

Investigation of Host and Viral Genetic Factors Influencing HIV-1 Evolution across North America: Implications for Vaccine Design

**by
Laura Anne Cotton**

B.Sc., University of British Columbia, 2009

Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Masters of Public Health

in the
Population Health Program
Faculty of Health Science

**© Laura Anne Cotton 2012
SIMON FRASER UNIVERSITY
Spring 2012**

All rights reserved.
However, in accordance with the *Copyright Act of Canada*, this work may
be reproduced, without authorization, under the conditions for
“Fair Dealing.” Therefore, limited reproduction of this work for the
purposes of private study, research, criticism, review and news reporting
is likely to be in accordance with the law, particularly if cited appropriately.

Approval

Name: Laura Cotton
Degree: Masters of Public Health (Population Health)
Title of Thesis: *Investigation of Host and Viral Genetic Factors
Influencing HIV-1 Evolution across North America:
Implications for Vaccine Design*

Examining Committee:

Chair Of Defense: **Dr. Charles Goldsmith**
Professor
Faculty of Health Sciences

Senior Supervisor: **Dr. Zabrina Brumme**
Assistant Professor
Faculty of Health Sciences

Supervisor: **Dr. Mark Brockman**
Associate Professor
Faculty of Health Sciences

Supervisor: **Dr. Art Poon**
Associate Research Scientist
BC Centre for Excellence in HIV/AIDS

External: **Dr. Robert Hogg**
Professor
Faculty of Health Sciences

Date Defended/Approved: April 16, 2012

Partial Copyright License



The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the "Institutional Repository" link of the SFU Library website (www.lib.sfu.ca) at <http://summit/sfu.ca> and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2011

Abstract

Human leukocyte antigen (HLA) class I restricted cytotoxic T-lymphocyte (CTL) responses drive HIV-1 evolution through the selection of immune escape mutations, however, the extent to which population-level patterns of immune escape have changed over the course of the HIV-1 epidemic in North America remains incompletely known. The objective of this thesis was to explore how immune selection pressures mediated through host HLA-restricted CTL responses have shaped the genomic and functional evolution of the HIV-1 *gag* gene over the course of the epidemic in North America. Results support the continued dynamic adaptation of HIV-1 as it passes through human hosts rather than the substantial accumulation of escape mutations over the course of the epidemic in North America. Additionally, only modest increases in *gag*-mediated replication capacity between pre- and post-1985 sequences were observed. Although the mechanism(s) behind these increases remain unknown, they may be attributable to factors other than CTL-driven immune responses.

Keywords: human immunodeficiency virus type 1 (HIV-1); human leukocyte antigen (HLA); cytotoxic T-lymphocyte (CTL); immune escape; replication capacity; historic

Acknowledgements

Firstly, I would like to sincerely thank my supervisor, Dr. Zabrina Brumme, for all of her support and encouragement as well as the many opportunities she has given me to present this research. I would also like to thank my committee members, Dr. Mark Brockman and Dr. Art Poon, for their additional support and mentorship over the last year. Furthermore, none of this research would have been possible without the guidance and advice of my fellow students and coworkers in the laboratory: Dr. Bemulu Baraki, Dr. Denis Chopera, Tristan Markle, Eric Martin, Manal Rahman, Kali Penney, Tallie Kuang and Anh Le. A special thanks to Dr. Denis Chopera and Kali Penney and for the contributions they have made to this project and their continual positive influence.

Table of Contents

Approval.....	ii
Partial Copyright License	iii
Abstract.....	iv
Acknowledgements.....	v
Table of Contents.....	vi
List of Tables.....	viii
List of Figures	ix
1. Chapter 1: Introduction	1
1.1. The Emergence of Aquired Immunodeficiency Syndrome and Discovery of Human Immunodeficiency Virus	1
1.2. The Origin of Human Immunodeficiency Virus	3
1.3. Nomenclature and Global Diversity	6
1.4. Global Impact.....	8
1.5. HIV Prevention and Treatment	9
1.6. HIV Pathogenesis	11
1.7. Brief Overview of the HIV Life Cycle and Genomic Organization	12
1.8. Host Immune Responses to HIV Infection	14
1.9. Overview of Cytotoxic T-lymphocyte Mediated Immune Escape.....	16
1.10. Thesis Objectives and Organization	20
1.11. Overview of Data Sources and Population-based Cohorts.....	21
1.12. References.....	23
2. Chapter 2: HLA Class I Sequence-based Typing using DNA Recovered from Frozen Plasma.....	35
2.1. Abstract.....	35
2.2. Introduction	36
2.3. Methods	37
2.3.1. Source of samples.....	37
2.3.2. DNA extraction and quantification	37
2.3.3. Amplification of HLA-A, B and C loci	38
2.3.4. DNA sequencing	41
2.3.5. Sequence analysis and HLA allele interpretation.....	42
2.4. Results and Discussion.....	43
2.5. Conclusions	48
2.6. References.....	49
3. Chapter 3: Immune Escape Pathways and Replication Capacity in HIV-1 Gag have Remained Largely Consistent over the Past 30 Years of the Epidemic in North America	52
3.1. Abstract.....	52
3.2. Introduction	53
3.3. Methods	57
3.3.1. Cohort description	57
3.3.2. Total nucleic acid isolation	58
3.3.3. Amplification of HIV-1 gag.....	58

3.3.4.	Amplification of HLA-A, B and C loci	60
3.3.5.	Sequencing	60
3.3.6.	Identification of HLA-associated polymorphisms in the HIV-1 gag gene	61
3.3.7.	Immune escape maps	62
3.3.8.	Estimating the time of the most recent common ancestor	63
3.3.9.	Generation of HIV-1 gag clones	63
3.3.10.	Generation of chimeric virus and replication assays	64
3.3.11.	Statistical analyses of replication capacity	66
3.4.	Results	66
3.4.1.	Identification of HLA-associated gag polymorphisms in the historic cohort	66
3.4.2.	Timing of the most recent common ancestor	73
3.4.3.	Replication capacity in gag between historic and modern cohorts	73
3.5.	Discussion.....	86
3.6.	Conclusion	93
3.7.	References.....	93
4.	Chapter 4: General Conclusions	101
4.1.	Public Health Implications.....	101
4.2.	References.....	104
	Appendices.....	106
Appendix A.	HLA Class I Sequencing Reference Standards.....	107

List of Tables

Table 2.1. Nested PCR and sequencing primers used for HLA-A, B and C typing.	40
Table 2.2. HLA class I alleles represented in the test panel.	47
Table 3.1. Amino acids in gag associated with either an increase or decrease in in vitro viral replication capacity.	85

List of Figures

Figure 1.1. Distribution of HIV-1 group M subtypes.	8
Figure 2.1. Amplification of HLA-A, B and C using PBMC vs. plasma-derived genomic DNA.	44
Figure 2.2. Quality comparison of sequence chromatograms for HLA-A, B and C alleles from matched PBMC vs. plasma-derived amplicons.	45
Figure 3.1. Historic gag escape map.	68
Figure 3.2. Frequency of HLA-associated polymorphisms identified in gag in the general population.	70
Figure 3.3. Median fold-change in HLA-associated polymorphisms identified in gag compared to fold-difference in HLA class I allele frequencies.	72
Figure 3.4. Confirmation of gag clone origin by phylogenetic analysis.	76
Figure 3.5. Confirmation of re-sequenced gag clone origin after recombination and chimeric virus generation by phylogenetic analysis.	79
Figure 3.6. Comparison of HIV-1 replication capacity over time.	81
Figure 3.7. HLA-associated changes in gag-mediated HIV-1 replication capacity over time.	83

1. Chapter 1: Introduction

1.1. The Emergence of Acquired Immunodeficiency Syndrome and Discovery of Human Immunodeficiency Virus

Prior to 1980, a little over thirty years ago, acquired immunodeficiency syndrome (AIDS) and its etiologic agent, human immunodeficiency virus (HIV), were unknown. The first recognized cases of AIDS appeared between October 1980 and May 1981 when hospitals throughout Los Angeles began reporting cases of *Pneumocystis carinii* pneumonia (PCP), a rare form of pneumonia, in young, homosexual men who had previously been healthy (1). In a report released on June 5th 1981, in *Morbidity and Mortality Weekly Report* (MMWR), the U.S. Centres for Disease Control and Prevention (CDC) confirmed the deaths of two of these individuals (1). Less than a month later, the CDC released another report of 10 additional cases of PCP diagnosed in Los Angeles and the San Francisco Bay area as well as 26 cases of Kaposi's sarcoma (KS), a rare form of skin cancer, appearing across New York and California (2). In the months that followed hundreds of new cases of PCP, KS and other rare opportunistic infections typically associated with immunosuppression were reported across the US (3). Initially, the majority of cases were reported in homosexual men, many dying within 20-24 months of diagnosis; however, later reports indicated that heterosexual Haitian immigrants, intravenous drug users and hemophiliacs were also at risk of developing what would become known as AIDS (4, 5).

Shortly after the sudden increases in PCP and KS incidence were reported, blood work revealed low numbers of CD4+ T helper cells in these young men affected (6-9). It appeared that an unknown “agent” was depleting the immune system of these individuals, leaving them vulnerable to infection (10). By early 1982 this became known as acquired immunodeficiency syndrome or AIDS. With the cause of AIDS still unknown, the medical community remained divided as to whether an infectious agent (11), environmental factor (9) or combination of factors (12) were responsible for the observed symptoms.

On May 20th 1983, two separate research teams led by Robert Gallo and Luc Montagnier published reports in the journal *Science* declaring the identification of a novel retrovirus in patients displaying characteristics associated with AIDS (13, 14). Both papers discussed the ability of this novel virus to infect T cells and its close similarity to the human T-cell leukemia virus (HTLV) (13, 14). Over the next several years, more reports surfaced strengthening the evidence of a viral agent as the cause of AIDS (15-21), however, by 1985 new research was beginning to show that this novel retrovirus was not closely related to HTLV (22). In January of 1985 a complete genomic sequence was published in the journal *Nature* highlighting the differences between HTLV-III (the original proposed name for HIV-1) and other members of the HTLV family (23). Rather, this novel retrovirus displayed characteristics similar to members of the genus *Lentiviridae* belonging to the Retroviridae family (10). Subsequently, the name human immunodeficiency virus or HIV was agreed upon. In less than five years the medical community characterized the agent of an emerging epidemic, an impressive feat for the scientific technology of the time.

1.2. The Origin of Human Immunodeficiency Virus

Current evidence suggests that HIV (HIV-1 and HIV-2) originated in west-central Africa from simian immunodeficiency virus (SIV) infections of non-human primates, entering the human population as the result of multiple zoonotic transmission events over the last 100 years (24). Using phylogenetic methods, HIV-1 group M, the group responsible for a majority of the HIV infections worldwide, and group N are found to be most closely related to SIVcpz commonly found in the chimpanzee subspecies *Pan troglodytes troglodytes* (25-27), while HIV-1 group O and the newly identified group P are most closely related to SIVgor found in Gorillas (*Gorilla gorilla*) (28, 29). Similarly, HIV-2 is most closely related to SIVsm isolated from the sooty mangabey (*Cercocebus atys*) (30). Sequence analysis has revealed that the most diverse forms of HIV are found in Democratic Republic of Congo (DRC) corresponding with the natural habitat of these primates and further supporting the hypothesis that the HIV epidemic began in west-central Africa (31). It is believed that SIV may have jumped species as the result of close contact between humans and primates, most likely through open wounds, bites or handling of primate meat for human consumption (32, 33). Indeed, a considerable number of primates in west-central Africa are infected with SIV (34). As human populations have expanded into primate territory, human/primate contact has become more frequent, thus increasing the likelihood of transmission events.

To date, most estimates put the timing of the most recent common ancestor of the HIV-1 group M pandemic strains near the beginning of the twentieth century. One of the first studies to apply phylogenetic methods to a large data set in order to predict the timing of the most recent common ancestor was a paper published in 2000 by Korber et al. (35) in the journal *Science*. In this study, Korber et al. (35) used HIV-1 sequences

collected between 1983 and 1998 to generate a maximum-likelihood phylogenetic tree and date the internal nodes using molecular clock models. These models can assume a strict uniform rate of evolution as a function of time (“strict molecular clocks”) or allow rate variation across lineages (“relaxed molecular clock”) (35). Unique to this study was the inclusion of a historic HIV-1 sequence recovered from west-central Africa in 1959, the oldest known HIV-1 sequence on record (35). Historic sequences are key to phylogenetic reconstructions of the HIV-1 epidemic as they are used as “calibration” points to understand evolutionary distance as a unit of time (36). Based on their analyses, Korber et al. (35) dated the timing of the most recent common ancestor of the HIV-1 group M strains to 1931 [CI: 1915-1941]. In a subsequent paper published in 2008, Worobey et al. (36) used similar phylogenetic methods and an additional historic HIV-1 sequence recovered from a paraffin-embedded tissue sample from west-central Africa in 1960, to again estimate the timing of the most recent common ancestor of HIV-1 group M. Their analysis predicted the most recent common ancestor to date back to 1908 [CI: 1884-1924] (36). Importantly, Worobey et al. (36) also observed extensive genetic differences between sequences generated from the 1959 HIV-1 sample used by Korber et al. (35) and sequences generated from the 1960 HIV-1 sample, indicating that HIV-1 diversification in west-central Africa was well underway before AIDS was even recognized.

Although phylogenetic evidence indicates that HIV-1 group M, N, O and P strains entered the human population through four independent zoonotic transmission events from SIV, the reason why HIV-1 group M sparked a pandemic is still up for debate. In 1992 an article was published in Rolling Stone magazine outlining a hypothesis that suggested SIVcpz-contaminated oral polio vaccinations (OPV) administered in Central Africa in the late 1950s might have contributed to the spread of HIV-1 (37). Journalist

Edward Hooper and supporters of the OPV hypothesis claimed that chimpanzee kidney cells were used to propagate the polio vaccine stocks subsequently used in mass vaccination campaigns in central Africa around the late 1950s (38), coinciding with what was thought to be the earliest cases of HIV-1. Subsequent analysis of historic polio vaccine stocks from the late 1950s has revealed no contamination with SIV or HIV-1, while mitochondrial analysis has further confirmed that macaque rather than chimpanzee kidney cells were used to propagate the vaccine (39-41). Additionally, evidence generated in studies like the one by Worobey et al. (36) date the most recent common ancestor to around the turn of the twentieth century and indicate substantial genetic diversity prior to 1959, supporting the idea that HIV-1 was established within human populations well before polio vaccination campaigns in the late 1950s. Accumulating evidence has since resulted in rejection of the oral polio vaccine hypothesis (39-42).

Recently, a new theory put forward by Canadian professor Jacques Pepin suggests that the re-use of syringes and needles during large-scale public health treatment campaigns for endemic infectious diseases undertaken prior to 1951 may have inadvertently led to the transmission of HIV-1 in central Africa (43). Using hepatitis C virus (HCV) and human T cell lymphotropic virus type 1 (HTLV-1) infections as proxies for infection with HIV-1, Pepin et al. (43) found that pre-1951 intramuscular treatment for trypanosomiasis (parasitic infection) was associated with an increased risk of HCV and HTLV-1 infection. These results coupled with the observed excess number of deaths in trypanosomiasis patients treated prior to 1951 indicate that treatment campaigns using unsterilized equipment may have fueled the HIV-1 epidemic (43). Needless to say, the early spread of HIV-1 was likely influenced by multiple factors, potentially including public health campaigns, changing population dynamics, the rise of cities and increasing

global travel. Indeed, the right conditions needed to fuel an HIV-1 pandemic were all in place in the early part of the twentieth century.

1.3. Nomenclature and Global Diversity

On a global level, HIV ranks among the most diverse pathogens known. Belonging to the *Lentivirus* genus within the Retroviridae family, there are two recognized types of HIV: HIV-1, which is associated with the global HIV epidemic and HIV-2, which is predominantly restricted to West Africa and is not as easily transmitted (10, 44). HIV-1 and HIV-2 strains are further broken down into groups and subtypes based on phylogenetic relatedness (45, 46). For the purposes of this review, HIV-2 will not be discussed. Within HIV-1 there are 4 groups: M (main), N (non-main), O (outlier) (45) and the newly identified P (28), each the likely result of a unique zoonotic transmission event (as discussed in section 1.2) (24). Group M is responsible for the vast majority of HIV-1 infections and is further broken down into nine subtypes (A, B, C, D, F, G, H, J and K), four sub-subtypes (A1, A2, F1 and F2) (45) and 51 circulating recombinant forms (CRF) as of February 2012 ((47); <http://www.hiv.lanl.gov/>). CRFs are generated when an individual is co-infected with two different HIV-1 strains that recombine to produce a new mosaic virus (24, 46). In order to be classified as a CRF at least three epidemiologically unlinked cases must be identified (45). HIV-1 group M subtypes and CRFs are distributed differently around the world.

Central Africa, the presumed epicenter of the HIV-1 group M pandemic, still remains the center of HIV-1 diversity. HIV-1 groups N and P have been identified in only a small number of Cameroonians and remain restricted to west-central Africa (24, 28). Although group O infections have also been centralized in Cameroon, spread to

neighbouring countries has been reported (24). On the other hand HIV-1 group M has spread around the world with different subtypes prevalent in different geographic regions. Figure 1.1 shows the distribution of the different group M subtypes across the world that have been published on Los Alamos National Laboratories HIV Database ((47); <http://www.hiv.lanl.gov/>). Currently, the highest variation within group M remains in the Congo River basin, the hypothesized origin of the initial zoonotic transmission event (24). The western world is dominated by subtype B viral variants with more diversity in areas of South America and Europe (24). Subtype A variants dominate regions of Russia and are highly prevalent in countries throughout the Middle East along with subtype B and C (24). In Asia there are multiple subtypes in circulation including a variation of subtype B known as Thai B and a high prevalence of subtype C (24). Areas of Southeast Asia also have a prominent proportion of a CRF known as CRF01_AE, a recombinant developed from a subtype A virus and a subtype E virus (24, 46). Interestingly, subtype E continues to exist only within a CRF and has altogether disappeared as a unique subtype (46). The enormous global diversity in HIV-1 sequences remains a primary challenge in vaccine development, highlighting the importance of achieving a deeper understanding of the factors influencing HIV-1 evolution.

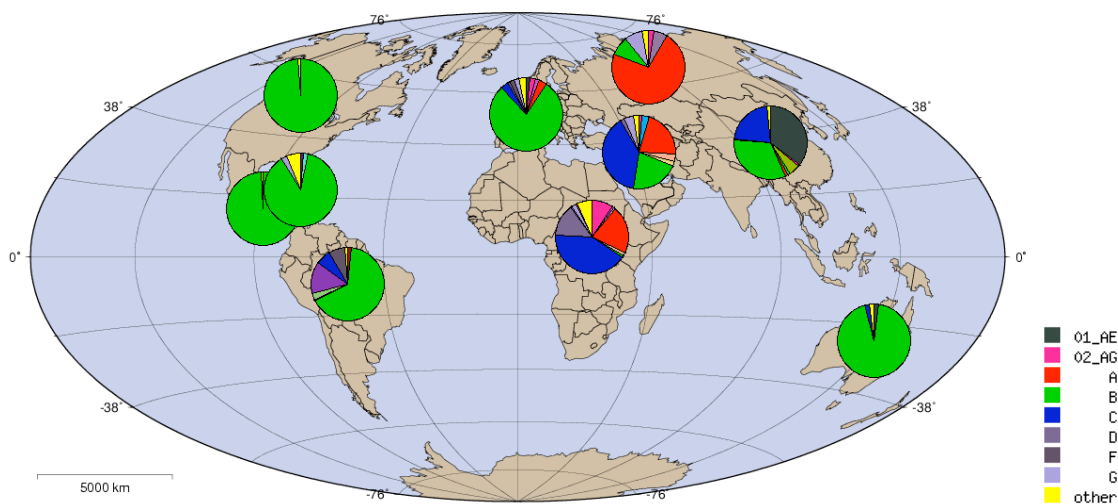


Figure 1.1. Distribution of HIV-1 group M subtypes.

((47); <http://www.hiv.lanl.gov/>)

1.4. Global Impact

Since its first recognized emergence in the early 1980s, HIV-1 has achieved pandemic status with a reported 34 million people estimated to be living with HIV/AIDS worldwide at the end of 2010; a 17% increase since 2001 (48). While global prevalence continues to rise, HIV incidence, defined as the number of new HIV-1 infections per year, has been decreasing over the last decade after peaking in the late-1990s (49). At the end of 2010 there were an estimated 2.7 million new cases of HIV-1 reported worldwide, down from an estimated 3.1 million new cases reported in 2001 (48). However, in many countries HIV-1 incidence continues to rise (48, 49). To date more than 60 million people have become infected with HIV-1 contributing to a cumulative HIV/AIDS associated death toll of nearly 30 million (50). Current projections estimate up to 75 million HIV/AIDS related deaths by 2030 (51). In 2010 alone there were a reported 1.8 million deaths related to HIV/AIDS, down from an estimated 1.9 million deaths in

2001 (48). More than 16.6 million children worldwide have lost both parents as a result of HIV/AIDS-related deaths (50).

The impact of HIV/AIDS extends around the world despite the highest burden of infection remaining in areas of Sub-Saharan Africa. Current estimates indicate approximately 1.3 million adults and children living with HIV/AIDS across North America with 20 thousand deaths reported in 2010 (48). While the number of new infections per year has decreased worldwide, HIV-1 incidence continues to increase in North America. At the end of 2010 there were an estimated 58 thousand new cases of HIV-1 reported in North America, up slightly from an estimated 49 thousand new cases reported in 2001 (48). The most recent reports from the Public Health Agency of Canada indicate approximately 65 thousand Canadians were living with HIV/AIDS at the end of 2008 (52), 26% of whom were unaware of their infection (53). Unprotected sex among men who have sex with men and injection drug use continue to represent major risk factors for HIV-1 transmission in North America with a large number of HIV-1 cases concentrated in racial and ethnic minorities (49).

1.5. HIV Prevention and Treatment

Presently, behaviour change is the single most important factor contributing to the decline in worldwide HIV-1 incidence rates. Public health initiatives aimed at raising awareness have resulted in enhanced knowledge and comprehension regarding the manner in which HIV-1 is spread and the ways to stay protected (49). This in turn has resulted in increased condom use, decreases in the number of sexual partners and delayed onset of sexual activity in youth (49). Furthermore, public health campaigns have fought to reduce social and economic barriers while improving access to condoms,

sterile needles and safe injection sites (49, 54). Despite these initiatives and reductions in incidence, millions of new cases of HIV-1 are still reported every year. Changing behaviour on a global scale is incredibly challenging, particularly when HIV/AIDS continues to remain a very stigmatized disease.

Another important advance in HIV-1 prevention and treatment has been the introduction of triple combination antiretroviral therapy, also known as highly active antiretroviral therapy (HAART), in the mid-1990s. HAART reduces viral loads below detectable limits and has resulted in dramatic reductions in HIV-related morbidity and mortality among infected individuals (49, 55-57). More recent studies have additionally described the preventative applications of HAART, demonstrating that undetectable levels of HIV-1 RNA in the plasma and sexual fluids of individuals using HAART decreases the risk of HIV-1 transmission (58-60). Unfortunately, upon discontinuing HAART HIV-1 positive individuals rapidly regain detectable plasma viral loads (61), indicating that antiretroviral drugs work to suppress viral replication but do not eliminate HIV-1 infection (57, 62-65). Coupled with the immune system's inability to effectively eliminate infection (66), HIV-1 remains a lifelong disease once contracted.

Given the challenges associated with behaviour change and the fact that HIV-1 cannot be cured, development of an effective HIV-1 vaccine has become a central focus in public health. Unfortunately, we have yet to develop an effective HIV-1 vaccine despite more than 30 years of research (67). Traditionally, vaccines typically aim to achieve "sterilizing immunity" whereby inoculated individuals generate effective long-lasting immune responses that protect them from infection in the event of future exposure. However, vaccines that are not able to block infection but can attenuate pathogenesis in the circumstance that infection is established, could also be highly

beneficial from a public health perspective (67). Indeed, given current knowledge gaps in HIV-1 immunobiology, the latter type of vaccine may represent a more realistic goal (67). During the course of a natural HIV-1 infection, plasma viral load correlates positively with rate of disease progression as well as the likelihood of transmission (68), therefore, developing a vaccine that induces an immune response to reduce viral loads could yield major public health benefits (67, 69). Such a vaccine would not only favour disease prognosis in the infected individual but would also decrease the risk of transmission, thus reducing the number of new infections. Dramatic reductions in global incidence rates are needed to curb the increasing global prevalence of HIV-1 infection and change the course of the epidemic, a primary goal of current vaccine strategies.

1.6. HIV Pathogenesis

HIV-1 is a pathogenic retrovirus that infects CD4⁺ T-lymphocytes and macrophages, essential effector cells of the human immune system (70-73). Transmission and infection with HIV-1 occurs when bodily fluids (such as blood, semen, vaginal fluid or breast milk) containing HIV-1 come into direct contact with a mucus membrane, abrasions on the skin or the bloodstream (10). Exposure can occur horizontally through sexual contact, blood transfusions, organ transplantations and needle sharing, or vertically when an infected mother transmits the virus to her child in utero, during delivery or through breastfeeding (10). If left untreated, HIV-1 begins to deplete the host immune system leaving the individual susceptible to a variety of opportunistic bacterial, viral and/or fungal infections at the onset of AIDS. As classified by the U.S. CDC in 1992, AIDS is defined as the presence of an AIDS-defining illness or a CD4⁺ T-lymphocyte count below 200 cells/mm³ of blood (74). Without access to

antiretroviral treatment, most patients die within several years once HIV-1 has progressed to AIDS.

1.7. Brief Overview of the HIV Life Cycle and Genomic Organization

The HIV-1 genome is composed of two identical copies of a single positive (+)-sense RNA strand, approximately 9700 base-pairs long (23). Characteristic of all known retroviruses, the HIV-1 genome is organized into three major open reading frames: from 5' to 3' are encoded *gag* (group-specific antigen), *pol* (polymerase) and *env* (envelope) (10). The *gag* region encodes the precursor protein Pr55^{*gag*}, which is cleaved by HIV-1 protease to release the major structural proteins: matrix protein (MA, p17), capsid protein (CA, p24), nucleocapsid protein (NC, p7) and a small accessory protein known as p6^{*gag*} (10). Following *gag*, the *pol* region encodes three primary enzymes including protease, reverse transcriptase and integrase (10). *Pol* genes are only expressed as part of a *gag-pol* polyprotein precursor (Pr160^{*gag-pol*}), generated by a ribosomal frameshift during translation of viral mRNA that is later cleaved by the viral protease enzyme into individual proteins (10). Ribosomal frameshifting allows the *gag* and *pol* open reading frames to overlap one another, producing different proteins from the same genomic region (10). Lastly, the *env* region encodes the precursor protein gp160, which unlike *gag* is cleaved by a host protease to release the major envelope surface protein (SU, gp120) and transmembrane protein (TM, gp41) (10). In addition, the HIV-1 genome encodes for a number of regulatory and accessory proteins that are required for viral replication and increase viral pathogenicity (not discussed).

