Examination of HIV evolution in response to host pressures

by

Anh Quang Le

B.Sc., Simon Fraser University, 2012

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Master of Science Program
Faculty of Health Sciences

SIMON FRASER UNIVERSITY

Summer 2015

Approval

Name: **Anh Quang Le**

Degree: **Master of Science**

Title: ***Examination of HIV evolution in response to host pressures***

Examining Committee: **Chair:** Dr. Masahiro Niikura
Associate Professor

Dr. Zabrina Brumme
Senior Supervisor
Assistant Professor

Dr. Art Poon
Supervisor
Adjunct Professor

Dr. Ryan Morin
Supervisor
Assistant Professor
Department of Molecular Biology and Biochemistry

Dr. William Small
Supervisor
Assistant Professor

Dr. Ralph Pantophlet
External Examiner
Associate Professor

Date Defended/Approved: June 12, 2015

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

update Spring 2010

Abstract

The overarching aim of this thesis was to study the evolution of HIV-1 in response to host pressures. The main data chapter comprises a detailed HIV-1 transmission study where we identified a putative case of X4 HIV-1 transmission from a CCR5-wt/wt donor to a recipient homozygous for the naturally-occurring 32 base pair deletion in the CCR5 gene (CCR5-Δ32/Δ32). This rare genotype confers resistance to infection by CCR5-using ("R5") HIV-1 strains not CXCR4-using ("X4") strains. Using ultradeep sequencing and phylogenetic analysis, we estimate the number of founder viruses that established infection in both donor and recipient (one in each case), reconstruct their sequences, and study within-host HIV-1 evolution and coreceptor usage. Notably, results suggest that HIV-1 infection in the recipient was initiated by transfer of an infected cell (*i.e.* not a virion) from the donor, and reveal differential HIV-1 evolution in both members of the pair.

Keywords: HIV-1, evolution, next-generation sequencing (NGS), ancestral reconstruction, CCR5-Δ32/Δ32, transmission

Dedication

I would like to dedicate this to my parents Dung Le and Do Nguyen and my sisters Hong and Kieu for everything I have achieved. Without their encouragement and support none of this would be possible. I would also like to dedicate this to my partner Vanessa Ho for always sticking by me throughout the years.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Zabrina Brumme for her continued support, guidance, and the various opportunities her laboratory provided me during my graduate and undergraduate degree. I would also like to thank my committee members Dr. Art Poon, Dr. Ryan Morin, and Dr. Will Small for their guidance and mentorship.

I would also like to thank past and present peers who have offered continued support and advice: Tallie, Gursev, Eric, Tristan, Anna, Bemulu, Philip, Aniqa, Arthur, Laura, Natalie, and others. I would also like to thank my research collaborators at the British Columbia Centre for Excellence in HIV/AIDS for guidance, knowledge, and expertise.

Table of Contents

List of Figures

List of Acronyms

BNAb	Broadly neutralizing antibody
CCR5	C-C chemokine receptor type 5
CtC	Cell-to-cell
CTL	Cytotoxic T-Lymphocyte
CXCR4	C-X-C chemokine receptor type 4 or fusin
HIV-1	Human immunodeficiency virus 1
HLA	Human leukocyte antigen
NAb	Neutralizing antibody
NGS	Next-generation sequencing
NK-cell	Natural killer cell
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
PVL	Plasma viral load
R5	CCR5 using HIV-1
R5/X4	Dual-tropic HIV-1
RT-PCR	Reverse transcriptase polymerase chain reaction
T-cell	T lymphocyte
T/F	Transmitted founder
X4	CXCR4 using HIV-1

Chapter 1.

Introduction to HIV-1

1.1. Introduction

Human Immunodeficiency Virus Type 1 (HIV-1), the causative agent of Acquired Immunodeficiency Syndrome (AIDS), was first identified as a novel pathogen in 1983 [1,2]. To date, a cumulative total of 74 million people have been infected with HIV-1 [3]. A total of 39 million have died, and approximately 35 million currently live with HIV-1 [3]. More than half of HIV-1 infected persons globally reside in Sub-Saharan Africa [3]. In North America, an estimated 1.3 million individuals were living with HIV/AIDS in 2013 [4]. In Canada, an estimated 72,000 individuals are HIV-positive, with 25% of these individuals unaware of their HIV status [4].

1.2. Discovery

The first cases of the syndrome later to be known as AIDS were reported between October 1980 and May 1981 in Los Angeles, USA when two young homosexual men were diagnosed with *Pneumocystis carinii pneumonia* (PCP), a rare form of pneumonia [5]. In the following months, the U.S. Centers for Disease Control and Prevention (CDC) reported additional cases of PCP, Kaposi's sarcoma (KS) (a rare skin cancer), and other rare opportunistic infections appearing primarily in men who have sex with men across the United States [6]. Most of these patients died shortly thereafter. The same syndrome was subsequently identified in heterosexual Haitian immigrants, injection drug users, and hemophiliacs, suggesting a blood borne pathogen [7,8]. Analysis of blood cells from persons with AIDS revealed low numbers of CD4+ T-cells, an observation that provided the first clue that the etiologic agent damaged the immune

system [9]. In 1983, two separate research teams, led by Dr. Robert Gallo of the National Cancer Institute in Maryland, USA and Dr. Luc Montagnier of the Pasteur Institute in Paris, France reported the discovery of a novel retrovirus that infected T cells in AIDS patients [1,2]. Originally classified as the third member of the Human T-Lymphotropic Virus (HTLV) family [1,2], this novel retrovirus was later determined to be a member of the genus *Lentiviridae* belonging to the *Retroviridae* family and given the name Human Immunodeficiency Virus (HIV) [10].

1.3. Origin and diversity

HIV originates from zoonotic transmissions of Simian Immunodeficiency Viruses (SIV) that are found in non-human primates [11]. HIV is one of the most genetically diverse pathogens known and can be classified into two types, HIV-1 and HIV-2. HIV-1 is further classified into four groups: M (main), N (non-main), O (outlier), and P [12-14]. HIV-1 groups M, N, O, and P as well as HIV-2 strains each originate from a separate zoonotic transmission from a non-human primate host harbouring a species-specific SIV infections. All of these cross-species transfers occurred over the last century [11]. There are over 45 variants of SIV [15], each infecting a specific nonhuman primate species [16,17]. Using phylogenetic methods, it has been inferred that three of these – specifically, SIV infecting chimpanzees (*Pan troglodytes troglodytes*; SIVcpz), gorillas (*Gorilla gorilla*; SIVgor) and sooty mangabeys (*Cercocebus atys*; SIVsmm) were transferred to humans. HIV-1 group M and N are most closely related to SIVcpz [18], HIV-1 group O and P likely originated from SIVgor found in south-western and central Cameroon respectively [14], and HIV-2 can be divided into 8 groups (A-H) each arising from a separate zoonotic transmission event from contact with *Sooty mangabeys* infected with SIVsmm. HIV-1 group N, O, and P account for approximately 100,000 infections globally [14], whereas HIV-2 accounts for approximately 1-2 million infections in western-Africa with few reported cases globally [19]. These will not be discussed further here.

HIV-1 group M strains are the most genetically diverse and widespread globally and are responsible for the HIV pandemic [11]. HIV-1 group M can be further subdivided into 9 distinct subtypes (A, B, C, D, F, G, H, J, and K), and 4 sub-subtypes (A1, A2, F1,

and F2) (Figure 1.1) [12]. In addition, as of April 2015 there are 72 circulating recombinant forms (CRFs) [20,21]. CRFs arise when an individual is infected with two different HIV-1 strains which combine to create a recombinant virus that subsequently outcompetes its parent strains within a single host and is transmitted to others [11]. Strains are classified as CRFs if at least three epidemiologically unlinked cases are identified [22]. HIV-1 subtype E was initially classified as a distinct subtype through phylogenetic analysis of the *env* gene, but further analysis of other gene regions of viruses classified as subtype E revealed similarities to subtype A. Thus, subtype E was reclassified as CRF01_AE, the first identified CRF [23].

HIV-1 group M subtypes are differentially distributed globally. The greatest HIV-1 group M genetic diversity worldwide is observed in the proposed epicenter of the HIV epidemic in the Congo River basin: here, essentially all HIV-1 group M subtypes can be found. In contrast, the predominant subtype in North America and the Western world is B [11]. Subtype A variants dominate throughout regions of Russia and the middle east along with subtypes B and C [11]. In Asia multiple subtypes are in circulation including subtype B, C, and CRF01_AE [11]. HIV-1 group M subtype C accounts for 50% of group M infections globally with the majority occurring in Sub-Saharan Africa [24]. Subtype A, B, D account for ~26%, ~12%, and ~5% of infections worldwide with others occurring in lower frequencies throughout the world [24]. HIV-1 extensive genetic diversity and its unequal distribution globally represent major challenges to the development of an effective HIV-1 vaccine. As such, a deeper understanding of the factors affecting the diversity, evolution, and transmission of HIV-1 are extremely important.

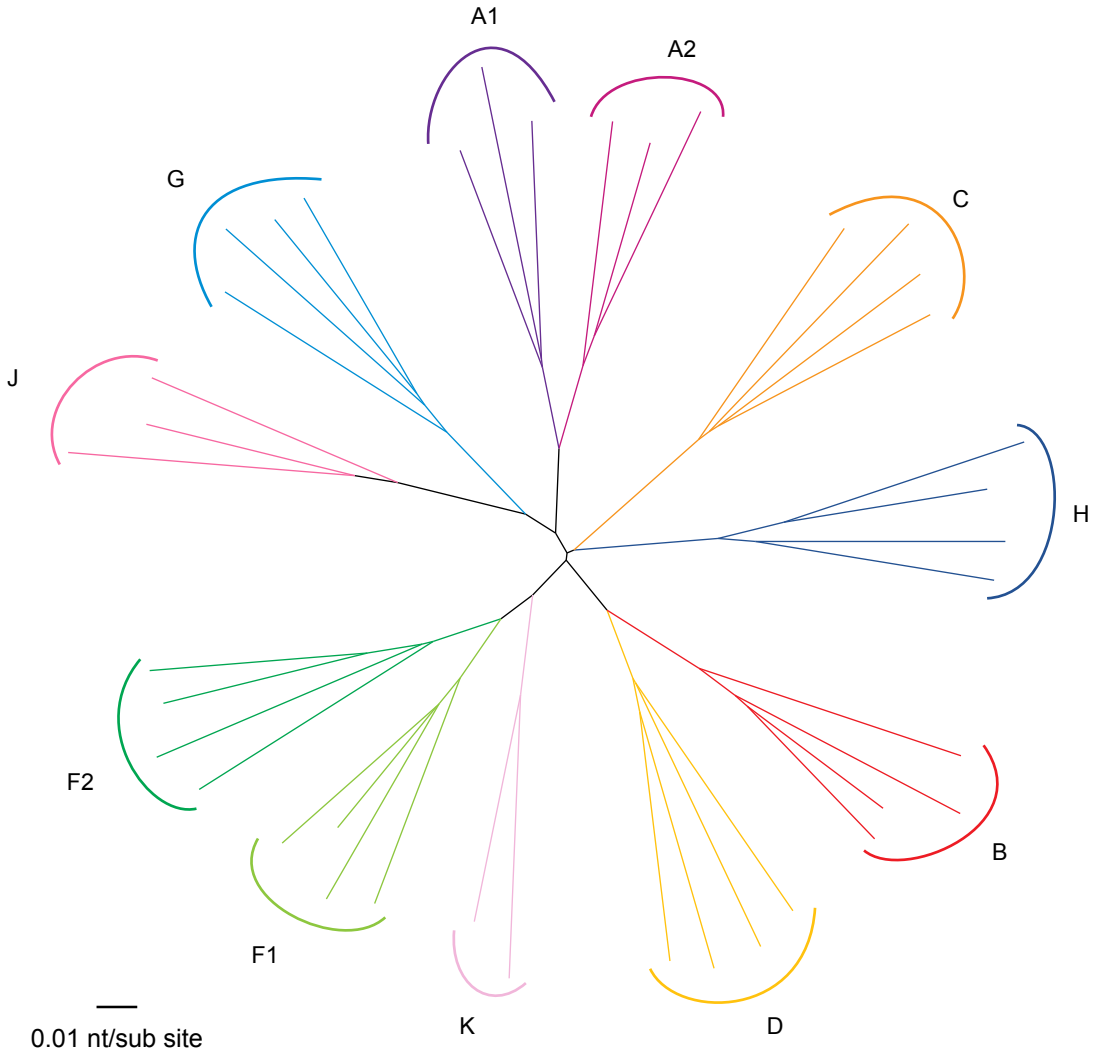

Figure 1.1. HIV-1 group M subtype diversity
Maximum likelihood phylogenetic tree constructed using the envelope sequence of HIV-1 group M subtype reference sequences for A1, A2, B, C, D, F1, F2, G, H, J, and K obtained from the HIV Los Alamos Database.

1.4. Pathogenesis

HIV-1 is a pathogenic retrovirus that infects CD4+ T-cells, macrophages, and other immune regulatory cells such dendritic cells [25-27]. Infection occurs when HIV-1 infected bodily fluids come into contact with mucosal membranes, abrasions on the skin, or are introduced directly into the bloodstream [10]. Infection can occur via sexual contact, unsafe injection practises, or from mother to child in utero, during delivery, or

during breastfeeding [10]. Historically, prior to the availability of tests to screen the blood supply, HIV-1 could also be transmitted via blood transfusions or organ transplants. If left untreated, HIV-1 progressively damages the host immune system, and over an average of 5 to 10 years eventually leads to AIDS, a syndrome defined as the presence of an AIDS defining illness or opportunistic infection (such as Kaposi's sarcoma or *Pneumocystis carinii* infection) or a CD4+ T-cell count of less than 200 cells/mm^3 [28].

However, major advances in antiretroviral therapy and treatment regimes since the mid-1990s have greatly reduced levels of HIV-related morbidity and mortality, transforming HIV-1 infection into a chronic manageable condition [29]. Modern combination antiretroviral therapy (cART) involves the use of multiple antiretroviral drugs (usually 3 from multiple drug classes) to inhibit various stages of the viral life cycle. Modern cART can suppress plasma viral loads (pVL) to undetectable levels for prolonged periods [30,31]. Once initiated however, cART must be maintained for life in order to sustain viral suppression and prevent the emergence of drug resistance mutations.

CART also serves as prevention [30,32]. Specifically, studies of serodiscordant couples [30] and mother-to-child transmission [32] indicate that if an HIV-infected person is fully virally suppressed on cART, the risk of transmitting HIV-1 approaches zero. CART can also be used as post-exposure prophylaxis (PEP) in cases where individuals are exposed to HIV-1 (e.g. occupational exposure in medical settings and infants born to HIV-1 infected mothers [33]). Lastly, current guidelines also approve daily antiretroviral therapy as a pre-exposure prophylaxis (PrEP) for HIV-negative individuals at high risk for HIV-1 [34].

1.5. Genetic organization and life cycle

The HIV-1 particle or "virion" houses two single positive sense RNA strands approximately 10,000 base-pairs long. Like all retroviruses, HIV-1 is organized into three major genes, *gag* (group-specific antigen), *pol* (polymerase), and *env* (envelope) (Figure 1.2). Gag is transcribed as a single precursor polyprotein pr55gag, which is then cleaved by HIV-1 protease into various structural proteins: matrix protein (MA, p17), capsid

protein (CA, p24), spacer peptide 1 (SP1; p2), nucleocapsid protein (NC, p7), spacer peptide 2 (SP2, p1), and p6 protein. The *pol* region encodes the viral enzymes protease (PR, prot), reverse transcriptase (RT, p51), and its ribonuclease H domain (RNaseH; p15), and integrase (int, p31). Pol is expressed as a part of a gag-pol polyprotein precursor (pr160gag-pol), generated by a ribosomal frameshift during translation of viral mRNA. Pol is also cleaved by HIV-1 protease into individual proteins. The *env* gene encodes precursor protein gp160, which is cleaved by the host protease furin to yield the envelope surface glycoprotein (SU, gp120) and the transmembrane glycoprotein (TM, gp41). In addition, the HIV-1 genome encodes additional regulatory and accessory proteins vif, vpr, vpu, tat, rev, and nef.

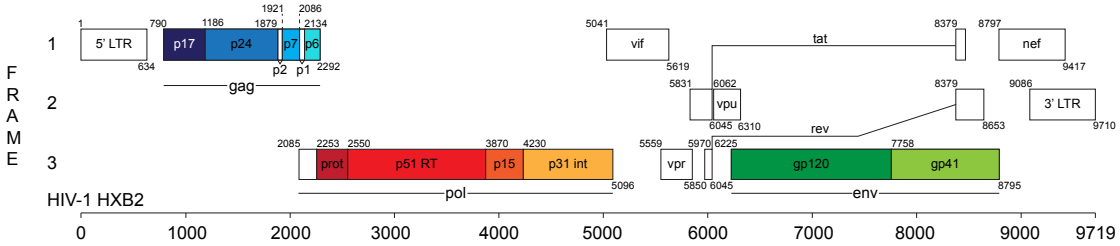

Figure 1.2. HIV-1 HXB2 genetic map
HIV-1 HXB2 reference standard genetic map

HIV-1 entry into the host cell begins by sequential interactions between HIV-1 envelope glycoproteins gp120 and gp41 and host cellular proteins. Before envelope glycoprotein mediated entry can occur, the virus particle first must come into close contact with the target host cell. This process can be mediated by the viral glycoproteins interacting with or host cell membrane polysaccharides such as heparin sulphate or proteins such as DC-SIGN [35,36]. These interactions allow HIV-1 to come in close contact with the target cell and recruit CD4 and the necessary coreceptors (Figure 1.3) [35,36]. After this, gp120 binds to the host CD4 receptor [35,36]. This binding causes gp120 to undergo a conformational change that exposes the coreceptor binding site envelope-V3 (V3). Gp120 is composed of 5 constant regions (C1-C5) and 5 variable loops (V1-V5). The V3 loop is the principle genetic determinant of coreceptor binding [37] and functions alongside the V1, V2, and C4 regions of gp120 to bind to the host coreceptor [35,36]. Depending on the strain, the V3 loop is able to bind to CCR5 and/or CXCR4 on the host cell surface [38]. After coreceptor binding to the V3 loop, gp120

undergoes further conformational change, allowing gp41-mediated membrane fusion [35,39].

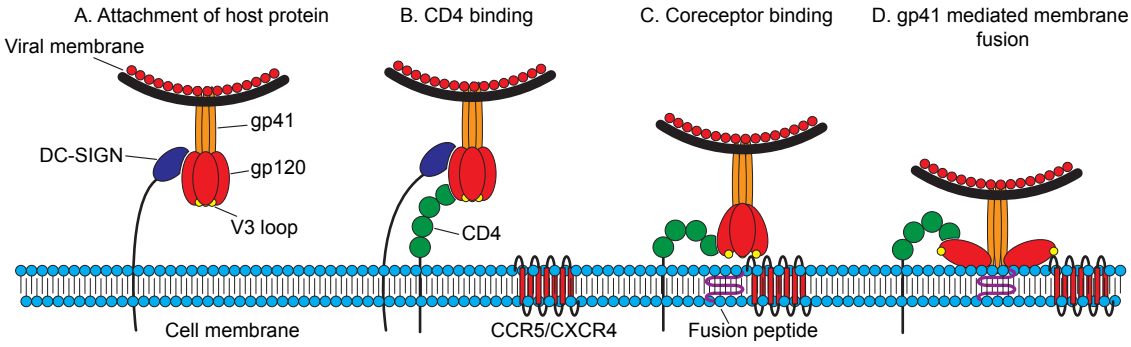

Figure 1.3. HIV-1 entry
A) Host cellular proteins such as DC-SIGN binding to gp120 bringing it closer to the target cell. B) Attachment of CD4 to the CD4 binding groove of gp120 causes a conformational change exposing the V3 loop. C) The V3 loop on the surface of gp120 subsequently binds to the host coreceptor (CCR5 or CXCR4) resulting in another conformational change. D) Lastly membrane fusion is facilitated by gp41 and fusion proteins found in the host plasma membrane.

After membrane fusion, the virion contents enter the target cell. Here, viral RNA is converted to a double-stranded DNA copy by the viral reverse transcriptase enzyme (which was packaged inside the virion) [10]. This DNA copy is then translocated into the host cell nucleus and integrated into the genome by HIV-1 integrase and other host and viral proteins. The integrated HIV-1 DNA is referred to as a “provirus”. The HIV-1 provirus is then transcribed by host cell polymerase to yield full-length genomic RNAs. Some of these full-length RNAs serve as mRNA templates for the production of viral proteins [10]. In these cases, some are multiply spliced (to generate messages for *tat and rev*) some are singly spliced (to generate *env, vif, vpr, vpu and nef*) and some serve directly a templates for translation (to generate *gag* and *pol*). Yet other full-length RNA copies serve as viral genomes for packaging inside new virions. The final stages of the viral life cycle include assembly of new viral particles. Gag proteins play a crucial role in coordinating the packaging of viral RNA and accessory proteins into virions. These new virus particles bud from the host cell membrane and go on to infect other target cells [10].

It is also important to note that, in a small minority of HIV-1 infected cells, HIV-1 integrates its genome into the host cells, but the provirus remains in a transcriptionally

inactive latent state that can persist for very long periods. Because these latently-infected cells are transcriptionally inactive they are undetectable by the host immune system, but they may reactivate and produce new virions at any time. As such, latently-infected cells represent a major hurdle for HIV cure.

1.6. Thesis objective and overview

The overarching aim of this thesis is to examine HIV evolution in response to selection pressures imposed by its human host. These include pressures imposed by host immune responses (e.g. Chapter 2) as well as other host genetic factors (e.g. Chapter 4). By examining HIV evolution at an individual, population, and global level I hope to draw attention to the complex interplay of host and viral genetic factors and how we may leverage this knowledge to inform biomedical intervention strategies.

This thesis is organized into five chapters. This first chapter provides a brief overview of the molecular epidemiology, genetics, life cycle and pathogenesis of HIV-1 and outlines the objectives of this thesis. Chapter 2 provides an in depth review of HIV-1 adaptation to host immune responses. Chapter 3 provides a brief overview of HIV-1 receptors and coreceptors. Chapter 4 is an original research chapter that addresses the thesis' primary aim: to use phylogenetic and molecular techniques to characterize the transmission and evolution of a CXCR4-using HIV strain in a pair of individuals where one member of the pair, the recipient, possesses two copies of a rare human genetic mutation that renders them resistant to CCR5-using HIV strains. Lastly, chapter 5 provides a brief conclusion and discusses implications of the work presented here. This thesis is prepared according to a manuscript-based format. Chapter 2 is modified from a review published in SpringerLink Encyclopedia of AIDS and chapter 4 will be submitted to an international peer-reviewed journal.

1.7. References

1. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, et al. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220: 868-871.

2. Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, et al. (1983) Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science 220: 865-867.

3. WHO (2014) HIV/AIDS: Global situation and trends. World Health Organization.

4. UNAIDS (2013) Global report: UNAIDS report on the global AIDS epidemic 2013.

5. Centers for Disease C (1981) Pneumocystis pneumonia--Los Angeles. MMWR Morb Mortal Wkly Rep 30: 250-252.

6. Centers for Disease C (1981) Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. MMWR Morb Mortal Wkly Rep 30: 305-308.

7. Centers for Disease C (1982) Update on acquired immune deficiency syndrome (AIDS)--United States. MMWR Morb Mortal Wkly Rep 31: 507-508, 513-504.

8. Hymes KB, Cheung T, Greene JB, Prose NS, Marcus A, et al. (1981) Kaposi's sarcoma in homosexual men-a report of eight cases. Lancet 2: 598-600.

9. Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, et al. (1981) Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N Engl J Med 305: 1425-1431.

10. Richman DD, editor (2003) Human Virus Guides 2: Human Immunodeficiency Virus. 2 ed. London: International Medical Press.

11. Tebit DM, Arts EJ (2011) Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. Lancet Infect Dis 11: 45-56.

12. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, et al. (2000) HIV-1 nomenclature proposal. Science 288: 55-56.

13. Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, et al. (2009) A new human immunodeficiency virus derived from gorillas. Nat Med 15: 871-872.

14. D'Arc M, Ayouba A, Esteban A, Learn GH, Boue V, et al. (2015) Origin of the HIV-1 group O epidemic in western lowland gorillas. Proc Natl Acad Sci U S A 112: E1343-1352.

15. Peeters M, Courgnaud V, Abela B, Auzel P, Pourrut X, et al. (2002) Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. Emerg Infect Dis 8: 451-457.

16. Peeters M, Courgnaud, V. (2002) Overview of Primate Lentiviruses and Their Evolution in Non-human Primates in Africa.

17. Peeters M, D'Arc M, Delaporte E (2014) Origin and diversity of human retroviruses. AIDS Rev 16: 23-34.

18. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, et al. (1999) Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature 397: 436-441.

19. Gottlieb GS, Eholie SP, Nkengasong JN, Jallow S, Rowland-Jones S, et al. (2008) A call for randomized controlled trials of antiretroviral therapy for HIV-2 infection in West Africa. AIDS 22: 2069-2072; discussion 2073-2064.

20. Pessoa R, Carneiro Proietti AB, Busch MP, Sanabani SS (2014) Identification of a Novel HIV-1 Circulating Recombinant Form (CRF72_BF1) in Deep Sequencing Data from Blood Donors in Southeastern Brazil. Genome Announc 2.

21. LANL H (2015) HIV Circulating Recombinant Forms (CRFs).

22. Carr JK, Salminen MO, Albert J, Sanders-Buell E, Gotte D, et al. (1998) Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. Virology 247: 22-31.

23. Murphy E, Korber B, Georges-Courbot MC, You B, Pinter A, et al. (1993) Diversity of V3 region sequences of human immunodeficiency viruses type 1 from the central African Republic. AIDS Res Hum Retroviruses 9: 997-1006.

24. Osmanov S, Pattou C, Walker N, Schwardlander B, Esparza J, et al. (2002) Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. J Acquir Immune Defic Syndr 29: 184-190.

25. Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, et al. (1984) The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature 312: 763-767.

26. Klatzmann D, Champagne E, Chamaret S, Gruest J, Guetard D, et al. (1984) T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature 312: 767-768.

27. Barroca P, Calado M, Azevedo-Pereira JM (2014) HIV/dendritic cell interaction: consequences in the pathogenesis of HIV infection. AIDS Rev 16: 223-235.

28. (1992) 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. MMWR Recomm Rep 41: 1-19.

29. Deeks SG, Lewin SR, Havlir DV (2013) The end of AIDS: HIV infection as a chronic disease. Lancet 382: 1525-1533.

30. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, et al. (2011) Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 365: 493-505.

31. Perelson AS, Essunger P, Cao Y, Vesanen M, Hurley A, et al. (1997) Decay characteristics of HIV-1-infected compartments during combination therapy. Nature 387: 188-191.

