HIV-1 Viral Genetics and Superinfection within a Cohort of Commercial Sex Workers:

Epidemiological, Virological and Clinical Consequences

BY

TAMMY LEIGH STUART

A thesis submitted to the faculty of Graduate Studies of
the University of Manitoba
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Medical Microbiology and Infectious Diseases
University of Manitoba
Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

HIV-1 Viral Genetics and Superinfection within a Cohort of Commercial Sex Workers: Epidemiological, Virological and Clinical Consequences

by

Tammy Leigh Stuart

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

of

Doctor of Philosophy

ABSTRACT

Understanding HIV-1 genetic diversity and host-virus relationship is essential to the development of new treatments and vaccine design. A cohort of HIV-1 positive, drug-naïve commercial sex workers from Kenya was studied in order to investigate these issues. The research was driven by three global hypotheses:

1) Single HIV-1 infections would reveal; an array of genetically distinct viruses, a dynamic distribution of viral subtypes, and differences in disease progression.

2) Infection with HIV-1 would provide protection against HIV-1 superinfection.

3) Rapid disease progression would be seen in superinfected/dually infected individuals and recombinant viruses would be present in most cases.

Analysis of HIV-1 sequence data revealed multiple circulating subtypes, the distribution of which significantly changed over time. Recombinant forms were detected early in the epidemic in this cohort and diversity was seen within the amino acid sequences of a major drug target (HIV-1 protease). No associations between viral subtypes and disease progression were found, but non-subtype associated viral genetic characteristics may impact upon disease progression. These data highlight the importance of a candidate vaccine capable of providing immunity against an array of viral subtypes and the need for continued monitoring of the viral composition within populations.

The discovery of only four confirmed cases of superinfection within a nested cohort study revealed that HIV-1 infection has a protective effect against the acquisition of a second HIV infection (Hazard ratio=0.14 [CI95% 0.05-0.39]). This provides hope that a cross protective vaccine may be possible.

In total 11 of 213 sex workers had evidence of superinfection/dual infections. Recombinant viruses were found in 10/11 cases (90.9%) and became the predominant form in 60% of these cases. This extreme ability of HIV to evolve creates new challenges for treatment, diagnostic tests and vaccine design.

Where it was possible to estimate the timing of superinfection, an increase in viral load and decrease in CD4+ T cells were observed concurrently. Contrary to the hypothesis, infection with multiple viruses (superinfection/dual infection cases) did not always lead to rapid progression. Given that the clinical progression of dually/superinfected individuals was not homogeneous the true impact of multiple virus infections is likely dependent on both host genetics and the viruses involved.

The studies reported in this thesis serve to better define viral genetics within a sex worker cohort and to more clearly define the host-virus relationship.

DEDICATION

I would like to dedicate this thesis to my parents, Linda and Larry Stuart.

You have always encouraged me in everything that I do, including the writing of this thesis. By example you have taught me the importance of perserverance, dedication and hard work in reaching any goal.

The most important things I have learned in life so far do not lie within the pages of this thesis but in the lessons I have learned from you along the way.

Thank you for supporting my every dream.

ACKNOWLEDGEMENTS

I would first like to thank my supervisors Frank Plummer and Kevin Coombs. To Frank; your enthusiasm for research is contagious and your leadership and dedication to the pursuit of improving global health has been such a huge inspiration to me. I have tremendous respect for you and am constantly in awe of how you manage to do it all! I am incredibly thankful for the opportunities that you have provided me with over the years. They have certainly shaped me both as a scientist and a person. You have always made me feel like you were "in my corner". Your belief in me really gave me the confidence I needed to complete this thesis.

To Kevin; thank you is really not enough! You have been an ever constant source of encouragement, support and wisdom. You are an exceptional scientist and wonderful person and I am so thankful to have you as a role model. Thank you for teaching me how to problem solve and to think scientifically. Thank you for also reminding me (on more than one occasion) that sometimes we all need to "head back into the woods". Your advice has been invaluable - somehow you always manage to be right! This journey would have been impossible without you.

To my other committee members Joanne Embree, Grant McClarty and Deborah Court; I am very thankful for the encouragement you have given me and truly appreciate the role that each of you has played in my development as a scientist. Thank you for the time and effort you have put into being a part of my committee.

Thank you also to the Natural Sciences and Engineering Reasearch Council of Canada (NSERC), Manitoba Health Research Council (MHRC) and the Canadian Institute of Health Research (CIHR) for financial support over the years.

I have had the privilege of working with many fantastic people within the HIV group over the years. Ma Luo, thank you for always cheering me on and for your eternal optimism! Xiaojuan Mao, your willingness to help and the assistance you have given me in the lab has been invaluable. To Suzie Ramdahin, Leslie Slaney and Steve Wayne; thank you for your friendship and for the help and guidance you have given me over the years. To Blake Ball and Keith Fowke; thank you for the time and effort you have put into making the entire HIV research program so much better for everyone!

To Lawrence Elliott from Community Health Sciences; thank you for broadening my horizons and for helping me to find an answer to the dreaded "what are you going to do after grad school?" question. Thank you also to Mary Chaeng for sharing her statistical expertise.

I would also like to thank the individuals who have worked so hard at the lab and in the clinic in Nairobi, Kenya over the years. To Joshua Kimani, Charles Wachici, Jane Njoki, Jane Kamene, Elizabeth Bwibo and Edith Amatiwa; *Asante sana.*

The friendships I have enjoyed with fellow grad students have certainly been one of the best things to come out of my time in grad school. I would like to thank, Megan and Israel Mendez and Maggie Swanson from the Coombs lab as well as Melanie Murray, Allison Land, Julius Oyugi, Hezhao Ji and Francoise Vouriot from the HIV group.

I would also like to give an extra special thank you to my fellow HIV lab grad students Paul McLaren, Sandy Koesters, Jill Waruk, Lyle McKinnon and Shehzad Iqbal with whom I have been on this PhD grad school journey the longest!! You have been wonderful colleagues, but friends first and foremost. Your truly genuine caring and support have made all the difference. I feel very lucky to have shared this grad school experience with each of you.

Thank you also to Emma Illife-Lee, Kulvinder Gill, Lisa Baspaly and Julie Boutilier for your words of encouragement along the way. They have meant so much.

To Tom Dowden, Nicole Lafreniere, Mei-Ling Mah and Lori Cherwinski; thank you for the reminders that life is a balancing act and also that it is perfectly okay to "colour outside the lines!"

To my parents, Linda and Larry Stuart, my sister Susan, brothers Clinton and Tom, my Grandma and Grandpa Lawson and my Grandma Stuart; thank you for your love and encouragement and for being a constant reminder of what is really important.

To Ryan Zarychanski; thank you for your endless encouragement and loving support, especially over this past year. Thank you for putting up with the long hours, stressful days and nights, and for also being the first person to read my entire thesis!

I will end this now with one final and profoundly important thank you. I would like to thank the women of the Pumwani sex worker cohort who have allowed the HIV research group to study them for so long and without whom this research would clearly have been impossible. They have also allowed us to put some faces and stories behind the statistics that we so often recite and are a constant reminder of why health research is so important.

TABLE OF CONTENTS

LIST OF FIGURES

List of Tables

INTRODUCTION

Overview:

The HIV-1 pandemic is a global health crisis of unprecedented magnitude. The number of people living with HIV continues to climb and has reached a reported 40.3 million (36.7- 45.3 million) as of the end of 2005, with a staggering 4.9 million (4.3-6.6) new infections and 3.1 million (2.8-3.6 million) deaths due to the acquired immunodeficiency syndrome (AIDS) within the year 2005 alone (1). Sub-Saharan Africa is undeniably the hardest hit region where nearly two-thirds (23.8-28.9 million) of those living with HIV reside. The impact of the epidemic on women continues to increase within this region where females currently make up 57% of adults living with HIV (2). It has thus been said that AIDS in Africa wears "a woman's face" (3).

These infection rates and death statistics present us with a picture of the toll that the epidemic has taken on human life. The implications though are even farther reaching with the resulting social and economic burden on families and communities as well as the creation of political strife. There are already millions of orphans created throughout the world by HIV and it is estimated that by 2010 the number of orphaned children will reach 25 million (4). Those who are not infected with the deadly virus are unquestionably still very much affected by the devastation and havoc it has and continues to create.

Unfortunately the global response has not been equivalent to the enormity of the HIV pandemic. HIV prevention strategies have had great successes, are highly cost-effective and have been predicted to curb the epidemic in developing countries if expanded (5); however, despite these facts, it has been estimated that less than one in five individuals at risk of acquiring HIV have access to prevention initiatives (1).

Antiretroviral pharmaceutical development has been very successful and has turned HIV into a treatable disease within the developed world but has not been a reality for the majority of individuals living in developing nations (6).

Global research efforts directed towards a better understanding of HIV-1 have been greater than for any other infectious disease and yet the development of an efficacious vaccine continues to evade the scientific community. The development of microbicides that can significantly decrease virus transmission and allow women to protect themselves are desperately needed (7). Ultimately though, an effective vaccine against HIV is paramount to halting the pandemic.

Discovery of HIV-1: the Etiologic Agent of AIDS:

It is doubtful that when AIDS was first discovered anyone could have fathomed the catastrophic impact that HIV, the causative agent, would have on global health. It was in June of 1981 that this syndrome was first described by the Centers for Disease Control (CDC) in the Morbidity and Mortality Weekly Report. Five previously healthy homosexual men from Los Angeles, California presented with severe immunodeficiency and *Pneumocystis carinii* pneumonia (PCP), an opportunistic infection (8). Additional cases of PCP and other opportunistic infections as well as cases of Kaposi sarcoma (resulting from Human Herpesvirus type 8 (HHV-8)) and generalized lymphadenopathy (enlargement of the lymph nodes) were soon reported in other previously healthy individuals. These individuals had in common a depletion of CD4+ T cells and a resulting deficiency in cell-mediated immunity (9-12) . This syndrome would soon become known as acquired immunodeficiency syndrome, or AIDS.

Although the initial identification of AIDS was in homosexual men (8) the syndrome was soon observed within many diverse groups of individuals including Haitians (13;14), hemophiliacs (15;16), injection drug users (IDUs) (12;17), sexual partners of individuals in high risk groups (18-20), as well as in the children of mothers within high risk groups (21-23). The available epidemiological evidence pointed towards an etiologic agent that was an infectious microorganism which could be spread through sexual contact or blood products. At the time, human T-cell leukemia/lymphoma viruses (HTLV) had recently been identified (24;25). Due to the T cell tropism of these viruses and the presence of antibodies in patients that could react with HTLV-I the hunt for a retrovirus in individuals with AIDS was initiated.

In 1983 Luc Montagnier and his group at the Pasteur institute (Paris, France) isolated a virus from the lymph nodes of a patient who presented with lymphadenopathy. Lymphadenopathy virus as they called it had reverse transcriptase activity and specifically killed CD4+ T cells (26;27). Soon after, Gallo and his research team at the National Institutes of Health also reported the isolation of a retrovirus from AIDS patients which they referred to as HTLV-III (28). Additionally, Jay Levy et al. isolated a retrovirus from individuals with AIDS and from those that were asymptomatic but in high risk groups, and named the virus AIDS-associated retrovirus (26;29). The viruses isolated by these groups were the same and would later be renamed HIV-1 (30). In 1984 both Gallo and Montagnier published landmark studies establishing HIV as the causative agent of AIDS (31;32).

Shortly after the discovery of HIV it was predicted that a vaccine would soon become readily available (33). Now in 2006, 23 years since the discovery of HIV-1,

despite the extraordinary efforts of the global scientific community; no such vaccine exists. However, the undisputed global impact of this virus has been the impetus which has led to extensive research and great advances in virology, immunology and the resulting host-virus dynamic.

HIV: Virus Structure, Replication and Genetics

HIV Structure:

The discovery of HIV, and the scientific investigation that has followed, has lead to the acquisition of a tremendous amount of information regarding the viral structure, replicative processes and genetics of HIV. HIV is a member of the *Retroviridae* family of viruses, which are characterized by their reverse transcriptase ability. It is classified within the *Lentivirus* genus of this family, a genus well known for its ability to infect specific types of hematopoietic cells (especially lymphocytes and macrophages) and cause slowly progressive, but fatal diseases (34;35).

HIV is a 110-120 nm enveloped virus with a cone-shaped core. The envelope is host derived and contains host proteins as well as viral envelope (Env) glycoproteins, gp120 (surface glycoprotein (SU)) and gp41 (transmembrane glycoprotein (TM)). HIV envelope glycoprotein complexes are present on the cell surface in the form of trimers composed of noncovalently bound gp120-gp41 pairs. The core consists of the capsid (CA) protein (p24), the matrix (MA) protein (p17) which is located on the inside of the envelope, and the nucleocapsid (NC) protein (p7) which is associated with the viral ribonucleic acid (RNA). The HIV genome is linear, + sense, single stranded RNA of which there are two identical copies found within each virion. The HIV-1 proviral DNA

version of the genome is approximately 9.7 kilobases (kb) in length and has a 5' to 3' arrangement of the gag, pol and env genomic regions as well as long terminal repeats (LTR) at each end (see Figure 1). The gag (group specific antigen) genomic region encodes for the proteins p24, p17 and p7, described above, as well as the p6 protein. The env region encodes for a precursor (gp160) that is cleaved to form the mature envelope glycoproteins gp120 and gp41. The pol region encodes for the viral protease (PR), reverse transcriptase (RT) and integrase (IN) proteins, all of which have important enzymatic roles. An additional 6 open reading frames (ORFs) encode for the accessory proteins, Vif, Vpr, Nef, Vpu, Tat and Rev (34;36;37). As is stated by Frankel et al., this virus is simply "15 proteins and an RNA"(38).

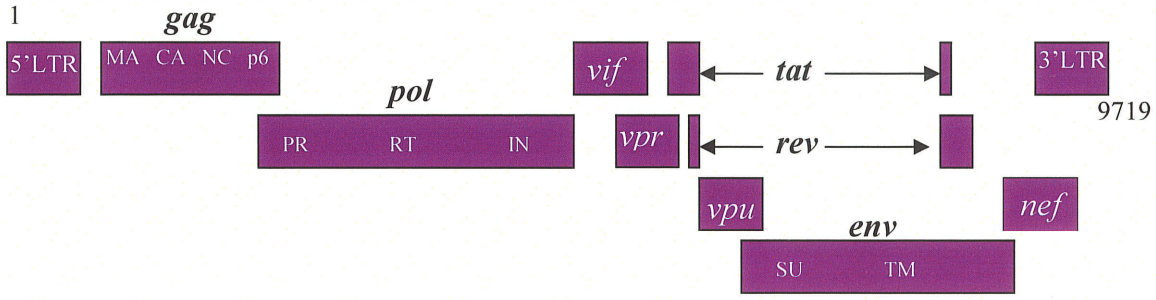

Figure 1: HIV-1 Genomic Arrangement (5'-3'). Numbering based on HXB2 positioning is 1-9719. Numbering according to HXB2 (gene start) is as follows: 5'LTR (1), MA/p17 (790), CA/p24 (1186), NC/p7 (1921), p6 (2134), PR (2253), RT (2550), IN (4230), vif (5041), vpr (5559), tat (5831), rev (5970), vpu (6062), SU/gp120 (6225), TM (gp41) 7758, nef (8797), 3'LTR (9086) (39).

HIV Replication:

However simple the structure itself may sound, the complexity of this virus and the challenges it presents must be emphasized. Soon after its discovery, scientists tried to understand how the virus enters and replicates within the body. We now know that HIV can infect CD4 + T cells and macrophages as well as some dendritic cell populations. Its tropism is regulated by the cellular proteins required for host cell entry (40). In 1984 the primary receptor for HIV was identified as the CD4 molecule (41;42). Membrane fusion and entry of the virus into cells however also requires the use of additional receptors, or co-receptors. Macrophage (M-tropic) viruses require the chemokine receptor CCR5 while T cell (T-tropic) viruses use CXCR4 (41;43-47). Binding of gp120 to the appropriate co-receptor induces a conformational change in gp41 that reveals a fusion peptide which in turn enables the viral envelope and cellular plasma membrane to fuse and thus allows virus entry to occur (34).

Following virus entry, partial uncoating of the virus takes place in the cytoplasm where the viral RNA and proteins are released. Reverse transcription of the RNA genome by the viral RT yields double stranded deoxyribonucleic acid (DNA) which is transported, as part of a nucleoprotein preintegration complex, into the nucleus. Once in the nucleus, the viral integrase protein catalyzes integration of the proviral DNA into the chromosomal DNA. This integrated viral DNA can then act as a template for the production of messenger RNA (mRNA) by host RNA polymerase II. Gene expression can be divided into an early and a late stage. In the early stage, short multiply spliced mRNAs encoding the regulatory proteins Tat, Rev and Nef are produced. These mRNAs are transported into the cytoplasm where translation occurs. Late stage gene expression

(env, gag, pol genes) includes structural gene expression and full length genome production. The Rev protein functions to initiate this late stage by enabling the export of singly spliced and unspliced RNA. Following translation within the cytoplasm, Gag and Gag-Pol polyproteins are transported to the plasma membrane. Translation of gp160 (Env) occurs in the endoplasmic reticulum followed by transport to the Golgi where it is cleaved by cellular proteases into the envelope glycoproteins gp120 and gp41. Complexes of gp120-gp41 are transported to and can integrate into the cytoplasmic membrane. Dimers of full length viral RNA can then associate with Gag and Gag-Pol polyproteins to form immature virions along with the membrane integrated gp120 and gp41 (34;36;38;48).

The next step is budding and release of the virus. During this process HIV protease is responsible for proteolytic processing of the Gag ($Pr55^{Gag}$) and Gag-Pol ($Pr160^{Gag\text{-}Pol}$) polyproteins. Protease releases itself from the Gag-Pol polyprotein via autocatalysis and functions as a dimeric enzyme to cleave various sites in the polyproteins. It releases the reverse transcriptase and integrase products from the Gag-Pol polyprotein; while cleavage of the Gag polyprotein releases the mature Gag proteins p17 (MA), p24 (CA), p7 (NC) and p6. This proteolytic process results in a change in virus morphology and subsequently the formation of a mature viral particle (34;36;38;48).

HIV-1 Genetics

Mechanisms and driving forces of HIV diversity:

Like HIV structure and virus replication the genetics of HIV have also been extensively studied. Based on studies of other retroviruses, including HTLVs, it was

initially thought that all HIV-1 viruses would be genetically uniform, but this was soon shown to be erroneous (34). We have now come to recognize that genetic diversity is a hallmark of HIV. Variability occurs as a result of a number of mechanisms. Reverse transcriptase, responsible for synthesis of the viral complementary DNA (cDNA), lacks proofreading activity which can result in point mutations as well as deletions and insertions (49-51). This error prone reverse transcriptase activity combined with a high replication rate leads to extensive sequence variation (52). In 1995 it was recognized that recombination was another mechanism driving HIV diversity (53). Recombinant viruses can be created when an individual is co-infected with different viral subtypes and RNA from each are packaged within the same virion. When this virus infects another cell the reverse transcriptase can switch between the different RNA strands and results in the formation of a recombinant virus (54-56). Recombination allows for extensive evolutionary changes to occur quickly (57). It has further been shown that viral evolution itself may occur in response to human leukocyte antigen (HLA) restricted immune responses (58). In addition to immune pressure, genetic diversity may also be driven by antiviral drugs (59). All of these factors have contributed to or driven the extraordinary diversity that is characteristic of HIV.

HIV classification and the extent of genetic diversity:

The incredible genetic diversity of HIV allows for its classification into different types, groups and into distinct subtypes within groups, resolved through phylogenetic analysis. Variability is also seen within each subtype and even amongst the quasi-species seen within individuals.

AIDS in humans can be caused by two distinct types of HIV, HIV-1 and HIV-2 which can be distinguished based on genome organization and phylogenetic relationships (60). HIV-2 was discovered in 1986, is most prevalent in west Africa and is far less pathogenic than HIV-1(61-63). The origin of HIV-2 is thought to be through zoonotic transmission of simian immunodeficiency virus (SIV) from sooty mangabeys (SIVsm) to humans while the origin of HIV-1 in the human population is due to transmission of SIV from chimpanzees (SIVcpz) (64). HIV-1 and HIV-2 can both be further divided into distinct groups based on phylogenetic analyses (60). Genetic differences allow for HIV-1 to be divided into groups M (major), N (non-M or O) and O (outlier). A single cross species transmission event of SIVcpz in west-central Africa, some time in the early part of the 20th century, appears to have given rise to HIV-1 group M, which is responsible for the current pandemic (60;65). Group M can be further separated into nine genetically defined lineages , or subtypes, A,B,C,D,F,G,H, J and K. Subtypes A and F are further delineated into sub-subtypes A1, A2 and F1, F2 respectively (66-68). It has been suggested that B and D should actually be classified as sub-subtypes as well but the current classification scheme is unlikely to change (69). To complicate matters further, there are currently 20 circulating recombinant forms (CRFs) that have been identified as well as unique recombinant forms (URFs). To be considered a CRF the recombinant virus must be sequenced and detected in at least three epidemiologically unrelated individuals. Consequently, recombinant viruses that do not fulfill these requirements are referred to as unique recombinant forms (URFs) (53;70). The HIV-1 subtypes and recombinant forms that are seen are not equally distributed throughout the world. On most continents there are a limited number of subtypes circulating, but in Africa almost

all subtypes and CRFs can be found (69). Subtype A and A/G viruses are the predominant types found in west and central Africa while subtype B is the predominant virus in the Americas and in Europe. Subtype C is found largely in southern and eastern Africa as well as in India and Nepal. East and central Africa are the primary locations where subtype D is found although it has been detected in southern and western Africa. Subtype E has not been detected on its own, only as part of A/E viruses which are found in central Africa, China, Thailand and the Philipines. G and A/G viruses are mainly in central Europe and western and eastern Africa. H and J are rare subtypes that have been found in central Africa and central America respectively. Lastly, subtype K has been localized to the Democratic Republic of Congo and Cameroon (71).

In addition to the accrued genetic diversity that allows for the designation of different HIV-1 subtypes, variability is also observed within subtypes and within individuals. Compared to HIV-1 Group M *intra*subtype variability, which is about 20-30% within the envelope genomic region, *inter*subtype variation within this region is 10-15% (71). Many different genetic variants or quasispecies are also found *in vivo* in infected patients (72). Studies have even shown that tissue compartmentalization and subsequent viral evolution may occur, resulting in diverse yet phylogenetically related sequences within different sites in individuals (73-78). HIV diversity is evident between types and subtypes and within subtypes and individuals. This diversity has served as both a tool and a challenge in HIV research.

Implications of genetic diversity:

HIV's genetic variability may impact vaccine development, antiretroviral treatment, disease progression and virus transmission. It is therefore important to expand our understanding of the consequences of virus diversity and also to monitor the variability within geographic regions and within patients.

Extensive genetic diversity presents a major challenge for the development of a vaccine capable of eliciting an immune response that is broadly protective against all HIV-1 variants (79). Vaccine design specific for different subtypes or geographic regions may be necessary (80;81). However, recent reports of cellular immune responses that are cross-subtype reactive provide hope that the development of a cross-protective vaccine may be achievable (82;83). It is currently uncertain if the clinical successes of antiretroviral chemotherapeutics and the development of resistance will be influenced by subtype diversity. Although subtypes C and A are the predominate subtypes worldwide, current antiretrovirals have been designed based on subtype B, which prevails in North America (84). One study has suggested that certain subtype G viruses may be less susceptible to protease inhibitors (85). Moreover, now that antiretroviral therapy is becoming more common in regions with multiple subtypes, recombination events may create new multiple drug resistant viruses (86). Diversity within the nucleotide and resulting amino acid sequences of protease and reverse transcriptase (the proteins that are the targets of most of the current antiretrovirals) for all subtypes is therefore important to delineate and may impact upon patient management (71).

Differences in viral genetics may also impact both disease progression and transmission. One example where viral genetic factors have been shown to play a role in

disease progression is in an Australian study where individuals infected with an HIV-1 virus with a significantly altered nef gene had very slow disease progression (87). Another study, which investigated the virus population at initial infection, found that diversity within this initial population was associated with faster disease progression (88). Although it is well established that HIV-1 is more pathogenic than HIV-2 (89) the role of subtype specific differences on disease progression is contentious as it has revealed discordant results (90;91). Biological differences between subtypes such as their co-receptor usage and ability to form syncytia have been suggested which supports the notion that differences in disease progression between subtypes may exist (91). Some studies have reported differences in disease progression (92-94). One such study looking at individuals in Uganda compared those infected with A and D type viruses and found that infection with subtype D viruses resulted in a more rapid progression to death (92). Furthermore, a cross-sectional study found individuals infected with subtype C viruses to have the lowest CD4 counts compared to those infected with A or D subtypes (94). Other investigations have suggested that there are no differences in disease progression between subtypes (95-97). One such study looked at individuals in Sweden whose ethnicity was Swede or African and found no difference in the rate of CD4 decline or in clinical progression from first positive date among individuals infected with A, B, C or D viruses. However, some of the individuals in this study were receiving therapy (95). In summary, the body of literature on subtypes and disease progression is difficult to interpret as a whole. This is largely due to the different groups of individuals studied, differing access to treatment and varying strategies of measuring disease progression. Moreover, different subtypes may cause differing rates of disease progression within populations of different

genetic backgrounds. It is prudent to continue to try to discern differences in disease progression in individuals infected with varying subtypes. The influence of viral genetics on transmission also remains uncertain as studies have not shown consistent results (98-100).

Changes in viral distribution

Understanding viral genetics is critical for effective vaccine development and disease treatment. Viral genetics may play a role in disease progression and virus transmission; therefore, it remains important to monitor changes in the prevalence of different viral forms within geographical regions. Within some populations where HIV infection has been well established, the subtype distribution has not remained constant. Minor variants may increase, new genetic forms may be introduced or recombinants created and spread, resulting in alterations of the subtype distribution (70). In Brazil for example, where subtype B predominates, subtype C has been increasing in the south and is now the predominate form in some regions (101). An increase in subtype C has also been seen in Kinshasa (Democratic Republic of Congo) between 1997 and 2002 (102). Another example of alterations in the genetic forms present involves injection drug users (IDU) in Thailand. In this population subtype B was the founding virus but the CRF01_AE form now dominates (103). There are many explanations for why subtype distributions do not remain static. Within a population different genetic forms may be unequal in their ability to be transmitted or to cause disease. Although debatable, it has been suggested that subtype C may result in slower clinical progression than subtype B. This may result in a longer period of time for the virus to be transmitted (101) . Over time

some virus types may adapt to a population's HLA composition (58) possibly creating fluctuations in the viruses that are present. Another scenario which could explain changes in viral distributions is that there may be a dramatic increase in new infections that occurs within a group that was seeded with a different genetic form.

Host-Virus Relationship:

In addition to the genetics of HIV, investigations have also advanced our understanding of aspects of the host-virus relationship including the modes of transmission, the subsequent immune response and the eventual clinical course.

HIV-1 Transmission:

It is now well established that the most common mode of HIV transmission is through sexual contact (heterosexual and homosexual). Intravascular inoculation with infected blood or blood products is the second most common route of transmission and can occur as a result of injection drug use or through contaminated transfusions. Lastly, mother to child transmission can happen *in utero*, during the birthing process or through breast milk (104).

The Natural History of HIV Infection:

Primary HIV infection occurs following the transmission event, when HIV from one individual first infects the cells of another individual. The cells that are first infected are likely CD4+ T cells or macrophages but it is the dendritic cells that bind to virions and move with them into local lymph nodes. The virus can then infect CD4+ cells within

the lymph nodes, resulting in viral replication that can be detected within days after the initial exposure (105). A fraction of newly infected individuals have an acute influenza like illness. Viral loads at this time are often greater than a million copies per milliliter and a decrease in CD4+T cells is seen in the blood (106). Recent studies suggest that infected individuals have extensive loss of CD4+ T cells in this acute phase but that this occurs primarily at mucosal sites (104;107). Virus spread into peripheral lymphoid tissues occurs and an adaptive immune response develops which is able to partially control viral replication. Approximately 3 months after the initial exposure a person's viral load generally declines to very low levels and can only be detected by sensitive assays. A second phase called the clinical latency period or the chronic asymptomatic phase begins in which chronic infection is established and no clinical signs and symptoms are seen. During this phase low level virus replication and relatively stable CD4+T cell counts are seen within the blood (106). Over the course of years there is a progressive decline in CD4+ T cells within both lymphoid tissues and in the circulation. Progression to the advanced stage of disease, or AIDS, occurs when circulating CD4+ T cell levels fall below 200 cells/mm^3. These individuals are susceptible to acquiring opportunistic infections. Virus levels may increase significantly, and death usually occurs within five years (105).

Disease Progression:

The period of time between initial infection and the development of AIDS can vary significantly between individuals. Time to AIDS without treatment averages 10-11 years in industrialized countries (106) but is faster in developing nations (104;107;108).

Some individuals progress to AIDS dramatically faster or slower than the mean. Those who progress very quickly to AIDS are referred to as rapid progressors and are individuals whose CD4+ T cell counts drop to below 200 cells/mm^3 within 2-3 years following seroconversion (109;110). Conversely long term non-progressors have also been reported. Although definitions vary, typical features include survival for long periods of time (>7 years) with CD4+ T cell counts that remain above 400-500 cells/mm^3 (111-113). Variables influencing an individual's rate of progression to AIDS are not completely understood but are likely multifactoral. This variation in disease progression may be due to factors which include host genetics (114;115), viral genetics (87), nutrition and stress (13;116;116;117), age at time of infection (118) as well as the impact of other infections (119). For those individuals who have access, antiretroviral treatment significantly alters disease progression (6).

The Immune Response to Infection:

The immune system reacts to foreign pathogens with both innate and adaptive immune responses. Once physical barriers are breached the innate defense mechanisms targeted against foreign pathogens are already in place and respond very quickly. This response includes the recognition of molecular structures common to pathogens and involves both cellular and biochemical mechanisms. The adaptive immune response needs time to develop in response to an infection and, unlike the innate response, is highly antigen specific and has classical immune memory (120;121) . The adaptive response includes both humoral (B cell) and cell-mediated (T cell) responses. T helper cells (CD4+ T cells) are key players in the adaptive response and can be thought of as the

conductors of an immune system cell orchestra. Following recognition of processed foreign antigens bound to major histocompatibility complex (MHC) class II molecules, T helper cells can function to facilitate the production of antibodies and the recruitment and stimulation of effector cells including CD8+ cytotoxic T lymphocytes (CTL) (122).

Humoral Immune Response:

Antibody responses against viral antigens can be seen one to three months following infection (120;121;123). Much of this response however may be directed at 'virion debris' (nonphysiological forms of envelope) and so many antibodies may be functionally inefficient (124;125). Antibodies that can aid in the lysis of infected cells are also present in infected individuals (126;127). However, the role that antibody-dependent cell mediated cytotoxicity plays in preventing infection and in disease progression remains equivocal (128-131). *In vitro* studies have shown that some antibodies are capable of binding to and neutralizing HIV variants, thus inhibiting the infection of cells (37;132). Neutralizing antibodies (nAb) are present at low titers in almost all HIV infected individuals. They generally target envelope, and have been shown to exert selective immune pressure in infected patients (133). This being so, escape from nAbs occurs very rapidly after infection (37;134). Furthermore, most antibodies are non-neutralizing and direct evidence that *in vitro* neutralizing antibodies can suppress viral replication in infected individuals has been lacking (133;135-137). There has also been no correlation found between the titers of these antibodies and clinical progression. In addition, the development of a neutralizing antibody response does not correlate with the initial partial containment of the virus and so it is doubtful that antibodies play much of a

role in early virus control in patients (138;139). Given this information the potential for nAbs as inhibitors of infection may seem dismal, but other evidence suggests the contrary. A recent human trial provides promising findings that viral control among infected individuals using neutralizing antibodies might be possible. The study involved the administration of a high dose combination of three neutralizing antibodies after antiretroviral therapy was stopped and found that viral rebound was slowed (140;141). Several studies involving passive antibody transfer, at very high doses, in the macaque model and infection with SIV or an HIV/SIV hybrid (SHIV) have shown that preexisting neutralizing antibodies can mitigate or even prevent viral infection (132;140-144). These neutralizing antibodies have garnered much attention in the scientific literature. Evidence that suggests nAbs can protect against HIV underscores the importance they may have in developing a strategy to prevent infection (145).

Cell Mediated Immunity:

The cell mediated immune response against infection depends upon the presentation of foreign antigens in combination with host proteins, which are encoded by genes in a locus referred to as the major histocompatibility complex (MHC). The MHC has over 200 genes which are highly polymorphic and in humans, are referred to as human leukocyte antigen (HLA) genes (146). There are three main MHC class I molecules (HLA-A, HLA-B and HLA-C) and three classical MHC class II molecules (HLA-DR, HLA-DP and HLA-DQ). They function to bind processed antigen for presentation on cell surfaces and recognition by antigen specific T cells. These antigen specific T cells can only recognize a particular peptide when it is associated with a

specific HLA molecule. Within the peptide being presented there are residues that are important for binding to the MHC molecule and also those that are important for T cell recognition (121).

The class II MHC molecules bind exogenous antigens 10-30 amino acids (aa) in length (12-16 is optimal) and are present mainly on specialized antigen presenting cells including macrophages, dendritic cells and B lymphocytes (146). CD4+ helper T cells are capable of recognizing peptides that are bound to these MHC class II molecules. Recognition of an MHC class II bound viral antigen by virus specific CD4+ T cells results in the production of cytokines which are critical in aiding both the cell mediated and humoral responses (123). A correlation can be seen between the ability of T helper cells to proliferate and produce cytokines and the disease progression of infected humans or monkeys (147;148). Loss of CD4+T cells in HIV infected patients ultimately occurs although the cause of this decline remains controversial (40).

In contrast to the exogenous antigens bound by MHC class II molecules, MHC class I molecules bind endogenous antigens that are 8-11 aa in length (146). CD8+ cytotoxic T lymphocytes (CTL) recognize MHC class I bound viral antigens and results in a CTL response. CTLs use a number of different mechanisms to inhibit virus replication. They are able to lyse HIV infected cells through the release of proteases and perforin and thereby prevent further replication of the virus. (149-152). Moreover, interaction between Fas ligand on the CTL and Fas molecules on infected cell can result in apoptosis of the infected cell (153). Following interaction with an antigen CTLs are also able to produce cytokines that have an antiviral function, such as gamma interferon (IFN-γ) and tumor necrosis factor alpha (TNF-α) (154). Production of other soluble

factors include β-chemokines MIP-1α, MIP-1β and RANTES which are all CCR5 ligands that can function to reduce HIV replication by preventing the infection of CD4+ cells by R5 HIV-1 viruses (152;152;155;155;156;156). The production of additional soluble factors may also play a role in controlling viral replication (150;157).

The CTL response targets both structural and regulatory HIV proteins (158-160). It takes 2-3 weeks for this response against HIV to be detected and it reaches a zenith within 9-12 weeks of acute infection (105) . This peak CTL response is temporally correlated with control of viremia. Subsequently, virus levels and the CTL response are inversely related, lending support to the importance of CTL in controlling viral replication (161-163). *In vitro* studies have also shown that viral replication in patients' CD4+ T cells can be prevented by autologous CTL (151). Further convincing evidence for the role of CTL in controlling HIV replication comes from animal studies using SIV infected rhesus monkeys. Monoclonal antibodies against CD8 used to deplete CD8+ CTL resulted in a reproducible increase in viral replication (164;165). When CD8+ T cells were depleted for more than 28 days initial virus replication went uncontrolled and the animals subsequently died (164). It has also been shown that rhesus monkeys vaccinated to produce CTL responses had less severe disease when challenged with SIV or SHIV than did those who were unvaccinated (166-168). Studies on HIV resistant sex workers have shown that HIV specific CTL responses may play a key role in their protection, further highlighting the importance of CTL (169-171). The development and selection of CD8+ T cell escape mutant viruses in patients, as described below, lends even further credence to the role of CTL in controlling HIV infection (172-175).

Immune Evasion:

Despite the immune response elicited by HIV, ultimately replication of the virus cannot be halted. There are many reasons for why immune control ultimately does not occur but one of the most well documented reasons is immune escape. Immune escape variants can arise and can soon predominate due to mutations that occur which alter epitope regions important for recognition by the immune system. Both the humoral and cellular immune responses are responsible for exerting selective pressure (138). Escape from CTL responses is known to occur, and evidence linking this escape to disease progression has been reported (172;173;175;176). Escape mutants can be formed by amino acid (aa) mutations within or at sites flanking an epitope that are important for the processing or presentation of the epitope, or at sites that are needed for recognition by the CTL (138). Viral escape variants that evade neutralizing antibodies can also arise in patients and seem to occur soon after infection (134). HIV has developed various ways of evading these antibodies; in addition to antigenic variation, alterations in envelope glycosylation and changes in the conformation of the envelope glycoproteins also play a role (137;177). It is unclear however if there is any connection between disease progression and escape from neutralizing antibodies (138).

There are other ways in which HIV may evade the immune response including the ability of the Nef protein to down regulate some of the HLA class I molecules (HLA-A and HLA-B) on the surface of infected cells. This down regulation can then lead to avoidance of the CTL response (178-180). Some of the other contributing factors to immune evasion include provirus latency, upregulation of FasL, sequestration reservoirs and switching of co-receptor usage from CCR5 to CXCR4 such that CC chemokines can

no longer assist in controlling infection (181). The impairment and loss of CD4+ T cells, which are important for both cellular and humoral responses, may ultimately result in an immune system too compromised to control the infection (105).

Superinfection:

Studying HIV-1 superinfection and the rates of such infections, is another means by which the effectiveness of the HIV elicited immune response can be explored. Superinfection may have implications for vaccine development as well as public health consequences (182;183).

What is Superinfection? :

Individuals may be dually infected with more than one HIV-1 virus. If this is due to the transmission of more than one virus directly from a dually infected host, or by the acquisition of a second infection shortly after the first but prior to the development of an established immune response this can be referred to as HIV-1 co-infection. If infection with a second virus occurs after an immune response to the first infection has already been established this event is referred to as superinfection. HIV-1 superinfection can thus be defined as the reinfection of an HIV-1 positive individual with another HIV-1 virus following an already established primary infection (184-186). Individuals infected with more than one HIV-1 virus have been well documented but until recently it was thought that this was only the result of co-infections, as described above, and that HIV-1 superinfection did not occur in humans (182).

Superinfection has been the topic of much debate over the past few years.

Evidence that superinfection was possible was first demonstrated in a chimpanzee model in 1987 (187). Human cases of superinfection have only been recently reported (184). These cases have fueled questions regarding the potential ability to develop a cross-protective vaccine and have raised public health concerns among those infected with HIV-1 (182;182;182;183;183;183). The first reported cases of superinfection involved injection drug users (IDUs) from Thailand. Ramos et al. reported two cases of superinfection in IDUs and provided evidence of superinfection via injection (185). Thoughts that perhaps direct intravenous (IV) inoculation was the only way in which superinfection could occur were soon dismissed by reports of a second case which demonstrated superinfection via sexual transmission in a homosexual male who was on highly active antiretroviral therapy (HAART). This individual was initially infected with an AE type virus, was on therapy, and had a viral load that was nearly undetectable. During an interruption in treatment he traveled to Brazil and had a rebound in viral load which was found to be due to superinfection with a B type virus. He also became infected with hepatitis C at this time (184). A third case demonstrating intrasubtype superinfection involved a homosexual male infected with a B type virus and who was treated with an antiretroviral drug regimen soon after becoming infected. The treatment effectively controlled his HIV levels. He was later placed on a 'stop-start' treatment regimen in which treatment is deliberately stopped and then started again when viral loads increase. A rapid increase in viral load during a treatment interruption resulted in further investigation which identified a B type virus that was different from the original infecting virus. This was felt to represent a new infection (188). The possibility that this virus may have been present from the start however cannot be completely excluded

(189). Additional cases of superinfection have also been described (190-194). It appears obvious that superinfection is not confined to a particular risk group or to a single route of transmission.

The Superinfection Debate: Does HIV-1 Infection Confer Protection?

Evidently superinfection can occur among HIV infected individuals and immune responses against HIV-1 are not fully protective against superinfection with inter- or intrasubtype HIV-1 infection in all cases. What is perhaps a more important question is how common superinfection is and if infection with HIV-1 provides any protection against superinfection. There are clearly two sides to this debate. One hypothesis is that HIV-1 provides protection against HIV-1 superinfection. The belief that the immune response generated by an initial infection may protect against superinfection is supported by the fact that the reduction in viral load seen shortly after acute infection has been attributed to a vigorous virus specific CTL immune response (161;162;164). One study published in 1995 examined the protective effect of HIV-2 in a group of commercial sex workers found that although those infected with HIV-2 had a higher incidence of infection with other sexually transmitted diseases, HIV-2 infection provided approximately 70% protection against infection with HIV-1 (195). This lends support to the theory that infection with HIV-1 will provide protection against HIV-1 superinfection.

Experimental animal models provide evidence that superinfection can occur but also substantiate the hypothesis that an initial infection may provide protection against superinfection (196-200). Shibata et al. found that HIV-1 naïve chimpanzees were easily

infected when challenged with HIV-1 DH12 (subtype B) but that those animals who were previously infected with HIV-1 IIIB resisted infection despite repeated challenge (198). Another study which used macaques and HIV-2 showed that infection with a second HIV-2 subtype was not possible 8 weeks after the initial infection suggesting that there may be a narrow window of opportunity during which a second infection can occur. This could partly explain the existence of dual infections but supports the idea of protection from challenge with a second virus after an established immune response (197).

The opposing side to this debate is the hypothesis that HIV-1 infection does not provide any protection against superinfection. The superinfection cases that have been seen are thought by some to be the "tip of the iceberg" and that superinfection is in fact rampant (183). Certain animal studies have also provided support for this belief. Superinfection has been reported in cats (feline immunodeficiency (FIV)), macaques (SIV) and in chimpanzees (HIV-1). Fultz suggests that protection will not be provided against superinfection in humans based on his studies of HIV-1 infected chimpanzees who typically have uncompromised immune systems and yet can still become superinfected (201).

The concept of superinfection is not novel to HIV infection and Levy draws our attention to evidence from other viral infections that have gone relatively unacknowledged in the discussions surrounding HIV-1 superinfection (202). For example, individuals may be simultaneously infected by more than one strain of viruses including herpesvirus (203), human papillomaviruses (204) and Epstein-Barr virus (205). It has further been suggested that the high prevalence of recombinant viruses are the ultimate evidence that superinfection rates in the population are high (182;206;207).

Whether the high number of recombinant viruses seen are due to superinfection events or infection from partners who have either dual infections or recombinant viruses remains uncertain (183;208).

There are clearly two opposing hypotheses surrounding the question of whether HIV-1 infection provides protection against HIV-1 superinfection. Reports of individual cases of superinfection have slowly been increasing in the literature but in order to assess whether HIV infected individuals are more likely to ward off a superinfection event than to succumb to one, population based studies are warranted. Several studies have looked for superinfection within specific groups with varying results. Among IDUs one study found no evidence of superinfection after 215 person years of observation (n=37) (209), while another study found an intersubtype superinfection incidence of 2.2 per 100 person years (209;210). Superinfection rates among IDUs are not well established and what conclusions can be drawn from injection drug user superinfection rates and then applied to transmission via sexual exposure is equivocal. Gonzales et al. studied patients from northern California and found no evidence of superinfection despite 1072 person-years of observation. The extent of exposure to HIV within this group however was unknown (211). Another recent investigation looked at 14 couples in which partners had been independently infected with different viral strains and, despite counseling, participated in high risk behaviour. Multiple samples over the course of 1-4 years for each individual were studied but superinfection was not detected (212). Still another study that looked at 20 high risk women has suggested that superinfection may be very common and that an immune response to the initial infection provides little protection (213). It is undisputable that the question of protection provided by initial HIV-1 infection, and superinfection

rates, remains unanswered (207). Large population studies in untreated, well characterized cohorts with data on HIV-1 incidence and exposure rates, with an unbiased selection of individuals investigated, are needed to accurately assess the question of protection provided by HIV-1 infection against superinfection.

Clinical Consequences of Superinfection and Dual Infections:

As for superinfection rates and the questionable protection provided by an initial HIV infection, the clinical consequences of both superinfection and dual infections are also unclear. We might predict that an increase in disease progression could occur in the context of dual infections or following a superinfection event since infection with multiple strains could allow for a more diverse range of viral epitopes potentially able to escape the immune response (86). With respect to superinfection the very occurrence of this event dictates that the immune response to the first infection could not prevent the second infection, so an increased rate of disease progression might be expected (183). It is possible that only viruses more pathogenic than the initial virus are capable of superinfecting. Investigation into this topic is complicated by the debate that a decrease in CD4+ T cell counts might be caused by HIV-1 superinfection but may also account for an individual's susceptibility to such an event. Additionally, some individuals may be predisposed to acquiring a dual infection or becoming superinfected and those same individuals might also be predisposed to a particular type of disease progression. Animal studies do indicate that experimental superinfection in monkeys does not necessarily lead to rapid disease progression (198;214;215). One study in macaques infected with HIV-2 and then superinfected with SIV even revealed that the superinfected animals progressed

more slowly than those solely infected with SIV (216). The only way to truly begin to unravel the complexities of HIV disease progression in humans who are dually or superinfected is to continue to study the human cases that have been identified.

Unfortunately, reported human superinfection cases are few and investigations into the impact of such infections on disease progression are even more limited. Despite a larger number of reports of dual infection cases, the clinical consequences of such infections have also been poorly addressed. In many of the reported human HIV-1 superinfection and dual/multiple infection cases, an increase in clinical progression has been described. In these studies measurements of increased progression include an increase in viral load, high viral loads, or a decrease in CD4+ T cell counts (186;217;218). One of the largest and most recent studies to look at the consequences of these infections on disease progression describes four cases of intrasubtype dual infection and one case of same subtype superinfection. Viral loads in these individuals were only slightly higher compared to controls. All individuals had a time from seroconversion to an AIDS defining event of less than 3.4 years and to a CD4 count below 200 cells/ mm^3 of less than 3.1 years, which was much faster than the control group who were infected with only one virus type. However, the CD4+ T cell decline in the superinfected individual occurred prior to the superinfection event, indicating that the drop may not be a consequence of superinfection but was what resulted in susceptibility to it (186). Jost et al. described a case of superinfection in which an individual had a superinfection event and a drop in CD4+ T cells (of 300 cells/ mm^3) that occurred within four months of detection of the second virus, in addition to an increase in viral load. This suggests that these changes are a direct consequence of the superinfection event. It should be noted that

this individual's immune response to HIV may have been limited by early initiation of HAART (184).

Additional superinfection and dual infection cases have also been described which report an increase in clinical progression (218). It should be mentioned however that many of the described superinfection cases are complicated by patients receiving therapy early in infection or by interruptions in that therapy. These factors may have influenced both their clinical progression and the generation of HIV-1 specific immune responses. Several of the currently reported multiple infection and superinfection cases may also have been identified due to a change in a patient's clinical status or a notable rapid progression from infection and so there may be some bias in determining the consequences of such infections. Larger and more rigorously structured studies investigating both superinfection and dual infections are needed so that we may more completely understand the impact of these events as they pertain to clinical outcome and disease progression.

