et al. (2013) A Next-Generation Cleaved, Soluble HIV-1 Env Trimer, BG505 SOSIP.664 gp140, Expresses Multiple Epitopes for Broadly Neutralizing but Not Non-Neutralizing Antibodies. *PLoS Pathogens*. **9**(9):e1003618.

Sarma, J. and Ward, P. (2011) The complement system. *Cell Tissue Research*. **343**:227 – 235.

Sarzotti-Kelsoe, M. et al. (2014) Optimization and Validation of a Neutralizing Antibody Assay for HIV-1 in A3R5 Cells. *Journal of Immunological Methods.* **409**:147 – 160.

Sato, K. et al. (2010) Remarkable Lethal G-to-A Mutations in *vif*-Proficient HIV-1 Provirus by Individual APOBEC3 Proteins in Humanized Mice. *Journal of Virology*. **84**(18):9546 – 9556.

Schmidt, B. et al. (2005) HIV-infected cells are major inducers of plasmacytoid dendritic cell interferon production, maturation, and migration. *Virology*. **343**:256 – 266.

Schneider, W. et al. (2014) Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annual Review of Immunology*. **32**:513 – 545.

Schochetman, G. (1992) Diagnosis of HIV infection. *Clinica Chimica Acta*. **211**:1 – 26.

Schoggins, J. (2019) Interferon-Stimulated Genes: What Do They All Do? *Annual Review of Virology*. **6**:567 – 584.

Schorcht, A. et al. (2020) Neutralizing Antibody Reponses Induced by HIV-1 Envelope Glycoprotein SOSIP Trimers Derived from Elite Neutralizers. *Journal of Virology*. **94**(24):e01214-20.

Schwartz, J. et al. (2018) Characterization of the Plasmacytoid Dendritic Cell Response to Transmitted/Founder and Nontransmitted variants of HIV-1. *Journal of Virology*. **92**(19):e00157-18.

Seclén, E. et al. (2011) Short Communication: Severe Immune Suppression in Patients Infected with R5-Tropic HIV-1 Strains Is Associated with Increased gp120 Net Charge at Variable Regions. *AIDS Research and Human Retroviruses*. **27**(9):965 – 967.

Seitz, R. (2016) Human Immunodeficiency Virus (HIV). *Transfusion Medicine and Hemotherapy*. **43**:203 – 222.

Sewald, X. et al. (2015) Retroviruses use CD169-mediated trans-infection of permissive lymphocytes to establish infection. *Science*. **350**(6260):563 – 567.

Shan, M. et al. (2007) HIV-1 gp120 Mannoses Induce Immunosuppressive Responses from Dendritic Cells. *PLoS Pathogens*. **3**(11):e169.

Sharma, D. et al. (1992) Derivation of Neurotropic Simian Immunodeficiency Virus from Exclusively Lymphocytetropic Parental Virus: Pathogenesis of Infection in *Macaques*. *Journal of Virology*. **66**(6):3550 – 3556.

Sharma, M. et al. (2019) Moving from Empirical to Rational Vaccine Design in the 'Omics' Era. *Vaccines*. **7**:89.

Sharp, P. and Hahn, B. (2011) Origins of HIV and the AIDS Pandemic. *Cold Spring Harbor Perspectives in Medicine*. **1**:a006841.

Shen, R. et al. (2009) Macrophages in Vaginal but Not Intestinal Mucosa Are Monocyte-Like and Permissive to Human Immunodeficiency Virus Type 1 Infection. *Journal of Virology*. **83**(7):3258 – 3267.

Shen, R. et al. (2010a) Dendritic cells transmit HIV-1 through human small intestinal mucosa. *Journal of Leukocyte Biology*. **87**:663 – 670.

Shen, R. et al. (2010b) GP41-Specific Antibody Blocks Cell-Free HIV Type 1 Transcytosis through Human Rectal Mucosa and Model Colonic Epithelium. *Journal of Immunology*. **184**:3648 – 3655.

Shen, R. et al. (2010c) Early HIV-1 Target Cells in Human Vaginal Ectocervical Mucosa. *American Journal of Reproductive Immunology*. **65**:261 – 267.

Shen, R. et al. (2014) HIV-1 Envelope Glycan Moieties Modulate HIV-1 Transmission. *Journal of Virology*. **88**(24):14258 – 14267.

Shepard, R. et al. (2000) Quantitation of Human Immunodeficiency Virus Type 1 RNA in Different Biological Compartments. *Journal of Clinical Microbiology*. **38**(4):1414 – 1418.

da Silva, R. et al. (2017) Evidence of bottleneck effect on hepatitis C virus transmission between a couple under interferon based therapy. *Infection, Genetics and Evolution*. **47**:87 – 93.

Simes, R. (1986) An improved Bonferroni procedure for multiple tests of significance. *Biometrika*. **73**(3):751 – 754.

Smalls-Mantey, A. et al. (2012) Antibody-Dependent Cellular Cytotoxicity against Primary HIV-Infected $CD4^+$ T Cells Is Directly Associated with the Magnitude of Surface IgG Binding. *Journal of Virology*. **86**(16):8672 – 8680.

Song, H. et al. (2016) Transmission of Multiple HIV-1 Subtype C Transmitted/founder Viruses into the Same Recipients Was not Determined by Modest Phenotypic Differences. *Scientific Reports*. **6**:38130.

Soper, A. et al. (2018) Type I Interferon Responses by HIV-1 Infection: Association with Disease Progression and Control. *Frontiers in Immunology*. **8**:1823.

Spire, B. et al. (1989) Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immunodeficiency virus. *Gene*. **81**:275 – 284.

Srinivasan, B. et al. (2015) TEER Measurement Techniques for *In Vitro* Barrier Model Systems. *Journal of Laboratory Automation*. **20**(2):107 – 126.

Stacey, A. et al. (2009) Induction of a Striking Systemic Cytokine Cascade prior to Peak Viremia in Acute Human Immunodeficiency Virus Type 1 Infection, in Contrast to More Modest and Delayed Responses in Acute Hepatitis B and C Virus Infections. *Journal of Virology*. **83**(8):3719 – 3733.

Stagg, A. (2018) Intestinal Dendritic Cells in Health and Gut Inflammation. *Frontiers in Immunology*. **9**:2883.

Stax, M. et al. (2009) Mucin 6 in seminal plasma binds DC-SIGN and potently blocks dendritic cell mediated transfer of HIV-1 to $CD4^+$ T-lymphocytes. *Virology*. **391**:203 – 211.

Stieh, D. et al. (2016) Th17 cells are preferentially infected very early after vaginal transmission of SIV in *macaques*. *Cell Host and Microbe*. **19**(4):529 – 540.

Stoletzki, N. and Eyre-Walker, A. (2011) Estimation of the Neutrality Index. *Molecular Biology and Evolution*. **28**(1):63 – 70.

Stone, M. et al. (2010) A Limited Number of Simian Immunodeficiency Virus (SIV) *env* Variants Are Transmitted to *Rhesus macaques* Vaginally Inoculated with SIVmac251. *Journal of Virology*. **84**(14):7083 – 7095.

Storim, J. et al. (2019) Antiretroviral therapy suppresses rectal HIV-RNA shedding despite inflammation in MSM with rectal *C. trachomatis* and *N. gonnorrhoeae* infections—a cross-sectional, single-center study. *Sexually Transmitted Infections*. **95**:95 – 98.

Straehley, I. et al. (2020) Horner Syndrome induced by toxoplasmosis infection in a patient with AIDS and disseminated herpes simplex virus. *American Journal of Ophthalmology Case Reports*. **18**:100679.

Stremlau, M. et al. (2004) The cytoplasmic body component TRIM5α restricts HIV-1 infection in Old World Monkeys. *Nature*. **427**(6977):848 – 853.

Sturdevant, C. et al. (2015) Compartmentalized Replication of R5 T Cell-Tropic HIV-1 in the Central Nervous System Early in the Course of Infection. *PLoS Pathogens*. **11**(3):e1004720.

Sugrue, E. et al. (2022) The apparent interferon resistance of transmitted HIV-1 is possibly a consequence of enhanced replicative fitness. *PLoS Pathogens*. **8**(11):e1010973.

Sun, Z. et al. (2022) Increasing awareness of HIV pre-exposure prophylaxis (PrEP) and willingness to use HIV PrEP among men who have sex with men: a systematic review and meta-analysis of global data. *Journal of the International AIDS Society*. **25**:e25883.

Suntharasamai, P. et al. (2008) Factors associated with incarceration and incident human immunodeficiency virus (HIV) infection among injection drug users participating in an HIV vaccine trial in Bangkok, Thailand, 1999 – 2003. *Addiction*. **104**:235 – 242.

Sutthent, R. et al. 2001) Diversity of HIV-1 subtype E in semen and cervicovaginal secretion. *Journal of Human Virology*. **4**(5):260 – 268.

Swindells, S. (2020) Long-Acting Cabotegravir and Rilpivirine for Maintenance of HIV-1 Suppression. *NEJM*. **382**(12):1112 – 1123.

Taborda, N. et al. (2015) Higher Frequency of NK and $CD4^+$ T-Cells in Mucosa and Potent Cytotoxic Response in HIV Controllers. *PLoS One*. **10**(8):e0136292.

Takeda, A. et al. (1988) Antibody-enhanced infection by HIV-1 via Fc receptor-mediated entry. *Science*. **242**(4878):580 – 583.