32. Connor EM, Sperling RS, Gelber R, Kiselev P, Scott G, et al. (1994) Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. N Engl J Med 331: 1173-1180.

33. Cardo DM, Culver DH, Ciesielski CA, Srivastava PU, Marcus R, et al. (1997) A case-control study of HIV seroconversion in health care workers after percutaneous exposure. Centers for Disease Control and Prevention Needlestick Surveillance Group. N Engl J Med 337: 1485-1490.

34. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, et al. (2010) Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 363: 2587-2599.

35. Wilen CB, Tilton JC, Doms RW (2012) Molecular mechanisms of HIV entry. Adv Exp Med Biol 726: 223-242.

36. Wilen CB, Tilton JC, Doms RW (2012) HIV: cell binding and entry. Cold Spring Harb Perspect Med 2.

37. De Jong JJ, De Ronde A, Keulen W, Tersmette M, Goudsmit J (1992) Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. J Virol 66: 6777-6780.

38. Berger EA, Murphy PM, Farber JM (1999) Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol 17: 657-700.

39. Melikyan GB (2014) HIV entry: a game of hide-and-fuse? Curr Opin Virol 4: 1-7.

Chapter 2.

HIV-1 mutational escape from host immunity

This chapter is a modified encyclopaedia entry in the Encyclopedia of AIDS and published as:

Le AQ, Shahid, A., Brumme, Z. L. (2014) HIV-1 Mutational Escape from Host Immunity. In: Hope TJ, Stevenson, M., Richman, D., editor. Encyclopedia of AIDS: Springer New York. pp. 1-19.

2.1. Introduction

Within an infected individual, HIV-1 develops specific mutations within its genome that allow it to escape detection by host immune responses. As such, host immunity represents a major selective force driving the evolution and diversification of HIV-1 at the individual and population levels. Here, we highlight HIV-1 mutational escape from adaptive, innate and vaccine-induced immune responses as highly specific and reproducible processes beginning rapidly following HIV-1 infection. The potential biological implications of immune escape, including viral fitness costs and population-level HIV-1 adaptation to host immunity, are also summarized.

2.2. Escape from CD8+ cytotoxic T-lymphocytes

CD8+ cytotoxic T-Lymphocytes (CTL) eliminate HIV-infected cells via the recognition of short, virus-derived peptide epitopes that are produced within the infected cell and presented at its surface by the highly polymorphic Human Leukocyte Antigen (HLA) class I molecules (HLA-A, B, and C) (Figure 2.1a). HLA-restricted CTL play a

major role in HIV-1 immune control. HIV-specific CTL first appear around acute-phase viremia decline and play an active role in its control to setpoint levels [1]. Experimental depletion of CD8+ T-cells in rhesus macaques results in their inability to control Simian Immunodeficiency Virus (SIV) infection [2]. Epidemiological links between host carriage of specific HLA class I alleles and HIV-1 disease progression have been demonstrated in natural history [3] and genome-wide association [4] studies. In particular, HLA-B*57 and B*27 are associated with lower viral loads and slower disease progression [3]. Independent effects of HLA-C expression level on HIV-1 control have also been demonstrated [5].

That HLA-restricted CTL exert pressure on HIV-1 *in vivo* is also demonstrated by the virus' ability to escape this pressure via mutation. CTL escape was first described in 1991 when researchers noted temporal shifts (and in some cases permanent loss of recognition) of HLA-B*08-restricted HIV-1 Gag epitopes targeted by patient-derived CTL over time, which coincided with the appearance of viral mutations within them [6]. Another key concept revealed by this study is the HLA-restricted nature of CTL escape, due to the requirement that epitopes be bound and presented by a specific HLA molecule for CTL recognition.

CTL escape mutations can be classified into three mechanistic categories. The most intuitive is escape via mutation(s) that reduce or abrogate viral epitope binding to HLA, thereby impairing CTL recognition of infected cells (Figure 2.1b). Such mutations usually occur at HLA-specific epitope "anchor" residues - typically peptide positions two and/or C-terminus. A well-known example is the B*27-associated R264K substitution selected at position 2 of the B*27-restricted KK10 epitope in Gag [7]. Escape via abrogation of peptide-HLA binding represents a predominant CTL escape mechanism *in vivo*, with escape conferring an average (predicted) ten-fold reduction in peptide-HLA binding affinity [8]. CTL escape can also act upon processes that occur prior to, or following, peptide-HLA binding. For example, CTL escape mutations can inhibit epitope formation by interfering with their proper intracellular processing. The first such "antigen processing escape mutation" to be mechanistically characterized was the B*57:03-restricted Gag-A146P substitution, occurring at the residue immediately upstream of the IW9 epitope, which acts via prevention of N-terminal aminopeptidase-mediated trimming

of this epitope [9]. Antigen processing mutations can also occur within the epitope. For example, a substitution at position 5 of a B*07-restricted epitope in a cryptic Gag reading frame acted via introduction of a proteasomal cleavage site, yielding reduced epitope formation [10]. The final category of “T-cell receptor (TCR) escape mutations” retain the capability to bind HLA, but reduce or abrogate recognition of the peptide-HLA complex by the TCR(s) expressed by the original selecting CTL(s). TCR escape mutations usually occur at central epitope positions. An example is the B*27-associated L268M substitution (selected at position 6 of the KK10 epitope) [6]. L268M-containing KK10 retains the ability to bind HLA-B*27, but abrogates its recognition by key B*27-restricted CTL clonotypes in the repertoire [11].

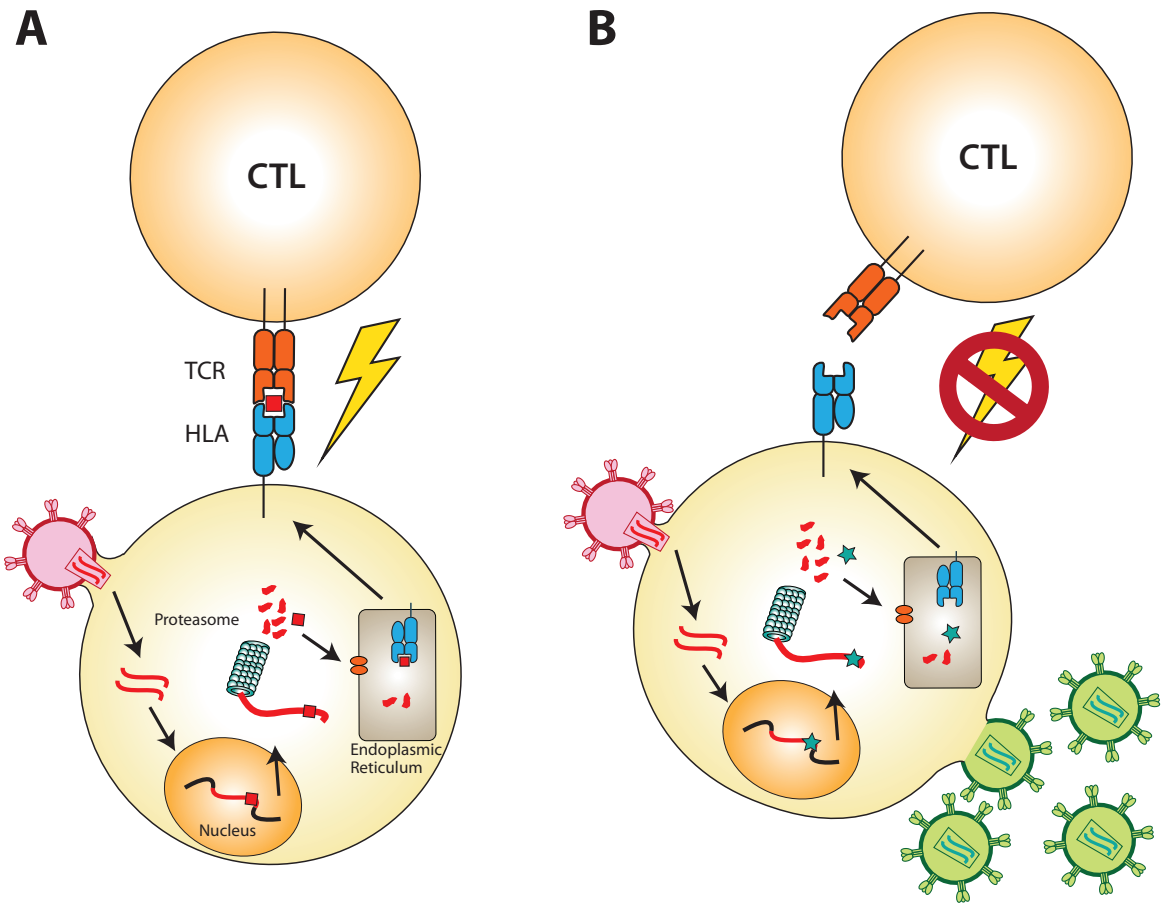

Figure 2.1. Escape from cytotoxic T-lymphocytes
***Panel A*:** HIV-1 proteins produced within an infected cell are processed into peptide epitopes by host cellular machinery and loaded onto HLA class I molecules for presentation at the cell surface. CTL eliminate infected cells via recognition of the viral peptide-HLA complex via their T-Cell Receptor (TCR). ***Panel B*:** HIV-1 mutations can arise during the process of reverse transcription, that are then translated into protein. Here, a mutation abrogates the ability of the encoded viral epitope to bind HLA, allowing the infected cell to avoid CTL-mediated killing. Mutant progeny viruses are then released. CTL escape mutations may also hinder CTL-mediated killing of infected cells by interfering with viral antigen processing or by abrogating TCR-mediated recognition of the viral peptide-HLA complex *(not shown).*

Despite substantial HIV-1 and host genetic variation, the mutational pathways of CTL escape are broadly predictable based on the HLA class I alleles expressed by the host. For example, three-quarters of HIV-1 subtype B infected persons expressing the protective HLA-B*57 allele will select the T242N substitution in Gag (position three of the $p24^{Gag}$ TW10 epitope) within weeks or months following infection [12], while fifty percent will also select G248A at position 9 of this epitope later on [8]. Together, these two mutations confer complete escape from B*57-restricted, TW10-specific CTL [12]. In

contrast, among HLA-B*27-expressing persons, targeting of the immunodominant B*27-restricted $p24^{Gag}$ KK10 epitope begins in early infection and is often sustained for years thereafter [13]. KK10 escape begins via selection of the L268M mutation at epitope position 6 that abrogates its recognition by certain autologous B*27-restricted CTL [11], but complete escape from KK10-expressing CTL, via the R264K anchor residue escape at position 2 of the epitope, does not usually occur until years later [7]. Notably, KK10 escape remains one of the few clear-cut examples where escape directly precedes loss of HIV-1 immune control [14].

The predictable nature of CTL escape has allowed the identification of HLA-associated viral polymorphisms by statistical association. These studies, undertaken in cross-sectional datasets of linked HIV-1 and host HLA genotypes, identify viral polymorphisms significantly over- (or under-) represented among persons expressing a given HLA class I allele, identifying these as likely escape mutations (and their associated immunologically susceptible forms), respectively. The first such study, published in 2002, identified nearly 100 HLA-associated polymorphisms in HIV-1 reverse transcriptase in a cohort of ~400 patients, illustrating the extensive impact of CTL pressures on HIV-1 [15]. In recognition of the potential confounding effects of viral lineage (or "founder") effects in such analyses, more recent studies incorporate "phylogenetic corrections" [16], as well as statistical corrections for the confounding effects of linkage disequilibrium between HLA class I alleles and HIV-1 amino acid co-variation.

Population-level studies have yielded comprehensive "immune escape maps" of the locations and mutational pathways of HLA-restricted CTL escape in HIV-1. These maps are most detailed for HIV-1 subtype B (e.g. [8]) and C (e.g. [17]). Population-level studies have also confirmed escape (and reversion, discussed later) as highly reproducible processes in context of host HLA. For example, the strongest HLA association in subtype B is the HLA-A*24:02-restricted Y135F escape mutation in Nef, where 81% of A*24:02-expressing persons harbor this substitution in chronic infection, compared to only 12% of persons who do not express an allele belonging to the A24 supertype [8]. Such a strong statistical association (in this case, an odds ratio of ~30 and

a p-value of $8x10^{-118}$ [8]) can only be achieved if the mutation is near-universally selected in persons harboring the HLA, and reverts consistently in individuals lacking it [18].

Escape is also highly HLA-specific. When population-level analyses are undertaken at various HLA resolution levels (*e.g.* supertype, type, subtype), the majority (>60%) of HLA-associated polymorphisms are identified as HLA subtype-specific while <10% are identified as shared across HLA supertypes [8]. This high HLA-specificity remains true even for closely related HLA alleles that present the same viral epitopes. For example, HLA-B*57:02, B*57:03 and B*58:01 all bind Gag-TW10, but they drive significantly different escape pathways within it [19]. Escape pathways can also be complex and varied. Escape at a given viral site may occur along multiple pathways under pressure by a given HLA – for example, B*08-driven escape at Nef codon 94, position 5 of the B*08-restricted FL8 epitope, can occur via K94E, M, N or Q [20]. A given HIV-1 site may be under selection by various HLA alleles that select different, sometimes opposing, substitutions. For example at Gag codon 147, HLA-A*25:01, B*13:02 and B*57:01 escape via selection of “L”, while B*14:02 and B*15:01 escape via selection of “I” (which also happens to be the subtype B consensus at this site) [8]. Identification of HLA-associated polymorphisms has also aided the discovery of novel CTL epitopes, including those in cryptic HIV-1 reading frames [21] .

2.3. HLA class II-driven immune escape

Effective antiviral immunity generally requires CD4+ T-lymphocyte help, but CD4+ responses rapidly become dysfunctional in HIV-1 infection, in part because of the specific elimination of virus-specific CD4+ T-cells [22]. As such, the contribution of CD4+ T-cells to HIV-1 control *in vivo* remains incompletely understood. Though some early studies supported the possibility of *in vivo* mutational escape from HIV-specific CD4+ T-cell responses [23], others did not [24]. Attempts to identify HLA class II-restricted viral polymorphisms by statistical association have yielded no strong evidence of their existence [25].

2.4. Escape from humoral (B-cell) immune responses

HIV-1 envelope glycoprotein evolves rapidly within a host after infection, and has diversified to an extraordinary extent at the population level [26]. Although CTL escape contributes to this process, the autologous neutralizing antibody (NAb) response is the major driver of envelope evolution. Beginning at approximately three months post-infection [27], HIV-infected individuals begin to develop antibodies capable of neutralizing their own virus (termed “autologous neutralizing antibodies”; NAbs) [28]. However, unlike acute-phase HIV-specific CTL, autologous NAbs do not appreciably contribute to virus containment, likely due to the rapid selection and outgrowth of neutralization-resistant escape mutants [27]. Initial NAb escape exposes novel envelope epitopes against which subsequent waves of autologous NAbs arise, driving further envelope glycoprotein evolution.

That antibodies and virus co-evolve in cycles of response and escape was first inferred via the ability of autologous sera to neutralize viral variants present in the infected individual 6 (or 12) months prior, but not those present at the time of serum sampling [29]. Early studies of HIV-1 neutralization escape hinted at a variety of escape pathways including the accumulation of amino acid changes in envelope [29] (suggestive of escape through the selection of specific point mutations), changes in N-linked glycosylation patterns [27] (Figure 2.2) and lengthening of certain hypervariable domains in gp120, notably V1/V2 [30]. However, the identification of specific genetic events conferring NAb escape began only recently (*e.g.* the first specific identification of an envelope glycoprotein escape mutation conferring neutralization escape at the single antibody level was not achieved until 2009 [31]).

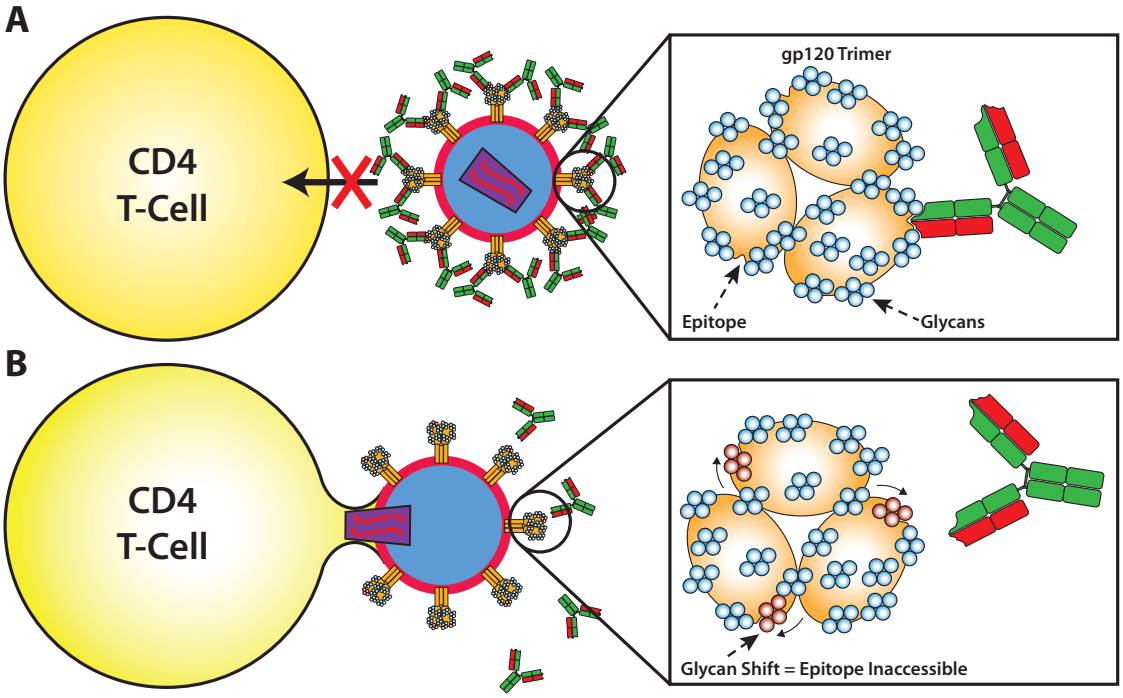

Figure 2.2. Neutralizing antibody escape .
***Panel A*:** Neutralizing antibodies (NAb) bind to epitopes (shown as an indent) on the viral envelope glycoprotein , blocking their ability to infect target cells. ***Panel B*:** Mutations in HIV-1 envelope – in this case leading to changes in N-linked glycosylation patterns that block NAb access to the epitope, confer NAb escape. NAb escape may also occur via point mutations, lengthening of certain gp120 hypervariable domains (notably V1/V2), cooperative interactions between different regions on a single or multiple members of the envelope trimer, or other mechanisms (*not shown*).

Unlike CTL epitopes whose (linear) sequences can be predicted from HLA anchor residue motifs in HIV-1 sequences without knowledge of the T-cell receptor sequence or structure, antibodies directly recognize three-dimensional epitopes whose sequences can span discontinuous sites on one or more members of the envelope trimer, rendering their locations difficult to predict based on viral sequence alone. Recent studies have therefore taken the approach of longitudinally characterizing envelope evolution while simultaneously attempting to isolate individual neutralizing antibodies (and/or the B-cell clonal lineages producing them; *e.g.* [32]) in individual patients.

From these studies, a central role of immune-driven envelope evolution in driving autologous neutralization breadth is emerging. In one individual, initial autologous NAbs were directed against epitopes in the first and second hypervariable loops of gp120

(V1/V2), and escape was achieved via point mutations in this region including one in V2 that created a putative N-linked glycosylation site conferring escape from two distinct monoclonal antibodies isolated from this patient [31]. In a second individual, escape from the initial NAb pool occurred via convergent evolutionary pathways (one involving changes in the V3-V5 gp120 outer domain and the other involving co-dependent changes in V1/V2 and gp41), whose lineage members subsequently oscillated in frequency [31]. NAb escape via distinct evolutionary pathways within a single host was confirmed in an individual in whom escape in a V3-proximal epitope occurred along three divergent viral lineages, each featuring a unique amino acid change [33]. A study of three acutely-infected individuals whose initial response was directed against different conformational epitopes in envelope, where each escaped along distinct pathways [34], also supports the strain- and host-specific nature of initial epitope targeting and autologous neutralization escape. That escape occurs via distinct mechanisms (*e.g.* point mutations, glycan shifts, and co-operative conformational changes between two domains) both within and among hosts indicates that HIV-1 employs multiple mutational strategies to escape early autologous NAbs [31]. However, the extent to which NAb epitopes - and their escape pathways - are shared across patients remains a key question. The observation that, compared to transmitted/founder viruses, chronic subtype C viruses are significantly enriched for a glycan at envelope codon 332 (whose presence can help trigger the evolution of broadly neutralizing antibodies against this key conserved region [32]), supports the idea of shared neutralization escape pathways.

In approximately 80% of infected individuals, this process of virus-NAb coevolution results in the continued production of NAbs that remain largely specific to the individual's evolving virus. However in approximately 20% of individuals, this process leads to the emergence of antibodies that are capable of neutralizing a broad range of HIV-1 isolates across subtypes. Though individuals producing such "broadly neutralizing antibodies" do not likely derive clinical benefit from them (presumably because their own virus has already escaped) [35], the evolutionary mechanisms driving their development are of paramount interest as an effective preventative HIV-1 vaccine will likely require their elicitation (along with effective cellular responses). This discovery has led to the hypothesis that this process could be recapitulated via vaccination with specific

transmitted/founder envelopes and their sequential escape variants [36], a strategy for which there is preliminary experimental support [37].

Non-neutralizing HIV-specific antibodies that mediate antibody-dependent cellular cytotoxicity (ADCC) through activation of effector cells bearing Fc receptors, notably Natural Killer (NK) cells, may also contribute to natural- and vaccine-induced HIV-1 immune control [38]. Some evidence also supports ADCC antibodies, including those that do not possess neutralizing activity, as drivers of evolution within HIV-1 envelope and possibly other viral regions [39].

2.5. Escape from innate immune responses: KIR-driven HIV-1 polymorphisms?

Innate immune responses, in particular Natural Killer (NK) cells, may also directly drive immune escape. NK cells express cell-surface receptors belonging to the polymorphic Killer cell Immunoglobulin-like Receptor (KIR) gene family, which comprise a variety of inhibitory and activating receptors that interact with HLA class I ligands on target cells. Engagement of activating KIR delivers a stimulatory signal, while engagement of inhibitory KIR delivers a tolerance signal. When the former overcome the latter, NK effector functions are initiated. Indeed, a major trigger for enhanced NK cell-mediated recognition of HIV-infected cells is the selective downregulation of their HLA-A and -B (though not C) ligands by the viral Nef protein [40], leading to a reduction in signalling through inhibitory KIR. Inhibitory KIR bind their HLA class I ligands in an allotype-specific manner. For example, KIR3DL1 receptors interact with HLA-B molecules belonging to the Bw4 allotype (determined by amino acids 77-83 of the HLA coding region), notably those harboring isoleucine at position 80 (Bw4-80I), and to a lesser extent those harboring threonine at this position (Bw4-80T) [41]. Some activating KIR also recognize HLA class I in an allotype-specific manner, though generally at lower avidity than their inhibitory counterparts [42].

KIR, alone and in combination with their allotype-specific HLA ligands, may modulate HIV-1 susceptibility and pathogenesis. HIV-infected individuals expressing the activating KIR3DS1 allele in combination with HLA-Bw4-80I exhibit lower viral loads [43]

and delayed clinical progression [44]. Higher frequencies of KIR3DS1 homozygosity [45] have been observed in HIV-1 exposed seronegative individuals, suggesting that activating KIR may also confer some level of protection against HIV-1 acquisition. Though protection via engagement of an activating KIR seems intuitive, the underlying mechanism remains unknown (KIR3DS1-expressing NK cells can suppress HIV-1 replication in Bw4-80I–expressing cells *in vitro* [46], but direct binding of KIR3DS1 to HLA-Bw4-80I has not been shown). KIR3DL1 alleles possessing a high-expression, high-inhibitory phenotype (termed KIR3DL1*h/*y) may also be protective. When present in combination with HLA-Bw4-80I alleles, notably HLA-B*57, KIR3DL1*h/*y alleles were associated with lower viral loads and conferred protection against HIV-1 disease progression [47]. KIR3DL1*h/*y-HLA-B*57 co-expression may also protect against HIV-1 acquisition [48]. That highly inhibitory KIR receptor-ligand interactions can be protective seems counterintuitive, especially given that the opposing signals of activating KIR may also be protective. Nevertheless, the data support a role, albeit incompletely elucidated, of KIR in HIV-1 control.

KIR-associated immune pressures may also drive the selection of viral polymorphisms that allow infected cells to evade NK-mediated killing. To shed light on how such mutations could arise, we must first briefly re-visit KIR-ligand binding. Though not antigen-specific in the classical sense, KIR receptor-ligand interactions are nevertheless modulated in part by HLA polymorphism (through their allotype-specificity) as well as the sequence of the HLA-bound peptide [49]. The idea that naturally-arising HIV-1 variants could affect KIR-HLA binding was supported by reduced *in vitro* binding of KIR3DL1 to its HLA B*57:03 ligand in the presence of the TW10 epitope harboring a G-to-E substitution at position 9 (though this was not claimed to be an *in vivo* NK-driven escape mutation, as failure to engage KIR3DL1 would render infected cells more, not less, susceptible to NK-mediated killing [50]). Rather, NK cell escape could theoretically be achieved via viral polymorphisms that reduce recognition by activating KIR, or enhance recognition by inhibitory KIR. Towards the identification of such mutations, statistical association approaches were applied to N=91 linked KIR/HIV-1 sequences, yielding 22 KIR-associated viral polymorphisms. Two linked polymorphisms in Vpu (71M/71H), located in a region that overlaps the Env reading frame, were overrepresented among KIR2DL2-expressing persons, in particular those KIR2DL2+

individuals homozygous for HLA-C group 1 alleles [51] (consistent with the greater affinity of KIR2DL2 for HLA-C group 1 ligands [52]). Researchers further showed *in vitro* that the presence of these polymorphisms enhanced the ability of the inhibitory KIR2DL2 to bind HIV-infected cells, that KIR2DL2+ NK cells failed to become activated in the presence of polymorphism-containing HIV-1, and that cells infected with polymorphism-containing HIV-1 were not inhibited by KIR2DL2+ NK cells [51]. These findings suggest that immune pressure by an inhibitory KIR could select *in vivo* escape mutations conferring enhanced binding of the inhibitory receptor to HIV-infected cells, thereby allowing them to escape NK cell-mediated elimination (Figure 2.3). The recent identification of an HLA-C*01:02-restricted $p24^{Gag}$ peptide variant that bound KIR2DL2, that conferred functional inhibition of KIR2DL2-expressing NK cells in vitro [53], provides theoretical support for this model.