Hypotheses and Objectives

Genetic diversity is a hallmark of HIV-1 and impacts vaccine development, the potential effectiveness of current treatment and the design of new drugs. An understanding of viral genetic diversity and the virus-host relationship is critical to the development of strategies to halt the pandemic.

A cohort of drug-naïve female commercial sex workers referred to as the ML cohort, from the Pumwani slum in Nairobi, Kenya has been the subject of investigations into numerous sexually transmitted infections. With a high HIV prevalence from the start

of the epidemic they became the focus of many studies exploring susceptibility to HIV infection and the HIV specific immune response.

This thesis describes how the sex workers were studied, at a population level, in order to investigate HIV-1 viral genetics and the host-virus relationship relevant to HIV-1 single virus infections, HIV-1 superinfection (defined as the acquisition of a second HIV-1 virus following the generation of an immune response to the first) and HIV-1 dual infections. The investigations reported within this thesis serve to advance our understanding of the virological, clinical and epidemiological implications of these infections.

Hypotheses:

There were three global hypotheses upon which the research reported here was based:

I.) That an investigation of HIV-1 genetics within the sex workers that have single HIV infections will reveal the following: **A.) a diverse array of genetically distinct viruses, B.) a dynamic distribution of viral subtypes, C.) viral genetic factors associated with differences in disease progression**

II.) Infection with HIV-1 will provide protection against HIV-1 superinfection

III.) The majority of individuals infected with more than one HIV-1 virus (superinfections or co-infections) will have evidence of recombinant viruses and they will also have a fast rate of disease progression.

Rationale, Objectives, Experimental Approach:

Section I:

Section I hypothesis: **That an investigation of viral genetics, within the sex workers that have single HIV infections, will reveal the following: A.) a diverse array of genetically distinct viruses, B.) a dynamic and heterogeneous distribution of viral subtypes, C.) viral genetic factors associated with differences in disease progression**

Addressing the initial hypothesis allowed for an understanding of single HIV infections within the cohort prior to engaging into the investigation of more complex HIV-1 infections including multiple subtype co-infections and superinfection events (which will be detailed in sections 2 and 3). In order to test the first hypothesis and achieve an understanding of the viruses infecting the women within this cohort, as well as the potential impact on disease progression, I addressed parts A-C separately as follows:

Section I A.) hypothesis rationale:

In the first hypothesis (part A) I predicted that a **diverse array of genetically distinct viruses** would be detected within the ML cohort. Reports of multiple circulating viruses within Kenya and surrounding countries exist and I had no reason to suspect that the ML cohort would be an exception to this observation. I also suspected that HIV co-infections and the spread of recombinant viruses would result in frequent identification of recombinant forms within the cohort. Further to this I predicted that naturally occurring

polymorphisms, which have been associated with protease inhibitor (PI) resistance in subtype B viruses, would be seen in the HIV-1 protease of non-B subtypes. In order to test the hypothesis and develop a molecular profile of the single HIV infections in the women within the cohort I set out the following **objectives:**

- To elucidate the subtypes infecting individuals within the cohort
- To determine if recombinant viruses are common
- To evaluate the baseline diversity (naturally occurring polymorphisms associated with drug resistance) within the HIV-1 protease gene

Section I A.) experimental approach:

The experimental approach included sequencing of a region of the HIV-1 genome of HIV infected women within the cohort. Recombination screening and phylogenetic analysis of the data was done to determine the viral subtypes and recombinant viruses present. To further test for recombination additional regions of the genome were analyzed for subsets of the women initially tested. Diversity within HIV-1 protease (one of the key targets of antiretrovirals) was assessed by comparison of the amino acid sequences of different viral subtypes to subtype B data. This was done to determine if naturally occurring polymorphisms associated with PI resistance were present.

Section I B.) hypothesis rationale:

In the next part of the first hypothesis I predicted that a **dynamic and heterogeneous distribution of viral subtypes would be seen**. In other words, that the

viral distribution will have changed over time and that the viral distribution seen will differ between distinct groups of individuals within the cohort. I suspected that many possible factors including the influx of different subtypes from surrounding regions as well as potential differences in the transmission and pathogenesis of different viral subtypes would result in changes in the viral subtype distribution over time. The ML cohort is also composed of a heterogeneous group of women with respect to both disease progression (108) and susceptibility to infection (108;219). There is a unique group of individuals within the cohort who have been termed "resistant" to infection. A resistant individual is defined as anyone who has been HIV negative for at least three years after entry into the cohort, despite exposure to HIV (219). Unfortunately this resistance is not necessarily life-long. Some of these women have become infected with HIV and are thus called "late seroconverters". The reasons behind this apparent resistance have not been fully elucidated but one theory supported by immunological studies is that they have a specific immune response to HIV due to continued exposure (170;220). I predicted that the late seroconverters would have a unique subtype profile due to a less effective immune response against and thus preferential infection with less common viral subtypes. To evaluate the hypothesis that a dynamic and heterogeneous distribution of viral subtypes would be seen, I proposed the following **objectives**:

- To determine fluctuations in subtype distribution over time
- To determine if a unique group, the late seroconverters, have a unique HIV-1 subtype molecular profile compared to the rest of the cohort

Section I B.) experimental approach:

Since the cohort was established 20 years ago, I was presented with the unique opportunity to determine fluctuations in viral distributions as the epidemic evolved. Subtype distribution within three separate time periods 1985-1989, 1990-1995, 1996-2003 was assessed. Sequence based subtyping of the late seroconverters was also conducted and compared to data from the rest of the women.

Section I C.) hypothesis rationale:

I hypothesized that I would find **viral genetic associated differences in disease progression.** As is more thoroughly discussed within the introduction, reasons for the variablility seen in disease progression are likely multifactoral but evidence suggests that differences in viral genetics are a contributing factor. HIV-2 is known to be less pathogenic than HIV-1 (89), and viruses with altered nef genes have been shown to result in long term non-progression, to name a couple of examples (87). I therefore predicted that viral genetic factors capable of influencing disease progression would be detected. This last component of the initial hypothesis guided the next **objective:**

- To determine if viral subtypes or potentially deleterious genetic mutations are associated with differences in disease progression

Section I C.) experimental approach:

To test this hypothesis I first assessed the impact of HIV-1 subtype on disease progression. Next, I attempted to determine if subtype impact on disease progression

might change over time. Lastly, I investigated an individual case that illustrates the potential impact of viral genetics on disease outcome. Given that all of the women included in this study were treatment naïve, and that treatment has only recently become available it was not possible to assess the impact of drug resistance associated mutations on treatment success and disease progression. This could be the focus of future investigations.

Section II

Section II hypothesis: **Infection with HIV-1 will provide protection against HIV-1 superinfection**

Section II hypothesis rationale:

Multiple circulating viruses, frequent exposure to HIV and well defined data on HIV incidence within the cohort make the ML cohort an ideal one to study superinfection. As was previously discussed, there is much debate surrounding whether or not infection with HIV-1 provides protection against superinfection. Evidence of a vigorous immune response against HIV and the ability of this immune response to control initial viral replication (161) supports the notion that HIV-1 infection will provide protection against superinfection. Additional studies, as discussed within the introduction, such as one which suggests that HIV-2 can provide protection against HIV-1 infection lend further support to this hypothesis (195). I predicted that although dual infections, prior to the establishment of an immune response (co-infections) may be common, that

superinfection events are rare. This would explain the high proportion of recombinant viruses seen and support a protective effect of HIV-1 infection against superinfection. In order to test the hypothesis that infection with HIV-1 will provide protection against HIV-1 superinfection I derived the following **objective:**

- To determine the incidence of superinfection and to compare this to the incidence of initial HIV-1 infections within commercial sex workers in the ML cohort.

Section II experimental approach:

The experimental approach involved a nested cohort study in which ML women were investigated over time for evidence of superinfection, based on viral sequencing at multiple time points. The data was analyzed using multiple Poisson regression and Cox's proportional hazard modeling to determine if a protective effect was evident.

Section III

Section III hypothesis: **The majority of individuals infected with more than one HIV-1 virus (superinfections or co-infections) will have evidence of recombinant viruses and they will also have a fast rate of disease progression.**

Section III hypothesis rationale:

The third hypothesis predicts that recombinant viruses will be present and that increased disease progression will be observed in individuals with confirmed infection

with more than one HIV-1 virus (superinfections or co-infections). Infection of individuals with more than one virus is required for recombinant viruses to be formed, either through superinfection or co-infection of individuals. I therefore predicted that, due to the existence and frequency of recombinant viruses within regions where multiple subtypes circulate that such forms would arise in most cases. The hypothesis that an accelerated clinical progression would be observed was based on the fact that a more diverse array of epitopes may allow for the increased possibility of immune evasion and subsequently disease progression. Alternatively, individuals who are able to be infected with multiple viruses may be more prone to rapid disease progression. The next **objective** was designed to test this third hypothesis:

- To analyze the viruses circulating in individuals with more complicated infections (including all confirmed superinfection cases and other identified dual infections) and to examine the clinical profiles of individuals with such infections

Section III experimental approach:

The experimental approach included an investigation into all superinfection/dual infection cases identified. To determine what viruses were present and at what proportions, viral clones from each individual at all available time points were analyzed. Assessment of disease progression was carried out using two approaches: i) case control studies examining CD4+ T cell data and ii) individual case descriptions of clinical disease progression.

MATERIALS AND METHODS

Materials

Subjects/ Biologic Samples:

Cohort:

A group of women who are commercial sex workers in the Pumwani slum of Nairobi, Kenya were the subjects of this study and form the Pumwani sex worker cohort (ML cohort). The cohort was established in 1985 to study the epidemiology and biology of a variety of sexually transmitted infections (STIs). To date the cohort includes over 2200 women, of whom approximately 60% were HIV seropositive at enrollment. The number of active participants at any particular time from whom samples can be taken is approximately 500 women. Studies using this cohort have been approved by the National AIDS Committee, the National Ethics and Scientific Review Committee of Kenyatta National Hospital and the University of Manitoba, Use of Human Subjects in Research review committee.

Enrollment:

Enrollment of women in the cohort includes informed consent, a clinical examination, completion of a questionnaire/interview and the assignment of a unique ML number and identification card. The questionnaire provides information on demographic data, general medical history, sexual practices, vaccinations and reproductive and contraceptive history. Blood is also drawn for HIV-1 serology and polymerase chain reaction (PCR) as well as for lymphocyte counts. These women are asked to return to the

Majengo STI clinic, located in the Pumwani slum, at least twice yearly during resurvey periods and whenever they need medical attention. Free clinical services and HIV prevention programs are provided.

Samples:

When the sex workers visit the clinic biological samples are obtained for research purposes. Peripheral blood mononuclear cells (PBMCs) and plasma samples from these women were used in this study.

Reagents:

See Appendix 1

Methods:

General Procedures:

Isolation of peripheral blood mononuclear cells (PBMCs) and plasma (Nairobi lab):

Peripheral blood samples were taken at the Pumwani clinic and were drawn into sodium heparin vacutainers (Becton Dickinson). PBMCs were isolated by layering the whole blood onto a one-third volume of Ficoll-Hypaque (Sigma) followed by centrifugation at 500 x gravity for 30 minutes at room temperature using a table top centrifuge. The plasma layer was aseptically transferred to sterile tubes and was used for HIV-1 serologic testing with cryopreservation of the remaining plasma. The PBMCs were transferred to a separate tube where they were washed with 50 milliliters (mL) of phosphate buffered saline (PBS) and centrifuged at 400 x gravity for 20 minutes. This

wash was done twice and then cells were stained with trypan blue and counted with use of a hemocytometer. Cells were then cryopreserved at -70°C (freezing media: 90% fetal calf serum, 10% dimethylsulfoxide) in aliquots of 5 x10^6 cells/mL.

HIV-1 serology:

Commercial enzyme immunoassays (EIA) (1985-1988: HTLV-III Elisa (Du pont); 1988-1990: Vironstika (Oranon Technika); 1990-1992: Detect HIV (IAF Biochem); 1991-onward: Enzygnost HIV-1/2 EIA (Behring)) were used for the detection of HIV-1 antibodies. Seroconversions were confirmed with the use of immunoblots prior to 1991 (Novapath Immunoblot, Bio-Rad) and then from 1991-onward a confirmatory immunoassay was used (Recombigen HIV-1/2 EIA (Cambridge Biotech)). These procedures were performed in the Nairobi, Kenya labs following the manufacturers' instructions.

CD4/CD8 counts:

Lymphocyte subsets (CD4+ and CD8+ T cell counts) were determined by fluorescence-activated cell sorting as previously described (221). CD4+ and CD8+ T cell counts are described as the number of cells per millimeter cubed (mm^3) of blood.

Viral loads:

Viral loads were performed by the National HIV-1 Retrovirology Labs (Health Canada) using the Organon Teknika Nuclisens HIV-1 QT method. This method was chosen because it can be used with samples that contain heparin.

Preparation of Template:

DNA isolation from PBMCs:

DNA was isolated from PBMC samples using QIAamp DNA Mini kits (Qiagen). Approximately $5x10^6$ cells were pelleted and resuspended in 200 microliters (μL) of PBS. 20 μL of Proteinase K (920 mg/mL) was added in order to lyse cells and DNA isolations were performed according to the manufacturer's instructions. The final product consisted of the DNA in 200 μL of a commercial buffer. Samples were stored at -20°C. All samples used in this study were DNA samples isolated from PBMCs unless otherwise stated.

RNA isolation from plasma:

Plasma was clarified by centrifugation at 400 x gravity for 10 minutes (room temperature) and then transferred to a sterile 1.5mL tube. Virions were collected by centrifugation at 20,000 x gravity in a table top centrifuge for 2 hours at 4°C, the supernatant was removed and the virions resuspended in 140 μL of PBS (sterile, RNase free). RNA was then isolated using QIAamp Viral RNA Mini Kit (Qiagen). Isolations were performed in accordance with the manufacturer's instructions. RNA was stored in 10 μl aliquots at -70°C.

Reverse transcriptase polymerase chain reaction (RT-PCR):

Since some samples were collected with heparin which is known to inhibit enzyme reactions, they were treated with a heparinase solution (HS) at an HS to sample ratio of 12:1 [HS: 100 μL heparin buffer (Sigma), 10 μL Hep I solution, (Sigma) 20 μL

RNAse out (Invitrogen)] and then incubated at room temperature for 2 hours prior to the RT-PCR step.

RT-PCR was performed using the Superscript III First Strand Synthesis System™ for RT-PCR kit (Invitrogen) with random hexamers. This was done using the kit as per the manufacturer's instructions with the maximum amount of RNA used per reaction. The RT-PCR step was done multiple times for each sample. The cDNA product was then ready for HIV PCR.

PCR (General Protocol):

DNA and RNA isolations were done in a separate amplicon free lab. In order to avoid contamination, first and second round PCR reactions were also performed in different locations with the use of separate pipetters and aerosol resistant tips. When performing HIV PCR a double distilled water (ddH_2O) control and an HIV negative DNA sample were used as negative controls. A positive control which consists of DNA isolated from cultured HIV-1 subtype B infected cells was also used in each PCR run. If the procedure began with the RNA or DNA isolation, negative and positive controls from the point of isolation were also included when PCR was first performed on the new set of samples. HLA PCR involved the use of ddH_2O as the negative control and DNA from an HIV negative individual for the positive control.

If at any time there was concern that laboratory error may have occurred, resulting in contamination, the entire PCR run was discarded. Additionally, if a false positive occurred in the PCR negative controls or a false negative occurred in the PCR positive controls the entire PCR run was aborted. The procedure recommenced from the point of

the PCR/RT-PCR step. If the set of samples being tested was one that had been newly isolated and the negative control from the beginning of the isolation step was a false positive then the procedure recommenced from the point of DNA/RNA isolation.

Unless otherwise stated PCR reactions included use of the Expand High Fidelity System™ (Roche Applied Science). Reactions included 2.6 units (per 50μL reaction) of the Expand High Fidelity Enzyme. Expand High Fidelity Buffer (10x concentration with 15mM $MgCl_2$) was used for a final concentration of 1x with1.5mM $MgCl_2$. A deoxynucleotide triphospate (dNTP) mixture of 1.25mM for each dNTP (for a final concentration of 200μM each) and a primer mix (final concentration of 0.5μM for each primer) were used along with approximately 5μL of DNA. Reactions were brought up to a volume of 50μL with ddH_2O. In nested PCR reactions, 1μL of the first round product was used in the second round. The cycling reactions were performed on GeneAmp PCR system 9700 or MJ machines.

In order to confirm successful amplification of the correct PCR product 5μL of sample was mixed with 2μL of tracking dye and run along with a 1.0 kB ladder on a 1.2% agarose gel with ethidium bromide (EtBr). Visualization and pictures were taken using the Gel Doc system (Bio-Rad).

DNA/cDNA Quality Testing:

Testing for presence of genomic DNA:

In order to confirm the presence and quality of DNA isolated, PCR reactions amplifying a part of the HLA DQα1 gene were performed (see Table 1a for primers) (222). A 25μL PCR reaction was used and, instead of the high fidelity system, Taq

Polymerase was used instead (1.25 units). 10x PCR buffer, $MgCl_2$, dNTPs and primers were all used with the same final concentrations as described above. Amplification conditions included 30 cycles of denaturing at 94 °C for 1 minute, annealing at 50° C for 1 minute 30 seconds, extension at 72° C for 1 minute followed by 72°C for 7 minutes to complete the program and finally drop to 4°C until removal from the machine. Product was run on an agarose gel and visualized using the Gel Doc (Bio-Rad).

Testing for presence of HIV provirus:

If HLA PCR described above was successful and ample DNA was available, samples were also initially tested for amplification of short regions of the provirus. Primers for a short segment of pol (part of integrase) (223), vif (224) or both were used (see Table 1b for primers). Reaction mixtures were as described for HLA PCR testing except that 2.5 units of Taq and 50μL reactions were used. The amplification program for pol (223) and for vif included 35 cycles of 94°C for 1 minute, 56°C for vif (50°C for pol) for 1 minute, and 72°C for 1 minute. The program is completed by reaching 72°C for 7 minutes and cooling to 4 °C. Nested reactions were done with the use of 1μL from the first reaction. The PCR reaction program was the same for the second round except that 30 instead of 35 cycles were used. (Note: If samples did not amplify they were still tested using the degenerate Gag/Pol primers described below.)

HIV PCR:

HIV PCR primer design:

Primers were needed that were capable of amplifying all possible HIV-1 subtypes. A reference group of HIV-1 sequences aligned in the HIV Sequence Compendium which

is available on the Los Alamos HIV Sequence Database website (http://www.hiv.lanl.gov/content/hiv-db/HTML/compendium.html) were referred to in order to design the best possible primers. Primers were designed in conserved regions of the viral genome in order to avoid differences in the ability to bind to different viral subtypes. The oligo selection program was used to help with the primer selection (225;226). Primers of cartridge purity were ordered from Qiagen and were reconstituted in TE buffer (10mM Tris 1mM EDTA) at a concentration of 50μM (see Table 2a and b for primers).

PCR for Gag/Pol (p1/p6/Protease/partRT) :

Nested degenerate primers were designed to amplify a product approximately 1000 base pairs (bp) in size that includes the p1, p6, the protease gene and part of reverse transcriptase (outer product: 1960-3034, inner product amplified: 2016-3019 (HXB2 numbering)) . Primers were tested and confirmed to be capable of amplifying all major subtype found in Kenya. The Expand High Fidelity PCR System™ (Roche) was utilized as described above. The outer primers were used for the first round PCR and the inner primers for the nested second round. The PCR program consisted of 95°C for 2 minutes followed by 35 cycles for the first round PCR, or 30 cycles for the second round of the following: 95°C for 20 seconds, 48°C for 30 seconds, 72°C for 1 minute 30 seconds. 72°C for 7 minutes and 4°C until removal from the thermocycler completed the program.

Table 1a: DQα Primers: Testing for Presence of Human Genomic DNA (225)

PRIMER NAME	PRIMER SEQUENCE (5'-3')	LOCATION (HUMAN GENOME)
DQαGH26	GTGCTGCAGGTGTAAACTTGTACCAG	DQα locus (2^{nd} exon)
DQαGH27	CACGGATCCGGTAGCAGCGGTAGAGTTG	DQα locus (2^{nd} exon)

Table 1b: Vif and Pol (Integrase): Testing for Presence of Provirus

Inner and outer primers for nested PCR of a small region of HIV-1 pol (part of integrase) (223) and a small region of vif (224). Locations are indicated as position within HXB2 HIV sequence.

PRIMER NAME	PRIMER SEQUENCE (5'-3')	LOCATION (HIV-1 HXB2)
Outer Vif A	CAGATGGCAGGTGATGATTGTG	5049-5070
Outer Vif C	CTCCAGTATGCAGACCCCAATATG	5266-5243
Inner Vif A	ATTGTGTGGCAAGTAGACAGGATGA	5065-5089
Inner Vif C	CTAGTGGGATGTGTACTTCTGAACT	5218-5194
Outer HPOL 4235	CCCTACAATCCCCAAAGTCAAGG	4653-4675
Outer HPOL 4538	TACTGCCCCTTCACCTTTCCA	4976-4956
Inner HPOL 4327	TAAGACAGCAGTACAAATGGCAG	4746-4767
Inner HPOL 4481	GCTGTCCCTGTAATAACCG	4919-4901

Table 2a: Gag/Pol Degenerate Primers Designed to Amplify all Subtypes

PCR PRIMER NAME	PRIMER SEQUENCE 5'-3'	LOCATION (HXB2 NUMBERING)
TS PRO 5' Outer	AAR TGT TTC AAY TGT GGC AAR GAA	1960-1983
TS PRO 3' Outer	AAR GGC TCT AAG ATT BTT GTC ATR CTA C	3061-3034
TS PRO 5' Inner	RAA AAR RGG CTG TTG GAA ATG TG	2016-2038
TS PRO 3' Inner	GGT GAT CCT TTC CAT CCY TGT G	3019-2998

Table 2b: Envelope Degenerate Primers Designed to Amplify all Subtypes

PCR PRIMER NAME	PRIMER SEQUENCE 5'-3'	LOCATION (HXB2 NUMBERING)
TSEnv1 Outer	AAT TCC YAT ACA TTA TTG TGC YCC AGC	6860-6886
TSEnv9 Outer	CGC CCA TMG TGC TTC CTG	7818-7801
TSEnv3 Inner	TGY TRA ATG GYA GTC TAG CAG AAG	7003-7026
TSEnv8 Inner	TTC CTG CTG CTC CYA AGA AC	7806-7787

PCR for Env (v3-end of gp120):

PCR and sequencing of a portion of gp120 including V3 to the end of gp120 was done for a subset of samples. Degenerate PCR primers were designed to allow for amplification of all subtypes. Primers were tested and confirmed to be capable of amplifying all major subtype found in Kenya. The outer PCR product is approximately 1000bp (6860-7818 by HXB2 numbering) and the inner approximately 800-900 bp (7003-7806 by HXB2 numbering). Deletions and insertions affect product size. The PCR cycling conditions for both first and second rounds included 95°C for 2 minutes followed by 30 cycles of 95°C for 20 seconds, 53°C for 30 seconds and 72°C for 1 minute and 30 seconds. This was followed by 72°C for 7 minutes and then 4°C until removal from the machine.

Purification of PCR products:

Purification of PCR products was done using 96 well Montage PCR96 plates (Millipore) with a vacuum system. PCR reactions were loaded into the plate and the vacuum applied. This was followed by a wash with TE buffer. Resuspension of the samples was achieved by addition of ddH_2O and the use of a shaker (10 minutes). Samples were then transferred to a 96 well plate and were ready for sequencing or for storage at -20°C.

Cycle Sequencing:

Sequencing reaction:

ABI PRISM BigDye™ Terminator v3.1 Cycle Sequencing Mix (Applied Biosystems) was employed for sequencing reactions. Sequencing reactions consisted of 3μL of Big Dye version 3.1, 1.5μL of primer (concentration: 3.2pmol/μL) (see Table 3 for sequencing primers) and 2μL of purified PCR product (or approx.150 nanograms (ng)). This was done in a 96 well plate format that was later sealed with a Microseal A Film (Bio-Rad). Cycle sequencing was performed on a GeneAmp or MJ thermocycler using the following program: 96° C for 3 minutes followed by 60 cycles of 96 °C for 30 seconds, [56° C TS PRO A SEQ primer/ 52° C TS PRO B SEQ primer or Melting temperature (Tm)-5°C for other primers] for 30 seconds, 60°C for 4 minutes. Following the 60 cycles the temperature decreases to 4°C until removal from the machine. The direct sequencing approach did not work well for some Gag/Pol products or for the Env products so TOPO TA cloning® followed by sequencing was necessary in those cases as described below.

Table 3: Sequencing Primers for Gag/Pol Product

SEQUENCING PRIMER NAME	PRIMER SEQUENCE 5'-3'	LOCATION
TS PRO A Seq 3'	TAT GGA TTT TCA GGC CCA ATT YTT G	2716-2692
TS PRO B Seq 5'	CCA GRC CAA CAG CCC	2146-2161

(Note: Inner PCR primers were also used as sequencing primers)

Purification and sequencing run:

The sequencing products were precipitated using ethanol/sodium acetate (1μL NaAc pH 4.8-5.2, 20μL 95% EtOH). Samples were left to sit for at least 2 hours with the ethanol/sodium acetate mixture and then centrifuged for one hour at 4000 rpm. After removal of the supernatant and two subsequent 70% ethanol washes, the samples were dried and resuspended in 20μL of formamide. Samples were then denatured by incubation at 95°C for 1.5 minutes and immediate immersion in ice. Samples were transferred to a MicroAmp 96 well sequencer plate (Applied Biosystems) and run on either a 3100 Genetic Analyzer (Applied Biosystems) or a 3730 DNA Analyzer (Applied Biosystems).

TOPO TA Cloning®

If direct sequencing of the purified Gag/Pol PCR products resulted in poor sequence data that was difficult to read or if double peaks were seen at 2% or more of the sites then TOPO TA cloning® (Invitrogen) and sequencing was performed. In these cases, high *in vivo* HIV-1 variability including deletions/insertions in some of the quasispecies present, or infection with multiple HIV-1 viruses (dual infections) were suspected.

Although direct population based sequencing of the env PCR products was first attempted using multiple primers, ultimately TA cloning was deemed a better approach in all cases due to the high amount of diversity within individuals in this region.

Cloning reaction and culturing:

The TOPO TA Cloning® Kit for Sequencing (Invitrogen) with One Shot® TOP10 chemically competent *E.coli* (Invitrogen) was used. Both the first and second round PCR reactions were repeated and once the second round PCR reaction was complete the products were placed on ice. 5μL of PCR product was then run on an EtBr agarose gel to confirm successful amplification. The cloning reaction could then be set up. This was done following the manufacturer's instructions and includes 4μL of fresh PCR product, 1μL of salt solution provided by Invitrogen and 1μL of the TOPO® vector. Following incubation at room temperature for 5 minutes, the reaction was placed on ice. One shot chemical transformation was carried out using the chemically competent *E.coli*. Briefly, this included addition of the cloning reaction to the cells followed by a heat shock step. SOC medium (provided in kit) was added to the tube and samples then shook at 37°C for one hour. For each sample three different amounts were spread onto three different 37 °C preincubated Luria-Bertani (LB) plates (plates contained 200μg/ml of ampicillin and were coated with 40μL of 2% X-gal solution prior to incubation). The plates were inverted and left at 37°C for at least 12 hours. The next day 32-48 creamy white colonies were picked and cultured in LB (with 200μg/mL of ampicillin) overnight in tubes or in a 96 well plate format. Cultures were incubated for 20-24 hours at 37°C with shaking at 300 rpm. Stocks were saved by mixing a 50μL aliquot with equal volume of glycerol and stored at -80°C.

Plasmid isolation:

Cultures were centrifuged to pellet bacteria and media was removed. Each bacterial pellet was then resuspended in Buffer PI (QIAprep Turbo Miniprep Kit (Qiagen)). Plasmids were then isolated using the QIAprep96 Turbo Miniprep Kit and vacuum system as per the manufacturer's protocol. (Note: TA cloning minipreps that were done for envelope PCR products were done without a kit using a labour intensive approach for isolation of the plasmids. The kit based approach replaced this method.)

Testing for the PCR insert was done in one of two ways. The first involved restriction digestion of the plasmids with the *Eco*RI enzyme. Following digestion the samples were run on an EtBr agarose gel to visualize the insert or lack thereof. An alternative method used was the dot-blot approach which can be done prior to plasmid minipreps. A biorobot was used to prepare the blot (up to four 96 well plates at a time). Detection of the insert was via the ECL™ direct nucleic acid labeling and detection system protocol (Amersham Pharmacia biotech) as per the manufacturer's instructions. Both approaches were used successfully to determine if the desired insert was present.

Sequencing of product:

DNA concentrations were determined using a spectrophotometer (Nanodrop). Plasmid specific primers M13 reverse (5'CAG GAA ACA GCT ATG AC) and T7 (5'TAA TAC GAC TCA CTA TAG GG) were used for sequencing. Sequencing was done as described previously with the use of 450 ng of plasmid DNA per reaction and the default Big Dye sequencing program as per the manufacturer's instructions.

Sample Identity Confirmation (HLA based):

HLA PCR and sequencing (for identity confirmation):

Sequence based HLA typing was done to confirm the identity of the samples that were suspected superinfection cases. Initial identity screening was done using the DRB allele. The HLA-B allele was chosen for ultimate screening purposes as it is the most diverse HLA allele. PCR and sequencing protocols were followed as previously described for DRB (227;228) and for HLA-B (229).

HLA typing analysis:

HLA typing analysis was done using a taxonomy based sequence analysis (TBSA) program (230). In brief, after samples of interest were sequenced each codon was compared to an HLA sequence database file. Samples which were labeled with the same ML number but were being tested to rule out sample mislabeling were required to have the same DRB and HLA-B alleles present to be considered a correct match. HLA results also had to correspond with different samples previously HLA typed for that ML in order to confirm that the sample's ML identity was correct.

General HIV Sequence Analysis:

A program called Sequencher™ (Gene Codes Corporation) was used to align sequences in both directions. Sequence clean up was done to clarify discrepancies by visual inspection of peaks and removal of inadequate sequence data. For all population based sequences and all clones a consensus sequence was made from a minimum of two overlapping sequences, one in each direction, and exported as a text file.

Testing for recombination:

Prior to subtyping by phylogenetic analysis, sequence data was subjected to an initial screening process to determine if recombination events were apparent. Testing for possible recombination was done using the RIP analysis program (Los Alamos Database website: http://www.hiv.lanl.gov/content/hiv-db/RIPPER/RIP.html). All population based sequences and every clone were individually screened for recombination using this program. In those cases where recombination was detected, Simplot (231) was used for fine mapping of the breakpoint as described later.

PCR, cloning and sequencing of recombinant viruses was repeated to confirm presence of the recombinant form. (Limiting dilution was performed if recombination during PCR was suspected.)

Subtyping analysis:

For initial rapid subtype screening the National Center for Biotechnology Information (NCBI) subtyping program (http://www.ncbi.nih.gov/projects/genotyping/formpage.cgi) and the HIV Blast tool were used (http://www.hiv.lanl.gov/content/hiv-db/BASIC_BLAST/basic_blast.html). Subtyping was ultimately confirmed by phylogenetic analysis using the 2003 and 2005 recommended reference sequences from the Los Alamos site (downloaded from: http://www.hiv.lanl.gov/content/hiv-db/SUBTYPE_REF/align.html) (see Table 4 below for a subset of these references that will be referred to in many of the figures within this thesis). Sequences were aligned in Mega 3.0 using Clustal W (232). Manual alignment was necessary for regions of extensive diversity and the 2003 HIV Sequence

Compendium was used as a guide (http://www.hiv.lanl.gov/content/hiv-db/HTML/2003compendium.html). Phylogenetic analysis using neighbour joining (NJ) trees and the Kimura 2 parameter model was done along with subtype reference sequences. Bootstrapping using 1000 replicates was performed. Bootstrap values are shown out of 100 and values of 50 or greater are shown on trees. (Values of 70 or higher are typically used as a cut off but this is thought to be a highly conservative measure. Hillis and Bull examined the relationship between a bootstrap value and the probability that subtyping is correct. They found that for branches with values of 70% that there is a probability of 95% that the clade is true. (233) Based on the phylogenetic relationship with the reference sequences a subtype was assigned (see Figure 2 for an example).

Subtype references:

Los Alamos subtype reference sequences from the 2005 and 2003 recommended lists were used as detailed above (see Table 4).

Recombination breakpoint analysis:

Fine mapping of recombination breakpoints was done using Simplot (231). A four sequence alignment consisting of the putative recombinant sequence, two potential parental sequences and an outlier was used. The choice of parental sequences depended on the sequence under investigation. Subtype consensus sequences and specific isolate reference sequences were both tested. If along with the recombinant form, parental sequence clones were also detected within an individual, then those sequences were used

as the parental references in the analysis of the recombinant virus to determine the breakpoint.

Informative site distribution was used to determine potential breakpoints. Informative sites in the four sequence alignment were used to yield three possible configurations, two of which show the recombinant sequence clustering with one of the two "parental" sequences. Then sites are assessed to see if a breakpoint placed along the alignment will result in a significant difference in the ratio of the site types on each side of the break. This is assessed by chi-square values. Breakpoint ranges were determined by attaining the maximum chi-square value. (Breakpoints were selected on a lower than maximum chi-square value for some clones if after examination of the sequence alignment it was evident that choosing the higher value would shift the breakpoint due to a single mutation (234)). NJ trees and bootstrapping of the different regions with reference sequences was then done to confirm findings. (Note: Although not ideal, if recombinant segments were too small this final step does not give reliable results and recombination assessment had to be based on RIP and Simplot outputs and phylogenetic trees with only a minimum number of references.)

Protease Amino Acid Sequence Analysis:

Differences from the protease amino acid sequence of the HIV-1 HXB2 virus were determined for sequences that were categorized by subtype (A, C, D, G). Protease nucleotide sequences were translated and analyzed using Sequencher and Bioedit programs. They were assessed for differences at resistance mutation sites according to the

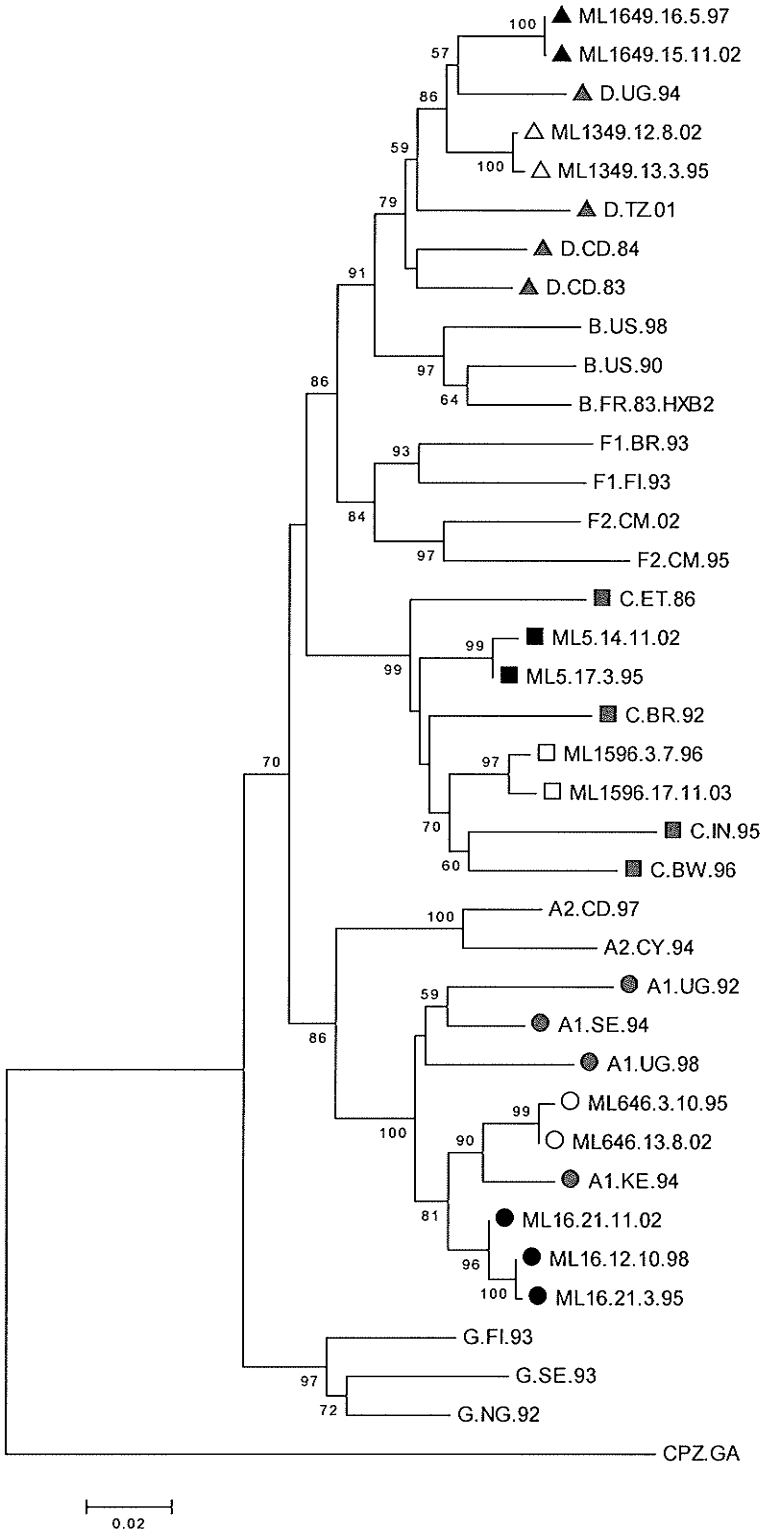

Figure 2: Subtyping of Sequence Data

Figure 2: Subtyping of Sequence Data by Phylogenetic analysis example: Circles=A1, Triangle=D, Square=C. Grey= reference sequences, Black and White= different sex workers (ML cohort). The number refers to the unique identifier each woman is given and the date that follows is the date the sample was taken from each woman. Numbers at nodes indicate the bootstrap values. This NJ dendogram reveals that ML5 and ML1596 are subtype C, ML1349 and ML1649 are subtype D and ML16 and ML646 are subtype A based on their relationships with the reference sequences. Sequences from each individual but from different dates are also most similar to each other.

Table 4: Selected HIV-1 Sequence References: from the Los Alamos 2003 and 2005 recommended reference sequence list. These will be referred to in figures that follow.

NAME	ACC #	Sampling Country /year	REFERENCE
A1.KE.Q23-17	AF004885	Kenya (KE)/1994	J Virol 72(10):8240-51 (1998)
A1.SE.SE7253	AF069670	SE/1994 (Somalia origin)	AIDS 13(14): 1819-1826 (1999)
A1.UG.92UG037	U51190	Uganda(UG)/1992	J Virol 70(3): 1651-1657 (1996)
A1.UG.98UG57136	AF484509	UG/1998	ARHR 18(17); 1281-90 (2002)
A1.UG.U455	M62320	UG/1985	ARHR 6(9):1073-1078 (1990)
A2.CD.97CDKFE4	AF286240	Cambodia (CD) /1997	ARHR 17(8): 675-688 (2001)
A2.CY.94CY017-41	AF286237	Cyprus (CY) /1994	ARHR 17(8): 675-688 (2001)
B.FR.HXB2	K03455	France (FR)/1983	Nature 313(6000):277-284 (1985)
B.US.WEAU160	U21135	United States (US)/1990	FEBS Letters 281: 77-80 (1991)
B.US.1058_11	AY331295	US/1998	J Virol 79(17); 11523-8 (2005)
C.BR.92BR025	U52953	Brazil (BR)/1992	J Virol 70(3):1651-1667 (1996)
C.BW.96BW0502	AF110967	Botswana (BW) /1996	J Virol 73(5): 4427-32 (1999)
C.ET.ETH2220	U46016	Ethiopia (ET) /1986	ARHR 12(14): 1329-1339 (1996)
C.IN.95IN21068	AF067155	India (IN) /1995	J Virol 73(1):152-160 (1999)
D.TZ.A280	AY253311	Tanzania (TZ) /2001	Arroyo, MA Unpublished
D.CD.ELI	K03454	CD/1983	Cell 46(1):63-74 (1986)
D.CD.84ZR085	U88822	CD/1984	J Virol 72(7):5680-98 (1998)
D.UG.94UG1141	U88824	UG/1994	J Virol 72(7):5680-98 (1998)
F1.BR.93BR020-1	AF005494	BR/1993	J Virol 72(7):5680-98 (1998)
F1.FI.FIN9363	AF075703	Finland (FI) /1993	Virology 269(1):95-104 (2000)
F2.CM.MP255	AJ249236	Cameroon (CM) /1995	ARHR 16(2): 139-151 (2000)
F2.CM.02CM_0016BBY	AY371158	CM/2002	ARHR 20(5); 521-30 (2004)
G.FI.HH8793-12-1	AF061641	FI/1993 (Kenya origin)	ARHR 8(9):1733-1742 (1992)
G.NG.92NG083	U88826	Nigeria (NG) /1992	J Virol 72(7):5680-98 (1998)
G.SE.SE6165	AF061642	Sweden(SE)/1993 (Democratic Republic of Congo origin)	Virology 247(1):22-31 (1998)
CPZ.GA.CPZGAB	X52154	Gabon (GA)	Nature 345(6273):356-359 (1990)

Stanford University HIV Drug Resistance Database (http://hivdb.stanford.edu/cgi-bin/PIResiNote.cgi- resistance matrix last updated 5/05) and are grouped into mutation types I-IV.

Superinfection/ Dual infection Analysis:

If two HIV-1 sequences, of the same genetic region, from a particular individual are no more similar to each other than to other epidemiologically unlinked HIV sequences then this can be defined as infection with two distinct viruses (186). These distinct viruses may be detected at a single time point or may be seen at different time points in the course of an individual's infection. Sequence data (Gag/Pol: p6, pro, part RT) from all individuals and references were aligned and a single phylogenetic tree constructed. (See Figure 3A for an example of a smaller tree which illustrates this point. Sequences from one individual do not cluster closest with each other when compared to both reference strains and HIV-1 sequences from other women within the ML cohort.) Patients whose samples did not cluster together in the context of sequences from both references and all other ML samples were further investigated for possible superinfection or dual infection. HLA typing was done as described above to confirm sample identities. PCR, TA cloning and sequencing was done on all available samples for women who were dually or superinfected. (Note: in cases where TA cloning and sequencing was done a region that encompasses p1, p6, protease and part of RT that is roughly 900 base pairs in size was used for analysis.) See Figure 3B for an example of a dually infected individual.

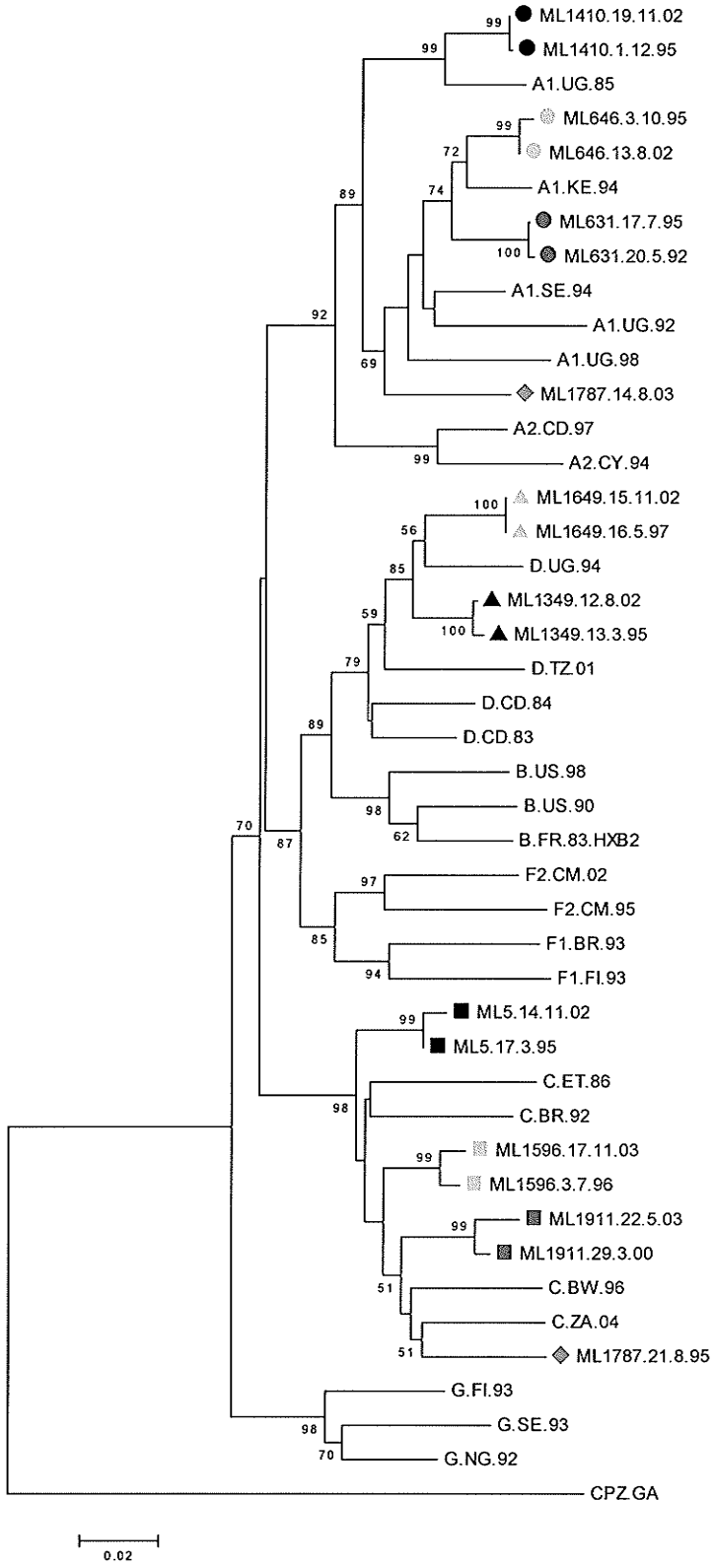

Figure 3A: Sequences from an Individual do not Cluster Together

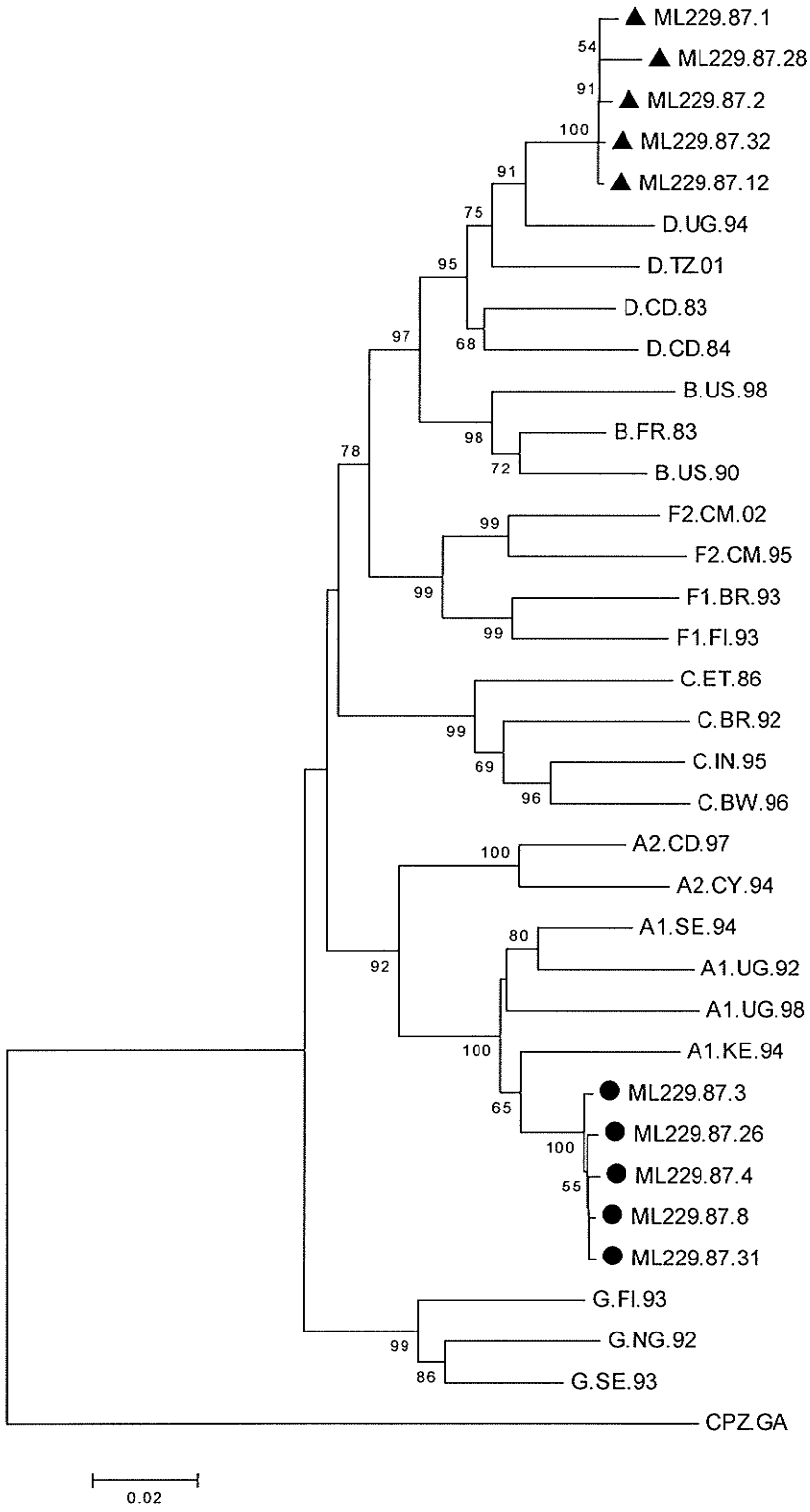

Figure 3B: Example of Two Distinct Viruses Found Within One Patient

Figure 3: A. Sequences from an Individual do not Cluster Together: This dendogram includes HIV-1 sequence data from Gag/Pol (2134-2687 by HXB2 numbering) HIV-1 reference sequences and sequences from HIV-1 infected ML women are included here. Matching symbols are used to show sequences that are from the same individual. Here all sequences from each individual cluster closest with each other except for ML 1787. This individual was later found to be a superinfection case. Her 2003 sample is actually an A/C recombinant virus. **B. Example of Two Distinct Viruses Found Within One Patient:** Phylogenetic analysis displays sequence clones from ML 229 (p1, p6, protease and part of RT sequence from 2086-2988 HXB2 numbering) obtained from a 1987 sample. The figure displays the point that some of the individual's sequences cluster more closely with other sequences than with each other. This individual is clearly dually infected with subtype A and subtype D viruses (triangle=D sequences from ML229, circle=A sequences from ML229) at the 1987 time point studied.