Takeda, A. and Ennis, F. (1990) FcR-Mediated Enhancement of HIV-1 Infection by Antibody. *AIDS Research and Human Retroviruses*. **6**(8):999 – 1004.

Tamamis, P. and Floudas, C. (2013) Molecular Recognition of CXCR4 by a Dual Tropic HIV-1 gp120 V3 Loop. *Biophysical Journal*. **105**:1502 – 1514.

Tamamis, P. and Floudas, C. (2014) Molecular Recognition of CCR5 by an HIV-1 gp120 V3 Loop. *PLoS One*. **9**(4):e95767.

Tezuka, H. and Ohteki, T. (2018) Regulation of IgA Production by Intestinal Dendritic Cells and Related Cells. *Frontiers in Immunology*. **10**:1891.

Thieblemont, N. et al. (1993a) CR1 (CD35) and CR3 (CD11b/CD18) mediate infection of human monocytes and monocytic cell lines with complement-opsonized HIV independently of CD4. *Clinical and Experimental Immunology*. **92**(1):106 – 113.

Thieblemont, N. et al. (1993b) Complement enhancement of HIV infection is mediated by complement receptors. *Immunopharmacology*. **25**:87 – 93.

Tomescu, C. et al. (2017) IFN-α augments natural killer-mediated antibody-dependent cellular cytotoxicity of HIV-1-infected autologous $CD4^+$ T cells regardless of major histocompatibility complex class 1 downregulation. *AIDS*. **31**:613 – 622.

Tortorec, A. et al. (2008) Infection of Semen-Producing Organs by SIV during the Acute and Chronic Stages of the Disease. *PLoS One*. **3**(3):e1792.

Trautz et al. (2016) The Antagonism of HIV-1 Nef to SERINC5 Particle Infectivity Restriction Involves the Counteraction of Virion-Associated Pools of the Restriction Factor. *Journal of Virology*. **90**(23):10915 – 10927.

Tremblay, M. et al. (1990) Complement Receptor 2 Mediates Enhancement of Human Immunodeficiency Virus 1 Infection in Epstein-Barr Virus-carrying B Cells. *Journal of Experimental Medicine*. **171**:1791 – 1796.

Trifonova, R. et al. (2018) Myeloid Cells in Intact Human Cervical Explants Capture HIV and Can Transmit It to CD4 T Cells. *Frontiers in Immunology*. **9**:2719.

Trischmann, H et al. (1995) Lymphocytotropic Strains of HIV Type 1 When Complexed with Enhancing Antibodies Can Infect Macrophages via FcγRIII, Independently of CD4. *AIDS Research and Human Retroviruses*. **11**(3):343 – 352.

Trkola, A. et al. (1996) CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature*. **384**(6605):184 – 187.

Tsai, L. et al. (2014) The number and genetic relatedness of transmitted/founder virus impact clinical outcome in vaginal R5 $SHIV_{SF162P3N}$ infection. *Retrovirology*. **11**:22.

Tully, D. et al. (2016) Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus. *PLoS Pathogens*. **12**(5):e1005619.

Tumba, N. et al. (2022) Immunization with HIV-1 trimeric SOSIP.664 BG505 or founder virus C (FVC_{ENV}) covalently complexed to two-domain $CD4^{S60C}$ elicits cross-clade neutralizing antibodies in New Zealand white rabbits. *Vaccine: X*. **12**:100222.

Turner, B. and Summers, M. (1999) Structural Biology of HIV. *Journal of Molecular Biology*. **285**:jmbi.1998.2354.

Tunthanathip, P. et al. (2009) Indicators for sexual HIV transmission risk among people in Thailand attending HIV care: the importance of positive prevention. *Sexually Transmitted Infections*. **85**(1):36 – 41.

Unger, R. et al. (1992) The *nef* Gene of Simian Immunodeficiency Virus SIVmac1A11. *Journal of Virology*. **66**(9):5432 – 5442.

Utay, N. and Douek, D et al. (2016) Interferons and HIV Infection: The Good, the Bad, and the Ugly. *Pathogens and Immunity*. **1**(1):107 – 116.

Varchetta, S. et al. (2013) Sialic acid-binding Ig-like lectin-7 interacts with HIV-1 gp120 and facilitates infection of $CD4^{pos}$ T cells and macrophages. *Retrovirology*. **10**:154.

Verweij, M. et al. (2021) Modulation of MHC-E transport by viral decoy ligands is required for RhCMV/SIV vaccine efficacy. *Science*. **371**(6541):eabe9233.

Wagh, K. et al. (2018) Completeness of HIV-1 Envelope Glycan Shield at Transmission Determines Neutralization Breadth. *Cell Reports*. **25**:893 – 908.

Walker, B. (2021) CMV, MHC-E, and the quest for an unconventional AIDS vaccine. *Science Immunology*. **6**:eabi5830.

Walter, M. (2020) The Role of Structure in the Biology of Interferon Signaling. *Frontiers in Immunology*. **11**:606489.

Wang, F-X. et al. (2008) APOBEC3G upregulation by alpha interferon restricts human immunodeficiency virus type 1 infection in human peripheral plasmacytoid dendritic cells. *Journal of General Virology*. **89**:722 – 730.

Wang, D. et al. (2011) Nonsynonymous substitution rate (Ka) is a relatively consistent parameter for defining fast-evolving and slow-evolving protein-coding genes. *Biology Direct*. **6**:13.

Wang, Y. et al. (2017) The V3 Loop of HIV-1 Env Determines Viral Susceptibility to IFITM3 Impairment of Viral Infectivity. *Journal of Virology*. **91**(7):e02441-16.

Wang, B. et al. (2023) $CD4^{+}$ Cytotoxic T Lymphocytes in Cancer Immunity and Immunotherapy. *Advanced Biology*. **7**:2200169.

Weissman, D. et al. (1995) Both a Precursor and a Mature Population of Dendritic Cells Can Bind HIV: However, only the Mature Population That Expresses CD80 Can Pass Infection to Unstimulated $CD4^{+}$ T Cells. *Journal of Immunology*. **155**:4111 – 4117.

van de Wijgert, J. et al. (2008) Bacterial Vaginosis and Vaginal Yeast, But Not Vaginal Cleansing, Increase HIV-1 Acquisition in African Women. *Journal of Acquired Immune Deficiency Syndromes*. **48**:203 – 210.

Wik, J. and Skålhegg, B. (2022) T Cell Metabolism in Infection. *Frontiers in Immunology*. **13**:840610.

Wilen, C. et al. (2011) Phenotypic and Immunologic Comparison of Clade B Transmitted/Founder and Chronic HIV-1 Envelope Glycoproteins. *Journal of Virology*. **85**(17):8514 – 8527.

Willey, S. et al. (2011) Extensive complement-dependent enhancement of HIV-1 by autologous non-neutralising antibodies at early stages of infection. *Retrovirology*. **8**:16.

Wood, N. et al. (2009) HIV Evolution in Early Infection: Selection Pressures, Patterns of Insertion and Deletion, and the Impact of APOBEC. *PLoS Pathogens*. **5**(5):e1000414.

Wu, L. et al. (1996) CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature*. **384**(6605):179 – 183.

Wu, X. et al. (2012) Selection Pressure on HIV-1 Envelope by Broadly Neutralizing Antibodies to the Conserved CD4-Binding Site. *Journal of Virology*. **86**(10):5844 – 5856.

Wymant, C. et al. (2022) A highly virulent variant of HIV-1 circulating in the Netherlands. *Science*. **375**(6580):540 – 545.

Xia, X. and Kumar, S. (2006) Codon-based detection of positive selection can be biased by heterogenous distribution of polar amino acids along protein sequences. *Computational Systems Bioinformatics*. **2006**:335 – 340.

Xing, S. et al. (2010) Increased Turnover of FoxP3^{high} Regulatory T Cells Is Associated With Hyperactivation and Disease Progression of Chronic HIV-1 Infection. *Journal of Acquired Immune Deficiency Syndromes*. **54**(5):455 – 462.

Yamada, E. et al. (2021) Intestinal microbial communities and *Holdemanella* isolated from HIV+/- men who have sex with men increase frequencies of lamina propria CCR5^{+} CD4^{+} T cells. *Gut Microbes*. **13**(1):e1997292.

Yang, H. et al. (2021) HLA-E-restricted, Gag-specific CD8^{+} T cells can suppress HIV-1 infection, offering vaccine opportunities. *Science Immunology*. **6**(57):eabg1703.

Yang, S. et al. (2022) Dynamic HIV-1 spike motion creates vulnerability for its membrane-bound tripod to antibody attack. *Nature Communications*. **13**(1):6393.

Yarchoan, R. and Uldrick, T. (2018) HIV-Associated Cancers and Related Diseases. *NEJM*. **378**(11):1029 – 1041.

Yonezawa, A. et al. (2003) Natural Alpha Interferon-Producing Cells Respond to Human Immunodeficiency Virus Type 1 with Alpha Interferon Production and Maturation into Dendritic Cells. *Journal of Virology*. **77**(6):3777 – 3784.

Yu, M. and Vajdy, M. (2010) Mucosal HIV transmission and vaccination strategies through oral compared with vaginal and rectal routes. *Expert Opinion on Biological Therapy*. **10**(8):1181 – 1195.