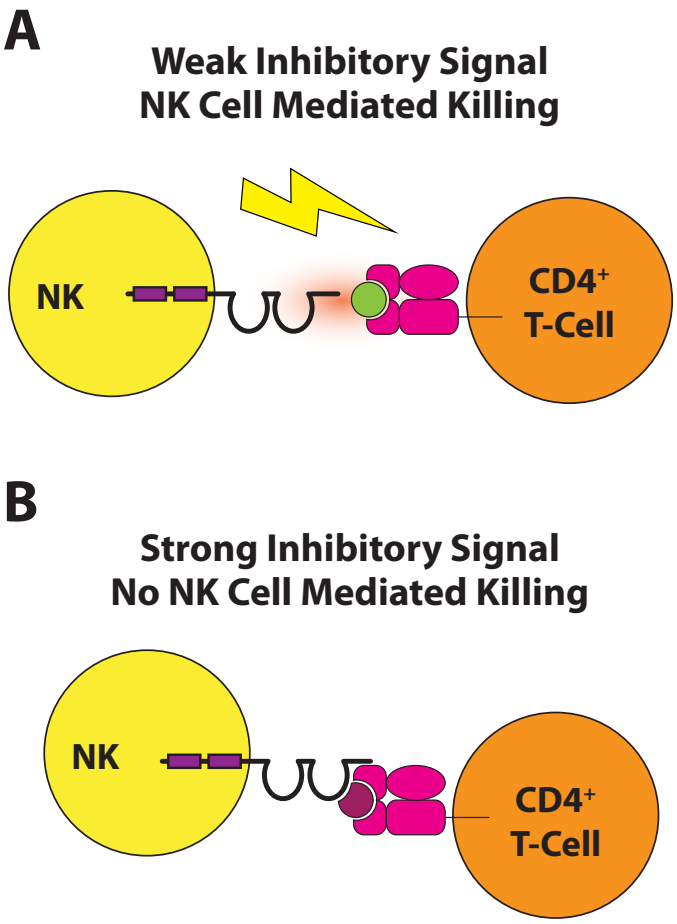

Figure 2.3. Escape from KIR
Mutations in HIV-1 could theoretically impair natural killer (NK) cell-mediated recognition of HIV-infected cells, thereby conferring escape from innate immunity. Based on a putative KIR2DL2 HIV-1 escape mutation described by Alter et al. [51] this figure illustrates how NK cell escape could occur. ***Panel A*****:** weak interactions between the inhibitory NK KIR2DL2 receptor and the viral peptide/HLA-C complex on the HIV-infected cell produce weak inhibitory signals, leading to NK cell mediated elimination of the infected cell. ***Panel B*****:** viral escape mutations that enhance KIR2DL2-mediated NK cell recognition of the peptide/HLA-C complex on the HIV-infected cell enhance binding of the inhibitory KIR2DL2 receptor, thus protecting the infected cell from NK cell mediated elimination. Image modified from [54].

2.6. Escape from vaccine-induced antiviral immunity

A challenge in designing vaccines against genetically diverse pathogens such as HIV-1 is the possibility that vaccine-induced immunity may protect against infection by strains similar to the vaccine immunogen(s), but not genetically divergent strains (Figure 2.4). Such “sieve effects” can be identified by retrospectively comparing HIV-1

sequences of vaccine vs. placebo trial participants who subsequently became infected (termed “breakthrough” sequences), to determine genetic differences between them [55]. The idea that vaccine-induced immunity could induce a partial barrier through which antigenically divergent HIV-1 strains could penetrate has been termed the “acquisition sieve effect” (Figure 2.4b), while the related - yet mechanistically distinct - possibility that vaccine-induced immunity would fail to block infection but would instead drive the rapid outgrowth of escape variants has been termed “postinfection sieve effect” [56] (Figure 2.4c). The latter is particularly relevant to CTL-based vaccines, as these are unlikely to block HIV-1 transmission. Notably, acquisition and post-infection sieve effects can be difficult to distinguish from one another, as both may occur before HIV-1 RNA can be reliably detected in the blood, and/or may manifest themselves via the presence of identical immune-associated polymorphisms.

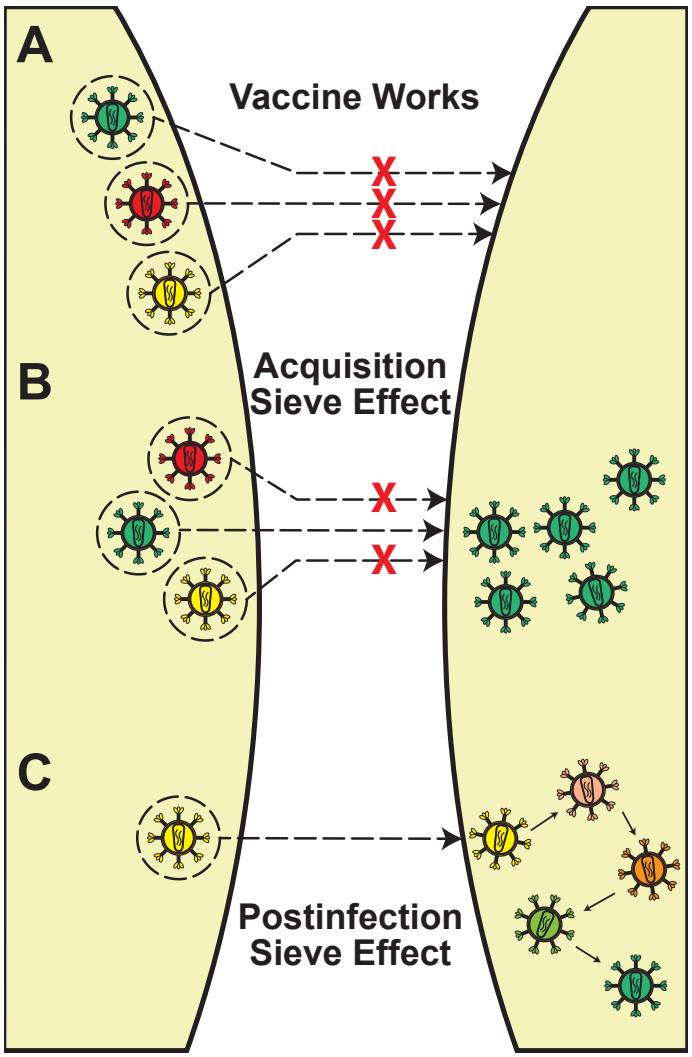

Figure 2.4. Escape from vaccines
Sieve effects demonstrate immunogenicity in HIV-1 vaccine trials. ***Panel A***: a sterilizing HIV-1 vaccine induces immune responses that block infection by any/all incoming HIV-1 strains. ***Panel B***: vaccination induces a partial barrier that blocks infection by HIV-1 strains similar to vaccine immunogen, but not those that are antigenically divergent from it (acquisition sieve effect). ***Panel C***: vaccination focuses immune responses on epitopes shared between founder virus and vaccine immunogen, leading to rapid in vivo escape (postinfection sieve effect). Although mechanistically distinct, acquisition and postinfection sieve effects are difficult to distinguish via the analysis of breakthrough HIV-1 sequences. Image modified from [54].

Analysis of vaccine trial data support sieve effects in HIV-1. Such effects were first suggested by the presence of atypical V3 amino acid motifs in HIV-1 *env* sequences from individuals vaccinated with recombinant HIV-1_{MN} gp120 [57]. Recent comparisons of founder HIV-1 strains from vaccine- and placebo recipients of the RV144 "Thai" vaccine trial identified differential amino acid frequencies at *env* V2 codons 169 and 181

between the two groups [58], suggesting that the vaccine preferentially blocked viruses harboring specific substitutions at these positions. Rapid selection of CTL escape mutations by vaccine-induced cellular immune responses may also have occurred in the failed STEP vaccine trial [59]. Inferred T-cell epitope sequences within Gag/Pol/Nef (the regions contained within the vaccine) from infected vaccine recipients exhibited greater genetic distances to the immunogen sequence compared to those of infected placebo recipients, presumably as a result of extensive and rapid immune escape [60]. The lack of such differences for epitopes within other HIV-1 proteins also supported this conclusion. HIV-1 sequences from vaccine recipients also exhibited substitutions at Gag codon 84 more frequently than placebo recipients, identifying this as a putative signature site of HIV-1 evolution in response to vaccine-induced CTL responses [60].

The implications of vaccine-induced immune responses on HIV-1 evolution are potentially profound. At the individual level, rapid vaccine-driven escape could accelerate disease progression [61], while the use of vaccines capable of blocking infection by only certain HIV-1 strains raises concerns regarding potential shifts in viral polymorphism frequencies and/or HIV-1 lineage distributions (and their clinical and pathogenic consequences) at the population level. That vaccine-induced immune responses (notably CTL) may target slightly different epitopes than those in natural infection [62] may further complicate this issue.

2.7. A note on the role of HIV-1 accessory proteins in immune evasion

Although beyond the scope of the present essay, certain HIV-1 proteins possess immune evasion functions that deserve brief mention. In particular, downregulation of HLA-A and -B (but not HLA-C) from the cell-surface by Nef [40] represents a major mechanism of immune evasion by HLA-restricted CTL, as it reduces their ability to recognize infected cells. HIV-1 Nef, Vpu and envelope also serve to remove CD4 from the infected cell surface [63]. The recent observation that interaction of HIV-1 envelope with CD4 on the infected cell surface is required to expose certain ADCC epitopes suggests that cell-surface CD4 downregulation could represent an immune evasion strategy to reduce ADCC-mediated elimination of infected cells [64].

2.8. Immune escape dynamics in early infection

Escape begins rapidly following HIV-1 infection. Recently however, detailed studies of intra-host HIV-1 evolution using single-genome amplification (*e.g.* [65]) or next-generation sequencing (*e.g.* [66]) have advanced our understanding of HIV-1 transmission and escape dynamics. HIV-1 transmission is characterized by a severe genetic bottleneck. An estimated 80% of heterosexual transmissions are productively initiated by a single transmitted/founder virus [65], while infection in persons who inject drugs is generally established by more than one closely related founder viruses [67]. In the days following infection, the transmitted/founder virus(es) undergo rapid population growth and star-like diversification [68], giving rise to a "quasispecies" swarm of related HIV-1 variants. This genetic pool becomes the evolutionary substrate upon which host immune responses exert pressure, driving the selection of escape mutations and the survival of viral lineages harboring them. The first CTL escape mutations appear during acute-phase viremia decline [1]; the selection (and in some cases fixation) of CTL escape variants has been observed as early as 21 days post-infection in humans [68]. Selection (and subsequent fixation) of the first NAb escape mutations also occurs relatively rapidly, though on a slightly longer timecourse than CTL escape [34].

The evolutionary pathways along which these early mutations arise have recently been elucidated in detail. The conceptually straightforward pathway whereby the first selected escape mutation gradually outcompetes the original transmitted form is likely to be true for only a minority of cases [1]. More commonly, the first escape variant tends to be rapidly followed by the emergence of numerous others, from which the "final" escape form is ultimately selected [1]. This is likely because the initially-appearing pool of low frequency mutants often retains some ability to be targeted by existing (or *de novo*) CTL [50]. This drives the selection of more effective escape variants, often at HLA-anchor residues, that ultimately outcompete both transmitted founder and initial variants. For example, in a B*57:03-expressing individual, initial escape within the $p24^{Gag}$ TW10 epitope occurred approximately 5 months post-infection via a transient, minority G-to-E mutation at position 9 (G248E) that retained the ability to bind B*57:03 and reduced CTL recognition only modestly [50]. By approximately 1.5 years post-infection, this mutation was outcompeted by variants expressing the canonical B*57-restricted G248A mutation

at this position (along with T242N and V247I at epitope positions 3 and 8). Similarly, multiple amino acids often transiently appear in the regions under NAb pressure, from which the final neutralization mutant(s) ultimately emerge [34]. Escape continues to occur (albeit at a slower rate [69]) over the infection course.

2.9. Immune escape as a major driver of HIV-1 diversity

Immune escape is a major driver of HIV-1 diversity within individuals and populations. CTL escape accounts for a major proportion of within-host HIV-1 evolution in the first year of infection. For example, a study of seven newly-infected individuals revealed that, six months post-infection, between 9 to 18 positively-selected substitutions were observed throughout the HIV-1 proteome [68]. Another population-based study estimated that a minimum of 30% of substitutions in Gag/Pol and 60% in Nef were attributable to HLA pressures [70]. Escape is also widespread throughout the HIV-1 proteome. A recent statistical association study identified over 2100 HLA-associated polymorphisms at ~35% of HIV-1's nonconserved codons [8], distributed somewhat unequally throughout the proteome. For example, Vpu exhibited evidence for HLA-mediated selection at one-quarter of its nonconserved sites, compared to ~70% of nonconserved sites in Nef [8]. The status of HLA as the most important host genetic factor influencing HIV-1 diversity was recently confirmed via genome wide association [71]. Similarly, HIV-1 Gag and Nef sites under HLA selection have diversified to the greatest extent over the past three decades of the North American epidemic, supporting a significant role of HLA in driving global HIV-1 diversification [72].

2.10. Fitness consequences of escape

Upon transmission, some immune escape mutations selected in the previous host will revert to the original (usually subtype consensus) amino acid [12]. While some CTL escape mutations, for example the B*57-associated Gag T242N, revert consistently and rapidly following transmission [12], most revert more slowly [73] and others rarely or not at all [74]. Reversion occurs because these mutations incur a fitness cost. Generally, escape mutations within conserved viral regions tend to be more fitness-costly, while

escape in more variable regions tends to be fitness-neutral [75]. An example of a fitness-costly mutation is the B*27-associated R264K substitution in the $p24^{Gag}$ KK10 epitope, which essentially abolishes *in vitro* viral replication when engineered alone into HIV-1_{NL4-3} [76]. Generally though, *in vitro* fitness costs of escape mutations observed *in vivo* tend to be subtler, often requiring multiple substitutions to reduce function. Alone, the B*57-driven Gag-T242N mutation reduces viral replicative capacity only modestly [77], but dose-dependent replicative reductions are observed when it is present alongside other common B*57-driven mutations in $p24^{Gag}$ [78]. Other examples of fitness-costly CTL escape mutations include B*13-associated mutations in Gag [79], Cw*05-driven mutations in integrase [80] and B*35-associated mutations in Nef [81]. Fitness costs ranging from 0% to 24% have been observed for early envelope escape mutants, indicating that NAb escape can also be fitness-costly [34]. Fitness costs of escape can be offset by the selection of compensatory mutations at secondary sites. Whereas most compensatory mutations occur in relatively close proximity to the primary escape site (e.g. S165N with A163G in B*5703-KF11 [73] in $p24^{Gag}$), others, such as S173A with R264K in B*27-KK10 [76], occur a substantial linear distance away, but may reside nearby in the folded protein structure.

In the case where escape can only occur at a functional and/or replicative cost, the virus' advantage gained via immune escape is offset in part by these costs, thus potentially conferring some residual biological benefit to the host in terms of lower viral loads. For example, the sustained protective effect of HLA-B*81 is believed to be due in part to selection of the fitness-costly Gag T186S escape mutation at position 7 of the immunodominant B*81-restricted TL9 epitope, which is difficult to compensate [82]. Relative clinical benefits of fitness-costly escape in HLA-mismatched individuals who have acquired HIV-1 with key Gag escape mutations have also been observed [83]. That HIV-1 sequences contain inherent determinants of pathogenesis is supported by the observation that set-point plasma viral load is to a certain extent "heritable" from one infection to the next [84], and that viral replication capacity correlates positively with viral load (and negatively with CD4+ T-cell count) at various infection stages (e.g. [85]). Indeed, acquisition of attenuated HIV-1, followed by further within-host selection of noncanonical fitness-costly escape mutations is likely to explain a portion of HIV-1 elite control [86], a rare phenotype where individuals are able to spontaneously suppress

plasma HIV-1 RNA to below limits of clinical detection without the need for antiretroviral therapy.

These observations have led to the idea that immune-mediated containment of HIV-1 replication to levels that slow disease progression and possibly reduce transmission might be achievable through the design of vaccines that stimulate CTL responses focused against critically conserved viral regions where escape can only occur at substantial fitness costs [87]. A related strategy would be to design immunogens featuring both susceptible and common escape variant forms - provided the latter retain the ability to bind the relevant HLA molecules - with the goal of generating broad, potent, variant-reactive CTL responses that, upon infection, will drive HIV-1 evolution down unconventional pathways not unlike those selected in elite controllers [88].

2.11. Population-level adaptation of HIV-1 to host immune pressures

As HIV-1 genomes residing in an individual exhibit adaptations to its host's immunogenetic profile, then HIV-1 sequences circulating in a given host population exhibit adaptations that reflect the distinct immunogenetic profile of that population (Figure 2.5). This is often referred to as "population-level" adaptation of HIV-1. For example, >50% of HLA-associated polymorphisms identified in HIV-1 subtype B sequences in Mexico [89] and nearly two-thirds of those identified in Japan [90] are distinct from those observed in subtype B-infected cohorts from Canada/USA/Australia, because the former populations exhibit HLA alleles unique to those populations (*e.g.* B*39 in Mexico and B*67:01 in Japan). The frequencies of HLA-associated polymorphisms will similarly vary according to the frequencies of their restricting HLA alleles in the population. The B*51-associated I135X mutation in Reverse Transcriptase (at the C-terminus of the B*51-TI8 epitope, RT codons 128-135) provides an example. In an analysis of 9 cohorts spanning 5 continents, HLA-B*51 and RT-I135X prevalence exhibited a strong positive correlation [91], indicating that the more frequent an HLA allele is in a population, the more frequent its associated adaptations will be observed in circulating HIV-1 sequences. Other host immune factors (*e.g.* variability in T-cell receptor

genetics) may also play a role in population-specific HIV-1 adaptation. For example, a recent comparative study of HIV-1 subtype B cohorts in Japan versus Canada/USA/Australia identified numerous cases where the same HLA allele selected significantly different escape pathways across cohorts [90], implying factors beyond HLA in driving these differences. HLA-driven escape pathways also differ across HIV-1 subtypes, presumably as a result of genetic differences in the viral backbone. For example, Gag-T242N is commonly selected by B*57 in HIV-1 subtypes B, C and D, but rarely in subtype A1 [92].

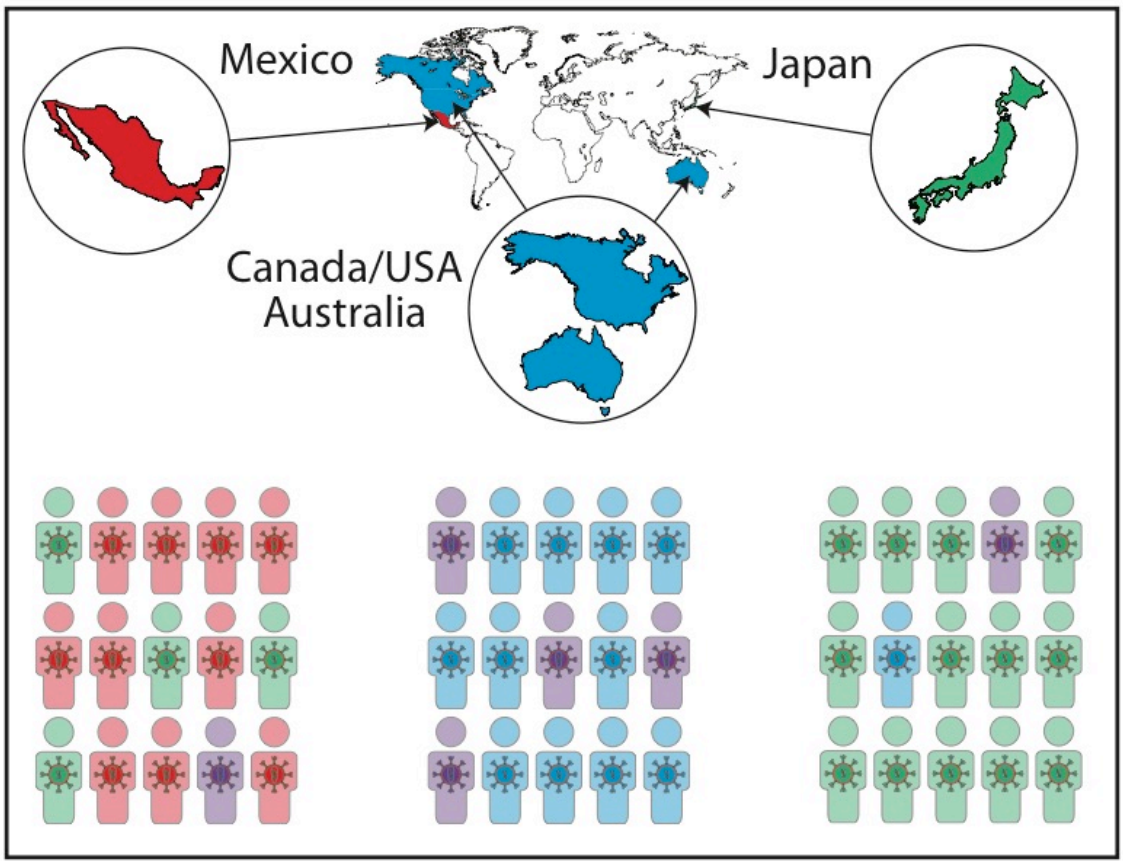

Figure 2.5. Population level escape
HLA is a major driver of global HIV-1 diversity. HIV-1 genomes in an infected individual will harbor adaptations specific to the HLA alleles expressed by that individual (denoted by matching colors of virus and host). By extension, HIV-1 sequences in a host population will harbor adaptations to HLA alleles expressed in that population. Host populations in Mexico, Canada/USA/Australia, and Japan exhibit very different HLA class I allele distributions (denoted by different host colors); as such, adaptations exhibited by HIV-1 sequences will also differ between these populations (denoted by different virus colors). HLA is therefore a major driver of global HIV-1 diversity. Note that, because reversion is neither instantaneous nor universal, escape mutations for particular HLA alleles will be found in some proportion of individuals lacking those alleles (not shown). Image modified from [54].

The persistence of certain immune escape mutations following transmission to a new host has led to a related concern – namely, that escape mutations could gradually spread throughout the population [93]. Analogous to the negative impact of transmitted drug resistance mutations on treatment efficacy, acquisition of "escape mutant" HIV-1 by persons expressing the relevant HLA could undermine the ability of their CTL to control infection; as such, escape mutant spread could gradually undermine host antiviral immune potential (and potentially diminish the protective effects of certain HLA alleles as the epidemic progresses). Indeed, the S173A compensatory mutation has been shown to stabilize the B*27-associated R264K mutation in $p24^{Gag}$ upon transmission [94], while the S165N compensatory mutation has been shown to stabilize B*57-associated mutations within the $p24^{Gag}$ KF11 epitope [73], supporting this concern.

The extent to which immune escape mutations are spreading in HIV-infected populations remains incompletely known, in part due to the scarcity of historic data. Nevertheless, it has been suggested that CTL epitopes in European HIV-1 sequences are being "lost" through mutational escape from HLA-B mediated selective pressures [95]; similarly, higher viral polymorphism frequencies have been reported in modern compared to historic HIV-1 subtype B and F sequences in South America [96]. The high frequency of the B*51-associated HIV-1 Reverse Transcriptase (RT) I135X mutation in Japan, a population where B*51 prevalence approaches 20%, is also suggestive of escape mutation accumulation [91] (though the possibility that the Japanese epidemic was founded by an HIV-1 sequence containing RT-I135X cannot be ruled out). A comparative study of historic (1979-1989) versus modern (2000+) HIV-1 subtype B cohorts in North America revealed modest spread of CTL escape mutations over the study period which occurred alongside an approximate twofold increase in HIV-1 diversity during this time [72]. Despite limited evidence of escape mutation spread in North America, rates of spread may be higher in populations with high HIV-1 prevalence, older epidemics, differential transmission dynamics and/or where host HLA diversity is relatively limited, and thus possess more immediate implications.

The gradual accumulation of CTL escape mutations in circulating HIV-1 sequences is paralleled by a similar phenomenon driven by humoral immunity. Two recent studies evaluating antibody neutralization resistance of historic versus modern

HIV-1 envelope sequences suggest that HIV-1 is drifting towards a more neutralization-resistant phenotype over time [97]. Furthermore, contemporary sera exhibited lower heterologous neutralizing activity than historic sera, consistent with a gradual undermining of humoral immunity as HIV-1 becomes increasingly neutralization resistant [98]. Taken together, evidence suggests that HIV-1 is becoming - albeit gradually - more "pre-adapted" to host immunity as immune escape mutations spread in circulation. Further studies are therefore warranted to explore the extent of HIV-1 adaptation to cellular and humoral immune pressures in different host populations as their respective epidemics increase in age and diversity, and the potential implications of this adaptation for natural (and vaccine-induced) immunity over time.

2.12. Conclusion

In conclusion, mutational escape from host immune responses represents a major selective force driving the evolution and diversification of HIV-1 within infected persons. By extension, mutational escape is also responsible for the diversification of HIV-1 globally, and for the continued evolution of the virus as the epidemic progresses. Continuing to advance our understanding of the dynamics and pathogenic implications of immune escape within individuals and populations – including how to recapitulate this process by vaccination as in the case of the generation of broadly-neutralizing antibodies – will be paramount to achieving our ultimate goal of an effective HIV-1 vaccine.

2.13. References

1. Goonetilleke N, Liu MK, Salazar-Gonzalez JF, Ferrari G, Giorgi E, et al. (2009) The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. J Exp Med 206: 1253-1272.

2. Matano T, Shibata R, Siemon C, Connors M, Lane HC, et al. (1998) Administration of an anti-CD8 monoclonal antibody interferes with the clearance of chimeric simian/human immunodeficiency virus during primary infections of rhesus macaques. J Virol 72: 164-169.

3. Carrington M, O'Brien SJ (2003) The influence of HLA genotype on AIDS. Annu Rev Med 54: 535-551.

4. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, et al. (2007) A whole-genome association study of major determinants for host control of HIV-1. Science 317: 944-947.

5. Apps R, Qi Y, Carlson JM, Chen H, Gao X, et al. (2013) Influence of HLA-C expression level on HIV control. Science 340: 87-91.

6. Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, et al. (1991) Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. Nature 354: 453-459.

7. Kelleher AD, Long C, Holmes EC, Allen RL, Wilson J, et al. (2001) Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. J Exp Med 193: 375-386.

8. Carlson JM, Brumme CJ, Martin E, Listgarten J, Brockman MA, et al. (2012) Correlates of protective cellular immunity revealed by analysis of population-level immune escape pathways in HIV-1. J Virol 86: 13202-13216.

9. Draenert R, Le Gall S, Pfafferott KJ, Leslie AJ, Chetty P, et al. (2004) Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. J Exp Med 199: 905-915.

10. Cardinaud S, Consiglieri G, Bouziat R, Urrutia A, Graff-Dubois S, et al. (2011) CTL escape mediated by proteasomal destruction of an HIV-1 cryptic epitope. PLoS Pathog 7: e1002049.