Epidemiological and Statistical Investigation

Epidemiological Data:

After data cleaning, several key variables from the questionnaire forms completed for the ML women were used in this study. Year of birth (yob) or age was recorded at each resurvey period. Data on condom use included a personal assessment of whether condoms were used always, most of the time, some of the time or never. In cases where this personal assessment information was missing a calculation was made by taking the number of condoms used per week divided by the number of sex partners per week. A score of 100%=always; >50%=most of the time; <50% but more than 0%=some of the time; and 0%=never. Information on the number of sex partners per day, the duration of prostitution or prostitution start date were used. In cases where data was missing or discrepancies were present for a particular individual all of the resurvey data for that sex worker was looked at and a median value was used.

HIV status was recorded as negative, positive on entry, or as seroconverters. Seroconversion dates were calculated as the midpoint between the last negative and the first positive date. Superinfection dates were calculated by taking the midpoint between the latest sample in which only the initial virus was detected and the earliest sample in which the second virus was detected.

Details of Statistical Analyses

Subtype distribution over time study:

Distribution changes over time were assessed based on subtyping data from Gag/Pol (p6/pro/partRT) region for all individuals. Those that entered the study positive

as well as seroconverters were included in this study. The dates used in this study were assigned to each individual as follows: date of entry was used as the HIV date for those that entered positive and the estimated seroconversion date was used for seroconverters (regardless of sampling date).

Subtypes were broken down into A, C, D or an "other" category. The "other" category included recombinants, unclassifiable or subtype G. Time was divided into 3 categories representative of early, mid and late stages of the pandemic as follows, 1985-1989, 1990-1995, 1996-2003. This time period breakdown was influenced by a number of factors, the first was that I wanted an early period to be defined in which HIV was first being detected and studied and during which time prevention and condom use promotion had just begun (chosen here as 1985-1989). I also wanted a later period to be distinguished (chosen here as 1996-2003) which encompassed the late 1990s, when the peak HIV prevalence was seen in Kenya (2). Another factor contributing to the time periods chosen was the establishment of a breakdown in which each time period would have a relatively equivalent number of individuals sampled. A statistical assessment of the changes in subtype distribution was performed using Chi-square and Maentel-Haenszel tests. A p-value of < 0.05 was considered significant (a trend was noted if p values were < 0.1).

CD4+ T cell decline (Kaplan-Meier survival curves):

Disease progression was addressed through an investigation of time to CD4+ T cell drop below 400 cells/mm^3 and below 200 cells/mm^3. This was measured by Kaplan-Meier survival curves (the event being CD4+ T cell counts falling below 400/200

cells/mm^3). This technique was used previously by members of our research group (235). Drop to below 200 cells/mm^3 was chosen as it is AIDS defining and results in an increased susceptibility to opportunistic infections (105). Drop to below 400 cells/mm^3 was chosen here as an indicator of a cell count that is below what is considered a normal CD4+ T cell count (normal range in adults is 400-1400 cells/mm^3) (27). Drop in CD4+ T cell counts was determined using the following criteria: Time to CD4+ T cell drop below 400 cells/mm^3 and time to CD4+ T cell drop below 200 cells/mm^3 required a count below 400 cells/mm^3 and 200 cells/mm^3 respectively for at least two consecutive dates. This was decided upon based on extreme fluctuations that were observed in many of the ML women but for whom such a decline, typically reported at a single time point, was not maintained. Only women who had an initial count above 400 cells/mm^3 and 200 cells/mm^3 were included in the time to below 400 cells/mm^3 and 200 cells/mm^3 studies respectively. If CD4+ T cell counts were above 400 cells/mm^3 or 200 cells/mm^3 for the first available cell count it was assumed that their counts were above 400 cells/mm^3 or 200 cells/mm^3 respectively prior to this as well. In these cases time from either entry into the cohort if positive, or seroconversion was thus included. Mean "survival" times as well as the point at which 50% of individuals in each group had dropped below the cell count of interest were also calculated. Kaplan-Meier survival curves with log-rank and Wilcoxon tests were used to determine if there were significant differences in the relationship between a specific factor (subtype) and the progression endpoints (outcomes of either drop below 200 cells/mm^3 or 400 cells/mm^3). This type of investigation, using these criteria, was used in a number of the analyses reported here.

Subtypes and disease progression

This study aim was to determine if subtype associated differences in clinical progression existed. The ML women were separated into one of four categories, those infected with A, C, D or a recombinant virus. This classification was based on subtyping data from the p6, pro, partRT region. All additional sequencing data available for individual patients within the ML cohort was also used. This included envelope data and data from a previously published study (236). Women who entered the cohort positive and those who seroconverted were included in this study.

A number of variables were investigated and are outlined below. The percentage of individuals with initial available CD4 counts that fell within the following ranges: >400, 201-399 or </= 200 for all groups A, C, D and recombinant (all individuals included n=202) was determined and differences assessed by Chi-square and Maentel-Haenszel Chi-square tests. Study groups which included those who met the criteria described above (in the time to drop below 200 cells/ mm^3 and 400 cells/ mm^3 description) were compared using the approach detailed above for the time to drop below 200 cells/ mm^3 and 400 cells/ mm^3. Within the study groups the percentages of individuals who dropped below 400 cells/ mm^3 and 200 cells/ mm^3 for each subtype were again compared by Chi-square and Mantel-Haenszel tests.

To address the question of whether changes in disease progression (as measured by time to drop below 200 cells/ mm^3 and 400 cells/ mm^3) have occurred for one particular subtype over time (85-89, 90-95, 96-04) analysis was carried out as described in the CD4 decline section. In order to determine if results were biased by different durations of infection prior to entry in the study the time from prostitution start to the first

CD4 value (if entered positive) or the time from seroconversion date to the first CD4 (if seroconverter) was determined for each individual. Mean times for each group (85-89, 90-95, 96-04) were compared using an ANOVA test.

Superinfection Investigation:

Nested cohort study

To test the hypothesis, that HIV-1 infection provides protection against superinfection, we attempted to study as many ML women as possible at multiple time points throughout the course of their infections. If for an individual, only one sample was successfully amplified and sequenced or multiple samples were not available they were not included in this superinfection "nested cohort" study. PCR of the Gag/Pol region was performed (using degenerate primers capable of amplifying all subtypes). Population based sequencing followed, as previously described. If the sequence data revealed numerous double peaks (2% or more) or if sequencing was unsuccessful then TA cloning and sequencing of at least 30 clones per sample was done. After recombination screening (using RIP) all of the sequences from the ML women along with the reference sequences were aligned and phylogenetic analysis performed. If samples from the same woman did not cluster together HLA typing was done to confirm identities.

Statistical assessment (of the protective effect of HIV-1 infection on superinfection):

To ascertain whether infection with HIV-1 provides protection against superinfection we employed two different types of analyses (Poisson regression and Cox's proportional Hazard modeling) as described below. Data on superinfection within the superinfection study cohort was compared to data on initial HIV infections within the

entire cohort. The entire cohort includes all individuals who entered the cohort negative between 1987 and 2003 (corresponds to time periods studied in the superinfection study group) and the superinfection study cohort includes HIV positive women studied at multiple time points for possible detection of superinfection.

Variables that were assessed using Poisson Regression and Cox's proportional Hazard modeling included the following:

Cohorts: Superinfection cohort compared to the entire cohort

Condom use: this was separated into two categories. The first is use of condoms some of the time or never and the second is use of condoms most of the time or always

Sex partners per day: Organized into two categories, 1-4 or =/>5

Prostitution duration: Prostitution duration at time of entry into ML cohort was broken down into three categories, 0-5, 6-15, >15 years

Age: Age at time of entry into the ML cohort was divided into age 25 or less, 26-35 or >35

Multiple Poisson regression:

Brief description: This type of analysis takes into account each time a person is seen while being followed and the events that are reported. Crude unadjusted values were first calculated. Then adjusted odds ratios (OR) with 95% confidence intervals (CI95%) and p-values were calculated for all variables and then further adjusted for only those variables that were significant.

Multivariate Cox's proportional hazard modeling:

Brief description: Cox's Proportional Hazard Modeling analysis allows for the analysis of the time in which superinfection or initial HIV infection occurred and what factors were present at that time. Prior to the hazard modeling, a crude assessment using Kaplan-Meier survival curves with log-rank and Wilcoxon tests was used to statistically compare the two groups and the percentage of individuals that remain infection/superinfection free over time. Hazard ratios with 95% confidence intervals and p-values were calculated taking all covariates into account and also using a parsimonious model where only significant variables were included. In this study condom use and sex partners per day were time-dependent variables while the others were only considered at initial visit.

Disease Progression in Dually or Superinfected Individuals:

Superinfection and dual infection (case control studies):

Assessment of clinical progression following superinfection/dual infection was analyzed by a number of methods. One of the methods used to evaluate these individuals was through a case control study, where each case was matched with two controls. This was performed to evaluate CD4+ T cell count data (rate of decline based on regression lines, intercepts (theoretical first CD4 counts) and survival curves of time to drop below 200 cells/mm^3 and 400 cells/mm^3 in the superinfected individuals compared to controls and in the dually infected individuals compared to controls. Cases included the women in the study who were dually or superinfected based on sequence analysis. Controls (women with single HIV-1 infections) were selected from the cohort of women whose samples

were sequenced and assessed for dual/superinfection at more than one time point. ML controls were also selected and matched to cases for age, HIV status (entered positive or seroconverted) and prostitution duration prior to study entry.

Individual assessments of consequences of dual/superinfection:

CD4+ T cell counts and viral load (VL) data are two important measures used to assess disease outcome (237). Increased clinical progression has been described in other studies of dual or superinfection cases as either, a high viral load, an increase in viral load, a drop in CD4+ T cells or rapid progression to CD4+ T cell counts below 200 cells/mm^3 (218). In addition to case control studies described above a case by case assessment of patient CD4+ T cell counts, viral load (VL) data and time to AIDS was also performed. Unfortunately a comparison group based on viral load data for all women infected with a single virus in this study was not available. Instead previously reported average viral load data from a group of Kenyan women (from the ML cohort or an antenatal cohort) were used as a reference point when discussing VL data (52,841+/-18,056 copies/mL) (238). Rapid progression was also addressed and the definition used was any individual who progressed to AIDS within 2 – 3 years following infection. Data concerning opportunistic infections and time to death would have been useful to look at here but were not available for all individuals under study.

RESULTS:

Section I:

The research described in section-I was directed at testing the initial hypothesis, **that an investigation of HIV-1 genetics, within the group of commercial sex workers having single HIV infections, will reveal the following: A.) a diverse array of genetically distinct viruses, B.) a dynamic and heterogeneous distribution of viral subtypes, C.) viral genetic associated differences in disease progression.**

A "single HIV-1 infection" is defined as an HIV-1 infection resulting from a single HIV-1 infection event with one distinct viral type and no signs of superinfection/dual infection. The aim was to better understand single HIV-1 infections within the cohort on a molecular epidemiological level as well as the clinical consequences of viral genetic differences. A thorough investigation and understanding of these HIV infections was an essential foundation for subsequent studies of more complex HIV-1 infections including multiple subtype infections and superinfection (SI) events.

Section I A: HIV-1 (Single Infections) Within the ML Cohort: a Molecular Profile

To test the hypothesis that an analysis of single HIV infections within the cohort would reveal a diverse array of genetically distinct viruses, three key elements of viral genetics were investigated. The first was the viral subtype distribution within the sex workers. The second was a study of the recombinant viruses present to determine if they are common and if they are a late or early introduction into the cohort. The third was an investigation into differences at drug mutation sites within the different viruses infecting these women. These data are important for present and future studies within the cohort.

A diverse mix of HIV-1 subtypes are present:

We hypothesized that a diverse array of viral subtypes would be present within the cohort. To determine the circulating subtypes within the ML sex worker cohort in the Pumwani slum of Nairobi, Kenya, degenerate primers were designed to amplify all possible subtypes. The p6 region of Gag, which is a diverse region, and the more conserved protease gene and part of the RT gene from Pol were sequenced (nucleotides (nt) 2134-2687 by HXB2 numbering- refer to Figure 1 for the HIV-1 genetic map) (For samples that required TA cloning sequence data also included p1 from Gag and additional RT sequence (nt 2086-2988 by HXB2 numbering). In total, 213 individuals were studied at one, two or several time points. Over 400 samples were sequenced of which nearly 25% required TA cloning.

All population based sequences and all clones (totaling approximately 3000 sequences) were examined for recombination and subtyped as described in the methods section. Subtyping and recombination analysis based on the described region (Gag/Pol) for 202 of the women revealed a diverse mix of subtypes including 138-A, 14-C, 43-D, 2-G, 4-recombinant (3 D/A and 1 D/unknown) and 1 unclassifiable virus (Figure 4). Eleven women were excluded from the denominator and this breakdown because sequence data revealed dual infections or superinfections. These cases will be discussed in detail in a later section.

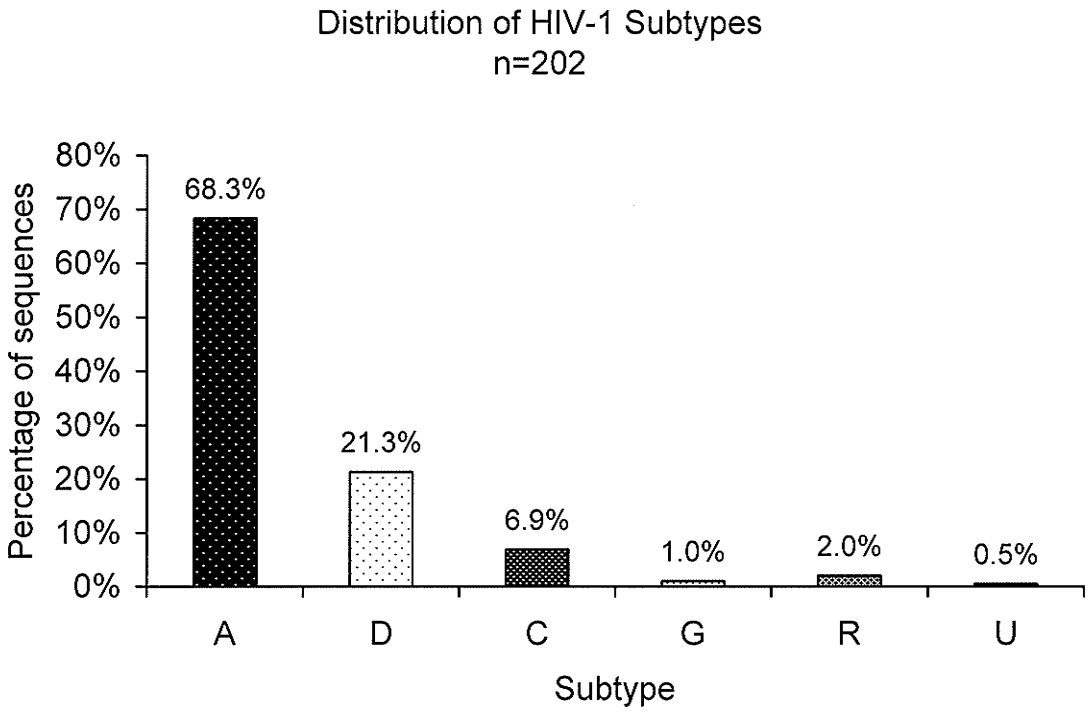

Figure 4: **HIV-1 Subtype Distribution** based on sequencing of Gag/Pol (p6, protease and part RT) for 202 women infected with a single virus. Subtypes A, D, C, G, R=recombinant, U=unclassifiable.

Recombinant viruses are common and present since early in the epidemic:

Data revealed a diverse mix of viral subtypes present within the ML cohort. Next I investigated recombinant viruses to determine if, as predicted, they would be commonly found within the ML cohort and also present since early in the epidemic within this group. Based on the Gag/Pol region described above, recombinant viruses were detected in 4 of the 202 women assessed. Three individuals had A/D recombinant viruses while one had a D/U (unclassified) recombinant virus. (If the individuals who were dually

infected or superinfected were included here, this number would increase substantially as recombinant viruses were detected in 10 of the 11 individuals.)

To determine how frequently additional recombinant viruses would be identified if other regions of the viral genome were also investigated, two additional regions were evaluated in subsets of the women studied. Subtyping results were then compared to the Gag/Pol subtype assignments. A region of the gp120 envelope glycoprotein gene spanning V3 to the end of gp120 (nt 7008-7817 by HXB2 numbering) was successfully cloned and sequenced for 17 individuals from the group (cloning and sequencing of this region was attempted for other individuals but the cloning technique utilized, at the time that this region was under investigation, was not very efficient). Multiple samples from different time points were cloned and sequenced for 15 of the women and 2 had discordant results when subtype assignments for Gag/Pol and the Env segment were compared. One of these individuals had a C subtype designation for Gag/Pol but was subtype D for envelope while the other had a D subtype in Gag/Pol and an A/D form in envelope. In each of these cases this finding was evident in both the initial and later dates studied for each individual. Two additional individuals were investigated only at one time point each and no discordance between regions studied was observed. In total, 2 of the 17 individuals (12%) harbored recombinant viruses based on the subtyping data from different parts of the genome (see Figure 5).

Viruses from 48 individuals who were also from the original group of 202 women were sequenced at one or two time points for yet another genetic region spanning nt 767-1921 (HXB2 designation) and covering p17 and p24 of Gag (Iversen, A. personal communication March 2006). When compared to the original Gag/Pol (p6, PR, RT)

subtyping results, 43 of the 48 women (89.6%) had concordant subtyping data while 5 of 48 (10.4%) had discordant results. Two were A/D in Gag but D in Gag/Pol, one was C in Gag and another was A but were subtypes A and C respectively in Gag/Pol. Lastly, one individual had an A/D type virus in Gag but went unclassified in Gag/Pol (most similar to A). Additional recombinant viruses were thus revealed (10.4% of individuals) when the subtyping data from these two different non-overlapping regions were compared (see Figure 5).

These data indicate that recombinant viruses are not uncommon within the cohort. Based on all available genome sequencing performed for each individual, a total of 11/202 (5.4%) of individuals were identified as having recombinant viruses. This number increased to 21/213 (9.9%) if all dual and superinfected individuals, discussed in sections II and III, are also included.

I next wanted to determine if the recombinant viruses had arisen early on or only recently within the commercial sex worker cohort. The dates of samples from the ML women in which recombinant viruses were identified were determined and ranged from 1991-2003. These individuals seroconverted or entered the study positive between 1985 and 2001. Seven of the 21 individuals (33.3%) in whom recombinant viruses were detected had seroconversion or HIV positive entry dates between the years 1985-1989 (those superinfection cases where recombinant viruses clearly arose later were excluded).

The investigation into recombination thus revealed that recombinant viruses are common within the cohort (9.9% of women based on partial genome sequencing alone) and that their existence is not just a recent phenomenon.

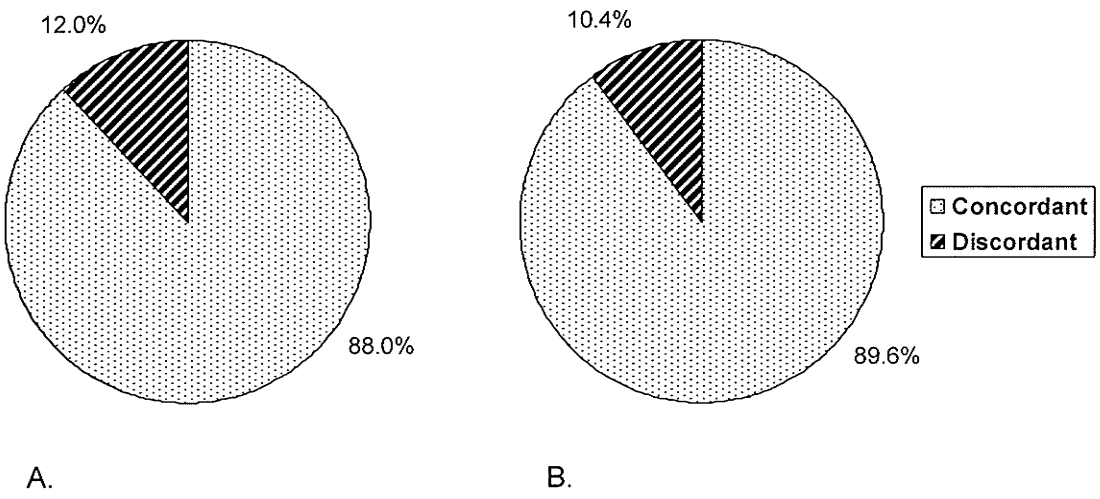

Figure 5: Recombinant Viruses were Detected Based on Sequencing of Additional Regions: Discordant and concordant results based on assessment of multiple regions are shown here. Discordant results indicate additional recombinant viruses. A.) Env vs. Gag-Pol (includes: p6, protease, part RT) B.) Gag (p17, p24) vs. Gag-Pol (p6, protease, part RT)

Some Type IV resistance mutations are natural polymorphisms within non-B subtypes:

The data so far have revealed that the women of the ML cohort have been infected with a diverse mix of HIV-1 viral subtypes including recombinant viruses which have been around since early in the epidemic. As part of this population's viral assessment I also wanted to determine the baseline prevalence of naturally occurring polymorphisms that are associated with drug resistance in subtype B viruses. The individuals in this study are drug-naïve but treatment is now being implemented in the

cohort; therefore, determining resistance mutations that may be natural variants in non-B subtypes was an important component to establishing a useful molecular profile. Such mutations may have an impact on the recently implemented antiretroviral treatment. I predicted that resistance mutations in positions critical to protease function would not be present in the drug-naïve individuals infected with different subtypes but that type III and IV (described below) resistance mutations may be natural variants for some subtypes. Sequences of the protease gene and the corresponding aa sequences were analyzed for 200 women (subtypes A, C, D and G) at their most recent time points at which they had not yet been receiving antiretrovirals.

The resistance sites assessed were those defined by the Stanford University HIV Drug Resistance Database (239). Type I, major mutations are those that alone can reduce susceptibility to one or more protease inhibitors (PI) (239). The sites at which type I major resistance mutations occur were relatively conserved in the non-B drug-naïve individuals except for the V82I mutation seen in some individuals (which does not cause drug resistance) and one D subtype individual who had G48R (effect unknown). Type II, flap mutation sites are the next most important and may cause resistance on their own but are usually accessory(239). The type II flap mutation sites were all non-polymorphic in the individuals studied. Type III mutations are additional protease inhibitor resistance mutations that are at sites that are usually non-polymorphic in drug naïve individuals (239). Within the type III sites one subtype D individual had the resistance mutation L33I. Amino acids valine (V) and phenylalanine (F) were also seen at this site; the effects of which are unknown. Type IV mutations are those that aid in resistance when they are combined with one or more category I-III mutations (239). The type IV mutations were

the ones in which the greatest numbers were seen in the non-B type viruses. The prevalence of these are as follows, L10I/V: 22.5% (A), and 6.5% (D) ; K20R/I: 25.4%(A), 8.7%(D) and both G sequences had K20I. M36I was present in 98.6% (A), 50% (D), 92.9% (C) and was also present in both G sequences. L63P was observed in 9.4% (A), 23.9% (D) and 21.4% C. V77I occurred in 1.4% (A), 6.5% (D), and 7.1% (C) (see Figure 6). Non-B subtype viruses seen within the ML cohort thus do not show a high proportion of type I-III mutations but type IV mutations are not uncommon. For some subtypes the type IV mutations (as defined for subtype B viruses) are the most common natural variant for that subtype, such as the M36I mutation described above where the majority of A and C type viruses studied contained this mutation.

Section IA Summary

In summary, women of the ML cohort are infected with a diverse array of genetically distinct viruses. Multiple subtypes are present within the cohort and recombinant viruses are common. Recombinant viruses have also been present since early in the epidemic. Type I-III resistance mutations were uncommon in the non-B subtype viruses infecting the drug–naïve commercial sex workers studied. Type IV mutations however were not rare and some were highly prevalent natural polymorphisms for non-B subtypes.

											♦									
Subtype	#	P 1	Q 2	I 3	T 4	L 5	W 6	Q 7	R 8	P 9	L 10	V 11	T 12	I 13	K 14	I 15	G 16	G 17	Q 18	L 19
A	138										11V 21I		2I 1P 1A 1K 1S 1V	116V	66R	24V	9E 3A	1E	1H	6I
D	46										3I			36V	6R	6V 1L	1E			2I 1T
C	14												8S 1A			14V	1A			6I 2T 1V
G	2													2V	1R					

Active Site

		♦		▲	▲							●	▲	▲			♦		
Subtype	#	K 20	E 21	A 22	L 23	L 24	D 25	T 26	G 27	A 28	D 29	D 30	V 32	L 33	E 34	E 35	M 36	N 37	L 38
A	138	32R 3I									1N			3V 1F		134D 2N	1L 136I	63D 4E 1Q 1Y 1S	
D	46	4R												3V 1I		3D	23I 2L	3D 1T 1S 1H	
C	14															2D	13I	6K 2D 1T 1S	
G	2	2I															2I	2S	

Flap Region

									■	■	●		●			■	■		
Subtype	#	P 39	G 40	R 41	W 42	K 43	P 44	K 45	M 46	I 47	G 48	G 49	I 50	G 51	G 52	F 53	I 54	K 55	V 56
A	138	1S 1Q		135K 1N	1 stop codon	1R		26R				1E							
D	46			45K		1R					1R								
C	14			6N 8K				2R											
G	2			2K															

								♦								♦		▲
Subtype	#	R 57	Q 58	Y 59	D 60	Q 61	I 62	L 63	I 64	E 65	I 66	C 67	G 68	H 69	K 70	A 71	I 72	G 73
A	138	123K		1F	15E	5E 1K	4V	13P 9S 3T 2I 2V 1A	1L	4D 1K		6Y	3E	136K 2Q	12R	1V	1V 1L	
D	46	1K			3E	1N 1D 1H	5V	11P 1H 2S 3T 1Q 1C	34V 1M	1D		1Y		12Y 1R 1Q 1K	4R		8V 1E 1M	
C	14				1E			2V 3T 3P 1S										
G	2				1E							1E 1S						

Sub-binding region

				▲	♦					●	●			▲			●		
Subtype	#	T 74	V 75	L 76	V 77	G 78	P 79	T 80	P 81	V 82	N 83	I 84	I 85	G 86	R 87	N 88	L 89	L 90	T 91	Q 92
A	138		1V		2I	1R				1I				2E			137M 1I		1C	
D	46				3I									1R			1F			
C	14	2S			1I					3I							11M			
G	2									1I							2M			

Subtype	#	I 93	G 94	C 95	T 96	L 97	N 98	F 99
A	138	2L 1M						
D	46	2L						
C	14	14L						
G	2							

Figure 6: Protease Amino Acids 1-99: An Assessment of Drug Resistance Sites for Different Subtypes. The aa sequence listed is for subtype B. Sites were assessed according to the Stanford University HIV Drug Resistance Database (resistance matrix last updated 5/2005) and were grouped into types I-IV. Type I, major mutations (position indicated by a circle) are those that alone can reduce susceptibility to one or more protease inhibitor. Type II, flap mutation sites (position indicated by a square) are the next most important and may cause resistance on their own but are usually accessory. Type III mutations (position indicated by a triangle) are additional protease inhibitor resistance mutations that are usually non-polymorphic in drug naïve individuals. Type IV mutations (position indicated by a diamond) are those that aid in resistance when they are combined with one or more category I-III mutations (239). The shaded regions indicate functionally important regions including the active site, flap region and sub-binding region.

Section I B: Viral Subtype Distributions Over Time and Within Unique Groups

HIV-1 diversity is present at a molecular level within the commercial sex workers of the ML cohort. I next wanted to test the hypothesis that a dynamic and heterogeneous distribution of viral subtypes would also be seen. Subtype distributions were studied over time to establish if changes had occurred throughout the course of the epidemic within the commercial sex workers. To determine if subtype distribution was heterogenous within different subsets of individuals within the cohort I studied a specific group of late seroconverters and compared their subtype distribution to the rest of the cohort studied.

Changes in subtype distribution over time

The hypothesis that a dynamic distribution of viral subtypes would be seen was tested by studying viral distributions in the ML cohort over time. The women in the study were broken down into groups by date of seroconversion, if available, or date of entry into the cohort, if entered positive. The time periods 1985-1989, 1990-1995 and 1996-2003 were used as representative of early, mid and late epidemic stages (n= 73, n=59, and n=70 respectively). As described within the methods section, this time period breakdown was selected such that each defined group (1985-1989, 1990-1995 and 1996-2004) would have a relatively equivalent number of individuals. I also wanted one time period to represent the early epidemic (chosen here as 1985-1989) and another (1996-2004) to contain the point at which HIV prevalence had peaked in Kenya (late 1990s) (2).

The proportions of the most common HIV types (A, D, C) and a group defined as "other" due to low numbers (recombinants, G subtype, unclassifiable), were determined for each time period based on the Gag/Pol region sequenced (see Figure 7). The subtype

distribution was significantly different over the three time periods (Chi-square p= 0.0324, Mantel-Haenszel p=0.0067). (Note: The proportion of individuals who were seroconverters within the three time periods was not significantly different (Chi-square p=0.6340, Mantel-Haenszel p= 0.3571).)

Changes in the proportion of each of A, D and C subtypes seen over time were also addressed separately. There was a significant difference in the percentage of circulating viruses that were subtype A over the three time periods; a consistent decrease was seen (Chi-square p= 0.0125, Mantel-Haenszel p= 0.0048). This decline was also looked at in another way, using an alternative breakdown in time. The midpoint of three year time intervals from 1985-1987 onward until 2002-2004 was used, and as time increased, individuals were shown to be less likely to be infected with an A type virus (OR=0.943, CI95% 0.895-0.993, p= 0.0253).

The number of C subtype viruses was very small. Although the proportion of C subtype viruses in the population increased in each time period the increase was not significant. However, a trend in an increase in subtype C was seen (Chi-square p= 0.1820, Mantel-Haenszel p= 0.0815).

The proportion of viruses that were subtype D was higher in 1990-1995 compared to 1985-1989 but a continued increase in subtype D was not evident in 1996-2003 (Chi-square p= 0.0999, Mantel-Haenszel p= 0.2041). When the initial period 1985-1989 was compared to 1990 or later; however, the proportion of subtype D viruses that were seen was significantly greater from 1990 onward (Chi-square p= 0.051).

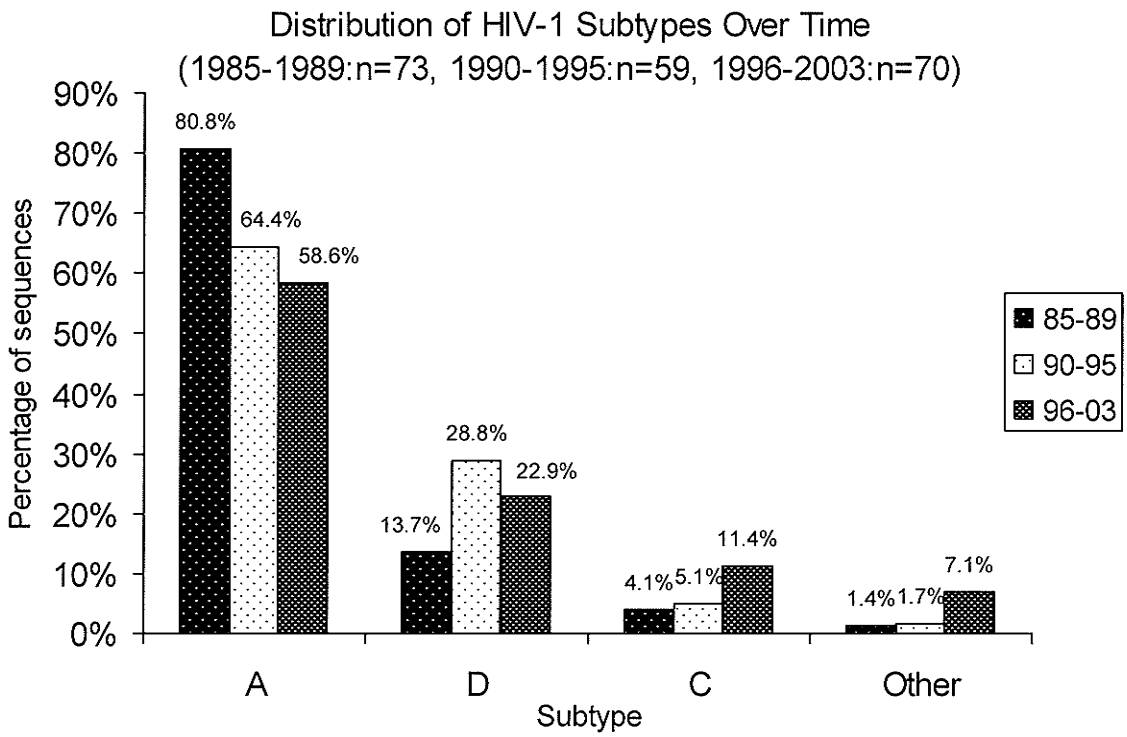

Figure 7: Distribution of HIV-1 Subtypes Over Time: The HIV-1 subtype distribution within the cohort for three different time periods (1985-1989 n= 73, 1990-1995 n=59, 1996-2003 n=70) is seen here. The breakdown is based on sequencing of Gag/Pol (p6, protease and part RT) for a total of 202 women. Subtypes A, D, C and Other: (subtype G, recombinant, unclassifiable) are shown here.

Late seroconverters do not have a significantly different subtype distribution:

The subtype distribution within the ML cohort was found to be dynamic; and significant changes were observed over the course of the epidemic. To test the hypothesis that subtype distribution was also heterogenous within different subsets of individuals within the cohort, I studied a specific group of late seroconverters and

compared their subtype distribution to the rest of the cohort. Late seroconverters, are defined as women who were once considered to be resistant (negative for at least three years) but then later seroconverted. There were 21 of these late seroconverters available for subtyping. Subtyping was determined based on sequencing of the p6/Pol/RT region. The proportion of late seroconverters infected with non-A subtypes were compared to the proportion seen for the rest of the women in the subtyping study. The proportion of those who were infected with non-A subtypes did not differ significantly from the rest of the cohort that was subtyped (Chi-square p= 0.2785).

These late seroconverters (n=21) were also divided into two groups, those who were "resistant" for 5 years or less (3-5 years) and those who were resistant for greater than 5 years. In the less than 5 years group (n=12), 8/12 (67%) were infected with an A subtype and 4/12 (33%) with a non-A subtype whereas in the group that was resistant for greater than 5 years (n=9) 4/9 (44%) were infected with subtype A and 5/9 (56%) with non-A subtypes. Numbers were limited and the differences seen between these two groups were not found to be significant (Chi-square p= 0.3085).

Section IB Summary:

In summary, a dynamic distribution of viral subtypes was seen within the ML cohort. Subtype distributions were significantly different within the different time periods studied. Contrary to the hypothesis these data did not indicate a significantly different subtype distribution within the subgroup tested compared to the general ML cohort. However, our sample size may not have been large enough to detect heterogeneous subtype distributions should they exist.

Section IC: Viral Genetics and the Impact on Disease Progression

A genetically diverse group of viruses were detected within the ML cohort. I next wanted to investigate whether these genetic differences would impact disease progression. The relationship between viral subtype and differences in disease progression was assessed. Disease progression in an individual infected with a virus containing a stop codon within HIV-1 protease gene was also evaluated. The impact of variability within HIV-1 protease, specifically at drug mutation sites, on treatment and clinical progression could not yet be determined as treatment is just now being made accessible within the cohort.

No subtype associated differences in disease progression:

An investigation followed to test the hypothesis that there would be viral genetic associated differences in disease progression. Specifically that infection with different subtypes would result in significant differences in disease progression. Based on the clinical information that was available, differences in disease progression were measured by Kaplan-Meier survival curves. For the purposes of analysis, an "event" was defined as CD4+ T cell count declining to below 400 cells/mm^3 and below 200 cells/mm^3 in specific individuals. A count of below 400 cells/mm^3 was used as a marker of progression because it falls below what is considered to be within the normal range of CD4+ T cell counts (normal range in adults is 400-1400 cells/mm^3) and also represents a value below which treatment is typically started (27) . A count of below 200 cells/mm^3 was used as it

is AIDS defining (105) (see methods for additional description of study criteria). I compared individuals infected with A, D, C subtype viruses or recombinant forms.

Individuals whose CD4+ T cell counts were above 400 cells/mm^3 at the first available date were included in the CD4+ T cell count decline to below 400 cells/mm^3 study. A total of 139 individuals met this initial criterion. The proportion of individuals within each of the A, C, D and recombinant groups who had CD4+ T cell counts that fell below 400cells/mm^3 was not significantly different (Chi-square p= 0.5491, Mantel-Haenszel p= 0.2713). The A subtype individuals had a mean time to drop below 400cells/mm^3 of 3735 days (point at which 50% were below was 3195 days), C subtype individuals had a mean of 1782 days (50% at 2361 days), D subtype had a mean time of 2189 days (50% at 2342 days) and the recombinants had a mean time of 2044 days (50% at 1996 days). Kaplan-Meier survival curve analysis showed no significant differences between A, C, D and the recombinant groups (Log Rank p= 0.0928, Wilcoxon p= 0.2542).

One hundred and eighty-three individuals were included in the study looking at CD4 + T cell counts that fall below 200 cells/mm^3. The percentage of individuals from each subtype whose CD4+ T cell counts fell below 200 cells/mm^3 was not significantly different (Chi-square p= 0.5307). The individuals infected with A subtype had a mean time of 4683 days (50% at 5118 days), subtype C individuals had a mean time of 4912 days (50% at 6329 days), D subtype mean time of 3609 days (50% at 3937 days) and recombinants had a mean time of 4315 days (50% at 5585 days). Kaplan-Meier survival curve analysis, of the event of CD4 + T cell counts declining to less than 200 cells/mm^3,

revealed no significant difference between the groups (Log Rank p= 0.9526, Wilcoxon p= 0.8899) (See Figure 8).

As I was studying individuals who were both seroconverters and those who entered the cohort positive, it was important to ensure that the data was not biased by significantly different initial CD4+ T cell counts within the different subtype groups. No significant differences were found between the average initial CD4+ T cell counts for subtypes A, C, D and R (recombinant) groups within the decline to below 400 cells/mm^3 study group (ANOVA p= 0.4215) or the decline to below 200 cells/mm^3 study group (ANOVA p=0.2245).

Conflicting data are present within the literature regarding potential differences in disease progression, particularly between subtype A and subtype D. Therefore, despite a relatively modest sample size, a comparison using just seroconverters was also done for those infected with subtype D and subtype A viruses. This analysis revealed no significant differences in the survival curves (event: drop below 400 cells/mm^3, A subtype n=22, D subtype n=5 (Log-Rank p= 0.1956, Wilcoxon p=0.5265); event: drop below 200 cells/mm^3, A subtype n=29, D n=6 (Log Rank p= 0.3368, Wilcoxon p= 0.6271)).

In summary, no significant differences between viral subtype and disease outcome were detected.

<u>Some evidence for changes in disease progression for subtype A over time:</u>

Kaplan-Meier survival analysis of the relationship between viral subtype and disease outcome, as measured by CD4+ T cell decline below 400 cells/mm^3 and below 200

cells/mm^3, revealed no significant differences. However, within the ML cohort we do know that significant fluctuations in viral distributions have occurred over time, particularly for subtype A. Since HIV has evolved over time, in part due to

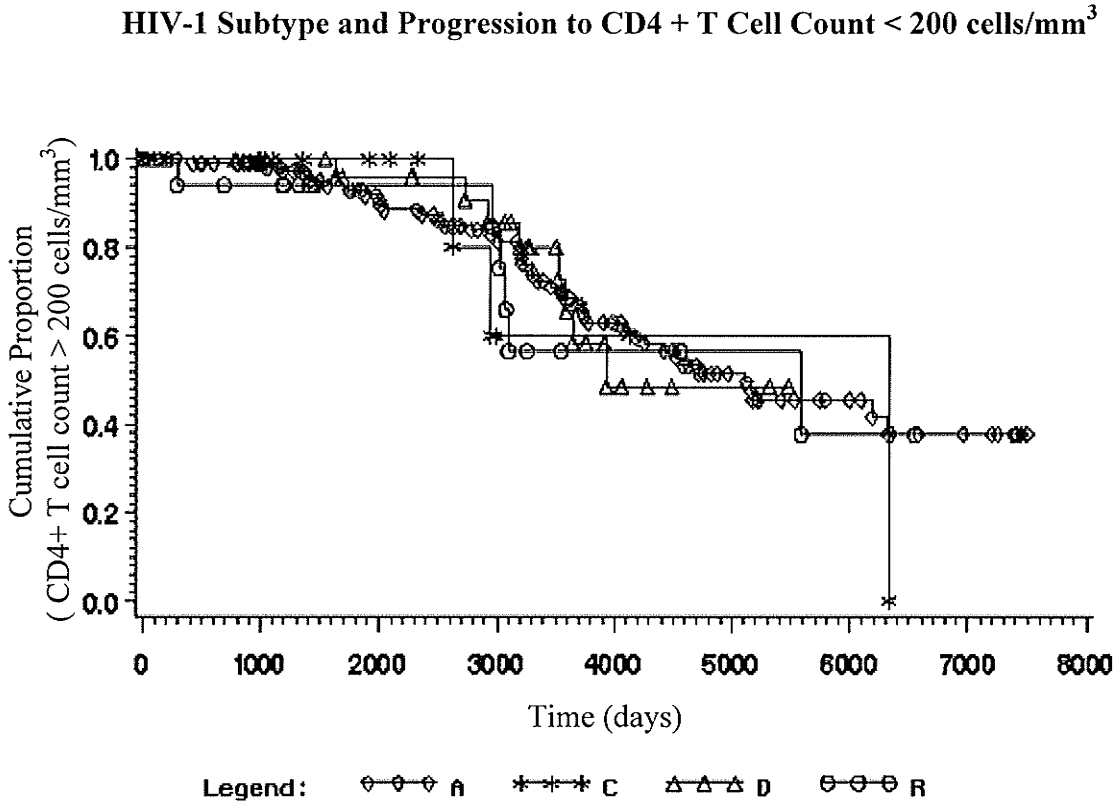

Figure 8: HIV-1 Subtype and Progression to CD4+ T Cell Count < 200 cells/mm^3

Kaplan-Meier Survival Analysis (outcome=drop below 200 cells/mm^3) for subtypes A, C, D, and recombinant viruses. No significant difference was found (Log Rank p= 0.9526, Wilcoxon p= 0.8899)

mutations away from CTL epitopes as a result of HLA restricted immune responses (58), it may follow that an increase in subtype A pathogenicity could also occur over time. An

investigation into differences in disease progression in individuals infected with subtype A within the different time periods was therefore performed. This was once again assessed by Kaplan-Meier survival analysis as described above.

A total of 92 individuals met the criteria for inclusion into the CD4+ T cell count decline below 400 cells/mm^3 study (1985-1989=33; 1990-1995=28; 1996-2003=31). Within the 1985-1989 time period the mean time to drop below 400 cells/mm^3 was 4569 days and time at which 50% of individuals within the study group were below was 3774 days. In the 1990-1995 time period the mean value was 2245 days with a median of 1890 days. The third period had a mean value of 1220 days. The time to drop below 400 cells/mm^3 was greatest for individuals within the 1985-1989 time period and was progressively less in the next two time periods. The Kaplan-Meier survival curve analysis revealed significant differences between the three time periods (Log rank p= 0.0281, Wilcoxon p= 0.0017) (See Figure 9).