Yu, Q. et al. (2010) The good and evil of complement activation in HIV-1 infection. *Cellular and Molecular Immunology*. **7**:334 – 340.

Yuan, T. et al. (2013) HIV-1 Envelope Glycoprotein Variable Loops Are Indispensable for Envelope Structural Integrity and Virus Entry. *PLoS One*. **8**(8):e69789.

Yuan, Z. et al. (2017) Characterization of founder viruses in very early SIV rectal transmission. *Virology*. **502**:97 – 105.

Yun, T. et al. (2021) Human plasmacytoid dendritic cells mount a distinct antiviral response to virus-infected cells. *Science Immunology*. **6**(58):eabc7302.

Zanini, F. et al. (2015) Population genomics of intrapatient HIV-1 evolution. *eLife*. **4**:e11282.

Zanini, F. et al. (2017) Error rates, PCR recombination, and sampling depth in HIV-1 whole genome deep sequencing. *Virus Research*. **239**:106 – 114.

Zarei, M. et al. (2016) A Molecular Approach to Nested RT-PCR Using a New Set of Primers for the Detection of the Human Immunodeficiency Virus Protease Gene. *Jundishapur Journal of Microbiology*. **9**(7):e30365.

Zhang, Z-Q. et al. (1999) Sexual Transmission and Propagation of SIV and HIV in Resting and Activated $CD4^+$ T Cells. *Science*. **286**(5443):1353 – 1357.

Zhang, Y. et al. (2015) Interferon-induced transmembrane protein-3 rs12252-C is associated with rapid progression of acute HIV-1 infection in Chinese MSM cohort. *AIDS*. **29**:889 – 894.

Zhang, J. et al. (2021) Rapid Clinical Progression and Its Correlates Among Acute HIV Infected Men Who Have With Men in China: Findings from a 5-Year Multicenter Prospective Cohort Study. *Frontiers in Immunology*. **12**:712802.

Zhang, J. et al. (2022) Global, regional, and national burdens of HIV and other sexually transmitted infections in adolescents and young adults aged 10 – 24 years from 1990 to 2019: a trend analysis based on the Global Burden of Disease Study 2019. *Lancet Child and Adolescent Health*. **6**:763 – 776.

Zhou, S. et al. (2021) Primer ID Next-Generation Sequencing for Analysis of a Broad Spectrum Antiviral Induced Transition Mutations and Errors Rates in a Coronavirus Genome. *Bio-protocol*. **11**(5):e3938.

Zhou, S. et al. (2022) Unique Molecular Identifiers and Multiplexing Amplicons Maximize the Utility of Deep Sequencing To Critically Assess Population Diversity in RNA Viruses. *ACS Infectious Diseases*. **8**:2505 – 2514.

Zhu, P. et al. (2008) Cryoelectron Tomography of HIV-1 Envelope Spikes: Further Evidence for Tripod-Like Legs. *PLoS Pathogens*. **4**(11):e1000203.

Zou, Z. et al. (2011) Siglecs Facilitate HIV-1 Infection of Macrophages through Adhesion with Viral Sialic Acids. *PLoS One*. **6**(9):e24559.