11. Iglesias MC, Almeida JR, Fastenackels S, van Bockel DJ, Hashimoto M, et al. (2011) Escape from highly effective public CD8+ T-cell clonotypes by HIV. Blood 118: 2138-2149.

12. Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, et al. (2004) HIV evolution: CTL escape mutation and reversion after transmission. Nat Med 10: 282-289.

13. Gao X, Bashirova A, Iversen AK, Phair J, Goedert JJ, et al. (2005) AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. Nat Med 11: 1290-1292.

14. Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, et al. (1997) Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nat Med 3: 212-217.

15. Moore CB, John M, James IR, Christiansen FT, Witt CS, et al. (2002) Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. Science 296: 1439-1443.

16. Bhattacharya T, Daniels M, Heckerman D, Foley B, Frahm N, et al. (2007) Founder effects in the assessment of HIV polymorphisms and HLA allele associations. Science 315: 1583-1586.

17. Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, et al. (2008) Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. PLoS Comput Biol 4: e1000225.

18. Fryer HR, Frater J, Duda A, Palmer D, Phillips RE, et al. (2012) Cytotoxic T-lymphocyte escape mutations identified by HLA association favor those which escape and revert rapidly. J Virol 86: 8568-8580.

19. Carlson JM, Listgarten J, Pfeifer N, Tan V, Kadie C, et al. (2012) Widespread Impact of HLA Restriction on Immune Control and Escape Pathways of HIV-1. J Virol 86: 5230-5243.

20. Brumme ZL, Brumme CJ, Heckerman D, Korber BT, Daniels M, et al. (2007) Evidence of Differential HLA Class I-Mediated Viral Evolution in Functional and Accessory/Regulatory Genes of HIV-1. PLoS Pathog 3: e94.

21. Berger CT, Carlson JM, Brumme CJ, Hartman KL, Brumme ZL, et al. (2010) Viral adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting epitopes in HIV frameshift sequences. J Exp Med 207: 61-75, S61-12.

22. Douek DC, Brenchley JM, Betts MR, Ambrozak DR, Hill BJ, et al. (2002) HIV preferentially infects HIV-specific CD4+ T cells. Nature 417: 95-98.

23. Harcourt GC, Garrard S, Davenport MP, Edwards A, Phillips RE (1998) HIV-1 variation diminishes CD4 T lymphocyte recognition. J Exp Med 188: 1785-1793.

24. Koeppe JR, Campbell TB, Rapaport EL, Wilson CC (2006) HIV-1-specific CD4+ T-cell responses are not associated with significant viral epitope variation in persons with persistent plasma viremia. J Acquir Immune Defic Syndr 41: 140-148.

25. Wright JK, Brumme ZL, Julg B, van der Stok M, Mncube Z, et al. (2012) Lack of association between HLA class II alleles and in vitro replication capacities of recombinant viruses encoding HIV-1 subtype C Gag-protease from chronically infected individuals. J Virol 86: 1273-1276.

26. Gaschen B, Taylor J, Yusim K, Foley B, Gao F, et al. (2002) Diversity considerations in HIV-1 vaccine selection. Science 296: 2354-2360.

27. Wei X, Decker JM, Wang S, Hui H, Kappes JC, et al. (2003) Antibody neutralization and escape by HIV-1. Nature 422: 307-312.

28. Richman DD, Wrin T, Little SJ, Petropoulos CJ (2003) Rapid evolution of the neutralizing antibody response to HIV type 1 infection. Proc Natl Acad Sci U S A 100: 4144-4149.

29. Frost SD, Wrin T, Smith DM, Kosakovsky Pond SL, Liu Y, et al. (2005) Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. Proc Natl Acad Sci U S A 102: 18514-18519.

30. Rong R, Gnanakaran S, Decker JM, Bibollet-Ruche F, Taylor J, et al. (2007) Unique mutational patterns in the envelope alpha 2 amphipathic helix and acquisition of length in gp120 hypervariable domains are associated with resistance to autologous neutralization of subtype C human immunodeficiency virus type 1. J Virol 81: 5658-5668.

31. Rong R, Li B, Lynch RM, Haaland RE, Murphy MK, et al. (2009) Escape from autologous neutralizing antibodies in acute/early subtype C HIV-1 infection requires multiple pathways. PLoS Pathog 5: e1000594.

32. Moore PL, Gray ES, Wibmer CK, Bhiman JN, Nonyane M, et al. (2012) Evolution of an HIV glycan-dependent broadly neutralizing antibody epitope through immune escape. Nat Med 18: 1688-1692.

33. Murphy MK, Yue L, Pan R, Boliar S, Sethi A, et al. (2013) Viral escape from neutralizing antibodies in early subtype A HIV-1 infection drives an increase in autologous neutralization breadth. PLoS Pathog 9: e1003173.

34. Bar KJ, Tsao CY, Iyer SS, Decker JM, Yang Y, et al. (2012) Early low-titer neutralizing antibodies impede HIV-1 replication and select for virus escape. PLoS Pathog 8: e1002721.

35. Euler Z, van Gils MJ, Bunnik EM, Phung P, Schweighardt B, et al. (2010) Cross-reactive neutralizing humoral immunity does not protect from HIV type 1 disease progression. J Infect Dis 201: 1045-1053.

36. Liao HX, Lynch R, Zhou T, Gao F, Alam SM, et al. (2013) Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature 496: 469-476.

37. Malherbe DC, Doria-Rose NA, Misher L, Beckett T, Puryear WB, et al. (2011) Sequential immunization with a subtype B HIV-1 envelope quasispecies partially mimics the in vivo development of neutralizing antibodies. J Virol 85: 5262-5274.

38. Wren L, Kent SJ (2011) HIV Vaccine efficacy trial: glimmers of hope and the potential role of antibody-dependent cellular cytotoxicity. Hum Vaccin 7: 466-473.

39. Chung AW, Isitman G, Navis M, Kramski M, Center RJ, et al. (2011) Immune escape from HIV-specific antibody-dependent cellular cytotoxicity (ADCC) pressure. Proc Natl Acad Sci U S A 108: 7505-7510.

40. Cohen GB, Gandhi RT, Davis DM, Mandelboim O, Chen BK, et al. (1999) The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. Immunity 10: 661-671.

41. Cella M, Longo A, Ferrara GB, Strominger JL, Colonna M (1994) NK3-specific natural killer cells are selectively inhibited by Bw4-positive HLA alleles with isoleucine 80. J Exp Med 180: 1235-1242.

42. Bashirova AA, Thomas R, Carrington M (2011) HLA/KIR restraint of HIV: surviving the fittest. Annu Rev Immunol 29: 295-317.

43. Qi Y, Martin MP, Gao X, Jacobson L, Goedert JJ, et al. (2006) KIR/HLA pleiotropism: protection against both HIV and opportunistic infections. PLoS Pathog 2: e79.

44. Martin MP, Gao X, Lee JH, Nelson GW, Detels R, et al. (2002) Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. Nat Genet 31: 429-434.

45. Boulet S, Sharafi S, Simic N, Bruneau J, Routy JP, et al. (2008) Increased proportion of KIR3DS1 homozygotes in HIV-exposed uninfected individuals. AIDS 22: 595-599.

46. Alter G, Martin MP, Teigen N, Carr WH, Suscovich TJ, et al. (2007) Differential natural killer cell-mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. J Exp Med 204: 3027-3036.

47. Martin MP, Qi Y, Gao X, Yamada E, Martin JN, et al. (2007) Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. Nat Genet 39: 733-740.

48. Boulet S, Kleyman M, Kim JY, Kamya P, Sharafi S, et al. (2008) A combined genotype of KIR3DL1 high expressing alleles and HLA-B*57 is associated with a reduced risk of HIV infection. AIDS 22: 1487-1491.

49. Malnati MS, Peruzzi M, Parker KC, Biddison WE, Ciccone E, et al. (1995) Peptide specificity in the recognition of MHC class I by natural killer cell clones. Science 267: 1016-1018.

50. Brackenridge S, Evans EJ, Toebes M, Goonetilleke N, Liu MK, et al. (2011) An early HIV mutation within an HLA-B*57-restricted T cell epitope abrogates binding to the killer inhibitory receptor 3DL1. J Virol 85: 5415-5422.

51. Alter G, Heckerman D, Schneidewind A, Fadda L, Kadie CM, et al. (2011) HIV-1 adaptation to NK-cell-mediated immune pressure. Nature 476: 96-100.

52. Moesta AK, Norman PJ, Yawata M, Yawata N, Gleimer M, et al. (2008) Synergistic polymorphism at two positions distal to the ligand-binding site makes KIR2DL2 a stronger receptor for HLA-C than KIR2DL3. J Immunol 180: 3969-3979.

53. Fadda L, Korner C, Kumar S, van Teijlingen NH, Piechocka-Trocha A, et al. (2012) HLA-Cw*0102-restricted HIV-1 p24 epitope variants can modulate the binding of the inhibitory KIR2DL2 receptor and primary NK cell function. PLoS Pathog 8: e1002805.

54. Carlson JM, Le AQ, Shahid A, Brumme ZL (2015) HIV-1 adaptation to HLA: a window into virus-host immune interactions. Trends Microbiol 23: 212-224.

55. Gilbert P, Self S, Rao M, Naficy A, Clemens J (2001) Sieve analysis: methods for assessing from vaccine trial data how vaccine efficacy varies with genotypic and phenotypic pathogen variation. J Clin Epidemiol 54: 68-85.

56. Edlefsen PT, Gilbert PB, Rolland M (2013) Sieve analysis in HIV-1 vaccine efficacy trials. Curr Opin HIV AIDS 8: 432-436.

57. Berman PW, Gray AM, Wrin T, Vennari JC, Eastman DJ, et al. (1997) Genetic and immunologic characterization of viruses infecting MN-rgp120-vaccinated volunteers. J Infect Dis 176: 384-397.

58. Rolland M, Edlefsen PT, Larsen BB, Tovanabutra S, Sanders-Buell E, et al. (2012) Increased HIV-1 vaccine efficacy against viruses with genetic signatures in Env V2. Nature 490: 417-420.

59. Buchbinder SP, Mehrotra DV, Duerr A, Fitzgerald DW, Mogg R, et al. (2008) Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. Lancet 372: 1881-1893.

60. Rolland M, Tovanabutra S, Decamp AC, Frahm N, Gilbert PB, et al. (2011) Genetic impact of vaccination on breakthrough HIV-1 sequences from the STEP trial. Nat Med 17: 366-371.

61. Betts MR, Exley B, Price DA, Bansal A, Camacho ZT, et al. (2005) Characterization of functional and phenotypic changes in anti-Gag vaccine-induced T cell responses and their role in protection after HIV-1 infection. Proc Natl Acad Sci U S A 102: 4512-4517.

62. Hertz T, Ahmed H, Friedrich DP, Casimiro DR, Self SG, et al. (2013) HIV-1 vaccine-induced T-cell responses cluster in epitope hotspots that differ from those induced in natural infection with HIV-1. PLoS Pathog 9: e1003404.

63. Chen BK, Gandhi RT, Baltimore D (1996) CD4 down-modulation during infection of human T cells with human immunodeficiency virus type 1 involves independent activities of vpu, env, and nef. J Virol 70: 6044-6053.

64. Veillette M, Desormeaux A, Medjahed H, Gharsallah NE, Coutu M, et al. (2014) Interaction with Cellular CD4 Exposes HIV-1 Envelope Epitopes Targeted by Antibody-Dependent Cell-Mediated Cytotoxicity. J Virol 88: 2633-2644.

65. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, et al. (2009) Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J Exp Med 206: 1273-1289.

66. Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, et al. (2012) Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. PLoS Pathog 8: e1002529.

67. Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, et al. (2010) Wide variation in the multiplicity of HIV-1 infection among injection drug users. J Virol 84: 6241-6247.

68. Herbeck JT, Rolland M, Liu Y, McLaughlin S, McNevin J, et al. (2011) Demographic processes affect HIV-1 evolution in primary infection before the onset of selective processes. J Virol.

69. Koibuchi T, Allen TM, Lichterfeld M, Mui SK, O'Sullivan KM, et al. (2005) Limited sequence evolution within persistently targeted CD8 epitopes in chronic human immunodeficiency virus type 1 infection. J Virol 79: 8171-8181.

70. Brumme ZL, Brumme CJ, Carlson J, Streeck H, John M, et al. (2008) Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. J Virol 82: 9216-9227.

71. Bartha I, Carlson JM, Brumme CJ, McLaren PJ, Brumme ZL, et al. (2013) A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. Elife 2: e01123.

72. Cotton LA, Kuang XT, Le AQ, Carlson JM, Chan B, et al. (2014) Genotypic and functional impact of HIV-1 adaptation to its host population during the North American epidemic. PLoS Genet 10: e1004295.

73. Crawford H, Prado JG, Leslie A, Hue S, Honeyborne I, et al. (2007) Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. J Virol 81: 8346-8351.

74. Leslie A, Kavanagh D, Honeyborne I, Pfafferott K, Edwards C, et al. (2005) Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. J Exp Med 201: 891-902.

75. Troyer RM, McNevin J, Liu Y, Zhang SC, Krizan RW, et al. (2009) Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response. PLoS Pathog 5: e1000365.

76. Schneidewind A, Brockman MA, Yang R, Adam RI, Li B, et al. (2007) Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. J Virol 81: 12382-12393.

77. Brockman MA, Schneidewind A, Lahaie M, Schmidt A, Miura T, et al. (2007) Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A. J Virol 81: 12608-12618.

78. Crawford H, Lumm W, Leslie A, Schaefer M, Boeras D, et al. (2009) Evolution of HLA-B*5703 HIV-1 escape mutations in HLA-B*5703-positive individuals and their transmission recipients. J Exp Med 206: 909-921.

79. Prado JG, Honeyborne I, Brierley I, Puertas MC, Martinez-Picado J, et al. (2009) Functional consequences of human immunodeficiency virus escape from an HLA-B*13-restricted CD8+ T-cell epitope in p1 Gag protein. J Virol 83: 1018-1025.

80. Brockman MA, Chopera DR, Olvera A, Brumme CJ, Sela J, et al. (2012) Uncommon pathways of immune escape attenuate HIV-1 integrase replication capacity. J Virol.

81. Ueno T, Motozono C, Dohki S, Mwimanzi P, Rauch S, et al. (2008) CTL-mediated selective pressure influences dynamic evolution and pathogenic functions of HIV-1 Nef. J Immunol 180: 1107-1116.

82. Wright JK, Naidoo VL, Brumme ZL, Prince JL, Claiborne DT, et al. (2012) Impact of HLA-B*81-Associated Mutations in HIV-1 Gag on Viral Replication Capacity. J Virol 86: 3193-3199.

83. Chopera DR, Woodman Z, Mlisana K, Mlotshwa M, Martin DP, et al. (2008) Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. PLoS Pathog 4: e1000033.

84. Alizon S, von Wyl V, Stadler T, Kouyos RD, Yerly S, et al. (2010) Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. PLoS Pathog 6.

85. Prince JL, Claiborne DT, Carlson JM, Schaefer M, Yu T, et al. (2012) Role of transmitted Gag CTL polymorphisms in defining replicative capacity and early HIV-1 pathogenesis. PLoS Pathog 8: e1003041.

86. Lobritz MA, Lassen KG, Arts EJ (2011) HIV-1 replicative fitness in elite controllers. Curr Opin HIV AIDS.

87. Altfeld M, Allen TM (2006) Hitting HIV where it hurts: an alternative approach to HIV vaccine design. Trends Immunol 27: 504-510.

88. Miura T, Brockman MA, Schneidewind A, Lobritz M, Pereyra F, et al. (2009) HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphotye recognition. J Virol 83: 2743-2755.

89. Avila-Rios S, Ormsby CE, Carlson JM, Valenzuela-Ponce H, Blanco-Heredia J, et al. (2009) Unique features of HLA-mediated HIV evolution in a Mexican cohort: a comparative study. Retrovirology 6: 72.

90. Chikata T, Carlson JM, Tamura Y, Borghan MA, Naruto T, et al. (2014) Host-specific adaptation of HIV-1 subtype B in the Japanese population. J Virol 88: 4764-4775.

91. Kawashima Y, Pfafferott K, Frater J, Matthews P, Payne R, et al. (2009) Adaptation of HIV-1 to human leukocyte antigen class I. Nature 458: 641-645.

92. McKinnon LR, Capina R, Peters H, Mendoza M, Kimani J, et al. (2009) Clade-specific evolution mediated by HLA-B*57/5801 in human immunodeficiency virus type 1 clade A1 p24. J Virol 83: 12636-12642.

93. Goulder PJ, Brander C, Tang Y, Tremblay C, Colbert RA, et al. (2001) Evolution and transmission of stable CTL escape mutations in HIV infection. Nature 412: 334-338.

94. Schneidewind A, Brumme ZL, Brumme CJ, Power KA, Reyor LL, et al. (2009) Transmission and long-term stability of compensated CD8 escape mutations. J Virol 83: 3993-3997.

95. Schellens IM, Navis M, van Deutekom HW, Boeser-Nunnink B, Berkhout B, et al. (2011) Loss of HIV-1-derived cytotoxic T lymphocyte epitopes restricted by protective HLA-B alleles during the HIV-1 epidemic. AIDS 25: 1691-1700.

96. Dilernia DA, Jones L, Rodriguez S, Turk G, Rubio AE, et al. (2008) HLA-driven convergence of HIV-1 viral subtypes B and F toward the adaptation to immune responses in human populations. PLoS ONE 3: e3429.

97. Bunnik EM, Euler Z, Welkers MR, Boeser-Nunnink BD, Grijsen ML, et al. (2010) Adaptation of HIV-1 envelope gp120 to humoral immunity at a population level. Nat Med 16: 995-997.

98. Bouvin-Pley M, Morgand M, Moreau A, Jestin P, Simonnet C, et al. (2013) Evidence for a Continuous Drift of the HIV-1 Species towards Higher Resistance to Neutralizing Antibodies over the Course of the Epidemic. PLoS Pathog 9: e1003477.

Chapter 3.

HIV receptors and coreceptors: a mini-review

3.1. Introduction

Identification of the CD4 receptor in 1984 [1] and the CXCR4 and CCR5 coreceptors in 1996 [2,3] were instrumental in furthering our understanding of HIV-1 transmission, evolution, and disease progression [4]. Moreover, the identification of viral coreceptors led to development of coreceptor antagonists, a class of drugs that inhibit viral entry via coreceptor blockade (e.g. the CCR5 antagonist Maraviroc) [5,6].

3.2. Identification of CD4 receptor and CXCR4 and CCR5 coreceptors

The discovery of CD4 as the primary HIV-1 receptor came hand in hand with the observation that HIV-1 infected individuals had decreased levels of CD4+ T-cells [7]. Specifically, in 1984 Klatzmann et al demonstrated that CD4 was the primary receptor for HIV-1 entry [1]. Using monoclonal antibodies that blocked only CD4, they showed that HIV-1 infection of immortalized T cell lines could be inhibited [1]. However, further research indicated that CD4 was not the sole receptor required for HIV-1 entry [8].

Two key observations led researchers to hypothesize that CD4 was not the only HIV-1 receptor necessary for viral entry into host cells. Firstly, researchers working with recombinant human CD4 molecules observed that HIV-1 was only able to use this receptor when it was expressed on human, but not murine cells [4]. Further experiments with cell lines concluded that non-permissiveness to HIV-1 infection in murine cells was not due to presence of an inhibitor in these cells, but rather to the lack of an unidentified cofactor necessary to complete the viral life cycle [4]. Secondly, there was evidence of

viral tropism in human cells: some HIV-1 strains were able to infect certain human cell types but not others [4]. This suggested that there may be more than one coreceptor involved in viral entry. With further experiments, researchers identified two HIV-1 phenotypes: some HIV-1 strains were able to efficiently infect primary T-cells while others were able to infect primary macrophages; thus these strains were named T-cell tropic (TCL-tropic) and macrophage tropic (M-tropic) strains, respectively. It was not until 1996 that the first HIV-1 coreceptor, CXCR4, was identified. Using a cDNA cloning strategy and sequence analysis, Feng et al identified a protein that was part of the seven transmembrane G-coupled protein receptor superfamily [2]. Although other laboratories had previously sequenced this protein, no ligand had yet been identified. As such, researchers initially named it “fusin” for its role in mediating HIV-1 fusion and entry [2], but its name was later changed to CXCR4. Further confirmation that CXCR4 served as an HIV-1 entry coreceptor came from subsequent functional studies that blocked CXCR4 or added CXCR4 to CD4-expressing cell types while exposing them to HIV-1 infection [4]. In addition it was noted that only TCL-tropic strains were able to infect CXCR4 expressing cells, indicating that M-tropic strains use another coreceptor.

The discovery of another HIV-1 coreceptor, CCR5, followed shortly in 1996. Early studies of host factors that mediate HIV-1 disease progression revealed that the host chemokines RANTES, MIP-1α, and MIP-1β secreted by CD8+ T-cells were major HIV-1 suppressive factors [9]. Furthermore, these chemokines suppressed M-tropic, but not T-tropic, strains of HIV-1 [3,4]. The receptor for these chemokines had previously been characterized as CCR5, another member of the G-coupled protein receptor family [3,9]. Again, confirmation that CCR5 acted as an HIV-1 coreceptor was done using CCR5-blocking antibodies. With the identification of the coreceptors for M- and TCL-tropic viruses, the picture was finally clear that HIV-1 entry required binding of CD4 and either CCR5 or CXCR4. Note that this historic nomenclature of TCL- and M-tropic is no longer used because M-tropic strains can also infect T-cells (and possibly vice-versa) – this is because CCR5 and CXCR4 are expressed on a variety of cell types. Instead, the modern nomenclature is based on the coreceptor usage of the viral isolate. For example, strains that enter by binding to CCR5 are termed CCR5-using (or “R5”), those that use CXCR4 are termed CXCR4-using (or “X4”), and those that are able to use either coreceptor are termed dual-tropic (or “R5X4”). Note that other HIV-1 coreceptors have

been identified *in vitro*, such as CCR2b, CCR3, and CX_3CR1, however their role in natural HIV-1 infection is poorly understood [4].

3.3. Determination of viral coreceptor use

HIV-1 coreceptor usage is relevant to HIV-1 pathogenesis. For example, the emergence of X4 viruses in an infected, untreated individual is associated with accelerated CD4 decline and faster progression to AIDS [10,11], whether X4 viruses are a cause or consequence of disease progression remains somewhat unclear. HIV-1 coreceptor usage is also relevant in the context of HIV-1 therapy. Specifically, the antiretroviral drug maraviroc is a CCR5 antagonist that binds to the host CCR5 receptor thereby blocking entry of R5 strains but not X4 or dual-tropic ones. Because maraviroc only protects against R5 variants, prior genetic testing of HIV-1 in the infected individual is vital to rule out the presence of X4 or dual-tropic strains. As such, determining the coreceptor usage of a given HIV-1 isolate is important in both a research as well as a clinical context. There are two primary methods to determine coreceptor use: phenotypic and genotypic, with slight variations of each in currently in use.

3.3.1. Phenotypic assays

Phenotypic co-receptor usage assays are cell culture based and can determine whether an HIV-1 strain is R5, X4, or dual-tropic. The most commonly used phenotypic assay is the Trofile assay (Monogram Biosciences, South San Francisco, California, USA) (Figure 3.1) [12,13]. Coreceptor usage is determined by amplifying env-gp160 sequences from the plasma HIV-1 RNA of infected patients using a nested reverse transcriptase polymerase chain reaction (RT-PCR). Amplicons are then cloned into an expression vector and amplified using bacterial cloning methods. A second plasmid, that contains a modified version of the HIV-1 reference strain NL4.3 genome where envelope-gp160 is replaced with a luciferase gene (such that light is emitted when it is expressed) is also used in the assay. Both plasmids are transfected into an immortalized cell line in order to produce HIV-1 NL4.3-luciferase “pseudoviruses” that express the patient-derived env gp160 on their surface. After harvesting, the pseudoviruses are used to infect specialized cell lines that express only CCR5 or CXCR4, in a single cycle

assay. If the pseudotyped virions are successfully able to infect the target cells they emit a luciferase light signal [12]. Viruses that are able to infect one or the other are R5 or X4 respectively, and viruses able to infect both cell lines are dual-tropic. When this assay is performed using single clonal sequences isolated from patients, it is able to determine if a sequence is truly R5, X4, or dual-tropic. However, if performed at the “bulk” level by amplifying all viruses present in a sample (thus producing a mixed stock of pseudoviruses) the assay is only able to discern if the sample is exclusively R5, X4, or contains a mixture of R5, X4, and/or dual-tropic viruses (“dual/mixed” or “D/M”). Phenotypic assays are considered the gold standard and are widely used in clinical and research applications [13]. However, phenotypic assays are time consuming, resource intensive and have a relatively substantial failure rate, especially on low viral load samples, due to the requirement that gp160 be amplified in its entirety.