The HIV-1 lifecycle can be broken down into a number of different phases. The initial step in replication begins when HIV-1 binds to CD4 on the host cell surface through an interaction with the viral envelope protein gp120 (70, 71, 75). Initially, CD4 was believed to be the only receptor required for cell entry, however, in the mid-1990s numerous publications appeared describing two important HIV-1 co-receptors: CCR5 and CXCR4 (76-80). The CXCR4 receptor is primarily expressed on activated CD4+ T-lymphocytes, while the CCR5 receptor is primarily expressed on naïve and memory CD4+ T-lymphocytes as well as macrophage cells (76, 78, 81). CCR5-using HIV-1 variants (also called M-tropic or R5 HIV-1 variants) tend to predominate early in infection where as CXCR4-using variants (also called T-tropic or X4 HIV variants) emerge in approximately one-half of infected individuals during later stage disease, a phenomenon termed “coreceptor switching” (82, 83). The appearance of CXCR4-using HIV-1 variants is associated with an increased rate of disease progression (83, 84). A review published by Moore et al. (85) summarizes the role of these co-receptors in HIV-1 pathogenesis.

After the binding of gp120 to the cell surface via CD4, a conformational change is induced within HIV-1 that exposes an additional co-receptor site on gp120 (10). The subsequent binding of gp120 to either CXCR4 or CCR5 triggers a viral-cell membrane fusion event that results in release of the viral capsid (10). Once inside the cell, the viral capsid begins to break down in a process known as “uncoating”, whereby the viral genome is released into the host cell cytoplasm (10). The HIV-1 enzyme reverse transcriptase then converts the single-stranded RNA genome into double-stranded DNA (dsDNA) that is then transported into the nucleus and integrated into the host cell’s genome with the help of HIV-1 integrase and other HIV/host proteins (10). Integrated HIV-1 DNA is referred to as a “provirus” and may remain inactive (latent) for many years. Once an infected host cell becomes activated, the HIV-1 provirus is transcribed by host

cell polymerases to yield full-length genomic RNAs (for inclusion in new virions and serve as templates for *gag-pol* polyproteins) as well as singly spliced and multiply-spliced viral mRNAs that serve as the template for production of other viral proteins (10).

The final stages of the HIV-1 lifecycle include assembly of new virions, the release of these particles at the host cell membrane and maturation. Formation of new virions is initiated when the *gag* precursor protein Pr55^{*gag*} and the *gag-pol* precursor protein Pr160^{*gag-pol*} move to the host cell plasma membrane and begin to assemble (10). *Gag* proteins play a crucial role in coordinating the packaging of viral RNA and accessory proteins into individual virions that subsequently bud from the host cell membrane (10). After release from the host cell, a final “maturation” step occurs whereby the protease domain of the *gag-pol* precursor protein cleaves and releases the individual proteins contained within the *gag* and *gag-pol* polyproteins (10). These mature virions are then capable of infecting and replicating within a new host cell. Studies appearing in the mid-1990s exploring the relationship between plasma viral load and initiation of antiretroviral therapy have further revealed important characteristics of the HIV-1 viral life cycle (86-88). With the use of mathematical modeling, estimates have revealed an average HIV-1 replication time of 1.2 days for actively infected CD4+ T-lymphocytes (88). Additionally, the lifespan of free plasma virions was shown to be less than twenty-four hours, indicating that high viral turnover is required for sustained plasma viremia (86, 88).

1.8. Host Immune Responses to HIV Infection

During the course of a natural HIV-1 infection, cytotoxic T-lymphocytes (CTL) play an essential role at reducing viral loads by eliminating infected cells presenting

virally derived peptides bound to human leukocyte antigen (HLA) class I molecules on the cell surface (89-91). HLA class I molecules, encoded by a cluster of genes known as the major histocompatibility complex (MHC), are a group of membrane glycoproteins found on the surface of all nucleated cells throughout the human body (92). The central role of these molecules is to bind peptides originating within the cell cytosol and display these at the cell surface for CTL recognition (92). This system of cell surface presentation allows cells to alert circulating CTL of a potential intracellular infection, targeting them for CTL induced cell death (92). In this way, CTL are able to eliminate cells presenting virally-derived peptide fragments at the cell surface. Although CTL responses are important in controlling viral loads during the early stages of HIV-1 infection, most infected individuals, if left untreated, experience gradual increases in plasma viremia (68, 69). These observations indicate that CTL responses are unable to sustain viral suppression and/or eliminate HIV-1 infection. As first described in 1991, this has been primarily attributed to a process known as CTL-driven “immune escape”, whereby HIV-1 evolves in response to HLA-coupled CTL responses in order to evade immune detection (93).

Similar to cellular immune responses, humoral immune responses are additionally important in HIV-1 infection. Antibodies serve to neutralize and destroy foreign antigen circulating in the body and are produced by activated B cells (94). Typically, antibodies specific for regions of HIV-1 envelope proteins appear in infected individuals within weeks after infection, however, these responses have primarily been associated with autologous virus (95). Broadly neutralizing antibodies do not emerge until later stage disease and have been associated with greater immune control in certain individuals labeled “elite neutralizers” (96). Nevertheless, like CTL immune selective pressures, HIV-1 is able to mutate in response to antibody pressures and

escape such that protective antibodies are no longer correlated with protection (95). Although there is a general consensus among the research community that an effective HIV-1 vaccine will likely have to stimulate both humoral and cellular immune responses (67), this thesis will specifically explore cellular immune responses.

1.9. Overview of Cytotoxic T-lymphocyte Mediated Immune Escape

Among the most challenging aspects of HIV-1 vaccine design is the ability of the virus to rapidly mutate and diversify. In approximately 80% of heterosexual transmissions (97), infection likely results from a single founder virus whereby transmission is characterized by a “genetic bottleneck” (98-101). However, upon establishment HIV-1 rapidly begins to diversify within the infected individual. A major part of HIV-1’s mutational capacity results from the error-prone nature of reverse transcriptase (102). Reverse transcriptase lacks the proofreading capability that most DNA polymerase enzymes possess, resulting in an estimated error rate of at least one mutation per replication cycle (102). Frequent recombination (103) and host factors such as APOBEC3G (104) also contribute to the generation of HIV-1 diversity in vivo. Combined with a high replication rate, within host variability begins to accumulate rapidly following infection such that over time multiple viral variants co-exist together in what is known as a quasispecies (105, 106). Subsequent immune responses are then thought to act as a major selective pressure driving the evolution of HIV-1 in a process referred to as “immune escape” (107-109).

Mutational immune escape represents a major mechanism allowing HIV-1 to evade immune detection. CTL-driven immune escape is typically characterized by the

selection of viral mutations within or near CTL epitopes specific for host HLA alleles (69). These mutations serve to disrupt HLA class I binding, intracellular epitope processing and/or CTL recognition of the epitope-HLA complex (69, 110-113). Consequently, viral variants that possess these escape mutations are able to continue replicating within the host cell undetected (or poorly detected) by the immune system. Because CTL responses require recognition of epitope-HLA complexes, antigen presentation is said to be “HLA-restricted”. Immune escape is thus similarly “HLA-restricted” (114), such that specific amino acid polymorphisms in HIV-1 are associated with certain HLA class I alleles expressed within the host (69). Given that a single individual expresses two (likely different) alleles at each of the three HLA class I loci: HLA-A, HLA-B and HLA-C, and that there are thousands of circulating HLA class I alleles throughout the human population (each recognizing a different repertoire of HIV-1 peptides) (92), CTL responses, and thus immune escape mutations, will vary between individuals. We now appreciate that CTL responses represent a significant selective pressure driving the evolution of key HIV-1 proteins within the host (69), and likely throughout the population as a whole (115).

HIV-1 mutational escape from CTL-mediated selection pressures was first described throughout the 1990s. In 1991 Phillips et al. (93) published results from a longitudinal study of HIV-1 positive patients showing that variation in proviral *gag* sequences could lead to escape from CTL responses whereby the host's immune system no longer recognized previous viral epitopes. Despite initial skepticism, subsequent studies supported the appearance of specific escape mutations throughout the HIV-1 proteome in response to HLA-associated CTL selection pressures (109, 116-122). In fact, studies have since identified a large number of CTL escape mutations that are reproducibly selected for in response to particular HLA class I alleles (123-126),

suggesting that HLA-associated immune selection pressures induce predictable patterns of immune escape. One such example is the well-characterized I135X mutation within an important reverse transcriptase epitope that is selected for in individuals expressing the HLA-B*51 allele (127). Individuals who do not express this allele do not actively select for this mutation; however, if the I135X mutation is acquired at transmission and does not result in a significant fitness cost to the virus, it may persist in the newly infected individual (127). It has been hypothesized that the persistence of escape mutations upon transmission could lead to the gradual accumulation of certain HLA-associated polymorphisms at the population-level (127). This has led to the development of the “HLA imprinting” hypothesis, which purposes that the HIV-1 consensus sequence circulating within a population may reflect selective pressures induced by the most commonly expressed HLA alleles for that given population (128-131). However, the extent to which these immune selection pressures have influenced HIV-1 evolution at the population-level over the course of the epidemic remains relatively uncharacterized.

Until recently, population-level analyses of HLA-associated polymorphisms have been limited by the integrity of statistical models. The first study to use population-level data to explore immune escape mutations in HIV-1 was published in 2002 by Moore et al. (132). Using more than 400 reverse transcriptase sequences derived from clinically infected patients, Moore et al. (132) identified 64 positive HLA-associated polymorphisms through multivariate logistic regression. However, one of the major limitations of this study was the inability of their model to account for confounding by viral lineage effects: the observation of shared viral sequences as a result of descent from a common ancestor rather than HLA-associated immune selection (133). In 2007 Bhattacharya et al. (133) attempted to characterize the impact of viral lineage effects in the assessment of HLA-associated polymorphisms by applying phylogenetically-

informed methods to a dataset similar to the one used by Moore et al. (132). Their results revealed that more than 75% of the HLA-associated polymorphisms identified without using phylogenetically-informed methods were likely due to viral lineage effects (133). These findings reiterate the importance of considering confounders in statistical analyses of association. Over the last several years, further statistical improvements in analyzing population-level datasets have included methods to correct for the confounding effects of linkage disequilibrium between HLA class I alleles and HIV-1 amino acid co-variation in HIV-1 genes (114, 134, 135). Additionally, phylogenetic software packages and statistical tools have enriched our ability to estimate the extent to which selection pressures shape HIV-1 evolution (136-140). Improvements in statistical modeling has resulted in the use of large cross-sectional cohorts of individuals with linked HLA class I and HIV-1 sequence data to map the distribution of population-wide escape mutations (127, 141-149). Despite these improvements, characterizing population-level HIV-1 adaptation over time remains limited by the scarcity of historic datasets.

Given the importance of characterizing HIV-1 evolution for the purposes of vaccine design, this thesis project will seek to address how CTL-derived immune selection pressures have shaped both the genomic sequence and functional properties of HIV-1 across the epidemic in North America. Unique to this study is the use of a historic cohort of individuals from Boston, New York and San Francisco for which serum samples were recovered dating between 1979 and 1989. These historic samples likely represent some of the first individuals infected with HIV-1 in North America. Although many HIV-1 protein products have the potential to elicit CTL immune responses, *gag* gene products have become a central focus in vaccine design because this region displays relatively high sequence conservation (150, 151) and is an important target of

the anti-HIV CTL response (152-157). Therefore, the objective of this thesis is to examine how immune selection pressures have influenced patterns of immune escape and replication capacity in the HIV-1 *gag* gene between historic and modern cohorts. Determining whether HIV-1 evolution within *gag* limits our ability to create an effective vaccine is important for informing public health initiatives aimed at HIV-1 prevention.

1.10. Thesis Objectives and Organization

This thesis is organized into of four chapters. This first chapter provides a general introduction to the history of HIV-1 infection, an update on current public health successes and shortcomings in HIV-1 prevention and treatment as well as a brief overview of CTL-mediated immune escape. It also outlines the relevance and objectives of this thesis research. Chapters 2 and 3 address the primary aims of this thesis project as outlined below and chapter 4 summarize the public health implications. This thesis is prepared according to a manuscript-based format where chapters 2 and 3 each represent stand-alone manuscripts. Chapter 2 is under review after submission to the *Journal of Immunological Methods*. Chapter 3 will be finalized and submitted to an international, peer-reviewed journal in the coming months.

The primary research question addressed by this thesis is the following: **How have immune selection pressures mediated through host cytotoxic T-lymphocyte recognition of viral derived peptides presented by HLA class I molecules shaped both the genomic and functional evolution of the HIV-1 *gag* gene across the epidemic in North America?** The host and viral genetic factors selected for investigation are factors known to be involved in HIV-1 adaptation and a central focus in vaccine research (67, 69). Chapter 3 is a population-based study conducted using

samples from a historic cross-sectional cohort of HIV-infected individuals to examine the impact of HLA-mediated immune selection pressures on patterns of immune escape and viral replication capacity compared to a previously characterized modern cohort. The specific aims of this thesis are:

1. To validate methods of HLA class I typing using plasma/serum samples as a starting material.
2. To examine the impact of HLA-driven immune selection pressures on patterns of escape mutations in the HIV-1 *gag* gene through time in North America.
3. Using both historic and modern serum samples from individuals in North America, to study the molecular epidemiology of the HIV-1 *gag* gene and incorporate the use of phylogenetic methods to estimate the timing of the initial founder virus.
4. To examine the functional properties of the HIV-1 *gag* gene to see whether differences in viral replication capacity differ between historic versus modern HIV-1 *gag* sequences obtained from North America.

This thesis will, therefore, explore both the individual-level and population-level complexities of HIV-1 evolution as well as the long-term implications of host and viral interactions on HIV-1 adaptation to human populations. The research outlined in this thesis has been approved by the Simon Fraser University Research ethics board.

1.11. Overview of Data Sources and Population-based Cohorts

As discussed in the following chapters, this thesis uses data from a number of cohorts and sources. Chapter 2 uses a set of 25 anonymized, matched peripheral blood mononuclear cells (PBMC)/plasma samples obtained from HIV-1 positive participants of the Vancouver Injection Drug Users Study (VIDUS), a cohort followed since 1996 and run through the B.C. Centre for Excellence in HIV/AIDS. Information regarding date of

infection, use of antiretroviral therapy, age and gender were not released for the purposes of this study. All 25 samples were collected between 2001-2007 and stored at -80°C until extraction. The median storage time and interquartile range [IQR] for these 25 samples was 8.66 [8.05-9.04] years. Although the VIDUS cohort samples used here were comprised of HIV-1 positive individuals, for the purposes of this study these samples were used only to validate the robustness of HLA class I typing methods from plasma as compared to typing from PBMC.

Linked HLA class I/HIV-1 genotype datasets from historic specimens used in Chapter 3 were generated using frozen serum samples recovered from three chronically-HIV-infected cohorts from the New York Blood Center (New York, NY), the Fenway Community Health Clinic (Boston, MA) and the San Francisco Department of Public Health (San Francisco, CA). Samples were recovered from historic repositories of frozen stored sera (-80°C), originally collected as part of observational cohort studies of Hepatitis B sero-incidence and sero-prevalence retrospectively identified to be HIV-1 positive (New York and San Francisco samples), and historic observational HIV-1 cohort studies (Boston samples). The median age of specimens was 24.43 [IQR: 23.89-26.44] years. HIV-1 *gag* amplification and sequencing was successful for 22 serum samples obtained from the New York Blood Center (1979-1984), 217 serum samples obtained from the Fenway Community Health Clinic in Boston (1985-1989) and 4 serum samples obtained from the San Francisco Department of Public Health (1984). Furthermore, these historic serum samples were used to perform HLA class I typing. In total, 187, 234 and 235 HLA-A, HLA-B and HLA-C loci respectively were successfully sequenced and typed. Information regarding antiretroviral therapy, age and gender was not provided. These three cohorts represent some of the oldest HIV-1 samples recovered in North America.

1.12. References

1. Pneumocystis pneumonia--Los Angeles. MMWR. Morbidity and mortality weekly report. 1981 Jun 5;**30**(21):250-2.
2. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. MMWR. Morbidity and mortality weekly report. 1981 Jul 3;**30**(25):305-8.
3. Follow-up on Kaposi's sarcoma and Pneumocystis pneumonia. MMWR. Morbidity and mortality weekly report. 1981 Aug 28;**30**(33):409-10.
4. Update on acquired immune deficiency syndrome (AIDS)--United States. MMWR. Morbidity and mortality weekly report. 1982 Sep 24;**31**(37):507-8, 13-4.
5. Hymes KB, Cheung T, Greene JB, et al. Kaposi's sarcoma in homosexual men-a report of eight cases. Lancet. 1981 Sep 19;**2**(8247):598-600.
6. Gottlieb MS, Schroff R, Schanker HM, et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. The New England journal of medicine. 1981 Dec 10;**305**(24):1425-31.
7. Masur H, Michelis MA, Greene JB, et al. An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. The New England journal of medicine. 1981 Dec 10;**305**(24):1431-8.
8. Wallace JM, Coral FS, Rimm IR, et al. T-cell ratios in homosexuals. Lancet. 1982 Apr 17;**1**(8277):908.
9. Goedert JJ, Neuland CY, Wallen WC, et al. Amyl nitrite may alter T lymphocytes in homosexual men. Lancet. 1982 Feb 20;**1**(8269):412-6.
10. Richman DD, editor. Human Virus Guides 2: Human Immunodeficiency Virus. 2 ed. London: International Medical Press; 2003.
11. Acquired immune deficiency syndrome (AIDS): precautions for clinical and laboratory staffs. MMWR. Morbidity and mortality weekly report. 1982 Nov 5;**31**(43):577-80.
12. Sonnabend J, Witkin SS, Purtilo DT. Acquired immunodeficiency syndrome, opportunistic infections, and malignancies in male homosexuals. A hypothesis of etiologic factors in pathogenesis. JAMA : the journal of the American Medical Association. 1983 May 6;**249**(17):2370-4.
13. Barre-Sinoussi F, Chermann JC, Rey F, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science. 1983 May 20;**220**(4599):868-71.
14. Gallo RC, Sarin PS, Gelmann EP, et al. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science. 1983 May 20;**220**(4599):865-7.

15. Evans AS. Hypothesis: the pathogenesis of AIDS. Activation of the T- and B-cell cascades. The Yale journal of biology and medicine. 1984 May-Jun;**57**(3):317-27.
16. Melbye M, Biggar RJ, Ebbesen P, et al. Seroepidemiology of HTLV-III antibody in Danish homosexual men: prevalence, transmission, and disease outcome. Br Med J (Clin Res Ed). 1984 Sep 8;**289**(6445):573-5.
17. Gallo RC, Salahuddin SZ, Popovic M, et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science. 1984 May 4;**224**(4648):500-3.
18. Brun-Vezinet F, Rouzioux C, Barre-Sinoussi F, et al. Detection of IgG antibodies to lymphadenopathy-associated virus in patients with AIDS or lymphadenopathy syndrome. Lancet. 1984 Jun 9;**1**(8389):1253-6.
19. Popovic M, Sarngadharan MG, Read E, Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science. 1984 May 4;**224**(4648):497-500.
20. Schupbach J, Popovic M, Gilden RV, Gonda MA, Sarngadharan MG, Gallo RC. Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. Science. 1984 May 4;**224**(4648):503-5.
21. Sarngadharan MG, Popovic M, Bruch L, Schupbach J, Gallo RC. Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS. Science. 1984 May 4;**224**(4648):506-8.
22. Sarngadharan MG, Bruch L, Popovic M, Gallo RC. Immunological properties of the Gag protein p24 of the acquired immunodeficiency syndrome retrovirus (human T-cell leukemia virus type III). Proceedings of the National Academy of Sciences of the United States of America. 1985 May;**82**(10):3481-4.
23. Ratner L, Haseltine W, Patarca R, et al. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature. 1985 Jan 24-30;**313**(6000):277-84.
24. Tebit DM, Arts EJ. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. The Lancet infectious diseases. 2011 Jan;**11**(1):45-56.
25. Gao F, Bailes E, Robertson DL, et al. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature. 1999 Feb 4;**397**(6718):436-41.
26. Corbet S, Muller-Trutwin MC, Versmisse P, et al. env sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. Journal of virology. 2000 Jan;**74**(1):529-34.
27. Keele BF, Van Heuverswyn F, Li Y, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science. 2006 Jul 28;**313**(5786):523-6.

28. Plantier JC, Leoz M, Dickerson JE, et al. A new human immunodeficiency virus derived from gorillas. *Nature medicine*. 2009 Aug;**15**(8):871-2.
29. Van Heuverswyn F, Li Y, Neel C, et al. Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature*. 2006 Nov 9;**444**(7116):164.
30. Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR. An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature*. 1989 Jun 1;**339**(6223):389-92.
31. Vidal N, Peeters M, Mulanga-Kabeya C, et al. Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *Journal of virology*. 2000 Nov;**74**(22):10498-507.
32. Kalish ML, Wolfe ND, Ndongmo CB, et al. Central African hunters exposed to simian immunodeficiency virus. *Emerging infectious diseases*. 2005 Dec;**11**(12):1928-30.
33. Wolfe ND, Prosser TA, Carr JK, et al. Exposure to nonhuman primates in rural Cameroon. *Emerging infectious diseases*. 2004 Dec;**10**(12):2094-9.
34. Peeters M, Courgnaud V, Abela B, et al. Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. *Emerging infectious diseases*. 2002 May;**8**(5):451-7.
35. Korber B, Muldoon M, Theiler J, et al. Timing the ancestor of the HIV-1 pandemic strains. *Science*. 2000 Jun 9;**288**(5472):1789-96.
36. Worobey M, Gemmel M, Teuwen DE, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature*. 2008 Oct 2;**455**(7213):661-4.
37. Curtis T. The Origin of AIDS: A startling new theory attempts to answer the question 'was it an act of God or an act of man?'. *Rolling Stone*. 1992:54-9, 61, 106, 8.
38. Hooper E. *The River: A Journey to the Source of HIV and AIDS*. Boston Little, Brown and Company; 1999.
39. Poinar H, Kuch M, Paabo S. Molecular analyses of oral polio vaccine samples. *Science*. 2001 Apr 27;**292**(5517):743-4.
40. Blancou P, Vartanian JP, Christopherson C, et al. Polio vaccine samples not linked to AIDS. *Nature*. 2001 Apr 26;**410**(6832):1045-6.
41. Berry N, Jenkins A, Martin J, et al. Mitochondrial DNA and retroviral RNA analyses of archival oral polio vaccine (OPV CHAT) materials: evidence of macaque nuclear sequences confirms substrate identity. *Vaccine*. 2005 Feb 25;**23**(14):1639-48.
42. Rambaut A, Robertson DL, Pybus OG, Peeters M, Holmes EC. Human immunodeficiency virus. Phylogeny and the origin of HIV-1. *Nature*. 2001 Apr 26;**410**(6832):1047-8.

43. Pepin J, Labbe AC, Mamadou-Yaya F, et al. Iatrogenic transmission of human T cell lymphotropic virus type 1 and hepatitis C virus through parenteral treatment and chemoprophylaxis of sleeping sickness in colonial Equatorial Africa. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2010 Oct 1;**51**(7):777-84.
44. Clavel F, Mansinho K, Chamaret S, et al. Human immunodeficiency virus type 2 infection associated with AIDS in West Africa. *The New England journal of medicine*. 1987 May 7;**316**(19):1180-5.
45. Robertson DL, Anderson JP, Bradac JA, et al. HIV-1 nomenclature proposal. *Science*. 2000 Apr 7;**288**(5463):55-6.
46. Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. The challenge of HIV-1 subtype diversity. *The New England journal of medicine*. 2008 Apr 10;**358**(15):1590-602.
47. Kuiken C, Foley B, Leitner T, et al. HIV Sequence Compendium. Los Alamos National Laboratory, Los Alamos, New Mexico: Theoretical Biology and Biophysics Group; 2010.
48. UNAIDS. UNAIDS: World AIDS Day Reprot, 2011. Geneva, Switzerland; 2011.
49. UNAIDS. Global Report: UNAIDS report on the global AIDS epidemic, 2010. Geneva, Switzerland 2010.
50. UNAIDS. UNAIDS Global Report: fact sheet. Geneva, Switzerland; 2010.
51. Bongaarts J, Pelletier F, Gerland P. Poverty, Gender, and Youth: Global Trends in AIDS Mortality. New York, USA: Population Council 2009.
52. PHAC. HIV/AIDS Epi Updates - July 2010. 2010.
53. PHAC. Surveillance Report to December 31, 2009. 2009.
54. VCH. Insite - Supervised Injection Site: A health-focused place for people to connect with health care services. 2011.
55. Perelson AS, Essunger P, Cao Y, et al. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature*. 1997 May 8;**387**(6629):188-91.
56. Palella FJ, Jr., Delaney KM, Moorman AC, et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *The New England journal of medicine*. 1998 Mar 26;**338**(13):853-60.
57. Ho DD, Bieniasz PD. HIV-1 at 25. *Cell*. 2008 May 16;**133**(4):561-5.
58. Montaner JS, Hogg R, Wood E, et al. The case for expanding access to highly active antiretroviral therapy to curb the growth of the HIV epidemic. *Lancet*. 2006 Aug 5;**368**(9534):531-6.