A total of 118 individuals fit the criteria for inclusion in the time to drop below 200 cells/mm^3 study (1985-1989= 49; 1990-1995=37; 1996-2003=32). Survival curve analysis (outcome: drop below 200 cells/mm^3) did not reveal significant differences between groups but a trend was seen (Log-Rank p= 0.1204 and Wilcoxon p= 0.0739). The same pattern was seen as for the outcome of drop below 400 cells/mm^3 with those in the 1985-1989 period taking the longest to decline below 200 cells/mm^3 (mean= 4943 days), followed by the next two time groups (1990-1995: mean= 3625 days 96-03: mean=1517 days).

I next wanted to determine if these data may have been confounded by the fact that the initial group could have been infected for a shorter period of time prior to entry

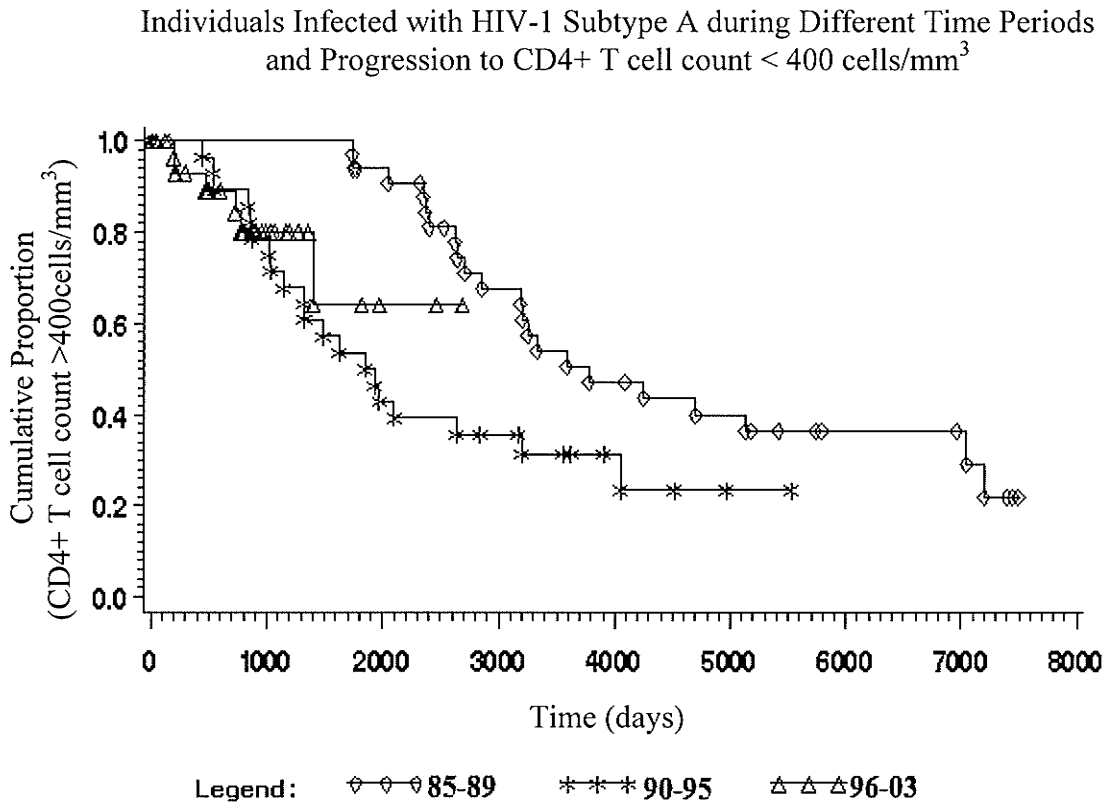

Figure 9: Subtype A Cohorts and HIV-1 Progression (Individuals Infected with HIV-1 at Different Time Periods and Progression to CD4+ T cell count < 400 cells/mm^3) Subtype A Kaplan-Meier survival curve analysis for individuals infected within three different time periods 85-89, 90-95, 96-03. Outcome event: CD4+ T cell count drop below 400 cells/mm^3 (Log-Rank p=0.0281, Wilcoxon test p=0.0017)

into the cohort (if entered positive). To address this issue, the time period from either the seroconversion date to first CD4 count (if a seroconverter) or from prostitution start date to first CD4 count was scrutinized to see if differences between the three groups existed and if any measured difference followed the same pattern as displayed by the survival curves. Overall the difference for women in the drop below 400 cells/mm^3 group was significantly different (ANOVA p=0.0069) but the pattern seen did not correspond to differences seen in the survival curves (1985-1989 mean time= 2479 days, 1990-1995 mean time = 1123 days, 1996-2004 mean time=1991 days). Moreover, when multiple comparisons were done, the difference between the 1985-1989 group and the 1996-2003 group was not found to be significantly different (ANOVA p=0.2451). For the drop below 200 cells/mm^3 group, the time lapse studied (mean time: 1985-1989: 2705 days, 1990-1995: 1072 days, 1996-2004: 1863 days) was also significantly different between groups (ANOVA p= 0.0001) but once again the pattern seen did not correspond to that seen for the survival curves.

In summary, no significant differences were found between the different viral genetic defined subtypes and disease progression. However, the data indicates that disease progression for subtype A may have changed over time.

Potential non-subtype related influence of viral genetics on progression:

Although no significant differences in disease progression based on subtype assignment were detected, other aspects of HIV-1 genetics may have a significant impact on clinical outcome. I decided to investigate disease progression in any individual in which a premature stop codon was detected within HIV-1 protease. Genetic analysis of

the protease aa sequence revealed only one individual, ML 1778, who had a premature stop codon. The codon was present at amino acid position 42 seen in both her 1995 and 2002 population based sequences. An assessment of her CD4+ T cell counts revealed that she is a long term non-progressor, defined as any individual whose CD4+ T cell counts remain above 400 cells/mm^3 for 8 years or more after seroconversion or entry into the cohort if entering positive. She entered the cohort positive in 1994. The following table (Table 5) reveals that her CD4+ T cell counts remain consistently well above 500 cells/mm^3, with a mean value of 799 cells/mm^3 over the last 11 years. Available viral load data reveals copies/mL that are well below average viral loads seen within the ML cohort (approximate average values are in the range of 10^4) (238). This stop codon, within a protein critical for virus particle maturation, may at least in part be responsible for this individual's non-progression. Studies here were conducted using proviral DNA. Future studies on this individual that include cloning and sequencing using plasma (viral RNA) samples are needed.

Section IC Summary:

Although no subtype specific differences in disease progression were seen, other viral genetic factors may significantly alter clinical progression.

Section I Key Findings:

A diverse molecular profile including multiple subtypes and frequent recombinant forms was observed. Furthermore, recombinant viruses were evident early in the epidemic within individuals of the ML cohort. Diversity between the amino acid

Table 5: Cell Counts and Viral Load Data for an Individual Infected with HIV-1 with a Stop Codon in the HIV-1 Protease Amino Acid Sequence (The stop codon was seen in 1995 and 2002 population based sequences). Dates, CD4+ T cell and CD8 + T cell counts as well as available viral load data are indicated.

Date (dd.mm.yy)	CD4 cells/mm^3	CD8 cells/mm^3	Available Viral Load Data (copies/mL)
04.10.94	770	770	9200
20.09.95	810	630	7600
29.11.95	690	630	590
15.04.96	660	700	3500
05.12.96	530	430	
04.09.98	780	610	440
12.02.99	1033	745	
21.02.00	842	1123	
31.10.00	993	901	
08.02.01	772	1196	
02.08.01	1138	916	<125
13.11.01	747	548	1200
01.08.02	623	492	<125
24.10.02	732	746	1500
07.08.03	657	560	6500
03.11.03	689	694	380
02.04.04	963	732	
17.06.05	952	887	

sequences of HIV-1 protease for the different subtypes, with respect to resistance mutations, was also observed. HIV-1 viral subtype distributions within the cohort were significantly different over time. The virus subtype profile seen for late seroconverters however did not differ significantly from the rest of the individuals within the cohort whose viruses were subtyped.

No associations between viral subtype (A, C, D, Recombinant) and disease progression were found but evidence does suggest that disease progression for subtype A may have changed over time. Non-subtype associated genetic characteristics (premature stop codon seen here) may result in long term non-progression. This first section has provided us with a baseline understanding of single subtype infections within women of the ML cohort. The next two sections will focus on an investigation of more complicated superinfection and dual infection cases.

SECTIONS II and III: Introduction

The results in Section-I provide an understanding of HIV-1 infections with single viruses within the ML cohort. Individuals however may also become dually infected with two distinct, unrelated viruses. This can occur either through HIV-1 co-infection or superinfection. As is described within the introduction, if it can be shown that more than one HIV-1 virus was acquired from a dually infected host at the time of infection or following an initial infection but prior to the establishment of an immune response, this can be referred to as HIV-1 co-infection (86). Superinfection however, is defined as the re-infection with a second HIV-1 virus only after the establishment of an immune

response to the first (184) . Sections II and III both focus on investigating infections with multiple HIV-1 viruses in women within the ML cohort.

SECTION II: Superinfection Study

This next section of the study was driven by the **hypothesis, that infection with HIV-1 will provide some protection against superinfection**. Multiple circulating subtypes, frequent exposure to HIV and well defined data on HIV incidence within the cohort make the ML cohort an ideal one to study HIV-1 superinfection. In order to test this hypothesis the first objective was to determine the incidence of superinfection using a nested cohort of the ML sex workers.

I successfully studied samples from 135 women at two or more time points during the period that they were infected with HIV. The study covered a total of 736.7 person years of observation. For the 135 women more than 270 samples were studied (>2 in many cases) and 73 of these samples, representing 44 of the women, required TA cloning. On average 30 clones were successfully sequenced per sample. Sequence data revealed that the cloning was mostly required due to variations in the diverse p6 region (insertions /deletions) that created difficulty in attaining good population based sequence data. In other cases cloning was necessary due to infection with more than one virus at a single time point. In total four superinfection cases, whose sample identities were confirmed by HLA typing, were detected. Additional putative superinfection cases were detected but were later found to be false due to HLA identity discrepancies at one or more of the time points studied. These cases were subsequently eliminated from the study (and are not included in the denominator).

Additionally, within this nested cohort study (n=135) one dual infection case (ML 293) was also identified but sequencing at seven different time points dating back as early as samples were available revealed two distinct viruses at all time points.

Superinfection Cases Detected Within Nested Cohort Study:

The four cases of superinfection that were identified as well as the one dual infection case found within the study group will be detailed below.

Superinfection case I:

Demographic and epidemiologic parameters for superinfection case I are listed in the table below (Table 6).

Table 6: Superinfection Case I: Demographic and Epidemiologic Parameters

ML	1787
Cohort entry date	07.02.1995 (positive on entry)
Year of birth	1961 : 34 at entry
Prostitution start date	1983 : 12 years on entry
Sex partners per day	Range: 2-4 per day
Condom use (range)	Most of the time-always (but not with regular partners)
HLA Type	HLA-A: A3001, A6802 HLA-B: B1302, B4201 HLA-C: C0602, C1701

This individual was studied at five different time points from 1995 to 2003. A minimum of thirty clones were successfully sequenced for a region of Gag/Pol encompassing p1, p6, protease and part of the RT gene for each time point. Phylogenetic and recombination analysis revealed that within this region all sequences were subtype C in 1995 and in 1997. A/C recombinant viruses were detected in 1998, 1999 and 2003.

Sequence similarity between the C subtype viruses detected in 1995 and 1997 and the C portion in the recombinant form suggests that an in vivo recombination event, and not a superinfection with a different A/C recombinant virus, occurred. Detection of the parental A subtype superinfecting virus in 1999 also lends support to this interpretation (see Figure 10).

In 1998, A/C recombinant viruses were first observed within the region under study. Initial recombination screening of clones was done using the RIP program and the Simplot program was subsequently used for fine mapping of the recombinant breakpoints (refer back to the methods section for a detailed description of recombination analysis). A subtype C consensus sequence was made from the subtype C clones detected in 1995 and 1997 and was used as the C parental reference in the recombination analysis. The A type virus detected in 1999 was used as the A parental reference in the recombination analysis. These references aided in determining the exact location of the recombination breakpoint. Recombination assessment was conducted for every clone that was sequenced from this individual and in the other cases that follow. Below is an example of one clone for ML1787 and the corresponding analysis. In this situation a B subtype virus (B.FR.83) was used as the outlier reference (See Figure 11A and 11B). Following determination of the recombination breakpoint phylogenetic analysis of the separate regions was performed (see Figure 12A, 12B, 12C).

Although the majority of recombinant viruses found were Type I (breakpoint at 293 within the sequence fragment or 2378 within HXB2), sequences with four additional different breakpoints were also found. These were all A/C recombinant viruses with ideal

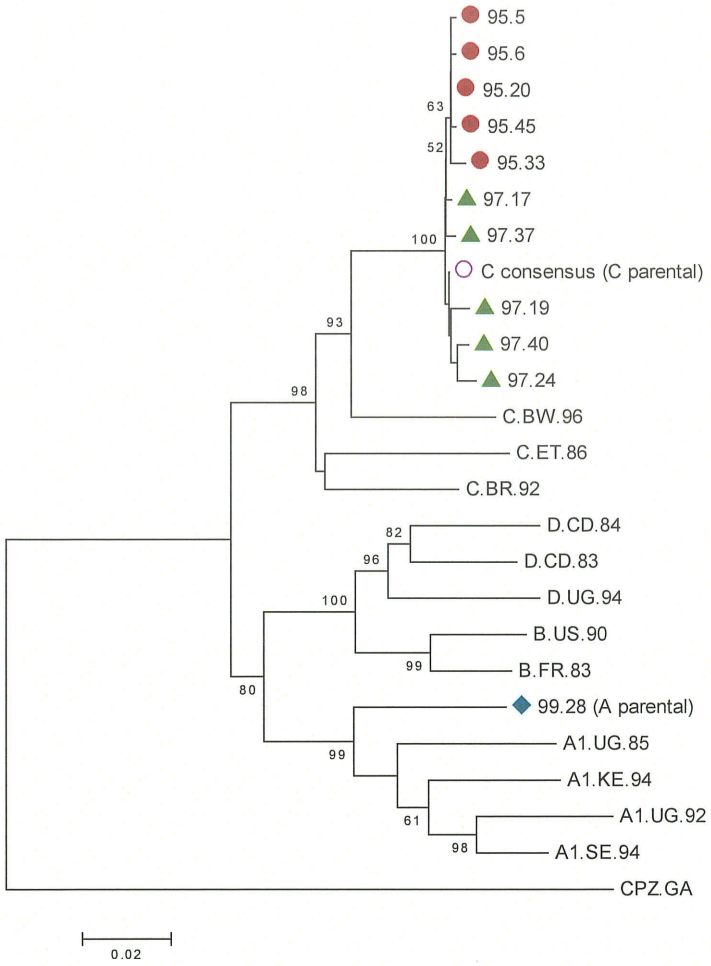

Figure 10: **Superinfection Case I Phylogenetic Analysis**: This figure displays clones from ML1787 for 1995(circle) and 1997 (triangle). They are all non-recombinant, subtype C viruses. Only a select number of representativeclones from each individual are shown here. A consensus for the subtype C virus was made based on 30 clones from 1995 and 30 clones from 1997 and is also displayed here (circle). The 99.28 is the A subtype virus detected in 1999 that is the A parental sequence (diamond) and was later used for the recombination analysis.

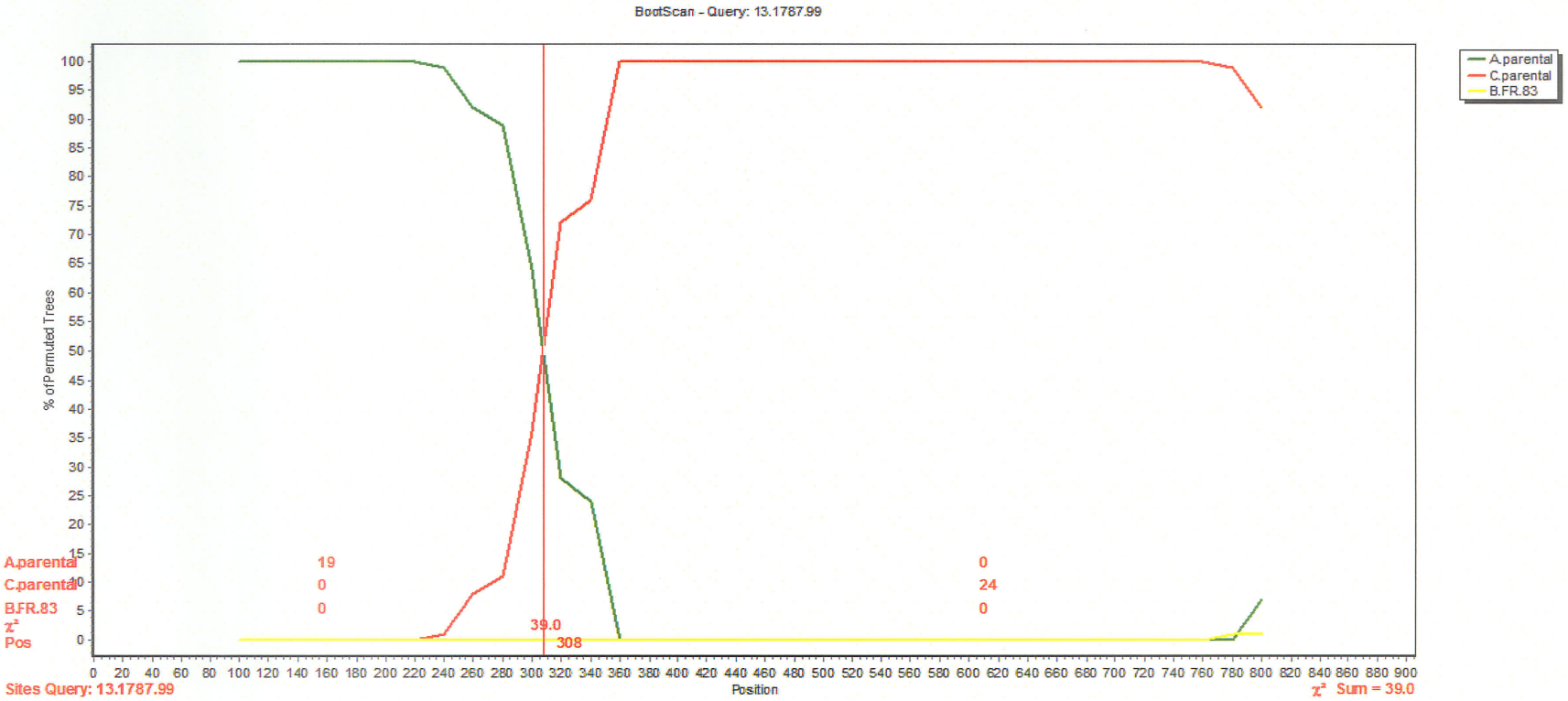

Figure 11A: **Superinfection Case I Simplot Output:** This figure displays the Simplot output comparison of our query sequence (clone from ML1787) to the parental A and parental C sequences with use of a subtype B variant as an outlier. This reveals that a maximum Chi-square value of 39.0 is reached ($p<0.0001$) at a breakpoint range of 276-314 within the fragment sequenced.

S=ML1787 1999clone 13 SequenceThere are 3 possible bifurcating trees:
A= A parental
C= C parental
O= Outlier (B.FR)

The following sites support the indicated trees:

Position	S	A	C	Outlier	Tree #	Position	S	A	C	Outlier	Tree #
26	G	G	A	A	1	383	C	T	C	T	2
50	C	C	T	T	1	404	G	A	G	A	2
62	T	T	C	C	1	452	A	T	A	T	2
66	A	A	G	G	1	458	A	T	A	T	2
89	T	T	G	G	1	459	C	T	C	T	2
93	T	T	C	C	1	485	A	G	A	G	2
94	G	G	A	A	1	554	G	A	G	A	2
97	A	A	T	T	1	574	C	A	C	A	2
110	T	T	C	C	1	582	G	A	G	A	2
112	G	G	A	A	1	583	A	C	A	C	2
118	T	T	C	C	1	609	A	T	A	T	2
137	G	G	A	A	1	635	T	C	T	C	2
141	G	G	A	A	1	638	C	T	C	T	2
197	T	T	C	C	1	659	A	G	A	G	2
207	G	G	A	A	1	662	G	A	G	A	2
213	G	G	A	A	1	665	G	A	G	A	2
230	A	A	G	G	1	677	G	A	G	A	2
248	G	G	A	A	1	701	G	A	G	A	2
275	C	C	A	A	1	734	G	A	G	A	2
314	A	G	A	G	2	755	C	T	C	T	2
353	A	G	A	G	2	835	G	A	G	A	2
354	G	A	G	A	2						

Figure 11B: Superinfection Case-I Informative Sites for Recombination Analysis: Simplot was used to determine the ideal breakpoint range for sequence clones. The breakpoint range was determined by choosing the position that gives the highest Chi-square (X^2) value. Here a X^2 sum value of 39.0 is seen (p-value < 0.0001) (Figure 10A). The informative sites shown above reveal that the breakpoint is between nucleotide position 276 and position 314 of the sequence fragment under evaluation. A midpoint value of nt 293 (or nt 2378 within the HIV HXB2 viral genome) is used when confirming each segment of the recombinant virus by phylogenetic analysis.

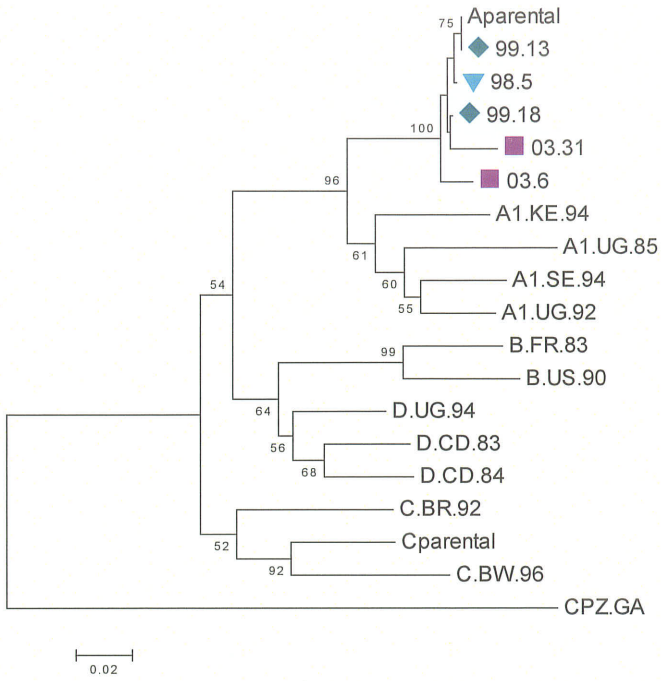

Figure 12A: SI Case I Dendogram for Recombination Analysis (part 1) revealing that the first part of the recombinant form is subtype A.

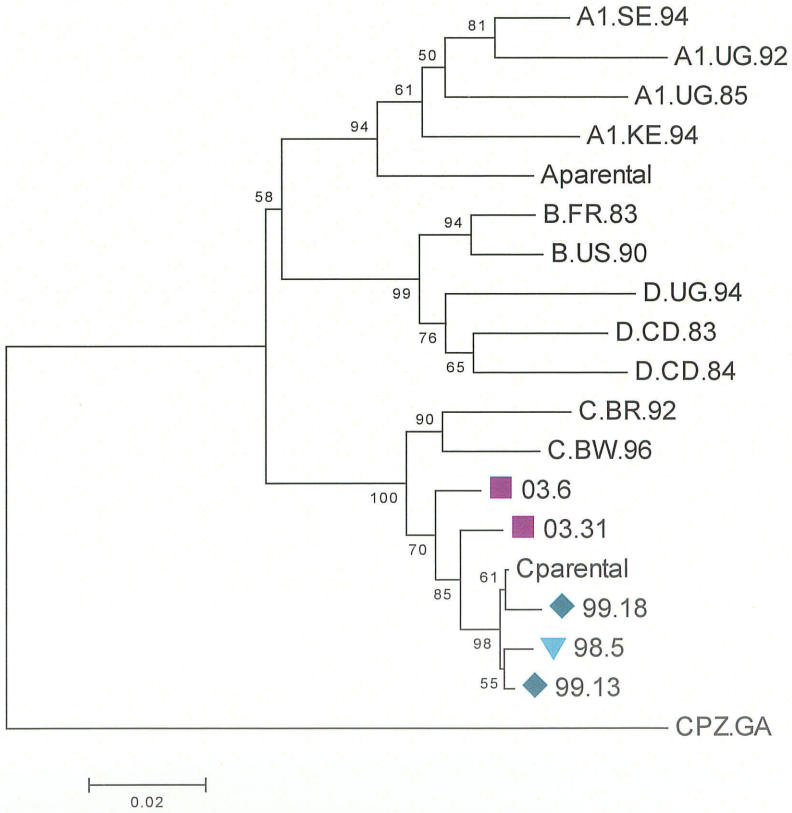

Figure 12B: SI Case I Dendogram for Recombination Analysis (part 2) revealing that the second part of the recombinant form is subtype C.

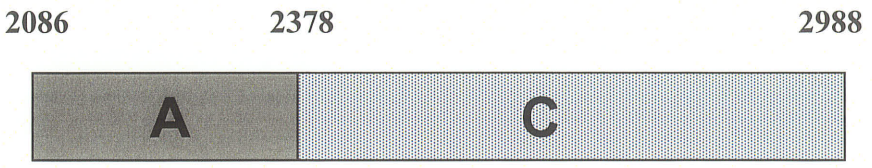

Figure 12C: SI Case I: Diagram of Recombinant Form A/C

Figure 12: In 1998, 1999 and 2003 recombinant viruses were detected that were subtype A from 1-292 and then subtype C from 293-903 within the region sequenced. **A.)** Dendogram with select clones and references reveals that the first section of the recombinant form (nucleotides 1-292 within the fragment) is subtype A. Different symbols and colours are used here and in later figures to simply aid in identifying clones from different time points studied. Here 2003 clones are indicated by a square, 1999 by a diamond and 1998 by an upside down triangle. **B.)** The dendogram contains select clones and references. It reveals that the second section of the recombinant form (nucleotides 293-903) is subtype C **C.)** Depicts the recombinant form and breakpoint at nucleotide 293 (nt 2378 by HXB2 numbering).

breaks at: Type II (breakpoint at 468), Type III (breakpoint at 603), Type IV (breakpoint at 534) and Type V (breakpoint at 379).

Based on sequencing of 30 clones at each time point recombinant viruses were initially detected in 1998 and made up 17% of the sequences seen. The subtype C virus accounted for the remaining 83%. In 1999, approximately 14 months later, the proportion of sequences that were recombinant increased to 94%. By 2003 the recombinant strain made up 100% of the sequences detected (see Figure 13).

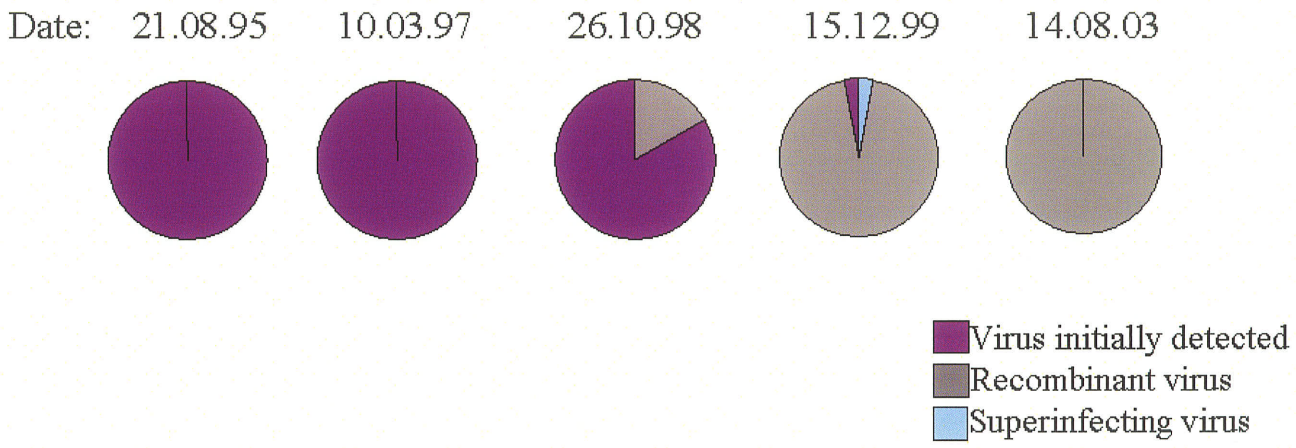

Figure 13: SI Case I: Proportion of Viral Forms Present at each Date:

The pie charts above depict the proportion of each viral type detected within this individual (ML 1787) over time at 5 different time points (dates: dd.mm.yy). At each time point 30 clones were analysed.

Superinfection case II:

Demographic and epidemiologic data on Superinfection Case II are detailed in the table below.

Table 7: Superinfection Case II: Demographic and Epidemiologic Parameters

ML	47
Cohort entry date (negative)	22.02.85
Seroconversion date	28.06.85
Year of birth	1955: 30 years old on entry
Prostitution start date	1982: 3 years on entry
Sex partners per day	3-7
Condom use (range)	Never to always
HLA Type	HLA-A: A2301, A7401 HLA-B: B5802, B8101 HLA-C: C0602, C1801

Three time points were available for study for this individual. Testing of additional dates using plasma samples stored at -20°C was attempted numerous times but was unsuccessful. In 1987 an A subtype virus was detected based on the Gag/Pol region under study but in 1995 and in 1999 an A/D recombinant form was found (see Figures 14 and 15). The recombinant form is depicted in Figure 16A. In 1987 only the A subtype form was detected based on over 30 clones but in 1995 and 1999 only the recombinant form was found (see Figure 16B). Sequence similarity between the A portion in the recombinant form and the initial A subtype suggests that a recombination event occurred following superinfection with a D type virus.

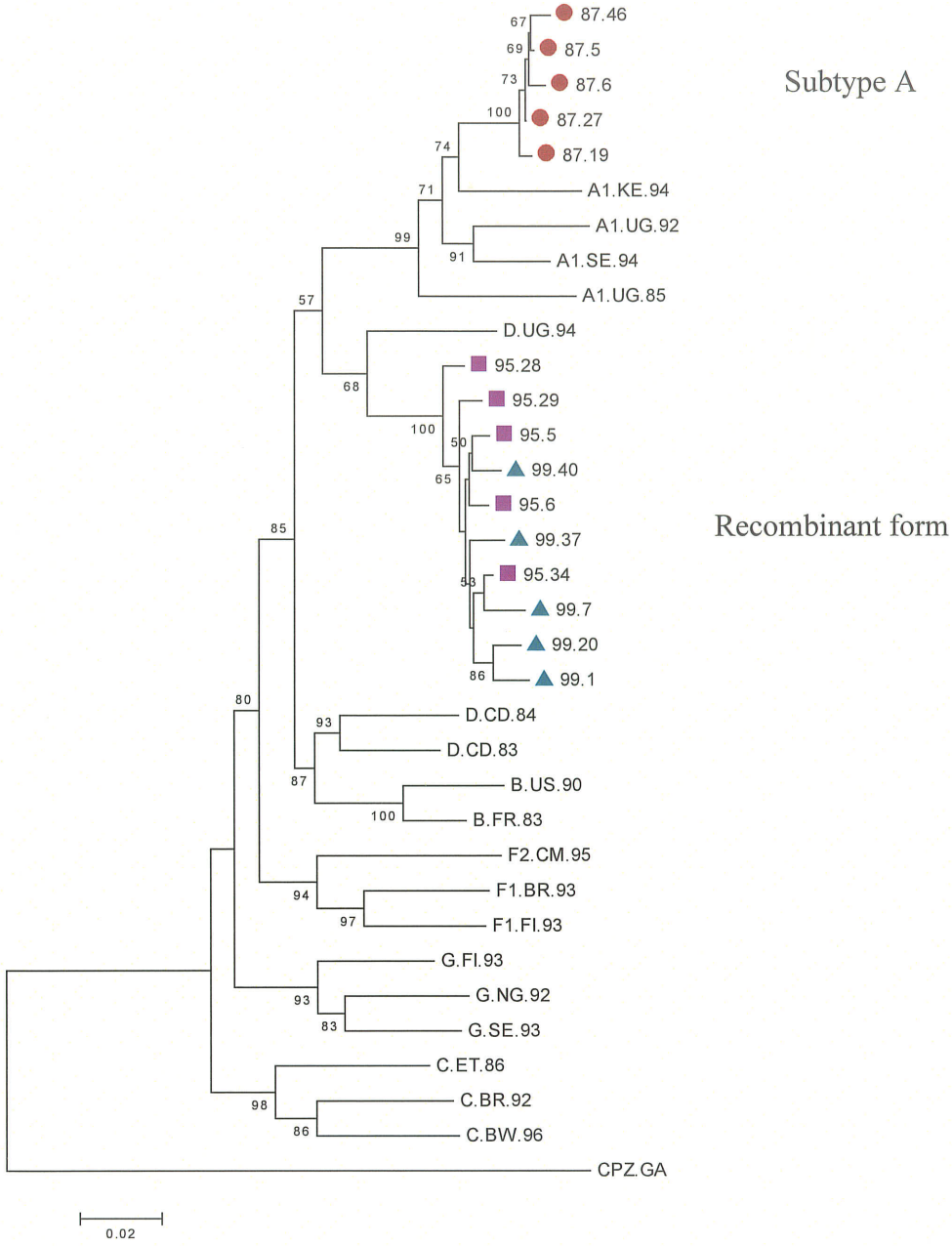

Figure 14: **Superinfection case II Phylogenetic Analysis**: the dendogram includes select clones from 1987 (circle), 1995(square), 1999(triangle) along with reference sequences. 1995 and 1999 samples are recombinant and therefore it was not possible to clearly classify these viruses based on phylogenetic analysis of the entire segment. Recombination analysis revealed viruses present in 1995 and 1999 were D/A recombinant viruses.

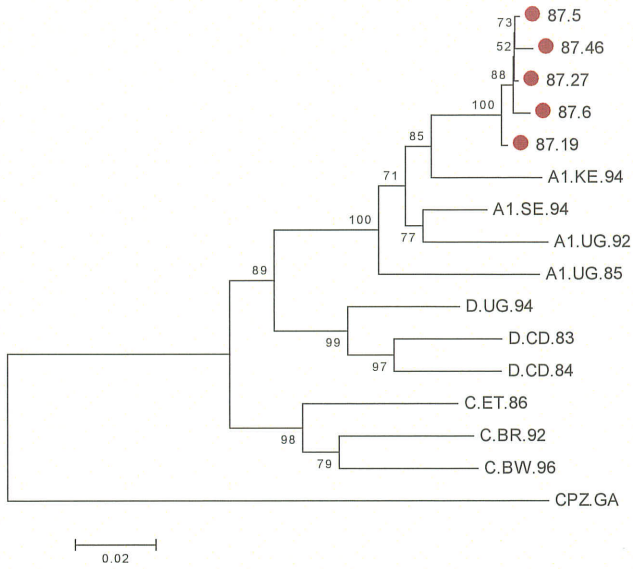

Figure15A: Superinfection Case II Dendogram for Full Length Clones (1987)

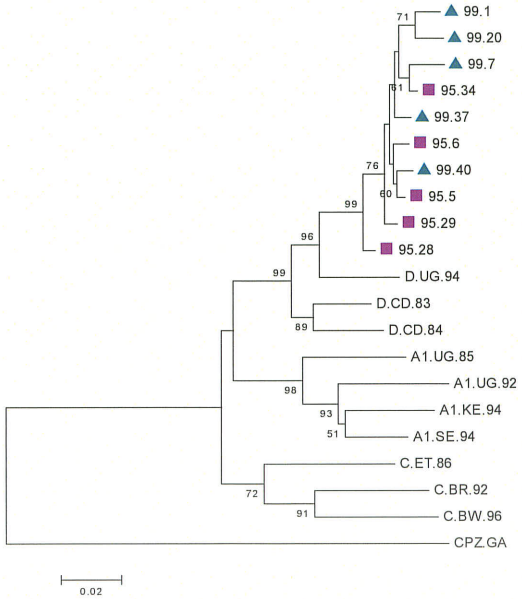

Figure 15B: Superinfection Case II Dendogram for Recombination Analysis (part 1): ML47 1995 and 1999 clones (nt 1-638)

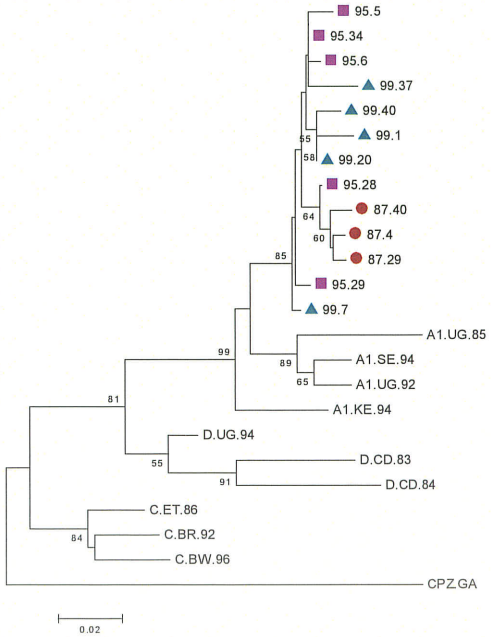

Figure 15C: Superinfection Case II Dendogram for Recombination Analysis (part 2): ML47 1987, 1995 and 1999 clones (nt 639-end)

Figure 15: **A.)** A few select clones from 1987 (circle) are shown here to display the fact that they are subtype A. **B.)** Dendogram with references reveals that the first part of the sequence fragment (nucleotides (nt) 1-638) for clones from 1995 (square) and 1999 (triangle) are subtype D. **C.)** Th second part of the sequences from 1995 and 1999 clones (nt 639-end) are subtype A. 1987 (circle) clones are included to illustrate how they cluster closely with the 1995 and 1999 clones.

Figure 16A: Superinfection Case II: Diagram of Recombinant Form D/A

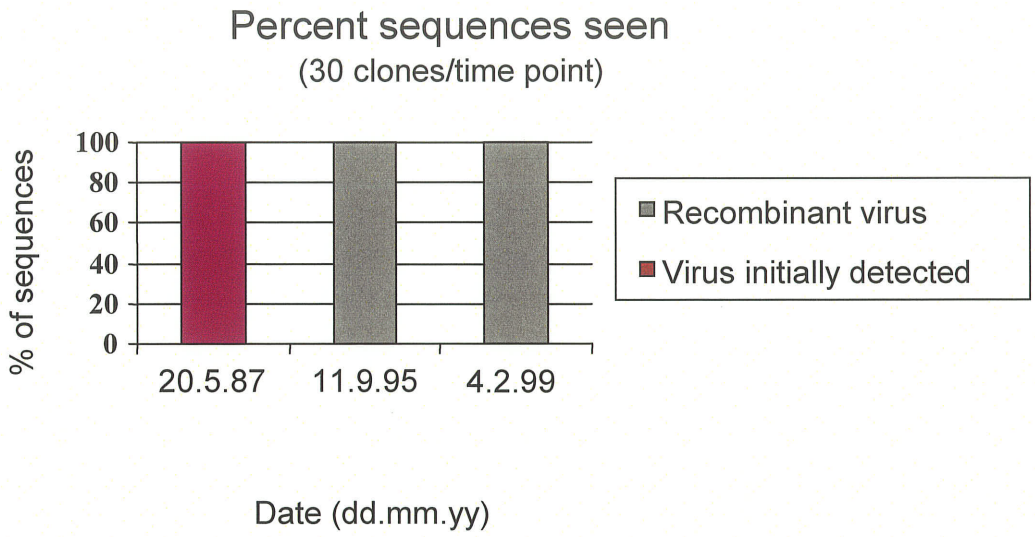

Figure 16B: Superinfection Case II: Proportion of Viral Forms Present at Each Date

Figure 16: **A.)** Depicts the recombinant form detected. It is a D/A recombinant within the region sequenced, with a breakpoint at nucleotide 2727 (HXB2 numbering). **B.)** Depicts the proportion of each virus form at the different time points. The non-recombinant form was seen in 1987 but in 1995 and 1999 the recombinant form is the only form detected.

Superinfection case III:

The demographic and epidemiological data for superinfection case III is detailed in the table below.

Table 8: Superinfection Case III: Demographic and Epidemiologic Parameters

ML	1449
Cohort entry date	11.02.92 (entered positive)
Year of birth	1964: 28 on entry
Prostitution start date	Mid 1981: approx.11 years on entry
Sex partners per day	5-10
Condom use (range)	Most of the time
HLA Type	HLA-A: A3201, Anew HLA-B: B4102, B4703 HLA-C: C0701, C1701

When sequences from this individual were aligned with references and sequences from all the other women in the study group phylogenetic analysis revealed that 1992 and 1996 samples from this individual did not cluster together. Two distinct viruses were evident (see Figure 17). In 1992 the sequence data reveals an A subtype virus. Sequence data from 1996 revealed that the first portion of the region sequenced was most similar to subtype D while the remainder was most similar to subtype A. Recombination analysis indicated a breakpoint at nucleotide 221 (2312 by HXB2 numbering) within the sequence. Phylogenetic analysis of the short region (nucleotides 1-220) revealed that sequences from 1992 clustered with A reference sequences while the 1996 sequences clustered closest with D references (bootstrap values are less than 70 because the region is not large enough to attain high values). Sequence data from both 1992 and 1996 was subtype A from nucleotide position 221 to the end of the fragment. Phylogenetic analysis revealed that sequences from the different dates for this region (nt 221 to the end) cluster

and intermingle with each other (see Figure 17B). This suggests that the individual was infected with an A type virus and later became superinfected with a D type virus. A recombination event then occured and resulted in the formation of an A/D recombinant virus. An analysis of all clones for this individual at both time points revealed that in 1992 the subtype A form was the only form present (100% of clones) while in 1996 the D/A recombinant form alone was detected (100% of clones).

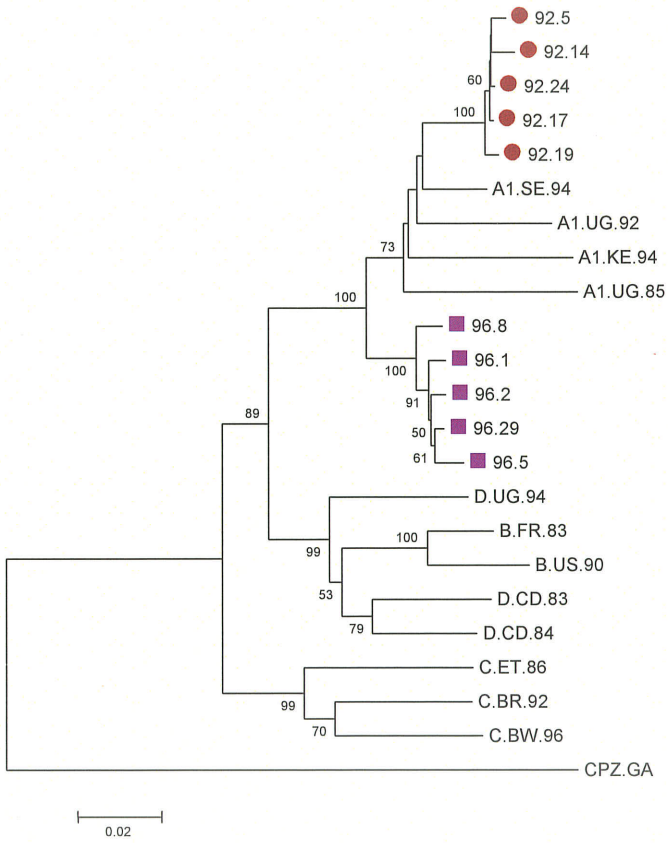

Figure 17: **Superinfection Case III Phylogenetic Analysis**: Select clones and reference sequences for the Gag/Pol region are shown here. The 1992 sequences (circle) are subtype A and the 1996 sequences (square) are A/D recombinant sequences as determined by recombination analysis.

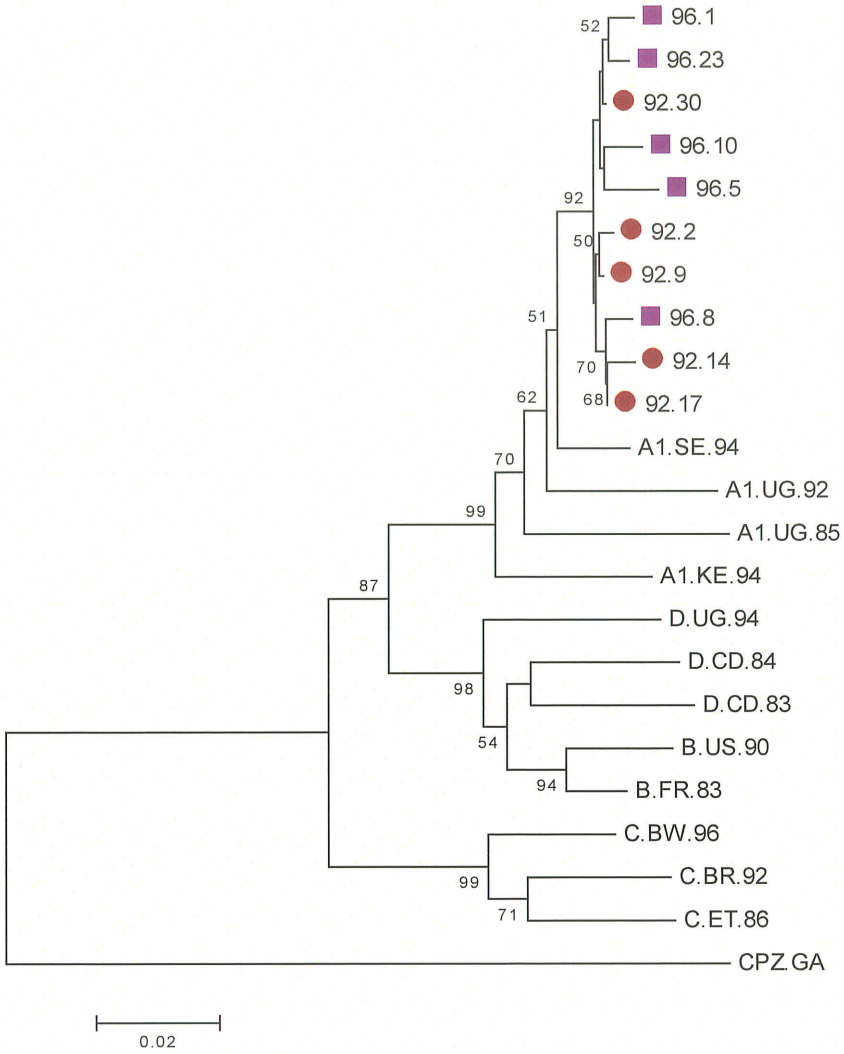

Figure 18: Superinfection case III: Subtype A Region of Recombinant Form

Region 221-end (nucleotides 2312-2988 by HXB2 numbering) was analyzed here with select references and clones from both 1996 (square) and 1992 (circle). This region is subtype A for both the 1996 and 1992 clones.

Superinfection case IV:

The demographic and epidemiologic data for superinfection case IV is detailed in the table below (Table 9).