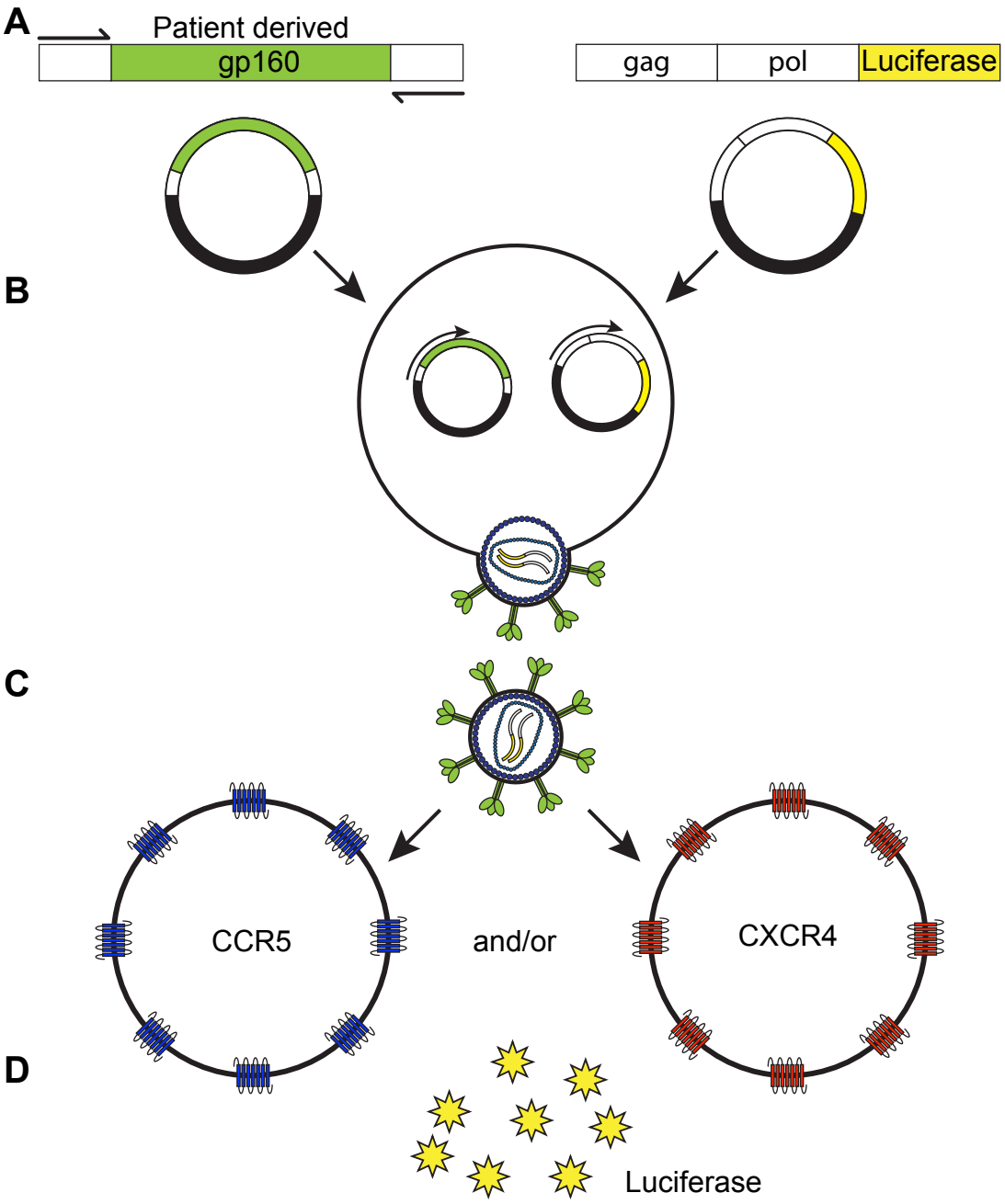

Figure 3.1. Phenotypic tropism assay (Trofile)
The above diagram shows the steps of the phenotypic tropism assay. ***Panel A*****:** (Left) Gp160 is amplified from patient plasma using RT-PCR and cloned into an expression vector. (Right) A modified NL4.3 reference strain where gp160 is replaced with a luciferase gene. ***Panel B*****:** Both plasmids are transfected into an immortalized cell line to produce "pseudoviruses" that carry HIV-1 NL4.3-luciferase genomes and express patient derived gp160 on the surface. ***Panel C*****:** Pseudoviruses infect specialized cell lines that express CD4 and CCR5 or CXCR4 exclusively. ***Panel D*****:** Upon infection the HIV-1 NL4.3-luciferase genome is expressed to produce a luciferase signal.

3.3.2. Genotypic assays

Genotypic assays infer HIV-1 coreceptor phenotype by analysis of the viral envelope sequence. The V3 region in gp120 is the principal genetic determinant for coreceptor binding [14]. Analysis of V3 sequences with linked phenotypic data originally allowed researchers to identify specific positions and amino acids in V3 that are associated with either CCR5 or CXCR4 usage [14,15]. For example, positively charged amino acids (*i.e.* arginine and lysine) at V3 codon 11 and/or 25 (subtype B; HXB2 codon 11 and/or 27) are associated with CXCR4 usage, while neutral or positively charged amino acids at position 11 and/or 25 are associated with CCR5 usage [14,16]. This historic and simple method used to determine coreceptor usage is termed the "11/25 rule". In addition to position 11 and/or 25 researchers have identified other amino acids within and outside of V3 that can be used to distinguish between R5 and X4 HIV-1 (Figure 3.2). Current coreceptor prediction algorithms are more advanced and generally consist of machine learning methods trained on very large linked datasets of genotypic and phenotypic information [17,18]. Some algorithms look at only the V3 region while others incorporate data from additional regions within gp120 [17,18]. Coreceptor prediction algorithms in widespread use include geno2pheno$_{[coreceptor]}$ [17] and WebPSSM [18]. Although the underlying algorithms differ, both are able to classify sequences as R5 or X4. Geno2pheno$_{[coreceptor]}$ can also incorporate other clinical parameters such as viral load and CD4 counts to improve coreceptor predictions [17]. Current genotypic methods paired with next-generation "deep" sequencing have comparable sensitivities and specificities to phenotypic assays [19,20] and are widely used in research and clinical applications [19,20].

Figure 3.2. Gp120 amino acid positions associated with CCR5/CXCR4 usage
Shown is the amino acid alignment of HXB2 envelope-gp120. Amino acid sequences highlighted in red and labeled above the alignment represent the V1-V5 loops. Amino acid sites highlighted in blue represent sites that can be used to distinguish between CCR5 and CXCR4 using strains.

3.4. Coreceptors, infection, and disease progression

The majority of HIV-1 infections worldwide are caused by R5 viruses [4]. R5 strains are also preferentially transmitted [21]. This may be due to their numerical dominance, but also their decreased susceptibility to host antiviral (e.g. defensins [22] and SDF-1 [23]) factors. More recent studies have revealed that 83-95% of primary subtype B infections comprise R5 viruses while the remainder comprise X4 or dual-tropic viruses [24,25]. In addition, approximately 50% of individuals infected with R5 viruses undergo a phenotypic “R5-to-X4” switch during chronic infection. This coreceptor switch is also temporally associated with a more rapid decline in CD4 counts and accelerated disease progression to AIDS [10,11]. Whether X4 HIV-1 is the cause or consequence of accelerated disease progression to AIDS is still unclear.

3.4.1. Genetic variation in the host CCR5 gene contributes to susceptibility to HIV-1 infection and disease progression

A naturally occurring 32 base-pair deletion in the host CCR5 gene (CCR5-Δ32) results in a defective CCR5 protein [26]. Rare individuals with two copies of this gene

are referred to as CCR5-Δ32/Δ32 homozygous. These individuals have no functional CCR5 receptor rendering them resistant to infection by R5 HIV-1 [26]. Although resistant to R5 HIV-1 strains, CCR5-Δ32/Δ32 homozygous individuals are still susceptible to infection by X4 and dual-tropic viruses [26,27]. Individuals with a single copy of the mutation are referred to as CCR5-wt/Δ32 heterozygotes. Heterozygosity does not protect against HIV-1 acquisition [27], however it is associated with slower disease progression [27,28]. Importantly, individuals homozygous or heterozygous for the Δ32 mutation display no detrimental side effects associated with reduced or non-expression of the CCR5 proteins [27]. The CCR5-Δ32 mutation is found at different frequencies throughout the world. In individuals with northern European descent the prevalence of CCR5-wt/Δ32 and CCR5-Δ32/Δ32 is approximately 20% and 1% respectively, but negligible in other ethnic populations [26,29].

There are other naturally-occurring mutations in the CCR5 gene and promoter. The m303 single nucleotide polymorphism (SNP) results in a truncated protein [30] and the CCR5 59029-G/A SNP in the promoter region influences protein expression [31]. Individuals who are homozygous for the CCR5-m303/m303 mutation or who are heterozygous in conjunction with the Δ32 mutation (CCR5-m303/Δ32) are resistant to R5 infection but remain susceptible to X4 and dual-tropic infections [30]. The CCR5 m303 mutation is understudied because it is even rarer than the Δ32 mutation [30]. The CCR5 59029-G/A SNP is however more common (allele frequency of 43-68% depending on ethnicity) [31]. In a historic cohort of untreated HIV-1 infected individuals, those who had the 59029-G/G SNP progressed to AIDS on average 3.8 years slower than individuals who had 59029-A/A SNP [31]. Of note, there are no known natural mutations in the CXCR4 gene that affect HIV-1 acquisition or disease progression. There is however a common polymorphism in SDF1 (the natural ligand for CXCR4) named SDF1-3'UTR-801G-A (abbreviated to SDF1-3'A) that has been associated with a delayed onset to AIDS [32].

3.5. Targeting coreceptors therapeutically

The discovery of natural mutations in the CCR5 gene that render individuals "resistant" to HIV-1 [26-28] has led to various strategies to replicate this phenotype

therapeutically. These include blockade of CCR5 (e.g. the CCR5 antagonist maraviroc) [6], modification of the CCR5 gene (e.g. using zinc finger nucleases) [33], or, in the case of the “Berlin patient” who is believed to be cured of HIV-1, a stem cell transplant from a tissue antigen-matched, CCR5-Δ32/Δ32 donor [34]. A brief discussion of these therapies follows here.

As mentioned, early studies showed that the cytokines RANTES, MIP-1α, and MIP-1β (the natural ligands for CCR5) can function as inhibitors of HIV-1 in vitro [9], though the ability of these of these cytokines to effectively block HIV-1 infection of target cells in vivo remains somewhat controversial. Some studies have shown a correlation between these chemokines and a delayed disease progression while others have not [4]. Nevertheless, these observations support strategies of therapeutic blockade of CCR5 in HIV-1 treatment and prevention. Moreover, the observation that natural mutations identified in CCR5 (i.e. CCR5-Δ32 or CCR5 m303) abrogate protein expression with no adverse effects suggests that CCR5 blockade strategies may be minimally toxic [26-28]. Indeed, coreceptor antagonists are a class of drugs used against HIV-1. Maraviroc is the only FDA approved antiretroviral drug in this class; it blocks entry of R5 HIV-1 strains by competitively binding to CCR5. It does not, however, block X4 or dual-tropic strains [5,6]. As such, a coreceptor usage test must be performed prior to starting Maraviroc to decrease the likelihood of the emergence of X4 strains during treatment [35].

Other attempts to inhibit HIV-1 entry include experimental gene modification of the CCR5 gene using zinc finger nucleases (ZFN) [33]. Here, ZFNs are engineered to specifically recognize the CCR5 gene whereupon the nuclease domain creates a double-stranded break in the CCR5 gene. This break is then repaired by the cellular DNA repair mechanisms. When the double-stranded break is not repaired correctly, a non-functional protein results, thus mimicking the CCR5-Δ32 mutation [33]. Adoptive transfers of ZFN modified CD4+ T-cells to humanized HIV-1 infected mice have been shown to decrease viral load and increase CD4+ T-cell counts [33]. Lastly, a more dramatic example came in 2009 doctors performed a stem-cell transplant from a CCR5-Δ32/Δ32 homozygous donor to the “Berlin patient”, an HIV-1 infected individual with acute myeloid leukemia [34]. To this day this individual has remained HIV-1 negative

while remaining off antiretroviral therapy. However, subsequent CCR5-Δ32/Δ32 homozygous stem-cell transplants to HIV-1 infected individuals have been unsuccessful [36]: in 2014, an individual who had undergone this procedure subsequently had viral rebound while off antiretroviral therapy with X4 HIV-1 [36] .

3.6. Assessing HIV-1 sequence diversity

HIV-1 is a highly genetically diverse pathogen with a rapid mutation rate [37]. Upon infection with one or a few distinct transmitted/founder viruses, HIV-1 undergoes rapid evolution generating a within-host swarm of viral "quasispecies" that are highly similar but genetically distinct [38,39]. These viral variants are important to study because each has their own genotypic and phenotypic traits that may contribute to HIV-1 pathogenesis. I will here give a brief overview of historic and next-generation DNA sequencing technologies and how they have helped us study HIV-1 sequence diversity.

Traditional "bulk" (direct) Sanger sequencing of HIV-1 genomes amplified from an individual via (RT-)PCR yields a single "composite" sequence that captures HIV-1 diversity in that individual. In a bulk sequence, nucleotide mixtures will be observed at positions where variants in the quasispecies differ from one another. Although this is useful for studies that require a single representative sequence per individual, there are two limitations to this approach. Firstly, if a sequence contains more than one site harboring a mixture, it is not possible to determine the exact nucleotide sequence of the individual template sequences in the original pool. Secondly, bulk Sanger sequencing can only reliably detect HIV-1 variants that are above a frequency of 20% in the original pool, [40]. For example, in the context of viral coreceptor usage prediction, Sanger sequencing provides an overall average prediction for all amplified sequences in an individual. Consequently, genotypic X4 prediction of some sequences will be more accurate than others. For example, if sequences contain nucleotide mixtures at crucial coreceptor-determining positions (e.g. V3 codon 11 and/or 25) it will result in less a confident prediction. As an example, if the nucleotide sequence at codon 11 was "A(A/T)A", it would translate to either lysine (which is highly associated with X4 phenotype) or isoleucine (which is associated with R5 phenotype), and the overall coreceptor prediction would be less confident compared to a sequence with no mixtures.

To capture individual unique sequences using Sanger sequencing, templates must first be separated and amplified individually prior to sequencing – either by employing a molecular cloning step or via single genome amplification (SGA) methods, both of which are time consuming procedures [39,41]. It is estimated that the sensitivity of genotypic coreceptor prediction using "bulk" sequences is approximately 40-60% [42,43] of that of phenotypic coreceptor usage assays [17,43]; this increases to 70-80% when using clonal sequence data.

Next-generation sequencing (NGS) technologies have significantly advanced HIV-1 research. NGS has allowed researchers to capture the vast diversity of HIV-1 at an individual sequence level within a single person. It has also allowed researchers to study a larger number of sequences [44] in greater detail including minority variants below the limit of detection by Sanger sequencing [45]. NGS is a term used to describe a variety of high throughput technologies that are able to sequence a large number of sequences simultaneously and can yield thousands to millions of sequencing reads per sample per patient (depending on platform and preparation protocol). Generally, in the context of HIV-1, NGS involves PCR amplification of all viral genetic material in a single or multiple sample(s). Next is the preparation of an amplicon library. The protocol differs depending on amplicon size: if the amplicon is small (250-400 nucleotides, though this is somewhat platform-dependent), then amplicons can be uniquely tagged and moved directly into the pipeline. However, longer amplicons must first be cut up into smaller fragments. Next, all sequences in the library are clonally amplified and sequenced directly. What results are thousands to millions of clonal sequencing reads per sample per patient, again depending on the platform and protocol. With respect to genotypic coreceptor prediction, NGS produces many clonal sequences per individual but it can also improve predictions by capturing low frequency non-R5 variants (below the limit of detection of phenotypic assays) allowing for more accurate coreceptor predictions [46].

Examples of NGS platforms include Roche 454, Illumina MiSeq, Ion Torrent, and SOLiD sequencing. My thesis research utilized the Roche 454 platform; as such, a brief description of this technology follows here. To sequence the env-V3 loop, I employed the amplicon sequencing method. V3 sequence specific primers were used to amplify envelope V3 from patient samples in an initial round of RT-PCR; this was followed by a

second round of PCR amplification using nested primers that contained unique multiplex identification (MID) tags. Two sample libraries containing full-length V3 sequences were created by combining sequences with unique MID tags in equal proportions. Each sequence is then tagged with specialized primer adapters that enable the amplicons to bind to a specialized reaction bead. Next, the amplicons, now attached to reaction beads, are added to a single emulsion PCR (emPCR) reaction mix that encapsulates each bead in a "microreaction" bubble that contains all the necessary enzymes and reagents for clonal PCR amplification of each amplicon on a bead. EmPCR produces beads coated with multiple copies of the original V3 amplicon. These beads are now deposited onto a picotiter plate to undergo pyrosequencing. Nucleotides are then automatically flowed over the plate one at a time (adenine [A], cytosine [C], thymine [T], or guanine [G]) in a cyclic fashion. When the nucleotide being flowed over the plate complements the nucleotide in the template sequence, it binds and emits a specific unit of light that is recorded. If more than one nucleotide is added in a single cycle, the light signal produced is designed to be proportional to the amount of nucleotides added. For example, if three nucleotides are added, the recorded amount of light should be triple the amount compared to a single nucleotide.

Roche 454 was initially favoured as a platform due to its longer (300-700 nucleotide) sequencing reads. However, its major limitation is its high error rates in areas with single nucleotide repeats (homopolymer repeats) [47]. Specifically, in these regions the strength of the emitted light is no longer proportional to the number of bases incorporated, resulting in missed bases or erroneous deletions [47]. This is particularly problematic because the HIV-1 genome contains many homopolymer repeats (usually stretches of adenines [A]) [47], making it difficult to determine between sequencing error and actual sequence variation. Currently these errors are dealt with bioinformatic post processing. More recently the Illumina MiSeq platform has become more popular than 454. Although its sequencing reads are slightly shorter, this technology has significantly improved error rates [48] requiring less post-processing. Additional novel sequencing technologies are also in development. These "third generation" technologies aim to offer an even higher throughput at a lower cost – an example is nanopore DNA sequencing. Currently this method has high error rates, but offers to sequence genomes without the need to previously PCR amplify them.

Regardless of technology and platform, DNA sequencing has proven to be an invaluable to assess HIV-1 sequence diversity and evolution at a greater level of detail. In the specific context of viral coreceptor usage prediction, NGS paired with bioinformatic coreceptor predictions allow the identification and interpretation of each virus sequence amplified from an individual patient. Unlike Sanger sequencing, NGS can identify low frequency variants that can then be used to determine viral tropism which are crucial in antiretroviral therapy settings [19].

3.7. References

1. Klatzmann D, Champagne E, Chamaret S, Gruest J, Guetard D, et al. (1984) T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature 312: 767-768.

2. Feng Y, Broder CC, Kennedy PE, Berger EA (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science 272: 872-877.

3. Alkhatib G, Combadiere C, Broder CC, Feng Y, Kennedy PE, et al. (1996) CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science 272: 1955-1958.

4. Berger EA, Murphy PM, Farber JM (1999) Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol 17: 657-700.

5. Fatkenheuer G, Nelson M, Lazzarin A, Konourina I, Hoepelman AI, et al. (2008) Subgroup analyses of maraviroc in previously treated R5 HIV-1 infection. N Engl J Med 359: 1442-1455.

6. Gulick RM, Lalezari J, Goodrich J, Clumeck N, DeJesus E, et al. (2008) Maraviroc for previously treated patients with R5 HIV-1 infection. N Engl J Med 359: 1429-1441.

7. Centers for Disease C (1981) Pneumocystis pneumonia--Los Angeles. MMWR Morb Mortal Wkly Rep 30: 250-252.

8. Berger EA, Doms RW, Fenyo EM, Korber BT, Littman DR, et al. (1998) A new classification for HIV-1. Nature 391: 240.

9. Cocchi F, DeVico AL, Garzino-Demo A, Arya SK, Gallo RC, et al. (1995) Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. Science 270: 1811-1815.

10. Richman DD, Bozzette SA (1994) The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression. J Infect Dis 169: 968-974.

11. Koot M, Keet IP, Vos AH, de Goede RE, Roos MT, et al. (1993) Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS. Ann Intern Med 118: 681-688.

12. Whitcomb JM, Huang W, Fransen S, Limoli K, Toma J, et al. (2007) Development and characterization of a novel single-cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor tropism. Antimicrob Agents Chemother 51: 566-575.

13. Low AJ, McGovern RA, Harrigan PR (2009) Trofile HIV co-receptor usage assay. Expert Opin Med Diagn 3: 181-191.

14. De Jong JJ, De Ronde A, Keulen W, Tersmette M, Goudsmit J (1992) Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. J Virol 66: 6777-6780.

15. Fouchier RA, Groenink M, Kootstra NA, Tersmette M, Huisman HG, et al. (1992) Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. J Virol 66: 3183-3187.

16. Fouchier RA, Brouwer M, Broersen SM, Schuitemaker H (1995) Simple determination of human immunodeficiency virus type 1 syncytium-inducing V3 genotype by PCR. J Clin Microbiol 33: 906-911.

17. Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R (2007) Bioinformatics prediction of HIV coreceptor usage. Nat Biotechnol 25: 1407-1410.

18. Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, et al. (2003) Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. J Virol 77: 13376-13388.

19. Knapp DJ, McGovern RA, Poon AF, Zhong X, Chan D, et al. (2014) "Deep" sequencing accuracy and reproducibility using Roche/454 technology for inferring co-receptor usage in HIV-1. PLoS One 9: e99508.

20. Swenson LC, Dong, W., Mo, T., Woods, C., Zhong, X., Thielen, A., Jensen, M., Biswas, P., Ellery, S., Lewis, M., James, I., Chapman, D., Valdez, H., Harrigan,

R. “Deep” sequencing to identify treatment-experienced patients who respond to maraviroc (MVC); 2009; Cologne, Germany. European AIDS Clinical Society.

21. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, et al. (2008) Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A 105: 7552-7557.

22. Feng Z, Dubyak GR, Lederman MM, Weinberg A (2006) Cutting edge: human beta defensin 3--a novel antagonist of the HIV-1 coreceptor CXCR4. J Immunol 177: 782-786.

23. Agace WW, Amara A, Roberts AI, Pablos JL, Thelen S, et al. (2000) Constitutive expression of stromal derived factor-1 by mucosal epithelia and its role in HIV transmission and propagation. Curr Biol 10: 325-328.

24. Ceresola ER, Nozza S, Sampaolo M, Pignataro AR, Saita D, et al. (2015) Performance of commonly used genotypic assays and comparison with phenotypic assays of HIV-1 coreceptor tropism in acutely HIV-1-infected patients. J Antimicrob Chemother.

25. Raymond S, Saliou A, Nicot F, Delobel P, Dubois M, et al. (2013) Characterization of CXCR4-using HIV-1 during primary infection by ultra-deep pyrosequencing. J Antimicrob Chemother 68: 2875-2881.

26. Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, et al. (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 86: 367-377.

27. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, et al. (1996) Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. Science 273: 1856-1862.

28. Carrington M, Dean M, Martin MP, O'Brien SJ (1999) Genetics of HIV-1 infection: chemokine receptor CCR5 polymorphism and its consequences. Hum Mol Genet 8: 1939-1945.

29. Martinson JJ, Chapman NH, Rees DC, Liu YT, Clegg JB (1997) Global distribution of the CCR5 gene 32-basepair deletion. Nat Genet 16: 100-103.

30. Quillent C, Oberlin E, Braun J, Rousset D, Gonzalez-Canali G, et al. (1998) HIV-1-resistance phenotype conferred by combination of two separate inherited mutations of CCR5 gene. Lancet 351: 14-18.

31. McDermott DH, Zimmerman PA, Guignard F, Kleeberger CA, Leitman SF, et al. (1998) CCR5 promoter polymorphism and HIV-1 disease progression. Multicenter AIDS Cohort Study (MACS). Lancet 352: 866-870.

32. Winkler C, Modi W, Smith MW, Nelson GW, Wu X, et al. (1998) Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. ALIVE Study, Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC). Science 279: 389-393.

33. Perez EE, Wang J, Miller JC, Jouvenot Y, Kim KA, et al. (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotechnol 26: 808-816.

34. Hutter G, Nowak D, Mossner M, Ganepola S, Mussig A, et al. (2009) Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. N Engl J Med 360: 692-698.

35. Low AJ, Swenson LC, Harrigan PR (2008) HIV coreceptor phenotyping in the clinical setting. AIDS Rev 10: 143-151.

36. Kordelas L, Verheyen J, Beelen DW, Horn PA, Heinold A, et al. (2014) Shift of HIV tropism in stem-cell transplantation with CCR5 Delta32 mutation. N Engl J Med 371: 880-882.

37. Tebit DM, Arts EJ (2011) Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. Lancet Infect Dis 11: 45-56.

38. Carlson JM, Schaefer M, Monaco DC, Batorsky R, Claiborne DT, et al. (2014) HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. Science 345: 1254031.

39. Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, et al. (2010) Wide variation in the multiplicity of HIV-1 infection among injection drug users. J Virol 84: 6241-6247.

40. Schuurman R, Demeter L, Reichelderfer P, Tijnagel J, de Groot T, et al. (1999) Worldwide evaluation of DNA sequencing approaches for identification of drug resistance mutations in the human immunodeficiency virus type 1 reverse transcriptase. J Clin Microbiol 37: 2291-2296.

41. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, et al. (2009) Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J Exp Med 206: 1273-1289.

42. Skrabal K, Low AJ, Dong W, Sing T, Cheung PK, et al. (2007) Determining human immunodeficiency virus coreceptor use in a clinical setting: degree of correlation between two phenotypic assays and a bioinformatic model. J Clin Microbiol 45: 279-284.

43. Sing T, Low AJ, Beerenwinkel N, Sander O, Cheung PK, et al. (2007) Predicting HIV coreceptor usage on the basis of genetic and clinical covariates. Antivir Ther 12: 1097-1106.

44. Tsibris AM, Korber B, Arnaout R, Russ C, Lo CC, et al. (2009) Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. PLoS One 4: e5683.

45. Quinones-Mateu ME, Avila S, Reyes-Teran G, Martinez MA (2014) Deep sequencing: becoming a critical tool in clinical virology. J Clin Virol 61: 9-19.

46. Swenson LC, Moores A, Low AJ, Thielen A, Dong W, et al. (2010) Improved detection of CXCR4-using HIV by V3 genotyping: application of population-based and "deep" sequencing to plasma RNA and proviral DNA. J Acquir Immune Defic Syndr 54: 506-510.

47. Gilles A, Meglecz E, Pech N, Ferreira S, Malausa T, et al. (2011) Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics 12: 245.

48. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30: 434-439.

Chapter 4.

Longitudinal deep sequencing and phylogenetic reconstruction of CXCR4 HIV-1 transmission to an individual homozygous for the CCR5-Δ32 mutation

The research presented in this chapter has been submitted to an international peer-reviewed journal in a modified form.

4.1. Abstract

Individuals homozygous for a naturally-occurring 32 base pair deletion in the CCR5 gene (CCR5-Δ32/Δ32) are resistant to infection by CCR5-using ("R5") HIV-1 strains but remain susceptible to CXCR4-using ("X4") strains. We identified a putative case of X4 HIV-1 transmission from a CCR5-wt/wt donor to a CCR5-Δ32/Δ32 recipient by injection drug use. We characterized this rare event by longitudinal envelope-V3 deep sequencing and phylogenetic ancestral reconstruction. Despite high (5.1 $\log_{10}$ copies/ml) donor pVL, ancestral reconstruction supported transmission of a single transmitted/founder (T/F) X4 virus to the recipient. This T/F virus differed by only one V3 residue from that originally acquired by the donor. Moreover, analysis of both plasma and PBMC V3 sequences suggested that HIV-1 infection in the recipient could have been initiated by transfer of an infected cell (*i.e.* not a free virion) from the donor. Whereas the donor's HIV-1 population gradually reverted from 100% X4 to ~60% R5 over ~4 years, the recipient's HIV-1 remained consistently X4 despite substantial within-host diversification. Our observations underscore the influence of host genetics on HIV-1 evolution and support cellular transfer as a mode of transmission.

4.2. Introduction

Entry of human immunodeficiency virus type-1 (HIV-1) into target cells occurs via binding of the viral envelope protein gp120 to the host CD4 receptor [1] followed by binding to chemokine coreceptors CCR5 or CXCR4 on the host cell surface [2,3]. HIV-1 strains that utilize CCR5 or CXCR4 are termed “R5” and “X4” respectively; those capable of utilizing either coreceptor are termed “R5/X4” (or dual-tropic) [4]. Coreceptor usage of a given HIV-1 isolate can be determined using cell-culture based phenotypic assays [5]. Alternatively, as the principal genetic determinants of viral coreceptor usage lie in the third variable (V3) loop of HIV-1 envelope gp120 (HXB2 codons 296-331) [6,7], coreceptor usage can also be predicted genotypically, via bioinformatic algorithms trained on large linked V3 sequence/phenotype datasets. Examples of widely-used V3 genotypic prediction algorithms include the support vector machine-based geno2pheno$_{[coreceptor]}$ (g2p) [8] and the position-specific scoring matrix-based WebPSSM [9].