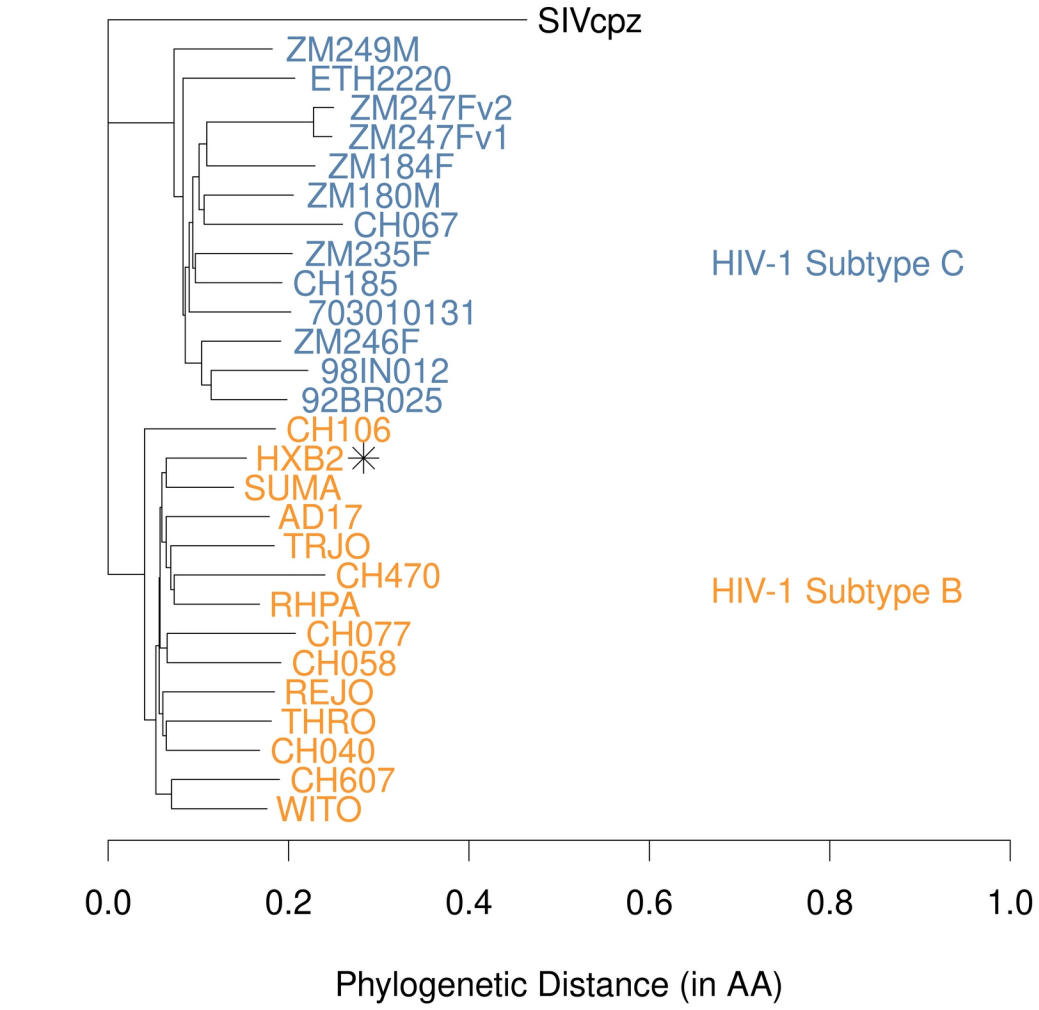
APPENDIX

Appendix 1: Table of HIV-1 Env proteins from documented T/Fs

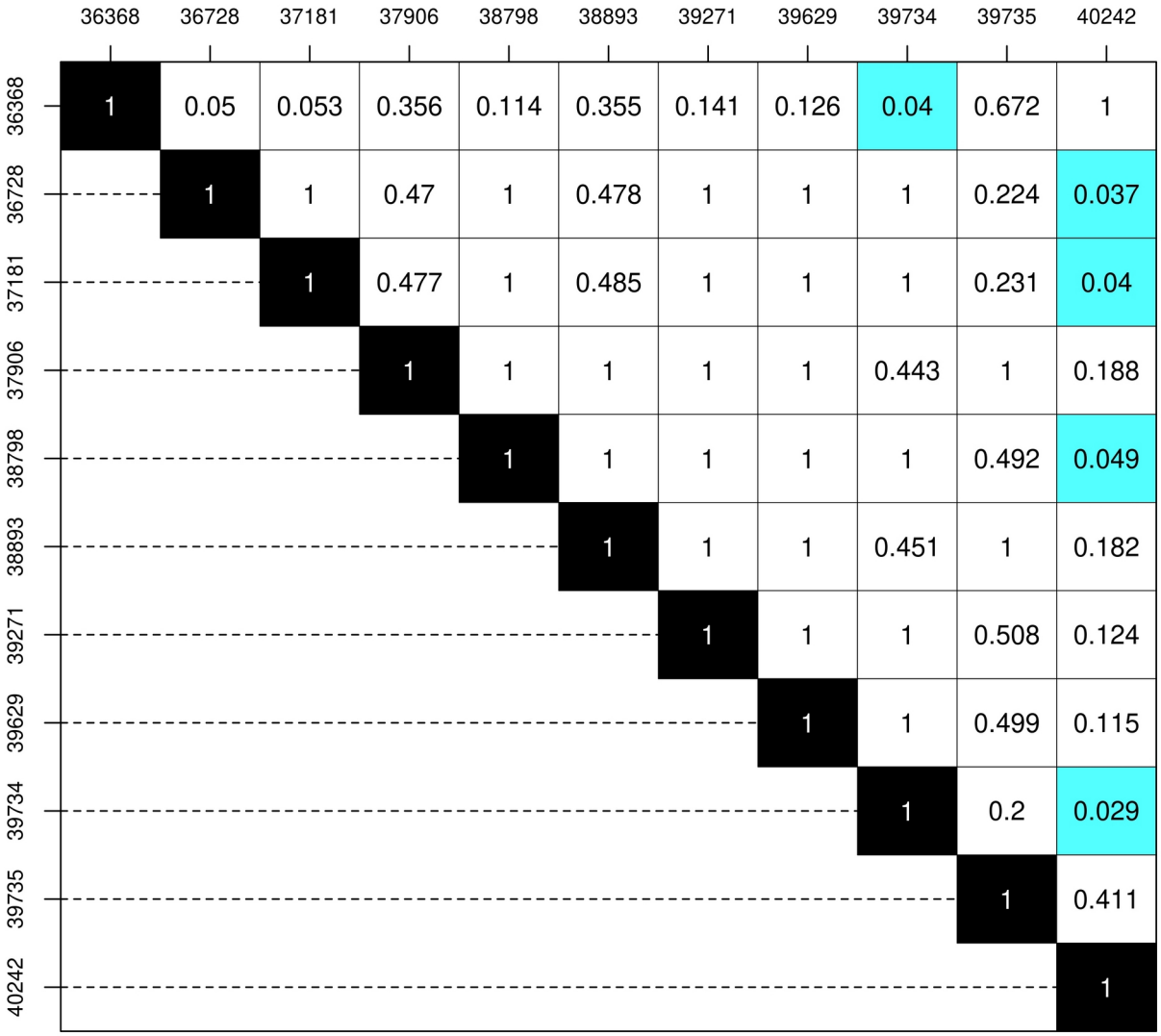
Patient Code	GenBank Accession #	HIV-1 Subtype	Source
CH106	AET76312.1	B	Ochsenbauer et al. [2012]
HXB2*	AAB50262.1	B	Ratner et al. [1985]
SUMA	AET76339.1	B	Keele et al. [2008]
AD17	AEO84467.1	B	Li et al. [2010]
TRJO	ACE70720.1	B	Keele et al. [2008]
CH470	AGG91976.1	B	Liu et al. [2013]
RHPA	AET76330.1	B	Keele et al. [2008]
CH077	AET76304.1	B	Ochsenbauer et al. [2012]
CH058	AET76296.1	B	Ochsenbauer et al. [2012]
REJO	ACE69207.1	B	Keele et al. [2008]
THRO	AET76348.1	B	Keele et al. [2008]
CH040	AET76289.1	B	Ochsenbauer et al. [2012]
CH607	AGG93086.1	B	Liu et al. [2013]
WITO	AET76364.1	B	Keele et al. [2008]
ZM249M	ABW85100.1	C	Salazar-Gonzalez et al. [2008]
ETH2220	AAB36507.1	C	Asmal et al. [2015]
ZM247Fv1	ABW85430.1	C	Salazar-Gonzalez et al. [2008]
ZM247Fv2	ABW85521.1	C	Salazar-Gonzalez et al. [2008]
ZM184F	ABW85902.1	C	Salazar-Gonzalez et al. [2008]
ZM180M	ABW95267.1	C	Salazar-Gonzalez et al. [2008]
CH067	AGF30468.1	C	Asmal et al. [2015]
ZM235F	ABW85240.1	C	Salazar-Gonzalez et al. [2008]
CH185	AGF30504.1	C	Asmal et al. [2015]
703010131	AGV38407.1	C	Ping et al. [2013]
ZM246F	ACR53189.1	C	Salazar-Gonzalez et al. [2008]
92IN012	AAK31033.1	C	Asmal et al. [2015]
92BR025	AAB61124.1	C	Asmal et al. [2015]
SIVcpz**	WLD25288.1	N/A	N/A

* HIV-1 Subtype B Refseq used for positional numbering in gp120

** Outgroup used in the phylogenetic tree for gp160 of T/Fs



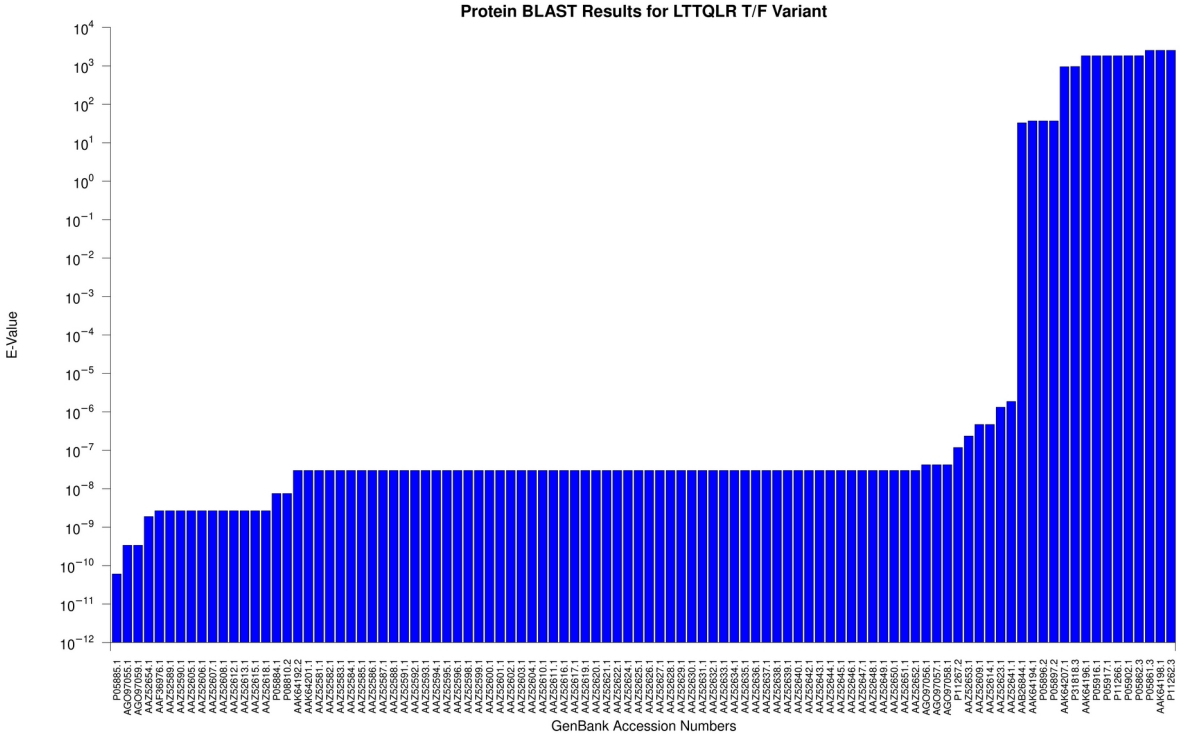
Appendix 2: Phylogenetic tree for gp160 precursor protein of T/F variants. Maximum-likelihood tree was constructed for the gp160 precursor protein alignment of 13 documented clade-B and clade-C T/F variants. Outgroup is the gp160 precursor from SIVcpz, the most recent common ancestor of all known HIV-1 groups and clades. HXB2, indicated by the asterisk, is the refseq used for numbering gp120 amino acid positions in analyzed sequences.



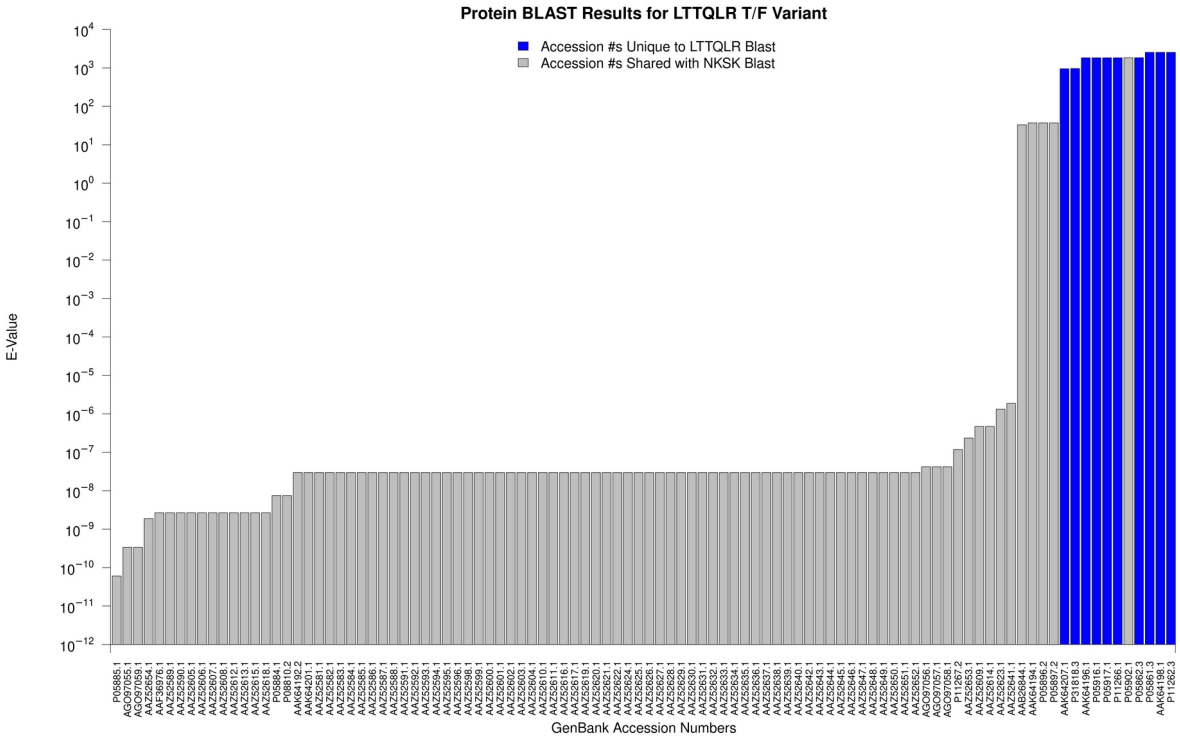
Appendix 3: Fisher's exact test results for comparisons in the distribution of the numbers of SGA-sequenced variants represented by the major (T/F) and minor (transient) lineages in each animal. The numbers of variants were enumerated on the basis of whether they were part of the major T/F lineage or a minor transient (non-T/F) lineage in each animal. Comparisons were examined statistically using Fisher's exact test (without Bonferroni correction) to evaluate whether lineage distribution between each animal was statistically different. Boxes colored in cyan indicate statistically significant differences between the two animals represented at each intersection. Boxes colored in black indicate equal comparisons between the same animals, necessarily resulting in p-values equal to 1.