59. Granich R, Crowley S, Vitoria M, et al. Highly active antiretroviral treatment for the prevention of HIV transmission. *Journal of the International AIDS Society*. 2010;**13**:1.
60. Granich R, Gilks CF, Dye C, De Cock KM, Williams BG. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. *Lancet*. 2009 Jan 3;**373**(9657):48-57.
61. Garcia F, Plana M, Vidal C, et al. Dynamics of viral load rebound and immunological changes after stopping effective antiretroviral therapy. *AIDS*. 1999 Jul 30;**13**(11):F79-86.
62. Finzi D, Hermankova M, Pierson T, et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science*. 1997 Nov 14;**278**(5341):1295-300.
63. Ramratnam B, Ribeiro R, He T, et al. Intensification of antiretroviral therapy accelerates the decay of the HIV-1 latent reservoir and decreases, but does not eliminate, ongoing virus replication. *J Acquir Immune Defic Syndr*. 2004 Jan 1;**35**(1):33-7.
64. Tobin NH, Learn GH, Holte SE, et al. Evidence that low-level viremias during effective highly active antiretroviral therapy result from two processes: expression of archival virus and replication of virus. *Journal of virology*. 2005 Aug;**79**(15):9625-34.
65. Chun TW, Carruth L, Finzi D, et al. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature*. 1997 May 8;**387**(6629):183-8.
66. Han Y, Wind-Rotolo M, Yang HC, Siliciano JD, Siliciano RF. Experimental approaches to the study of HIV-1 latency. *Nature reviews. Microbiology*. 2007 Feb;**5**(2):95-106.
67. Walker BD, Burton DR. Toward an AIDS vaccine. *Science*. 2008 May 9;**320**(5877):760-4.
68. Mellors JW, Rinaldo CR, Jr., Gupta P, White RM, Todd JA, Kingsley LA. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science*. 1996 May 24;**272**(5265):1167-70.
69. Goulder PJ, Watkins DI. HIV and SIV CTL escape: implications for vaccine design. *Nature reviews. Immunology*. 2004 Aug;**4**(8):630-40.
70. Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*. 1984 Dec 20-1985 Jan 2;**312**(5996):763-7.
71. Klatzmann D, Champagne E, Chamaret S, et al. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature*. 1984 Dec 20-1985 Jan 2;**312**(5996):767-8.
72. Ho DD, Rota TR, Hirsch MS. Infection of monocyte/macrophages by human T lymphotropic virus type III. *The Journal of clinical investigation*. 1986 May;**77**(5):1712-5.

73. Gyorkey F, Melnick JL, Sinkovics JG, Gyorkey P. Retrovirus resembling HTLV in macrophages of patients with AIDS. *Lancet*. 1985 Jan 12;**1**(8420):106.
74. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control*. 1992 Dec 18;**41**(RR-17):1-19.
75. McDougal JS, Nicholson JK, Cross GD, Cort SP, Kennedy MS, Mawle AC. Binding of the human retrovirus HTLV-III/LAV/ARV/HIV to the CD4 (T4) molecule: conformation dependence, epitope mapping, antibody inhibition, and potential for idiotypic mimicry. *J Immunol*. 1986 Nov 1;**137**(9):2937-44.
76. Feng Y, Broder CC, Kennedy PE, Berger EA. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science*. 1996 May 10;**272**(5263):872-7.
77. Cocchi F, DeVico AL, Garzino-Demo A, Arya SK, Gallo RC, Lusso P. Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science*. 1995 Dec 15;**270**(5243):1811-5.
78. Deng H, Liu R, Ellmeier W, et al. Identification of a major co-receptor for primary isolates of HIV-1. *Nature*. 1996 Jun 20;**381**(6584):661-6.
79. Choe H, Farzan M, Sun Y, et al. The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell*. 1996 Jun 28;**85**(7):1135-48.
80. Dragic T, Litwin V, Allaway GP, et al. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature*. 1996 Jun 20;**381**(6584):667-73.
81. Alkhatib G, Combadiere C, Broder CC, et al. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science*. 1996 Jun 28;**272**(5270):1955-8.
82. Doranz BJ, Rucker J, Yi Y, et al. A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. *Cell*. 1996 Jun 28;**85**(7):1149-58.
83. Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR. Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. *The Journal of experimental medicine*. 1997 Feb 17;**185**(4):621-8.
84. Schuitemaker H, Koot M, Kootstra NA, et al. Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytopathic to T-cell-tropic virus population. *Journal of virology*. 1992 Mar;**66**(3):1354-60.
85. Moore JP, Kitchen SG, Pugach P, Zack JA. The CCR5 and CXCR4 coreceptors--central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection. *AIDS research and human retroviruses*. 2004 Jan;**20**(1):111-26.

86. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*. 1995 Jan 12;**373**(6510):123-6.
87. Wei X, Ghosh SK, Taylor ME, et al. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature*. 1995 Jan 12;**373**(6510):117-22.
88. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*. 1996 Mar 15;**271**(5255):1582-6.
89. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *Journal of virology*. 1994 Sep;**68**(9):6103-10.
90. Koup RA, Safrit JT, Cao Y, et al. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *Journal of virology*. 1994 Jul;**68**(7):4650-5.
91. Musey L, Hughes J, Schacker T, Shea T, Corey L, McElrath MJ. Cytotoxic-T-cell responses, viral load, and disease progression in early human immunodeficiency virus type 1 infection. *The New England journal of medicine*. 1997 Oct 30;**337**(18):1267-74.
92. Klein J, Sato A. The HLA system. First of two parts. *The New England journal of medicine*. 2000 Sep 7;**343**(10):702-9.
93. Phillips RE, Rowland-Jones S, Nixon DF, et al. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature*. 1991 Dec 12;**354**(6353):453-9.
94. Harwood NE, Batista FD. Early events in B cell activation. *Annual review of immunology*. 2010 Mar;**28**:185-210.
95. Wei X, Decker JM, Wang S, et al. Antibody neutralization and escape by HIV-1. *Nature*. 2003 Mar 20;**422**(6929):307-12.
96. Stamatatos L, Morris L, Burton DR, Mascola JR. Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nature medicine*. 2009 Aug;**15**(8):866-70.
97. Abrahams MR, Anderson JA, Giorgi EE, et al. Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *Journal of virology*. 2009 Apr;**83**(8):3556-67.
98. Salazar-Gonzalez JF, Salazar MG, Keele BF, et al. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *The Journal of experimental medicine*. 2009 Jun 8;**206**(6):1273-89.

99. Derdeyn CA, Decker JM, Bibollet-Ruche F, et al. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science*. 2004 Mar 26;**303**(5666):2019-22.
100. Haaland RE, Hawkins PA, Salazar-Gonzalez J, et al. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS pathogens*. 2009 Jan;**5**(1):e1000274.
101. Keele BF, Giorgi EE, Salazar-Gonzalez JF, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America*. 2008 May 27;**105**(21):7552-7.
102. Malim MH, Emerman M. HIV-1 sequence variation: drift, shift, and attenuation. *Cell*. 2001 Feb 23;**104**(4):469-72.
103. Burke DS. Recombination in HIV: an important viral evolutionary strategy. *Emerging infectious diseases*. 1997 Jul-Sep;**3**(3):253-9.
104. Wood N, Bhattacharya T, Keele BF, et al. HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. *PLoS pathogens*. 2009 May;**5**(5):e1000414.
105. Goodenow M, Huet T, Saurin W, Kwok S, Sninsky J, Wain-Hobson S. HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. *J Acquir Immune Defic Syndr*. 1989;**2**(4):344-52.
106. Kamp C. A quasispecies approach to viral evolution in the context of an adaptive immune system. *Microbes and infection / Institut Pasteur*. 2003 Dec;**5**(15):1397-405.
107. Liu Y, McNevin J, Cao J, et al. Selection on the human immunodeficiency virus type 1 proteome following primary infection. *Journal of virology*. 2006 Oct;**80**(19):9519-29.
108. Goonetilleke N, Liu MK, Salazar-Gonzalez JF, et al. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *The Journal of experimental medicine*. 2009 Jun 8;**206**(6):1253-72.
109. Wolinsky SM, Korber BT, Neumann AU, et al. Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science*. 1996 Apr 26;**272**(5261):537-42.
110. Draenert R, Le Gall S, Pfafferoth KJ, et al. Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. *The Journal of experimental medicine*. 2004 Apr 5;**199**(7):905-15.
111. Allen TM, Altfeld M, Yu XG, et al. Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. *Journal of virology*. 2004 Jul;**78**(13):7069-78.

112. Yokomaku Y, Miura H, Tomiyama H, et al. Impaired processing and presentation of cytotoxic-T-lymphocyte (CTL) epitopes are major escape mechanisms from CTL immune pressure in human immunodeficiency virus type 1 infection. *Journal of virology*. 2004 Feb;**78**(3):1324-32.
113. Klenerman P, Rowland-Jones S, McAdam S, et al. Cytotoxic T-cell activity antagonized by naturally occurring HIV-1 Gag variants. *Nature*. 1994 Jun 2;**369**(6479):403-7.
114. Brumme ZL, Poon AEY, Carlson JM, Walker BD. Identifying HLA-Associated Polymorphisms in HIV-1. *HIV Molecular Immunology* 2010;**2010**:31-44.
115. Carlson JM, Brumme ZL. HIV evolution in response to HLA-restricted CTL selection pressures: a population-based perspective. *Microbes and infection / Institut Pasteur*. 2008 Apr;**10**(5):455-61.
116. Rowland-Jones SL, Phillips RE, Nixon DF, et al. Human immunodeficiency virus variants that escape cytotoxic T-cell recognition. *AIDS research and human retroviruses*. 1992 Aug;**8**(8):1353-4.
117. Couillin I, Culmann-Penciolelli B, Gomard E, et al. Impaired cytotoxic T lymphocyte recognition due to genetic variations in the main immunogenic region of the human immunodeficiency virus 1 NEF protein. *The Journal of experimental medicine*. 1994 Sep 1;**180**(3):1129-34.
118. Borrow P, Lewicki H, Wei X, et al. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nature medicine*. 1997 Feb;**3**(2):205-11.
119. McMichael AJ, Phillips RE. Escape of human immunodeficiency virus from immune control. *Annual review of immunology*. 1997;**15**:271-96.
120. Goulder PJ, Phillips RE, Colbert RA, et al. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nature medicine*. 1997 Feb;**3**(2):212-7.
121. Price DA, Goulder PJ, Klenerman P, et al. Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proceedings of the National Academy of Sciences of the United States of America*. 1997 Mar 4;**94**(5):1890-5.
122. Koenig S, Conley AJ, Brewah YA, et al. Transfer of HIV-1-specific cytotoxic T lymphocytes to an AIDS patient leads to selection for mutant HIV variants and subsequent disease progression. *Nature medicine*. 1995 Apr;**1**(4):330-6.
123. Allen TM, Altfeld M, Geer SC, et al. Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. *Journal of virology*. 2005 Nov;**79**(21):13239-49.

124. Leslie AJ, Pfafferott KJ, Chetty P, et al. HIV evolution: CTL escape mutation and reversion after transmission. *Nature medicine*. 2004 Mar;**10**(3):282-9.
125. Kelleher AD, Long C, Holmes EC, et al. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. *The Journal of experimental medicine*. 2001 Feb 5;**193**(3):375-86.
126. Brumme ZL, Brumme CJ, Carlson J, et al. 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. *Journal of virology*. 2008 Sep;**82**(18):9216-27.
127. Kawashima Y, Pfafferott K, Frater J, et al. Adaptation of HIV-1 to human leukocyte antigen class I. *Nature*. 2009 Apr 2;**458**(7238):641-5.
128. McMichael A, Klenerman P. HIV/AIDS. HLA leaves its footprints on HIV. *Science*. 2002 May 24;**296**(5572):1410-1.
129. Leslie A, Kavanagh D, Honeyborne I, et al. Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. *The Journal of experimental medicine*. 2005 Mar 21;**201**(6):891-902.
130. Goulder PJ, Brander C, Tang Y, et al. Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature*. 2001 Jul 19;**412**(6844):334-8.
131. Brander C, Walker BD. Gradual adaptation of HIV to human host populations: good or bad news? *Nature medicine*. 2003 Nov;**9**(11):1359-62.
132. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science*. 2002 May 24;**296**(5572):1439-43.
133. Bhattacharya T, Daniels M, Heckerman D, et al. Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science*. 2007 Mar 16;**315**(5818):1583-6.
134. Carlson JM, Brumme ZL, Rousseau CM, et al. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. *PLoS computational biology*. 2008 Nov;**4**(11):e1000225.
135. Poon AF, Swenson LC, Dong WW, et al. Phylogenetic analysis of population-based and deep sequencing data to identify coevolving sites in the nef gene of HIV-1. *Molecular biology and evolution*. 2010 Apr;**27**(4):819-32.
136. Delport W, Poon AF, Frost SD, Kosakovsky Pond SL. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*. 2010 Oct 1;**26**(19):2455-7.
137. Poon AF, Frost SD, Pond SL. Detecting signatures of selection from DNA sequences using Datamonkey. *Methods Mol Biol*. 2009;**537**:163-83.

138. Kosakovsky Pond SL, Poon AF, Zarate S, et al. Estimating selection pressures on HIV-1 using phylogenetic likelihood models. *Statistics in medicine*. 2008 Oct 15;**27**(23):4779-89.
139. Kosakovsky Pond SL, Posada D, Stawiski E, et al. An evolutionary model-based algorithm for accurate phylogenetic breakpoint mapping and subtype prediction in HIV-1. *PLoS computational biology*. 2009 Nov;**5**(11):e1000581.
140. Poon AF, Lewis FI, Frost SD, Kosakovsky Pond SL. Spidermonkey: rapid detection of co-evolving sites using Bayesian graphical models. *Bioinformatics*. 2008 Sep 1;**24**(17):1949-50.
141. Avila-Rios S, Ormsby CE, Carlson JM, et al. Unique features of HLA-mediated HIV evolution in a Mexican cohort: a comparative study. *Retrovirology*. 2009;**6**:72.
142. Boutwell CL, Essex M. Identification of HLA class I-associated amino acid polymorphisms in the HIV-1C proteome. *AIDS research and human retroviruses*. 2007 Jan;**23**(1):165-74.
143. Brumme ZL, Brumme CJ, Heckerman D, et al. Evidence of differential HLA class I-mediated viral evolution in functional and accessory/regulatory genes of HIV-1. *PLoS pathogens*. 2007 Jul;**3**(7):e94.
144. Brumme ZL, John M, Carlson JM, et al. HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. *PLoS one*. 2009;**4**(8):e6687.
145. John M, Heckerman D, James I, et al. Adaptive interactions between HLA and HIV-1: highly divergent selection imposed by HLA class I molecules with common supertype motifs. *J Immunol*. 2010 Apr 15;**184**(8):4368-77.
146. Matthews PC, Prendergast A, Leslie A, et al. Central role of reverting mutations in HLA associations with human immunodeficiency virus set point. *Journal of virology*. 2008 Sep;**82**(17):8548-59.
147. Wang YE, Li B, Carlson JM, et al. Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in highly conserved regions of human immunodeficiency virus type 1. *Journal of virology*. 2009 Feb;**83**(4):1845-55.
148. Rousseau CM, Daniels MG, Carlson JM, et al. HLA class I-driven evolution of human immunodeficiency virus type 1 subtype c proteome: immune escape and viral load. *Journal of virology*. 2008 Jul;**82**(13):6434-46.
149. Brumme ZL, Tao I, Szeto S, et al. Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection. *AIDS*. 2008 Jul 11;**22**(11):1277-86.
150. Novitsky V, Smith UR, Gilbert P, et al. Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? *Journal of virology*. 2002 Jun;**76**(11):5435-51.

151. Rolland M, Nickle DC, Mullins JI. HIV-1 group M conserved elements vaccine. *PLoS pathogens*. 2007 Nov;**3**(11):e157.
152. Edwards BH, Bansal A, Sabbaj S, Bakari J, Mulligan MJ, Goepfert PA. Magnitude of functional CD8+ T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma. *Journal of virology*. 2002 Mar;**76**(5):2298-305.
153. Geldmacher C, Currier JR, Herrmann E, et al. CD8 T-cell recognition of multiple epitopes within specific Gag regions is associated with maintenance of a low steady-state viremia in human immunodeficiency virus type 1-seropositive patients. *Journal of virology*. 2007 Mar;**81**(5):2440-8.
154. Kiepiela P, Ngumbela K, Thobakgale C, et al. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nature medicine*. 2007 Jan;**13**(1):46-53.
155. Riviere Y, McChesney MB, Porrot F, et al. Gag-specific cytotoxic responses to HIV type 1 are associated with a decreased risk of progression to AIDS-related complex or AIDS. *AIDS research and human retroviruses*. 1995 Aug;**11**(8):903-7.
156. Klein MR, van Baalen CA, Holwerda AM, et al. Kinetics of Gag-specific cytotoxic T lymphocyte responses during the clinical course of HIV-1 infection: a longitudinal analysis of rapid progressors and long-term asymptomatics. *The Journal of experimental medicine*. 1995 Apr 1;**181**(4):1365-72.
157. Zuniga R, Lucchetti A, Galvan P, et al. Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control. *Journal of virology*. 2006 Mar;**80**(6):3122-5.

2. Chapter 2: HLA Class I Sequence-based Typing using DNA Recovered from Frozen Plasma

2.1. Abstract

We describe a rapid, reliable and cost-effective method for intermediate-to-high-resolution sequence-based HLA class I typing using frozen plasma as a source of genomic DNA. The plasma samples investigated were a median 8.5 years old. Total nucleic acids were isolated from matched frozen PBMC (~2.5 million cells) and plasma (500µl) samples from a panel of 25 individuals using commercial silica-based kits. Extractions yielded median [IQR] nucleic acid concentrations of 85.7 [47.0-130.0] and 2.2 [1.7-2.6] ng/µl from PBMC and plasma, respectively. Following extraction, ~1000 base pair regions spanning exons 2 and 3 of HLA-A, B and C were amplified independently via nested PCR using universal, locus-specific primers and sequenced directly. Sequence chromatogram analysis was performed using commercial DNA sequence analysis software and allele interpretation was performed using a free web-based tool. HLA-A, B and C amplification rates were 100% and chromatograms were of uniformly high quality with clearly distinguishable mixed bases regardless of DNA source. Concordance between PBMC and plasma-derived HLA types was 100% at the allele and protein levels. At the nucleotide level, a single partially discordant base (resulting from a failure to call both peaks in a mixed base) was observed out of >46,975 bases sequenced (>99.9% concordance). This protocol has previously been used to perform HLA class I typing from a variety of genomic DNA sources including: PBMC,

whole blood, granulocyte pellets and serum, from specimens up to 30 years old. This method provides comparable specificity to conventional sequence-based approaches and could be applied in situations where cell samples are unavailable or DNA quantities are limiting.

2.2. Introduction

The human leukocyte antigen (HLA) system plays a fundamental role in immune recognition and response to foreign antigen. Located on chromosome 6, the HLA class I and class II genes represent two of the most polymorphic regions in the human genome (1), with 5,518 HLA class I and 1,612 class II allelic variants identified as of January 2012 ((2); <http://hla.alleles.org/nomenclature/stats.html>). HLA typing represents a fundamental clinical tool in tissue transplantation and, more recently, pharmacokinetic screening for HLA-mediated, medication-induced hypersensitivity reactions (3). As a growing number of molecular epidemiology studies recognize the links between the HLA system and autoimmune (4-6) and infectious (7-9) disease risk, as well as the impact of host immunogenetic variation on the evolution of pathogen genomes (10-12), high-resolution HLA typing has emerged as an essential research tool.

Traditionally, the most common source of genomic DNA for HLA typing is peripheral blood mononuclear cells (PBMC). However, PBMC are not always available for study; thus, typing approaches utilizing alternative, lower-yield sources of DNA could represent a powerful research tool. Genomic DNA can be reliably recovered from plasma or serum for downstream genotyping (13-21), and molecular methods for HLA class II typing using plasma and/or serum have been described (16, 17, 21). Methods for detection of soluble class I proteins in serum or plasma have also been described (22),

however, to our knowledge, no validated procedure for HLA class I sequence-based typing using plasma as source of DNA has been published. Here, we describe an ultrasensitive sequence-based HLA class I typing method that uses genomic DNA extracted from frozen plasma as starting material.

2.3. Methods

2.3.1. *Source of samples*

This study used a retrospective test panel comprised of matched frozen PBMC and plasma samples from 25 participants of the Vancouver Injection Drug Users Study-I (VIDUS-I) (23) collected as far back as 2001 (median age 8.5 years; interquartile range [IQR]; 8.1-9.0 years) and stored at -80°C until use. All participants provided written informed consent.

2.3.2. *DNA extraction and quantification*

Total nucleic acids were extracted from ~2.5 million PBMC resuspended in 200µl of H₂O and 500µl of plasma using a commercially-available silica-based kit (Invitrogen PureLink™ Genomic DNA kit; Life Technologies, Burlington, ON, Canada) according to the manufacturer's 'Mammalian Cells and Blood Lysate' protocol, however, the RNase treatment step was omitted. All extracts were assigned numeric ID codes and subsequent manipulations were performed in a blinded fashion. Total nucleic acids were eluted into 55µl elution buffer (supplied by manufacturer). Median and interquartile ranges [IQR] of total nucleic acid yields from PBMC and plasma were 85.7 [47.0-130.0] ng/µl and 2.2 [1.7-2.6] ng/µl, respectively, as measured on a NanoDrop™ 2000 spectrophotometer (Thermo Scientific Wilmington, DE, USA).

2.3.3. Amplification of HLA-A, B and C loci

Independent nested polymerase chain reactions (PCR) were used to amplify ~1000 base pair (bp) regions spanning exons 2 and 3 of HLA-A, B and C using universal, locus-specific primers taken directly from (or modified from) the published literature (24-26). The primers and their published origins are detailed in Table 2.1. The use of nested PCR ensures that sufficient template is generated for downstream sequencing and also reduces the amplification of nonspecific products, yielding improved specificity and amplicon purity that results in a higher-quality DNA sequence. All primers were purified by standard desalting (Integrated DNA Technologies, Coralville, IA, USA); in our experience HPLC, PAGE or other oligonucleotide purification methods were unnecessary. PCR reactions were prepared using the Roche Expand™ High Fidelity PCR system (Roche Applied Science, Laval, PQ, Canada).

First round PCR master mixes for HLA-A, B and C amplification were prepared as follows, for a final reaction volume of 50µl per sample: 27.5µl of molecular-grade H₂O, 5µl of 10X Expand™ High Fidelity Buffer containing 15mM MgCl₂ (supplied by manufacturer), 2µl of 25mM MgCl₂ (supplied by manufacturer), 0.5µl of 25mM dNTPs (Roche Applied Science, Laval, PQ, Canada), 3µl each of 10µM forward and reverse primers (Table 2.1) and 1µl Expand™ High Fidelity Enzyme Mix (3.5U/µl; supplied by manufacturer). Eight µl nucleic acid extract was added to each reaction, representing median template input quantities of 685.6 [IQR: 376.0-1040.0] (PBMC-derived) and 17.6 [IQR: 13.6-20.8] (plasma-derived) ng total nucleic acid per reaction. Negative (water) controls were included for each locus in every run. PCR amplification was performed using an ABI 9700 thermal cycler (Applied Biosystems, Foster City, CA, USA); cycling consisted of an initial denaturation step of 3 minutes @ 95°C, followed by 35 cycles of

(20 seconds @ 95°C, 20 seconds @ 60°C, 50 seconds @ 72°C) and finishing with a single 7 minute extension at 72°C.

Second round PCR master mixes for HLA-A, B and C were prepared as follows, for a final volume of 20µl per sample: 10µl of molecular-grade H₂O, 2µl 10X Expand™ High Fidelity Buffer containing 15mM MgCl₂, 0.8µl 25mM MgCl₂, 2µl of 60% Sucrose Buffer containing 0.08% Cresol Red (note: this reagent is optional and can be substituted with molecular-grade H₂O; its purpose is to serve as a pH indicator as well as to allow amplicons to be loaded directly onto an agarose gel without the addition of loading buffers), 0.16µl of 25mM dNTPs, 1.1µl of 10 µM forward and reverse primers (Table 2.1) and 0.29µl Expand™ High Fidelity Enzyme Mix. A total of 2.6µl of first round PCR product was transferred to each reaction as template. Second round PCR consisted of an initial denaturation of 3 minutes @ 95°C, followed by 35 cycles of (20 seconds @ 95°C, 20 seconds @ 62°C, 50 seconds @ 72°C) and finishing with a 7 min extension at 72°C. Eight µl of each 2nd round reaction was loaded directly onto a 1% agarose gel to confirm successful amplification prior to DNA sequencing (Figure 2.1).

Table 2.1. Nested PCR and sequencing primers used for HLA-A, B and C typing.