R5 strains predominate globally [10] as well as during all infection stages [11]. R5 strains are also preferentially transmitted [10,12]. This may be due in part to their numerical dominance, their lower susceptibility to certain antiviral factors (*e.g.* defensins [13], SDF-1 [14]), their decreased N-linked glycosylation in V3 [12,15,16], and/or the enrichment of CD4+CCR5+ target cells at sites of primary infection [10,15]. Indeed, recent studies of acute infection estimate that R5 strains comprise approximately 83%-97% of primary infections, with X4/dual tropic strains making up the remaining 3%-17% [17,18]. If left untreated, X4 variants tend to emerge alongside their R5 counterparts over a timeline of years in up to 50% of individuals infected with HIV-1 subtype B, whereas the remainder of individuals will exclusively harbor R5 variants throughout their infection course [19,20].

Major advances have recently been made in our understanding of HIV-1 transmission [21-23]. This event is characterized by a severe genetic bottleneck, where an estimated 80% of heterosexual infections are productively initiated by a single transmitted/founder (T/F) variant [21,24]. Infection via injection drug use is generally initiated by an average of 3 closely-related T/F viruses [25]. Much effort has been

devoted to the genetic characterization of T/F viruses. This is traditionally achieved by creating a consensus sequence from single-template (*e.g.* clonal, deep-sequenced or single-genome amplified) HIV-1 sequences sampled from plasma shortly after infection [21,25]. Alternatively, phylogenetic ancestral reconstruction techniques can been applied to longitudinal single-template HIV-1 sequence datasets – even those sampled weeks or months following infection – to estimate infection dates, reconstruct T/F virus sequences and study within-host HIV-1 evolution in detail [26-28]. For example, phylogenetic techniques were applied to longitudinal within-host deep V3 sequences to reconstruct the timing and emergence of ancestral X4 lineages in patients who underwent a R5-to-X4 "switch" over the infection course [27].

Another (host) genetic determinant of HIV-1 transmission is the naturally-occurring 32 base pair deletion in the human CCR5 gene (CCR5-Δ32), that results in a non-functional CCR5 protein [29]. Rare individuals homozygous for this deletion ("CCR5-Δ32/Δ32") – who comprise approximately 1% of individuals of European descent [29,30] – are effectively resistant to infection by R5 strains but remain susceptible to infection by X4 or dual tropic strains [29,31-33]. Although X4 infections in CCR5-Δ32/Δ32 individuals have been documented [34-46], no studies of these rare transmission cases have combined next-generation sequencing and phylogenetic approaches to identify the T/F virus and characterize intra- and inter- host HIV-1 evolution in detail.

In the present study, we identify a putative case of X4 transmission from a CCR5 wt/wt donor to a homozygous CCR5-Δ32/Δ32 recipient from among the participants of the Vancouver Injection Drug Users Study (VIDUS) [47]. We combine deep sequencing of the V3 region in plasma and PBMC with phylogenetic ancestral reconstruction to study within-host HIV-1 evolution in donor and recipient, including characterization of the T/F virus in both instances. Results reveal transmission of a single X4 variant from donor to recipient, presumably via transfer of an infected cell, and differential HIV-1 coreceptor usage evolution in these two hosts with different genetic backgrounds.

4.3. Methods

4.3.1. Vancouver Injection Drug Users Study (VIDUS)

Founded in 1996, the original Vancouver Injection Drug Users Study (VIDUS) was a longitudinal cohort of 1603 active injection drug users 18 years or older recruited from Vancouver's Downtown East Side through street outreach [47]. At baseline and semi-annual follow-up visits, participants completed a semi-structured, interviewer-administered questionnaire and provided a blood sample that was separated into plasma and peripheral blood mononuclear cells (PBMC) and stored at -80°C until use. PBMC pellets were frozen directly (*i.e.* not cryopreserved); as such, cell separation and viral outgrowth assays were not possible. Of 1603 VIDUS participants recruited, 325 (20.3%) were HIV-1 positive (seroprevalent) at study entry whereas 141 (8.8%) seroconverted during follow-up; all other participants did not register an HIV-positive test during follow-up. The present study made use of available bulk plasma HIV-1 RNA and/or DNA sequences spanning Gag, Integrase, V3 and Nef from 115 (of 141, 82.3%) seroconverters and 124 (of 325, 38.2%) seroprevalent VIDUS participants (total 239).

4.3.2. Ethics statement

This study was approved by the Research Ethics boards at Providence Health Care/University of British Columbia and Simon Fraser University.

4.3.3. Amplification and bulk sequencing of HIV-1 RNA and DNA from VIDUS participants

Total nucleic acids were extracted from plasma and PBMC pellets collected from VIDUS participants using standard methods. HIV-1 Gag, Integrase, V3 and Nef were amplified by nested RT-PCR (for HIV-1 RNA; Invitrogen SuperScript III One-Step RT-PCR System) or nested PCR (for HIV-1 DNA; Roche Expand High Fidelity PCR System) using primers optimized for HIV-1 subtype B sequences. Amplicons were bidirectionally sequenced on a 3130xl or 3730xl automated DNA sequencer (Applied Biosystems). Chromatograms were analyzed using Sequencher v5.0.1 (Genecodes) or custom software RECall [48] with nucleotide mixtures called if the height of the secondary peak

exceeded 25% of the dominant peak height (Sequencher) or 20% of the dominant peak area (RECall). Alignment to the HIV-1 subtype B reference strain HXB2 (for Gag, Integrase and Nef) or a modified subtype B reference sequence (for V3) was done using an in-house alignment tool based on the HyPhy platform [49]. Maximum likelihood phylogenetic trees were constructed using PhyML 3.0 [50]. Patristic (tip-to-tip) genetic distances, expressed in terms of substitutions per nucleotide site (sub/nt site), were extracted from maximum-likelihood Newick treefiles using PATRISTIC [51]. Trees were visualized using Figtree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

4.3.4. Identification of the putative transmission pair

A putative transmission pair was retrospectively identified via phylogenetic analysis of population-level bulk HIV-1 sequences: for all genes analyzed, this pair exhibited the shortest patristic distances in the VIDUS cohort (see results). Clinical estimated dates of infection (calculated as the midpoint between the last HIV-negative and first positive sample) were March 2000 for the donor and August 2001 for the recipient. The inferred transmission date was considered “time-zero”, with all donor and recipient timepoints expressed as months relative to this date, and all specimens named according to their source (“D” for donor, “R” for recipient) and sample type (plasma or PBMC). For example, R0M refers to the transmission event (for which no specimen was available) while D-13M^{Plasma} refers to the donor plasma sample collected 13 months prior to transmission. For the donor, paired plasma and PBMCs were available at -13, -7, -1, and +35 months from the transmission event, while recipient plasma samples were available at +5 months, and paired plasma/PBMCs at +6, and +12 months from this date.

The donor, initially antiretroviral naïve, began highly active antiretroviral therapy (HAART) in late August 2001, shortly after the estimated date of transmission, and remained intermittently on HAART over the course of study followup. The recipient remained antiretroviral-naive over the course of study followup.

4.3.5. CCR5-Δ32 and HLA class I genotyping

Host CCR5-Δ32 genotyping of VIDUS participants was performed as described previously [52]. Briefly, a ~172 bp region spanning the deletion site was amplified by nested PCR from plasma and/or PBMC-derived DNA and visualized on a 2% agarose gel. To confirm the genotype, 2nd round amplicons were bidirectionally sequenced and chromatograms were visually assessed for length and the presence of the prolonged mixed-base motif characteristic of heterozygous CCR5-wt/Δ32 genotypes. In doing so, the putative donor and recipient were identified as homozygous CCR5-wt/wt and CCR5Δ32/Δ32 respectively. The recipient's CCR5Δ32/Δ32 status was confirmed by genotyping 9 additional specimens collected longitudinally from this individual. High resolution HLA class I typing was performed by sequence-based typing [53].

4.3.6. Longitudinal deep-sequencing of HIV-1 V3 RNA and DNA sequences from donor and recipient

Prior to deep-sequencing, the V3 region was amplified in triplicate from all donor and recipient plasma and PBMC-derived nucleic acid extracts. Nested second round amplification was performed using forward and reverse primers incorporating one of 12 multiplex identifier (MID) tags and a linker sequence at the 5' end and visualized on a 1% agarose gel. Amplicons were quantified with the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) on a DTX 880 Multimode Detector (Beckman Coulter), pooled in equal proportions, purified, re-quantified, and deep-sequenced using the GS Junior Titanium Sequencing Kit on a GS Junior instrument (Roche/454). To avoid low-level, intra-run sequence cross-contamination by genetically similar amplicons, we sequenced each donor and recipient amplicon on a separate GS-Junior run (as each GS-Junior run typically included 24 V3 amplicons, this means that each run contained one amplicon from the donor or recipient, and 23 V3 amplicons from patients unrelated to the present study). The one exception was donor sample D+35M$^{Plasma/PBMC}$, where data are derived from an initial run that included other donor and recipient samples. Inclusion of data from D+35M$^{Plasma/PBMC}$ from this run was deemed appropriate after quality-control experiments confirmed that HIV-1 sequences and their distributions obtained from separate vs. combined runs were highly concordant (not shown).

4.3.7. Processing of deep sequencing data

Raw sequences were processed, aligned, and trimmed to a modified HIV-1 HXB2 V3 reference standard (HXB2 gp120 codons 296-331) using an iterative process as described previously [54]. Briefly, identical sequences were collapsed and annotated with read counts. Sequences were discarded if the MID or primer sequence was a mismatch to the one assigned to the sample or the sequence did not align to the V3 reference standard. Sequences that were identical except for 1-2 gap characters (attributable to erroneous indels introduced during sequencing) were merged, and read counts updated. The remaining sequences were re-aligned to generate a sample-specific consensus sequence, which was used as the reference standard in subsequent steps.

After realigning all sequences to the specimen-specific consensus, any gap characters followed by ≥3 instances of the same nucleotide were replaced with that nucleotide (to correct for the GS-Junior platform's difficulty in sequencing homopolymer repeats), and insertions/deletions (indels) were moved to be in-frame. Identical sequences were again merged and read counts updated. A multiple alignment was performed on all remaining sequences, and sequences observed at frequencies of <1% that still contained a single gap character were discarded. To remove any low-level sequence contamination from other patient-derived amplicons sequenced in the same run, an intra-run cross contamination check was performed. To do this, the 5 most frequent sequences within each run (that were observed at a >10% overall prevalence) were identified. Every sequence in our sample was then compared against this list and discarded if it represented an exact match. Lastly, nucleotide sequences with read counts of ≤2, those not divisible by 3 after removal of gap characters, those not encoding cysteines (C) as the starting and final V3 residues, and those <96 or >189 base pairs were discarded as invalid prior to final analysis [55]. Overall, approximately 8 to 32% of raw sequences were discarded as a result of this processing pipeline.

4.3.8. Ancestral phylogenetic reconstructions

Ancestral phylogenetic reconstructions of intra-host HIV-1 evolution, including the estimation of transmitted/founder (T/F) sequences and dates, were performed using

deep sequence data from the three donor and three recipient plasma timepoints surrounding the transmission event: D-13M^{Plasma}, D-7M^{Plasma}, D-1M^{Plasma} for the donor and R+5M^{Plasma}, R+6M^{Plasma}, and R+12M^{Plasma} for the recipient. To maximize information incorporated into the phylogeny, a ~250bp sequence encompassing V3 and flanking regions (mapping approximately to HXB2 genomic nucleotides 7086-7336) was used. Processing of V3 deep sequence data for ancestral reconstructions was done using an in-house pipeline described previously [27]. Briefly, raw sequences were grouped by their unique MID tag, and nucleotides with low quality scores (as reported by Roche GS-Junior software) were trimmed from the 5' and 3' ends. Identical sequences were temporarily collapsed and annotated with read counts. These were subsequently aligned using a custom sequence alignment algorithm in HyPhy [49] that adjusts for the high indel rates observed with the GS-Junior platform by aligning all three reading frames to a reference protein standard spanning HXB2 gp120 codons 278-375. This algorithm assumes that a true V3 sequence will encode a single open reading frame, with any frameshifts attributable to erroneous indels introduced during sequencing. Aligned sequences were then re-expanded by their read counts and annotated with sample dates expressed in terms of days elapsed since January 1, 1990.

Reconstructed, time-stamped phylogenies were created using Bayesian Evolutionary Analysis Sampling Trees (BEAST) v1.6.1 [56] using parameters described previously with some modifications [27]. Briefly, 100 sequences were randomly sampled from each donor and recipient plasma timepoint, for a total of 600 sequences included in each reconstruction. These 600 sequences were aligned using MUSCLE v3.8.31 [57] and alignments were manually curated using Se-Al (http://tree.bio.ed.ac.uk/software/seal/). Alignments were converted into a BEAST XML file with the following parameter settings: Tamura-Nei [58] nucleotide substitution model; uncorrelated lognormal molecular clock; Bayesian skyline model with 5 population size classes; and a chain length of 10^8 with chain states written to log files at intervals of 10^5 with a burn-in period of 2×10^7 (20%). The resulting trees were then thinned down to 100 sampled at regular intervals. Convergence of chain states was assessed using Gelman and Rubin's convergence diagnostic implemented in the R package *coda* [59]. For each tree, a Muse-Gaut codon substitution model crossed with a general time-reversible model of nucleotide substitution (implemented in HyPhy [49]) was fit to every

tree. Ancestral sequences were generated by sampling 100 character states from the posterior distributions reconstructed at every node of the tree. In total, 10 independent ancestral reconstructions, each randomly sampling 100 sequences per timepoint for a total of 600 sequences, were performed.

4.3.9. Assessing V3 sequence divergence and diversity

Within-host HIV-1 genetic divergence was over time was assessed by calculating patristic (tip-to-tip) phylogenetic distances between each host's reconstructed T/F virus and all the sequences observed in their plasma and PBMC specimens thereafter, taking into consideration the frequency of each sequence. Donor and recipient HIV-1 diversity, calculated as per-codon differences in Shannon entropy, were calculated from V3 amino acid alignments from the earliest and latest plasma timepoints using Entropy-Two (http://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy.html) using 1000 randomizations with replacement.

4.3.10. Inference of HIV-1 coreceptor usage

HIV-1 coreceptor usage (R5 vs. X4) was predicted from bulk and deep HIV-1 V3 sequences using geno2pheno$_{[coreceptor]}$ (g2p) [8]. This algorithm assigns each sequence a false-positive rate (FPR) that represents the probability of classifying an R5-virus falsely as X4. In the present study, we employed a false positive rate (FPR) cutoff of 5.75% that was derived from retrospective analysis of clinical data from the MOTIVATE clinical trials [60]. Using this cutoff, V3 sequences with FPR ≤5.75% and >5.75% were classified as X4 and R5, respectively.

4.4. Results

4.4.1. Identification of the putative transmission pair

Maximum-likelihood phylogenies were constructed using one bulk HIV-1 plasma RNA or PBMC DNA Gag, Integrase, V3, and Nef sequence per individual for 239 VIDUS participants (Figure 4.1). Computation of patristic (tip-to-tip) genetic distances within

these phylogenies consistently identified a participant pair who exhibited the lowest overall distances for all HIV-1 genes examined: these were 0.0027 substitutions per nucleotide site (sub/nt site) in gag (compared to a cohort median of 0.064 [IQR 0.055-0.070]), 0.0023 for integrase (cohort median of 0.034 [IQR 0.025-0.041]), 0.010 for V3 (cohort median of 0.087 [IQR 0.056-0.12]) and 0.023 for nef (cohort median of 0.10 [IQR 0.081-0.11]). The overall prevalence of X4 HIV-1 among all VIDUS seroconverters and seroprevalent participants studied, inferred from bulk V3 sequences, was 14% (12% among seroconverters sequenced within the first year of infection). Over one-third of these resided in a single large cluster. Of interest, the putative transmission pair resided within this cluster (Figure 4.1): donor and recipient bulk V3 sequences were predicted as X4 with low g2p FPR values of 1.7% and 2.8%, respectively. Moreover, CCR5 genotyping revealed that the putative donor was CCR5-wt/wt whereas the putative recipient was homozygous CCR5-Δ32/Δ32. Neither donor nor recipient expressed classical "protective" HLA class I alleles: their types were A*03:01/A*31:01, B*07:02/B*51:01, C*07:02/C*14:02 (donor) and A*23:01/A*25:01, B*35:08/B*44:02, C*04:01/C*05:01 (recipient).

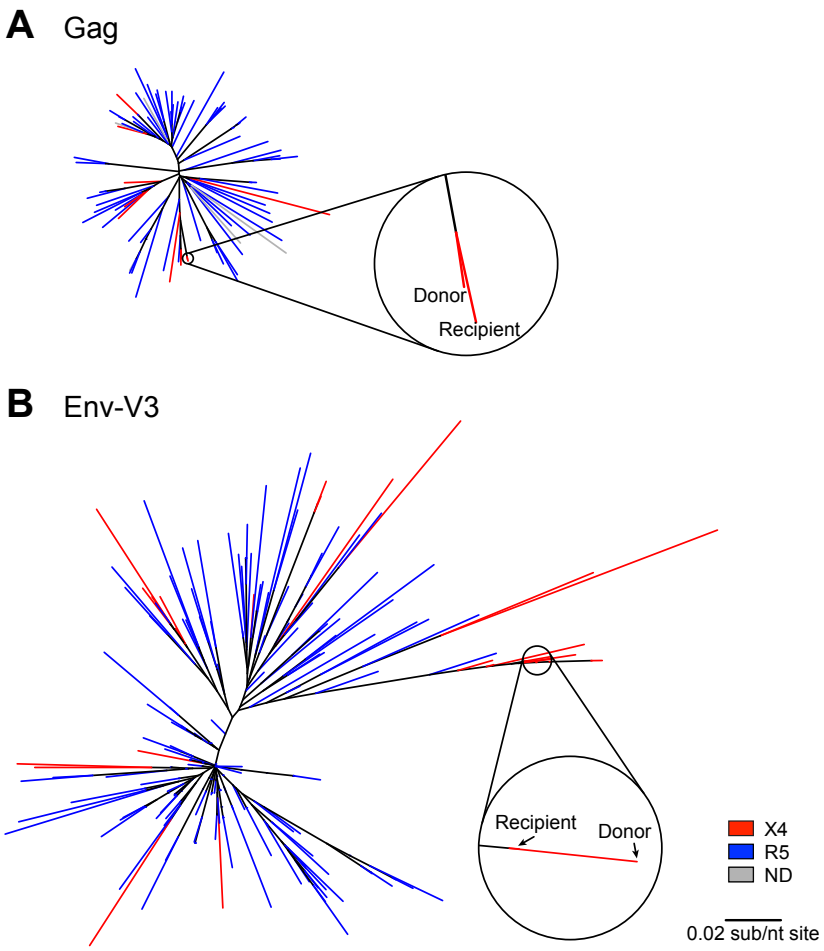

Figure 4.1. Maximum likelihood phylogenies of bulk HIV-1 Gag and V3 sequences from VIDUS participants

Maximum likelihood phylogenetic trees were constructed using available bulk Gag (***panel A***) and V3 (***panel B***) sequences from acute and chronically infected participants of the Vancouver Injection Drug Users Study. Putative donor and recipient sequences are shown in the zoomed-in window. Tree tips are coloured according to coreceptor usage predicted using V3 genotypes: red for X4-using, blue for R5-using sequences and gray for Gag sequences for which no corresponding V3 sequence was available for coreceptor prediction (ND; not determined). Gag and V3 sequences from putative donor and recipient exhibited the lowest pairwise genetic distances in the cohort.

HIV-1 transmission from the CCR5-wt/wt donor to the CCR5-Δ32/Δ32 recipient (rather than vice-versa) was inferred by estimating their infection dates (midpoint between their last HIV-negative and first HIV-positive tests): these were March 2000 (timepoint "D-17M", see methods) for the donor and August 2001 (transmission timepoint "R0M") for the recipient (Figure 4.2). Note that, despite harboring near-identical bulk HIV-1 sequences, confirmation of transmission (*e.g.* via participant contact) was not

possible due to the retrospective nature of the analysis and ethics guidelines. Nevertheless, this represents a rare opportunity to study intra- and inter-host evolutionary dynamics of a near-identical HIV-1 strain in genetically distinct individuals.

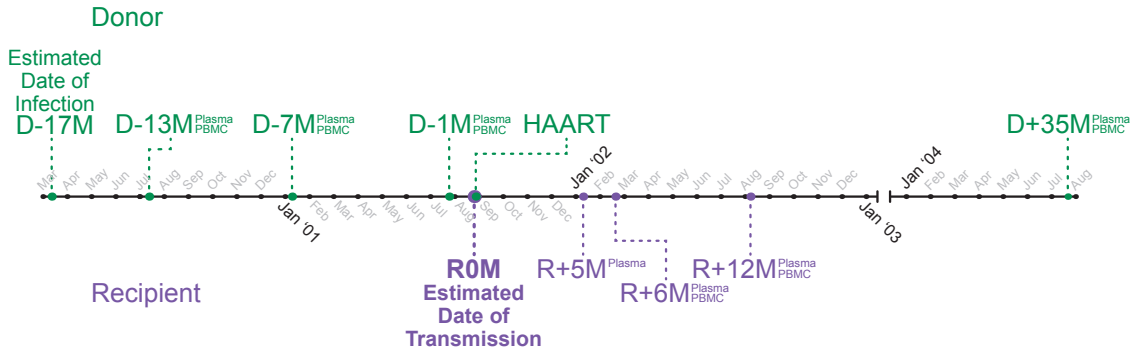

Figure 4.2. Sampling timeline for Donor and Recipient
Donor samples are in green and recipient samples are in purple. Putative infection/transmission dates for donor (March 2000) and recipient (August 2001) were estimated as the midpoint between their last HIV-negative and first HIV-positive tests. The recipient's estimated transmission date was set as "Month 0" (R0M); all patient samples were named according to their source ("D" for donor, "R" for recipient), their sampling time in months from the transmission date and their type (plasma or PBMC). Donor plasma and PBMC were available -13, -7, -1, and +35 months from the estimated date of transmission; recipient plasma was available +5, +6, and +12 months and PBMC +6 and +12 months from the estimated date of transmission.

4.4.2. Donor and recipient differences in nadir CD4 T-cell count

We first analyzed available pre-therapy clinical measurements (Figure 4.3a and 4.3b). The donor's highest plasma viral load (pVL), 5.1 $\log_{10}$ HIV-1 RNA copies/ml, and nadir CD4 T-cell count, 20 cells/mm^3, were observed 17 months postinfection. The donor initiated HAART <1 month thereafter. The recipient's highest pVL, observed 4.5 months postinfection, was 4.7 $\log_{10}$ HIV-1 RNA copies/ml whereas the nadir CD4 count, observed 9 months postinfection, was 270 CD4 cells/mm^3. The recipient's CD4 count subsequently rebounded to >400 cells/mm^3 and this individual remained HAART-naïve throughout followup.

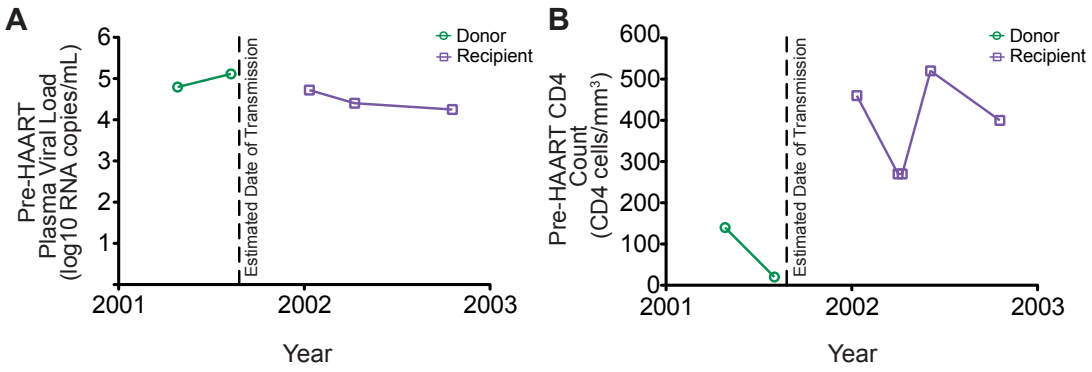

Figure 4.3. Clinical histories for donor and recipient
***Panel A*:** Available pre-HAART plasma viral loads (pVL) for donor (green) and recipient (purple). Maximum pVL for donor and recipient were 5.1 and 4.7 $\log_{10}$ RNA copies/mL, respectively. ***Panel B*:** Available pre-HAART CD4 counts for donor (green) and recipient (purple). Nadir CD4 counts for donor and recipient were 20 and 270 CD4 cells/mm^3 respectively.

4.4.3. Deep sequencing and ancestral reconstruction

Deep sequencing of the HIV-1 V3 region was performed on all available donor and recipient plasma and PBMC samples using the Roche GS-Junior Platform. A median of 3143 (range 1905-7248) high quality sequences per sample were analyzed. A total of 10 phylogenetic ancestral reconstructions were performed using 100 randomly sampled sequences from each of the three donor and recipient plasma timepoints closest to transmission. All 10 ancestral reconstructions supported transmission of a single T/F viral strain from the CCR5-wt/wt donor to CCR5-Δ32/Δ32 recipient between May 2001 and August 2001 (Figure 4.4), a range which corroborated the clinically estimated transmission date .The T/F virus sequence (estimated as the consensus of all 10 ancestral reconstructions performed) was predicted as X4 (median g2p FPR 2.6%, range 1.7-3.8%) (Figure 4.5).