Table 9: Superinfection Case IV: Demographic and Epidemiological Parameters

ML	1295
Cohort entry date	03.07.90 (entered positive)
Year of birth	1967: 23 on entry
Prostitution start date	1990: had just started on entry
Sex partners per day	2-5
Condom use (range)	Most of the time –always
HLA Type	HLA-A: A6601, A7401 HLA-B: B070201, B570301 HLA-C: C1701, C new

Clones sequenced for ML1295 from 1996 and 2002 samples were aligned along with reference sequences and sequences from all the other ML women within this study and phylogenetic analysis was carried out. Clones from 1996 and some of the clones from 2002 were more closely related to each other than any other ML or reference sequence. These sequences were subtype A (see Figure 19A). However, it was discovered that other 2002 clones, found witin a different part of the dendogram, clustered closest with other ML A/D recombinant viruses. Recombination analysis revealed a likely A/D recombinant form with a breakpoint at nucleotide 212 within the fragment (once again additional phylogenetic analysis was complicated by the small size of the fragment). Below is a dendogram which depicts the two distinct virus types seen in 2002, it consists of clones from 1996 and 2002 along with some of the reference sequences (see Figure 19B).

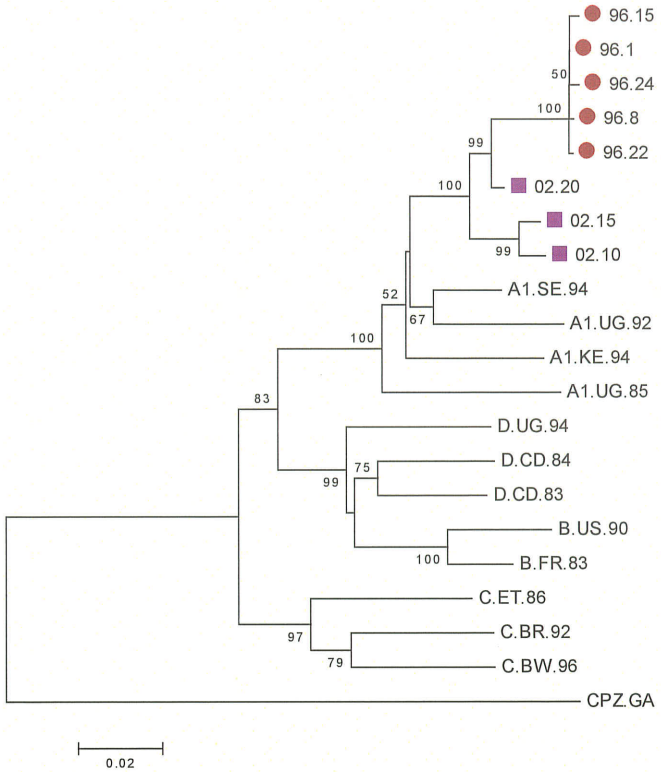

Figure 19A: Superinfection Case IV Non-Recombinant Forms

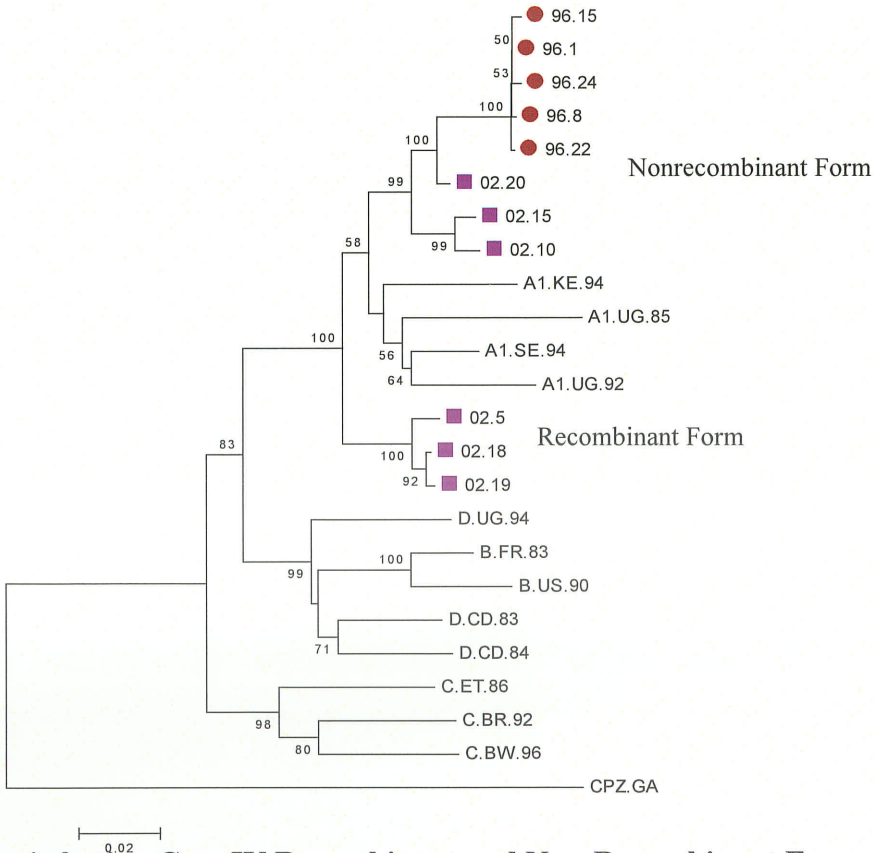

Figure 19B: Superinfection Case IV Recombinant and Non-Recombinant Forms

Figure 19:**A.)** Superinfection case IV (ML1295): The dendogram displays a few select non-recombinant subtype A clones from 1996 (circle) and 2002 (square) along with reference sequences. **B.)** A dendogram with both 2002 recombinant (A/D) forms and 2002 and 1996 non-recombinant (A) forms. Only the non-recombinant form is present in 1996 but both recombinant and non-recombinant forms are present in 2002.

Within this individual only subtype A was seen in 1996; however, 50% of the clones sequenced in 2002 were subtype A and the other 50% were an A/D recombinant form (Figure 20).

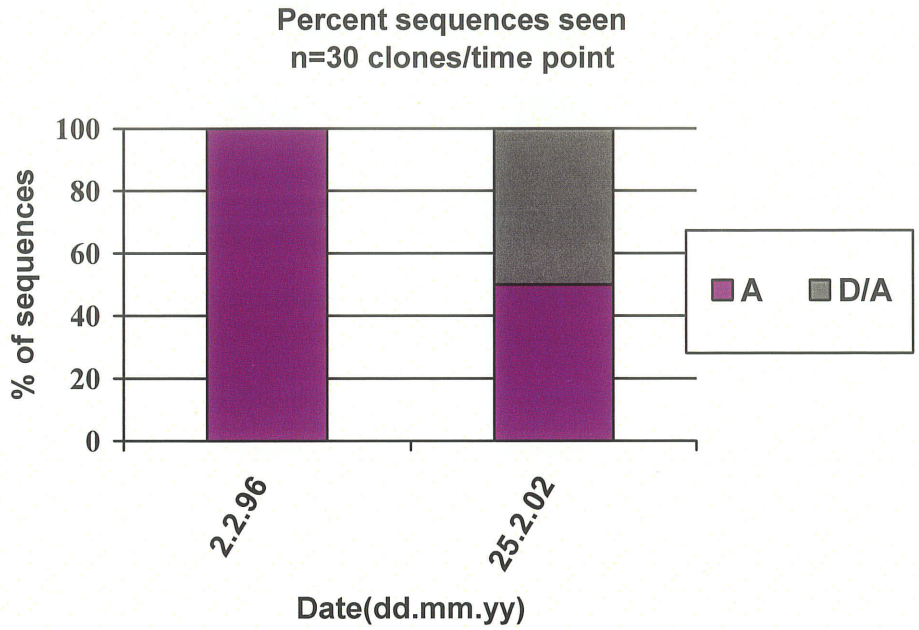

Figure 20: Superinfection Case IV: Proportion of Viral Forms Present at Each Date (subtype A and the D/A recombinant forms) 100% of clones were subtype A in 1996, but in 2002 50% were recombinant.

Superinfection Case Summary:

Table 10: Superinfection Case Summary

Superinfection Cases	Initial Sequence subtype (p1,p6,pro, part RT)	Sequences seen later	Recombinant viruses seen
I	C	A/C,A,C (only A/C at last time point)	Yes
II	A	A/D	Yes
III	A	A/D	Yes
IV	A	A, A/D (equal proportions)	Yes

Dual Infection (DI) Case I (detected in nested cohort study):

In addition to the four superinfection cases that were detected within the nested cohort study one individual was identified who was dually infected. Two different virus types were detected at every time point tested with the exception of the most recent sample. Demographic and epidemiologic data for this individual is detailed in the table below.

Table 11: Dual Infection Case I: Demographic and Epidemiologic Parameters

ML	293
Cohort entry date	1985
Year of birth	1963: 22 at entry
Prostitution start date	1985: had just started on entry
Sex partners per day	1-10
Condom use (range)	Never –always
HLA type	HLA-A: A030201, A3201 HLA-B: B5802, B8101 HLA-C: C0602, C0804

This individual, ML293, cannot be classified as a superinfection case based on the samples available for study. Multiple peripheral blood mononuclear cell and plasma samples were available for this individual and samples were traced back to the earliest available dates. PBMCs from which DNA was isolated were available for August 1995, November 1995, 1996, 1997 and 1998. RNA was also successfully isolated from the 1994 and 1993 plasma samples of this patient. In total, this individual was studied at 7 time points and more than 30 clones were sequenced at each point. As mentioned, two distinct viruses were detected within this individual (see Figure 21). Type 1 is an A type virus (There is a possible small segment of subtype D seen near the beginning of the region under study (81-175) but this could not be definitively confirmed by phylogenetic

analysis, therefore the virus will be classified as subtype A). Type 2 is a D/A recombinant virus. A breakpoint range of nucleotides 444-473 with a midpoint value at 459 within the fragment was determined based on Simplot analysis as described previously (see Figure 22A). Ideal breakpoint ranges differed slightly, but overlapped, between clones.

We studied the distribution of these two virus types within this individual over time and found that both viruses were present at every time point except in 1998 (the last sample) where only the Type II recombinant form was detected (see Figure 22B). Fluctuations were seen in the proportion of each virus type that was present but at all time points except one, the Type II recombinant form predominated.

Superinfection Nested Cohort Study Confirmatory Assay:

A collaboration was established with Dr. A. Iversen from Oxford University which allowed a confirmatory heteroduplex tracking mobility assay (HTMA) to be performed on a subset of the samples tested. Unfortunately, The HTMA was unsuccessful due to the variability seen between the Kenyan sequences and the standard subtype references. Confirmatory analysis was done instead with additional sequencing of a second region of the viral genome (nucleotides 767-1921 by HXB2 designation). Paired samples from nearly 25% of the women in this study (39 women) were investigated revealing no additional cases of superinfection.

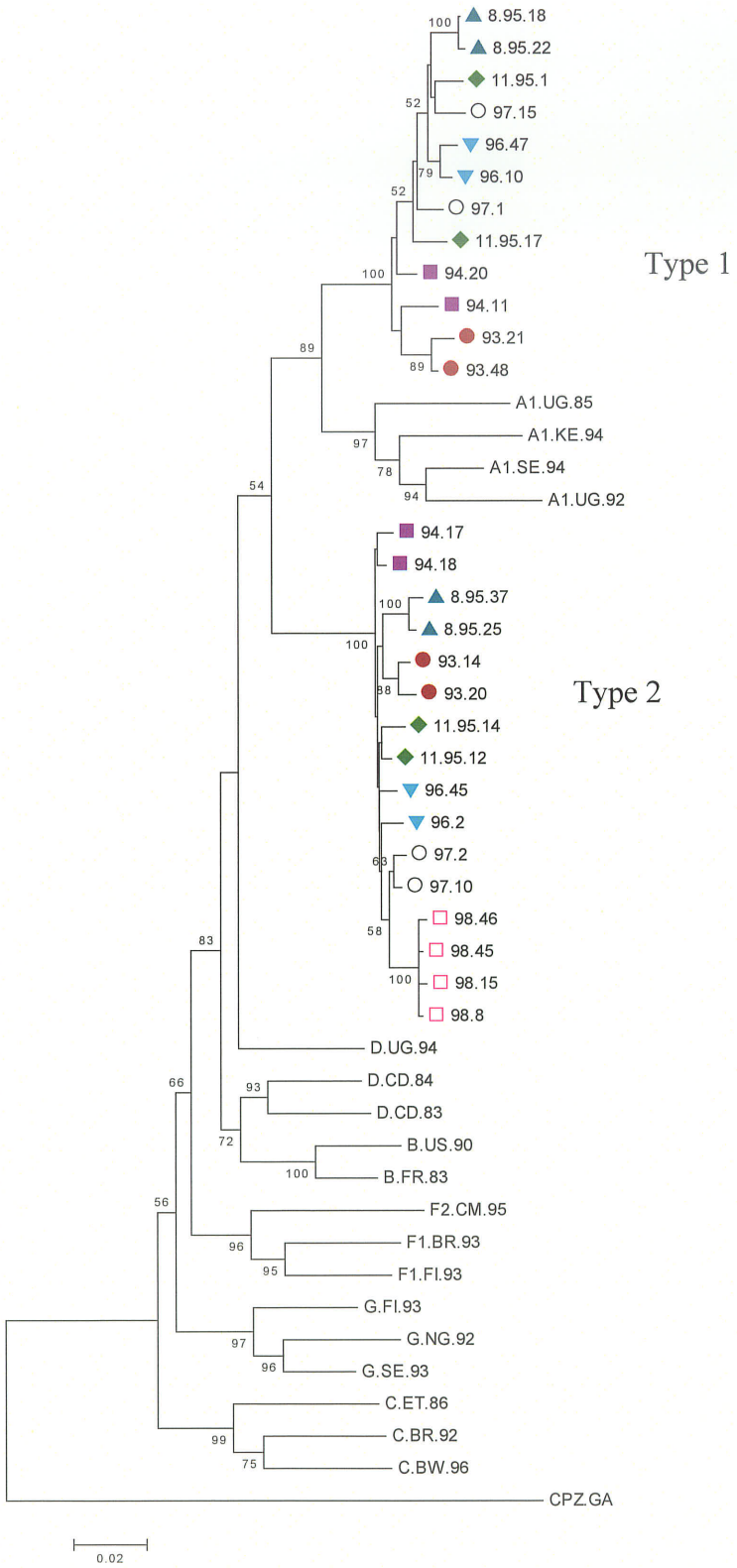

Figure 21: Dual infection Case I Phylogenetic Analysis

Figure 21: Dual Infection Case I Phylogenetic Analysis: Select clones from 7 different dates for ML293 are displayed along with reference sequences. Clones are assigned numbers and are labeled by year with each date studied given a different symbol/colour for ease of discriminating between the different dates. The ML293 clones cluster within two different groups. Type II is an A/D recombinant form while Type I is subtype A.

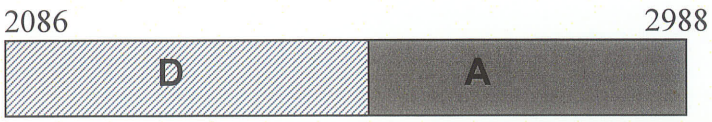

Figure 22A: Dual Infection Case I: Diagram of Recombinant Form D/A

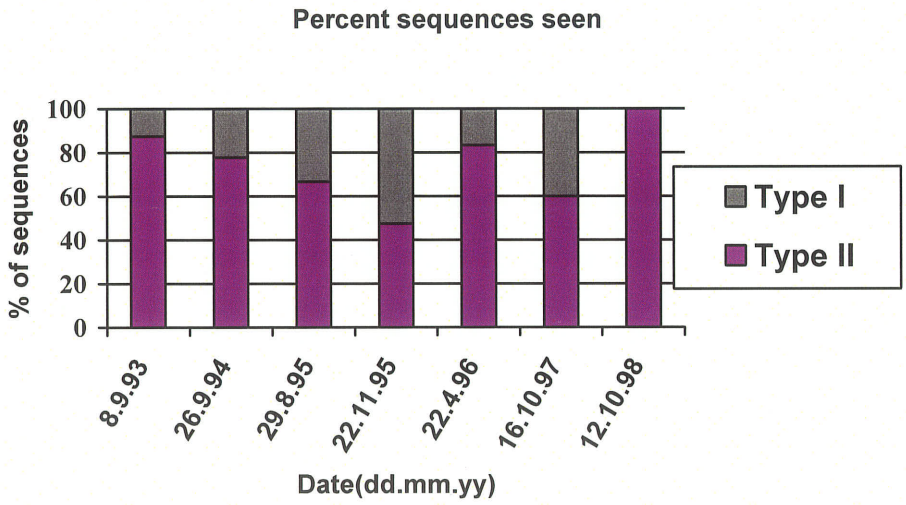

Figure 22B: Dual infection Case I: Proportion of Viral Forms Present at Each Date (Types I and II as described in text.)

Figure 22: A.) The D/A recombinant form (Type II) found within this individual

B.) The proportions of the two different virus types found over time. This is based on sequencing of more than 30 clones at each time point.

HIV-1 Infection Provides Some Protection Against Superinfection:

In total, 528 individuals were part of the entire cohort (this refers to those who entered the cohort HIV seronegative) and 124 of these individuals seroconverted over follow-up. For those who remained seronegative there were 619,793 person days of observation and 104,479 person days of observation for those who seroconverted giving a total of 724,272 person days of observation within the entire cohort. The superinfection study cohort was composed of 135 individuals (individuals who were HIV positive when first studied) and 4 of them became superinfected. Within the superinfection study cohort 261,363 person days were observed for those who did not become superinfected and 7527 person days were observed for those who did experience a superinfection event. There was a total of 268,890 person days of observation within this group. Epidemiological parameters pertaining to the entire cohort and the superinfection study cohort are detailed in the table below.

Table 12: Comparison of Parameters for the Entire and Superinfection Study Cohorts

Parameter	Entire Cohort (n= 528)	SI Study Cohort (n=135)
Initial age of sex worker at time of entry into cohort (years)	31.3+/- 6.8	34.3+/- 5.4
Prostitution duration at time of entry into cohort (years)	5.1 +/- 5.2	8.8 +/- 5.0
Follow-up (years)	3.0 +/- 3.9	5.0 +/- 2.9

Poisson regression:

Poisson regression was used to compare the incidence of superinfection within the superinfection cohort to the incidence of initial HIV infection in the entire cohort. This revealed an unadjusted odds ratio (OR)= 0.16, (CI 95%: 0.11-0.25), p<0.0001. This indicates that being HIV infected has a protective effect. In other words, HIV uninfected women are 6.14 times more likely to seroconvert compared to an HIV positive individual becoming superinfected with a second virus.

In order to control for multiple variables, multiple Poisson regression was done. Variables that were adjusted for were categorized as described in the methods section. The results from this analysis can be seen in Table 13.

Table 13: Multiple Poisson Regression Adjusted for all Variables

Variables	Adjusted Odds Ratio (95% CI)	p-value
HIV infection	0.18 (0.12-0.28)	<0.0001
Seldom/never use condoms	3.27 (2.60-4.11)	<0.0001
Sex partners/day (>/=5)	1.76 (1.44-2.15)	<0.0001
Prostitution duration (<5 yrs.)	1.20 (0.84-1.71)	0.32
Prostitution duration 5-15 years	1.42 (0.98-2.06)	0.07
Age <25	0.94 (0.72-1.24)	0.68
Age >35	0.38 (0.28-0.51)	<0.0001

The adjusted odds ratio, relating an initial HIV infection to protection from acquisition of a second HIV infection, was calculated to be 0.18 (CI 95%: 0.12-0.28). Individuals are 5.59 times more likely to acquire an initial HIV-1 infection than to become superinfected thus revealing that HIV infection provides a protective effect. During the times in which

individuals never or seldom used condoms they were 3.27 times more likely to become infected or superinfected. When individuals had 5 or more sex partners per day this behaviour was associated with a 1.76 fold increase in acquiring an initial HIV infection or a superinfection. Compared to a reference group of individuals that had a prostitution duration of >15 years (at time of entry into the cohort) a trend was seen which suggested that individuals who had a prostitution duration of 5-15 years were more likely to be infected or superinfected. Moreover, those who were older than 35 at time of entry were less likely to become infected or superinfected compared to a reference group of sex workers age 26-35.

Poisson regression was then rerun using those factors that were significant or had shown a trend (see Table 14). The protective effect of an initial HIV infection on superinfection remains evident (OR=0.18 (CI 95%: 0.12-0.28)).

Table 14: Multiple Poisson Regression Adjusted only for Significant Variables

Variables	Odds Ratio (95% CI)	p-value
HIV infection	0.18 (0.12-0.28)	<0.0001
Seldom/never use condoms	3.27 (2.60-4.10)	<0.0001
Sex partners/day >/=5	1.77 (1.44-2.16)	<0.0001
Prostitution duration (5-15 yrs.)	1.23(0.98-1.53)	0.07
Age >35	0.37 (0.28-0.49)	<0.0001

Cox's proportional hazard modeling

Although one type of statistical analysis would have been sufficient I decided to analyze the data using a different statistical method to further substantiate the findings. This next type of statistical analysis was Cox's proportional hazard modeling of risk

factors for HIV-1 seroconversion/superinfection. Before this analysis was performed a crude assessment using Kaplan-Meier survival curve analysis was used to statistically compare the two cohorts and the percentage of individuals that remained free of infection/ superinfection over time. The curves revealed significant differences between the two groups (Log-Rank $p<0.0001$ and Wilcoxon $p<0.0001$) (see Figure 23). A greater proportion of individuals acquired an initial HIV infection over time than an HIV-1 superinfection.

In the Cox's multivariate proportional hazard model, condom use and sex partners per day were used as time dependent covariates. Once again, being HIV infected remained an independent protective factor when all other covariates were controlled for (HR=0.15 (CI95%: 0.05-0.41)) (see Table 15).

Next, only covariates that had significant associations with either an increase or decrease in infection/superinfection were included in a parsimonious model (see Table 16). When all other significant variables are controlled for, being HIV positive was still shown to provide a protective effect (HR= 0.14 (CI95%: 0.05-0.39) against infection. Prostitution duration of less than 5 years is shown to be a risk factor for infection/superinfection while condom use and older age (>35) are associated with a decreased risk of infection.

Superinfection Study Analysis:

Both the multiple Poisson regression and Cox's multivariate proportional hazard modeling revealed that even after all other variables were adjusted for, being HIV infected provided for significant protection against HIV-1 superinfection.

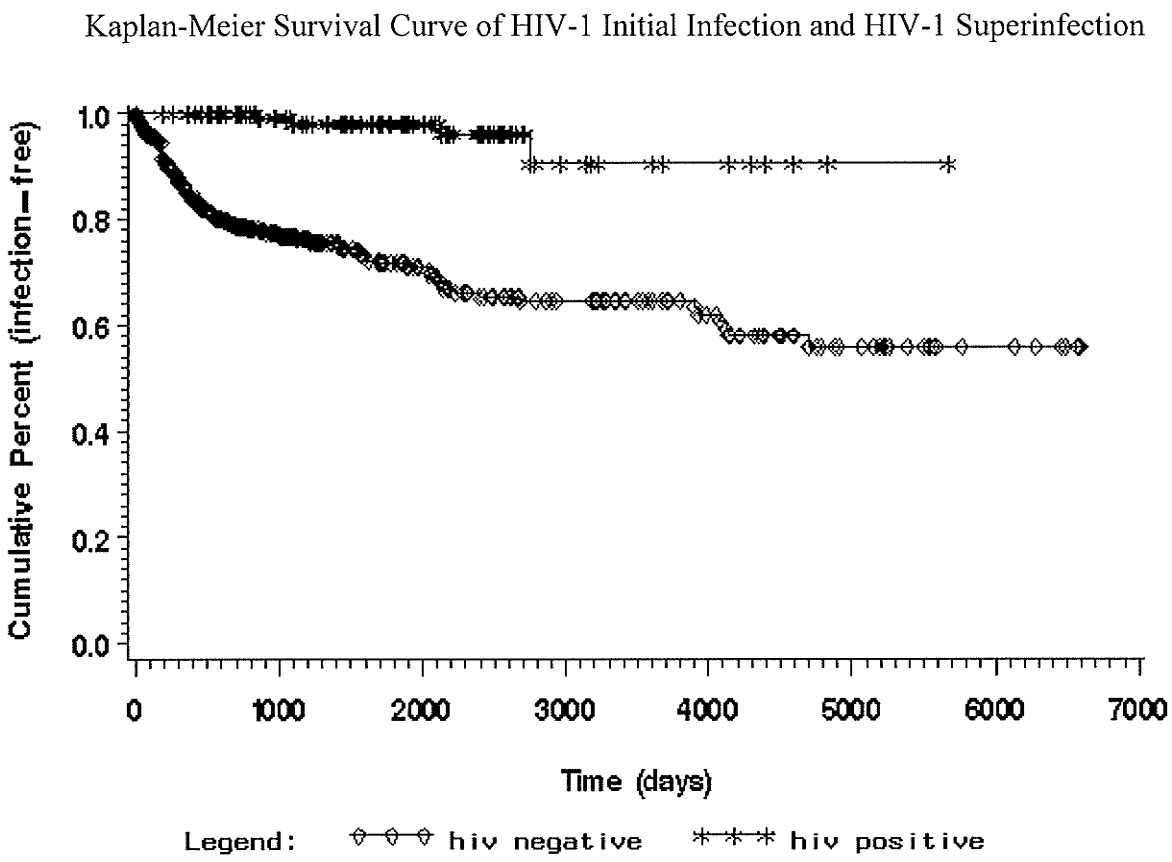

Figure 23: **Initial HIV Infection and Superinfection Kaplan-Meier Survival Curve Analysis** (Log-Rank $p<0.0001$ and Wilcoxon $p<0.0001$). HIV infection was investigated in the group that was originally HIV negative and HIV-1 superinfection was investigated in the group that was originally HIV positive.

Table 15: Multivariate Cox's Proportional Hazard Modeling

Variables	Hazard Ratio (95% CI)	p-value
HIV infected	0.15 (0.05-0.41)	0.0002
Prostitution duration (<5 years)	3.31 (1.94-5.65)	<0.0001
Prostitution duration (5-15 years)	1.40 (0.85-2.30)	0.18
Age (<25)	0.96 (0.46-2.03)	0.92
Age (>35)	0.24 (0.13-0.42)	<0.0001
Condom use	0.38 (0.26-0.56)	<0.0001
Sex partners per day(>/=5)	1.00 (0.70-1.43)	0.99

Table 16: Multivariate Cox's Proportional Hazard Modeling Using Only Significant Variables

Variables	Hazard Ratio (95% CI)	p-value
HIV infected	0.14 (0.05-0.39)	0.0002
Prostitution duration (<5 years)	2.92 (1.83-4.66)	<0.0001
Age (>35)	0.24(0.14-0.43)	<0.0001
Condom use	0.35 (0.24-0.51)	<0.0001

Conservative Estimate of Protection:

Further analysis of samples within the ML cohort did not reveal additional confirmed superinfection cases but other individuals who could only be defined as having a dual infection were identified. Using the entire group of women for which any viral sequencing had been performed revealed four confirmed superinfection cases (previously discussed) as well as a total of seven dual infection cases (one of which was discussed above). A crude assessment was done in which all of these cases were classified as potential superinfection cases, yielding 11 cases in the 213 individuals. This was compared to the 528 individuals who entered the cohort negative and of whom 124

became HIV infected. Chi-square analysis demonstrated a significantly higher proportion of individuals who acquired an initial infection compared to a superinfection (p-value <0.0001). Assessment of risk revealed that individuals were still 5.65 times more likely to acquire an initial infection compared to a superinfection (OR= 0.18 (CI 95%: 0.09-0.34)). The overall conclusion, that HIV-1 provides protection against superinfection remains consistent.

Section II Key Finding:

Infection with HIV-1 provides some protection against HIV-1 superinfection.

Section III:

As was the case for section II, this last section of research also focused on multiple HIV-1 infections. The aim was to test the hypothesis that **the majority of individuals with dual infections (superinfections/co-infections) will have evidence of recombinant viruses and that an increased disease progression will be seen.** To address the first part of this hypothesis, that recombinant viruses would be present in most dual infection cases, we analyzed the HIV sequence data from such individuals and also determined the proportion of recombinant and non-recombinant forms present.

There were four confirmed superinfection cases and one dually infected individual that were identified in the nested cohort study discussed in section II, in which individuals were studied at multiple time points. There were additional individuals studied but who were not part of the nested cohort study, and this was because only one sample was either available or successfully analyzed for these women. Within this group

of individuals another six women were identified as having dual infections. These additional cases will be discussed within this section. In total, of 213 individuals studied, at one or multiple time points, 11 women (5.16%) had evidence of infection with more than one virus (four of which were confirmed to be superinfection cases).

Dual Infections:

Among individuals for whom only one sample was available or successfully analyzed an additional 6 dual infections were identified. These cases will be described below starting with dual infection case II (dual infection case I, ML293, was described within section II (nested cohort study)).

Dual infection case II

ML825 was identified as being dually infected with both A and D type viruses based on sequencing of the Gag/Pol region (nucleotides 2086-2988 by HXB2) (see Figure 24). In addition to the A and D type viruses seen, two different recombinant forms were also detected (see Figure 25A). The A and D parental sequences within this individual were used to assess recombination breakpoints. Recombinant form Type 1 had a breakpoint range of nucleotides 626-689 (midpoint 658 (2743 HXB2 numbering)) within the sequence fragment and recombinant form Type 2 had a breakpoint range of nucleotides 501-580 (midpoint 541 (2625 HXB2 numbering)). Based on the analysis of 30 clones, 50% were subtype A, 40% subtype D and 10% were recombinant. A region of envelope (v3-end of gp120, 7008-7817 HXB2) was also sequenced (30 clones) for this

individual. Subtype A and D forms at a frequency of 16.7% and 83.3% respectively were seen but no recombinants were detected within this region (see Figure 25B).

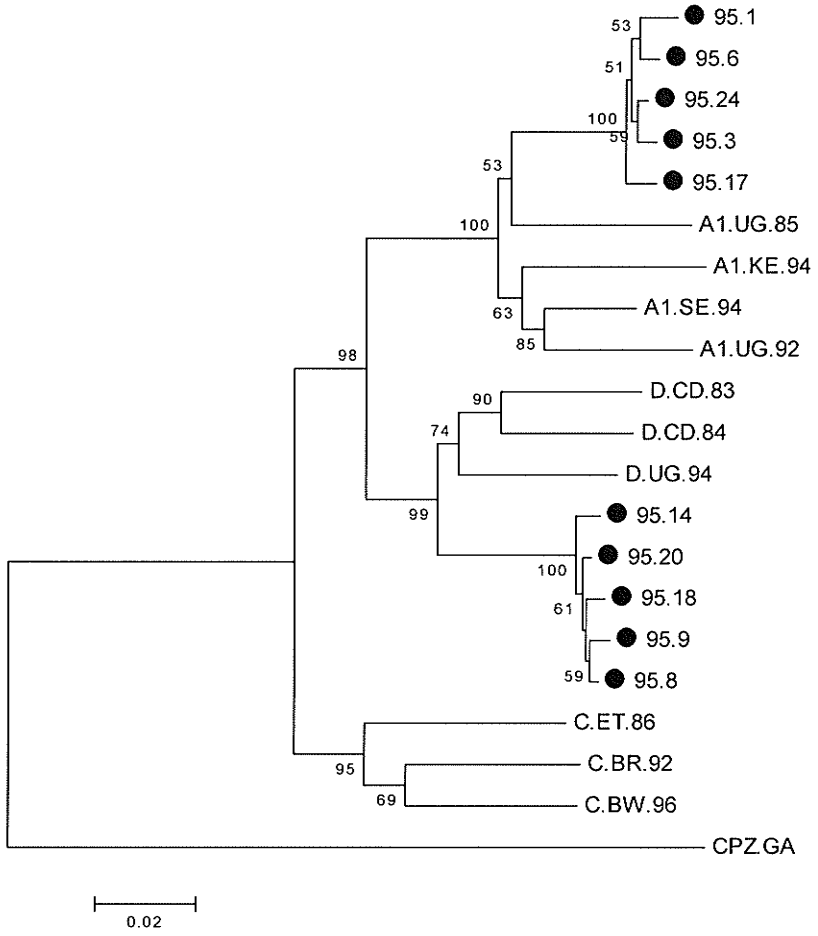

Figure 24: Dual Infection Case II Phylogenetic Analysis: A few select clones from ML825 (1995 sample) are displayed here with subtype references of interest. The Gag/Pol clones (nucleotides 2086-2988 by HXB2 numbering) sequenced revealed both A and D subtype viruses. Recombinant viruses were not included within this dendogram.

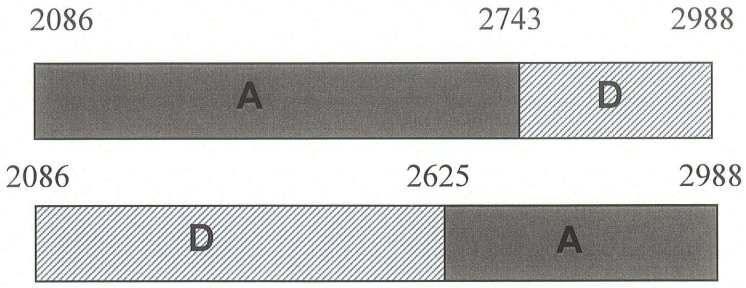

Figure 25A: **Dual Infection Case II: Diagrams of Recombinant Forms** (Two different A/D recombinant forms were seen)

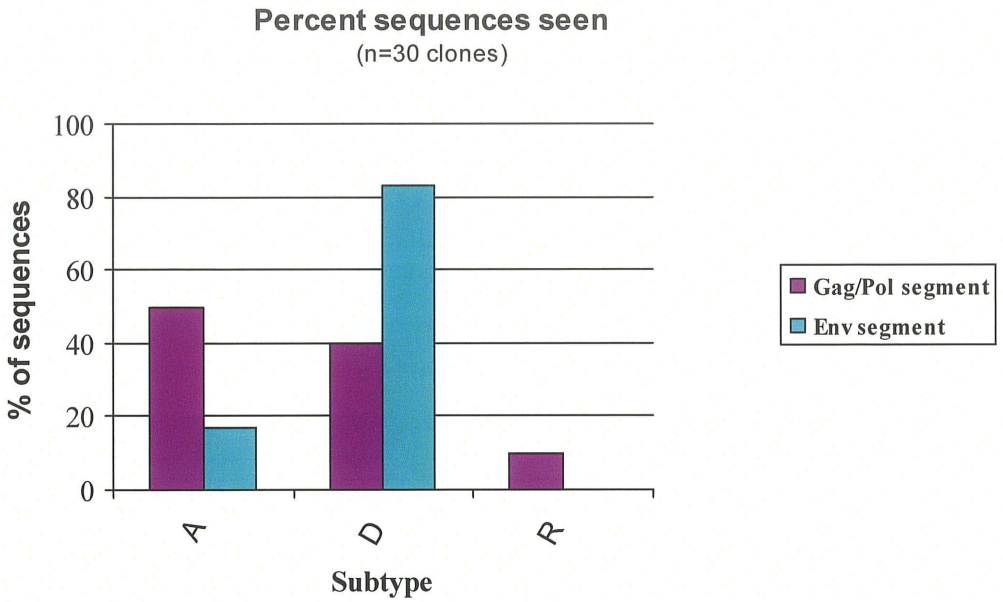

Figure 25B: Dual Infection Case II: Proportion of Viral Forms Present (Gag/Pol and Env sequence data)

Figure 25: Dual infection case II (ML825, sample year: 1995): **A.)** Two different recombinant forms were seen for the Gag/Pol segment (nucleotides 2086-2988 HXB2 numbering). An A/D form with a break at position 2743 (HXB2) and a D/A form with a break at 2625 (HXB2) were detected. **B.)** Subtype A, D and recombinant forms were detected based on the Gag/Pol clones that were sequenced (n=30). Subtype A was the most prevalent followed by D and then recombinants. Envelope (7008-7817 HXB2) sequence data revealed only A and D sequences with D being the dominant form seen.

Dual infection case III:

Dual infection case III (ML1088) was identified as being infected with three different genetic forms within the Gag/Pol region sequenced (nucleotides 2086-2988 by HXB2 numbering) - subtype A and two different recombinant forms A/D and C/D. Three different genetic forms were evident from phylogenetic analysis but such a dendogram also emphasizes the importance of recombination screening of all clones (see Figure 26). Depending on the breakpoints present and the length of each subtype within a fragment some recombinants may be missed without proper screening. The C/D recombinant form had a breakpoint range of nucleotides 230-329 within the sequence fragment (midpoint of 280 (nucleotide 2377 by HXB2 numbering)). The A/D form had a breakpoint range of 482-509 (midpoint of 496, 2581 by HXB2 numbering). Analysis of 30 clones revealed 80% were C/D, 13.3% A/D and 6.7% A (see Figures 27A and 27B).

Dual infection case IV:

Dual infection case IV, ML 1197, was identified as having A, D and A/D recombinant viruses within the Gag/Pol region sequenced (2086-2988 HXB2 numbering) (see Figure 28). The A and D viruses were used as parental sequences to determine the breakpoint for the recombinant form and a breakpoint range of nucleotides (nt) 483-647 within the fragment (midpoint nt 565 which corresponds to nt 2653 of HXB2) was identified (see Figure 29A). Analysis of 31 clones revealed that the majority were A/D recombinant (74%) and the parental subtypes A and D were equally represented (13% each) (see Figure 29B).

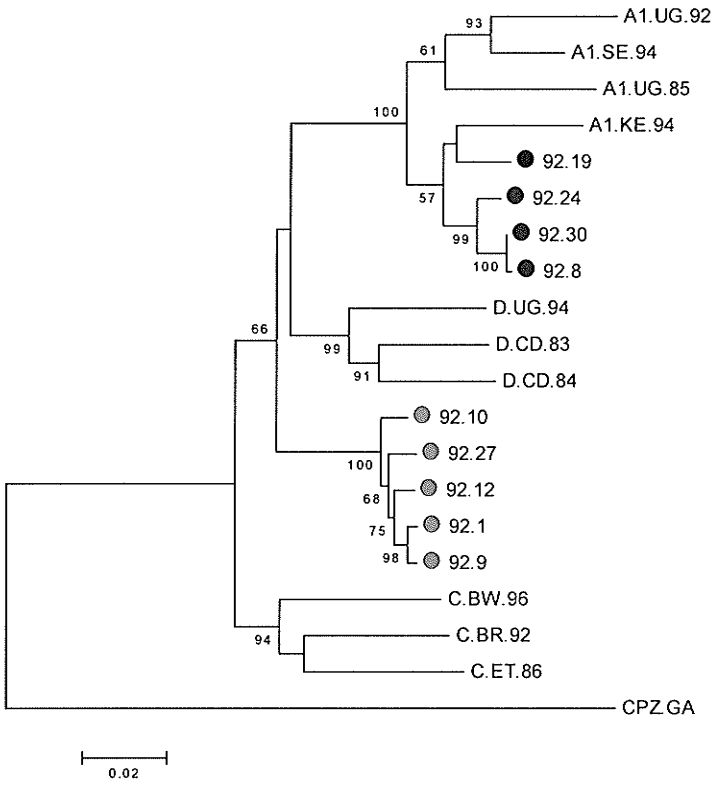

Figure 26: Dual Infection Case III Phylogenetic Analysis : This dendogram displays three different genetic forms of the Gag/Pol region (2086-2988 HXB2 numbering) based on phylogenetic and recombination analysis. A few select clones are shown and are represented by circles. Different colours are used to identify the three different forms. Clone 19 is subtype A, clones 24, 30, 8 are A/D recombinant forms and clones 10, 27, 12, 1, 9 are the C/D recombinant form. Relevant references are also included.

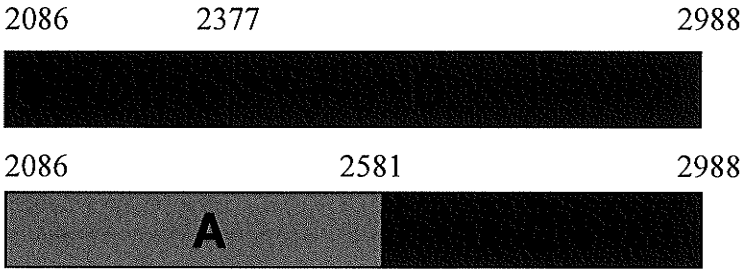

Figure 27A: Dual Infection Case III Diagrams of Recombinant Forms (A/D and C/D recombinant forms detected).

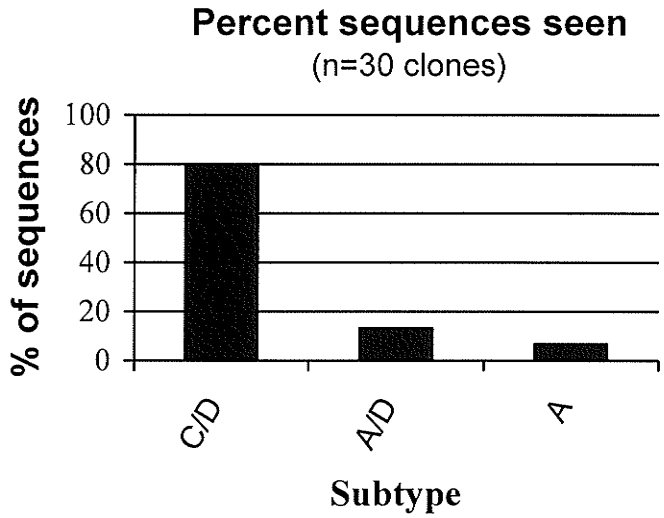

Figure 27 B: Dual Infection Case III: Proportion of Viral Forms Present

Figure 27: A.) Dual infection case III (ML1088- 1992 sample): Three different genetic forms were identified; an A subtype, a C/D recombinant (breakpoint at nucleotide 2377 by HXB2 numbering) and an A/D recombinant (break point at nucleotidet 2581 HXB2). **B.)** The graph shows the proportion of each variant present at the time of study (12.3.1992). The C/D form is the most prevalent followed by A/D and then A.

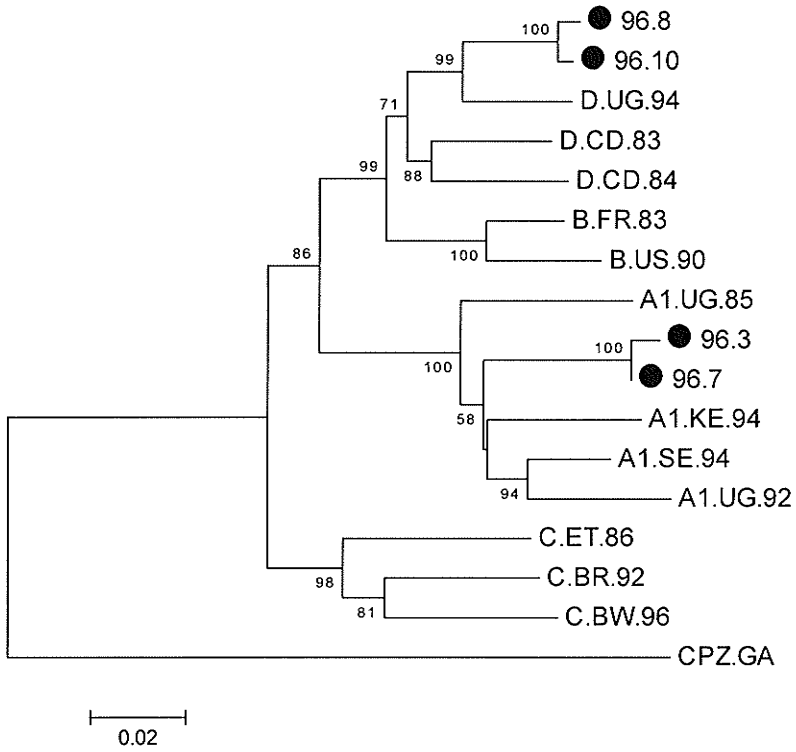

Figure 28: Dual Infection Case IV Phylogenetic Analysis: The dendogram displays representative subtype A and D sequence clones (circles) along with select references. Recombinant forms were not included.

Figure 29A: Dual Infection Case IV Diagram of Recombinant Form D/A

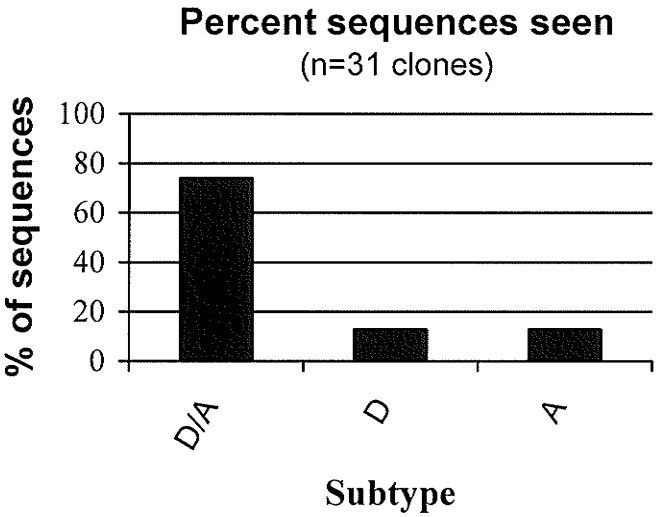

Figure 29B: Dual Infection Case IV Proportion of Viral Forms Present

Figure 29: A.) The recombinant form found within ML1197 based on the Gag/Pol region (nucleotides(nt) 2086-2988 by HXB2 numbering) is displayed above and has a breakpoint at nt 2653 (HXB2 numbering) **B.)** Parental subtype A and subtype D viruses as well as an A/D recombinant were detected within this individual. The recombinant was the dominant form seen.

Dual infection case V:

Dual infection case V (ML 229) was identified as having a subtype A and subtype D dual infection with no evidence of recombination based on Gag/Pol sequence data (see Figure 30). Analysis of 30 clones revealed 22.6% were subtype A and 77.4% were subtype D (see Figure 31).