HLA	1 st Round PCR	2 nd Round PCR	Sequencing Primer
A	AF: GAAACSGCCTCTG YGGGAGAGAAGCAA (109-134) ^a	AS1F: AGCCGCGCCCKGGASGAGGGTC (168-188) ^c	AS2R: GGCCCGTCGCTGGGGGATGAG (525-545) ^e
	AR: TGTGGGTCCCAATTGTCTCCCTC (1070-1093) ^b	AS4R: TGTGGAGGCCAGCCCGGGGAGA (1046-1067) ^d	AS4R (same as 2 nd round primer; see left)
			AS5F: CTACGTGACGACACGAGTT (295-315)
			AS6F: ACAGTCTCCGGGTCCGAGAT (555-574)
B			AS7R: TTTGGCCTMACYSAAAATGAAAC (638-661)
			AS8R: TTCCCGTTCTCCAGGTRTCTGC (964-985)
	BF_mod: GGGAGGAGMRAGGGGACCCSAG (409-430) ^f	BnewF: GCAGGGGGGGCGCGCAGGACC (427-446) ^h	BS1F: GGAGCCGCGCCGGGAGGAGGGTC (450-472) ^j
	BR: GGAGGCCCATCCCCGGGGACCTAT (1329-1351) ^g	BS4R: GGAGATGGGGAAGGCTCCCCACT (1305-1327) ⁱ	BS2R: GGATGGGAGTCTGTGACCT (797-815) ^k
			BS3F: ACKGKGTGACCGCGGGG (990-1007)
			BS11R: CGGGGTCACTACCCGKCCTC (766-785)
C			BS12F: GGCKGGGCCAGGGTCTCAC (1006-1025)
			BS13R: ACTGCCCTGTGTACCMGCGC (1288-1307)
	CF: AFGAGGKGCCKCCCGGCGA (398-418) ^m	CS1F: GGAGCCGCGCAGGGAGGWWGGTC (434-456) ^o	CS1F: (same as 2 nd round primer; see left)
	CR: GGAGATRGGGAAGGCTCCCCACT (1290-1312) ⁿ	CS7R: GGCTCCCCACTGCCCCYTGGTAC (1279-1300)	CS7R: GGCTCCCCACTGCCCCYTGGTAC (1279-1300)
			CS8F: CGGGTCTCAGCCMCTCCTC (460-478)
			CS9R: TGGATCTCAGACSGGGGAGACT (828-848)
			CS10F: GGGCTGACCRGGGGGGCG (978-995)
			CS11F: CGGGGCCAGGKTCTCACAYC (994-1013)

All primers are listed in 5'→3' orientation. Numbers in parentheses represent primer binding coordinates with respect to genomic reference sequences for HLA-A*01:01:01:01 (Genbank accession number AJ278305), HLA-B*07:02:01 (Genbank AJ292075) and HLA-C*01:02:01 (Genbank HM543696), respectively. ^a formerly 5AIn1-46 (24), ^b formerly 3AIn3-66 (24), ^c formerly 5AIn1-99 (26), ^d formerly 3AIn3-41 (26), ^e formerly 3AIn2-37 (26), ^f modified from 5BIn1-57 (24) to improve binding to B*44 and B*57 alleles, ^g formerly 3BIn3-37 (24), ^h modified from 5BIn1CG and 5BIn1TA (25) to be universal for B locus, ⁱ formerly 3BCIn3-12 (24), ^j formerly 5Bin1-77 (26), ^k formerly 3B44 (26), ^l formerly 5BInt2 (26), ^m modified from 5CIn1-61 (24), to improve binding to C*07 alleles, ⁿ modified from 3BCIn3-12 (24) to improve binding to C*15 alleles, ^o formerly 5CIn1-77 (26)

2.3.4. DNA sequencing

Second round amplicons were diluted approximately 15-fold by adding 180µl of molecular grade H₂O directly to the remaining 12µl of amplicon in the 2nd round PCR tubes. Dilution renders amplicon concentrations appropriate for DNA sequencing, and represents a rapid and cost-saving alternative to PCR purification using commercial kits. Amplicons were directly sequenced in both 5' and 3' directions using a set of six sequencing primers per locus. The sequencing primers are comprised of published (24, 26), modified and novel oligonucleotides (Table 2.1). Small-volume sequencing reactions (6µl final volume per sample) consisted of 2.6µl of 2µM sequencing primer, 2.1µl dilution buffer (175mM Trizma HCl, 1.25mM MgCl₂, pH=9.0) and 0.3µl of ABI Prism® BigDye® Terminator v3.1 sequencing mix (Applied Biosystems, Foster City, CA, USA), to which 1µl diluted 2nd round amplicon was added as a template. Ninety-six well plates were capped, vortexed and briefly centrifuged to ensure reactions were positioned at the bottom of each well. Thermal cycling for DNA sequencing consisted of 25 cycles of (10 seconds @ 96°C, 5 seconds @ 50°C and 55 seconds @ 60°C). Following amplification, sequencing reactions were purified via ethanol precipitation by adding 4µl of 1X EDTA-Sodium Acetate solution (containing 375mM Sodium Acetate and 62.5mM EDTA) and 40µl of chilled (-20°C) 100% ethanol to each well. Plates were incubated at -20°C for 30 minutes and then centrifuged for 20 minutes at 2000g after which the wash was discarded. A second wash using 155µl of 100% chilled ethanol was performed without centrifugation, and wells were allowed to air dry. DNA pellets were resuspended in 10µl of Hi-Di™ Formamide per well (Applied Biosystems, Foster City, CA, USA) and denatured by heating to 90°C for 2 min. Sequencing data was collected using either an

ABI 3130xl or 3730xl automated DNA analyzers (Applied Biosystems, Foster City, CA, USA).

2.3.5. *Sequence analysis and HLA allele interpretation*

Chromatograms were aligned to reference standards spanning exon 2, intron 2 and exon 3 for HLA-A (787 bp) and exons 2 and 3 separately (270 and 276 bp, respectively) for HLA-B and C, and analyzed using Sequencher® v4.9 (Gene Codes Corporation, Ann Arbor, MI, USA). Sequencing reference standards (derived from A*01:01:01:01, HLA-B*07:02:01 and HLA-C*01:02:01 respectively), are provided as supplementary material in Appendix A. Note that the presence of length polymorphisms within intron 2 among some HLA-B and C alleles will yield chromatogram phasing problems in this region during bulk sequencing. In these cases it is necessary to analyze exons 2 and 3 separately; for this reason, separate reference standards are provided for exons 2 and 3 for HLA-B and HLA-C. A minimum lower peak height of 25% was employed in Sequencher® v4.9 for nucleotide mixture detection. Analyzed sequences were exported as text files in FASTA format.

HLA allele interpretation was performed using a free web-based HLA class I sequence-based typing tool hosted at dbMHC at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/projects/gv/rbc/sbt.cgi?cmd=main>). In the case of HLA-A, the full 787 bp fragment was interpreted. HLA-B and C sequences for which intron 2 read-through was possible (i.e. cases where intron 2 was the same length for both alleles) were also interpreted as a single fragment; these ranged from 790-796 bp depending on the intron length. HLA-B and C sequences for which intron 2 read-through was not possible due to phasing problems (i.e. cases where intron 2 was a different

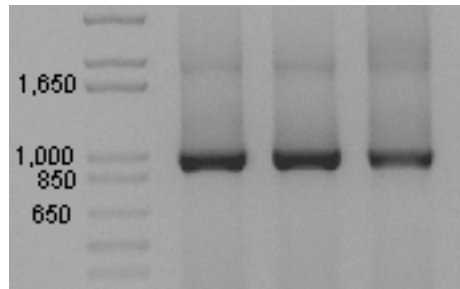
length for the two alleles expressed) were uploaded as individual fragments and interpreted as a pair.

2.4. Results and Discussion

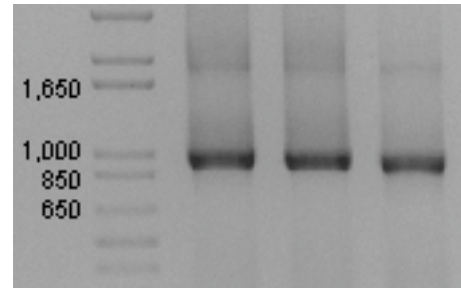
Total nucleic acids were successfully extracted, amplified and sequenced from PBMC and plasma samples in all 25 individuals. The median and interquartile range [IQR] total nucleic acid concentrations of PBMC and plasma-derived extracts were 85.7 [47.0-130.0] and 2.2 [1.7-2.6] ng/μl respectively. Note that total nucleic acid concentrations reflect the presence of both cellular and plasma HIV-1 viral RNA in these extracts (RNase treatment was not performed during the isolation procedure), therefore, genomic DNA concentrations are likely to be substantially lower.

Nested PCR amplification yielded high quality bands for HLA-A, B and C for both PBMC and plasma-derived DNA in all cases (Figure 2.1). Sequence chromatograms were of uniformly high quality with clearly distinguishable mixed bases regardless of DNA source (Figure 2.2). Out of a total of >46,975 nucleotides analyzed, comparison of PBMC and plasma-derived sequences revealed only one partial base mismatch in a single specimen within HLA-C exon 2 (not shown); here, the plasma-derived sequence revealed a C where the PBMC-derived sequence revealed the correct mixed base call of S (C/G mixture), yielding >99.9% concordance at the nucleotide level. This single discrepancy did not affect the HLA-C interpretation of this specimen.

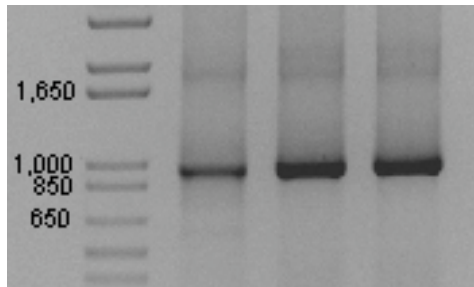
(a) PBMC HLA-A



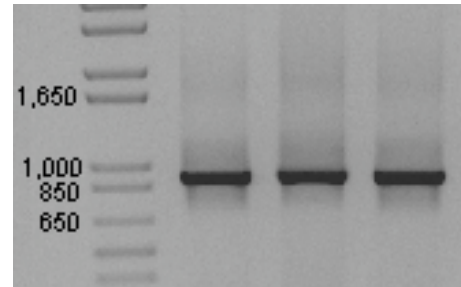
(b) Plasma HLA-A



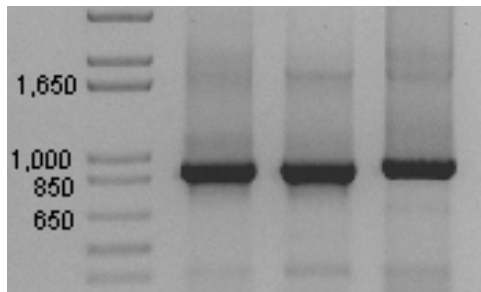
(c) PBMC HLA-B



(d) Plasma HLA-B



(e) PBMC HLA-C



(f) Plasma HLA-C

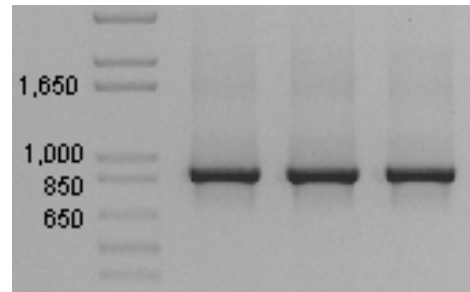


Figure 2.1. Amplification of HLA-A, B and C using PBMC vs. plasma-derived genomic DNA.

Representative amplicons for HLA-A (panels a and b), HLA-B (panels c and d) and HLA-C (panels e and f) were resolved on a 1% agarose gel and bands were visualized under UV light alongside DNA size standards (Invitrogen 1kb+ DNA ladder; Life Technologies, Burlington, ON, Canada) after staining with SYBR safe DNA Gel Stain (Invitrogen). Amplification of HLA-A, B and C was uniformly successful regardless of genomic source.

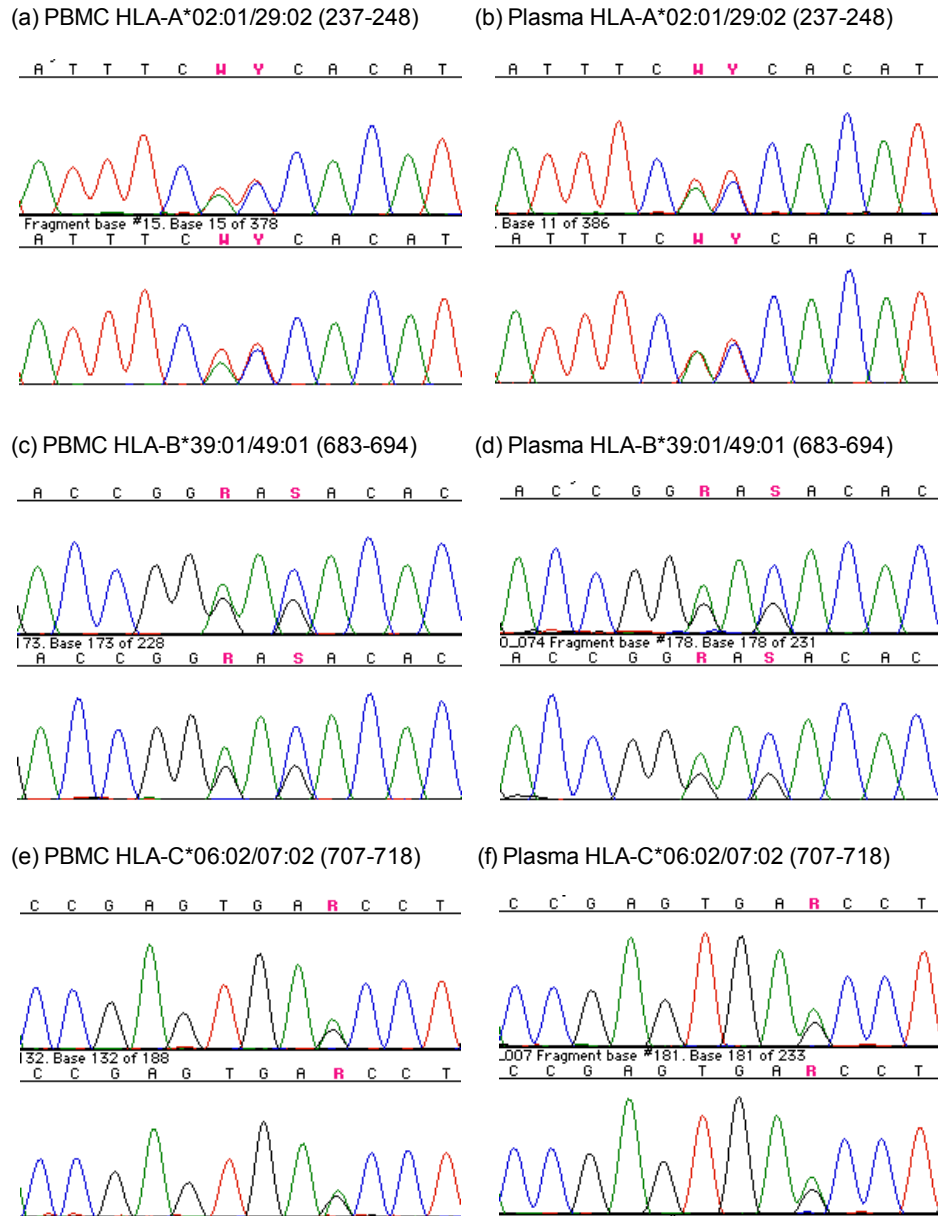


Figure 2.2. Quality comparison of sequence chromatograms for HLA-A, B and C alleles from matched PBMC vs. plasma-derived amplicons.

Representative chromatograms depict a 12 base region within exon 2 for HLA-A (panels a and b), HLA-B (panels c and d) and HLA-C (panels e and f) loci from the same patient. Numbers in parentheses represent nucleotide coordinates with respect to genomic reference sequences for HLA-A*01:01:01:01 (Genbank accession number AJ278305), HLA-B*07:02:01 (Genbank AJ292075) and HLA-C*01:02:01 (Genbank HM543696), respectively. Upper and lower chromatograms represent two independent sequence reads for each sample, yielding a minimum of double coverage over the entire sequence. Mixed bases are denoted using International Union of Pure and Applied Chemistry (IUPAC) ambiguity codes: W (A/T), Y (C/T), R (A/G) and S (C/G). Chromatograms are uniformly in phase, of high quality and mixed bases are clearly distinguishable, regardless of genomic DNA source.

Interpretation of HLA class I sequences using the NCBI-hosted dbMHC SBT bioinformatics tool yielded 100% concordance at both the HLA allele and protein levels between PBMC and plasma-derived sequences for all loci. Our test panel of 25 matched specimen pairs comprised a total of 12, 16 and 11 unique HLA-A, B and C alleles at the group level, and 16, 22 and 14 unique HLA-A, B and C alleles at the protein level, respectively (Table 2.2), indicating that this method is capable of amplifying and detecting a wide variety of different HLA class I alleles at high-level resolution. Note that where ambiguous allele combinations arose at the protein level (either due to identical sequences over exons 2 and 3 or identical patterns of mixed bases in the bulk sequence), Table 2.2 lists the likeliest allele pair as defined by allele frequency. For this reason we conservatively state that our method achieves intermediate-to-high level resolution, although the latter is usually achieved. Moreover, the availability of web-based bioinformatic HLA completion algorithms can be used to predict the most likely allele(s) at the protein level, given mixed-resolution or missing data at one or more loci ((27); <http://atom.research.microsoft.com/HLACompletion/>).

We have also investigated the reliability of this HLA typing method using different starting volumes of plasma (not shown). Reliable (>90%) amplification of all loci is generally achieved using 200µl plasma as starting material. Attempts to HLA class I type using 50ul of plasma as starting material generally yielded reliable (>90%) amplification for HLA-B and HLA-C but poor (<50%) amplification for HLA-A (not shown). In addition, the risk of amplifying only one of the two alleles at each locus, yielding a false homozygous result, tends to increase with declining DNA input. We, therefore, recommend that a minimum of 200µl of plasma be used as starting material for this procedure, and that homozygous results be confirmed by re-extraction, re-amplification and re-sequencing.

Table 2.2. HLA class I alleles represented in the test panel.

Specimen ID	HLA-A1	HLA-A2	HLA-B1	HLA-B2	HLA-C1	HLA-C2
1	A*11:01	A*68:02	B*14:02	B*27:05	C*01:02	C*08:02
2	A*02:01	A*29:02	B*07:02	B*18:01	C*07:01	C*07:02
3	A*02:01	A*29:02	B*39:24	B*44:03	C*07:01	C*16:01
4	A*01:01	A*24:23	B*08:01	B*40:01	C*03:04	C*07:01
5	A*02:01	A*31:01	B*27:05	B*35:01	C*02:02	C*04:01
6	A*01:01	A*01:01	B*07:02	B*37:01	C*06:02	C*07:02
7	A*24:02	A*24:02	B*35:01	B*35:01	C*04:04	C*04:04
8	A*24:23	A*31:01	B*51:01	B*51:01	C*03:04	C*15:02
9	A*02:01	A*68:01	B*39:01	B*57:01	C*06:02	C*07:02
10	A*02:01	A*68:01	B*15:10	B*44:02	C*03:04	C*05:01
11	A*11:01	A*25:01	B*07:02	B*58:01	C*07:01	C*07:02
12	A*02:01	A*32:01	B*39:11	B*57:01	C*06:02	C*07:02
13	A*01:01	A*11:01	B*08:01	B*15:01	C*03:04	C*07:01
14	A*02:01	A*29:02	B*27:05	B*44:03	C*01:02	C*16:01
15	A*02:01	A*31:01	B*35:01	B*39:01	C*02:02	C*07:02
16	A*24:02	A*24:02	B*40:02	B*40:02	C*03:04	C*03:04
17	A*31:01	A*68:02	B*51:01	B*53:01	C*04:01	C*15:09
18	A*02:01	A*02:06	B*39:01	B*39:01	C*07:02	C*07:02
19	A*02:01	A*26:08	B*39:01	B*49:01	C*07:01	C*07:02
20	A*23:01	A*25:01	B*35:08	B*44:02	C*04:01	C*05:01
21	A*24:02	A*31:01	B*40:01	B*40:02	C*03:04	C*03:04
22	A*03:01	A*24:23	B*07:02	B*40:02	C*03:04	C*07:02
23	A*03:01	A*31:01	B*18:01	B*39:01	C*05:01	C*12:03
24	A*01:01	A*11:100	B*08:01	B*35:01	C*04:01	C*07:01
25	A*02:01	A*32:01	B*15:01	B*51:01	C*05:01	C*12:03

A previous version of this protocol (which differed slightly in the concentration of PCR primers used) was validated against International Histocompatibility Working Group (IHWG) reference panels (28) and has been used to perform HLA class I typing from a

variety of specimen types including: PBMC, whole blood, granulocyte cell pellets, plasma and sera, plus archived sera over 30 years old (11, 28-32).

In the mid-1990s, when the first universal locus-specific PCR primers on which this method is based were published (24), a total of 78, 173 and 42 HLA-A, B and C alleles had been reported by the World Health Organization (33). As of January 2012, this list has grown to 1757, 2338 and 1304 HLA-A, B and C alleles respectively ((2); <http://hla.alleles.org/nomenclature/stats.html>). In the mid 2000s when our laboratory first started to develop the present protocol, we modified some of the original published primers by adding one or more degenerate bases to improve the amplification efficiency for certain alleles including members of the B*07, B*44, B*57, B*07 and C*15 allele groups (Table 2.1). Comparison of our PCR primer sequences to all currently available HLA class I sequences ((2); <http://www.ebi.ac.uk/imgt/hla/align.html>, accessed January 27, 2012) indicates near perfect coverage for all allele members at each locus; HLA-A*11:50Q and possibly HLA-B*73:01 represent the only cases where published intronic sequences mismatch one or more primers sufficiently to compromise amplification.

2.5. Conclusions

In conclusion, we have described a relatively rapid, cost-effective and accurate sequence-based method to perform intermediate-to-high-resolution HLA class I typing using frozen plasma as a source of genomic DNA. This method provides comparable performance to conventional sequence-based approaches requiring substantially higher template DNA concentrations, and can also be used on genomic DNA from PBMC, blood and other sources. This method could be useful in research studies, including

retrospective molecular epidemiological surveys, where cell samples are unavailable or DNA quantities are limiting.

2.6. References

1. Howell WM, Carter V, Clark B. The HLA system: immunobiology, HLA typing, antibody screening and crossmatching techniques. *J Clin Pathol*. 2010 May;**63**(5):387-90.
2. Robinson J, Mistry K, McWilliam H, Lopez R, Parham P, Marsh SG. The IMGT/HLA database. *Nucleic acids research*. 2011 Jan;**39**(Database issue):D1171-6.
3. Mallal S, Phillips E, Carosi G, et al. HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med*. 2008 Feb 7;**358**(6):568-79.
4. Reveille JD. The genetic basis of spondyloarthritis. *Annals of the rheumatic diseases*. 2011 Mar;**70 Suppl 1**:i44-50.
5. Bax M, van Heemst J, Huizinga TW, Toes RE. Genetics of rheumatoid arthritis: what have we learned? *Immunogenetics*. 2011 Aug;**63**(8):459-66.
6. Jones EY, Fugger L, Strominger JL, Siebold C. MHC class II proteins and disease: a structural perspective. *Nature reviews. Immunology*. 2006 Apr;**6**(4):271-82.
7. Carrington M, O'Brien SJ. The influence of HLA genotype on AIDS. *Annual review of medicine*. 2003;**54**:535-51.
8. Fellay J, Shianna KV, Ge D, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science*. 2007 Aug 17;**317**(5840):944-7.
9. Kuniholm MH, Kovacs A, Gao X, et al. Specific human leukocyte antigen class I and II alleles associated with hepatitis C virus viremia. *Hepatology*. 2010 May;**51**(5):1514-22.
10. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science*. 2002 May 24;**296**(5572):1439-43.
11. Brumme ZL, Brumme CJ, Heckerman D, et al. Evidence of differential HLA class I-mediated viral evolution in functional and accessory/regulatory genes of HIV-1. *PLoS pathogens*. 2007 Jul;**3**(7):e94.
12. Timm J, Li B, Daniels MG, et al. Human leukocyte antigen-associated sequence polymorphisms in hepatitis C virus reveal reproducible immune responses and constraints on viral evolution. *Hepatology*. 2007 Aug;**46**(2):339-49.

13. Andolfatto S, Namour F, Garnier AL, Chabot F, Gueant JL, Aimone-Gastin I. Genomic DNA extraction from small amounts of serum to be used for alpha1-antitrypsin genotype analysis. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology*. 2003 Feb;**21**(2):215-9.
14. Blomeke B, Bennett WP, Harris CC, Shields PG. Serum, plasma and paraffin-embedded tissues as sources of DNA for studying cancer susceptibility genes. *Carcinogenesis*. 1997 Jun;**18**(6):1271-5.
15. Dixon SC, Horti J, Guo Y, Reed E, Figg WD. Methods for extracting and amplifying genomic DNA isolated from frozen serum. *Nature biotechnology*. 1998 Jan;**16**(1):91-4.
16. Martin M, Carrington M, Mann D. A method for using serum or plasma as a source of DNA for HLA typing. *Human immunology*. 1992 Feb;**33**(2):108-13.
17. Fowke KR, Plummer FA, Simonsen JN. Genetic analysis of human DNA recovered from minute amounts of serum or plasma. *Journal of immunological methods*. 1995 Mar 13;**180**(1):45-51.
18. Lin Z, Floros J. Genomic DNA extraction from small amounts of sera to be used for genotype analysis. *BioTechniques*. 1998 Jun;**24**(6):937-40.
19. Sandford AJ, Pare PD. Direct PCR of small genomic DNA fragments from serum. *BioTechniques*. 1997 Nov;**23**(5):890-2.
20. Abbott WG, Tukuitonga CF, Ofanoa M, Munn SR, Gane EJ. Low-cost, simultaneous, single-sequence genotyping of the HLA-A, HLA-B and HLA-C loci. *Tissue Antigens*. 2006 Jul;**68**(1):28-37.
21. Cinek O, Wilkinson E, Paltiel L, Saugstad OD, Magnus P, Ronningen KS. Screening for the IDDM high-risk genotype. A rapid microtitre plate method using serum as source of DNA. *Tissue antigens*. 2000 Oct;**56**(4):344-9.
22. Doxiadis I, Grosse-Wilde H. Typing for HLA class I gene products using plasma as source. *Vox sanguinis*. 1989;**56**(3):196-9.
23. Wood E, Li K, Miller CL, et al. Baseline self-perceived risk of HIV infection independently predicts the rate of HIV seroconversion in a prospective cohort of injection drug users. *International journal of epidemiology*. 2005 Feb;**34**(1):152-8.
24. Cereb N, Maye P, Lee S, Kong Y, Yang SY. Locus-specific amplification of HLA class I genes from genomic DNA: locus-specific sequences in the first and third introns of HLA-A, -B, and -C alleles. *Tissue Antigens*. 1995 Jan;**45**(1):1-11.
25. Cereb N, Yang SY. Dimorphic primers derived from intron 1 for use in the molecular typing of HLA-B alleles. *Tissue antigens*. 1997 Jul;**50**(1):74-6.
26. Dunn PP, Day S, Williams S, Bendukidze N. DNA sequencing as a tissue-typing tool. *Methods Mol Med*. 2004;**91**:233-46.

27. Listgarten J, Brumme Z, Kadie C, et al. Statistical resolution of ambiguous HLA typing data. *PLoS computational biology*. 2008 Feb;**4**(2):e1000016.
28. Brumme ZL, Brumme CJ, Chui C, et al. Effects of human leukocyte antigen class I genetic parameters on clinical outcomes and survival after initiation of highly active antiretroviral therapy. *The Journal of infectious diseases*. 2007 Jun 1;**195**(11):1694-704.
29. Brumme ZL, John M, Carlson JM, et al. HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. *PloS one*. 2009;**4**(8):e6687.
30. Brumme ZL, Brumme CJ, Carlson J, et al. 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. *Journal of virology*. 2008 Sep;**82**(18):9216-27.
31. Miura T, Brumme ZL, Brockman MA, et al. Impaired replication capacity of acute/early viruses in persons who become HIV controllers. *Journal of virology*. 2010 Aug;**84**(15):7581-91.
32. Brumme CJ, Chan B, Markle T, et al. Molecular and Functional Consequences of Immune-Mediated Evolution in HIV-1 Nef During the North American Epidemic. *Aids Res Hum Retrov*. 2011 Oct;**27**(10):A60-A1.
33. Bodmer JG, Marsh SG, Albert ED, et al. Nomenclature for factors of the HLA system, 1996. *Tissue antigens*. 1997 Mar;**49**(3 Pt 2):297-321.