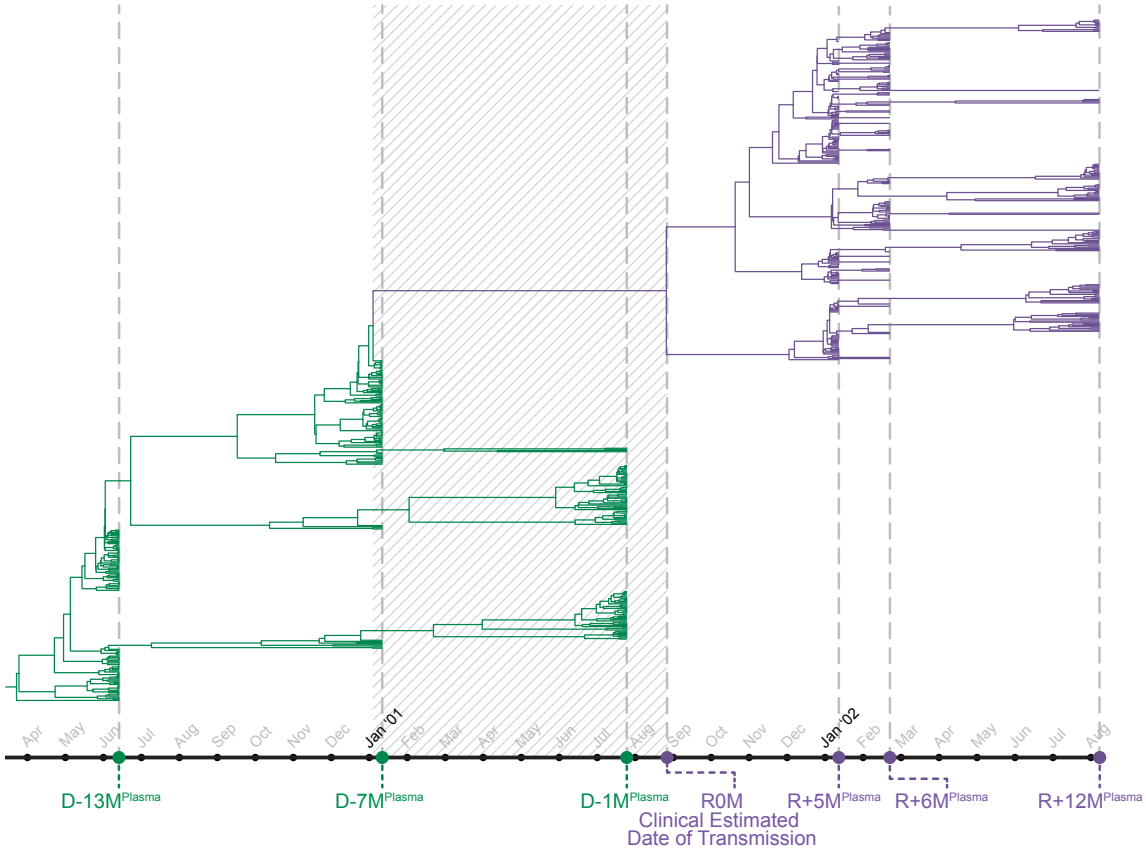

Figure 4.4. Ancestral phylogenetic reconstruction of HIV-1 V3 transmission/evolution in donor and recipient

N=10 ancestral phylogenetic reconstructions were performed by sampling 100 *plasma* HIV-1 RNA-derived ultradeep sequences per timepoint for the three donor (green) and three recipient (purple) timepoints closest to the estimated transmission date. A representative reconstructed phylogeny is shown. Reconstruction supports transmission of a single founder virus from donor to recipient at a timepoint between Jan and August 2001 (hatched grey area), which coincides with the clinical estimated date of transmission. All nine other phylogenetic reconstructions were also consistent with transmission of a single T/F virus; in addition, 8 of 10 reconstructions yielded transmission date ranges that coincided with the clinical estimated transmission date.

Phylogenetic ancestral reconstruction of the T/F virus that originally infected the donor indicated that this individual had also been infected with a single X4 variant (median g2p FPR 1.7%, range 1.7-3.2%) (Figures 4.4 and 4.5). The reconstructed V3 T/F sequence originally acquired by the donor and that transmitted to the recipient differed by only one amino acid at V3 codon 24 (Figure 4.5b). At this residue, the donor T/F virus harbored arginine (R) while the recipient's harbored lysine (K).

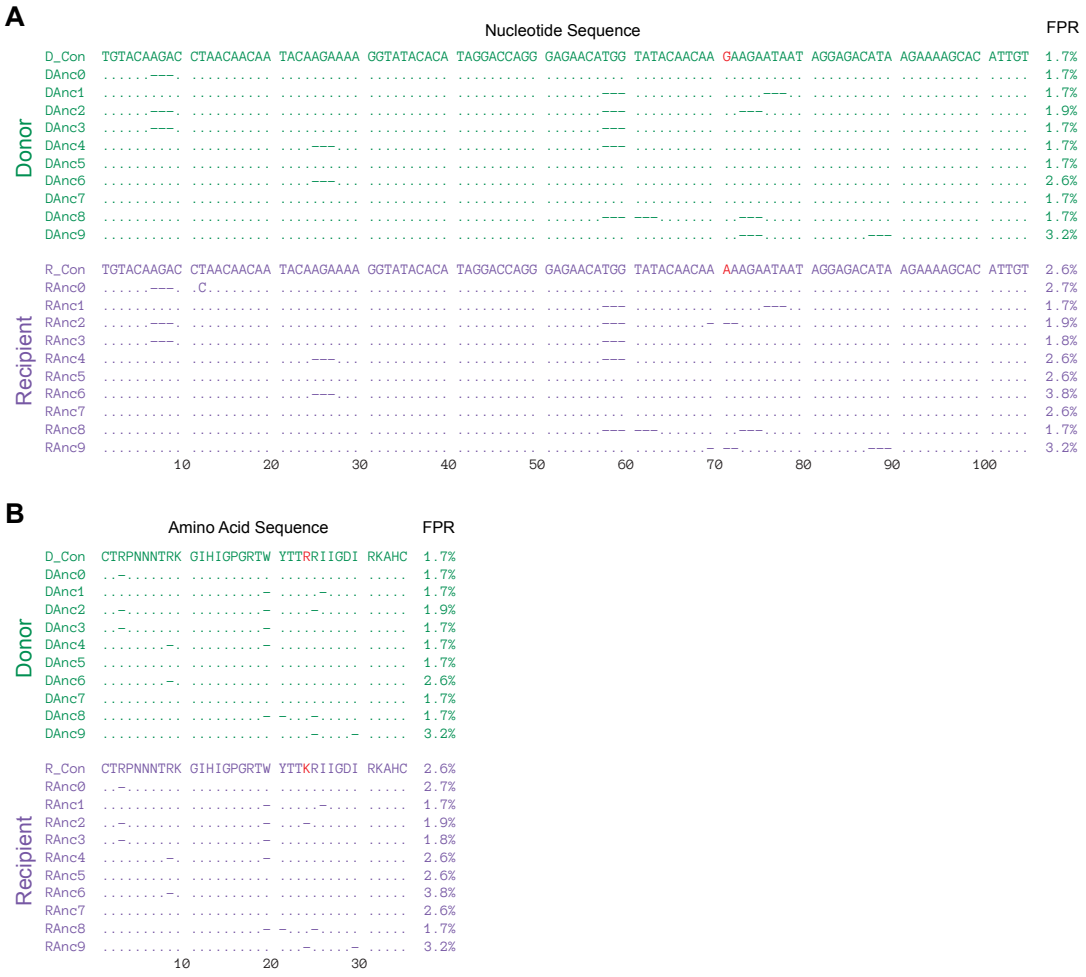

Figure 4.5. Nucleotide and protein alignments of reconstructed transmitted/founder viruses in donor and recipient

Nucleotide and amino acid sequence alignments of all reconstructed T/F viruses infecting the donor (green) and recipient (purple). The consensus sequence of all 10 ancestral reconstructions (labeled "D_Con" and "R_Con" for donor and recipient) is used as the reference sequence. Periods (".") indicate positions where the sequence is the same as the reference and dashes ("-") indicate deletions. The "FPR" value following each reconstructed T/F sequence denotes its false-positive rate assigned by geno2pheno$_{[coreceptor]}$[8]; sequences with FPR ≤5.75% are considered X4. ***Panel A*****:** Donor (top, green) and recipient (bottom, purple) nucleotide acid alignments. Consensus nucleotide differences between donor and recipient are shown in red. ***Panel B*****:** Donor (top, green) and recipient (bottom, purple) amino acid alignments. The single amino acid difference between donor and recipient T/F virus (at codon 24) is shown in red.

Reconstruction of the T/F virus sequence transmitted from donor to recipient allowed us to track its frequency in both hosts over time. In the donor, this sequence was first detected 7 months prior to transmission (D-7M) at 0.2% and 1.1% frequency in plasma and PBMCs, respectively (Figure 4.6a and b). One month prior to transmission

(D-1M), this sequence remained at low frequency (0.1%) in plasma but co-dominated (33.5%) in PBMC. By 35 months following transmission (D+35M), this sequence was not detected in donor plasma and was observed at only 0.8% in PBMC. The observation that the frequency of the T/F virus was negligible in plasma but co-dominant in PBMC suggests the possibility that 1) transmission from donor to recipient occurred via transfer of an infected cell (rather than a free virion) and 2) that the T/F virus may represent an archived X4 sequence in the donor.

In the recipient, the T/F virus sequence remained co-dominant in plasma for 6 months following transmission (36.2% in R+5M^{Plasma}; 50.0% in R+6M^{Plasma}), but was no longer detected in plasma by 12 months following infection (Figure 4.6c and d). By contrast, the T/F virus sequence remained co-dominant in recipient PBMCs over the entire course of follow-up (35.6% in R+6M^{PBMC} and 32.5% in R+12M^{PBMC}), suggesting archiving of this sequence.

We also tracked the frequency of the X4 T/F virus that originally infected the donor (not shown). It dominated in plasma (86.9%) at the donor's earliest studied timepoint (D-13M), continued to co-dominate up until transmission (42.4% in D-7M^{Plasma} and 34.1% in D-1M^{Plasma}), but was undetectable in plasma 35 months thereafter. Concomitantly, the frequency of this sequence steadily decreased in donor PBMCs, from 51.7% at D-13M^{PBMC} to 17.2% at D-7M^{PBMC} and then to low/undetectable levels thereafter (0% at D-1M^{PBMC} and 1.0% at D+35M^{PBMC}).

In summary, despite its dominance in donor plasma at the time of transmission, the donor's originally-acquired T/F sequence was not transmitted to the recipient. Rather, a viral variant that co-dominated in donor PBMC at that time was transmitted. Taken together, these data support transmission of HIV-1 from donor to recipient via transfer of an HIV-1 infected cell, or a very small minority plasma variant.

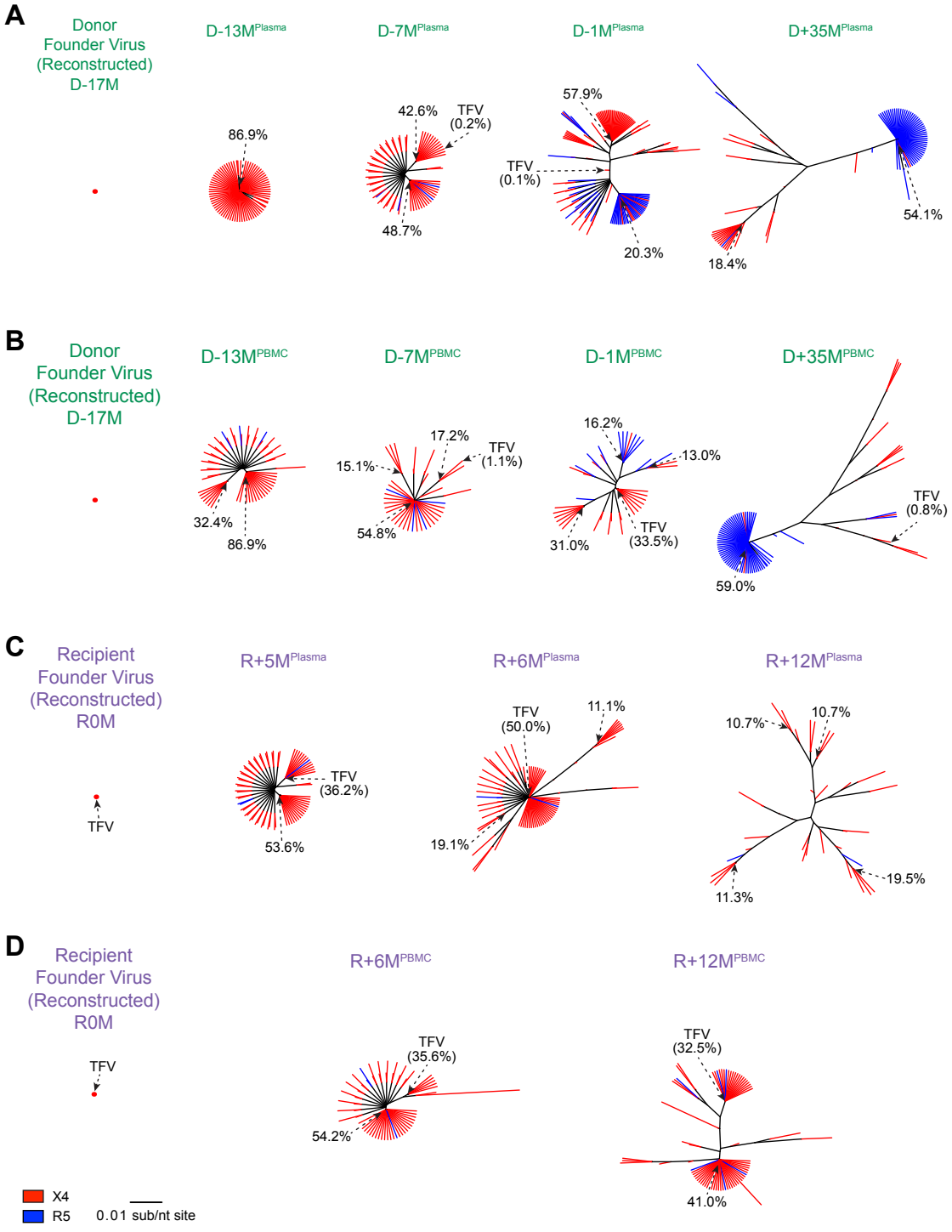

Figure 4.6. Increasing HIV-1 V3 diversification over time in donor and recipient

Maximum likelihood phylogenetic trees constructed from unique plasma and PBMC deep V3 sequences from donor (**panel A and B**) and recipient (**panel C and D**). Branches are colored by predicted coreceptor usage: red for X4; blue for R5. Prevalent sequences are labeled with their observed frequencies. "TFV" denotes the transmitted/founder virus transmitted from donor to recipient; its presence and frequency is tracked throughout donor and recipient trees. All phylogenies are drawn on the same genetic distance scale.

4.4.4. Divergence from the reconstructed T/F virus in the donor and recipient

We next wished to compare the extent to which plasma HIV-1 RNA V3 sequences in the donor and recipient initially diverged over time from their respective T/F viruses (Figure 4.7). Analysis was restricted to the donor's first 10 months of infection (pre-HAART) and a comparable follow-up time within the recipient. During this time, the mean divergence from the donor's T/F virus was 0.0027 sub/nt site, which translated to a rate of divergence of 0.00065 sub/nt site per month. In contrast, mean initial divergence of plasma V3 sequences from the T/F virus in the CCR5-Δ32/Δ32 recipient was 0.031 sub/nt site, which translated to a rate of divergence of 0.0036 sub/nt site per month. This value is 5.5-fold higher than that observed in the donor, indicating that V3 evolution in the recipient was not constrained by their CCR5-Δ32/Δ32 genotype.

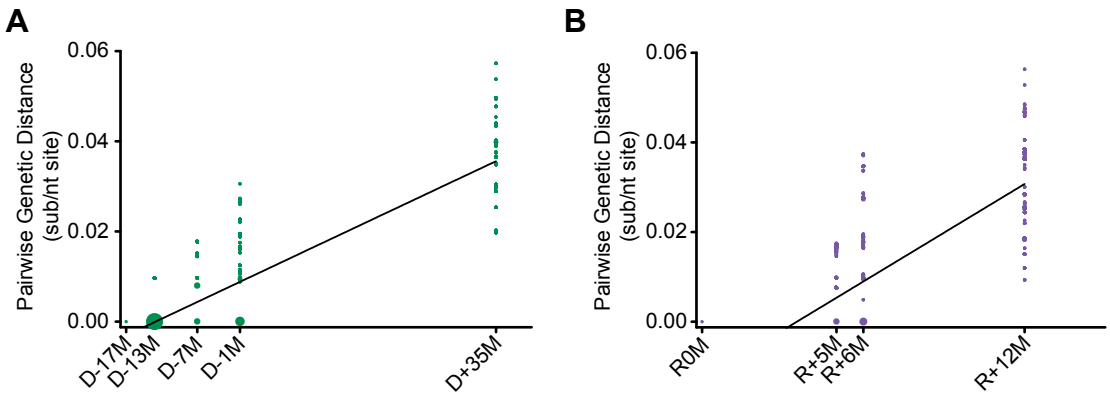

Figure 4.7. Increasing divergence from the transmitted/founder HIV-1 V3 sequence in both donor and recipient

Panel A: Pairwise genetic distances between the donor's estimated V3 T/F sequence and all subsequently-observed plasma HIV-1 RNA sequences, measured in terms of substitutions per nucleotide site (sub/nt site). Datapoint sizes reflect observed sequence frequencies, with the largest point representing ~5000 sequences. ***Panel B***: Corresponding genetic distances between the recipient's T/F and subsequent plasma V3 sequences.

4.4.5. Differential HIV-1 coreceptor usage evolution in donor and recipient

We next investigated HIV-1 coreceptor usage evolution in donor and recipient. In the donor, over a total of 52 months followup, R5 V3 sequences gradually emerged alongside their X4 counterparts (Figures 4.6a, 4.6b, 4.8a). The first R5 variants were detected 4 months following infection in PBMC (timepoint D-13M^{PBMC}): these early R5 variants comprised 0.2% of all sequences in this sample and exhibited g2p FPRs in the marginal range (6.6%-10.8%). No R5 variants were detected in plasma at this timepoint. By 10 months following infection (D-7M), R5 variants were detected at 0.4% frequency in plasma and 0.6% in PBMC, again with marginal FPRs (5.8%-10.8%). However, 16 months after infection and one month prior to transmission, R5 variants reached frequencies of 41.0% in plasma and 17.6% in PBMC, though FPRs remained marginal (median 8.7% in both compartments). By 35 months post-transmission, R5 sequences dominated in donor plasma (59.9%) and PBMCs (74.4%), with median FPR scores of 18.9% in both compartments (Figure 4.8a). Based on the trajectory of R5 emergence in the donor, at the time of transmission to the recipient, the donor's plasma virus was predicted to contain 79.0% X4 and 21.0% R5 variants (not shown).

In contrast, in the CCR5-Δ32/Δ32 recipient, essentially all (14743 of 14809; 99.6%) plasma and PBMC HIV-1 sequences remained X4 throughout followup (median FPR 2.6% in both compartments) (Figure 4.6c, 4.6d, 4.8b). The remaining minority (66 of 14809; 0.4%) of sequences were technically R5, but these exhibited marginal g2p FPRs (range 5.8%-8.7%). Moreover, unlike the donor, the frequencies of sequences with FPRs in this range did not increase over time in the recipient.

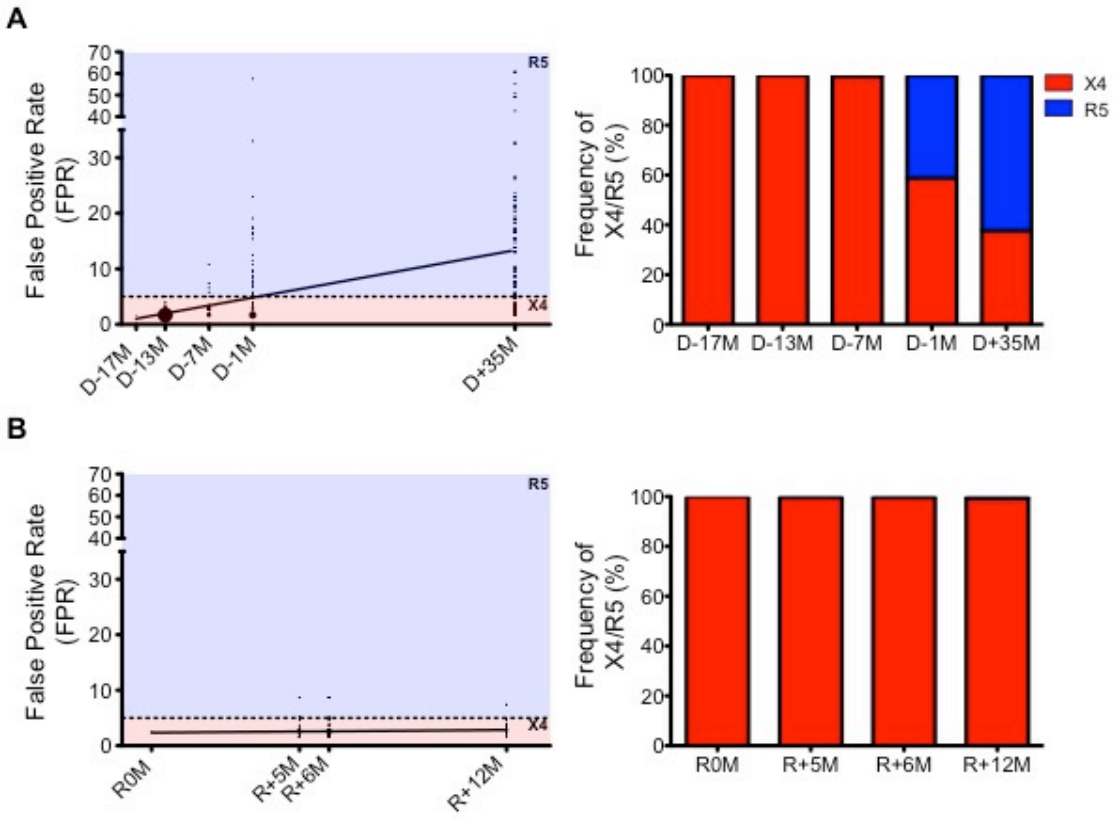

Figure 4.8. Marked differences in the evolution of coreceptor usage in CCR5-wt/wt donor vs. CCR5-Δ32/Δ32 recipient

Panel A left: The false-positive rate (FPR) of HIV-1 coreceptor usage prediction for the donor's T/F virus (D-17M) and each unique plasma HIV-1 RNA sequence collected thereafter. The horizontal dotted line denotes FPR=5.75%; sequences with values at or below this threshold are considered X4. ***Panel A right:*** summarizes the data in terms of the % of total sequences displaying X4 (red) vs. R5 (blue) usage at each timepoint. ***Panel B:*** Corresponding analyses for the CCR5-Δ32/Δ32 recipient.

We also investigated V3 codon substitutions over time (Figure 4.9). For the donor, comparison of the earliest (D-13M^{Plasma}) and latest (D+35M^{Plasma}) V3 sequences identified five codons (5, 24, 25, 27, and 34) that diversified significantly and three (8, 18, and 26) that contracted modestly during this time (Figure 4.9a) ($p<0.001$). Codon 25 diversified to the greatest extent, with the X4-associated arginine (R) decreasing from 99.8% to 38.7% frequency. In the CCR5-Δ32/Δ32 recipient, comparison of earliest (R+5M^{Plasma}) and latest (R+12M^{Plasma}) V3 sequences identified nine diversifying (4, 9, 24-27, 29, 30, 32) and three contracting codons (10, 13 and 18) ($p<0.001$) (Figure 4.9b). In the recipient, codon 25 also ranked among the most highly diversifying, with the

dominant X4-associated arginine (R) giving way to a 63.5/35.7% mixture of lysine (K)/arginine (R).

Taken together, V3 coreceptor evolution in the CCR5-wt/wt donor was markedly different than in the CCR5-Δ32/Δ32 recipient who remained consistently X4 despite considerable virus variation.

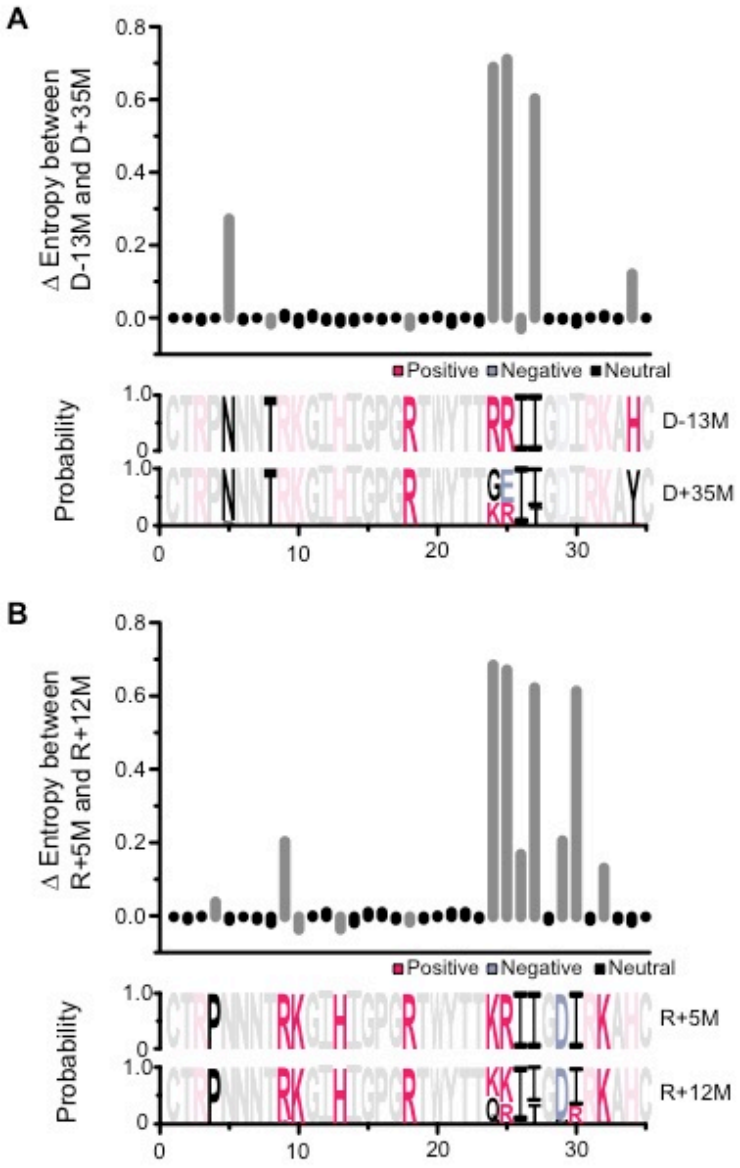

Figure 4.9. V3 sequences in both donor and recipient exhibit marked diversification at key coreceptor tropism determining sites

Panel A Top: Differences in Shannon Entropy (Δ Entropy) between the latest (D+35M) and earliest (D-13M) V3 plasma HIV-1 amino acid alignments from the CCR5-wt/wt donor. Positive values denote residues that exhibit higher entropy in the later vs. the earlier timepoint (negative values denote the opposite). Significant ($p<0.001$) values are shown in Gray. ***Panel A bottom*****:** Corresponding plasma V3 amino acid frequencies at these two timepoints. Positive, negative and neutrally-charged residues are in pink, grey and black respectively, with significantly-changing sites in bright colors and non-significantly-changing sites in dull colors. ***Panel B*****:** Corresponding analysis for the latest (R+12M) vs. earliest (R+5M) plasma V3 sequences from the recipient.