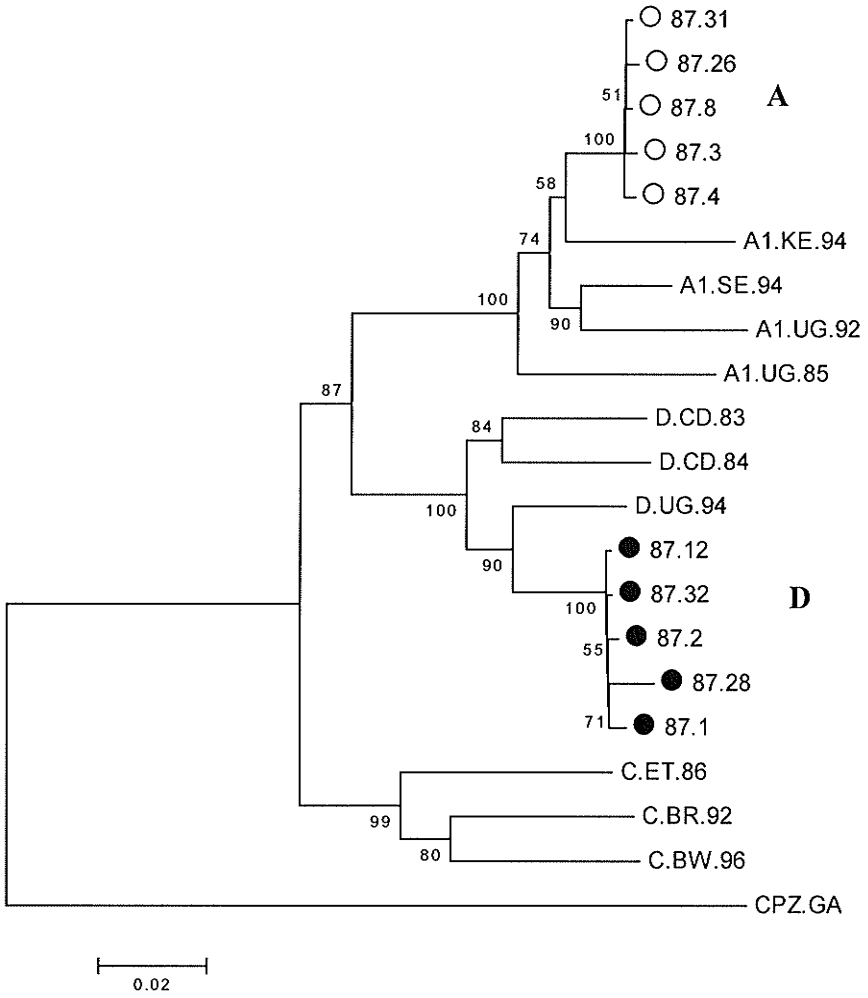

Figure 30: Dual Infection Case V Phylogenetic Analysis: This dendogram includes representative subtype A (open circle) and subtype D (solid circle) sequence clones from ML 229 (16.7.1987 sample) along with select references.

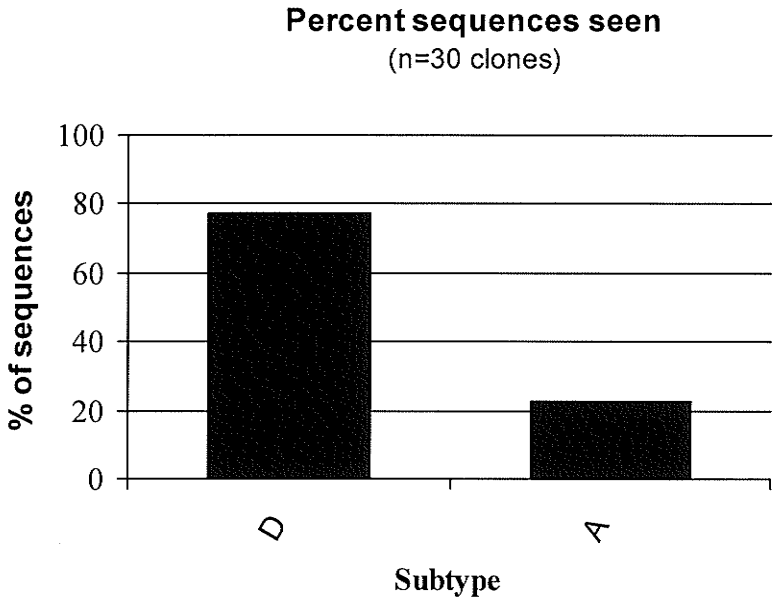

Figure 31: Dual Infection Case V: Proportion of Viral Forms Present

Subtype D and A forms within Gag/Pol were detected within ML229 (1987) at a frequency of 77.7%- D and 22.6%- A.)

Dual infection case VI:

Dual infection case VI (ML1744) was identified as having both subtype D and A/D recombinant viruses within Gag/Pol (nucleotide (nt) 2086-2988 by HXB2 numbering). Figure 32 displays a few select clones from this individual that are subtype D. Screening for recombination revealed that the second distinct group of sequences seen in Figure 33A are in fact an A/D recombinant form with a breakpoint at 2346 HXB2 (see Figure33B). Based on sequencing of 30 clones, a near equal distribution of the two different forms was seen, 58% were D and 42 % A/D (see Figure 33C).

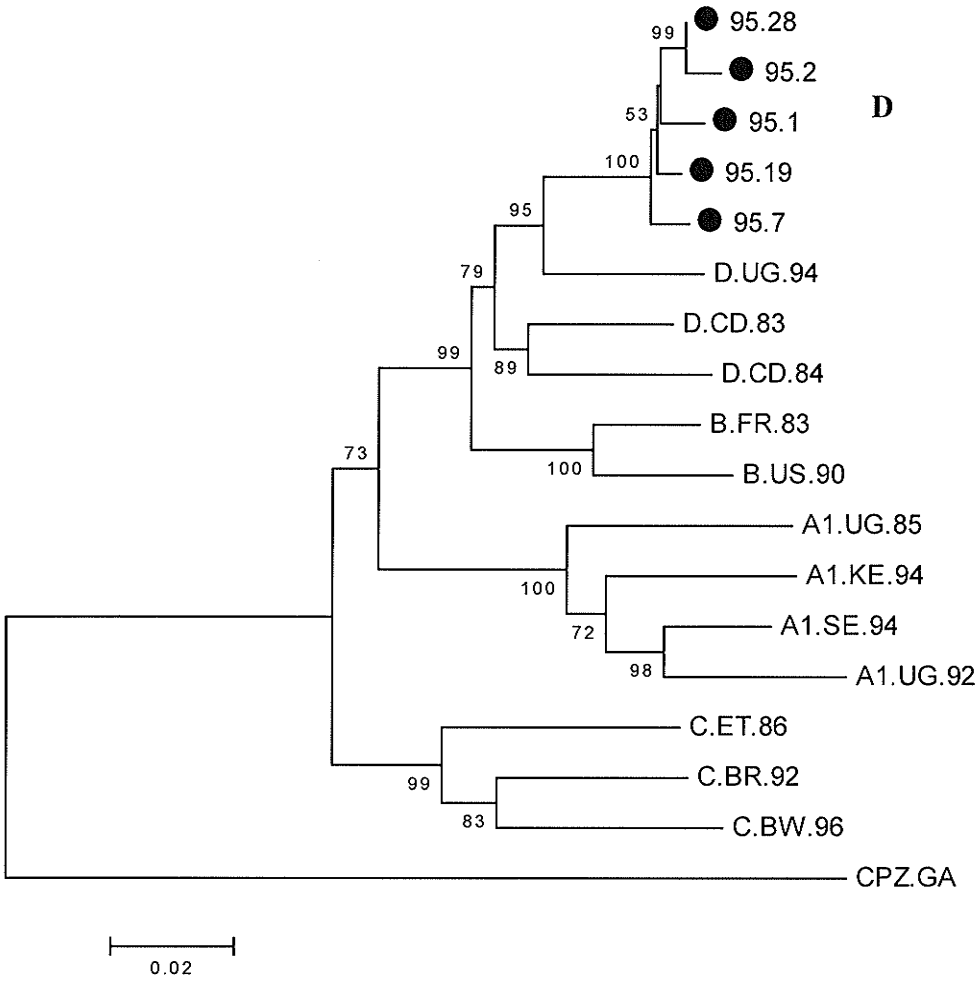

Figure 32: Dual Infection Case VI Phylogenetic Analysis for Non-Recombinant Form: References and a few select clones from ML 1744 (red circles) revealed that one of the infecting virus types is subtype D (fragment is 2086-2988 by HXB2 numbering).

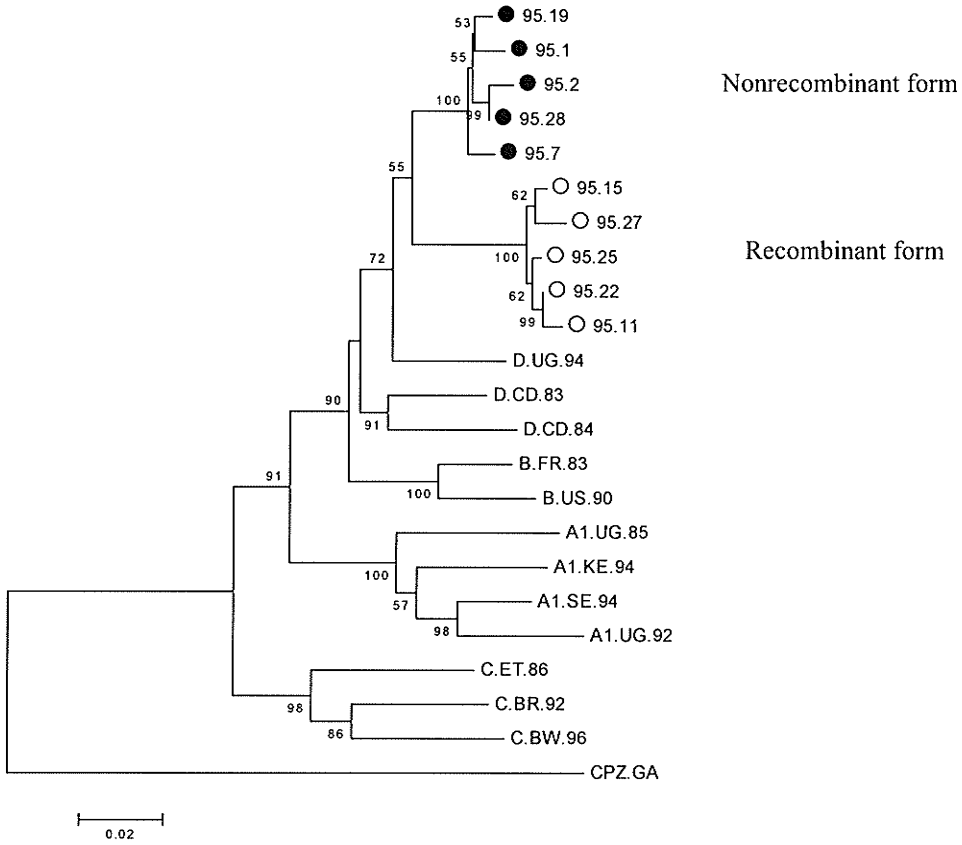

Figure 33A: Dual Infection Case VI Phylogenetic Analysis for Non-Recombinant and Recombinant Forms

Figure 33B: Dual Infection Case VI Diagram of Recombinant Form A/D

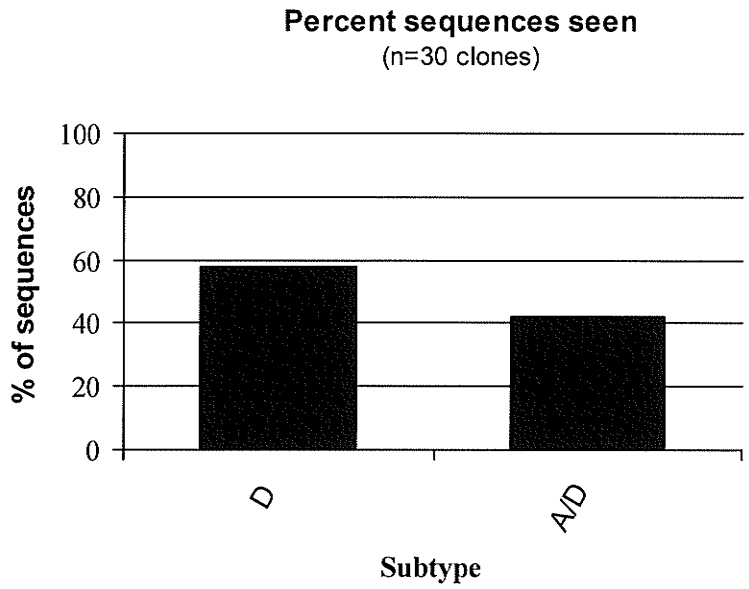

Figure 33C: Dual Infection Case VI: Proportion of Viral Forms Present

Figure 33: **A.)** Dual infection case VI (ML1744): Dendogram with references and two distinct viruses from ML1744 (region= nucleotides 2086-2988 by HXB2). One form (open circle) is an A/D recombinant virus and the other (closed circle) (as displayed in Figure 32) is subtype D. **B.)** The recombinant form is an A/D virus. It is predominantly D which is why it trees closely to the pure D form in figure A. The breakpoint is at nucleotide 2346 (HXB2 numbering). **C.)** The two forms are present at almost equal proportions, D (58%) and A/D (42%).

Dual infection case VII:

Dual infection case VII (ML45) was identified as being infected with a subtype A virus and an A/D recombinant virus (20.4.1995) (See Figure 34). The A/D recombinant form was predominantly D with a recombination breakpoint at position 2802 (HXB2 numbering) (see Figure 35A). Within this individual the subtype A virus predominated, and comprise 96.6% of the viruses seen while the recombinant form accounted for the remaining 3.3% (see Figure 35B).

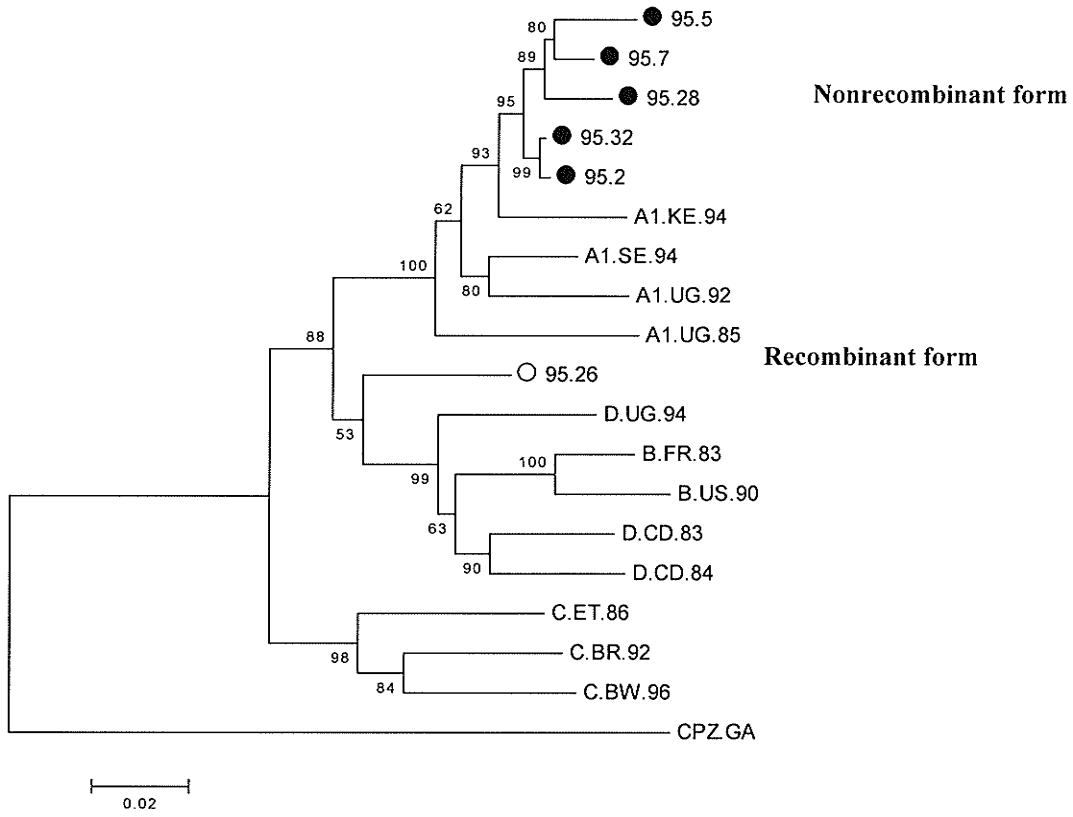

Figure 34: Dual Infection Case VII Phylogenetic Analysis: The dendogram includes subtype references and a few select clones (2086-2988 HXB2 numbering). One of the infecting virus types is subtype A (solid circle). The other virus type is a recombinant D/A form (open circle).

Figure 35A: Dual Infection Case VII: Diagram of Recombinant Form D/A

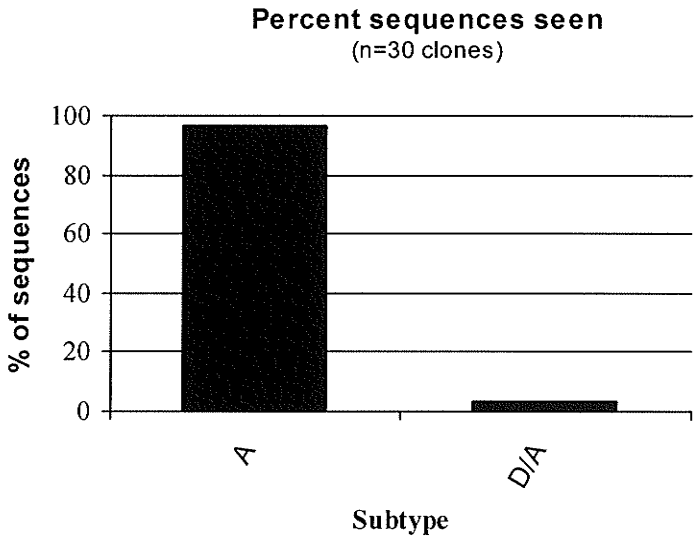

Figure 35B: Dual Infection Case VII: Proportion of Viral Forms Present

Figure 35: Dual infection case VII (ML 45): **A.)** This individual has an A/D recombinant virus based on sequencing of Gag/Pol nucleotides (nt) 2086-2988 by HXB2 numbering. It is predominantly subtype D with a recombination breakpoint at nt 2802 (HXB2 numbering) **B.)** Subtype A was the dominant form detected and it accounted for 96.6% of the viruses identified. The recombinant form was responsible for the remaining 3.3% of clones.

Summary: Dual Infections and Superinfections - Virological Consequences

The following table (Table 17) is a summary of the viral genetic data from the 6 dual infection cases presented in this section as well as the 5 dual/superinfection cases described in the nested cohort study (section II). In 10 of 11 cases an A subtype virus was detected (91%), C subtype in 1 of 11 cases (9%) and D subtype in 4 of 11 cases (36%). Recombinant viruses were detected in 10 of these 11 individuals (91%) of which 1 individual had 2 very different forms (C/D and A/D). These 10 individuals thus accounted for 11 different recombinant viruses. Eight of the recombinant forms were A/D recombinants (81.8%), 1 was a C/D form (9.1%) and the last was an A/C form (9.1%). In 3 of the 10 cases (30%) in which recombinants were detected non-recombinant parental strains that could possibly account for the formation of recombinants within the individual were identified. Within the 10 individuals that harbored recombinant viruses 6/10 (60%) had recombinant strains that predominated at the single time point studied or most recent date investigated. Two individuals (20%) had recombinant and non-recombinant forms at equal or near equal proportions and 2 other individuals (20%) had a non-recombinant form that predominated.

Table 17: Dual Infection/ Superinfection (SI): Virus Summary

Case	Viruses Seen Within the Individual (includes data from all available time points)						Virus at Greatest Proportion (either at single time point tested or at the most recent date)
	A	**C**	**D**	**A/D**	**A/C**	**C/D**	
ML47: SI-II	**X**			**X**			A/D
ML1295:SI-IV	**X**			**X**			A and A/D equal
ML1449: SI-III	**X**			**X**			A/D
ML1787: SI-I	**X**	**X**			**X**		A/C
ML45: Dual	**X**			**X**			A
ML229: Dual	**X**		**X**				D
ML293: Dual	**X**			**X**			A/D
ML825: Dual	**X**		**X**	**X**			A
ML1088: Dual	**X**			**X**		**X**	C/D
ML1197: Dual	**X**		**X**	**X**			A/D
ML1744: Dual			**X**	**X**			D (almost equal)

Disease Progression in Superinfection/Dual Infection Cases:

The majority of individuals who were dually infected or superinfected had recombinant viruses (91%) and in 60% of those individuals the recombinant virus became the predominant form in the time they were followed. Besides an investigation into the virological consequences of dual infections and superinfections I also wanted to investigate the clinical consequences of such infections. To test the hypothesis, **that an increased clinical progression would be seen in the majority of individuals with dual infections (superinfections/co-infections),** CD4 + T cell counts, viral loads and survival times were examined. Case control studies and individual assessments were used to investigate clinical progression.

Case control studies:

Due to the small number of dually and superinfected individuals who met the study criteria for Kaplan-Meier survival analysis pertaining to the time to CD4+ T cell counts below 400 or 200 cells/mm^3, the case control studies did not produce meaningful results. Small study groups also limited the usefulness of my attempts to measure the differences in the rates of CD4+ T cell decline and the CD4+ T cell intercepts for the superinfection case control study and the dual infection case control study.

Individual disease assessments for superinfection cases:

As the case control studies did not produce meaningful results an assessment of each individual on a case by case basis by evaluating CD4+ T cell counts, survival time and viral load data was done.

ML47 (Superinfection case II) seroconverted in 1985 and experienced a superinfection event some time between 28.05.1987 and 11.09.1995. Unfortunately this window could not be narrowed down further due to sample inavailability. Based on the midpoint value, the estimated time of superinfection was therefore reported as 20.07.1991. Viral load data in 1995 was only slightly higher than the average viral load data from HIV-1 positive Kenyan women (52 841 (+/- 18 056) copies/mL (238)) (see Table 18). A significant decrease in CD4+ T cells was seen following 29.10.91 (shortly after the estimated seroconversion date) from 612cells/mm^3 to 397cells/mm^3. However, this individual was not a rapid progressor. In 2000 (the last available date (205 cells/mm^3)), 15 years after seroconversion and an estimated 8 years after superinfection, CD4+ T cell counts had not fallen below 200cells/mm^3 (see Figure 36).

Table 18: SI Case II (ML47) Viral Load Data

Date	Viral Load (copies/mL)
11.09.1995	88,000
18.03.1996	61,000

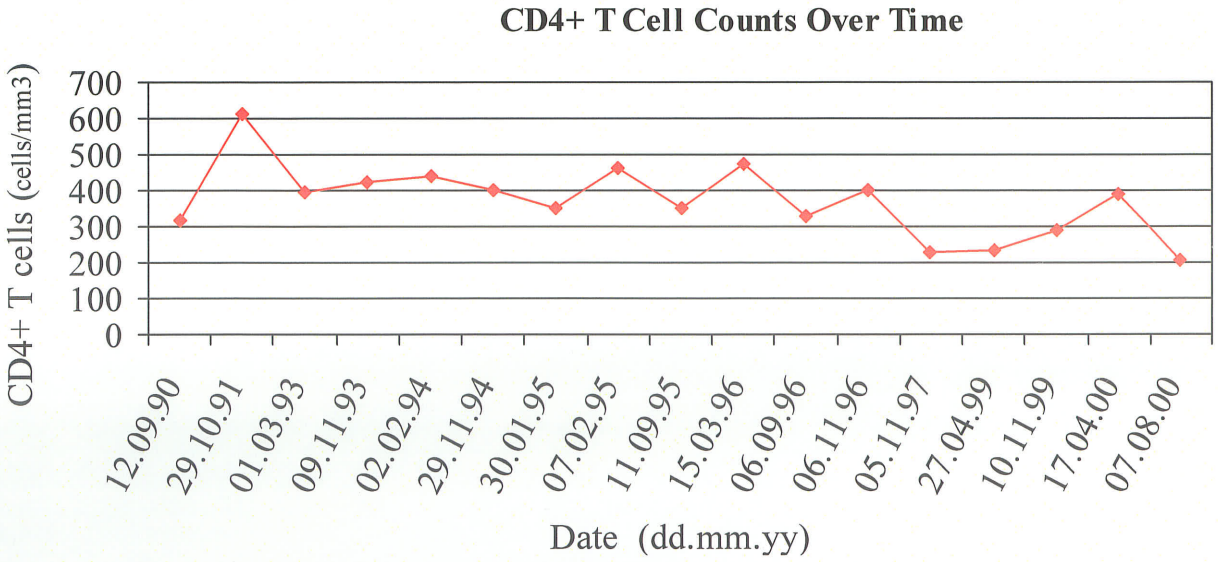

Figure 36: Superinfection Case II (ML47) CD4+ T Cell Counts Over Time

ML1295 (Superinfection case IV) entered the cohort positive on 03.07.1990. A superinfection event occurred between 02.02.1996 and 28.05.2002 at an estimated date of 13.02.1999. Viral load available prior to the superinfection event was low (1,600 copies/mL) but was dramatically higher following the event as determined in 2002 (90,000 copies/mL) (see Table 19). This value is also higher than the available average viral load mentioned above. CD4+ T cell counts dropped to below 200cells/mm^3 on 9.2.01 (see Figure 37). This individual however was still alive as of 2004, 14 years after entering the cohort positive and 5 years after the estimated superinfection date.

Table 19: Superinfection Case IV (ML1295) Viral Load Data

Date	Viral Load (copies/mL)
02.02.1996	1,600
28.05.2002	90,000

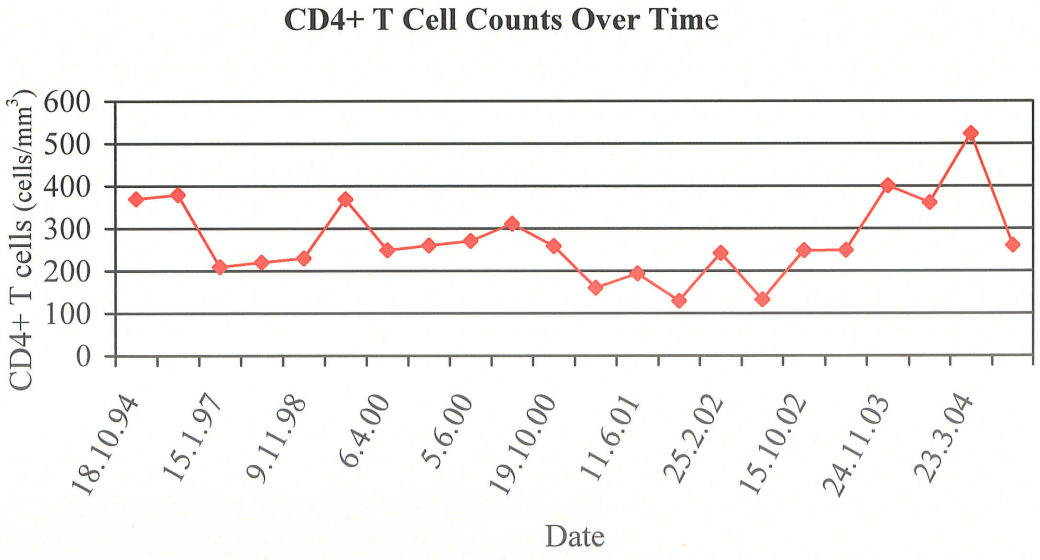

Figure 37: Superinfection Case IV (ML1295) CD4+ T Cell Counts Over Time

ML1449 entered the cohort HIV infected on 11.02.1992 and had an estimated superinfection date of 29.03.1994. A low viral load count of 6,900 copies/mL was seen in 1992 but by 1994 (following the estimated superinfection date) a very substantial increase was detected (viral load of 250,000 copies/mL). Viral loads continued to increase and reached 480,000 copies /mL by 1996 (see Table 20). Furthermore, the superinfection event was temporally associated with a significant drop in CD4+ T cells. Concentrations dropped from 397cells/mm^3 (17.11.92) to 180cells/mm^3 by 16.8.1994 followed by a constant decline and final CD4+ T cell count of 60cells/mm^3 by 5.2.1996 (see Figure 38). Since a seroconversion date was not available I could not determine if this individual was a rapid progressor, but a poor outcome following superinfection was certainly observed.

Table 20: Superinfection Case III (ML1449) Viral Load Data

Date	Viral Load (copies/mL)
21.04.92	6,900
10.11.94	250,000
05.03.96	480,000

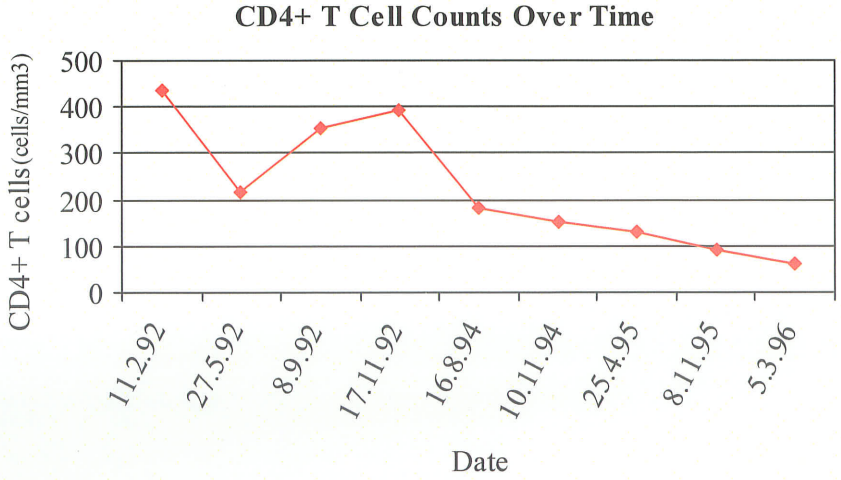

Figure 38: SI Case III (ML1449) CD4+ T Cell Counts Over Time

Superinfection case I (ML1787) provided the best opportunity to investigate superinfection and the CD4+ T cell and viral load data surrounding this event. This was because of the narrow window in which we could predict that the superinfection event occurred. Based on the sequencing data the superinfection event occurred between 10.12.1997 and 26.10.1998, with an estimated superinfection date of 19.05.1998. Viral loads were not above average but values more than doubled between the dates that most closely surrounded the superinfection event. By 2003, when the recombinant viral form was the only detectable variant, the viral load was at a maximum (see Table 21). A drop in CD4+ T cell counts between the 1997 and 1999 dates (1998 not available) was observed. A partial recovery in CD4+ T cell counts occurred within 2000. The first date at which the CD4+ T cell count was below 200 cells/mm^3 was in 2003 (see Figure 39). Despite superinfection this woman survived for a minimum of 5 years subsequent to the superinfection event.

Table 21: Superinfection Case I Viral Load Data

Date	Viral Load (copies/mL)
21.08.1995	<125
27.06.1996	<125
10.12.1997	<357
26.10.1998	1,100
15.12.1999	6,500
14.08.2003	53,000

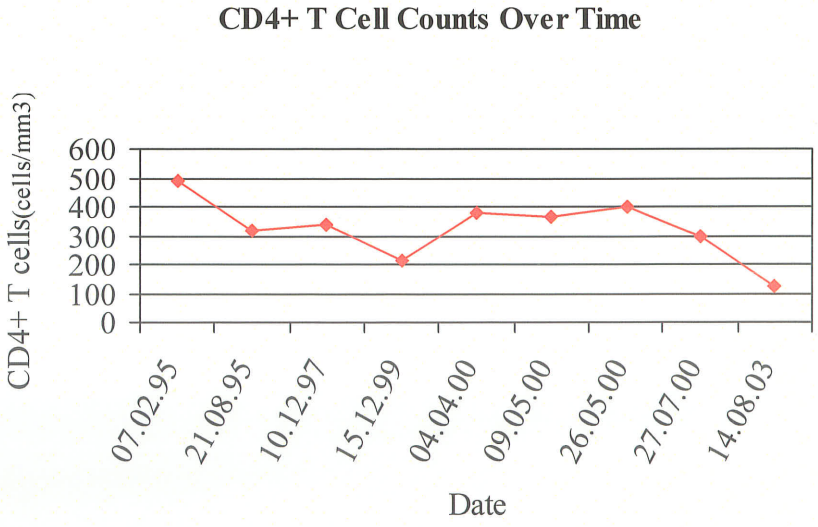

Figure 39: Superinfection Case I (ML1787) CD4+ T Cell Counts Over Time

Individual disease assessments of dual infection cases:

In most cases of dual infection, limited sample availability did not permit a definitive determination of whether the dual infection occurred due to co-infection early on, or was due to a superinfection event. CD4+ T cell data, viral load data and seroconversion dates make some cases more informative than others, but all cases will be discussed below.

Dual infection case II (ML825) seroconverted at an estimated date of 1.5.1992. A seroconversion date and available sample soon after seroconversion provided an ideal case in which to study the clinical impact of a dual infection. This dual infection was most likely the result of co-transmission of multiple viruses and not a superinfection event (based on seroconversion and sample date). The sample date at which a dual infection was detected was 22.09.1992 (also the first sample that tested positive for HIV). Viral load data was not above average for this individual but CD4 + T cell data indicated that progression to a CD4 + T cell count below 200cells/mm^3 occurred by 10.4.1995. This woman can therefore be defined as a rapid progresser as her CD4+ cell counts dropped to below 200 cells/mm^3 in less than three years following seroconversion.

Table 22: Dual Infection Case II Viral Load Data

Date	Viral Load (copies/mL)
10.04.1995	21,000
21.11.1995	11,000

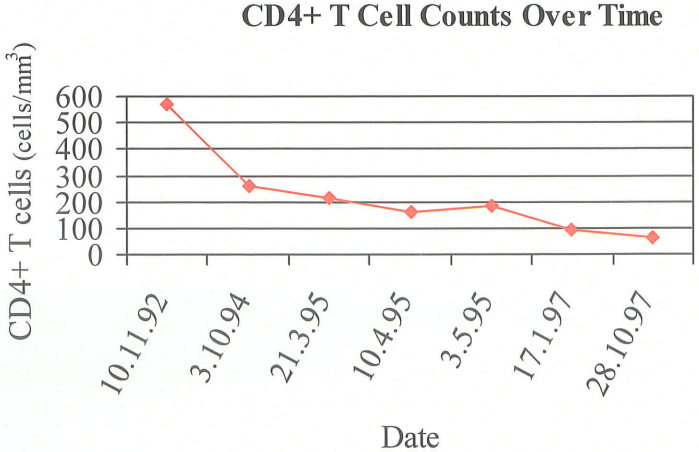

Figure 40: Dual Infection Case II (ML825) CD4+ T Cell Counts

Dual infection case V (ML229) entered the cohort positive on 7.3.1985 and a sample dated 16.7.1987 was analysed and revealed an A subtype, D subtype dual infection. Viral load data from 1987 was not high (10.9.1987=2,900 copies/mL). CD4+ T cell data only became available in 1990 at which point CD4+ T cell counts were 182 cells/mm^3 and by 1992 a count of 30 cells/mm^3 was seen. Since we do not have earlier CD4+ T cell data or an exact date of seroconversion it is not possible to determine if this individual is a rapid progresser as per our definition. However, the time from confirmed dual infection to confirmed CD4+ T cell counts below 200 cells/mm^3 was definitely less than three years. This individual is certainly not a long term non-progressor.

Dual infection case III (ML1088) was an individual with a reported seroconversion date of 09.08.1989. Testing of the earliest available sample (12.3.1992) revealed infection with multiple viruses (A, A/D, C/D). Viral load data however was not above average and no reported CD4+ T cell counts were below 200 cells/mm^3 (see Table 23 and Figure 41). The dual infection could have resulted if this individual was co-infected from the start, or if superinfection occurred at some point prior to 1992. To thoroughly analyze possible implications of both assumptions I assessed the progression from seroconversion and progression from a predicted superinfection date. From this individual's seroconversion date to the most recent date at which CD4+ T cell data was collected five years had passed and CD4+ T cell levels were never below 200 cells/mm^3. If this were a superinfection event, with a predicted superinfection date of 25.11.1990, over three years would have passed during which CD4+ T cell counts remained above 200 cells/mm^3.

Table 23: Dual Infection Case III (ML1088) Viral Load Data

Date	Viral Load (copies/mL)
12.03.1992	11,000
03.11.1992	6,400

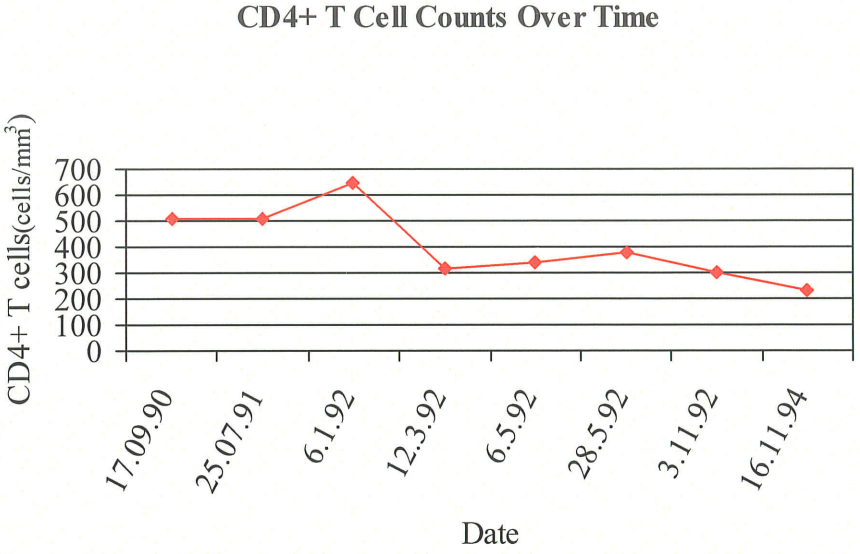

Figure 41: Dual Infection Case III (ML1088) CD4+ T Cell Counts

Dual infection case I (ML293) entered the cohort positive on 8.3.1985. Viral load data was well above average at every time point tested with values all in the 10^5 copies/mL range (see Table 24). A sample trace back to 1993 (plasma sample) revealed the presence of two distinct viruses as early as sample availability allowed. CD4+ T cell counts did not drop below 200cells/mm^3 until 5.2.1998, a minimum of 13 years since infection. Five years had passed even from the most conservative estimate of the time of dual infection (the confirmed date of 1993) to CD4+ T cells below 200 cells/mm^3 (see

Figure 42). Although this individual had a very high viral load she can certainly not be classified as a rapid progressor.

Table 24: Dual Infection Case I (ML293) Viral Load Data

Date	Viral Load (copies/mL)
29.08.1995	410,000
22.04.1996	280,000
12.10.1998	280,000

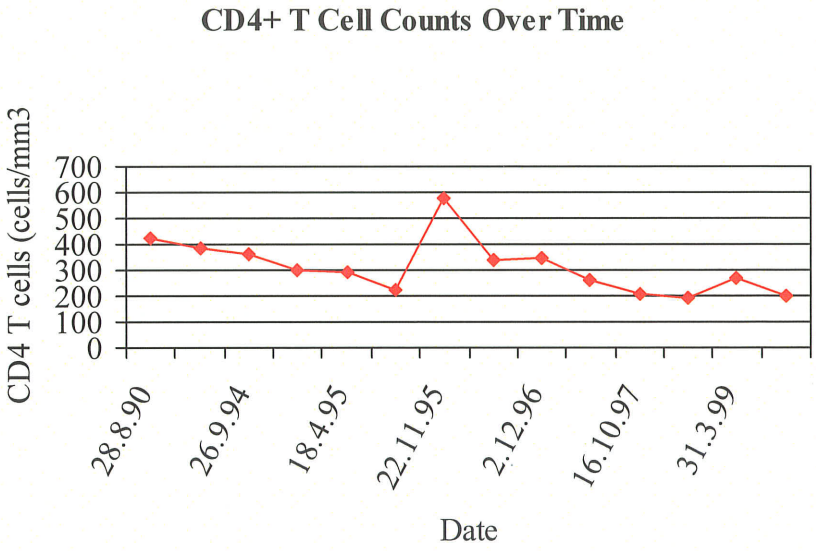

Figure 42: Dual Infection Case I (ML293) CD4+ T Cell Counts Over Time

ML1197 entered the cohort positive on 28.11.1989. The sample from which dual infection was detected was 15.4.1996. Viral load data was relatively low at the time points tested (9,500 and 2,100 copies/mL on 15.4.1996 and 10.11.1997 respectively). Six years following entry into the cohort, the CD4+ T cell counts dropped below

200cells/mm^3. This woman was still alive in 2000 but at that time her measured CD4+ T cell concentration was at 150 cells/mm^3.

ML 45 was HIV infected on entry into the cohort (22.02.1985). Dual infection was detected using a sample dated 20.4.1995. The viral load determined for 23.4.1995 was 43,000 copies/mL (not above average). CD4+ T cell counts fluctuated within this individual but the time at which two consecutive counts were below 200cells/mm^3 was in 1997, more than 12 years after infection and at least 2 years after confirmation of dual infection.

ML1744 entered the cohort positive on 18.01.1994 and the first available sample for testing, dated 8.8.1995, revealed a dual infection. The viral load was 73,000 copies/mL on 16.11.1995. This is slightly above the average range used in this analysis as a reference. CD4+ T cell data was limited as this individual was last seen in the clinic only 3 months after the initial CD4+ T cell count was taken (8.8.1995: 530 cells/mm^3, 12.10.1995: 280 cells/mm^3, 16.11.1995: 350 cells/mm^3). It is therefore difficult to draw concrete conclusions regarding progression but absence from the clinic may indicate a detrimental outcome.

Section III Key Findings:

Recombinant viruses were seen in 10 out of the 11 dual/superinfected individuals and in most cases the recombinant form predominated. In the superinfection cases, where the window in which the event occurred could be reasonably estimated, increases in viral loads and decreases in CD4 + T cell counts were seen. Rapid progression was seen in some of the superinfection/dual infection cases but this was certainly not the case for all individuals studied.

DISCUSSION

A comprehensive understanding of HIV viral genetics and the resulting host-virus relationship is paramount to the development of an effective vaccine against HIV-1. The purpose of the work reported within this thesis was to better characterize single HIV-1 infections with respect to viral genetics, clinical consequences associated with viral diversity and the strength of the immune response generated by such infections (as measured by the incidence of superinfection events). It further aimed to elucidate both the virological and clinical consequences of multiple (superinfection/dual infection) HIV-1 infections as well as to inspire the pursuit of future investigations into aspects of the viral genetics-host relationship.

Section I

The circulating HIV-1 genetic forms that predominate within different geographic regions and populations may differ. This is a result of numerous factors including the introduction of different, genetically distinct, founding viruses, the influx of new genetic forms and the increased expansion of some forms over others. The Joint United Nations Programme on HIV/AIDS warns that within sub-Saharan Africa "there is no single African epidemic". There is extensive diversity seen within different geographic sub-regions and populations with respect to the impact seen, the rate of spread as well as in prevention and treatment strategies (1). Prior to pursuing any additional studies I thus felt it important to define the epidemic within our cohort under study, specifically with respect to both the circulating viruses present and any differences in disease progression seen between different genetic forms.

The hypothesis that was first set out to test was that an investigation of HIV-1 genetics within the sex workers that have single HIV infections would reveal the following: **a.) a diverse array of genetically distinct viruses, b.) a dynamic and heterogeneous distribution of viral subtypes, c.) viral genetics associated differences in disease progression.**

Molecular Epidemiological Profile of Single HIV-1 Infections within the ML Cohort:

Diverse set of viral subtypes within the ML cohort

Sequencing of a region of Gag/Pol revealed the presence of numerous viral subtypes within the cohort (n=202) including subtypes A, C, D, G, unclassifiable (U) as well as recombinant forms (R). Subtype A was observed to be the most prevalent, followed by D and then C (68.3% A, 6.9% C, 21.3% D, 2% G, 0.5% U, 1.0% R). The order of prevalence between A, D and C subtypes seen here is consistent with other population studies within Kenya (94;240). Neilson *et al.* (1999) investigated a group of Kenyan women in a breast feeding transmission study. They used heteroduplex mobility assays (HMA) for envelope (sequencing only when HMA was unsuccessful) and reported a distribution of 70.3% A, 6.9% C, 20.5% D, 0.3% G and 2.2% recombinant (94;100). It appears that a diverse mix of viruses are seen in the ML sex worker cohort but this subtype diversity is not significantly different from reports of groups with lower exposure to HIV in Kenya.

Recombinant viruses are common

Although recombinant viruses were detected based on the Gag/Pol region sequenced this is most likely an underestimation of the recombinant viruses that are present. Full length genome sequencing for every woman would likely reveal additional recombinant viruses. An alternative approach to achieve a better understanding of the recombinant viruses present, is to sequence more than one region of the HIV genome for individuals, which is more feasible than full length sequencing. The existence of recombinant viruses was thus further assessed by determining how many individuals had discordant versus concordant subtyping results for different genetic regions (103;240;241). Sequencing a region of Env for a subset of women, and Gag for a different subset of individuals revealed that 12% and 10.4% of women within these groups respectively had discordance between the new regions investigated and the data from Gag/Pol. These data provide evidence that recombinant viruses within the cohort are common. (Inclusion of dually/superinfected individuals would increase the number of recombinants seen even further, as is discussed later).

The recombinant viruses that are observed within individuals may be due to *in vivo* recombination events resulting from a dual infection or due to the direct transmission of recombinant forms. The detection of dually (superinfected/co-infected) infected individuals within the cohort and the development of recombinant viruses is explored in sections II and III of this thesis. The recombinant viruses detected within the ML women (described above) were isolated from samples as early as 1991. Seroconversion dates or entry dates (if entered positive) of individuals from whom recombinants were identified are as early as 1985-1989. Although recombinant viruses

were first identified in 1995 (53) it is clear that they have been present within the ML cohort since early in the epidemic. Recombinant forms have thus likely had time to expand as the epidemic has progressed. Recent studies within different groups of individuals in Africa also support the finding that recombinant viruses have been present, and possibly even have been founding viruses of the early African epidemic (103;242). The prevalence of recombinant viruses seen within the cohort, and in Kenya in general, is therefore in part due to the expansion and transmission of recombinant strains within the population, since early in the epidemic. The diversity of circulating subtypes, described above, and infection with multiple forms, certainly contributes to the ongoing development and prevalence of recombinants (discussed in sections II and III).

Diversity within HIV-1 protease

As part of the viral assessment of single HIV-1 infections we determined the baseline prevalence of naturally occurring polymorphisms for HIV-1 protease which have been associated with drug resistance. Diversity was identified within the amino acid sequence of protease for the various subtypes present (A, C, D, G) compared to subtype B, against which antiretrovirals were designed. Although no type I (major mutations) were seen, type IV mutations, which can contribute to antiretroviral resistance when combined with one or more type I-III mutation(s), were very common among the different subtypes (see Figure 5). Some of the type IV mutations, determined for subtype B, are actually the most common amino acids seen within the protease sequence of different HIV-1 subtypes. Theoretically, the presence of these amino acids could result in faster development of resistance to protease inhibitors. Protease amino acid diversity is

certainly seen between the different subtypes. Now that we are on the cusp of seeing all ML women receive antiretroviral treatment this data will serve as an important baseline assessment. The differences seen in comparison to subtype B may impact the effectiveness of treatment and the development of resistance. This can be addressed in future studies.

It is evident that extensive viral genetic diversity exists within this cohort. At the subtype level, multiple subtypes circulate and recombinant forms are prevalent. Genetic variability within protease, an important antiretroviral drug target, is also evident, within the women of this cohort. Continued monitoring of the diversity with respect to subtypes, recombinants and antiretroviral targets will be important for both treatment and vaccine strategies as well as the design of future scientific studies.