3. Chapter 3: Immune Escape Pathways and Replication Capacity in HIV-1 *Gag* have Remained Largely Consistent over the Past 30 Years of the Epidemic in North America

3.1. Abstract

Immune-driven viral evolution occurs along broadly predictable pathways in HIV-1 infected individuals. However, the extent to which population-level immune escape patterns have changed over the course of the epidemic and the impact of this evolution on viral fitness remains incompletely understood. Here we examine population-level patterns of immune escape within HIV-1 *gag* using a unique historic cohort of HIV-positive individuals from various sites across North America (1979 to 1989).

Phylogenetically-informed methods were used to estimate the date of the original founder virus in North America and identify HLA-associated polymorphisms in historic HIV-1 *gag* sequences (N=239) for comparison to those previously identified in modern HIV-1 *gag* sequences ($q \leq 0.2$). Additionally, we generated chimeric viruses in an HIV-1 NL4-3 backbone using cloned *gag* sequences from 80 historic (1979-1989) and 58 modern (2002-2008) patient samples and assessed their ability to replicate in vitro. Timing of the most recent common ancestor dates the original founder virus in North America to 1960. Comparison of historic and modern patterns of immune escape ($q \leq 0.2$) revealed relatively consistent patterns of adaptation rather than substantial accumulation

of escape mutations over time. Furthermore, we observed no major differences in replication capacity when comparing historic *gag* sequences with modern *gag* sequences, although there was a modest trend towards increased replication capacity for sequences sampled after 1985 compared to those sampled earlier. Taken together, our results suggest that patterns of HIV-1 adaptation have remained largely consistent over the last 30 years of the epidemic such that CTL immune responses targeted towards conserved regions within *gag* will continued to provide protection against the most commonly circulating strains of HIV-1 in North America.

3.2. Introduction

Human leukocyte antigen (HLA) class I restricted cytotoxic T-lymphocytes (CTL) play a key role in controlling viral replication in the early stages of HIV-1 infection (1-5). However, CTL control is quickly diminished by the selection of escape mutations throughout the viral genome that serve to disrupt HLA class I binding, interfere with intracellular epitope processing and/or abrogate CTL recognition of the epitope-HLA complex (6-16). Characterization of these mutations has revealed broadly predictable patterns of escape such that specific amino acid polymorphisms in HIV-1 are associated with the expression of certain HLA class I alleles by the host (17). More recently, HLA-restricted immune escape mutations across the HIV-1 proteome have been mapped using large cross-sectional cohorts of individuals with linked HLA class I and HIV-1 sequences (18-31). However, the extent to which population-level immune escape patterns have changed over the course of the HIV-1 epidemic remains largely unknown. Developing a better understanding of host-driven viral adaptation is important for vaccine design.

Incomplete understanding of the HLA-restricted CTL responses that contribute to effective and prolonged suppression of HIV-1 replication remains a key knowledge gap in vaccine research. The appearance of immune escape mutations throughout the HIV-1 proteome underscores the importance of CTL responses in initial immune control of HIV-1 infection (32, 33), however, the contributions of these responses (and the consequences of escape) to long term HIV-1 immune control remains unclear. Enrichment of protective HLA class I alleles such as HLA-B*27 and HLA-B*57 in individuals with favourable HIV-1 clinical outcomes (including elite controllers) additionally supports the antiviral activity exerted by CTLs (34-41), but again the precise mechanisms underlying these protective effects remain incompletely known (42). To date efforts to develop CTL-based vaccines against HIV-1 have met with little success (43-45). The enormous diversity of HIV-1 and its ability to escape from protective immune responses remains a key obstacle in vaccine design (15, 32). One approach in overcoming this barrier is to target epitopes within the HIV-1 proteome that are relatively conserved such that escape would result in a significant fitness cost to the virus (46). Presumably, mutations that diminish viral replication capacity and/or protein function could be selected only with great difficulty, even in the presence of strong immune selection pressures. Thus, developing a vaccine that stimulates CTL responses toward key highly conserved epitopes, thereby forcing the selection of fitness-reducing mutations, may represent a strategy to achieve long lasting, vaccine-induced immune suppression (47, 48).

Currently, vaccine strategies have favoured targeting the major HIV-1 structural proteins encoded by *gag*, most notably the highly conserved yet immunogenic p24 capsid protein (34, 36, 49-56). *Gag* proteins play a crucial role in uncoating during viral entry and assembly of HIV-1 virions at the host cell surface during budding (57, 58).

Furthermore, the *gag* capsid protein p24 ranks among the most conserved regions of the HIV-1 genome (59). Consequently, key escape mutations in *gag* have been shown to incapacitate proper *gag* function and result in significant fitness costs to the virus (41, 60-64). An example of this is the well-characterized T242N escape mutation that occurs at position 3 of the TW10 (*gag* residues 240-249) epitope of p24 *gag* in individuals expressing the “protective” alleles HLA-B*57 and/or HLA-B*5801 (65, 66). Selection of the T242N escape mutation is thought to confer a modest yet biologically significant cost to viral replication capacity (61, 62, 64, 66). Further evidence to support a fitness cost of the T242N mutation is its tendency to revert following transmission to a recipient that does not express HLA-B*57 and/or HLA-B*5801 (65). In vivo reversion of escape mutations following transmission suggest that mutations in epitopes that confer a fitness cost are not sustainable at the population-level (28, 67-69). However, an important exception to this generalization is cases where fitness-reducing mutations are compensated through mutations at secondary sites (70).

Compensatory mutations have been shown to partially or fully restore fitness costs of escape at a primary site (61, 63) such that reversion upon transmission may not occur (71). A well-known example is the R264K escape mutation that occurs within the KK10 (*gag* residues 263-272) epitope of p24 *gag* in individuals expressing HLA-B*27 (14, 63). Appearance of the R264K mutation is closely associated with an upstream compensatory S173A mutation that restores viral replication in vitro (63). Although HLA-B*27-restricted epitopes are associated with long-term immune control (63), reversion of the R264K mutation upon transmission to a host that does not express HLA-B*27 is not always observed in vivo due to the presence of the S173A mutation (9, 70, 72). If indeed certain escape mutations were to persist upon transmission to hosts lacking the restricting HLA allele, it is possible that over time, these mutations could accumulate to

the point of fixation in circulating HIV-1 sequences. In other words, HIV-1 could be adapting to common HLA alleles (and/or alleles mounting strong immune selection pressures) in a given population over time. In fact, this is precisely what the “HLA imprinting” hypothesis proposes.

HLA “imprinting” is the idea that HIV-1 escape mutations selected by the most commonly circulating HLA alleles may accumulate over time such that they become fixed in a given population (72-74). A major consequence of the HLA imprinting hypothesis would be the “disappearance” of key immunogenic epitopes from circulating HIV-1 sequences as a result of accumulating escape mutations (75). Accordingly, CTL responses upon infection would be substantially diminished, potentially resulting in increasing levels of plasma viremia in infected individuals as the epidemic progresses. Furthermore, the protective effects of HLA alleles such as HLA-B*27, HLA-B*57 and HLA-B*5801 could cease to provide benefits in future generations (75). HLA-imprinting is particularly concerning in HIV-1 vaccine development, as the fixation of immune escape mutations could render CTL-based vaccine strategies ineffective (32). However, the relationship between individual-level and population-level HIV-1 evolution is complex and incompletely understood. Characterization of population-level HIV-1 adaptation over time, and its potential consequence to viral fitness, has to date been limited by the scarcity of historic host-viral genetic datasets.

The purpose of this study, therefore, was to explore the role of host immune selection pressures on population-level HIV-1 adaptation using a unique historic cohort of HIV-1 positive individuals from New York, Boston and San Francisco dating from 1979 to 1989. Historic *gag* sequences were used to estimate the date of the initial founder virus, compare historic versus modern escape patterns and assess the functional

capacity of historic versus modern *gag* sequences. To our knowledge, this is the first population-level study to use a cohort of HIV-1 positive individuals dating as far back as 1979 to compare immune escape and replicative capacity over the course of the epidemic in North America.

3.3. Methods

3.3.1. Cohort description

Immune-mediated viral evolution studies: HLA-associated polymorphisms in *gag* were identified in a historic cohort of 239 recovered serum samples from antiretroviral naïve, chronically-HIV-1-subtype B infected individuals dating from 1979-1989. Twenty-two of these samples were recovered from the New York Blood Center in New York, USA (1979-1984) and 217 were recovered from the Fenway Community Health Clinic in Boston, USA (1985-1989). Median age and interquartile range [IQR] of samples was 24.43 [23.88-26.39] years. HLA-associated polymorphisms in *gag* were compared to a modern cohort (year 2000+) of 534 antiretroviral naïve, chronically-HIV-1-subtype B infected individuals from the baseline time point of the AIDS Clinical Trials Group (ACTG) 5142/5128 studies undertaken across the USA (76). HLA-associated polymorphisms in this modern cohort have previously been published as part of the International HIV Adaptation Collaborative (IHAC), a multicenter international cohort (22).

Viral replication capacity studies: Replication capacity of recombinant viruses expressing patient-derived clonal *gag* sequences was assessed using a total of 80 historic specimens. Twenty clones were generated from specimens recovered from the New York Blood Center in New York (1979-1984), 56 from specimens recovered from

the Fenway Community Health Clinic in Boston (1985-1989) and 4 from specimens recovered from the San Francisco Department of Public Health (1984). Median age of samples was 25.78 [IQR: 24.15-27.51] years. Phylogenetic analysis of historic clones, assessed through visual evaluation of sequence dispersion, revealed no substantial clustering according to geographic region (New York, Boston, San Francisco) so we treated these samples as one cohort. Historic replication capacity was then compared to replication capacity in a modern control group of plasma samples from 58 antiretroviral naïve, chronically-HIV-1-subtype B infected individuals from New York and Boston. Of these 58 modern samples, 15 were from the Aaron Diamond AIDS Research Center in New York (2003-2006) and 43 were from the Massachusetts General Hospital in Massachusetts (2002-2008). Median sample age was 8.12 [IQR: 6.15-9.08] years. Historic and modern samples were assigned alphanumeric identification codes starting with either a Y (historic) or E (modern).

3.3.2. *Total nucleic acid isolation*

Total nucleic acids were manually extracted from 500µl of serum or plasma using the Invitrogen PureLink™ Genomic DNA kit (Life Technologies, Burlington, ON, Canada) according to the manufacturer's 'Mammalian Cells and Blood Lysate' protocol. The RNase treatment step was omitted to recover total nucleic acids. Samples were eluted into 55µl elution buffer and stored at -80°C.

3.3.3. *Amplification of HIV-1 gag*

HIV-1 *gag* (HXB2 nt 790-2292) was amplified from plasma HIV-1 RNA using nested reverse transcription polymerase chain reaction (RT-PCR) with the following 1st and 2nd round primers: forward 1st round, AAATCTCTAGCAGTGGCGCCCGAACAG

(HXB2 nt 643-649), reverse 1st round, TTTAACCCTGCTGGGTGTGGTATYCCT (HXB2 nt 2851-2825), forward 2nd round, GCAGGACTCGGCTTGCTGAA (HXB2 nt 691-710), and reverse 2nd round, TATCATCTGCTCCTGTATC (HXB2 nt 2343-2325). First round RT-PCR was performed using the Invitrogen SuperScript® III One-Step RT-PCR system with Platinum® *Taq* DNA Polymerase High Fidelity (Life Technologies, Burlington, ON, Canada). Each reaction mixture was prepared to a final volume of 40µl as follows: 13.4µl of molecular-grade H₂O, 20µl SuperScript® 2x Reaction Mix: 0.4mM of each dNTP, 2.4mM MgSO₄ (supplied by manufacturer), 1µl each of 10µM forward and reverse primers, 0.6µl of SuperScript® HiFi enzyme and 4µl of RNA. First round thermal cycling conditions consisted of an initial denaturation step of 30 minutes @ 55°C and 2 minutes @ 94°C, followed by 35 cycles of (15 seconds @ 94°C, 30 seconds @ 55°C, 2 minutes @ 68°C) and finishing with a 7 minute extension at 68°C. Second round PCR was performed using the Roche Expand™ High Fidelity PCR system (Roche Applied Science, Laval, PQ, Canada). A final volume of 21µl was prepared as follows: 14.4µl of molecular-grade H₂O, 2µl of 10X Expand™ High Fidelity Buffer containing 15mM MgCl₂ (supplied by manufacturer), 0.8µl of 25mM MgCl₂ (supplied by manufacturer), 0.16µl of 25mM dNTPs (Roche Applied Science, Laval, PQ, Canada), 0.375µl each of 10µM forward and reverse primers, 0.29µl Expand™ High Fidelity Enzyme Mix (3.5U/µl) and 2.6µl of first round PCR product. Second round PCR conditions included an initial 2 minute denaturation step at 94°C, followed by 35 cycles of (15 seconds @ 94°C, 30 seconds @ 55°C, 2 minutes @ 72°C) and finishing with a 7 minute extension at 72°C.

3.3.4. Amplification of HLA-A, B and C loci

Total nucleic acids extracted using 200µl of serum or plasma were also used for HLA class I amplification using the methods described in Chapter 2.

3.3.5. Sequencing

Second round HLA-A, B and C PCR products were directly sequenced using the methods described in Chapter 2. HIV-1 *gag* amplicons were bulk sequenced in both 5' and 3' directions using a set of eight sequencing primers. In the 5' direction the following primers were used: TTGACTAGCGGAGGCTAGAAGG (HXB2 nt 762-783), GTTAAGTGTTCATTGTGG (HXB2 nt 1957-1977), CATCAATGAGGAAGCTGCAGATGGGA (HXB2 nt 1401-1427) and CACTAGAAGAAATGATGACAGC (HXB2 nt 1814-1835). In the 3' direction the following primers were used: TTATCTAAAGCTTCCTTGGTGTCT (HXB2 nt 1097-1074), GGTTCTCTCATCTGGCCTGG (HXB2 nt 1481-1462), AAAATAGTCTTACAATCTGG (HXB2 nt 1790-1771) and CAGCCCTTTTTCCTAGGGGCCYTGC (HXB2 nt 2027-2003). As discussed in Chapter 2, amplicons were sequenced using either an ABI 3130xl or 3730xl automated DNA analyzers (Applied Biosystems, Foster City, CA, USA) and chromatograms were analyzed using Sequencher® v4.9 (Gene Codes Corporation, Ann Arbor, MI, USA) where a minimum lower peak height of 25% was employed for nucleotide mixture detection. HLA allele interpretation was performed using a free web-based HLA class I sequence-based typing tool hosted at dbMHC at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/projects/gv/rbc/sbt.cgi?cmd=main>). HIV-1 *gag* sequence alignment was performed using MUSCLE and phylogenetic analysis was performed

using phyML as implemented in the “one-step” phylogenetic analysis package available at (http://www.phylogeny.fr/version2_cgi/index.cgi).

3.3.6. *Identification of HLA-associated polymorphisms in the HIV-1 gag gene*

Previously, the identification of HLA-associated polymorphisms in large cross-sectional cohorts has been challenged by three confounding factors: HIV-1 viral lineage effects, HIV-1 amino acid co-variation and linkage disequilibrium between HLA class I alleles (19, 25). Statistical tests that do not address these confounders may result in spurious associations and the inability to discriminate between primary escape sites (“direct” selection pressures) and secondary/compensatory mutations (“indirect” selection pressures). Here we adopted a phylogenetically-informed analysis method previously developed by Carlson et al. (25) that simultaneously corrects for HIV-1 amino acid co-variation and HLA linkage disequilibrium. Briefly, HIV-1 *gag* sequences from our historic cohort were used to create a maximum-likelihood phylogenetic tree. In order to correct for viral lineage effects, a likelihood ratio test was used for each observed amino acid at each *gag* codon to test whether a model incorporating both phylogenetic structure and HLA-derived selection pressure explained the data significantly better than the null hypothesis (a model assuming neutral evolution according to the tree alone). HLA-associated polymorphisms were then identified on a codon-by-codon basis whereby the presence or absence of a given HLA class I allele was correlated with the presence or absence of a given amino acid at every codon in *gag*. Analysis of polymorphisms was performed at both the HLA allele and protein levels and the association that achieved the lowest p-value was reported.

Statistical significance was reported using the false discovery rate (FDR) or q-value to correct for multiple comparisons (77). This method measures the expected proportion of false positives among a set of significant results at a given p-value threshold. In the immune escape maps presented here we used a threshold of $q \leq 0.2$, meaning that we expect 20% of the observed HLA-associated polymorphisms on our maps to represent false positive associations. Statistically significant HLA-associated polymorphisms identified in the historic cohort were then compared to HLA-associated polymorphisms previously identified as part of the IHAC cohort, using modern HIV-1 samples (22).

3.3.7. *Immune escape maps*

HLA-associated polymorphisms identified in *gag* with a $q \leq 0.2$ were organized into an “immune escape map”, consistent with methods used in previous studies (22, 30). Two types of associations are presented on these maps: amino acids significantly enriched in the presence of a specific HLA class I allele (adapted or “escape mutant” forms) and amino acids significantly depleted in the presence of a specific HLA class I allele (nonadapted or “wild type” forms). If an HLA-associated polymorphism fell within three amino acids of a published CTL epitope (as identified by the “best-defined CTL epitopes” on the Los Alamos National Laboratories HIV Immunology database) ((78); <http://www.hiv.lanl.gov/content/immunology>), this epitope and its HLA restriction were also indicated on the map. Mapping within three amino acids of a known CTL epitope accounts for escape mutations potentially involved in antigen processing (79, 80).

3.3.8. Estimating the time of the most recent common ancestor

Using a randomly sampled subset of 77 historic bulk *gag* sequences from New York and Boston, the date of the most recent common ancestor was estimated using the phylogenetic software package BEAST version 1.6.1 (Bayesian Evolutionary Analysis by Sampling Trees) (81). Briefly, all sequences were confirmed to be HIV-1 subtype B using RIP ((82); <http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html>). Sample names were annotated with tip dates, specified as days elapsed since January 1, 1900. The nucleotide substitution model TN93 was chosen using a relaxed molecular clock: uncorrelated Lognormal. Tree prior settings were set to “coalescent: Bayesian Skyline” with the number of groups equal to 10. Priors were subsequently set to maximum and minimum values for all model parameters and statistics. Markov chain Monte Carlo chain lengths were set to 10^7 . The analysis was run overnight on two desktop computers and results of the duplicate runs were combined using Tracer v1.5.dmg.

3.3.9. Generation of HIV-1 gag clones

Clonal *gag* sequences were generated from historic and modern PCR amplicons using the Invitrogen TOPO® TA Cloning® kit system (Life Technologies, Burlington, ON, Canada). Bulk second round PCR *gag* amplicons were directly cloned into a pCR™ 2.1-TOPO® TA vector using Invitrogen One Shot® Stbl3™ Chemically Competent *E. coli* (Life Technologies, Burlington, ON, Canada). Each ligation reaction was incubated at room temperature for 30 minutes and consisted of the following components: 1.5µl of molecular-grade H₂O, 1µl salt solution, 0.5µl TOPO® TA vector and 3µl of second round PCR product. Following ligation, products were transformed into competent *E. coli* cells following the Invitrogen One Shot® Stbl3™ Chemically Competent *E. coli* transformation procedure with a 30 second heat-shock at 42°C. Transformed cells were then

propagated using agar-plates containing ampicillin and X-gal for visualization of re-circularized plasmid. Clones were isolated using an E.Z.N.A.® Plasmid Mini Kit I protocol (Omega Bio-Tek, Norcross, GA, USA) and subsequently digested using EcoR1 (New England BioLabs, Pickering, ON, Canada) to verify *gag* inserts. All clones were sequenced to confirm patient origin and verify that each clone contained an intact *gag* reading frame with start and stop codons.

3.3.10. Generation of chimeric virus and replication assays

Gag-deleted pNL4-3 was developed by inserting BstEII restriction sites at the 5' and 3' ends of the *gag* gene (HXB2 nt 782-788 and 2294-2300, respectively) using the QuikChange® XL site-directed mutagenesis kit (Agilent Technologies, Mississauga, ON, Canada). *Gag* was then removed from the pNL4-3 backbone by BstEII (New England BioLabs, Pickering, ON, Canada) digestion, yielding a Δgag pNL4-3 plasmid backbone. *Gag* chimeric NL4-3 virus generation, titration and replication assays were performed using previously described methods (41, 61, 64). Briefly, cloned *gag* DNA was used as a template for PCR using the following primer set: forward,

GACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAG
TACGCCAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGG (HXB2 nt 695 -
794), and reverse,

TCATTTTTGGTTTCCATCTTCCTGGCAAATTCATTTCTTCTAATACTGTATCATCTGCT
CCTGTATCTAATAGAGCTTCCTTTAATTGCCCCCTATC (HXB2 nt 2391-2294).

These primers were designed to match the pNL4-3 sequence directly upstream and downstream of the 5' and 3' ends of *gag*, respectively, thus generating an amplicon with long overhanging ends to facilitate subsequent in vitro recombination with the Δgag pNL4-3 plasmid backbone. Each 50 μ l reaction was prepared as follows: 35 μ l of

molecular-grade H₂O, 5µl of 10X Expand™ High Fidelity Buffer containing 15mM MgCl₂ (Roche Applied Science, Laval, PQ, Canada), 2µl of 25nM MgCl₂ (supplied by manufacturer), 1µl of 25mM dNTPs (Roche Applied Science, Laval, PQ, Canada), 1µl each of 10µM forward and reverse primers, 2µl Expand™ High Fidelity Enzyme Mix (3.5U/µl) and 3µl of purified plasmid DNA. Thermal cycling consisted of an initial denaturation step of 2 minutes at 94°C, followed by 35 cycles of (30 seconds @ 94°C, 30 seconds @ 55°C, 2 minutes @ 72°C) and finishing with a final 7 minute extension at 72°C.

These long patient-derived *gag* amplicons were then cotransfected with linearized Δ*gag* pNL4-3 vector via electroporation into a tat-driven green fluorescent protein (GFP) reporter T cell line (GXR cells) (83). Cells were incubated for 14 days at 37°C to allow for homologous recombination and generation of chimeric virus. GFP expression was subsequently monitored using a guava® easyCyte™ 8HT Flow Cytometry System (Millipore, Billerica, MA, USA) every 1 to 2 days after day 14 and supernatants were harvested once GFP expression among viable cells reached 15% or higher. Harvested viral supernatants were stored at -80°C until titration and use in replication assays. Viral stocks were titrated by infecting GXR cells with 400µl of chimeric virus stock and measuring GFP expression by flow cytometry two days post infection. Replication assays were subsequently performed in duplicate starting at a multiplicity of infection (MOI) of 0.003 (0.3% infected cells) on day 2 followed by consistent monitoring of GFP expression from day 2 through to day 6, as previously described (41). An MOI of 0.003 was chosen to ensure that clear exponential growth was observed between days 2 and 6 (41). All replication assays were performed in duplicate and featured multiple negative (cells only) and positive (HIV-1 NL4-3 infected) controls. Post data collection,

the natural log of the slope of the percent GFP-expressing cells was calculated between days 3 and 6 provided that day 5 did not exceed 11% infected cells; if day 5 exceeded 11% infected cells the natural log of the slope percent GFP-expressing cells was calculated between days 3 and 5 only. Results were then normalized to NL4-3 by dividing the replication capacity of each virus by the mean slope of NL4-3 replication, such that a replication capacity value of 1.0 indicates a rate of growth equal to that of NL4-3 while replication capacity values <1.0 and >1.0 indicate rates of growth less than and greater than that of NL4-3 respectively.

After the generation of chimeric viruses, the *gag* region from a random subset of N=42 recombinant viruses were re-sequenced to rule out sample mixups and to verify the extent (if any) of in vitro viral evolution during culture.

3.3.11. *Statistical analyses of replication capacity*

Statistical analyses of viral replication capacity data were performed using standard nonparametric tests implemented in GraphPad Prism version 5.0.

3.4. Results

3.4.1. *Identification of HLA-associated gag polymorphisms in the historic cohort*

HLA-associated polymorphisms were identified in a historic cohort (1979-1989) of HIV-1 subtype B infected individuals using serum samples obtained from repositories in New York and Boston. In total, plasma HIV-1 RNA *gag* genotyping was successful for 22 samples from New York (1979-1984) and 217 samples from Boston (1985-1989), yielding a final historic dataset of 239 *gag* sequences. HLA-A, B and C typing were successful for 183 (76.5%), 230 (96.2%) and 231 (96.7%) of these individuals

respectively. In cases where HLA data were missing for an individual, especially in the case of HLA-A where amplification is often poor, the associated *gag* sequences were excluded from analyses of the locus missing. Phylogenetically-informed methods that correct for HIV-1 phylogeny, HLA linkage disequilibrium and HIV-1 codon covariation were subsequently used to identify HLA-associated polymorphisms across the 500 amino acids in *gag* (25).

Using a false discovery rate threshold of $q \leq 0.2$, a total of 19 HLA-associated polymorphisms at 13 unique amino acid positions were identified in *gag*. Results were organized into an immune escape map which indicates the location, HLA- restriction, amino acid(s) and direction (adapted vs. non-adapted) of each HLA-associated polymorphism with respect to the most recent HIV-1 subtype B *gag* consensus sequence ((82); <http://www.hiv.lanl.gov/>), Figure 3.1. Of the 19 associations identified, 16 (84.2%) fell within a published CTL epitope as indicated by the epitope boundary arrows above the sequence. Six HLA-associated polymorphisms (31.6%) were identified as adapted (“escape”) forms and 13 (68.4%) were identified as nonadapted (“wild type”) forms. The overrepresentation of nonadapted associations may reflect a lack of statistical power. Sometimes, a single HLA allele can select for multiple escape variants such that there is enough power to detect the nonadapted (“wild-type”) form but insufficient power to detect each individual adapted (“escape”) form (21). Identified HLA-associated polymorphisms were restricted by 12 different alleles, of which HLA-B alleles accounted for nearly half (47.4%) of the total number of associations, while HLA-A and HLA-C accounted for 26.3% and 26.3% respectively. Note that 14 associations were identified at the HLA allele level (2-digit) while 5 were identified at the protein level (4-digit). Our methods look for HLA-associated polymorphisms at both the allele and protein levels and include the association that achieves the highest statistical

significance, which in some cases is at the allele level and in other cases is at the protein level.



Figure 3.1. Historic gag escape map.

HLA-associated polymorphisms identified across the 500 amino acids in *gag* have been organized into an immune escape map. Using the most recent HIV-1 subtype B consensus sequence as a reference, the locations, residues and HLA restrictions of each association are identified. Shaded vertical bars are used to highlight blocks of 10 amino acid segments to facilitate visualization. Polymorphisms falling directly in or within three amino acids of a published CTL-restricted *gag* epitope are mapped within the designated region as indicated by the above arrows. Adapted and nonadapted forms are mapped beside their restricting allele and distinguished by colour where red represents the adapted or “escape” forms and blue represents the nonadapted or “wild-type” forms. Polymorphisms associated with the same restricting allele and found in close proximity are boxed together in yellow. Uppercase letters indicate direct HLA-associated polymorphisms while lowercase letters indicate indirect associations, before and after correction for codon covariation respectively. This map shows all the statistically significant HLA class I associated polymorphisms in *gag* with a $q \leq 0.2$.