Appendix 4 – 15: Protein BLAST results shared and unique to antibody-selected and non-selected T/Fs. Protein BLAST searches were performed for each selected polymorphism from T/Fs selected during antibody opsonization and the consensus amino acid identities for respective polymorphisms. Sequence queries for the BLAST search of consensus T/F variants are "WFLNWVEDRNLTLTTQKLRE" (LTTQLR, Appendices 4 and 5), "YSQQIHTQQ" (I743, Appendices, 8 and 9), and "VYQILQPILQGLSATL" (G812, Appendices 12 and 13). Sequence queries for the BLAST search of antibody-selected T/F variants/mutants are "WFLNWVEDRN—LTNKKSKE" (NKSK, Appendices 6 and 7), "YSQQTHTQQ" (I743T, Appendices 10 and 11), and "VYQILQPILQRLSATL" (G812R, Appendices 14 and 15). BLAST searches are displayed on the x-axis with E-values displayed on the y-axis in $\log_{10}$ scale. Search results often showed overlapping accession numbers shared by both consensus and antibody-selected T/F mutations due to similarities in the sequences other than the single amino acid differences between the consensus and mutants. Analysis was performed to identify accession numbers that are either shared among consensus and mutant search queries (indicated by gray bars) or unique to each consensus and mutant search queries (indicated by blue bars) in Appendices 5, 7, 9, 11, 13, and 15. Accession numbers with the lowest E-values or those unique to search queries were used to identify scientific publications that reported similar polymorphisms as those identified in the current study.