Changes in subtype distribution over time

The description of genetic diversity provided above characterizes the genetic variability present within the ML cohort but does not obviate the reality that changes may have occurred over time as the epidemic evolved. The ML cohort was established in 1985 and its continued growth has provided a rare opportunity to investigate changes in subtype distribution over an extensive time span within a single cohort. Indeed, the subtype distribution seen in the ML cohort has changed significantly over the course of the ever growing epidemic (Chi-Square p= 0.0324, Mantel-Haenszel p=0.0067). Specifically, although subtype A remains the most prevalent subtype, it has decreased consistently over the three time periods investigated (Chi-square p= 0.0125 Mantel-Haenszel p= 0.0048). Subtype A prevalence dropped from 80.8% in the first time period

to 58.6% in the most recent one (1985-89: 80.8%, 1990-95: 64.4% and 1996-2003: 58.6%). Using the midpoint of 3 year time intervals it was shown that as time increased individuals were less likely to get infected with an A type virus (OR=0.943, (CI95%=0.895-0.993), p= 0.0253). While the frequency of subtype A has decreased over time other non-A subtypes have increased. Subtype D shows a significant increase from the early epidemic (1985-89) compared to the collapsed time frame of 1990-2003, while subtype G viruses were only detected in the 1996-2003 time period.

There are many possible explanations for the variability in subtype distribution that is seen within the cohort. The epidemic has expanded in Africa. Clients of the Pumwani sex workers may increasingly represent individuals already HIV positive and whose infections may have occurred in parts of Africa where other subtypes predominate, such as in the surrounding countries of Tanzania and Uganda where subtypes C and D are more prevalent. These clients may have brought with them an increased number of non-A subtype viruses resulting in a rise in non-A infections within the sex worker cohort. Conversely, as the epidemic has evolved in Africa many of the sex workers from the later time periods may have been initially infected in regions where non-A subtypes predominate, thereby influencing viral distributions in the cohort population. Following introduction of different subtypes it is also possible that variability in transmission efficiencies may affect distribution dynamics over time. Another hypothesis is supported by reports that changes in viral genetics may be driven by HLA restricted immune responses at a population level (58). Given that subtype A is the most common subtype, and likely the founding virus in Kenya, it may also have had the greatest amount of time to adapt to common HLA types within the population. This adaptation could allow for

increased immune evasion and more rapid disease progression in individuals with common HLA types, resulting in less opportunity for continued spread. These are just a few possible projections as for why subtype distributions have changed over time.

The prevalence of subtype C within the cohort is worthy of discussion in view of the expansion of this subtype seen in other geographic regions. Subtype C is currently the most prevalent subtype worldwide and is responsible for more than half of all HIV infections (243;244). This specific subtype has also become the predominant strain in epidemics that were initially established by other subtypes, such as in the Democratic Republic of Congo (1997-2002) (102) and in southern Brazil (101). Interestingly, although the proportion of C subtype viruses seen in the ML cohort did increase over time (from 4.1% (1985-89) to 5.1% (1990-1995) to 11.4% (1996-2002) of infections) the observed differences were not found to be significant. Clearly, although present in the early time period of study (1985-89), subtype C has not shown an explosive growth within this region of Kenya.

Within our study we found that subtype distributions showed significant fluctuations over time. One limitation however is the fact that not all individuals within the three time periods studied were seroconverters with defined infection dates. Therefore although it is definite that those individuals within the first group (1985-89) became infected prior to 1990 and represent the early epidemic, exact infection dates for individuals who entered the cohort positive within the other two groups (1990-95) and (1996-03) and who were not seroconverters cannot be confirmed. Regardless, we suggest that the three groupings remain representative of different increasing segments of time within the growing epidemic.

The significant changes seen in the ML cohort's viral distribution emphasize the constant need to monitor the circulating viral subtypes both regionally and globally. Vaccine development strategies, antiretroviral treatment, and the investigation and subsequent interpretation of research questions may all be impacted by alterations in subtype distributions within populations.

Unique group of late seroconverters does not have a unique virus profile

With evidence of extensive genetic diversity and changes in this diversity over time, it seemed possible that the distribution of such subtypes may not necessarily be homogeneous between all groups of individuals within the cohort. We predicted that a unique group of individuals, the late seroconverters, would have a different subtype distribution compared to the entire cohort. This was based on the observation that their resistance to HIV, prior to seroconversion, has in part been attributed to an HIV specific CTL response (170;171;221). We predicted that they would therefore be more likely to become infected with non-A subtype viruses to which they likely have less exposure to and perhaps a less effective response against. An analysis of the viruses infecting 21 of these women however did not reveal any significant differences between the proportions of each virus type seen and those seen for the rest of the women within the study. This may be explained by a number of factors. Firstly, seroconversion may only occur once the CTL response has waned due to a decrease in sex work and a consequent decrease in HIV exposure (169). Additionally the host immune response may not be significantly more effective against one subtype over the others and consequently preferential infection with less common subtypes would not be seen. One limitation to this analysis is

that since we do not have full length sequence data we cannot rule out the possibility that some of these viruses may be recombinant strains.

No differences in disease progression associated with subtype

The diverse mix of viruses identified within the ML cohort and available CD4+ T cell count data allowed for an investigation into the possible existence of subtype specific associations with disease progression. Disease progression for subtypes A, D, C and recombinant viruses (R) were compared using Kaplan-Meier survival analysis. Survival curves of time to the event of CD4+ T cell counts below 400 cells/mm^3 and 200 cells/mm^3 were compared. As is detailed in the results section, we found an absence of any significant correlation between subtype and these markers of disease progression. Previously reported studies that investigated the role of viral subtype on disease progression show variable results (90;91); however 2 studies in different parts of Africa have reported increased disease progression in those infected with subtype D compared to subtype A (92;245). Although we found that subtype D had the shortest mean time to a CD4+ T cell count below 200 cells/mm^3 compared to A, C and recombinant viruses, differences in the survival curves were not statistically significant. It is possible that an even larger study could increase the small differences observed but we would argue that if differences do exist they are likely inconsequential. Further to this, different populations have varied host genetics and different region specific subtype variants and thus subtype-disease progression associations may be region dependent.

One possible limitation to our study is that CD4+ T cell counts could not be measured from the start of infection (seroconversion) for all individuals. However,

average initial CD4+ T cell counts were not significantly different between groups and so we suggest that the findings are not significantly impacted by this limitation.

Our investigation of subtypes and disease progression included individuals who entered the cohort from1985- 2003, and although we did not see significant differences in progression between subtypes, we decided to investigate the possibility that changes in disease progression could occur for a single subtype over time. We addressed this question for subtype A whose prevalence over time decreased significantly within the cohort. Indeed, we found that the Kaplan-Meier curves for time to drop below 400 cells/mm^3 over the three time periods (85-89, 90-95, 96-03) were significantly different. Time to progression decreased with each period. Differences were not significant for the time to drop below 200 cells/mm^3 curves but a trend was observed which followed the same pattern. One factor that might have biased our results was the time that individuals were infected prior to entry into the study. Seroconversion dates were not available for most individuals. We therefore evaluated the time between prostitution start and the first CD4+ T cell count or seroconversion date and the first CD4+ T cell count as measures of possible infection time prior to entry in the cohort. Differences in the mean infection times for the individuals studied in the different time periods (1985-89,1990-95, 1996-2003) for both the 400 cells/mm^3 and 200 cells/mm^3 study groups were statistically significant. However, the pattern seen did not follow that which was seen for the Kaplan-Meier curve analysis. These findings therefore suggest that disease progression may have increased in individuals infected with subtype A over the three time periods. Interestingly, this pattern parallels the decrease in the proportion of viruses that are subtype A within the cohort over the same three time frames. One possible explanation

for this finding is that, as mentioned earlier, evolution of the virus driven by HLA restricted immune responses might have occurred over time. This could result in faster disease progression and consequently decreased transmission. HLA has been shown to play a role in viral evolution at a population level (58). Although not detailed within this thesis, I have found significant associations between HLA alleles and changes, or protection from change, at particular amino acid positions within protease of subtype A viruses at a population level. These data are intriguing and should be expanded and investigated further.

Although subtype associated differences in disease progression were not found this need not diminish the role that viral genetics may play in disease outcome. Subtype distinction is perhaps a less important contributor to disease progression than other viral genetic factors that are not specifically associated with subtype classification. One of the well established examples includes viruses with altered nef genes that are associated with long term non-progression (87). Population based sequence data revealed that one ML woman had a stop codon in the protease amino acid sequence at both 1995 and 2002 dates studied. This individual had both low viral loads and was a long term non-progressor with CD4+ T cell counts remaining well above 500 cells/mm^3 for over 11 years. This individual entered the cohort positive in 1994 but showed no signs of progressive CD4+ T cell decline (in 2005 a count of 952 cells/mm^3 was seen). A stop codon would render the viral protease inactive. The consequent lack of protease activity, or diminished activity (as some functional viruses likely exist without the stop codon) would result in an inability to form mature virus particles as the Gag and Gag-Pol polyproteins could not be cleaved. Additional factors may contribute to this individual's

non-progression but a deficient viral protease could well explain the observed low viral loads and non-progression.

Testing of the first hypothesis has revealed that there are multiple circulating subtypes within the cohort and that recombinant forms are common. Additionally, differences in disease progression based on subtype were not seen. These data set the stage for the following two sections pertaining to superinfection and the virological and clinical consequences of dual and superinfections.

Section II

Superinfection

HIV-1 infection provides protection against superinfection

There is much debate about whether infection with HIV-1 will provide protection against infection with a second HIV-1 virus following the establishment of an immune response to the first. This next section of the study attempted to address this question by testing our hypothesis, **that infection with HIV-1 will provide protection against HIV-1 superinfection.** As it is critical to the understanding of the remainder of the discussion a few key definitions will be revisited. HIV-1 dual infections, in which more than one virus is seen within an individual, may occur due to co-infection with two different viruses at the time of transmission or due to a second infection shortly thereafter but before an immune response is generated. Dual infections may also occur due to HIV-1 superinfection. Superinfection is the acquisition of a second HIV-1 variant only *after* the generation of an immune response to the first. (Typically the full establishment of an immune response takes 2-3 months.) A nested cohort study was designed to test the

hypothesis. This study used serial sample testing to follow individuals (minimum of 2 samples/patient) instead of a cross-sectional analysis. This approach was chosen due to fluctuations which may occur in viral types over time. Individuals were also not selected based on clinical progression status as we did not want to bias our assessment of clinical consequences which are addressed in section III. Furthermore, the investigation did not skew the assessment of general superinfection rates within the population through the investigation of individuals infected with only rarer, non-A initial infections. The women in our study were all also treatment naïve. Many of the individual superinfection cases described to date occurred in individuals who were or had received treatment (184;188). Such treatment may impact upon the immune response that is generated and therefore make these individuals more susceptible to superinfection than would naturally occur.

Our nested cohort superinfection study revealed four cases of superinfection over 736.7 person years of observation. Two different types of statistical analyses were used to compare superinfection results to HIV infection within the general cohort of women who entered HIV negative. Poisson regression revealed a crude unadjusted odds ratio of (OR=0.16, (CI95% 0.11-0.25), $p < 0.0001$) for the superinfection cohort compared to the general cohort. This indicates that those in the general cohort were 6.14 times more likely to seroconvert compared to an HIV positive individual becoming superinfected and that being HIV infected has a protective effect against acquisition of a subsequent HIV-1 infection. Given that there are other variables which may impact upon infection, multiple poisson regression was used to adjust for these factors (see Table 21 and Table 22). When adjusted for only those parameters that were found to be significant, based on the initial multiple poisson regression analysis (condom use, sex partners per day, prostitution

duration (5-15 years) and age >35), HIV positivity was still shown to be protective against HIV-1 superinfection (OR=0.18 (CI95% 0.12-0.28) p< 0.0001). We were also able to demonstrate that when condoms were never or seldom used individuals were 3.27 times more likely to become infected or superinfected. Similarly when sex partners per day were 5 or more, individuals were 1.77 times more likely to be infected. As expected, less protection and increased sexual exposure are associated with increased infection. Individuals who were older than 35 years at time of entry into the cohort were also less likely to become infected /superinfected when compared to individuals who were younger (26-35). This analysis further revealed that prostitution duration of 5-15 years, prior to entry in the cohort, was associated with a tendency to become infected/superinfected; compared to those with a prostitution duration of greater than 15 years. Fowke *et al.* have previously reported that, within the ML cohort, an individual's risk of infection decreases with time. This has been attributed to unequal susceptibility to infection or acquired immunity such that some individuals within the ML cohort are resistant to HIV infection (219). Had the data been available for all individuals it would also have been useful to look at both contraceptive use and other sexually transmitted infections, as both have been associated with an increased risk of HIV-1 infection (Plummer 1991).

A crude assessment using Kaplan-Meier survival curves revealed highly statistically significant differences between the curves for HIV positive individuals (to event of superinfection) and HIV negative individuals (to event of seroconversion) (Log-Rank p<0.0001, Wilcoxon p<0.0001). The cumulative percentage of individuals who remained infection/superinfection free over time was much greater for the superinfection

(HIV positive) cohort. Multivariate Cox's proportional hazard modeling, controlling for all significant variables also revealed that being HIV positive provides protection against superinfection (HR=0.14 (CI95% 0.05-0.39) p= 0.0002). Condom use, which was used as a time dependent variable once again revealed protection against infection (HR=0.35 (CI95% 0.24-0.51) p<0.0001). Sex partners per day was also used as a time dependent variable but no significant association was found, this can likely be attributed to the fluctuation seen in these numbers for each individual over time. Being older than 35 years at time of entry was shown to be protective against infection/superinfection (HR=0.24 (CI95% 0.14-0.43) p<0.0001) while prostitution duration of less than 5 years was a risk factor for infection/superinfection (HR=2.92 (CI95% 1.83-4.66) p <0.0001). As discussed above the association between prostitution duration for a shorter period of time and infection can be attributed to heterogeneity with respect to susceptibility to HIV infection within the cohort (219).

In summary, two types of statistical analyses, multiple Poisson regression and Cox's proportional hazard modeling both drew a definitive conclusion regarding protection against superinfection. As predicted by the study hypothesis and after all other significant variables were adjusted for, **being HIV positive provides significant protection against HIV-1 superinfection.** Although we know that superinfection can occur, our study provides indirect evidence that the immune response generated against an HIV-1 infection provides some protection against a subsequent one. In more cases than not, individuals are able to ward off a superinfection event than succumb to one.

Reports of same subtype superinfection events have lead to the argument by some that any protection against superinfection with different subtypes is extremely unlikely.

Our findings suggest that this is not the case. Based on the data from section I, we know that there exists a number of circulating viruses to which individual's within the cohort would be exposed. Given the frequency of the various subtypes and the predicted number of unprotected exposures each year (60 or more), exposure to multiple subtypes is inevitable. The emergence of recombinant forms, which are common within the cohort, and discussed in Section I, creates an even further diverse array of viruses to which individuals may be exposed. Still, very few superinfection events were detected. Our findings also gain support from other studies on the women of the ML cohort. There is evidence that many of the ML women are capable of generating multi-subtype responses. McKinnon et al. investigated CD8+ T cell responses to gp120 (Env) and found that 81% of women (n=47) responded to a minimum of 2 out of 4 subtypes (A,B,C,D) while 30% responded to all four (83). These data lend further credence to the idea that the immune response against one infection can protect against another of a different subtype. Moreover, women resistant to infection exist within the ML cohort. One theory to explain this resistance, supported by immunological studies, is that resistant women have a specific immune response to HIV-1 due to continued exposure (170;171;221). Some of these resistant sex workers have subsequently become infected after temporary cessation from sex work (169). This discovery suggests that constant exposure to HIV-1 may be needed to maintain the protective CTL response. Within this same ML cohort there is evidence of both multi-subtype responses and of the ability of HIV specific CTL responses to protect against infection.

In all cases reported in this thesis, superinfection events occurred with different viral types than were present at the time of initial infection. In one case (ML47) an

individual was initially infected with a subtype A virus and a D/A recombinant form was later detected. Sequence similarity reveals that this individual was likely infected with a subtype D virus that later resulted in the formation of a D/A recombinant form (figure 13). In the case of ML1787, infection with a subtype C virus was followed by superinfection with a subtype A virus with the consequent formation of recombinant viruses. In both ML 1449 and 1295 initial A subtype viruses were seen followed by detection of D/A recombinant forms. It could be argued that intrasubtype infections would be more difficult to detect and perhaps missed if the superinfection was caused by a virus extremely similar to the first. This is certainly a possibility that cannot be completely disregarded. However, with the exception of dual/superinfection cases, the viral sequence data obtained from each individual at multiple time points were more similar to each other than to the sequence data from any other ML or reference strain. This suggests that intrasubtype superinfection did not occur and is less common than intersubtype superinfection.

The argument that superinfection is a frequent event has been largely driven by the high prevalence of recombinant viruses seen and by select superinfection studies within animal models. Although chimpanzees can be superinfected following an established initial infection the relevance of these models is doubtful. Extremely high virus titer challenges were used in these studies and the infection efficiency was far greater than what has been observed for humans (208). Additionally, there are explanations for the high prevalence of recombinant viruses that are seen which do not argue against protection from superinfection. Recombinant viruses identified within the ML cohort, as discussed above, indicate that recombinants have been present since early

in the epidemic. They have had time to expand and disseminate within the population. One report in the literature even reveals that recombinant viruses were likely founding viruses of the early epidemic in central Africa (246). Since recombinant viruses are also highly prevalent in non high risk groups (247) this supports the notion that recombinants are widespread within the population and are not isolated to groups that are at greater risk of dual/superinfection. The replacement of non-recombinant founder viruses, within some regions, with an increasing number of recombinant forms also suggests that they may be fitter in most instances. The very fact that recombinant viruses, formed within individuals, frequently replace the non-recombinant viruses (as discussed in section III) also lends credence to this idea.

Although superinfection events certainly contribute to the formation and spread of recombinant viruses, co-infections may be responsible for the majority of those that are formed and disseminate within the population. The high exposure of ML sex workers to HIV-1 make it probable that they would encounter multiple viruses within the time frame in which an immune response to a first HIV infection has not yet developed. These sex workers average 5 sex partners per day and the period of time to reach a peak CTL response is 9-12 weeks; this allows plenty of opportunity for co-infections and the formation of recombinant strains to occur. Further to this, since it is well established that transmission during acute infection is 8-10 times more likely than during the asymptomatic phase (248), the spread of recombinant/dual infections may be primarily due to co-infected individuals or those initially infected with recombinant forms. Evidence supporting the transmission of more than one virus from a dually infected individual to his wife, through heterosexual contact and then through vertical

transmission to their child, illustrates that co-transmission of multiple viruses occurs within the population (249). Based on reports in the literature, we know that dual infections are not restricted to high risk sex workers within Kenya (250). Additionally, mathematical modeling by one research group provides theoretical evidence that infection with a second virus restricted to early infection could account for the high proportion of recombinant viruses that are seen and also the predominance of single subtypes. They also suggest that single subtypes would not be maintained if superinfection was widespread (208).

Studies on macaques have revealed only a small window of opportunity (8 weeks) in which these animals could be infected with a second virus. This finding sustains the idea that dual infections are most likely the result of co-infections prior to the establishment of an immune response (197). Our studies however indicate that superinfection events, although not common, can occur even well after an established infection. In the four cases identified, all individuals had been infected for a minimum time of two years prior to the superinfection event. We therefore propose that additional periods of opportunity may exist in which superinfection is more likely to occur, such as during a drop in CD4 + T cell counts. Additionally, there may be heterogeneity within the population making some individuals more or less susceptible to superinfection. This is certainly the case that pertains to the acquisition of an initial infection within women of the ML cohort (219).

The data contained within this thesis have provided a greater understanding of the viral consequences relating superinfection events but also illuminates the limitations which are inherent to studies on both superinfection and dual infections. It would be

remiss not to provide a discussion of these here. The in depth investigation of 4 cases of superinfection provided the opportunity to observe both the *in vivo* formation of recombinant viruses and also the fluctuations in viral types which may occur. In particular, the investigation into the superinfection event within ML1787 provided a detailed description of the *in vivo* viral composition and expansion of a recombinant form such that over the course of years parental viruses could no longer be detected. The progressive increase in a recombinant form was seen. An initial recombinant prevalence of 17% and the progressive expansion to 100% of the viral population, based on extensive cloning and sequencing, was seen. The change in prevalence of the recombinant form emphasizes the concept that expansion of one form can be so great that others become undetectable. It is not difficult to postulate the possible limitations inherrant to any such study on viral genetics. One cannot completely rule out the possibility that viral forms go undetected because all present and past cells within a host cannot possibly be tested. Cells are continually eliminated and sequestered viral reservoirs may exist. Another limitation is that recombinant viruses are in fact being formed but go undetected because the region examined in this study remained the same following the recombination event. This could therefore lead to missed detection of some cases.

In an attempt to mitigate the limitations detailed above serial testing of individuals was carried out in this study (in contrast to cross-sectional analyses which have also been reported) and where possible multiple intermediate samples were tested for individuals. To further lessen the concern of missed superinfection cases and undetected recombinants a confirmatory assay analyzing a separate segment of the genome was also

performed on nearly 25% of individuals. No additional cases of superinfection were detected within the samples analyzed.

It is important to mention that one individual in the nested cohort study (ML293) was dually infected and that this dual infection was observed as early as sample availability allowed. We cannot however completely rule out the possibility that this dual infection was not the result of a superinfection event. Regardless, inclusion of this individual as a fifth superinfection case would not alter the study conclusions. Still, as a confirmatory measure I considered all dual and superinfection cases identified within the entire cohort (11/213), regardless of inclusion in the nested cohort study, as possible superinfection cases. When compared to initial HIV-1 infections, superinfections were still found to be much less frequent.

The HIV-1 specific immune response is one way by which protection against superinfection may occur. The finding that HIV-1 infection provides some protection against superinfection lends hope for future vaccine efforts. It provides indirect evidence that the HIV-specific immune response is capable of warding off an HIV infection. Nonetheless, evidence of superinfection exists and the data within this thesis provide further support that such events do occur. This need not discourage vaccine efforts, and it is important to remember that the individuals under study are those with already compromised immune systems.

Caution should be exercised before drawing parallels between superinfection studies conducted for HIV-1 and other viral systems. Important factors to consider include the main mode of viral transmission and differences in the immune response elicited. It should be noted however that in addition to HIV-1, there is evidence from

other viral systems that individuals may be infected with more than one type of the same virus at one time. Examples of this include infections with Epstein-Barr virus, cytomegalovirus and human papillomavirus (HPV) (205;251;252). However, the timing of these multiple infections, and prevalence of such events has not been sufficiently examined. Other than HIV the majority of studies on superinfection, at the organism level, have been done on hepatitis C virus (HCV). One study which looked at twenty-five newly HCV infected injection drug users over a one year period found that 20% had a superinfection event following seroconversion. Since the HCV incidence rate within the cohort was 25%, the authors concluded that, within the first year of HCV chronic infection, no cross-protective immunity develops (253). Larger and more rigorously designed studies on HCV superinfection events are required before any definitive conclusions can be made.

Although bearing a similar name to superinfection at the organism level, "superinfection resistance (SIR)" as it is commonly termed, refers to superinfection at the cellular level. HIV superinfection resistance is described as the protection of an infected cell from being re-infected with a virus closely related to the first (254). The mechanisms responsible for such protection for HIV-1 have not been completely elucidated (254;255). Superinfection exclusion at the cellular level has also been reported for other viral infections, such as for vaccinia virus (256), but has not been found to occur in reovirus (257).

The role that superinfection exclusion may play in HIV-1 superinfection at the organism level is uncertain. The mechanisms of superinfection resistance are only capable of protecting already infected cells and not all potential target cells within an

individual. Additionally, very few CD4+ cells, compared to available ones, are infected during the chronic stage of infection, leaving many cells susceptible to infection with a second virus (254). The role that superinfection exclusion plays in protection against superinfection at the organism level is therefore likely minimal.

Section III

Virological and clinical consequences of multiple infections

We hypothesized that **the majority of individuals with dual infections (superinfections or co-infections) will have evidence of recombinant viruses and an increased rate of disease progression.** In order to address this hypothesis all cases of confirmed superinfections and dual infections were studied in detail with respect to the circulating viral forms present in addition to an assessment of clinical data. A discussion pertaining to the data presented in this thesis as they relate to this hypothesis follows below.

Recombinant viruses were detected in almost all superinfection/dual infection cases

The virological consequences of confirmed cases of superinfection were investigated at multiple time points. In all cases initial non-recombinant forms were detected (based on the region sequenced) but recombinant forms were detected in all individuals at the latter time points tested. In two of the superinfection cases subtype A viruses are initially detected. In one case an A and A/D are later found while in the other case an A/D form alone was detected. There is the possibility that the A/D viruses may have been the superinfecting strains, but they may also have arisen due to recombination

following infection with a D subtype virus. In the case of ML1787, however, the detection of both parental forms rules out the possibility that superinfection occurred with the recombinant form itself. This is a rare example in which both parental sequences, and the formation of a recombinant form are detected, which is then followed by a progressive increase in the recombinant form (Figure 11). Within a little over a year the recombinant form increased from 17% to 83%. Testing of a sample almost 4 years later revealed that the recombinant form was the only variant that could be detected. ML47 is yet another case in which an A subtype virus was initially detected but then at two later time points testing revealed the existence of a D/A form alone. Although no subtype D parental virus was detected here, sequence similarity between the original subtype A and the subtype A portion of the recombinant form indicates that an *in vivo* recombination event likely occurred following infection with a D subtype virus. Recombination is undoubtedly a virological consequence of superinfection. Additionally in 3 of the 4 cases described, the recombinant form became the only detectable variant over time while in one instance A and A/D were equally represented at the last date sampled.

In addition to the confirmed superinfection cases, additional dual infection cases were also investigated. The proportions of viral forms detected in dually infected women were determined. One of these individuals, ML 293, was investigated at 7 different time points. Within the first 6 samples dating from 1993 – 1997 slight fluctuations in the two forms (A and A/D) were seen but, in all cases except one, the A/D recombinant form predominated. Incidentally, the single time point in which the recombinant form does not predominate (22.11.1995) also correlates with the highest CD4+ T cell count seen in this individual over the course of 10 years (see Figure 41). Finally, at the last date tested

(1998) the recombinant form alone was detected. We therefore conclude that of the 5 individuals that were studied at multiple time points, 4 of them had recombinant viruses which expanded to eventually become the only detectable form (at the last dates tested). The other case revealed an equal distribution of recombinant and non-recombinant forms at the last time point tested, as discussed above. The eventual predominance of recombinant forms may indicate a greater capacity to replicate *in vivo*, to infect new cells and to evade the immune response generated. Interestingly, data from section I do not indicate that recombinant viruses are associated with a faster progression of disease.

The proportions of each virus type present within dually infected individuals, for whom only one time point was available were also investigated. Of these 6 individuals, a recombinant form was found in 5 of them, with one individual having two different recombinant viruses (A/D and C/D). In two of these individuals recombinant forms predominated. In another case the recombinant and non-recombinant forms were near equal. The other 2 individuals had non-recombinant forms which predominated. One of these individuals was ML825 who was infected with subtypes A, D and A/D in the Gag/Pol region and it was the A subtype that predominated. Additional sequencing of a region of envelope that was conducted on this individual indicated a higher proportion of subtype D viruses to A. We cannot rule out the possibility of A/D recombinant forms within this individual or within the other individuals that may have occurred within different regions of the genome. It should also be revealed that within these dually infected individuals we do not know how they became infected with multiple viruses. Both viruses may have been transmitted upon infection such that the recombinant forms seen may have originated within a different individual. Another possibility includes

infection with a second virus prior to the establishment of an immune response to the first. Lastly the possibility that a superinfection event occurred cannot be ruled out. The lack of additional data on later samples did not allow us to determine if recombinant forms would eventually predominate in all of the cases seen here. Additionally, viral dynamics seen over time may differ between co-infected and superinfected individuals.

As predicted by the study hypothesis, the majority of superinfected and dually infected individuals harboured recombinant viruses. In total, of the 11 individuals who had either confirmed superinfections or were dually infected, 10 (91%) had recombinant viruses, the majority of which were A/D.

Variability in disease progression was observed

The hypothesis predicted that **increased disease progression would be seen in individuals who were superinfected/dually infected**. This prediction was based on previous observations suggesting that exposure to a more diverse array of viral epitopes may result in a greater ability to escape the immune response. A more rapid progression might also be observed given that individuals prone to faster disease progression may also be more susceptible to multiple infections.

In section I no subtype-associated differences in disease progression were observed but I also wanted to investigate the possible impact of multiple versus single virus infections. Case control studies designed to compare the dual and superinfection cases to control individuals with single virus infections resulted in study groups that were too small to draw meaningful conclusions (regarding Kaplan-Meier survival curve analysis to the events of cell counts falling below 400cells/mm^3 and 200cells/mm^3 as

described within the methods). The rate of CD4+ T cell decline was also investigated and the y axis intercepts of regression lines (theoretical CD4 starting points: as initial lower CD4+ T cell counts have been associated with poor disease outcome (258)) scrutinized but this study approach was limited in the conclusions that could be drawn. Disease progression for each superinfected and/or dually infected individual was therefore investigated by looking at CD4+ T cell data, viral loads and survival data on a case by case basis.

An individual assessment of each of the 11 cases of superinfection/dual infection and clinical outcomes were described within the results. Many of the superinfection cases described within the literature report a rise in viral load and a drop in CD4+ T cell counts surrounding the time of superinfection (218). Although these indicators may describe an initial negative impact it is also important to measure disease outcome since long term consequences have rarely been discussed within the literature. Within the four superinfection cases detected in the ML cohort, two were more informative than the others due to narrower windows in which the superinfection events occurred. The period in which ML 1787 experienced a superinfection event could be narrowed down to less than one year. Both a rise in viral load and a drop in CD4+ T cell counts around the time of superinfection were detected. It was 5 years following the superinfection event before CD4+ T cell counts were determined to be below 200 cells/mm^3. It should also be mentioned that the date at which CD4 counts were found to be below 200 cells/mm^3 was also the time point at which the recombinant virus had expanded to the point that it was the only detectable form. Within ML 1449 the estimated superinfection event date (29.03.1994) was also associated with both a significant increase in viral load and a drop

in CD4+ T cell counts at subsequent dates tested. Within this individual the drop below 200cells/mm^3 was detected at the first date following the estimated superinfection and cell counts continued to decline, reaching 60 cells/mm^3 by 1996. Here we have one case in which the development of AIDS (based on CD4 data) was nearly simultaneous to the superinfection event and another case in which 5 years passed before cell counts dropped below 200 cells/mm^3. In the other two cases the superinfection events could not be as narrowly defined. With respect to infection outcome for ML 47, CD4+ T cell counts below 200 cells/mm^3 were never reported. Even 15 years following seroconversion and 8 years following the estimated superinfection date CD4+ counts remained above 200 cells/mm^3. ML 1295 had CD4+ counts that fluctuated significantly over the course of infection but she was still alive 14 years following entry into the cohort HIV positive and 5 years after the estimated superinfection date. Although individuals who become superinfected may experience an initial toll with respect to drops in CD4+ T cell counts and viral load increases as described above, survival times and time to CD4+ T cell counts below 200 cells/mm^3 vary significantly between individuals.

One limitation in the study of clinical consequences of superinfection events is that it is nearly impossible to be certain which came first, the superinfection event or an impairment of the immune system. In other words, a drop in CD4+ T cells might make an individual more susceptible to a superinfection event, but conversely, a superinfection event might be the driving force behind a negative outcome. Other limitations in such studies could include the fact that only more pathogenic viruses may be able to superinfect, or that individuals who succumb to a superinfection event may be prone to a particular type of disease progression. Nonetheless, our results illustrate that disease

outcome and progression is not uniform. Although rapid progression to AIDS following superinfection may occur, so may survival for extensive periods of time prior to a drop in CD4 + T cell counts below 200 cells/mm^3.

Seven additional individuals were defined as having dual infections based on the samples investigated. Available CD4+ T cell, viral load and seroconversion data left certain cases more instructive than others. None of these individuals were long term non-progressors; however there was also clear evidence that many were not rapid progressors. ML1088 who was infected with multiple viral types (A, A/D, C/D), had low viral loads and did not have any reported CD4+ T cell counts below 200cells/mm^3 over the time period studied. She seroconverted in 1989 and the last available reported CD4+ T cell count was in 1994. Analyses based on the assumptions of either a co-infection from time of seroconversion or on an estimated superinfection date both reveal that CD4+ T cells remained above 200cells/mm^3 for a minimum of 5 and 3 years respectively prior to being lost to follow up. Another example involves ML293 (DI case I), whose dual infection was detected in all samples tested (1993-1998). This individual was HIV infected upon entry into the cohort in 1985 but the earliest available sample was from1993. Viral loads were high (ranging from 280,000-410,000 copies/mL for dates tested) in this individual. Her CD4+ T cell counts however remained above 200cells/mm^3 for 13 years from time of entry into the cohort. Five years passed even since the most conservative estimate of time of dual infection (1993) to the point at which CD4+ T cells went below 200cells/mm^3. Given these results, dual infections do not uniformly result in rapid clinical progression.

The most well defined case within the dual infection group was ML 825 (DI case II). In her case the sample date tested and the estimated date of seroconversion reveal that

co-infection is likely responsible for the observed dual infection. This individual rapidly progressed. The time from seroconversion to CD4 + T cell count below 200cells/mm^3 was less than 3 years. Thus although we can confidently conclude that dual infections are not synonymous with rapid progression, this case highlights some important factors. Firstly, HIV-1 co-infections may result in rapid disease progression. Secondly, that rapid progression from time of seroconversion, if co-infected, would consequently make detection of these individuals extremely difficult. Individuals who have known dates of seroconversion are limited. Also, the very fact that co-infection could result in rapid progression would limit the number of samples for testing, in comparison to long term survivors or individuals who progress more slowly. Dually infected rapid progressors may therefore be less likely to be detected in sampling of individuals within the population.

One of the largest and arguably most well described studies in the literature on disease outcome of multiple infections investigated 4 cases of HIV co-infections and one confirmed superinfection case. In this study the investigators found that all individuals progressed to CD4+ T cell counts below 200cells/mm^3 within 3.1 years of initial infection (186). Our study suggests that while rapid progression in co-infected individuals (as detailed for ML825) can occur, individuals who do *not* progress rapidly also exist. As described above our study may be less likely to detect individuals who are co-infected and who rapidly progress based on random sampling of individuals. It could however be argued that the study by Goetlieb *et al.* might be less likely to detect those who do not progress rapidly. Only dual infection cases in individuals with both seroconversion dates and AIDS endpoint data were evaluated in their study (186). These criteria may have

biased the study towards selection of individuals who survived shorter periods of time and progressed to AIDS quickly and for whom such data was thus most complete. In other words, there may have been a greater chance that long term survivors or long term non-progressors were lost to follow-up prior to reaching an AIDS endpoint and therefore not included within their study.

Individual disease progression outcomes almost certainly result from a combination of the particular viruses involved, the development of recombinant forms, HLA genetics and how cross reactive the immune response generated may be. As is the case for individuals within the ML cohort who acquire single infections this study also reveals heterogeneity in disease progression within individuals who are infected with more than one virus.

Based on the data reported here, which illustrates that superinfection events are associated with a decrease in CD4 + T cell counts and increased viral loads, and given the evidence that some dually /superinfected individuals do progress quickly to disease, safe sex practices in HIV positive concordant sexual partners should be strongly encouraged. Further to this, women within the ML cohort are just now beginning to receive antiretroviral medications. Treatment, although beneficial, may make individuals more susceptible to superinfection. This is because treatment may reduce the HIV specific immune response generated and may also increase risk taking behaviour over time. This only further emphasizes the importance of communicating safe sex practices. Moreover, the evidence presented here which reveals that recombination is very common in dually/superinfected individuals presents additional concerns. New recombinant

antiretroviral resistant strains may develop within individuals that may severely impact their treatment and health.

Summary and Future Directions:

Data on single HIV-1 infections within the cohort of commercial sex workers highlight the importance of a candidate vaccine capable of providing immunity against a diverse array of viral subtypes and recombinant forms. These data also suggest that comprehensive monitoring of the epidemic within a population is required since the viral composition can evolve over time. The design and implementation of a vaccine or any therapeutic strategy will require an understanding of a population's viral distribution. Furthermore, the existence of natural polymorphisms in non-B subtype viruses within this population, that are known to be drug resistance mutations for subtype B viruses, may be of great importance when monitoring responses to treatment and when considering the development of future pharmaceuticals.

Despite the genetic diversity seen, these data further reveal that HIV-1 infection results in protection against superinfection. When the very first superinfection case was detected researchers felt the outlook for the development of a vaccine was dismal. However, this research suggests that such cases are less common than those individuals who are able to ward off superinfection events. This observation engenders new hope for the development of an effective vaccine against HIV-1.

Recombinant viruses were detected amongst the majority of the dually and/or superinfected individuals and such forms predominated in most cases. It is possible that recombination may result in new strains that are more pathogenic or more easily

transmitted and, in the face of anti-microbial pressure, may serve as a conduit by which different drug resistant strains can rapidly arise.

It is unclear whether the decline in CD4+ T cell numbers seen in some individuals around the time of superinfection is a consequence of the superinfection event itself or what predisposes an individual to such an event. Since the protective effect that HIV-1 confers against superinfection is not absolute, HIV-1 positive individuals should be cautioned against unsafe sexual practices even with HIV-1 concordant individuals.

It is clear that disease progression in dually infected individuals is not uniform. Athough rapid progression was seen in some individuals, it certainly does not occur in all cases. The evidence therefore does not suggest that clinicians can predict disease outcome based on detection of a dual infection. Given that the clinical progression of dually/superinfected individuals was not homogeneous, the true impact of multiple virus infections is likely dependent on both host genetics and the viruses involved.

The studies reported in this thesis serve to better define viral genetics within a sex worker cohort and to answer numerous questions regarding the host-virus relationship. The data generated also provided a foundation for future avenues of study. A particular future interest will be the influence of newly implemented treatment strategies on the incidence of superinfection. Treatment may make individuals more susceptible to superinfection events as they may have a lesser HIV specific immune response. There is also the possibility that the treated individuals may be less susceptible to superinfection as CD4+ T cell counts will remain higher for longer. Future studies may also investigate the impact of natural polymorphisms within HIV-1 protease of non-B subtype viruses known to be drug resistance mutations for subtype B viruses. The impact of these

mutations on treatment outcomes and the development of drug resistance will be of considerable interest. The emergence and predominance of recombinant forms seen in dually infected individuals may also spur investigations into the impact of such infections on disease progression and the development of resistant strains in the context of treatment. Studies specifically targeting and following seroconverters may better aid in the detection of HIV-1 co-infections (as well as superinfection events) and may help to further elucidate the clinical consequences of such infections. Extensive serial sampling of these individuals may also allow for a more comprehensive understanding of the immune system dynamics that immediately surround superinfection events. Since some individuals become superinfected while others are able to avoid such events, studies designed to establish the correlates of immune protection with respect to superinfection are of significant interest.

The virus-host relationship is complex but studies that continue to elucidate elements of this relationship are critical for the development of strategies to halt the pandemic.

Reference List

(1) UNAIDS. 2004 Report on the Global AIDS Epidemic. 2004. Geneva.

(2) UNAIDS/WHO. UNAIDS/WHO AIDS Epidemic Update: December 2005. 2005.

(3) Annan K. In Africa, AIDS has a woman's face. New York Times 2002 Dec 29.

(4) UNAIDS, Unicef. Children on the Brink 2002 : A joint report on orphan estimates and program strategies. 2002.

(5) Stover J, Walker N, Garnett GP, Salomon JA, Stanecki KA, Ghys PD et al. Can we reverse the HIV/AIDS pandemic with an expanded response? Lancet 2002; 360(9326):73-77.

(6) Pomerantz RJ, Horn DL. Twenty years of therapy for HIV-1 infection. Nat Med 2003; 9(7):867-873.

(7) Veazey RS, Shattock RJ, Pope M, Kirijan JC, Jones J, Hu Q et al. Prevention of virus transmission to macaque monkeys by a vaginally applied monoclonal antibody to HIV-1 gp120. Nat Med 2003; 9(3):343-346.

(8) *Pneumocystis* Pneumonia: Los Angeles. MMWR Morb Mortal Wkly Rep 30[21], 250-252. 1981. CDC.

(9) Kaposi's sarcoma and *Pneumocystis* pneumonia among homosexual men: New York City and Califormia. MMWR Morb Mortal Wkly Rep 30, 305-308. 1981.

(10) Borkovic SP, Schwartz RA. Kaposi's sarcoma presenting in the homosexual man -- a new and striking phenomenon! Ariz Med 1981; 38(12):902-904.

(11) Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA et al. *Pneumocystis* carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N Engl J Med 1981; 305(24):1425-1431.

(12) Masur H, Michelis MA, Greene JB, Onorato I, Stouwe RA, Holzman RS et al. An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. N Engl J Med 1981; 305(24):1431-1438.

(13) Laverdiere M, Tremblay J, Lavallee R, Bonny Y, Lacombe M, Boileau J et al. AIDS in Haitian immigrants and in a Caucasian woman closely associated with Haitians. Can Med Assoc J 1983; 129(11):1209-1212.

(14) Vieira J, Frank E, Spira TJ, Landesman SH. Acquired immune deficiency in Haitians: opportunistic infections in previously healthy Haitian immigrants. N Engl J Med 1983; 308(3):125-129.

(15) Epidemiologic Notes and Reports *Pneumocystis carinii* Pneumonia among persons with Hemophilia A. MMWR Morb Mortal Wkly Rep 31[27], 365-367. 1982.

(16) Gill JC, Menitove JE, Wheeler D, Aster RH, Montgomery RR. Generalized lymphadenopathy and T cell abnormalities in hemophilia A. J Pediatr 1983; 103(1):18-22.

(17) Gold KD, Thomas L, Garrett TJ. Aggressive Kaposi's sarcoma in a heterosexual drug addict. N Engl J Med 1982; 307(8):498.

(18) Epidemiologic Notes and Reports Immunodeficiency among Female Sex Partners of Males with Acquired Immune Deficiency Syndrome(AIDS): New York. MMWR Morb Mortal Wkly Rep 31[52], 697-698. 1983.

(19) Harris C, Small CB, Klein RS, Friedland GH, Moll B, Emeson EE et al. Immunodeficiency in female sexual partners of men with the acquired immunodeficiency syndrome. N Engl J Med 1983; 308(20):1181-1184.

(20) Kreiss JK, Kitchen LW, Prince HE, Kasper CK, Essex M. Antibody to human T-lymphotropic virus type III in wives of hemophiliacs. Evidence for heterosexual transmission. Ann Intern Med 1985; 102(5):623-626.

(21) Joncas JH, Delage G, Chad Z, Lapointe N. Acquired (or congenital) immunodeficiency syndrome in infants born of Haitian mothers. N Engl J Med 1983; 308(14):842.

(22) Oleske J, Minnefor A, Cooper R, Jr., Thomas K, dela CA, Ahdieh H et al. Immune deficiency syndrome in children. JAMA 1983; 249(17):2345-2349.

(23) Ragni MV, Urbach AH, Kiernan S, Stambouli J, Cohen B, Rabin BS et al. Acquired immunodeficiency syndrome in the child of a haemophiliac. Lancet 1985; 1(8421):133-135.

(24) Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc Natl Acad Sci U S A 1980; 77(12):7415-7419.

(25) Poiesz BJ, Ruscetti FW, Reitz MS, Kalyanaraman VS, Gallo RC. Isolation of a new type C retrovirus (HTLV) in primary uncultured cells of a patient with Sezary T-cell leukaemia. Nature 1981; 294(5838):268-271.

(26) Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). 1983. Rev Invest Clin 2004; 56(2):126-129.

(27) Montanier J, Phillips P, Montessori V. Principles of Management in the Developed World. In: Cohen, Powderly, editors. Infectious Diseases. Mosby, 2004.

(28) Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS et al. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science 1983; 220(4599):865-867.

(29) Levy JA, Hoffman AD, Kramer SM, Landis JA, Shimabukuro JM, Oshiro LS. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. Science 1984; 225(4664):840-842.

(30) Coffin J, Haase A, Levy JA, Montagnier L, Oroszlan S, Teich N et al. What to call the AIDS virus? Nature 1986; 321(6065):10.

(31) Gallo RC, Montagnier L. Historical essay. Prospects for the future. Science 2002; 298(5599):1730-1731.

(32) Montagnier L. Historical essay. A history of HIV discovery. Science 2002; 298(5599):1727-1728.

(33) Weiss RA. HIV and AIDS: looking ahead. Nat Med 2003; 9(7):887-891.

(34) Freed EO, M.A.Martin. HIVs and Their Replication. Fields Virology 4th Edition. Lippincott Williams & Wilkins, 2001.

(35) Levy JA. Pathogenesis of human immunodeficiency virus infection. Microbiol Rev 1993; 57(1):183-289.

(36) Chinen J, Shearer WT. Molecular virology and immunology of HIV infection. J Allergy Clin Immunol 2002; 110(2):189-198.

(37) Frankel SS, Steinman RM, Michael NL, Kim SR, Bhardwaj N, Pope M et al. Neutralizing monoclonal antibodies block human immunodeficiency virus type 1 infection of dendritic cells and transmission to T cells. J Virol 1998; 72(12):9788-9794.

(38) Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. Annu Rev Biochem 1998; 67:1-25.

(39) Korber B, Foley BTKC, Pillaisk SK, Sodroski J.G. Numbering Positions in HIV Relative to HXB2CG. The 1998 HIV Sequence Compendium. 1998: 102.

(40) Stevenson M. HIV-1 pathogenesis. Nat Med 2003; 9(7):853-860.

(41) 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; 312(5996):763-767.

(42) Klatzmann D, Champagne E, Chamaret S, Gruest J, Guetard D, Hercend T et al. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature 1984; 312(5996):767-768.

(43) Alkhatib G, Combadiere C, Broder CC, Feng Y, Kennedy PE, Murphy PM et al. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science 1996; 272(5270):1955-1958.

(44) Choe H, Farzan M, Sun Y, Sullivan N, Rollins B, Ponath PD et al. The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. Cell 1996; 85(7):1135-1148.

(45) Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M et al. Identification of a major co-receptor for primary isolates of HIV-1. Nature 1996; 381(6584):661-666.

(46) Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, Nagashima KA et al. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. Nature 1996; 381(6584):667-673.

(47) 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; 272(5263):872-877.

(48) Emerman M, Malim MH. HIV-1 regulatory/accessory genes: keys to unraveling viral and host cell biology. Science 1998; 280(5371):1880-1884.

(49) Bebenek K, Abbotts J, Roberts JD, Wilson SH, Kunkel TA. Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. J Biol Chem 1989; 264(28):16948-16956.

(50) Boyer JC, Bebenek K, Kunkel TA. Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. Proc Natl Acad Sci U S A 1992; 89(15):6919-6923.

(51) Preston BD, Poiesz BJ, Loeb LA. Fidelity of HIV-1 reverse transcriptase. Science 1988; 242(4882):1168-1171.

(52) Coffin JM. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science 1995; 267(5197):483-489.

(53) Robertson DL, Sharp PM, McCutchan FE, Hahn BH. Recombination in HIV-1. Nature 1995; 374(6518):124-126.

(54) Goodrich DW, Duesberg PH. Retroviral recombination during reverse transcription. Proc Natl Acad Sci U S A 1990; 87(6):2052-2056.

(55) Hu WS, Temin HM. Retroviral recombination and reverse transcription. Science 1990; 250(4985):1227-1233.

(56) Stuhlmann H, Berg P. Homologous recombination of copackaged retrovirus RNAs during reverse transcription. J Virol 1992; 66(4):2378-2388.