HLA-associated polymorphisms identified in this historic cohort were further compared with published lists of HLA-associated polymorphisms in modern circulating HIV-1 subtype B sequences as part of the IHAC study (22). Of the 19 HLA-associated

polymorphisms identified in the historic cohort, 18 mapped to positions under HLA-mediated selection in modern sequences. Of these 18, 14 represent exact matches to an HLA-associated escape pathway in the modern sequences. The remaining 4 likely also represent exact matches, however, due to inherent difficulties in disentangling HLA alleles in tight linkage disequilibrium they mapped to a linked HLA allele rather than the primary restricting allele; for example, in the historic cohort *gag* position 62 was associated with immune pressure exerted by HLA-B*49 whereas in the modern cohort position 62 was associated with immune pressure exerted by the tightly linked HLA-C*14 allele. Only 1 polymorphism identified in the historic cohort was not observed in the modern cohort: escape mutation T84V restricted by HLA-C*14. This may be an example of an escape mutation that is unique to historic sequences. However, a false discovery rate threshold of $q \leq 0.2$ means that 20% of the HLA-associated polymorphisms identified in our historic cohort are likely false positives. Therefore, it is plausible that this association represents a false positive.

Next we wanted to see whether the frequency of any of these HLA-associated escape mutations have increased in the general population over time. Using *gag* subtype B sequences from our historic cohort and modern *gag* subtype B sequences obtained from the ACTG 5142/5128 studies, the frequency of known escape mutations (22) was compared between the two cohorts (Figure 3.2). A Spearman correlation coefficient of $R=0.86$ ($p<0.0001$) revealed a robust statistical relationship between the frequency of escape mutations observed in the historic cohort compared with the modern cohort, indicating that the frequency of most HLA-associated escape mutations have remained relatively stable over time. However, there are escape mutations that do appear to have a higher frequency in the modern cohort compared to the historic cohort as represented by the data points high on the Y-axis and low on the X-axis (Figure 3.2).

Overall, the median fold-change in the frequency of all investigated escape mutations between modern and historic cohorts was 1.34 [IQR: 0.92-2.33], indicating that escape mutations have generally increased in frequency, at least to a modest extent, over the last 30 years.

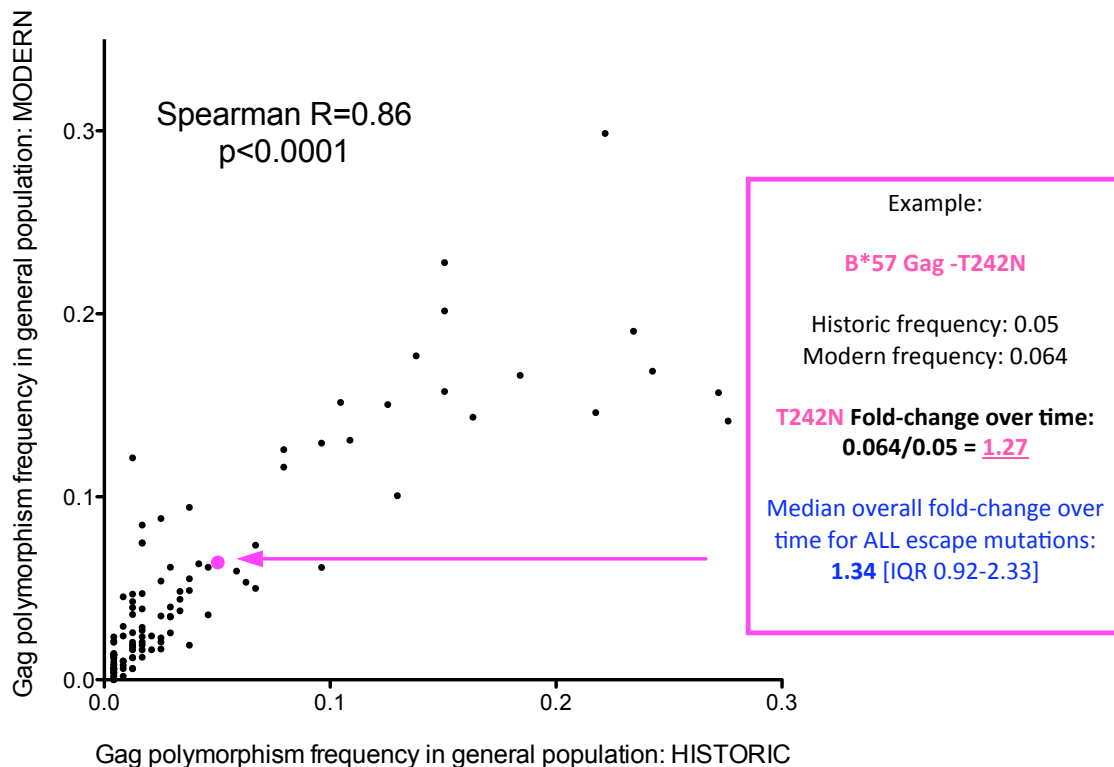


Figure 3.2. Frequency of HLA-associated polymorphisms identified in gag in the general population.

Modern *gag* sequences obtained from the ACTG cohort and historic sequences generated in the present study were scanned for previously identified HLA-associated polymorphisms in *gag*. Each dot represents one of N=109 HLA-associated polymorphisms known to occur in *gag* in modern HIV-1 subtype B sequences. A Spearman correlation analysis revealed a correlation coefficient of 0.86 (p<0.0001), indicating that escape mutations identified in the historic cohort are significantly correlated with escape mutations identified in the modern cohort. For each dot the fold-change in polymorphism frequency between modern and historic sequences can be calculated as shown above. The median overall fold-change of HLA-associated polymorphisms identified in *gag* over time for all N=109 escape mutations is 1.34 [IQR: 0.92-2.33].

Although these data appear to support a modest increase in escape mutations in the general population over time, we next wanted to see whether this could be attributed to differences in HLA class I allele frequencies between our two populations. Indeed, analysis of fold-difference in HLA class I allele frequencies compared with the median fold-change in escape mutation frequencies restricted by each allele revealed a robust statistical correlation ($R=0.75$, $p<0.0001$) (Figure 3.3). Taken together, these results suggest that observed increases in escape mutation frequencies between historic and modern cohorts appear to be largely driven by shifts in the frequencies of certain HLA class I alleles in the North American HIV-infected population (likely as a result of changing demographics of HIV-1 infection over time) rather than the steady accumulation of these mutations as a result of immune selection.

Given that the HLA-C*14 associated T84V escape mutation was unique to the historic cohort, we further evaluated this mutation with respect to its frequency in the general population. If in fact this escape mutation has accumulated over time such that it is no longer detected in modern populations, we would expect to see a dramatic increase in the frequency of the T84V mutation in the general population between our historic and modern cohorts. In the historic cohort the frequency of the T84V mutation was 0.177 (17.7%) compared to a frequency of 0.381 (38.1%) in the modern cohort, resulting in an overall fold-change of 2.16. Additionally, the allele frequency of HLA-C*14 was comparable between historic and modern cohorts with an overall fold-difference of 1.066. These data suggest a two-fold increase in the prevalence of the T84V escape mutation in the general population over time. Indeed, this may represent an escape mutation that has been steadily accumulating over the course of the epidemic in North America.

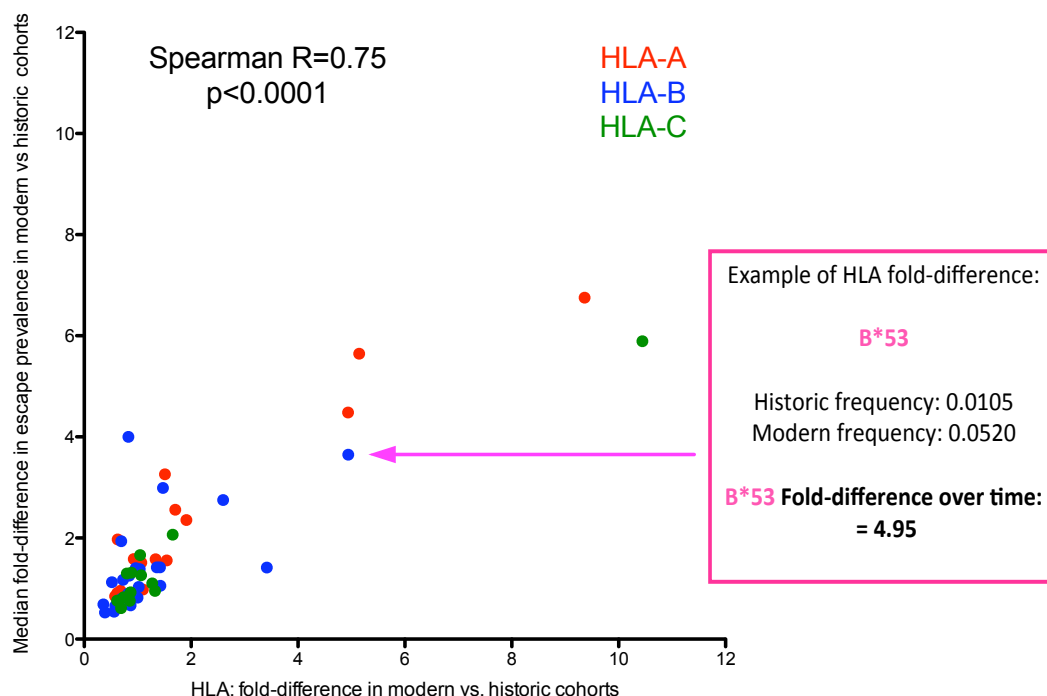


Figure 3.3. Median fold-change in HLA-associated polymorphisms identified in *gag* compared to fold-difference in HLA class I allele frequencies.

Each dot represents one of $N=54$ HLA class I alleles represented in both the historic and modern cohorts with HLA-A represented in red, HLA-B represented in blue and HLA-C represented in green. For each allele, the fold-difference in its frequency between modern and historic cohorts is plotted against the median fold-change in the prevalence of the HLA-associated polymorphisms between modern and historic sequences restricted by the given allele. A Spearman correlation analysis revealed a correlation coefficient of 0.75 ($p<0.0001$), indicating that the median fold-change in HLA-associated polymorphisms between modern and historic cohorts is significantly correlated with fold-difference in HLA class I allele frequencies over time.

Lastly, historic *gag* amino acid sequences ($N=239$) were used to generate a consensus sequence that was subsequently compared to the current published modern (2004) subtype B consensus sequence ((82); <http://www.hiv.lanl.gov/>). Alignment of our historic consensus sequence with the modern consensus sequence revealed only two mismatched amino acids out of 500 amino acids analyzed, a difference of 0.4%. At *gag* position 67, the historic consensus sequence featured alanine (A), whereas the modern consensus sequence features serine (S). Furthermore, at *gag* position 403, the historic consensus sequence featured arginine (R) whereas the modern consensus sequenced

features lysine (K). A 99.6% similarity at the amino acid level between historic and modern HIV-1 consensus *gag* sequences, indicates little change in the “average” HIV-1 sequence over the last 30 years and suggests that CTL-driven immune selection pressures have not led to substantial increases in the frequency of immune escape mutations in circulating HIV-1 sequences over time.

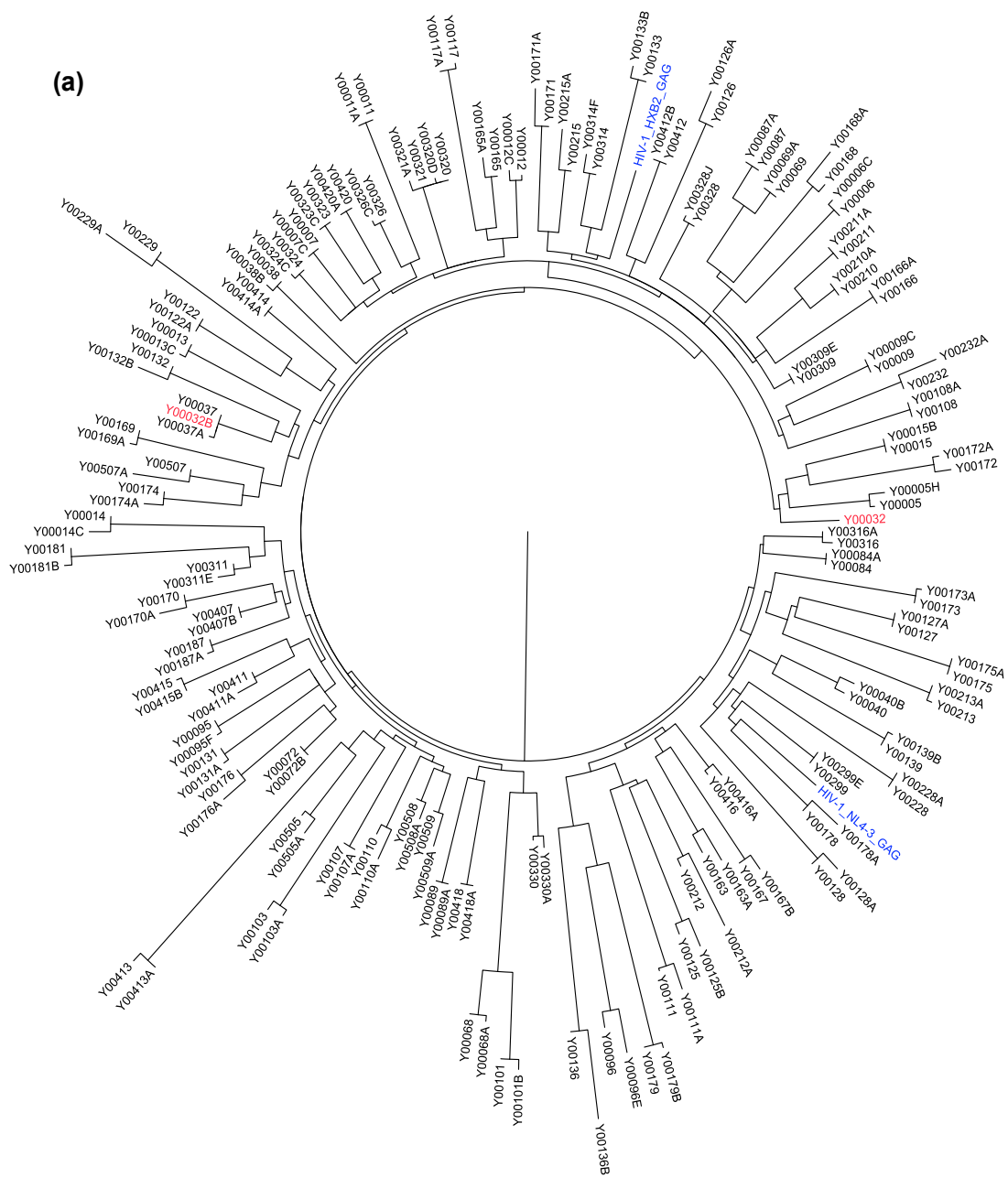
3.4.2. *Timing of the most recent common ancestor*

The phylogenetic software package BEAST was used to estimate the timing of the most recent common ancestor in North America. BEAST is a program that creates rooted, time-stamped phylogenetic trees via strict or relaxed molecular clock methods using a Bayesian Markov chain Monte Carlo platform (81). As described in the methods section, we estimated the date of the most recent common ancestor using a randomly sampled subset of 77 historic bulk *gag* sequences annotated with their sampling dates (1979-1989) using a TN93 nucleotide substitution model and a relaxed molecular clock. Two runs of chain length 10^7 were performed and outputs combined. Results revealed a mean root height of 10,531.8 days (with upper and lower 95% highest posterior density (HPD) bounds of 17,219.1-6,141.7 days, respectively), allowing us to estimate the date of the most recent common ancestor as 1960. These estimates using historic *gag* sequences are consistent with previously published estimates dating the most recent common ancestor of North American subtype B sequences (84).

3.4.3. *Replication capacity in gag between historic and modern cohorts*

Replication capacity of historic and modern *gag* chimeric NL4-3 virus was assessed by inserting clonal patient-derived HIV-1 RNA *gag* sequences into an NL4-3 backbone. Previously, studies from our laboratory have used gene-specific bulk PCR

sequences to generate chimeric viruses in an attempt to capture quasispecies diversity (41, 61, 64). However, a potential limitation in this approach is the genetic bottleneck that occurs during the in vitro homologous recombination step such that the resulting viral stocks may not represent the initial diversity of HIV-1 sequences obtained from the patient. More importantly, the ability to capture quasispecies diversity through PCR may differ based on the sample quality (41): RNA recovered from historic serum samples that have been stored for up to three decades may not fully capture HIV-1 diversity within the original host as a result of RNA decay. To reduce biases associated with capturing quasispecies diversity in historic samples, we cloned a single *gag* sequence for each patient. *Gag* cloning was successful for 87 historic (21 from New York, 62 from Boston and 4 from San Francisco) and 73 modern (17 from New York and 56 from Boston) patient samples. The origin of clonal sequences were subsequently verified against their original bulk sequences using phylogenetic analysis, revealing one historic clone (Y00032B) and two modern clones (E00466E, E00643B) that did not cluster as expected (Figure 3.4). As these were likely the results of labeling errors or accidental contamination, they were eliminated from further analyses. Additionally, *gag* sequences from HIV-1 NL4-3 and HXB2 were included in both phylogenetic trees to rule out contamination with these two commonly used laboratory strains.



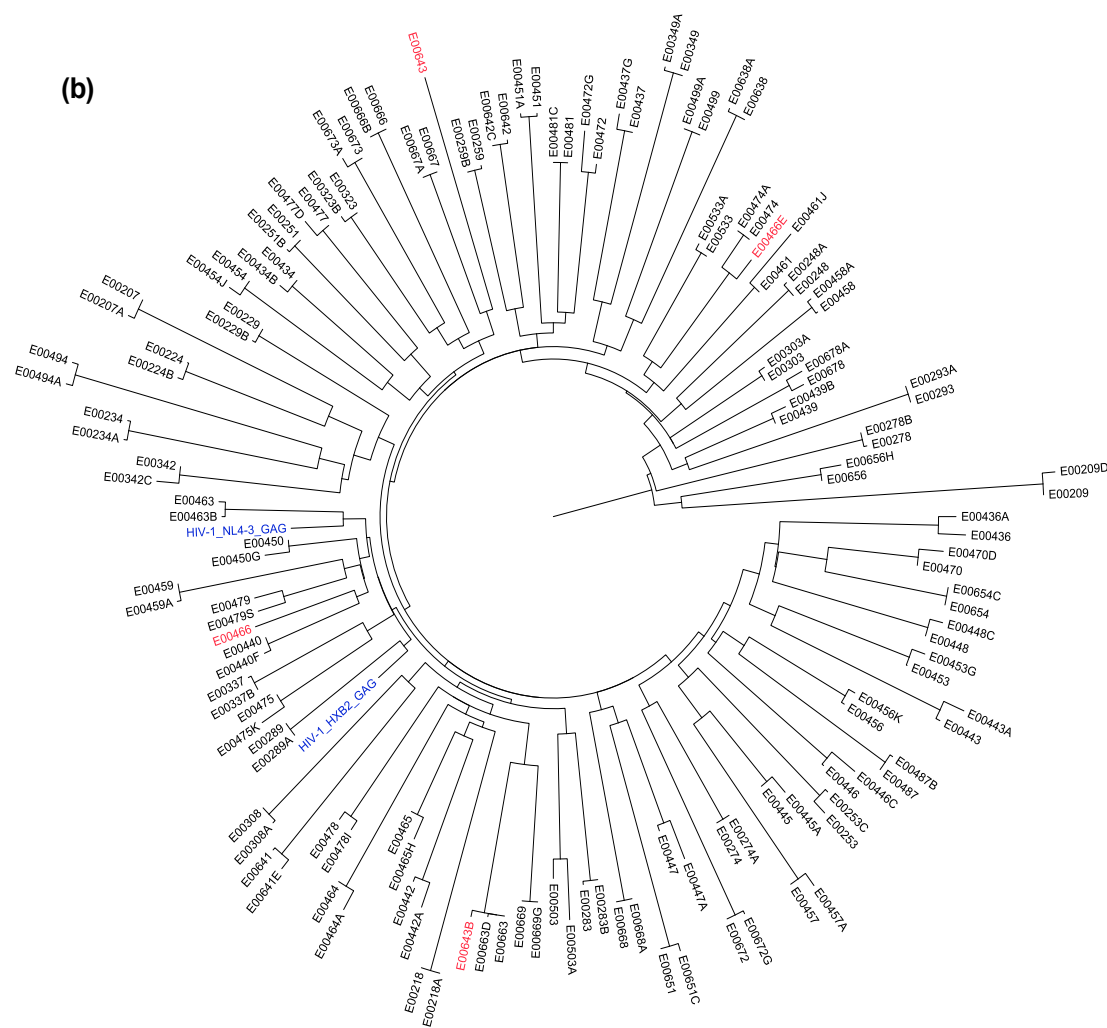


Figure 3.4. Confirmation of gag clone origin by phylogenetic analysis.

Maximum-likelihood phylogenetic trees were generated (online tool: phyML) using historic (a) and modern (b) *gag* clone sequences as well as the original patient bulk *gag* sequences. Clonal sequences are represented by a letter (indicating the specific clone) directly after the alphanumeric identification code while bulk sequences do not contain letters at the end of their identification code. Laboratory HIV-1 strains HXB2 and NL4-3 *gag* sequences are highlighted in blue and clones that failed to map to the same branch as their original patient sequence are highlighted in red. Note that the one clone highlighted in red in the historic tree (a) and the two clones highlighted in red in the modern tree (b) that did not map with their original patient sequence were excluded from further analyses.

Cotransfection of a total of 157 *gag* clones with linearized Δ *gag* pNL4-3 into GXR cells resulted in the successful production of 80 historic (20 from New York, 56 from Boston and 4 from San Francisco) and 58 modern (15 from New York and 43 from

Boston) recombinant viruses. Median time from cotransfection to harvest of recombinant virus stock at $\geq 15\%$ GFP-expressing cells was 16 days [range: 13-25 days]. Of note, 6 historic and 13 modern *gag* clones (total 19; 12.1%), failed to generate recombinant virions despite two independent attempts. Phylogenetic analysis of failed and successful clonal sequences revealed no significant clustering of those that failed (not shown), allowing us to rule out specific sequence features associated with a single viral lineage as the cause of these failures. Additionally, amino acid frequency analyses of failed vs. successful clonal sequences using the Viral Epidemiology Signature Pattern Analysis (VESPA) tool ((82); <http://www.hiv.lanl.gov/content/sequence/VESPA/vespa.html>), revealed no significant enrichment or depletion of specific amino acids at any of the 500 codons in *gag* (not shown). Taken together, these results suggest that failure to generate recombinant viruses was not due to shared sequence features associated with descent from a single viral lineage, nor was it due to particular polymorphism(s) in *gag* present in these samples.

After the generation of chimeric viruses, a subset of 43 viral stocks (22 historic and 21 modern) were re-extracted, amplified for *gag* and sequenced. These sequences were then compared with the original *gag* clone sequences to confirm their patient origin and assess the extent (if any) of in vitro evolution. All 43 re-sequenced clones mapped to their original clonal sequence (Figure 3.5). Out of the 43 chimeric virus sequences analyzed, 24 (55.8%) had at least one nucleotide difference compared to the original clone sequence resulting in a total nucleotide difference of 0.060%. The majority of these changes (74.4%) were the result of mixed bases in the chimeric *gag* sequence, whereas the remainder were complete base changes (25.6%). Interestingly, A to G mutations accounted for 59.0% of the observed nucleotide differences between clonal sequences and re-sequenced harvested recombinant viral stocks. Of the observed

nucleotide changes, 48.7% resulted in an amino acid change (nonsynonymous mutation) such that the total observed difference at the protein level was 0.088%. No significant difference in the number of nucleotide or amino acid changes amongst clone and recombinant virus was observed between historic and modern sequences (not shown). These data indicate that a small amount of evolution did indeed occur in vitro, however, these observations are unlikely to affect the overall analysis.

In order to assess viral replication capacity in vitro, *gag* recombinant viral stocks were used to infect a GFP-reporter CD4+ T cell line. The percentage of GFP-expressing cells was quantified daily between days 3 to 6 post infection by flow cytometry, and the results were reported as the natural log of the slope of this exponential curve, normalized to that of the reference strain NL4-3. Replication capacity was assessed in duplicate for each clone and results were averaged. Overall, results were highly reproducible: the median percent difference between replicates in the historic cohort was 2.8% [IQR: 1.2%-6.1%] and the modern cohort was 2.3% [IQR: 1.2%-4.5%], with an overall difference of 2.6% [IQR: 1.2%-5.3%].

Stratification of viral replication capacities by historic (1979-1989) vs. modern (2002-2008) eras revealed no significant difference between the two groups (Mann-Whitney $p=0.4187$), Figure 3.6a. The median replication capacity for historic *gag* viruses was 0.9674 [IQR: 0.8389-1.033] and modern *gag* viruses was 0.9649 [IQR: 0.8297-1.095]. Both historic and modern *gag* recombinant viruses had a median replication capacity below 1.0, indicating that the “average” *gag* recombinants grew slower than NL4-3. When results were stratified by era (historic/modern) as well as geographic region (NY/Boston/San Francisco), a trend towards lower replication capacity was observed in the New York and San Francisco historic samples (Figure 3.6b); however, this did not achieve statistical significance (Kruskal-Wallis $p=0.1148$). Of interest, stratifying results by sample year (not shown) revealed no significant difference in replication capacity within the modern cohort (Kruskal-Wallis $p=0.1986$), but revealed a modest trend towards increasing fitness by year in the historic cohort, despite not reaching statistical significance (Kruskal-Wallis $p=0.1556$).

Given that a modest increase in replication capacity was observed in the historic cohort with year, replication capacity between historic and modern sequences were subsequently stratified into three groups according to year: 1979 to 1984, 1985 to 1989 and 2002-2008. When results were stratified in this manner a trend towards increasing viral replication between pre-1985 and post-1985 *gag* sequences was observed (Kruskal-Wallis $p=0.0465$, Figure 3.6c). These findings support the notion that *gag*-mediated viral replication capacity has modestly increased over time, where sequences generated from samples pre-1985 appear to replicate slightly slower than sequences generated from samples post-1985.