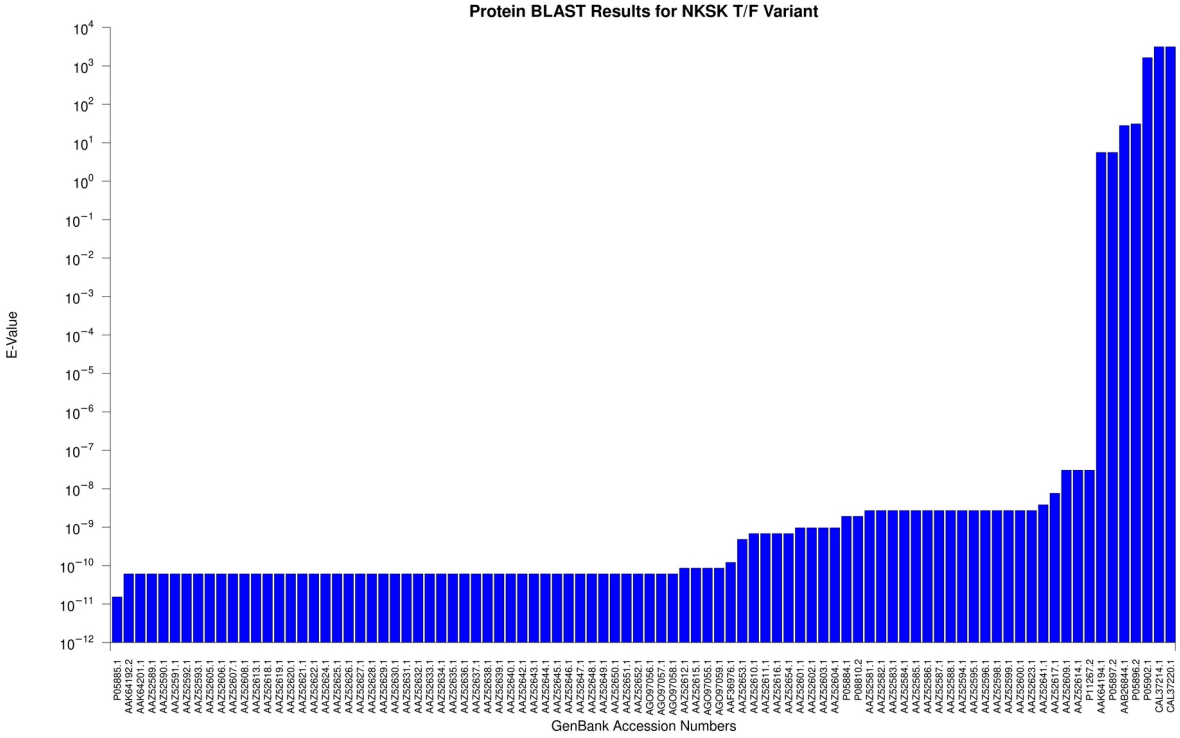
Appendix 4



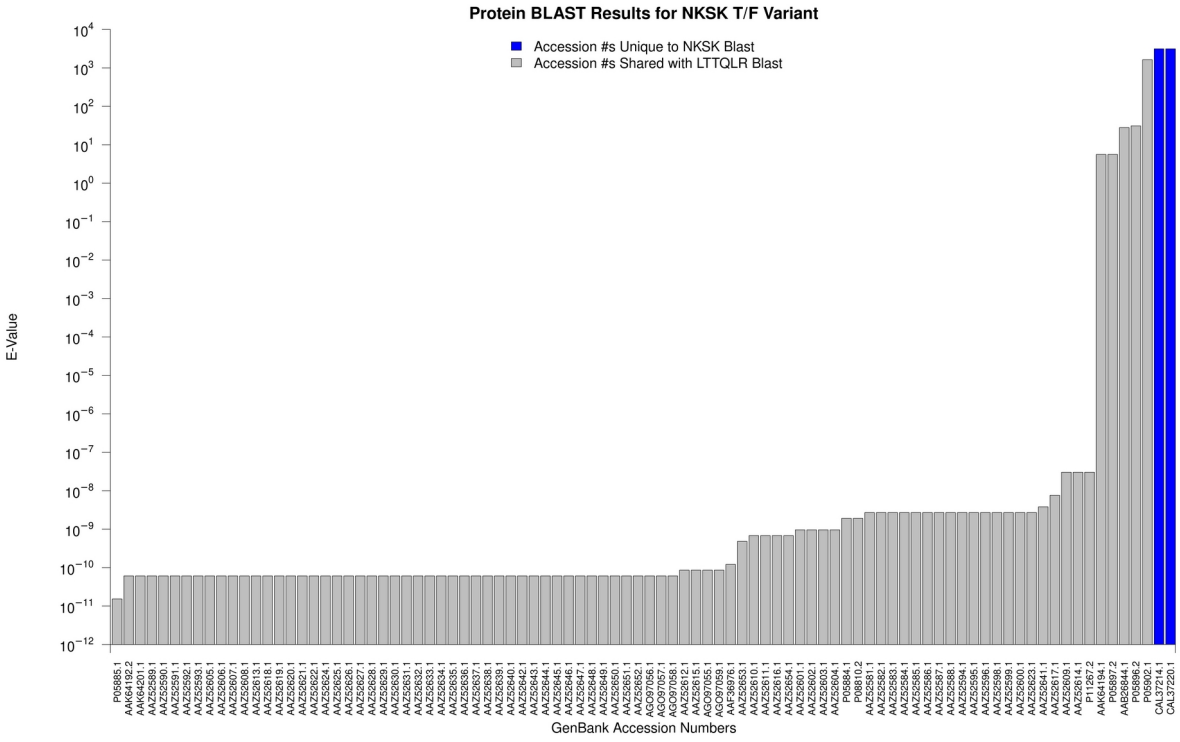
Appendix 5



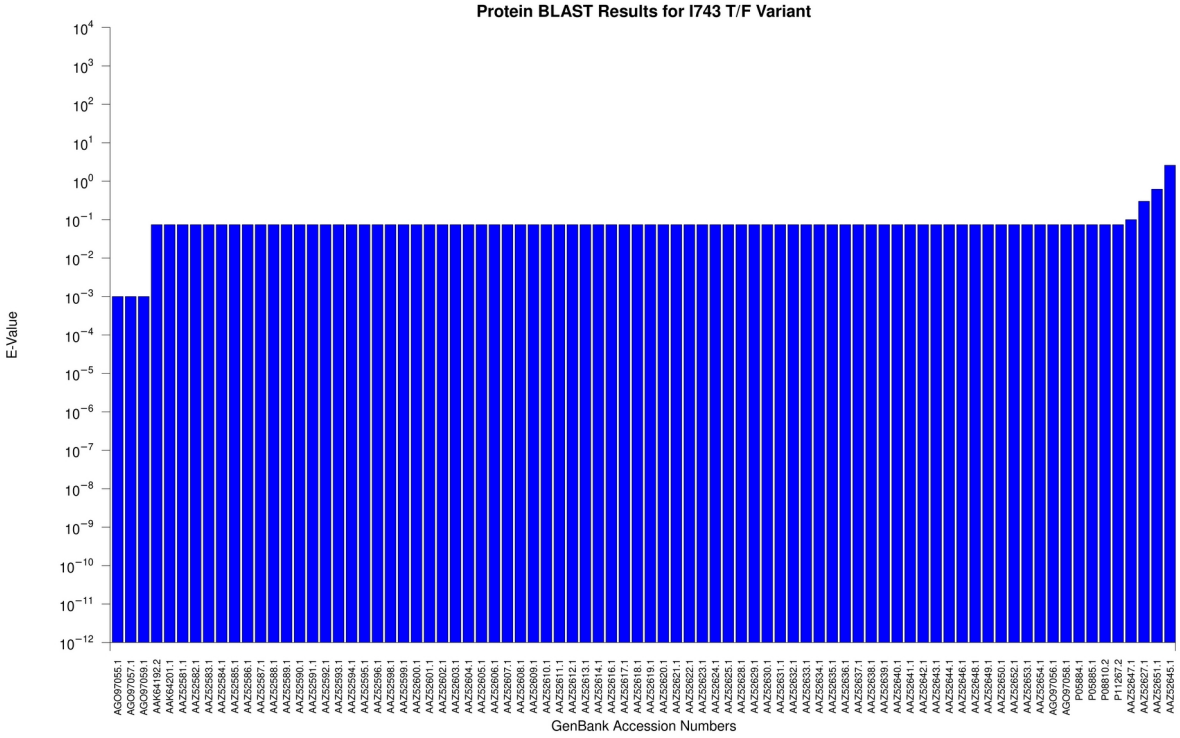
Appendix 6



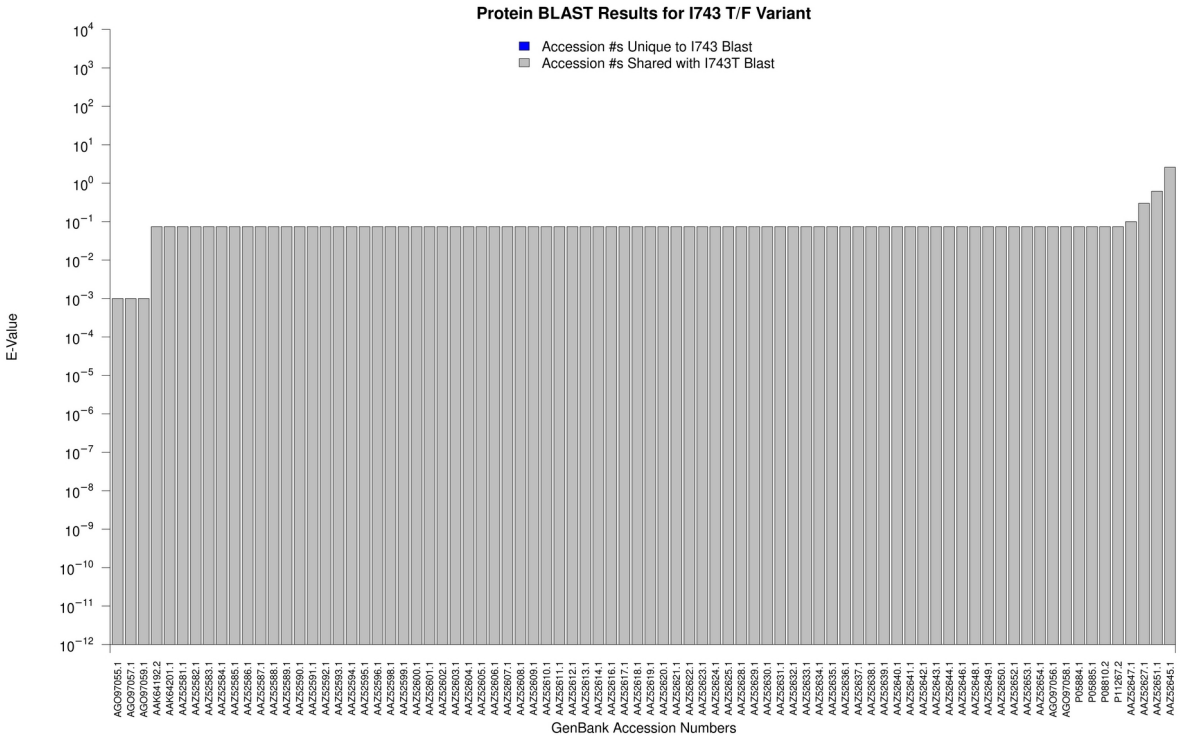
Appendix 7



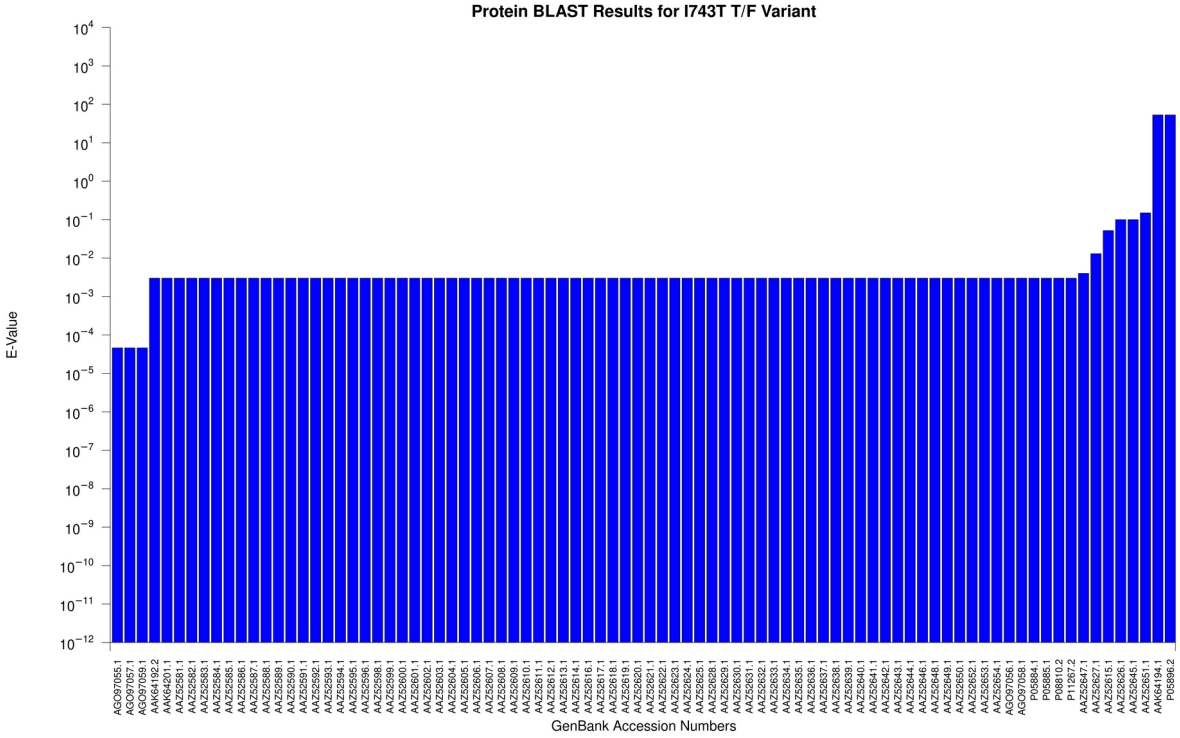
Appendix 8



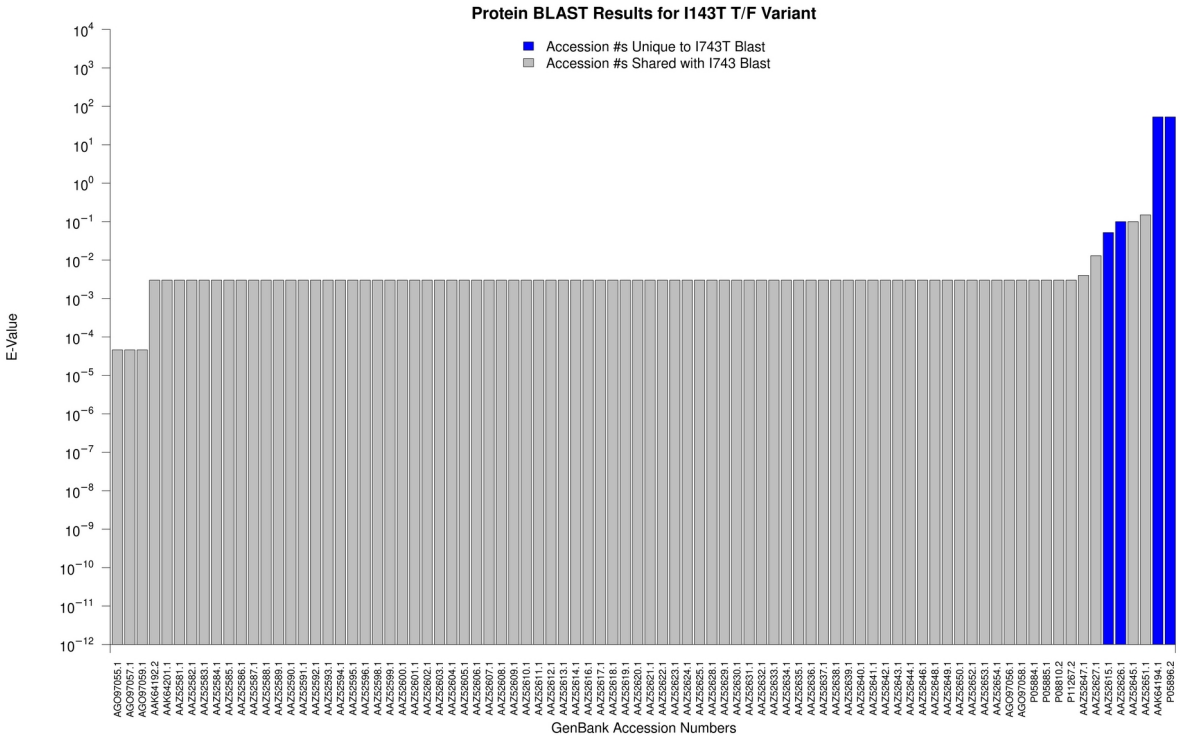
Appendix 9



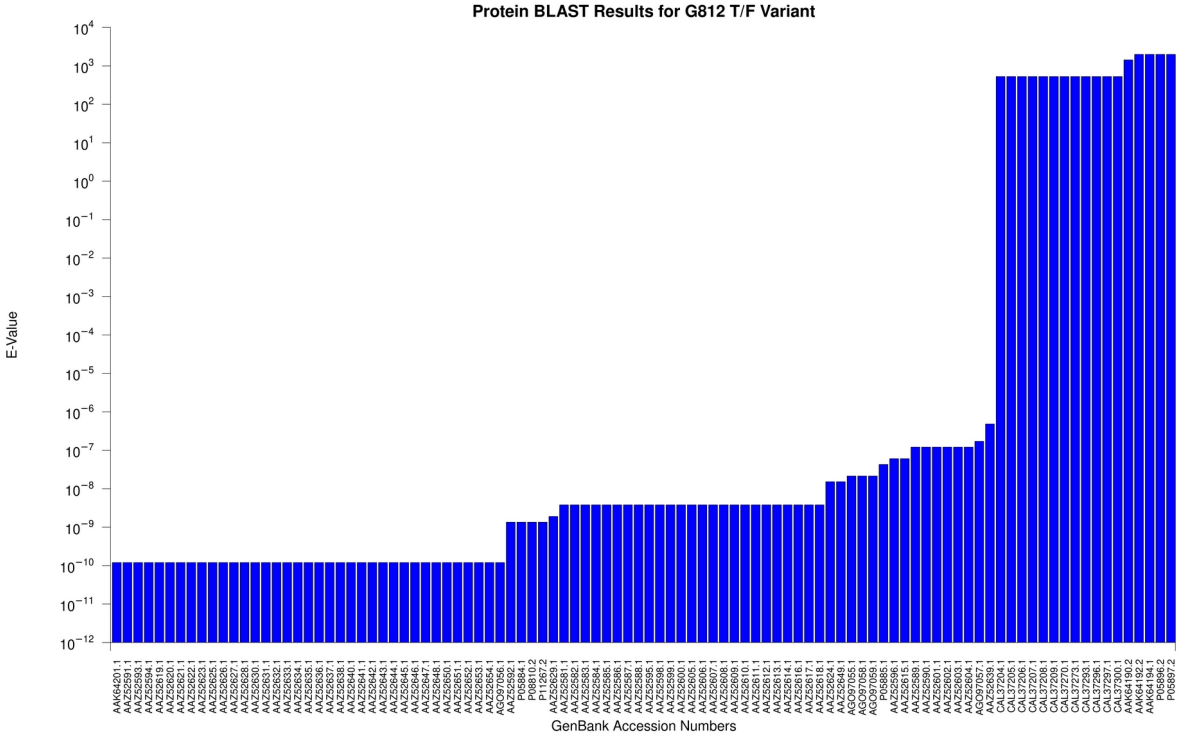
Appendix 10



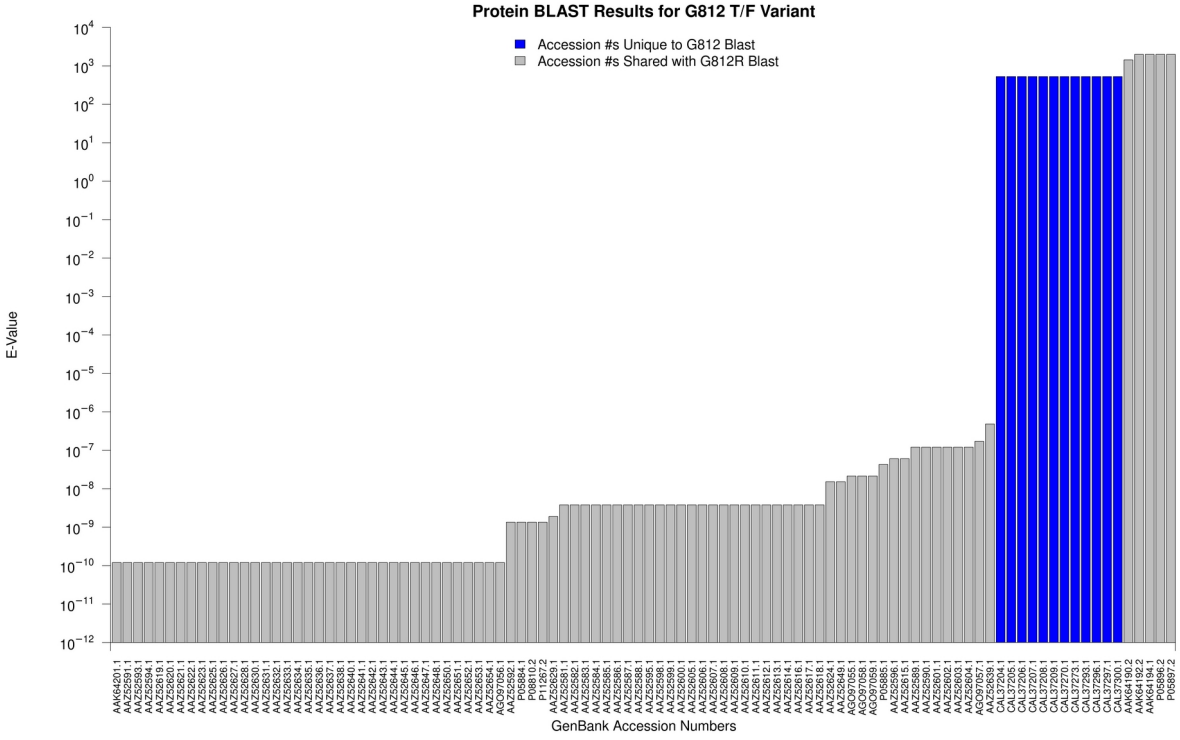
Appendix 11



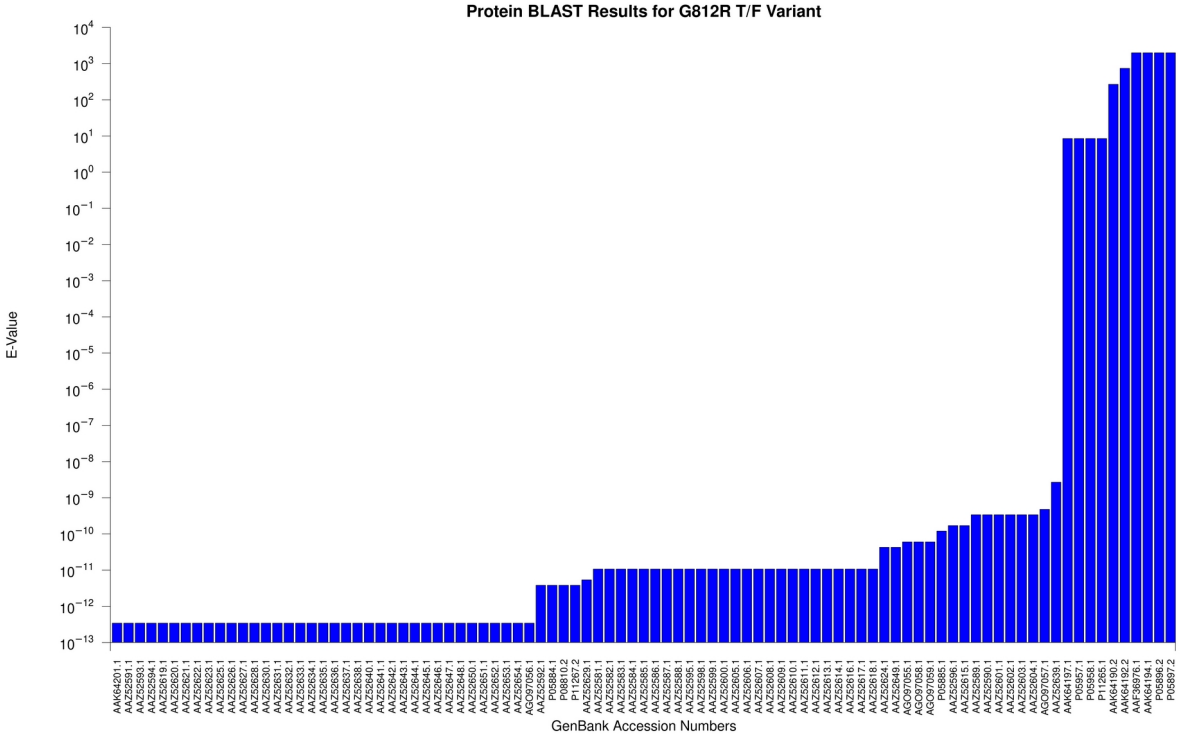
Appendix 12



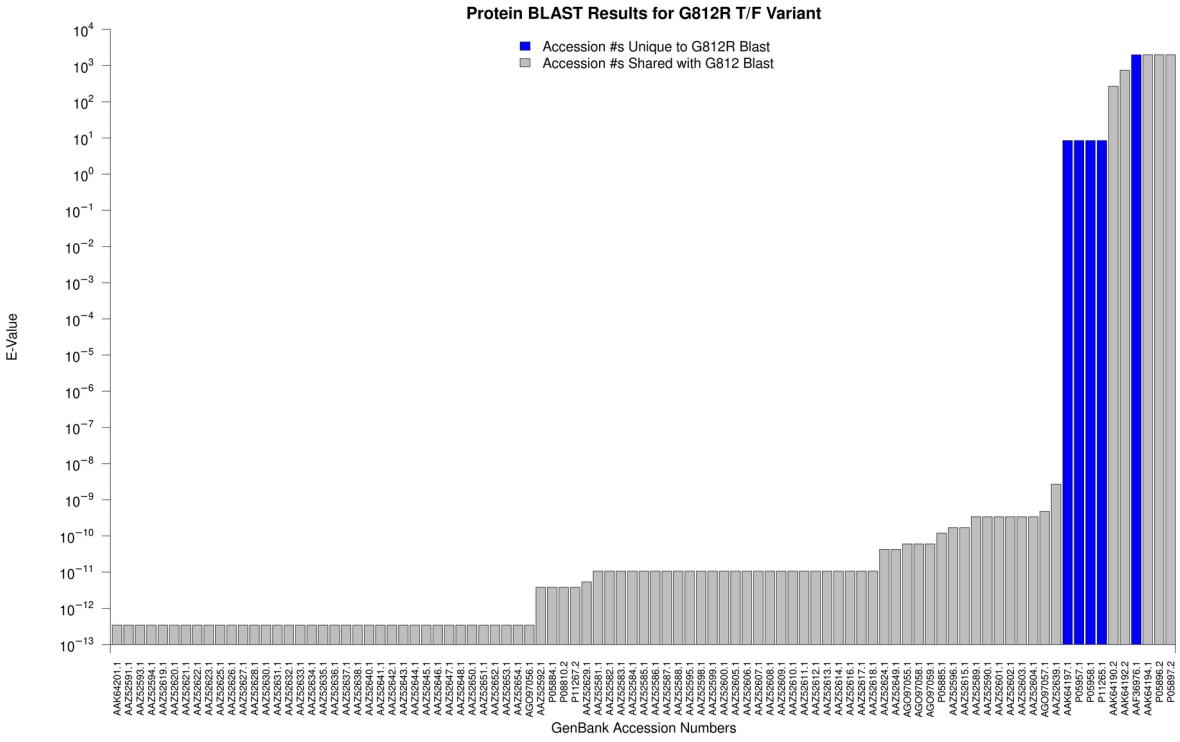
Appendix 13



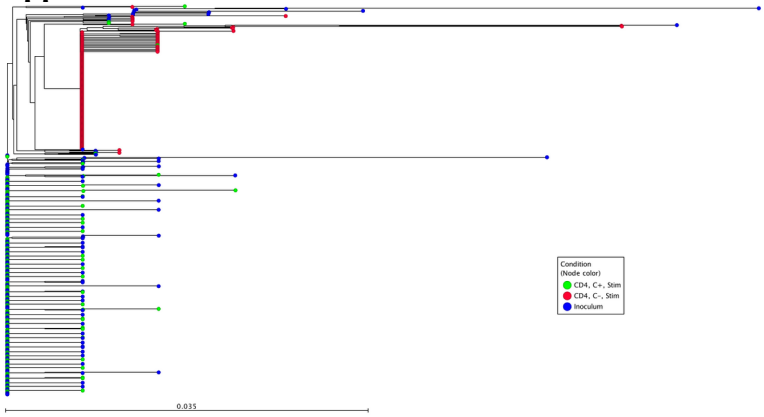
Appendix 14



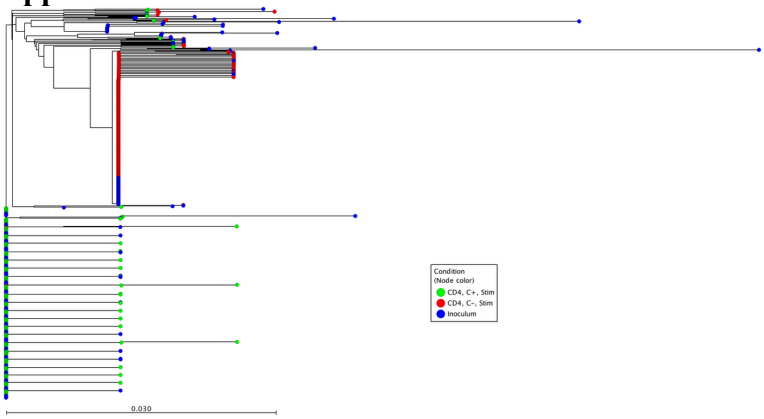
Appendix 15



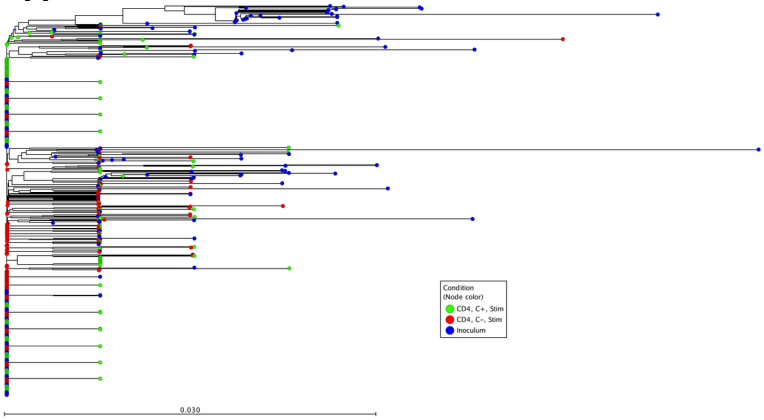
Appendix 16



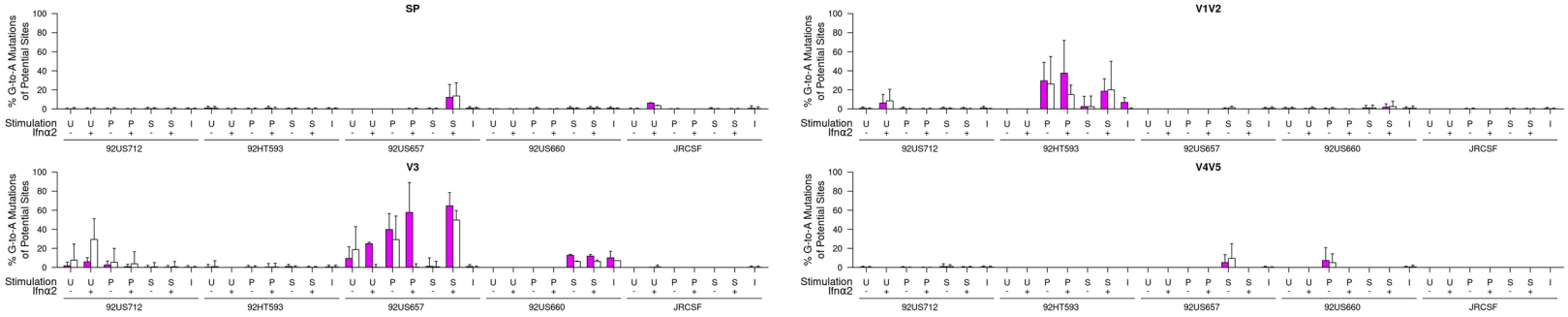
Appendix 17



Appendix 18