4.5. Discussion

We retrospectively identified a putative case of X4 HIV-1 transmission from a CCR5-wt/wt donor to a CCR5-Δ32/Δ32 recipient in a cohort of injection drug users and characterized this event by longitudinal V3 deep sequencing and ancestral phylogenetic reconstruction. A total of 10 independent ancestral reconstructions performed using plasma-derived deep sequences strongly suggested that the donor was originally productively infected with a single X4 virus, and that a variant differing by only one V3 residue was transmitted to the recipient. Although inference of a single T/F virus is consistent with the severe genetic bottleneck at transmission [21,24], this number is lower than the average for infection via injection drug use (N=3) [25]. That both donor and recipient acquired a single T/F virus suggests that this low multiplicity of infection (MOI) is not attributable to the latter's CCR5-Δ32/Δ32 genotype. Rather, we hypothesize that this is due to the transmission of an X4 (rather than the more common R5) variant, though additional studies would be required to confirm a relationship between HIV-1 coreceptor usage and MOI in different risk groups.

The virus that was transmitted to the recipient represented a minority (0.1%) variant in donor plasma but a co-dominant (33.5%) variant in PBMC at the timepoint closest to transmission, suggesting that productive HIV-1 transmission occurred via transfer of an infected cell. Co-dominance of the T/F variant in donor PBMC also suggests that it may have represented an archived variant maintained in one or more cell types but not propagated to plasma. That the donor and recipient T/F viruses differed by only one V3 amino acid (at codon 24) is also consistent with the observation that transmitted viruses tend to be genetically closer to "ancestral" donor viruses than those present in plasma at the time transmission [61,62]. Indeed, the genetic distance between the T/F virus and that originally acquired by the donor (0.01704 sub/nt site) was marginally yet significantly lower than that separating the T/F and donor plasma viruses present at transmission (0.01711 sub/nt site) ($p<0.0001$, Wilcoxon one-sample test), supporting the transmission bottleneck driving HIV-1 evolution towards ancestral states [61,62].

Donor and recipient within-host phylogenies were initially starlike (consistent with rapid evolution of a single T/F virus) but later exhibited a more asymmetrical appearance (characteristic of subsequent extinction of viral lineages via immune selection) [63]. Indeed, evidence of selection by neutralizing antibodies was observed in both hosts in the form of diversity loss at V3 codon 18, the final residue of the GPGR "crown motif" in a neutralizing antibody epitope [64,65]. That this occurred in both hosts is notable as it supports reproducible pathways and timecourse of antibody-driven escape in individuals acquiring genetically-similar viral strains [66].

The observation that the donor exhibited an extremely low nadir CD4+ T-cell count (20 cells/mm^3) within ~1.5 years of infection (whereas the recipient's CD4 counts were generally preserved) is consistent with rapid untreated HIV-1 progression in CCR5-wt/wt [41,67] - but not CCR5-Δ32/Δ32 [46] - individuals who acquire X4 infections (neither host expressed classical "protective" HLA class I alleles such as HLA-B*57 [68]). Consistent with X4-to-R5 "reversions" reported in CCR5-wt/wt hosts [69], R5 variants gradually emerged alongside their X4 counterparts beginning at 10 months post-infection in the donor and steadily increased to 60% frequency by 52 months post-infection. This supports higher fitness of R5 strains compared to their X4 counterparts [15], a pathway that was not available in the homozygous CCR5-Δ32/Δ32 recipient. In contrast, V3 evolution in the latter was characterized by an "exploration" of different genetic strategies that maintained the X4 phenotype: for example, at V3 codon 25 the initial X4-associated arginine gave rise to a lysine (observed at 63.5% frequency 12 months post-infection). In fact, initial rates of within-host V3 sequence divergence from the T/F strain were actually greater for the recipient compared to the donor, indicating that V3 evolution was not constrained by the latter's CCR5-Δ32/Δ32 genotype.

Some limitations of this study merit mention. We did not HIV-1 genotype the entire VIDUS cohort nor did we confirm direct transmission between donor and recipient – as such, we cannot rule out an intermediary host or a third individual who infected both donor and recipient at different times. Similarly, we are inferring HIV-1 transmission by injection drug use, but sexual transmission cannot be ruled out. Nevertheless, among 239 VIDUS patients examined, the individual harboring the next closest HIV-1 sequence to our CCR5-Δ32/Δ32 recipient exhibited mean genetic distances >5.6-fold (range 1.4-

107) greater than those separating our putative donor and recipient. The possibility of PCR amplification and/or template resampling bias is another limitation, as viral templates were not directly quantified (*e.g.* using "primer ID" techniques [70]). Nevertheless our approach of triplicate amplifying each extract, quantifying DNA and pooling resulting amplicons equally prior to deep sequencing reduces this bias [27]. Related to this issue is our reconstruction of donor and recipient T/F viruses and our inference that the recipient's infection was initiated by transfer of an infected cell from the donor. Several lines of evidence support this observation as not simply an artefact of PCR amplification bias. Firstly, the composition (in terms of unique sequences and their frequencies) is consistent across plasma and PBMC compartments over time in both hosts. Second, within-host plasma and PBMC phylogenies exhibit characteristic tree shapes over time. Thirdly and critically, ancestral reconstructions were performed using donor and recipient *plasma* sequences only – but despite this, we consistently reconstructed a T/F virus sequence that represented a minority variant in plasma yet co-dominated in PBMC. Finally, replicate sequence data from 8 of 13 samples analyzed yielded concordant results: for example, in the donor, the frequencies of the inferred T/F virus 7 months prior to transmission were 0.4% and 3.5% in plasma and PBMC (compared to replicate values of 0.2% and 1.1% in the same samples). Due to direct freezing of PBMC pellets, it was not possible to separate cell types prior to deep sequencing, as such, we cannot speculate what type of infected cell may have initiated productive infection in the recipient. Lastly, as this is a descriptive study of a rare transmission pair, our ability to draw broad conclusions is limited. Nevertheless, studies such as this represent an important step towards bridging our understanding of HIV-1 evolution at the within- and between-host scales, a key current challenge in phylogenetics [71].

In conclusion, this study highlights the utility of deep sequencing paired with phylogenetic ancestral reconstruction to study HIV-1 transmission dynamics and intra-/inter-host evolution. Our study of this rare donor/recipient pair supports infected cell transfer as the mode of transmission and highlights the influence of host genetics on HIV-1 pathogenesis and evolution.

4.6. References

1. Klatzmann D, Champagne E, Chamaret S, Gruest J, Guetard D, et al. (1984) T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature 312: 767-768.

2. Feng Y, Broder CC, Kennedy PE, Berger EA (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science 272: 872-877.

3. Samson M, Labbe O, Mollereau C, Vassart G, Parmentier M (1996) Molecular cloning and functional expression of a new human CC-chemokine receptor gene. Biochemistry 35: 3362-3367.

4. Berger EA, Doms RW, Fenyo EM, Korber BT, Littman DR, et al. (1998) A new classification for HIV-1. Nature 391: 240.

5. Coakley E, Reeves JD, Huang W, Mangas-Ruiz M, Maurer I, et al. (2009) Comparison of human immunodeficiency virus type 1 tropism profiles in clinical samples by the Trofile and MT-2 assays. Antimicrob Agents Chemother 53: 4686-4693.

6. De Jong JJ, De Ronde A, Keulen W, Tersmette M, Goudsmit J (1992) Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. J Virol 66: 6777-6780.

7. Fouchier RA, Groenink M, Kootstra NA, Tersmette M, Huisman HG, et al. (1992) Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. J Virol 66: 3183-3187.

8. Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R (2007) Bioinformatics prediction of HIV coreceptor usage. Nat Biotechnol 25: 1407-1410.

9. Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, et al. (2003) Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. J Virol 77: 13376-13388.

10. Berger EA, Murphy PM, Farber JM (1999) Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol 17: 657-700.

11. Carrington M, Dean M, Martin MP, O'Brien SJ (1999) Genetics of HIV-1 infection: chemokine receptor CCR5 polymorphism and its consequences. Hum Mol Genet 8: 1939-1945.

12. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, et al. (2008) Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A 105: 7552-7557.

13. Feng Z, Dubyak GR, Lederman MM, Weinberg A (2006) Cutting edge: human beta defensin 3--a novel antagonist of the HIV-1 coreceptor CXCR4. J Immunol 177: 782-786.

14. Agace WW, Amara A, Roberts AI, Pablos JL, Thelen S, et al. (2000) Constitutive expression of stromal derived factor-1 by mucosal epithelia and its role in HIV transmission and propagation. Curr Biol 10: 325-328.

15. Grivel JC, Shattock RJ, Margolis LB (2011) Selective transmission of R5 HIV-1 variants: where is the gatekeeper? J Transl Med 9 Suppl 1: S6.

16. Liu Y, Curlin ME, Diem K, Zhao H, Ghosh AK, et al. (2008) Env length and N-linked glycosylation following transmission of human immunodeficiency virus Type 1 subtype B viruses. Virology 374: 229-233.

17. Ceresola ER, Nozza S, Sampaolo M, Pignataro AR, Saita D, et al. (2015) Performance of commonly used genotypic assays and comparison with phenotypic assays of HIV-1 coreceptor tropism in acutely HIV-1-infected patients. J Antimicrob Chemother.

18. Raymond S, Saliou A, Nicot F, Delobel P, Dubois M, et al. (2013) Characterization of CXCR4-using HIV-1 during primary infection by ultra-deep pyrosequencing. J Antimicrob Chemother 68: 2875-2881.

19. Koot M, Keet IP, Vos AH, de Goede RE, Roos MT, et al. (1993) Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS. Ann Intern Med 118: 681-688.

20. Richman DD, Bozzette SA (1994) The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression. J Infect Dis 169: 968-974.

21. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, et al. (2009) Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J Exp Med 206: 1273-1289.

22. Bar KJ, Tsao CY, Iyer SS, Decker JM, Yang Y, et al. (2012) Early low-titer neutralizing antibodies impede HIV-1 replication and select for virus escape. PLoS Pathog 8: e1002721.

23. Carlson JM, Schaefer M, Monaco DC, Batorsky R, Claiborne DT, et al. (2014) HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. Science 345: 1254031.

24. Derdeyn CA, Decker JM, Bibollet-Ruche F, Mokili JL, Muldoon M, et al. (2004) Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. Science 303: 2019-2022.

25. Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, et al. (2010) Wide variation in the multiplicity of HIV-1 infection among injection drug users. J Virol 84: 6241-6247.

26. McCloskey RM, Liang RH, Harrigan PR, Brumme ZL, Poon AF (2014) An evaluation of phylogenetic methods for reconstructing transmitted HIV variants using longitudinal clonal HIV sequence data. J Virol 88: 6181-6194.

27. Poon AF, Swenson LC, Bunnik EM, Edo-Matas D, Schuitemaker H, et al. (2012) Reconstructing the dynamics of HIV evolution within hosts from serial deep sequence data. PLoS Comput Biol 8: e1002753.

28. Poon AF, McGovern RA, Mo T, Knapp DJ, Brenner B, et al. (2011) Dates of HIV infection can be estimated for seroprevalent patients by coalescent analysis of serial next-generation sequencing data. AIDS 25: 2019-2026.

29. Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, et al. (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 86: 367-377.

30. Martinson JJ, Chapman NH, Rees DC, Liu YT, Clegg JB (1997) Global distribution of the CCR5 gene 32-basepair deletion. Nat Genet 16: 100-103.

31. Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, et al. (1996) The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. Nat Med 2: 1240-1243.

32. Marmor M, Sheppard HW, Donnell D, Bozeman S, Celum C, et al. (2001) Homozygous and heterozygous CCR5-Delta32 genotypes are associated with resistance to HIV infection. J Acquir Immune Defic Syndr 27: 472-481.

33. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, et al. (1996) Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382: 722-725.

34. Biti R, Ffrench R, Young J, Bennetts B, Stewart G, et al. (1997) HIV-1 infection in an individual homozygous for the CCR5 deletion allele. Nat Med 3: 252-253.

35. O'Brien TR, Winkler C, Dean M, Nelson JA, Carrington M, et al. (1997) HIV-1 infection in a man homozygous for CCR5 delta 32. Lancet 349: 1219.

36. Balotta C, Bagnarelli P, Violin M, Ridolfo AL, Zhou D, et al. (1997) Homozygous delta 32 deletion of the CCR-5 chemokine receptor gene in an HIV-1-infected patient. AIDS 11: F67-71.

37. Gorry PR, Zhang C, Wu S, Kunstman K, Trachtenberg E, et al. (2002) Persistence of dual-tropic HIV-1 in an individual homozygous for the CCR5 Delta 32 allele. Lancet 359: 1832-1834.

38. Heiken H, Becker S, Bastisch I, Schmidt RE (1999) HIV-1 infection in a heterosexual man homozygous for CCR-5 delta32. AIDS 13: 529-530.

39. Kuipers H, Workman C, Dyer W, Geczy A, Sullivan J, et al. (1999) An HIV-1-infected individual homozygous for the CCR-5 delta32 allele and the SDF-1 3'A allele. AIDS 13: 433-434.

40. Theodorou I, Meyer L, Magierowska M, Katlama C, Rouzioux C (1997) HIV-1 infection in an individual homozygous for CCR5 delta 32. Seroco Study Group. Lancet 349: 1219-1220.

41. Sheppard HW, Celum C, Michael NL, O'Brien S, Dean M, et al. (2002) HIV-1 infection in individuals with the CCR5-Delta32/Delta32 genotype: acquisition of syncytium-inducing virus at seroconversion. J Acquir Immune Defic Syndr 29: 307-313.

42. Iversen AK, Christiansen CB, Attermann J, Eugen-Olsen J, Schulman S, et al. (2003) Limited protective effect of the CCR5Delta32/CCR5Delta32 genotype on human immunodeficiency virus infection incidence in a cohort of patients with hemophilia and selection for genotypic X4 virus. J Infect Dis 187: 215-225.

43. Bhaskaran K, Brettle R, Porter K, Walker AS, Collaboration C (2004) Systemic non-Hodgkin lymphoma in individuals with known dates of HIV seroconversion: incidence and predictors. AIDS 18: 673-681.

44. Gray L, Churchill MJ, Keane N, Sterjovski J, Ellett AM, et al. (2006) Genetic and functional analysis of R5X4 human immunodeficiency virus type 1 envelope glycoproteins derived from two individuals homozygous for the CCR5delta32 allele. J Virol 80: 3684-3691.

45. Oh DY, Jessen H, Kucherer C, Neumann K, Oh N, et al. (2008) CCR5Delta32 genotypes in a German HIV-1 seroconverter cohort and report of HIV-1 infection in a CCR5Delta32 homozygous individual. PLoS One 3: e2747.

46. Henrich TJ, Hanhauser E, Hu Z, Stellbrink HJ, Noah C, et al. (2015) Viremic control and viral coreceptor usage in two HIV-1-infected persons homozygous for CCR5 Delta32. AIDS.

47. Strathdee SA, Patrick DM, Currie SL, Cornelisse PG, Rekart ML, et al. (1997) Needle exchange is not enough: lessons from the Vancouver injecting drug use study. AIDS 11: F59-65.

48. Woods CK, Brumme CJ, Liu TF, Chui CK, Chu AL, et al. (2012) Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. J Clin Microbiol 50: 1936-1942.

49. Pond SL, Frost SD, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21: 676-679.

50. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59: 307-321.

51. Fourment M, Gibbs MJ (2006) PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. BMC Evol Biol 6: 1.

52. Brumme ZL, Chan KJ, Dong W, Hogg R, O'Shaughnessy MV, et al. (2001) CCR5Delta32 and promoter polymorphisms are not correlated with initial virological or immunological treatment response. AIDS 15: 2259-2266.

53. Cotton LA, Abdur Rahman M, Ng C, Le AQ, Milloy MJ, et al. (2012) HLA class I sequence-based typing using DNA recovered from frozen plasma. J Immunol Methods 382: 40-47.

54. Knapp DJ, McGovern RA, Poon AF, Zhong X, Chan D, et al. (2014) "Deep" sequencing accuracy and reproducibility using Roche/454 technology for inferring co-receptor usage in HIV-1. PLoS One 9: e99508.

55. Swenson LC, Dong, W., Mo, T., Woods, C., Zhong, X., Thielen, A., Jensen, M., Biswas, P., Ellery, S., Lewis, M., James, I., Chapman, D., Valdez, H., Harrigan, R. “Deep” sequencing to identify treatment-experienced patients who respond to maraviroc (MVC); 2009; Cologne, Germany. European AIDS Clinical Society.

56. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.

57. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792-1797.

58. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10: 512-526.

59. Plummer M, Best, N., Cowles, K., Vines, K. (2006) CODA: convergence diagnosis and output analysis for MCMC. R News. pp. 7-11.

60. McGovern RA, Thielen A, Mo T, Dong W, Woods CK, et al. (2010) Population-based V3 genotypic tropism assay: a retrospective analysis using screening samples from the A4001029 and MOTIVATE studies. AIDS 24: 2517-2525.

61. Herbeck JT, Nickle DC, Learn GH, Gottlieb GS, Curlin ME, et al. (2006) Human immunodeficiency virus type 1 env evolves toward ancestral states upon transmission to a new host. J Virol 80: 1637-1644.

62. Sagar M, Laeyendecker O, Lee S, Gamiel J, Wawer MJ, et al. (2009) Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. J Infect Dis 199: 580-589.

63. Herbeck JT, Rolland M, Liu Y, McLaughlin S, McNevin J, et al. (2011) Demographic processes affect HIV-1 evolution in primary infection before the onset of selective processes. J Virol.

64. Stanfield RL, Gorny MK, Williams C, Zolla-Pazner S, Wilson IA (2004) Structural rationale for the broad neutralization of HIV-1 by human monoclonal antibody 447-52D. Structure 12: 193-204.

65. Zolla-Pazner S, Zhong P, Revesz K, Volsky B, Williams C, et al. (2004) The cross-clade neutralizing activity of a human monoclonal antibody is determined by the GPGR V3 motif of HIV type 1. AIDS Res Hum Retroviruses 20: 1254-1258.

66. Draenert R, Allen TM, Liu Y, Wrin T, Chappey C, et al. (2006) Constraints on HIV-1 evolution and immunodominance revealed in monozygotic adult twins infected with the same virus. J Exp Med 203: 529-539.

67. Raymond S, Delobel P, Mavigner M, Cazabat M, Encinas S, et al. (2010) CXCR4-using viruses in plasma and peripheral blood mononuclear cells during primary HIV-1 infection and impact on disease progression. AIDS 24: 2305-2312.

68. O'Brien SJ, Gao X, Carrington M (2001) HLA and AIDS: a cautionary tale. Trends Mol Med 7: 379-381.

69. Baroncelli S, Galluzzo CM, Andreotti M, Pirillo MF, Fragola V, et al. (2013) HIV-1 coreceptor switch during 2 years of structured treatment interruptions. Eur J Clin Microbiol Infect Dis 32: 1565-1570.

70. Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R (2011) Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci U S A 108: 20166-20171.

71. Frost SDW, Pybus, O. G., Gog, J. R., Viboud, C., Bonhoeffer, S., Bedford, T. (2014) Eight challenges in phylodynamic inference. Epidemics.

Chapter 5.

Concluding remarks

The goal of my MSc was to study HIV evolution in response to selection pressures imposed by the host. Achieving an improved understanding of HIV genetic diversity and evolution may assist in the development of future biomedical intervention strategies such as an HIV vaccine.

The original research project undertaken in this thesis was to characterize a rare HIV-1 transmission pair using next generation sequencing and phylogenetic approaches. Using Roche 454 next generation sequencing, HIV-1 sequences derived from the CCR5-wt/wt donor and CCR5-Δ32/Δ32 homozygous recipient were characterized longitudinally and used to study HIV-1 evolution within and between these individuals. Sequence comparison revealed differential rates and patterns of evolution in each patient. For example, many of the descendants of the original CXCR4-using strain that infected the CCR5-wt/wt donor gradually evolved towards R5 usage over a period of 4 years (though a substantial minority of circulating viruses retained X4 status). Although the timing of X4-to-R5 "reversion" in this individual was similar to that reported by Baroncelli et al [1], more cases need to be examined to fully understand this phenomenon. Broadening of coreceptor use could allow HIV-1 to take advantage of a greater pool of target cells (*i.e.* those that express CCR5/CD4 as well as those that express CXCR4/CD4), a strategy that would aid the dissemination of HIV-1 and have implications for disease pathogenesis and progression. A starting point to investigate this question would be to use similar methods described in my thesis (next-generation sequencing and phylogenetic methods) to study the dynamics of HIV-1 evolution in individuals who undergo X4-to-R5 phenotypic switch. In doing so, we might be able to learn how often this occurs, how long the X4-to-R5 "reversion" takes, and the evolutionary pathways whereby R5 viruses arise in X4 infected individuals. For example,

it is hypothesized that the R5-to-X4 phenotypic switch that occurs in approximately ~50% of persons who are infected with an R5 strain could be a predetermined process based on viral sequence that occurs at a consistent rate in all individuals, or that it may be a stochastic event [2,3]. The former model is characterized by the accumulation of mutations where intermediary strains are more fit than their previous counterparts and can occur over a number of years [2]. The latter model (also known as the "fitness valley" model) assumes that intermediary strains are less fit than either R5 or X4 strain and traversal of this "fitness valley" can only be achieved by the accumulation of multiple mutations in a short time [2,3]. In my thesis, the gradual and incomplete reversion from X4-to-R5 in the CCR5-wt/wt donor suggests that this was a gradual event, potentially characterized by greater fitness of intermediary R5 variants [2]. However, whether this is true for all cases of X4-to-R5 phenotypic switch remains to be determined.

Using ancestral phylogenetic reconstruction, we were able to infer the V3 sequences of the viruses that established a productive infection both the donor and recipient. This allowed us to track the presence of these sequences longitudinally in both hosts. Our results suggested that transmission of X4 HIV-1 from the donor to recipient potentially occurred via cell-to-cell contact. Cell-to-cell spread/transmission (or transfer) occurs when an HIV-1 infected cell comes into close proximity to an uninfected cell via binding to multiple host and viral factors to create a viral synapse [4]. The viral synapse allows for the fast and efficient transfer of HIV-1 virions to the uninfected target cell [4-6]

Although sometimes used interchangeably it is important to note that cell-to-cell (CtC) spread and transmission refer to two different processes. The former refers to the dissemination of HIV-1 within a single individual (*i.e.* from one HIV-1 infected cell to another in a given host). The latter refers to the transmission of HIV-1 from one individual to another. The mode of HIV-1 spread/transmission could have treatment and vaccine implications. For example, CtC spread may facilitate transfer of HIV-1 to cell types that are not classical targets for HIV-1 (*e.g.* monocytes [7]). These newly infected cells may then travel to other anatomical compartments (*e.g.* the testes [8] and central nervous system [9]) and establish viral reservoirs [10] that can remain undetected. In addition, drug penetration into these compartments may also be incomplete, reducing the effectiveness of antiretroviral therapy (ART) in these areas [11]. For example, *in vitro*

analysis revealed that tenofovir was able to decrease cell free infection of PBMCs by 30-fold, whereas it was only able to decrease cell-to-cell infection of PBMCs by 2-fold [12]. In addition, CtC transfer of HIV-1 may also act as mechanism to evade neutralizing antibodies (bNAbs) elicited by vaccines or entry inhibitors (*e.g.* T-20 [enfuvirtide]) due to the extreme proximity of cell membranes in the viral synapse. Therefore, it could undermine the efficacy of a vaccine or treatment [13]. For example, Abela et al demonstrated that CD4 binding bNAb (VRC01) and gp41 entry inhibitor (T-20) had a decreased activity in blocking HIV-1 infection in a cell-to-cell infection model versus a cell free infection model [13].

The extent to which CtC transmission occurs *in vivo* is largely unknown. If it does occur frequently there are many more questions that need to be answered. For example, how does disease progression and outcome differ from infection via cell free virus? Also, are different routes of transmission (*e.g.* IDU vs heterosexual transmission) more prone to cell-to-cell transmission?

Taken together, the research presented here further highlight the complex interplay between host pressures and HIV-1 escape and evolution. Further research to elucidate the different mechanisms of HIV-1 transmission, and their possible implications for future intervention strategies, is warranted.

5.1. References

1. Baroncelli S, Galluzzo CM, Andreotti M, Pirillo MF, Fragola V, et al. (2013) HIV-1 coreceptor switch during 2 years of structured treatment interruptions. Eur J Clin Microbiol Infect Dis 32: 1565-1570.

2. Poon AF, Swenson LC, Bunnik EM, Edo-Matas D, Schuitemaker H, et al. (2012) Reconstructing the dynamics of HIV evolution within hosts from serial deep sequence data. PLoS Comput Biol 8: e1002753.

3. da Silva J, Wyatt SK (2014) Fitness valleys constrain HIV-1's adaptation to its secondary chemokine coreceptor. J Evol Biol.

4. Sattentau Q (2008) Avoiding the void: cell-to-cell spread of human viruses. Nat Rev Microbiol 6: 815-826.

5. Jung A, Maier R, Vartanian JP, Bocharov G, Jung V, et al. (2002) Recombination: Multiply infected spleen cells in HIV patients. Nature 418: 144.

6. Agosto LM, Uchil PD, Mothes W (2015) HIV cell-to-cell transmission: effects on pathogenesis and antiretroviral therapy. Trends Microbiol.

7. Cameron PU, Lowe MG, Sotzik F, Coughlan AF, Crowe SM, et al. (1996) The interaction of macrophage and non-macrophage tropic isolates of HIV-1 with thymic and tonsillar dendritic cells in vitro. J Exp Med 183: 1851-1856.

8. Le Tortorec A, Le Grand R, Denis H, Satie AP, Mannioui K, et al. (2008) Infection of semen-producing organs by SIV during the acute and chronic stages of the disease. PLoS One 3: e1792.

9. Canestri A, Lescure FX, Jaureguiberry S, Moulignier A, Amiel C, et al. (2010) Discordance between cerebral spinal fluid and plasma HIV replication in patients with neurological symptoms who are receiving suppressive antiretroviral therapy. Clin Infect Dis 50: 773-778.

10. Costiniuk CT, Jenabian MA (2014) Cell-to-cell transfer of HIV infection: implications for HIV viral persistence. J Gen Virol 95: 2346-2355.

11. Agosto LM, Uchil PD, Mothes W (2015) HIV cell-to-cell transmission: effects on pathogenesis and antiretroviral therapy. Trends Microbiol 23: 289-295.

12. Sigal A, Kim JT, Balazs AB, Dekel E, Mayo A, et al. (2011) Cell-to-cell spread of HIV permits ongoing replication despite antiretroviral therapy. Nature 477: 95-98.

13. Abela IA, Berlinger L, Schanz M, Reynell L, Gunthard HF, et al. (2012) Cell-cell transmission enables HIV-1 to evade inhibition by potent CD4bs directed antibodies. PLoS Pathog 8: e1002634.