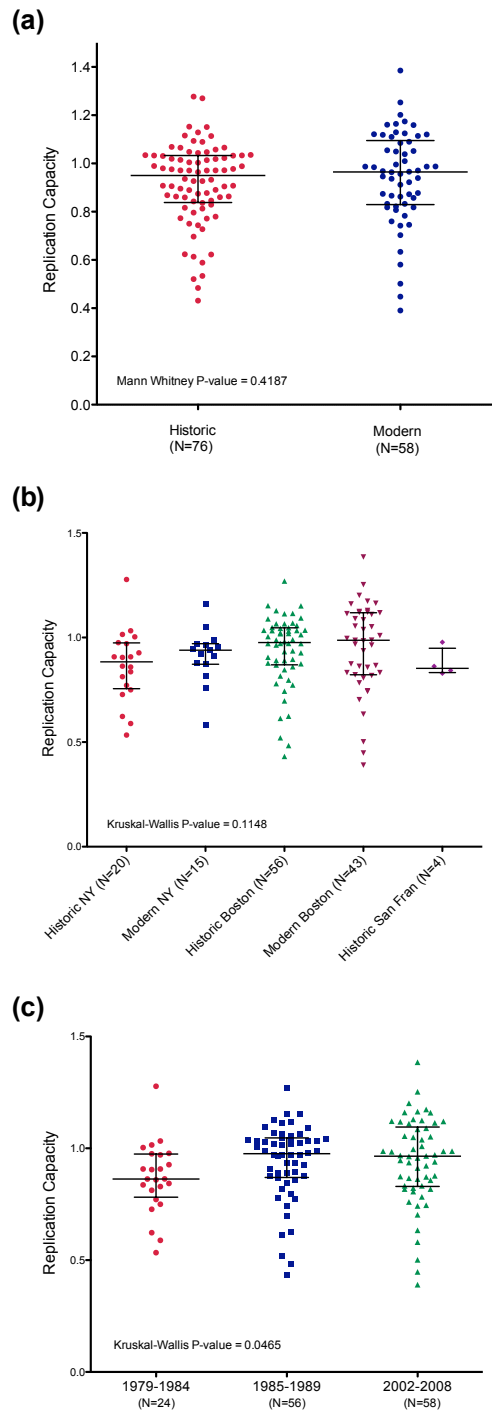


Figure 3.6. Comparison of HIV-1 replication capacity over time.

Chimeric *gag* viral replication capacity was compared overall between historic and modern (a) cohorts as well as stratified by region (b) and year (c). Median and interquartile ranges for each group are indicated along with the p-value for the given test of significance.

Next, we wished to examine whether *gag*-mediated viral replication capacity differed by specific HLA class I alleles expressed by the host, in historic and modern cohorts (Figure 3.7). In this analysis, only HLA-A, B and C alleles observed a minimum of three times in both historic and modern cohorts were included. In both historic and modern cohorts, stratification by host HLA-A, B and C alleles expressed revealed a range of viral replication capacities for a given locus, however, in all cases these differences were not statistically significant (Kruskal-Wallis $p < 0.05$). In a post-hoc analysis, we performed pairwise comparisons of viral replication capacity among persons expressing the same HLA class I allele, between modern and historic cohorts. Among all pair-wise comparisons, HLA-A*24 was the only HLA class I allele whose expression was associated with significantly different replication capacity between historic and modern *gag* recombinant viruses (Mann-Whitney $p = 0.0409$). The median viral replication capacity of viruses derived from HLA-A*24 expressing individuals in the historic and modern cohorts were 1.015 [IQR: 0.8777-1.034] and 1.085 [IQR: 0.9956-1.161], respectively. Interestingly, *gag* recombinant viruses from patients expressing HLA-B*27 and HLA-B*57, alleles commonly associated with slower HIV-1 disease progression, had lower replication capacities compared with other HLA-B alleles in the historic cohort. In the modern cohort HLA-B*27 was again associated with lower replication capacity compared with other HLA-B alleles, however, these observations were not statistically significant.

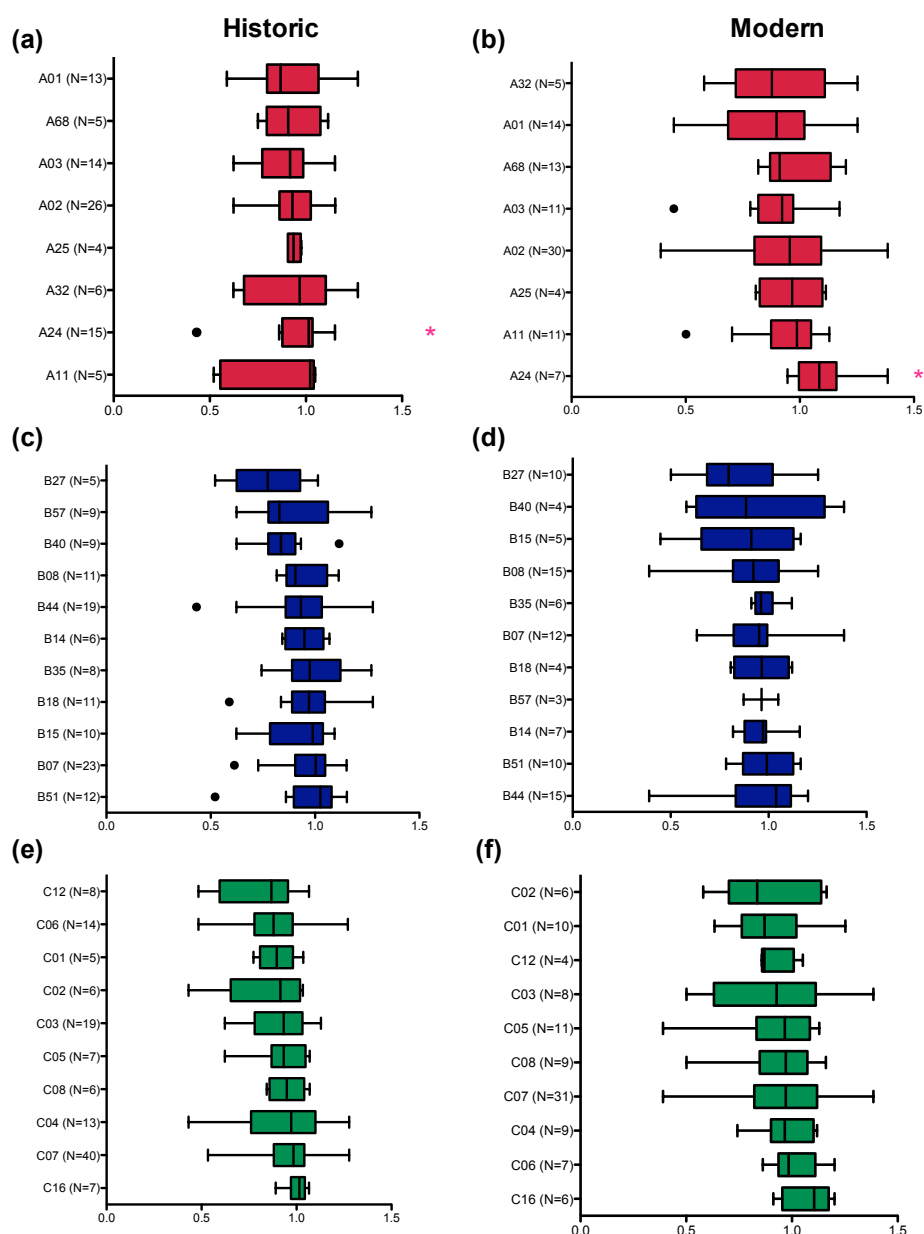


Figure 3.7. HLA-associated changes in gag-mediated HIV-1 replication capacity over time.

Replication capacity of chimeric viruses expressing patient-derived HIV-1 *gag* sequences from historic (panels a, c and d) and modern (panels b, d and f) cohorts were stratified according to HLA-A (panels a and b), HLA-B (panels c and d) and HLA-C (panels e and f) allele expression in the original patient. Results are represented as boxplots where the middle band and outer boundaries represent the median and interquartile range [IQR] respectively. Whiskers display maximum and minimum values and outliers are represented by dots outside the given range. A minimum of N=3 was required for inclusion in this analysis. HLA class I alleles associated with significantly different (Mann-Whitney $p < 0.05$) replication capacity between historic and modern cohorts are indicated with an asterisk.

Lastly, we wished to identify specific amino acids in historic and modern *gag* sequences associated with differences in replication capacity. At $p < 0.05$, codon-by-codon analysis of historic *gag* sequences using a Mann-Whitney test revealed 21 amino acids at 15 unique positions associated with differential replication capacities (Table 3.1), whereas analysis of combined historic and modern sequence datasets revealed a total of 25 amino acids at 18 unique positions associated with differential replication capacities (Mann-Whitney $p < 0.05$). However, it is important to note that no sites remained statistically significant after correction for multiple comparisons ($q \leq 0.2$). Nevertheless, several observations are notable. Of the 15 unique positions identified in the historic cohort, 12 fell directly in or within three amino acids of a previously defined HLA class I restricted CTL epitope ((78); <http://www.hiv.lanl.gov/content/immunology>). Interestingly, two of the polymorphisms associated with changes in replication capacity (28R and 30K) map to HLA class I associated polymorphisms identified in our historic cohort (Figure 3.1). Specifically, sequences harbouring 28R display reduced replication capacity compared to those harbouring K at this position. Polymorphism 28R is identified as an escape variant in individuals expressing HLA-A*0301, suggesting that escape to an R at this position confers a fitness cost. Furthermore, sequences harbouring 30K also display reduced replication capacity compared to historic sequences harbouring an R at this position. Interestingly, polymorphism 30K is identified as wild-type and 30R is an escape variant in individuals expressing HLA-A*24. These results suggest that in the historic cohort escape away from 30K increased viral fitness. Interestingly, HLA-A*24 was the only HLA class I allele whose expression was associated with increased viral replication in individuals in the modern cohort compared to individuals in the historic cohort. This may be a position of interest for future studies.

Table 3.1. Amino acids in gag associated with either an increase or decrease in in vitro viral replication capacity.

Amino Acid Position	Amino Acid	Median Replication Capacity		P-value	Q-value
		With Amino Acid (N) ^b	Without Amino Acid (N) ^b		
28 ^a	R	0.7947 (N=10)	0.9711 (N=70)	0.04024	0.49310
30 ^a	K	0.9077 (N=45)	1.0148 (N=35)	0.03247	0.49310
67	A	0.9077 (N=49)	1.0047 (N=31)	0.03949	0.49310
67	S	1.0047 (N=31)	0.9077 (N=49)	0.03949	0.49310
91 ^a	Q	0.7667 (N=6)	0.9711 (N=74)	0.007655	0.42266
219 ^a	H	0.9296 (N=76)	1.05055 (N=4)	0.01261	0.42266
219 ^a	Q	1.05055 (N=4)	0.9296 (N=76)	0.01261	0.42266
370 ^a	V	0.9268 (N=67)	1.0314 (N=12)	0.04183	0.49310
372 ^a	N	0.9296 (N=76)	1.0861 (N=4)	0.00919	0.42266
374 ^a	N	0.7732 (N=4)	0.9709 (N=76)	0.01817	0.42266
389	I	0.8777 (N=33)	0.9756 (N=47)	0.005593	0.42266
389	T	0.9737 (N=40)	0.8837 (N=40)	0.0117	0.42266
425 ^a	D	0.9268 (N=73)	1.0352 (N=7)	0.04019	0.49310
425 ^a	E	1.0352 (N=7)	0.9268 (N=73)	0.04019	0.49310
441 ^a	S	1.09275 (N=4)	0.9342 (N=76)	0.04455	0.49310
475	K	0.9709 (N=76)	0.8210 (N=4)	0.04455	0.49310
478 ^a	L	0.7501 (N=7)	0.971 (N=73)	0.008522	0.42266
486 ^a	L	0.9712 (N=71)	0.8162 (N=9)	0.008247	0.42266
486 ^a	S	0.8162 (N=9)	0.9712 (N=71)	0.008247	0.42266
490 ^a	K	0.9997 (N=20)	0.9173 (N=60)	0.03672	0.49310
490 ^a	R	0.9173 (N=60)	0.9997 (N=20)	0.03672	0.49310

^amaps within a previously defined CTL epitope, ^bminimum N=4 required for inclusion

Differences in replication capacity at *gag* codon 67 are also notable. In the historic cohort 67A was associated with decreased viral replication capacity whereas 67S was associated with higher viral replication capacity. Intriguingly, codon 67 is one of two sites where the historic consensus sequence (67A) differs from the modern

consensus sequence (67S). Initially, it was tempting to hypothesize that 67S-harboring sequences could have outcompeted their 67A-expressing counterparts over time, due to their enhanced replication capacity. However, analysis of modern sequences revealed the opposite relationship between *gag* codon 67 and replication capacity: in the modern cohort, 67S was associated with decreased viral replication capacity whereas 67A was associated with higher viral replication capacity. The observation that in each case the cohort consensus residue at codon 67 represents the lower fitness variant appears paradoxical. This may imply dynamic changes at this specific site such that modern sequences are continuing to change in a direction that optimizes viral fitness. However, this is just one potential explanation, further studies exploring this site are needed to explain these observations.

3.5. Discussion

The enormous genetic diversity of HIV-1, driven in large part by CTL-driven immune escape within infected individuals, remains a primary challenge in vaccine development (32). Overcoming these barriers will require a deeper understanding of the HIV-host interaction as well as the relationship between individual-level and population-level HIV-1 evolution. A major concern in vaccine research is the accumulation and persistence of immune escape mutations whereby at a population-level HIV-1 may be adapting to pressures induced by the most commonly expressed HLA alleles with a given population (73, 75). Referred to as “HLA imprinting”, this hypothesis purports that escape mutations are accumulating to the point of fixation in circulating HIV-1 sequences over time (85). If this were the case, CTL responses that were once associated with potent anti-HIV-1 activity would no longer be effective in controlling infection as the epidemic progresses (75). The consequences of HLA-imprinting are

additionally concerning with respect to the development of an effective HIV-1 vaccine, as the inability to stimulate a protective immune response against the most prevalent HIV-1 epitopes would render these strategies ineffective (73). Although examples of persisting escape mutations after transmission have been shown on an individual-level (20, 70, 72), the extent to which population-level immune escape patterns have changed over time, and the consequences of these pressures on viral fitness since HIV-1's first introduction into North America remain relatively uncharacterized.

Given the degree of high sequence conservation (59) and the importance of CTL responses targeted at HIV-1 structural proteins (51, 52), we chose to focus this study on the *gag* gene. We therefore examined how HLA-associated polymorphisms in *gag* have changed over time during the North American epidemic, and assessed whether there was any evidence supporting changes in *gag*-mediated viral replication capacity during this same period. To our knowledge this is the first study to use a unique set of historic HIV-1 sequences dating between 1979 and 1989 from three regions across North America (New York, Boston and San Francisco) to address these specific objectives.

Using historic sequences, our results indicate that the most recent common ancestor of the HIV-1 subtype B epidemic in North America existed in 1960. Moreover, the consensus sequence of our historic cohort, which likely closely resembles the sequence of the founder virus, is highly similar to the modern subtype B consensus sequence. Of 500 amino acids compared, only two codons were observed to differ between the historic and modern consensus sequences, a 99.6% sequence similarity. A fundamental argument outlined in the HLA "imprinting hypothesis" is the fixation of escape mutations within a given population such that the circulating consensus sequence would evolve over time to reflect selective pressures induced by the most

commonly expressed HLA alleles (74). If HLA imprinting were occurring to a large extent, then the circulating consensus sequence at the beginning of the epidemic would look different than the consensus sequence we see in modern populations today as a result of the continual accumulation of mutations reflected by CTL-driven immune selection pressures. The high degree of similarity between historic and modern consensus sequences observed here does not support major accumulation of escape mutations at the population-level. Furthermore, it is interesting to note that neither of the two amino acid changes between consensus sequences fell directly within a defined CTL epitope.

Moreover, we found that immune escape pathways in HIV-1 subtype B *gag* sequences have remained largely consistent over the last 30 years of the epidemic in North America. Using phylogenetically informed methods we observed a total of 19 HLA-associated polymorphisms in the historic cohort of which 18 were consistent with patterns of escape previously characterized in modern North American populations. Again these results support the continued dynamic adaptation of HIV-1 to its host as it is passed throughout a given population rather than the substantial accumulation of mutations driven by host CTL responses.

Further supporting the relative stability of HIV-1 immune escape pathways is our observation that the frequencies of escape mutations in the general population (defined as individuals lacking the restricting HLA class I allele) have remained relatively consistent over time. If in fact some escape mutations were accumulating at the population-level due to lack of reversion upon transmission (75), we would expect to see an increased fold-change in the frequency of these mutations in modern populations compared with historic populations. Overall, we did in fact find that HLA-associated

escape mutations in *gag* increased in frequency with a median fold-change of 1.34 (IQR: 0.92-2.33) between modern and historic eras, suggesting that escape mutations have increased on average. However, a major confounder in these analyses is differences in the frequency of certain HLA class I alleles between our historic and modern cohorts, arising as a result of the changes in the demographics of HIV-1 infected populations on this continent as the epidemic has progressed. When analyzed in context of HLA class I allele frequencies we observed in a majority of the cases that the escape mutations that have increased in frequency are those whose restricting alleles have also increased in frequency to the same extent. These data indicate that in fact fold-changes in escape mutation frequencies over time may largely be driven by shifts in HLA class I allele distributions between the cohorts as opposed to the accumulation of escape mutations under steady immune selection pressures.

The HLA-B*27-associated R264K mutation serves as an excellent example to put our observations into perspective. Previous studies have documented the persistence of the HLA-B*27-associated R264K escape mutation upon transmission to persons lacking HLA-B*27 (9, 70, 72), suggesting that this mutation may be increasing in prevalence throughout the general population as the epidemic progresses. Here we found that the allele frequency of HLA-B*27 was comparable in both historic and modern cohorts (~2.5%). Our data are consistent with a modest potential increase in prevalence of R264K over time, but not to the extent where it is rapidly accumulating to fixation. Firstly, *gag* codon 264 was identified as a site of strong HLA-B*27-mediated selection in both historic and modern cohorts, supporting substantial pressure to revert upon transmission consistently over the last 30 years. Furthermore, analysis of the prevalence of R264K in circulating HIV-1 sequences in the general population revealed an approximately 3-fold increase in its prevalence over time, from 0.4 to 1.3% between

historic vs. modern cohorts. Therefore, although the prevalence of R264K appears to have increased to a modest extent over time, the fact that its' prevalence still remains below 1.5% in the general population since the estimated introduction of HIV-1 subtype B to North America (year 1960) argues against major and/or rapid losses of natural or vaccine-induced HLA-B*27-restricted immune responses to HIV-1 during this time.

The second major objective of this paper was to measure viral replication capacity in vitro using cloned *gag* sequences from historic and modern HIV-1 subtype B infected patients to assess whether viral replication capacity has changed over time. Previous studies have highlighted the consequences of escape mutations within highly conserved regions of *gag* with respect to viral fitness (41, 60-64) but have yet to explore the long-term impacts these mutations have on viral replication over time. After generating *gag* clone recombinant viruses using an NL4-3 backbone, viral replication was not found to be significantly different between our historic and modern cohorts in either grouped analysis (Mann-Whitney $p=0.4187$) or pairwise comparisons after stratifying by geographic region and era of the epidemic (Kruskal-Wallis $p=0.1148$). However, after stratifying the historic cohort into two groups (1979-1984 and 1985-1989) and comparing these with the modern cohort (2002-2008) we observed modest yet significantly lower viral replication capacities in pre-1985 *gag* sequences compared to post-1985 *gag* sequences (Kruskal-Wallis $p=0.0465$). These results suggest that the *gag* sequences obtained from individuals infected with HIV-1 subtype B at the beginning of the epidemic in North America have slightly reduced fitness compared with later *gag* sequences. The lack of evidence supporting the accumulation of escape mutations in the present study suggests that the observed changes in viral replication capacity may be due to factors other than immune selection pressures.

To further explore the relationship between HLA immune pressures and viral replication capacity we stratified our historic and modern cohorts by HLA class I alleles expressed by the host. Comparisons of replication capacities revealed no significant differences between alleles of a particular locus (HLA-A, B and C; Kruskal-Wallis $p < 0.05$), although lower replication capacities of viruses from HLA-B*27 and HLA-B*57-expressing persons in the historic cohort, and HLA-B*27-expressing persons in the modern cohort were notable. Post-hoc, pair-wise comparisons between historic and modern HLA class I alleles revealed reduced replication capacity of viruses from HLA-A*24-expressing individuals in the historic cohort compared to those in the modern cohort (Mann-Whitney $p = 0.0409$). Interestingly, we found that HLA-A*24 selects for the K30R escape mutation. Of further interest, codon-by-codon analysis additionally revealed a substantial increase in replication capacity in modern sequences harbouring 30R, although this did not remain statistically significant after correction for multiple comparisons ($q < 0.2$). In Japanese populations, there is evidence that the K30R escape mutation is accumulating throughout the population as a result of the high prevalence of HLA-A*24 alleles (86), however, studies have failed to show a difference in replication capacity between wild-type 30K and escape 30R variants (80). Given the scope of this paper we were not able to further investigate this association, however, the subsequent appearance of the K30R escape mutation in modern populations suggests that these differences in viral replication may be driven by factors other than CTL-driven immune selection pressures.

Additionally, codon-by-codon analysis revealed an interesting association between variation at *gag* codon 67 and viral replication capacity. Historic *gag* sequences expressing 67A had reduced replication capacities whereas in modern *gag* sequences a seemingly opposite observation was made where 67S was associated with lower

replication capacities. Given that an A to S change at position 67 is observed between historic and modern consensus sequences, it may be that this site represents a region of variability that is dynamically changing. Note, however, that this site does not lie within a known CTL epitope and does not appear to be associated with selection pressure by a specific HLA class I allele. Therefore, selection pressures other than those driven by HLA-restricted CTL responses may be responsible for these observed shifts in viral fitness at this position. It was beyond the scope of this paper to further investigate these differences, but it may be that this position is the site of other unexplored evolutionary selective pressures or mechanisms such as genetic drift.

Although *gag* was the focus of the current study, it is plausible that results could differ if other less conserved regions of the HIV-1 proteome were investigated. For instance HIV-1 *nef* represents a highly immunogenic yet also highly variable protein (87-89). In order to fully understand the extent to which HLA-associated selection pressures are driving HIV-1 evolution it will be necessary to evaluate the structure and function of other HIV-1 genes. Additionally, our results are limited by the quantity of historic samples. Expanding this dataset to include more historic HIV-1 subtype B sequences from other regions across North America could further help unravel factors influencing population-level HIV-1 adaptation. Larger fitness datasets may also allow us to develop a more detailed picture of the specific amino acids and positions that are in fact associated with changes in replication capacity. These analyses may additionally help illuminate other factors involved in HIV-1 evolution aside from CTL-driven selection pressures. Finally, it must be noted that a limitation of this study is that our results only apply to the North American epidemic where genetic diversity, and thus HLA diversity, is relatively high among HIV-1 infected populations. However, in populations exhibiting lower HLA diversity, such as Japan, HIV-1 may be evolving in different ways (20, 86).

3.6. Conclusion

In conclusion, our observations that patterns of immune escape have remained largely consistent over time, and that the prevalence of known escape mutations in the general population have only modestly increased over the past 30 years argues against rapid and substantial “HLA imprinting” in circulating HIV-1 sequences in North America. Furthermore, only modest changes in viral replication were observed between pre-1985 and post-1985 *gag* sequences, however, these changes may be due to factors other than CTL-driven immune selection pressures. Indeed variation at *gag* codon 67, one of two residues differing between historic and modern subtype B consensus sequences, was associated with increased viral replication capacity in historic sequences despite a lack of evidence suggesting the involvement of this position in immune escape. It may be that these changes are the result of other host selection pressures or natural drift. Regardless, these results have profound implications for public health prevention and vaccine design. Consistent patterns of immune escape suggest that HIV-1 appears to be escaping and reverting in a cyclical process as it passes through human hosts such that CTL immune responses targeted at epitopes within *gag* continue to provide protection against the most commonly circulating strains of HIV-1. Our results suggest that vaccines strategies aimed at eliciting an immune response towards epitopes within *gag* should continue to remain an important goal.

3.7. References

1. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *Journal of virology*. 1994 Sep;68(9):6103-10.

2. Koup RA, Safrit JT, Cao Y, et al. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *Journal of virology*. 1994 Jul;**68**(7):4650-5.
3. Musey L, Hughes J, Schacker T, Shea T, Corey L, McElrath MJ. Cytotoxic-T-cell responses, viral load, and disease progression in early human immunodeficiency virus type 1 infection. *The New England journal of medicine*. 1997 Oct 30;**337**(18):1267-74.
4. Mellors JW, Rinaldo CR, Jr., Gupta P, White RM, Todd JA, Kingsley LA. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science*. 1996 May 24;**272**(5265):1167-70.
5. Ogg GS, Jin X, Bonhoeffer S, et al. Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. *Science*. 1998 Mar 27;**279**(5359):2103-6.
6. Phillips RE, Rowland-Jones S, Nixon DF, et al. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature*. 1991 Dec 12;**354**(6353):453-9.
7. Borrow P, Lewicki H, Wei X, et al. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nature medicine*. 1997 Feb;**3**(2):205-11.
8. Couillin I, Culmann-Penciolelli B, Gomard E, et al. Impaired cytotoxic T lymphocyte recognition due to genetic variations in the main immunogenic region of the human immunodeficiency virus 1 NEF protein. *The Journal of experimental medicine*. 1994 Sep 1;**180**(3):1129-34.
9. Goulder PJ, Phillips RE, Colbert RA, et al. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nature medicine*. 1997 Feb;**3**(2):212-7.
10. Koenig S, Conley AJ, Brewah YA, et al. Transfer of HIV-1-specific cytotoxic T lymphocytes to an AIDS patient leads to selection for mutant HIV variants and subsequent disease progression. *Nature medicine*. 1995 Apr;**1**(4):330-6.
11. McMichael AJ, Phillips RE. Escape of human immunodeficiency virus from immune control. *Annual review of immunology*. 1997;**15**:271-96.
12. Price DA, Goulder PJ, Klenerman P, et al. Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proceedings of the National Academy of Sciences of the United States of America*. 1997 Mar 4;**94**(5):1890-5.
13. Rowland-Jones SL, Phillips RE, Nixon DF, et al. Human immunodeficiency virus variants that escape cytotoxic T-cell recognition. *AIDS research and human retroviruses*. 1992 Aug;**8**(8):1353-4.
14. Kelleher AD, Long C, Holmes EC, et al. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. *The Journal of experimental medicine*. 2001 Feb 5;**193**(3):375-86.