Appendix 16 – 18: Phylogenetic trees of viruses from stimulated CD4$^+$ T-lymphocytes directly infected with a complement-opsonized or non-opsonized inoculum. T-lymphocytes were stimulated and challenged directly with a sample inoculum of clinical isolate 92US657 that was either incubated in the presence of activated human complement (opsonized) or inactivated human complement (non-opsonized). Output virus from these infections were PCR-amplified over the variable loop regions, sequenced, and a random sample of 200 reads from each condition and each amplicon were aligned using CLC Genomics Workbench. Neighbor-joining trees were constructed with the same program and shown here for V1V2 (Appendix 16), V3 (Appendix 17), and V4V5 (Appendix 18). Output virus from complement-opsonized infections are shown in gree tips and those from non-opsonized infections in red tips, and inoculum variants are shown in blue tips.



Appendix 19: Calculation of APOBEC-3G hypermutation rates and baseline G-A mutation rates in DNA reads from each infection condition. DNA reads from each infection condition (including the inoculum) were aligned, and APOBEC-3G hypermutation signatures and baseline G-A mutation signatures were detected and enumerated using the Hypermut tool from hiv.lanl.gov. Data is displayed as a percentage of detected mutation signatures over all possible mutation-possible sites in each sequence and shown in the bar plot as mean ± standard deviation. APOBEC-3G hypermutation signatures are shown in fuchsia while baseline G-A mutation signatures are in white.