15. Goulder PJ, Watkins DI. HIV and SIV CTL escape: implications for vaccine design. *Nature reviews. Immunology*. 2004 Aug;**4**(8):630-40.
16. Wolinsky SM, Korber BT, Neumann AU, et al. Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science*. 1996 Apr 26;**272**(5261):537-42.
17. Allen TM, Altfeld M, Geer SC, et al. Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. *Journal of virology*. 2005 Nov;**79**(21):13239-49.
18. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science*. 2002 May 24;**296**(5572):1439-43.
19. Bhattacharya T, Daniels M, Heckerman D, et al. Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science*. 2007 Mar 16;**315**(5818):1583-6.
20. Kawashima Y, Pfafferoth K, Frater J, et al. Adaptation of HIV-1 to human leukocyte antigen class I. *Nature*. 2009 Apr 2;**458**(7238):641-5.
21. Brumme ZL, Brumme CJ, Heckerman D, et al. Evidence of differential HLA class I-mediated viral evolution in functional and accessory/regulatory genes of HIV-1. *PLoS pathogens*. 2007 Jul;**3**(7):e94.
22. Brumme ZL, John M, Carlson JM, et al. HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. *PloS one*. 2009;**4**(8):e6687.
23. Avila-Rios S, Ormsby CE, Carlson JM, et al. Unique features of HLA-mediated HIV evolution in a Mexican cohort: a comparative study. *Retrovirology*. 2009;**6**:72.
24. Boutwell CL, Essex M. Identification of HLA class I-associated amino acid polymorphisms in the HIV-1C proteome. *AIDS research and human retroviruses*. 2007 Jan;**23**(1):165-74.
25. Carlson JM, Brumme ZL, Rousseau CM, et al. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. *PLoS computational biology*. 2008 Nov;**4**(11):e1000225.
26. John M, Heckerman D, James I, et al. Adaptive interactions between HLA and HIV-1: highly divergent selection imposed by HLA class I molecules with common supertype motifs. *J Immunol*. 2010 Apr 15;**184**(8):4368-77.
27. Matthews PC, Prendergast A, Leslie A, et al. Central role of reverting mutations in HLA associations with human immunodeficiency virus set point. *Journal of virology*. 2008 Sep;**82**(17):8548-59.

28. Wang YE, Li B, Carlson JM, et al. Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in highly conserved regions of human immunodeficiency virus type 1. *Journal of virology*. 2009 Feb;**83**(4):1845-55.
29. Rousseau CM, Daniels MG, Carlson JM, et al. HLA class I-driven evolution of human immunodeficiency virus type 1 subtype c proteome: immune escape and viral load. *Journal of virology*. 2008 Jul;**82**(13):6434-46.
30. Brumme ZL, Tao I, Szeto S, et al. Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection. *AIDS*. 2008 Jul 11;**22**(11):1277-86.
31. Poon AF, Swenson LC, Dong WW, et al. Phylogenetic analysis of population-based and deep sequencing data to identify coevolving sites in the nef gene of HIV-1. *Molecular biology and evolution*. 2010 Apr;**27**(4):819-32.
32. Walker BD, Burton DR. Toward an AIDS vaccine. *Science*. 2008 May 9;**320**(5877):760-4.
33. Carlson JM, Brumme ZL. HIV evolution in response to HLA-restricted CTL selection pressures: a population-based perspective. *Microbes and infection / Institut Pasteur*. 2008 Apr;**10**(5):455-61.
34. Migueles SA, Sabbaghian MS, Shupert WL, et al. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proceedings of the National Academy of Sciences of the United States of America*. 2000 Mar 14;**97**(6):2709-14.
35. Altfeld M, Addo MM, Rosenberg ES, et al. Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. *AIDS*. 2003 Dec 5;**17**(18):2581-91.
36. Migueles SA, Laborico AC, Imamichi H, et al. The differential ability of HLA B*5701+ long-term nonprogressors and progressors to restrict human immunodeficiency virus replication is not caused by loss of recognition of autologous viral gag sequences. *Journal of virology*. 2003 Jun;**77**(12):6889-98.
37. Kaslow RA, Carrington M, Apple R, et al. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nature medicine*. 1996 Apr;**2**(4):405-11.
38. Goulder PJ, Bunce M, Krausa P, et al. Novel, cross-restricted, conserved, and immunodominant cytotoxic T lymphocyte epitopes in slow progressors in HIV type 1 infection. *AIDS research and human retroviruses*. 1996 Dec 10;**12**(18):1691-8.
39. Streeck H, Lichterfeld M, Alter G, et al. Recognition of a defined region within p24 gag by CD8+ T cells during primary human immunodeficiency virus type 1 infection in individuals expressing protective HLA class I alleles. *Journal of virology*. 2007 Jul;**81**(14):7725-31.

40. Boutwell CL, Rowley CF, Essex M. Reduced viral replication capacity of human immunodeficiency virus type 1 subtype C caused by cytotoxic-T-lymphocyte escape mutations in HLA-B57 epitopes of capsid protein. *Journal of virology*. 2009 Mar;**83**(6):2460-8.
41. Miura T, Brockman MA, Brumme ZL, et al. HLA-associated alterations in replication capacity of chimeric NL4-3 viruses carrying gag-protease from elite controllers of human immunodeficiency virus type 1. *Journal of virology*. 2009 Jan;**83**(1):140-9.
42. Pereyra F, Jia X, McLaren PJ, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science*. 2010 Dec 10;**330**(6010):1551-7.
43. Steinbrook R. One step forward, two steps back--will there ever be an AIDS vaccine? *The New England journal of medicine*. 2007 Dec 27;**357**(26):2653-5.
44. Buchbinder SP, Mehrotra DV, Duerr A, et al. Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *Lancet*. 2008 Nov 29;**372**(9653):1881-93.
45. McElrath MJ, De Rosa SC, Moodie Z, et al. HIV-1 vaccine-induced immunity in the test-of-concept Step Study: a case-cohort analysis. *Lancet*. 2008 Nov 29;**372**(9653):1894-905.
46. Rolland M, Nickle DC, Mullins JI. HIV-1 group M conserved elements vaccine. *PLoS pathogens*. 2007 Nov;**3**(11):e157.
47. Allen TM, Altfeld M. Crippling HIV one mutation at a time. *The Journal of experimental medicine*. 2008 May 12;**205**(5):1003-7.
48. Chopera DR, Wright JK, Brockman MA, Brumme ZL. Immune-mediated attenuation of HIV-1. *Future virology*. 2011 Aug;**6**(8):917-28.
49. Almeida JR, Price DA, Papagno L, et al. Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. *The Journal of experimental medicine*. 2007 Oct 1;**204**(10):2473-85.
50. Geldmacher C, Currier JR, Herrmann E, et al. CD8 T-cell recognition of multiple epitopes within specific Gag regions is associated with maintenance of a low steady-state viremia in human immunodeficiency virus type 1-seropositive patients. *Journal of virology*. 2007 Mar;**81**(5):2440-8.
51. Klein MR, van Baalen CA, Holwerda AM, et al. Kinetics of Gag-specific cytotoxic T lymphocyte responses during the clinical course of HIV-1 infection: a longitudinal analysis of rapid progressors and long-term asymptomatics. *The Journal of experimental medicine*. 1995 Apr 1;**181**(4):1365-72.
52. Riviere Y, McChesney MB, Porrot F, et al. Gag-specific cytotoxic responses to HIV type 1 are associated with a decreased risk of progression to AIDS-related complex or AIDS. *AIDS research and human retroviruses*. 1995 Aug;**11**(8):903-7.

53. Zuniga R, Lucchetti A, Galvan P, et al. Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control. *Journal of virology*. 2006 Mar;**80**(6):3122-5.
54. Borghans JA, Molgaard A, de Boer RJ, Kesmir C. HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. *PLoS one*. 2007;**2**(9):e920.
55. Edwards BH, Bansal A, Sabbaj S, Bakari J, Mulligan MJ, Goepfert PA. Magnitude of functional CD8+ T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma. *Journal of virology*. 2002 Mar;**76**(5):2298-305.
56. Kiepiela P, Ngumbela K, Thobakgale C, et al. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nature medicine*. 2007 Jan;**13**(1):46-53.
57. Dismuke DJ, Aiken C. Evidence for a functional link between uncoating of the human immunodeficiency virus type 1 core and nuclear import of the viral preintegration complex. *Journal of virology*. 2006 Apr;**80**(8):3712-20.
58. Forshey BM, von Schwedler U, Sundquist WI, Aiken C. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *Journal of virology*. 2002 Jun;**76**(11):5667-77.
59. Novitsky V, Smith UR, Gilbert P, et al. Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? *Journal of virology*. 2002 Jun;**76**(11):5435-51.
60. Peyerl FW, Bazick HS, Newberg MH, Barouch DH, Sodroski J, Letvin NL. Fitness costs limit viral escape from cytotoxic T lymphocytes at a structurally constrained epitope. *Journal of virology*. 2004 Dec;**78**(24):13901-10.
61. Brockman MA, Brumme ZL, Brumme CJ, et al. Early selection in Gag by protective HLA alleles contributes to reduced HIV-1 replication capacity that may be largely compensated for in chronic infection. *Journal of virology*. 2010 Nov;**84**(22):11937-49.
62. Brockman MA, Schneidewind A, Lahaie M, et al. 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. *Journal of virology*. 2007 Nov;**81**(22):12608-18.
63. Schneidewind A, Brockman MA, Yang R, et al. 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. *Journal of virology*. 2007 Nov;**81**(22):12382-93.
64. Miura T, Brumme ZL, Brockman MA, et al. Impaired replication capacity of acute/early viruses in persons who become HIV controllers. *Journal of virology*. 2010 Aug;**84**(15):7581-91.

65. Leslie AJ, Pfafferott KJ, Chetty P, et al. HIV evolution: CTL escape mutation and reversion after transmission. *Nature medicine*. 2004 Mar;**10**(3):282-9.
66. Martinez-Picado J, Prado JG, Fry EE, et al. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. *Journal of virology*. 2006 Apr;**80**(7):3617-23.
67. Friedrich TC, Dodds EJ, Yant LJ, et al. Reversion of CTL escape-variant immunodeficiency viruses in vivo. *Nature medicine*. 2004 Mar;**10**(3):275-81.
68. Li B, Gladden AD, Altfeld M, et al. Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *Journal of virology*. 2007 Jan;**81**(1):193-201.
69. Allen TM, Altfeld M, Yu XG, et al. Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. *Journal of virology*. 2004 Jul;**78**(13):7069-78.
70. Schneidewind A, Brumme ZL, Brumme CJ, et al. Transmission and long-term stability of compensated CD8 escape mutations. *Journal of virology*. 2009 Apr;**83**(8):3993-7.
71. Crawford H, Prado JG, Leslie A, et al. 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. *Journal of virology*. 2007 Aug;**81**(15):8346-51.
72. Goulder PJ, Brander C, Tang Y, et al. Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature*. 2001 Jul 19;**412**(6844):334-8.
73. Brander C, Walker BD. Gradual adaptation of HIV to human host populations: good or bad news? *Nature medicine*. 2003 Nov;**9**(11):1359-62.
74. McMichael A, Klenerman P. HIV/AIDS. HLA leaves its footprints on HIV. *Science*. 2002 May 24;**296**(5572):1410-1.
75. Leslie A, Kavanagh D, Honeyborne I, et al. Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. *The Journal of experimental medicine*. 2005 Mar 21;**201**(6):891-902.
76. Haas DW, Wilkinson GR, Kuritzkes DR, et al. A multi-investigator/institutional DNA bank for AIDS-related human genetic studies: AACTG Protocol A5128. *HIV clinical trials*. 2003 Sep-Oct;**4**(5):287-300.
77. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*. 2003 Aug 5;**100**(16):9440-5.

78. Yusim K, Korber B, Brander C, et al. HIV Molecular Immunology. Los Alamos National Laboratory, Los Alamos, New Mexico: Theoretical Biology and Biophysics, Los Alamos, New Mexico.; 2009.
79. Draenert R, Le Gall S, Pfafferoth KJ, et al. Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. The Journal of experimental medicine. 2004 Apr 5;**199**(7):905-15.
80. Yokomaku Y, Miura H, Tomiyama H, et al. Impaired processing and presentation of cytotoxic-T-lymphocyte (CTL) epitopes are major escape mechanisms from CTL immune pressure in human immunodeficiency virus type 1 infection. Journal of virology. 2004 Feb;**78**(3):1324-32.
81. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC evolutionary biology. 2007;**7**:214.
82. Kuiken C, Foley B, Leitner T, et al. HIV Sequence Compendium. Los Alamos National Laboratory, Los Alamos, New Mexico: Theoretical Biology and Biophysics Group; 2010.
83. Brockman MA, Tanzi GO, Walker BD, Allen TM. Use of a novel GFP reporter cell line to examine replication capacity of CXCR4- and CCR5-tropic HIV-1 by flow cytometry. Journal of virological methods. 2006 Feb;**131**(2):134-42.
84. Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. The emergence of HIV/AIDS in the Americas and beyond. Proceedings of the National Academy of Sciences of the United States of America. 2007 Nov 20;**104**(47):18566-70.
85. Reveille JD. The genetic basis of spondyloarthritis. Annals of the rheumatic diseases. 2011 Mar;**70 Suppl 1**:i44-50.
86. Akahoshi T, Chikata T, Tamura Y, Gatanaga H, Oka S, Takiguchi M. Selection and accumulation of an HIV-1 escape mutant by three types of HIV-1-specific cytotoxic T lymphocytes recognizing wild-type and/or escape mutant epitopes. Journal of virology. 2012 Feb;**86**(4):1971-81.
87. Shugars DC, Smith MS, Glueck DH, Nantermet PV, Seillier-Moiseiwitsch F, Swanstrom R. Analysis of human immunodeficiency virus type 1 nef gene sequences present in vivo. Journal of virology. 1993 Aug;**67**(8):4639-50.
88. Kirchhoff F, Easterbrook PJ, Douglas N, et al. Sequence variations in human immunodeficiency virus type 1 Nef are associated with different stages of disease. Journal of virology. 1999 Jul;**73**(7):5497-508.
89. Huang Y, Zhang L, Ho DD. Characterization of nef sequences in long-term survivors of human immunodeficiency virus type 1 infection. Journal of virology. 1995 Jan;**69**(1):93-100.

4. Chapter 4: General Conclusions

4.1. Public Health Implications

The objective of this thesis was to explore how immune selection pressures mediated through host human leukocyte antigen (HLA)-restricted cytotoxic T-lymphocyte (CTL) responses have shaped the genomic and functional evolution of the HIV-1 *gag* gene over the course of the epidemic in North America. CTL-driven immune responses play a critical role in controlling HIV-1 infection (1, 2) and have, consequently, become an important target in HIV-1 vaccine strategies (3, 4). However, of increasing concern is the selection of escape mutations throughout the HIV-1 proteome such that CTL-inducing vaccine strategies may not be viable in the long run (5). Developing a deeper understanding of the host and viral genetic factors influencing HIV-1 evolution is invaluable for public health initiatives aimed at vaccine design.

Taken together, the findings of this thesis indicate that HLA-restricted CTL responses have continued to induce consistent patterns of immune escape in HIV-1 *gag* over the course of the North American epidemic. Furthermore, the prevalence of known escape mutations in the general population has only modestly increased over the past 30 years. These results argue against the rapid and substantial accumulation of escape mutations as outlined by the “HLA imprinting hypothesis” (6-8). Rather, these observations support the dynamic adaptation of HIV-1 over time such that the virus appears to be continually escaping and reverting as it passes through human hosts. Additionally, only modest increases in *gag*-mediated replication capacity between pre-

and post-1985 sequences were observed. Although the mechanism(s) behind these increases remain unknown, they may be attributable to factors other than CTL-driven immune responses. From a public health and vaccine perspective, the lack of substantial accumulation of escape mutations suggests that CTL immune responses targeted at epitopes within *gag* will continue to provide protection against the most commonly circulating strains of HIV-1 for a while to come. These results suggest that vaccine strategies aimed at eliciting an immune response towards epitopes within *gag* should continue to remain an important goal.

Over the past 30 years of the epidemic, HIV-1 vaccine strategies have been met with little success. In 2007, results from the STEP trial, a vaccine delivered using an adenovirus serotype 5 (Ad5) vector aimed at eliciting CTL responses towards *gag*, *pol* and *nef*, revealed adverse reactions where the incidence of HIV-1 infection in individuals receiving the vaccine treatment, most notably among uncircumcised men with pre-existing Ad5-seropositivity (9), was higher than the incidence in individuals receiving a placebo (10-12). Although the mechanisms behind this increased HIV-1 acquisition risk remains incompletely known (13), more hopeful results of the recent RV144 HIV-1 vaccine trial (14), continues to fuel vaccine initiatives. Regardless, an incomplete understanding of the immune responses associated with effective HIV-1 control suggests that an effective HIV-1 vaccine remains a relatively distant goal.

Recently, public health HIV-1 prevention strategies using antiretroviral drugs reformulated as biomedical prevention tools have met with substantial success. These strategies include PreExposure Prophylaxis (PrEP), microbicides and treatment as prevention. The aim of PrEP is to treat individuals at high risk of acquiring HIV-1 with orally administered antiretroviral drugs for the duration that they are at risk to decrease

their risk of acquiring HIV-1 (15). Microbicides featuring the antiretroviral drug Tenofovir have also shown efficacy in clinical trials in South Africa (16). These compounds are generally administered intra-vaginally or intra-rectally and are used to decrease the risk of transmission during sexual contact. Although pre-exposure therapy has been shown to reduce the rate of HIV-1 transmission, from a public health perspective there are substantial limitations. Finally, treatment as prevention has additionally demonstrated that dramatic reductions in HIV-1 RNA circulating in the plasma and sexual fluids of individuals receiving antiretroviral therapy decreases the risk of HIV-1 transmission (17, 18). Treating serodiscordant couples in the early stages of HIV-1 infection has resulted in drastic reductions in the number of linked transmission events (18).

Although PrEP, microbicides and treatment as prevention strategies have clearly shown efficacy in clinical trials, the primary limitation of such strategies is the responsibility of the individual to comply with specific regimes outlined by their physician. If antiretroviral drugs are not taken properly or applied in the correct manner, the protection of these therapies substantially diminishes. Methods of prevention that rely on individual behaviour are often hard to implement at a population-level to the extent of seeing attributable reductions in incidence. From a public health perspective developing a vaccine that achieves “sterilizing immunity” or attenuation of pathogenesis to the point of reducing transmission, remains an important goal (3). Not only would this provide immediate protection for high-risk individuals but would also continue to remain sustainable at the population-level.

The goal of this thesis was to address some of the purported issues associated with CTL vaccine design using molecular and evolutionary techniques to explore host-viral interactions while approaching these aims from a public health perspective. HIV-1

remains a global epidemic and continues to kill approximately 1.8 million people around the world on a yearly basis (19). It will be essential for public health and science to work together in the future to develop novel HIV-1 prevention strategies using new technologies and research. The research presented here reiterates the continuing importance of CTL-driven HIV-1 adaptation throughout populations in North America and will hopefully fuel further studies in other areas across the world. It is my hope that the current project will contribute to future vaccine approaches.

4.2. References

1. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *Journal of virology*. 1994 Sep;**68**(9):6103-10.
2. Koup RA, Safrit JT, Cao Y, et al. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *Journal of virology*. 1994 Jul;**68**(7):4650-5.
3. Walker BD, Burton DR. Toward an AIDS vaccine. *Science*. 2008 May 9;**320**(5877):760-4.
4. Goulder PJ, Watkins DI. HIV and SIV CTL escape: implications for vaccine design. *Nature reviews. Immunology*. 2004 Aug;**4**(8):630-40.
5. Leslie A, Kavanagh D, Honeyborne I, et al. Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. *The Journal of experimental medicine*. 2005 Mar 21;**201**(6):891-902.
6. Brander C, Walker BD. Gradual adaptation of HIV to human host populations: good or bad news? *Nature medicine*. 2003 Nov;**9**(11):1359-62.
7. Goulder PJ, Brander C, Tang Y, et al. Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature*. 2001 Jul 19;**412**(6844):334-8.
8. McMichael A, Klenerman P. HIV/AIDS. HLA leaves its footprints on HIV. *Science*. 2002 May 24;**296**(5572):1410-1.
9. D'Souza MP, Frahm N. Adenovirus 5 serotype vector-specific immunity and HIV-1 infection: a tale of T cells and antibodies. *AIDS*. 2010 Mar 27;**24**(6):803-9.

10. Buchbinder SP, Mehrotra DV, Duerr A, et al. Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *Lancet*. 2008 Nov 29;**372**(9653):1881-93.
11. McElrath MJ, De Rosa SC, Moodie Z, et al. HIV-1 vaccine-induced immunity in the test-of-concept Step Study: a case-cohort analysis. *Lancet*. 2008 Nov 29;**372**(9653):1894-905.
12. Steinbrook R. One step forward, two steps back--will there ever be an AIDS vaccine? *The New England journal of medicine*. 2007 Dec 27;**357**(26):2653-5.
13. Hutnick NA, Carnathan DG, Dubey SA, et al. Baseline Ad5 serostatus does not predict Ad5 HIV vaccine-induced expansion of adenovirus-specific CD4+ T cells. *Nature medicine*. 2009 Aug;**15**(8):876-8.
14. Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, et al. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *The New England journal of medicine*. 2009 Dec 3;**361**(23):2209-20.
15. Grant RM, Lama JR, Anderson PL, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *The New England journal of medicine*. 2010 Dec 30;**363**(27):2587-99.
16. Abdool Karim Q, Abdool Karim SS, Frohlich JA, et al. Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science*. 2010 Sep 3;**329**(5996):1168-74.
17. Montaner JS, Lima VD, Barrios R, et al. Association of highly active antiretroviral therapy coverage, population viral load, and yearly new HIV diagnoses in British Columbia, Canada: a population-based study. *Lancet*. 2010 Aug 14;**376**(9740):532-9.
18. Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *The New England journal of medicine*. 2011 Aug 11;**365**(6):493-505.
19. UNAIDS. UNAIDS: World AIDS Day Reprot, 2011. Geneva, Switzerland; 2011.

Appendices

Appendix A.

HLA Class I Sequencing Reference Standards

HLA-A_exon2&3_std

GCTCCCACTCCATGAGGTATTTCTTCACATCCGTGTCCCGGCCCGGCCGCGGGGA
GCCCCGCTTCATCGCCGTGGGCTACGTGGACGACACGCAGTTCGTGCGGTTTCGAC
AGCGACGCCGCGAGCCAGAAGATGGAGCCGCGGGCGCCGTGGATAGAGCAGGA
GGGGCCGGAGTATTGGGACCAGGAGACACGGAATATGAAGGCCCACTCACAGACT
GACCGAGCGAACCTGGGGACCCTGCGCGGCTACTACAACCAGAGCGAGGACGGT
GAGTGACCCCGGCCCGGGGCGCAGGTCACGACCCCTCATCCCCACGGACGGGC
CAGGTCGCCACAGTCTCCGGGTCCGAGATCCACCCCGAAGCCGCGGGACTCCG
AGACCCTTGTCCCGGGAGAGGCCAGGCGCCTTTACCCGGTTTCATTTTCAGTTTA
GGCCAAAATCCCCCGGGTGTGTCGGGGCGGGGCGGGGCTCGGGGGACTGGG
CTGACCGCGGGGTGCGGGCCAGGTTCTCACACCATCCAGATAATGTATGGCTGCG
ACGTGGGGCCGGACGGGCGCTTCTCCGCGGGTACCGGCAGGACGCCTACGACG
GCAAGGATTACATCGCCCTGAACGAGGACCTGCGCTCTTGGACCGCGGCGGACAT
GGCAGCTCAGATACCAAGCGCAAGTGGGAGGCGGTCCATGCGGCGGAGCAGCG
GAGAGTCTACCTGGAGGGCCGGTGGCTGGACGGGCTCCGCAGATACCTGGAGAA
CGGGAAGGAGACGCTGCAGCGCACGG

HLA-B_exon2_std

GCTCCCACTCCATGAGGTATTTCTACACCTCCGTGTCCCGGCCCGGCCGCGGGGA
GCCCCGCTTCATCTCAGTGGGCTACGTGGACGACACCCAGTTCGTGAGGTTTCGAC
AGCGACGCCGCGAGTCCGAGAGAGGAGCCGCGGGCGCCGTGGATAGAGCAGGA
GGGGCCGGAGTATTGGGACCAGAACACACAGATCTACAAGGCCAGGCACAGACT
GACCGAGAGAGCCTGCGGAACCTGCGCGGCTACTACAACCAGAGCGAGGCGG

HLA-B_exon3_std

GGTCTCACACCCTCCAGAGCATGTACGGCTGCGACGTGGGGCCGGACGGGGCGCC
TCCTCCGCGGGCATGACCAGTACGCCTACGACGGCAAGGATTACATCGCCCTGAA
CGAGGACCTGCGCTCCTGGACCGCCGCGGACACGGCGGCTCAGATACCCAGCG
CAAGTGGGAGGCGGCCCGTGAGGCGGAGCAGCGGAGAGCCTACCTGGAGGGCG
AGTGCGTGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGACAAGCTGGAGC
GCGCTG

HLA-C_exon2_std

GCTCCCACTCCATGAAGTATTTCTTCACATCCGTGTCCCGGCCTGGCCGCGGAGA
GCCCCGCTTCATCTCAGTGGGCTACGTGGACGACACGCAGTTCGTGCGGTTTCGAC
AGCGACGCCGCGAGTCCGAGAGGGGAGCCGCGGGCGCCGTGGGTGGAGCAGGA
GGGGCCGGAGTATTGGGACCAGGAGACACAGAAGTACAAGCGCCAGGCACAGAC
TGACCGAGTGAGCCTGCGGAACCTGCGCGGCTACTACAACCAGAGCGAGGCGG

HLA-C_exon3_std

GGTCTCACACCCTCCAGTGGATGTGTGGCTGCGACCTGGGGCCCGACGGGGCGCC
TCCTCCGCGGGTATGACCAGTACGCCTACGACGGCAAGGATTACATCGCCCTGAA
CGAGGACCTGCGCTCCTGGACCGCCGCGGACACCGCGGCTCAGATACCCAGCG

CAAGTGGGAGGCGGCCCGTGAGGCGGAGCAGCGGAGAGCCTACCTGGAGGGCA
CGTGCGTGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGCAGC
GCGCGG