(57) Malim MH, Emerman M. HIV-1 sequence variation: drift, shift, and attenuation. Cell 2001; 104(4):469-472.

(58) 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; 296(5572):1439-1443.

(59) Kantor R, Katzenstein DA, Efron B, Carvalho AP, Wynhoven B, Cane P et al. Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. PLoS Med 2005; 2(4):e112.

(60) Hahn BH, Shaw GM, De Cock KM, Sharp PM. AIDS as a zoonosis: scientific and public health implications. Science 2000; 287(5453):607-614.

(61) Clavel F, Guetard D, Brun-Vezinet F, Chamaret S, Rey MA, Santos-Ferreira MO et al. Isolation of a new human retrovirus from West African patients with AIDS. Science 1986; 233(4761):343-346.

(62) Clavel F, Brun-Vezinet F, Guetard D, Chamaret S, Laurent A, Rouzioux C et al. [LAV type II: a second retrovirus associated with AIDS in West Africa]. C R Acad Sci III 1986; 302(13):485-488.

(63) Schim van der Loeff MF, Aaby P. Towards a better understanding of the epidemiology of HIV-2. AIDS 1999; 13 Suppl A:S69-S84.

(64) Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF et al. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature 1999; 397(6718):436-441.

(65) Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A et al. Timing the ancestor of the HIV-1 pandemic strains. Science 2000; 288(5472):1789-1796.

(66) Gao F, Vidal N, Li Y, Trask SA, Chen Y, Kostrikis LG et al. Evidence of two distinct subsubtypes within the HIV-1 subtype A radiation. AIDS Res Hum Retroviruses 2001; 17(8):675-688.

(67) Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK et al. HIV-1 nomenclature proposal. Science 2000; 288(5463):55-56.

(68) Triques K, Bourgeois A, Vidal N, Mpoudi-Ngole E, Mulanga-Kabeya C, Nzilambi N et al. Near-full-length genome sequencing of divergent African HIV type 1 subtype F viruses leads to the identification of a new HIV type 1 subtype designated K. AIDS Res Hum Retroviruses 2000; 16(2):139-151.

(69) Peeters M, Toure-Kane C, Nkengasong JN. Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials. AIDS 2003; 17(18):2547-2560.

(70) Thomson MM, Najera R. Molecular epidemiology of HIV-1 variants in the global AIDS pandemic: an update. AIDS Rev 2005; 7(4):210-224.

(71) Wainberg MA. HIV-1 subtype distribution and the problem of drug resistance. AIDS 2004; 18 Suppl 3:S63-S68.

(72) Delwart EL, Sheppard HW, Walker BD, Goudsmit J, Mullins JI. Human immunodeficiency virus type 1 evolution in vivo tracked by DNA heteroduplex mobility assays. J Virol 1994; 68(10):6672-6683.

(73) Goulston C, Stevens E, Gallo D, Mullins JI, Hanson CV, Katzenstein D. Human immunodeficiency virus in plasma and genital secretions during the menstrual cycle. J Infect Dis 1996; 174(4):858-861.

(74) Haggerty S, Stevenson M. Predominance of distinct viral genotypes in brain and lymph node compartments of HIV-1-infected individuals. Viral Immunol 1991; 4(2):123-131.

(75) Overbaugh J, Anderson RJ, Ndinya-Achola JO, Kreiss JK. Distinct but related human immunodeficiency virus type 1 variant populations in genital secretions and blood. AIDS Res Hum Retroviruses 1996; 12(2):107-115.

(76) Poss M, Martin HL, Kreiss JK, Granville L, Chohan B, Nyange P et al. Diversity in virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1. J Virol 1995; 69(12):8118-8122.

(77) Poss M, Rodrigo AG, Gosink JJ, Learn GH, de Vange PD, Martin HL, Jr. et al. Evolution of envelope sequences from the genital tract and peripheral blood of women infected with clade A human immunodeficiency virus type 1. J Virol 1998; 72(10):8240-8251.

(78) Zhu T, Wang N, Carr A, Nam DS, Moor-Jankowski R, Cooper DA et al. Genetic characterization of human immunodeficiency virus type 1 in blood and genital secretions: evidence for viral compartmentalization and selection during sexual transmission. J Virol 1996; 70(5):3098-3107.

(79) Thomson MM, Perez-Alvarez L, Najera R. Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. Lancet Infect Dis 2002; 2(8):461-471.

(80) Novitsky V, Smith UR, Gilbert P, McLane MF, Chigwedere P, Williamson C et al. Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? J Virol 2002; 76(11):5435-5451.

(81) Williamson C, Morris L, Maughan MF, Ping LH, Dryga SA, Thomas R et al. Characterization and selection of HIV-1 subtype C isolates for use in vaccine development. AIDS Res Hum Retroviruses 2003; 19(2):133-144.

(82) Coplan PM, Gupta SB, Dubey SA, Pitisuttithum P, Nikas A, Mbewe B et al. Cross-reactivity of anti-HIV-1 T cell immune responses among the major HIV-1 clades in HIV-1-positive individuals from 4 continents. J Infect Dis 2005; 191(9):1427-1434.

(83) McKinnon LR, Ball TB, Kimani J, Wachihi C, Matu L, Luo M et al. Cross-clade CD8(+) T-cell responses with a preference for the predominant circulating clade. J Acquir Immune Defic Syndr 2005; 40(3):245-249.

(84) Julg B, Goebel FD. HIV genetic diversity: any implications for drug resistance? Infection 2005; 33(4):299-301.

(85) Descamps D, Apetrei C, Collin G, Damond F, Simon F, Brun-Vezinet F. Naturally occurring decreased susceptibility of HIV-1 subtype G to protease inhibitors. AIDS 1998; 12(9):1109-1111.

(86) Steain MC, Wang B, Dwyer DE, Saksena NK. HIV-1 co-infection, superinfection and recombination. Sex Health 2004; 1(4):239-250.

(87) Deacon NJ, Tsykin A, Solomon A, Smith K, Ludford-Menting M, Hooker DJ et al. Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients. Science 1995; 270(5238):988-991.

(88) Lazarus JP, Bhatia M, Shukla G, Padma MV, Tripathi M, Shrivastava AK et al. A study of nonepileptic seizures in an Indian population. Epilepsy Behav 2003; 4(5):496-499.

(89) Marlink R, Kanki P, Thior I, Travers K, Eisen G, Siby T et al. Reduced rate of disease development after HIV-2 infection as compared to HIV-1. Science 1994; 265(5178):1587-1590.

(90) Hu DJ, Buve A, Baggs J, van der GG, Dondero TJ. What role does HIV-1 subtype play in transmission and pathogenesis? An epidemiological perspective. AIDS 1999; 13(8):873-881.

(91) Spira S, Wainberg MA, Loemba H, Turner D, Brenner BG. Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. J Antimicrob Chemother 2003; 51(2):229-240.

(92) Kaleebu P, French N, Mahe C, Yirrell D, Watera C, Lyagoba F et al. Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda. J Infect Dis 2002; 185(9):1244-1250.

(93) Kanki PJ, Hamel DJ, Sankale JL, Hsieh C, Thior I, Barin F et al. Human immunodeficiency virus type 1 subtypes differ in disease progression. J Infect Dis 1999; 179(1):68-73.

(94) Neilson JR, John GC, Carr JK, Lewis P, Kreiss JK, Jackson S et al. Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. J Virol 1999; 73(5):4393-4403.

(95) Alaeus A, Lidman K, Bjorkman A, Giesecke J, Albert J. Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A-D. AIDS 1999; 13(8):901-907.

(96) Amornkul PN, Tansuphasawadikul S, Limpakarnjanarat K, Likanonsakul S, Young N, Eampokalap B et al. Clinical disease associated with HIV-1 subtype B' and E infection among 2104 patients in Thailand. AIDS 1999; 13(14):1963-1969.

(97) Galai N, Kalinkovich A, Burstein R, Vlahov D, Bentwich Z. African HIV-1 subtype C and rate of progression among Ethiopian immigrants in Israel. Lancet 1997; 349(9046):180-181.

(98) Kunanusont C, Foy HM, Kreiss JK, Rerks-Ngarm S, Phanuphak P, Raktham S et al. HIV-1 subtypes and male-to-female transmission in Thailand. Lancet 1995; 345(8957):1078-1083.

(99) Renjufi B, Gilbert P, Chaplin B, Msamanga G, Mwakagile D, Fawzi W et al. Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D. AIDS 2004; 18(12):1629-1636.

(100) Yang C, Li M, Newman RD, Shi YP, Ayisi J, van Eijk AM et al. Genetic diversity of HIV-1 in western Kenya: subtype-specific differences in mother-to-child transmission. AIDS 2003; 17(11):1667-1674.

(101) Soares EA, Martinez AM, Souza TM, Santos AF, Da H, V, Silveira J et al. HIV-1 subtype C dissemination in southern Brazil. AIDS 2005; 19 Suppl 4:S81-S86.

(102) Vidal N, Mulanga C, Bazepeo SE, Mwamba JK, Tshimpaka JW, Kashi M et al. Distribution of HIV-1 variants in the Democratic Republic of Congo suggests increase of subtype C in Kinshasa between 1997 and 2002. J Acquir Immune Defic Syndr 2005; 40(4):456-462.

(103) Kalish ML, Baldwin A, Raktham S, Wasi C, Luo CC, Schochetman G et al. The evolving molecular epidemiology of HIV-1 envelope subtypes in injecting drug users in Bangkok, Thailand: implications for HIV vaccine trials. AIDS 1995; 9(8):851-857.

(104) Cohen OJ, Fauci AS. Pathogenesis and Medical Aspects of HIV-1 Infection. Fields Virology 4th Edition. Lippincott Williams & Wilkins, 2001.

(105) Abbas A.K., Lichtman A.H. Congenital and Acquired Immunodeficiencies. Cellular and Molecular Immunology. 2003.

(106) Bart P, Pantaleo G. The Immunopathogenesis of HIV-1 Infection. In: Cohen, Powderly, editors. Infectious Diseases. Mosby, 2004.

(107) Brenchley JM, Price DA, Douek DC. HIV disease: fallout from a mucosal catastrophe? Nat Immunol 2006; 7(3):235-239.

(108) Anzala OA, Nagelkerke NJ, Bwayo JJ, Holton D, Moses S, Ngugi EN et al. Rapid progression to disease in African sex workers with human immunodeficiency virus type 1 infection. J Infect Dis 1995; 171(3):686-689.

(109) Farzadegan H, Henrard DR, Kleeberger CA, Schrager L, Kirby AJ, Saah AJ et al. Virologic and serologic markers of rapid progression to AIDS after HIV-1 seroconversion. J Acquir Immune Defic Syndr Hum Retrovirol 1996; 13(5):448-455.

(110) Liu SL, Schacker T, Musey L, Shriner D, McElrath MJ, Corey L et al. Divergent patterns of progression to AIDS after infection from the same source: human immunodeficiency virus type 1 evolution and antiviral responses. J Virol 1997; 71(6):4284-4295.

(111) Easterbrook PJ. Non-progression in HIV infection. AIDS 1994; 8(8):1179-1182.

(112) Easterbrook PJ, Schrager LK. Long-term nonprogression in HIV infection: methodological issues and scientific priorities. Report of an international European community-National Institutes of Health Workshop, The Royal Society, London, England, November 27-29, 1995. Scientific Coordinating Committee. AIDS Res Hum Retroviruses 1998; 14(14):1211-1228.

(113) Cao Y, Qin L, Zhang L, Safrit J, Ho DD. Virologic and immunologic characterization of long-term survivors of human immunodeficiency virus type 1 infection. N Engl J Med 1995; 332(4):201-208.

(114) Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, He T et al. The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. Nat Med 1996; 2(11):1240-1243.

(115) Mann DL, Murray C, Yarchoan R, Blattner WA, Goedert JJ. HLA antigen frequencies in HIV-1 seropositive disease-free individuals and patients with AIDS. J Acquir Immune Defic Syndr 1988; 1(1):13-17.

(116) Tang AM, Graham NM, Kirby AJ, McCall LD, Willett WC, Saah AJ. Dietary micronutrient intake and risk of progression to acquired immunodeficiency syndrome (AIDS) in human immunodeficiency virus type 1 (HIV-1)-infected homosexual men. Am J Epidemiol 1993; 138(11):937-951.

(117) Leserman J. The effects of stressful life events, coping, and cortisol on HIV infection. CNS Spectr 2003; 8(1):25-30.

(118) Pezzotti P, Phillips AN, Dorrucci M, Lepri AC, Galai N, Vlahov D et al. Category of exposure to HIV and age in the progression to AIDS: longitudinal study of 1199 people with known dates of seroconversion. HIV Italian Seroconversion Study Group. BMJ 1996; 313(7057):583-586.

(119) Blanchard A, Montagnier L, Gougeon ML. Influence of microbial infections on the progression of HIV disease. Trends Microbiol 1997; 5(8):326-331.

(120) Abbas A.K., Lichtman A.H. Introduction to Immunology. In: ., editor. Cellular and Molecular Immunology. 2003.

(121) Chaplin DD. 1. Overview of the human immune response. J Allergy Clin Immunol 2006; 117(2 Suppl):S430-S435.

(122) Goepfert PA. Making sense of the HIV immune response. Top HIV Med 2003; 11(1):4-8.

(123) Gandhi RT, Walker BD. Immunologic control of HIV-1. Annu Rev Med 2002; 53:149-172.

(124) Parren PW, Burton DR, Sattentau QJ. HIV-1 antibody--debris or virion? Nat Med 1997; 3(4):366-367.

(125) Poignard P, Sabbe R, Picchio GR, Wang M, Gulizia RJ, Katinger H et al. Neutralizing antibodies have limited effects on the control of established HIV-1 infection in vivo. Immunity 1999; 10(4):431-438.

(126) Ahmad A, Menezes J. Antibody-dependent cellular cytotoxicity in HIV infections. FASEB J 1996; 10(2):258-266.

(127) Lyerly HK, Reed DL, Matthews TJ, Langlois AJ, Ahearne PA, Petteway SR, Jr. et al. Anti-GP 120 antibodies from HIV seropositive individuals mediate broadly reactive anti-HIV ADCC. AIDS Res Hum Retroviruses 1987; 3(4):409-422.

(128) Jenkins M, Landers D, Williams-Herman D, Wara D, Viscarello RR, Hammill HA et al. Association between anti-human immunodeficiency virus type 1 (HIV-1) antibody-dependent cellular cytotoxicity antibody titers at birth and vertical transmission of HIV-1. J Infect Dis 1994; 170(2):308-312.

(129) Karnasuta C, Paris RM, Cox JH, Nitayaphan S, Pitisuttithum P, Thongcharoen P et al. Antibody-dependent cell-mediated cytotoxic responses in participants enrolled in a phase I/II ALVAC-HIV/AIDSVAX B/E prime-boost HIV-1 vaccine trial in Thailand. Vaccine 2005; 23(19):2522-2529.

(130) Ljunggren K, Moschese V, Broliden PA, Giaquinto C, Quinti I, Fenyo EM et al. Antibodies mediating cellular cytotoxicity and neutralization correlate with a better clinical stage in children born to human immunodeficiency virus-infected mothers. J Infect Dis 1990; 161(2):198-202.

(131) Pugatch D, Sullivan JL, Pikora CA, Luzuriaga K. Delayed generation of antibodies mediating human immunodeficiency virus type 1-specific antibody-dependent cellular cytotoxicity in vertically infected infants. WITS Study Group. Women and Infants Transmission Study. J Infect Dis 1997; 176(3):643-648.

(132) Mascola JR. Defining the protective antibody response for HIV-1. Curr Mol Med 2003; 3(3):209-216.

(133) Julg B, Goebel FD. What's new in HIV/AIDS? Neutralizing HIV antibodies: do they really protect? Infection 2005; 33(5-6):405-407.

(134) Albert J, Abrahamsson B, Nagy K, Aurelius E, Gaines H, Nystrom G et al. Rapid development of isolate-specific neutralizing antibodies after primary HIV-1 infection and consequent emergence of virus variants which resist neutralization by autologous sera. AIDS 1990; 4(2):107-112.

(135) Parren PW, Moore JP, Burton DR, Sattentau QJ. The neutralizing antibody response to HIV-1: viral evasion and escape from humoral immunity. AIDS 1999; 13 Suppl A:S137-S162.

(136) Richman DD, Wrin T, Little SJ, Petropoulos CJ. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. Proc Natl Acad Sci U S A 2003; 100(7):4144-4149.

(137) Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X et al. Antibody neutralization and escape by HIV-1. Nature 2003; 422(6929):307-312.

(138) Letvin NL, Walker BD. Immunopathogenesis and immunotherapy in AIDS virus infections. Nat Med 2003; 9(7):861-866.

(139) Pilgrim AK, Pantaleo G, Cohen OJ, Fink LM, Zhou JY, Zhou JT et al. Neutralizing antibody responses to human immunodeficiency virus type 1 in primary infection and long-term-nonprogressive infection. J Infect Dis 1997; 176(4):924-932.

(140) Montefiori DC. Neutralizing antibodies take a swipe at HIV in vivo. Nat Med 2005; 11(6):593-594.

(141) Trkola A, Kuster H, Rusert P, Joos B, Fischer M, Leemann C et al. Delay of HIV-1 rebound after cessation of antiretroviral therapy through passive transfer of human neutralizing antibodies. Nat Med 2005; 11(6):615-622.

(142) Baba TW, Liska V, Hofmann-Lehmann R, Vlasak J, Xu W, Ayehunie S et al. Human neutralizing monoclonal antibodies of the IgG1 subtype protect against mucosal simian-human immunodeficiency virus infection. Nat Med 2000; 6(2):200-206.

(143) Haigwood NL, Watson A, Sutton WF, McClure J, Lewis A, Ranchalis J et al. Passive immune globulin therapy in the SIV/macaque model: early intervention can alter disease profile. Immunol Lett 1996; 51(1-2):107-114.

(144) Mascola JR, Stiegler G, Vancott TC, Katinger H, Carpenter CB, Hanson CE et al. Protection of macaques against vaginal transmission of a pathogenic HIV-1/SIV chimeric virus by passive infusion of neutralizing antibodies. Nat Med 2000; 6(2):207-210.

(145) Letvin NL. Progress toward an HIV vaccine. Annu Rev Med 2005; 56:213-223.

(146) Abbas A.K., Lichtman A.H. The Major Histocompatibility Complex. Cellular and Molecular Immunology. 2003.

(147) McKay PF, Barouch DH, Schmitz JE, Veazey RS, Gorgone DA, Lifton MA et al. Global dysfunction of CD4 T-lymphocyte cytokine expression in simian-human immunodeficiency virus/SIV-infected monkeys is prevented by vaccination. J Virol 2003; 77(8):4695-4702.

(148) Rosenberg ES, Billingsley JM, Caliendo AM, Boswell SL, Sax PE, Kalams SA et al. Vigorous HIV-1-specific CD4+ T cell responses associated with control of viremia. Science 1997; 278(5342):1447-1450.

(149) Tsubota H, Lord CI, Watkins DI, Morimoto C, Letvin NL. A cytotoxic T lymphocyte inhibits acquired immunodeficiency syndrome virus replication in peripheral blood lymphocytes. J Exp Med 1989; 169(4):1421-1434.

(150) Walker CM, Moody DJ, Stites DP, Levy JA. CD8+ lymphocytes can control HIV infection in vitro by suppressing virus replication. Science 1986; 234(4783):1563-1566.

(151) Yang OO, Kalams SA, Rosenzweig M, Trocha A, Jones N, Koziel M et al. Efficient lysis of human immunodeficiency virus type 1-infected cells by cytotoxic T lymphocytes. J Virol 1996; 70(9):5799-5806.

(152) Yang OO, Kalams SA, Trocha A, Cao H, Luster A, Johnson RP et al. Suppression of human immunodeficiency virus type 1 replication by CD8+ cells: evidence for HLA class I-restricted triggering of cytolytic and noncytolytic mechanisms. J Virol 1997; 71(4):3120-3128.

(153) Hadida F, Vieillard V, Mollet L, Clark-Lewis I, Baggiolini M, Debre P. Cutting edge: RANTES regulates Fas ligand expression and killing by HIV-specific CD8 cytotoxic T cells. J Immunol 1999; 163(3):1105-1109.

(154) Fujiwara M, Takata H, Oka S, Tomiyama H, Takiguchi M. Patterns of cytokine production in human immunodeficiency virus type 1 (HIV-1)-specific human CD8+ T cells after stimulation with HIV-1-infected CD4+ T cells. J Virol 2005; 79(19):12536-12543.

(155) 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; 270(5243):1811-1815.

(156) Wagner L, Yang OO, Garcia-Zepeda EA, Ge Y, Kalams SA, Walker BD et al. Beta-chemokines are released from HIV-1-specific cytolytic T-cell granules complexed to proteoglycans. Nature 1998; 391(6670):908-911.

(157) Zhang L, Yu W, He T, Yu J, Caffrey RE, Dalmasso EA et al. Contribution of human alpha-defensin 1, 2, and 3 to the anti-HIV-1 activity of CD8 antiviral factor. Science 2002; 298(5595):995-1000.

(158) Addo MM, Altfeld M, Rosenberg ES, Eldridge RL, Philips MN, Habeeb K et al. The HIV-1 regulatory proteins Tat and Rev are frequently targeted by cytotoxic T lymphocytes derived from HIV-1-infected individuals. Proc Natl Acad Sci U S A 2001; 98(4):1781-1786.

(159) Walker BD, Chakrabarti S, Moss B, Paradis TJ, Flynn T, Durno AG et al. HIV-specific cytotoxic T lymphocytes in seropositive individuals. Nature 1987; 328(6128):345-348.

(160) Walker BD, Flexner C, Paradis TJ, Fuller TC, Hirsch MS, Schooley RT et al. HIV-1 reverse transcriptase is a target for cytotoxic T lymphocytes in infected individuals. Science 1988; 240(4848):64-66.

(161) 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. J Virol 1994; 68(9):6103-6110.

(162) Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W et al. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. J Virol 1994; 68(7):4650-4655.

(163) Ogg GS, Jin X, Bonhoeffer S, Dunbar PR, Nowak MA, Monard S et al. Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. Science 1998; 279(5359):2103-2106.

(164) Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA et al. Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. Science 1999; 283(5403):857-860.

(165) Jin X, Bauer DE, Tuttleton SE, Lewin S, Gettie A, Blanchard J et al. Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques. J Exp Med 1999; 189(6):991-998.

(166) Barouch DH, Santra S, Schmitz JE, Kuroda MJ, Fu TM, Wagner W et al. Control of viremia and prevention of clinical AIDS in rhesus monkeys by cytokine-augmented DNA vaccination. Science 2000; 290(5491):486-492.

(167) Seth A, Ourmanov I, Schmitz JE, Kuroda MJ, Lifton MA, Nickerson CE et al. Immunization with a modified vaccinia virus expressing simian immunodeficiency virus (SIV) Gag-Pol primes for an anamnestic Gag-specific cytotoxic T-lymphocyte response and is associated with reduction of viremia after SIV challenge. J Virol 2000; 74(6):2502-2509.

(168) Shiver JW, Fu TM, Chen L, Casimiro DR, Davies ME, Evans RK et al. Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity. Nature 2002; 415(6869):331-335.

(169) Kaul R, Rowland-Jones SL, Kimani J, Dong T, Yang HB, Kiama P et al. Late seroconversion in HIV-resistant Nairobi prostitutes despite pre-existing HIV-specific CD8+ responses. J Clin Invest 2001; 107(3):341-349.

(170) Kaul R, Rowland-Jones SL, Kimani J, Fowke K, Dong T, Kiama P et al. New insights into HIV-1 specific cytotoxic T-lymphocyte responses in exposed, persistently seronegative Kenyan sex workers. Immunol Lett 2001; 79(1-2):3-13.

(171) Fowke KR, Kaul R, Rosenthal KL, Oyugi J, Kimani J, Rutherford WJ et al. HIV-1-specific cellular immune responses among HIV-1-resistant sex workers. Immunol Cell Biol 2000; 78(6):586-595.

(172) Allen TM, O'Connor DH, Jing P, Dzuris JL, Mothe BR, Vogel TU et al. Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. Nature 2000; 407(6802):386-390.

(173) Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H et al. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. Nat Med 1997; 3(2):205-211.

(174) Evans DT, O'Connor DH, Jing P, Dzuris JL, Sidney J, da Silva J et al. Virus-specific cytotoxic T-lymphocyte responses select for amino-acid variation in simian immunodeficiency virus Env and Nef. Nat Med 1999; 5(11):1270-1276.

(175) Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO et al. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. Nature 1991; 354(6353):453-459.

(176) Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, Nowak MA et al. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nat Med 1997; 3(2):212-217.

(177) Kwong PD, Doyle ML, Casper DJ, Cicala C, Leavitt SA, Majeed S et al. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. Nature 2002; 420(6916):678-682.

(178) Cohen GB, Gandhi RT, Davis DM, Mandelboim O, Chen BK, Strominger JL et al. The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. Immunity 1999; 10(6):661-671.

(179) Collins KL, Chen BK, Kalams SA, Walker BD, Baltimore D. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. Nature 1998; 391(6665):397-401.

(180) Yang OO, Nguyen PT, Kalams SA, Dorfman T, Gottlinger HG, Stewart S et al. Nef-mediated resistance of human immunodeficiency virus type 1 to antiviral cytotoxic T lymphocytes. J Virol 2002; 76(4):1626-1631.

(181) McMichael AJ, Rowland-Jones SL. Cellular immune responses to HIV. Nature 2001; 410(6831):980-987.

(182) Allen TM, Altfeld M. HIV-1 superinfection. J Allergy Clin Immunol 2003; 112(5):829-835.

(183) Goulder PJ, Walker BD. HIV-1 superinfection--a word of caution. N Engl J Med 2002; 347(10):756-758.

(184) Jost S, Bernard MC, Kaiser L, Yerly S, Hirschel B, Samri A et al. A patient with HIV-1 superinfection. N Engl J Med 2002; 347(10):731-736.

(185) Ramos A, Hu DJ, Nguyen L, Phan KO, Vanichseni S, Promadej N et al. Intersubtype human immunodeficiency virus type 1 superinfection following seroconversion to primary infection in two injection drug users. J Virol 2002; 76(15):7444-7452.

(186) Gottlieb GS, Nickle DC, Jensen MA, Wong KG, Grobler J, Li F et al. Dual HIV-1 infection associated with rapid disease progression. Lancet 2004; 363(9409):619-622.

(187) Fultz PN, Srinivasan A, Greene CR, Butler D, Swenson RB, McClure HM. Superinfection of a chimpanzee with a second strain of human immunodeficiency virus. J Virol 1987; 61(12):4026-4029.

(188) Altfeld M, Allen TM, Yu XG, Johnston MN, Agrawal D, Korber BT et al. HIV-1 superinfection despite broad CD8+ T-cell responses containing replication of the primary virus. Nature 2002; 420(6914):434-439.

(189) McMichael AJ, Rowland-Jones SL. HIV: Bad news for stop--start therapy? Nature 2002; 420(6914):371, 373.

(190) Fang G, Weiser B, Kuiken C, Philpott SM, Rowland-Jones S, Plummer F et al. Recombination following superinfection by HIV-1. AIDS 2004; 18(2):153-159.

(191) Gerhardt M, Mloka D, Tovanabutra S, Sanders-Buell E, Hoffmann O, Maboko L et al. In-depth, longitudinal analysis of viral quasispecies from an individual triply infected with late-stage human immunodeficiency virus type 1, using a multiple PCR primer approach. J Virol 2005; 79(13):8249-8261.

(192) Koelsch KK, Smith DM, Little SJ, Ignacio CC, Macaranas TR, Brown AJ et al. Clade B HIV-1 superinfection with wild-type virus after primary infection with drug-resistant clade B virus. AIDS 2003; 17(7):F11-F16.

(193) McCutchan FE, Hoelscher M, Tovanabutra S, Piyasirisilp S, Sanders-Buell E, Ramos G et al. In-depth analysis of a heterosexually acquired human immunodeficiency virus type 1 superinfection: evolution, temporal fluctuation, and intercompartment dynamics from the seronegative window period through 30 months postinfection. J Virol 2005; 79(18):11693-11704.

(194) Yang OO, Daar ES, Jamieson BD, Balamurugan A, Smith DM, Pitt JA et al. Human immunodeficiency virus type 1 clade B superinfection: evidence for differential immune containment of distinct clade B strains. J Virol 2005; 79(2):860-868.

(195) Travers K, Mboup S, Marlink R, Gueye-Nidaye A, Siby T, Thior I et al. Natural protection against HIV-1 infection provided by HIV-2. Science 1995; 268(5217):1612-1615.

(196) Daniel MD, Kirchhoff F, Czajak SC, Sehgal PK, Desrosiers RC. Protective effects of a live attenuated SIV vaccine with a deletion in the nef gene. Science 1992; 258(5090):1938-1941.

(197) Otten RA, Ellenberger DL, Adams DR, Fridlund CA, Jackson E, Pieniazek D et al. Identification of a window period for susceptibility to dual infection with two distinct human immunodeficiency virus type 2 isolates in a Macaca nemestrina (pig-tailed macaque) model. J Infect Dis 1999; 180(3):673-684.

(198) Shibata R, Siemon C, Cho MW, Arthur LO, Nigida SM, Jr., Matthews T et al. Resistance of previously infected chimpanzees to successive challenges with a heterologous intraclade B strain of human immunodeficiency virus type 1. J Virol 1996; 70(7):4361-4369.

(199) Stebbings R, Berry N, Stott J, Hull R, Walker B, Lines J et al. Vaccination with live attenuated simian immunodeficiency virus for 21 days protects against superinfection. Virology 2004; 330(1):249-260.

(200) Wyand MS, Manson K, Montefiori DC, Lifson JD, Johnson RP, Desrosiers RC. Protection by live, attenuated simian immunodeficiency virus against heterologous challenge. J Virol 1999; 73(10):8356-8363.

(201) Fultz PN. HIV-1 superinfections: omens for vaccine efficacy? AIDS 2004; 18(1):115-119.

(202) Levy JA. Is HIV superinfection worrisome? Lancet 2003; 361(9352):98-99.

(203) Bale JF, Jr., Petheram SJ, Souza IE, Murph JR. Cytomegalovirus reinfection in young children. J Pediatr 1996; 128(3):347-352.

(204) Palefsky JM, Holly EA, Ralston ML, Da Costa M, Greenblatt RM. Prevalence and risk factors for anal human papillomavirus infection in human immunodeficiency virus (HIV)-positive and high-risk HIV-negative women. J Infect Dis 2001; 183(3):383-391.

(205) Srivastava G, Wong KY, Chiang AK, Lam KY, Tao Q. Coinfection of multiple strains of Epstein-Barr virus in immunocompetent normal individuals: reassessment of the viral carrier state. Blood 2000; 95(7):2443-2445.

(206) Blackard JT, Cohen DE, Mayer KH. Human immunodeficiency virus superinfection and recombination: current state of knowledge and potential clinical consequences. Clin Infect Dis 2002; 34(8):1108-1114.

(207) Taylor JE, Korber BT. HIV-1 intra-subtype superinfection rates: estimates using a structured coalescent with recombination. Infect Genet Evol 2005; 5(1):85-95.

(208) Gross KL, Porco TC, Grant RM. HIV-1 superinfection and viral diversity. AIDS 2004; 18(11):1513-1520.

(209) Tsui R, Herring BL, Barbour JD, Grant RM, Bacchetti P, Kral A et al. Human immunodeficiency virus type 1 superinfection was not detected following 215 years of injection drug user exposure. J Virol 2004; 78(1):94-103.

(210) Hu DJ, Subbarao S, Vanichseni S, Mock PA, Ramos A, Nguyen L et al. Frequency of HIV-1 dual subtype infections, including intersubtype superinfections, among injection drug users in Bangkok, Thailand. AIDS 2005; 19(3):303-308.

(211) Gonzales MJ, Delwart E, Rhee SY, Tsui R, Zolopa AR, Taylor J et al. Lack of detectable human immunodeficiency virus type 1 superinfection during 1072 person-years of observation. J Infect Dis 2003; 188(3):397-405.

(212) Chakraborty B, Valer L, De Mendoza C, Soriano V, Quinones-Mateu ME. Failure to detect human immunodeficiency virus type 1 superinfection in 28 HIV-seroconcordant individuals with high risk of reexposure to the virus. AIDS Res Hum Retroviruses 2004; 20(9):1026-1031.

(213) Chohan B, Lavreys L, Rainwater SM, Overbaugh J. Evidence for frequent reinfection with human immunodeficiency virus type 1 of a different subtype. J Virol 2005; 79(16):10701-10708.

(214) Stephens EB, Joag SV, Atkinson B, Sahni M, Li Z, Foresman L et al. Infected macaques that controlled replication of SIVmac or nonpathogenic SHIV developed sterilizing resistance against pathogenic SHIV(KU-1). Virology 1997; 234(2):328-339.

(215) Stahl-Hennig C, Dittmer U, Nisslein T, Pekrun K, Petry H, Jurkiewicz E et al. Attenuated SIV imparts immunity to challenge with pathogenic spleen-derived SIV but cannot prevent repair of the nef deletion. Immunol Lett 1996; 51(1-2):129-135.

(216) Petry H, Dittmer U, Stahl-Hennig C, Coulibaly C, Makoschey B, Fuchs D et al. Reactivation of human immunodeficiency virus type 2 in macaques after simian immunodeficiency virus SIVmac superinfection. J Virol 1995; 69(3):1564-1574.

(217) Smith DM, Wong JK, Hightower GK, Ignacio CC, Koelsch KK, Daar ES et al. Incidence of HIV superinfection following primary infection. JAMA 2004; 292(10):1177-1178.

(218) Smith DM, Richman DD, Little SJ. HIV superinfection. J Infect Dis 2005; 192(3):438-444.

(219) Fowke KR, Nagelkerke NJ, Kimani J, Simonsen JN, Anzala AO, Bwayo JJ et al. Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. Lancet 1996; 348(9038):1347-1351.

(220) Kaul R, Dong T, Plummer FA, Kimani J, Rostron T, Kiama P et al. CD8(+) lymphocytes respond to different HIV epitopes in seronegative and infected subjects. J Clin Invest 2001; 107(10):1303-1310.

(221) Kaul R, Plummer FA, Kimani J, Dong T, Kiama P, Rostron T et al. HIV-1-specific mucosal CD8+ lymphocyte responses in the cervix of HIV-1-resistant prostitutes in Nairobi. J Immunol 2000; 164(3):1602-1611.

(222) Erlich H, Lee JS, Petersen JW, Bugawan T, DeMars R. Molecular analysis of HLA class I and class II antigen loss mutants reveals a homozygous deletion of the DR, DQ, and part of the DP region: implications for class II gene order. Hum Immunol 1986; 16(2):205-219.

(223) Fransen K, Zhong P, De Beenhouwer H, Carpels G, Peeters M, Louwagie J et al. Design and evaluation of new, highly sensitive and specific primers for polymerase chain reaction detection of HIV-1 infected primary lymphocytes. Mol Cell Probes 1994; 8(4):317-322.

(224) Dawood MR, Allan R, Fowke K, Embree J, Hammond GW. Development of oligonucleotide primers and probes against structural and regulatory genes of human immunodeficiency virus type 1 (HIV-1) and their use for amplification of HIV-1 provirus by using polymerase chain reaction. J Clin Microbiol 1992; 30(9):2279-2283.

(225) Erlich HA, Bugawan TL. HLA DNA Typing. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, editors. PCR Protocols: A guide to methods and applications. New York: Academic Press, 1990.

(226) Hillier L, Green P. OSP: a computer program for choosing PCR and DNA sequencing primers. PCR Methods Appl 1991; 1(2):124-128.

(227) Luo M, Blanchard J, Maclean I, Brunham R. Identification of a novel DRB1 allele, DRB1*1112, by sequence-based DRB typing. Tissue Antigens 2001; 57(1):85-86.

(228) Luo M, Blanchard J, Brunham K, Pan Y, Shen CX, Lu H et al. Two-step high resolution sequence-based HLA-DRB typing of exon 2 DNA with taxonomy-based sequence analysis allele assignment. Hum Immunol 2001; 62(11):1294-1310.

(229) Luo M, Embree J, Ramdahin S, Ndinya-Achola J, Njenga S, Bwayo JB et al. HLA-A and HLA-B in Kenya, Africa: allele frequencies and identification of HLA-B*1567 and HLA-B*4426. Tissue Antigens 2002; 59(5):370-380.

(230) Luo M, Blanchard J, Pan Y, Brunham K, Brunham RC. High-resolution sequence typing of HLA-DQA1 and -DQB1 exon 2 DNA with taxonomy-based sequence analysis (TBSA) allele assignment. Tissue Antigens 1999; 54(1):69-82.

(231) Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 1999; 73(1):152-160.

(232) Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 2004; 5(2):150-163.

(233) Hillis DM, Bull JJ. An empirical Test of Bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst Biol 1993; 42(2):182-192.

(234) Salminen MO, Carr JK, Robertson DL, Hegerich P, Gotte D, Koch C et al. Evolution and probable transmission of intersubtype recombinant human immunodeficiency virus type 1 in a Zambian couple. J Virol 1997; 71(4):2647-2655.

(235) Vouriot FCM. The Role of Single Nucleotide Polymorphism of the CD4 Gene in HIV-1 Disease Progression and Mother to Child Transmission. Manitoba, Canada: University of Manitoba, 2004 (MSc thesis).

(236) Fang G, Kuiken C, Weiser B, Rowland-Jones S, Plummer F, Chen CH et al. Long-term survivors in Nairobi: complete HIV-1 RNA sequences and immunogenetic associations. J Infect Dis 2004; 190(4):697-701.

(237) Lyles RH, Munoz A, Yamashita TE, Bazmi H, Detels R, Rinaldo CR et al. Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. Multicenter AIDS Cohort Study. J Infect Dis 2000; 181(3):872-880.

(238) Koesters SA, Alimonti JB, Wachihi C, Matu L, Anzala O, Kimani J et al. IL-7Ralpha expression on CD4+ T lymphocytes decreases with HIV disease progression and inversely correlates with immune activation. Eur J Immunol 2006; 36(2):336-344.

(239) PI Resistance Notes. Stanford University HIV Drug Resistance Database . 2005.

(240) Yang C, Li M, Shi YP, Winter J, van Eijk AM, Ayisi J et al. Genetic diversity and high proportion of intersubtype recombinants among HIV type 1-infected pregnant women in Kisumu, western Kenya. AIDS Res Hum Retroviruses 2004; 20(5):565-574.

(241) Bartolo I, Epalanga M, Bartolomeu J, Fonseca M, Mendes A, Gama A et al. High genetic diversity of human immunodeficiency virus type 1 in Angola. AIDS Res Hum Retroviruses 2005; 21(4):306-310.

(242) Rainwater S, DeVange S, Sagar M, Ndinya-Achola J, Mandaliya K, Kreiss JK et al. No evidence for rapid subtype C spread within an epidemic in which multiple subtypes and intersubtype recombinants circulate. AIDS Res Hum Retroviruses 2005; 21(12):1060-1065.

(243) Esparza J, Bhamarapravati N. Accelerating the development and future availability of HIV-1 vaccines: why, when, where, and how? Lancet 2000; 355(9220):2061-2066.

(244) Osmanov S, Pattou C, Walker N, Schwardlander B, Esparza J. Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. J Acquir Immune Defic Syndr 2002; 29(2):184-190.

(245) Vasan A, Renjifo B, Hertzmark E, Chaplin B, Msamanga G, Essex M et al. Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. Clin Infect Dis 2006; 42(6):843-852.

(246) Kalish ML, Robbins KE, Pieniazek D, Schaefer A, Nzilambi N, Quinn TC et al. Recombinant viruses and early global HIV-1 epidemic. Emerg Infect Dis 2004; 10(7):1227-1234.

(247) Dowling WE, Kim B, Mason CJ, Wasunna KM, Alam U, Elson L et al. Forty-one near full-length HIV-1 sequences from Kenya reveal an epidemic of subtype A and A-containing recombinants. AIDS 2002; 16(13):1809-1820.

(248) Pilcher CD, Tien HC, Eron JJ, Jr., Vernazza PL, Leu SY, Stewart PW et al. Brief but efficient: acute HIV infection and the sexual transmission of HIV. J Infect Dis 2004; 189(10):1785-1792.

(249) Janini LM, Tanuri A, Schechter M, Peralta JM, Vicente AC, Dela TN et al. Horizontal and vertical transmission of human immunodeficiency virus type 1 dual infections caused by viruses of subtypes B and C. J Infect Dis 1998; 177(1):227-231.

(250) Steain MC, Wang B, Yang C, Shi YP, Nahlen B, Lal RB et al. HIV type 1 sequence diversity and dual infections in Kenya. AIDS Res Hum Retroviruses 2005; 21(10):882-885.

(251) Chandler SH, Handsfield HH, McDougall JK. Isolation of multiple strains of cytomegalovirus from women attending a clinic for sexually transmitted disease. J Infect Dis 1987; 155(4):655-660.

(252) Langenberg A, Cone RW, McDougall J, Kiviat N, Corey L. Dual infection with human papillomavirus in a population with overt genital condylomas. J Am Acad Dermatol 1993; 28(3):434-442.

(253) Herring BL, Page-Shafer K, Tobler LH, Delwart EL. Frequent hepatitis C virus superinfection in injection drug users. J Infect Dis 2004; 190(8):1396-1403.

(254) Nethe M, Berkhout B, van der Kuyl AC. Retroviral superinfection resistance. Retrovirology 2005; 2:52.

(255) Wildum S, Schindler M, Munch J, Kirchhoff F. Contribution of Vpu, Env, and Nef to CD4 down-modulation and resistance of human immunodeficiency virus type 1-infected T cells to superinfection. J Virol 2006; 80(16):8047-8059.

(256) Christen L, Seto J, Niles EG. Superinfection exclusion of vaccinia virus in virus-infected cell cultures. Virology 1990; 174(1):35-42.

(257) Keirstead ND, Coombs KM. Absence of superinfection exclusion during asynchronous reovirus infections of mouse, monkey, and human cell lines. Virus Res 1998; 54(2):225-235.

(258) Lavreys L, Baeten JM, Chohan V, McClelland RS, Hassan WM, Richardson BA et al. Higher set point plasma viral load and more-severe acute HIV type 1 (HIV-1) illness predict mortality among high-risk HIV-1-infected African women. Clin Infect Dis 2006; 42(9):1333-1339.

(259) Sambrook J., Russell D.W. Molecular Cloning: A Laboratory Manual. 3rd ed. New York: Cold Spring Harbor, 2001.

APPENDIX 1: Reagents (259)

Chemicals:

Laboratory chemicals were ordered from Sigma or Fisher Scientific.

General Molecular Biology Solutions

10x Tris-Borate EDTA Buffer (TBE Buffer)

121.1 grams (g) Tris Base, 61.8 g Boric Acid, 3.72 g Na_2EDTA, dissolved in double distilled water (ddH_2O). Bring up to 1 litre (L) volume, pH 8.3

10mM Tris 1 mM EDTA (TE Buffer)

10 millilitres (mL) of 1M TrisHCl (pH 8.0), 2ml of 0.5 M EDTA (pH 8.0), adjust to 1L volume

*10x Phosphate Buffered Saline (PBS)

80g NaCl, 2g KCl, 14.4g Na_2PO_4 and 2.4g KH_2PO_4 in 800mL ddH2O, adjust the pH to 7.4 with HCl and then take volume up to 1L

Sodium Acetate NaOAc

408.3g of sodium acetate.3H2O in 800mL of H2O, adjust the pH to 5.2 with glacial acetic acid then take up to volume of 1L with H2O, dispense into aliquots and sterilize

10x Gel Loading Dye

25 milligrams of bromophenol blue (0.25%), 1.5 grams Ficoll (Type 400 Pharmacia), dissolved in ddH_2O up to 10ml volume

Reagents for PCR

PCR System:

The Expand High Fidelity PCR system (Roche) was used. Buffer number 2 which contains 15mM $MgCl_2$ was used for PCR unless otherwise stated. The Expand High Fidelity Enzyme mix consists of Tgo DNA polymerase, a thermostable DNA polymerase with proofreading activity and thermostable Taq DNA polymerase.

Primers

Primers were ordered from Invitrogen and were at cartridge purity, reconstituted in TE buffer at a concentration of 50μM.

DNA Ladders:

One Kilobase Ladder (Invitrogen):

Fragment sizes (bps): 12,216, 11,198, 10,180, 9,162, 8,144, 7,126, 6,108, 5,090, 4,072, 3,054, 2,036, 1,636, 1,018, 517/506, 396, 344, 298, 220, 201, 154, 134, 75

100 base pair (bp) Ladder (Invitrogen):

Fragment sizes in increments of 100 bp starting at 100 through to 1500 plus a 2072 bp fragment

Reagents for Sequencing

ABI PRISM BigDye Terminator v3.1 Cycle Sequencing Mix was used (Applied Biosystems)

Reagents for Cloning

Luria-Bertani (LB) Medium:

Add 10g bacto-tryptone, 5g bacto-yeast extract, 10g NaCl to 950 mL ddH2O, adjust pH to 7.0 and then bring volume up to 1 L with ddH20 and autoclave to sterilize

Add ampicillin before use: 200μg/mL

LB Agar for plates:

Make LB medium as described above then before autoclaving add 15g/L of bacto-agar(for plates), autoclave to sterilize. Allow this to cool and then pour 30mL cooled agar/petri plate. Add 200μg/mL of ampicillin to the cooled agar prior too pouring plates.

APPENDIX 2: Abbreviations

aa	amino acid
Ab	antibody
nAb	neutralizing antibody
AIDS	Acquired Immune Deficiency Syndrome
bp	base pair
CA	capsid protein
CCR5	c chemokine receptor 5
CD4	cluster of differentiation 4
CDC	Centers for Disease Control
cDNA	complementary DNA
CI 95	confidence interval 95%
CMI	cell mediated immunity
CRFs	circulating recombinant forms
CTL	cytotoxic T-lymphocyte
CXCR4	cx-chemokine receptor 4
DI	dual infection
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
EIA	enzyme immunoassays
ELISA	enzyme-linked immunosorbent assay
Env	envelope
EtBr	ethidium bromide
FIV	feline immunodeficiency
g	gram
Gag	group specific antigen
gp120	glycoprotein 120 kDa
gp41	glycoprotein 41kDa
HAART	highly active antiretroviral therapy
HHV-8	human herpesvirus type 8
HIV-1	human immunodeficiency virus type 1
HIV-2	human immunodeficiency virus type 2
HIV+	HIV infected
HIV-	HIV uninfected
HLA	human leukocyte antigen
HR	hazard ratio
HS	heparinase solution
HTLV	human T cell lymphotropic virus
HTMA	heteroduplex tracking mobility assay
IDU	injection drug user
IN	integrase
IV	intravenous
Kb	kilobases
LTNP	long term non-progressor
LTR	long terminal repeat
M	Molar (moles/Litre)

MA	matrix
MHC	major histocompatibility complex
ML	commercial sex worker cohort
mL	milliliter
MMWR	Morbidity and Mortality Weekly Report
mRNA	messenger RNA
n	number
NC	nucleocapsid
NJ	neighbour joining
nt	nucleotide
OR	odds ratio
ORF	open reading frame
p	probability
PBMC	peripheral blood mononuclear cells
PBS	phosphate buffered saline
PCP	*Pneumocystis carinii* pneumonia
PCR	polymerase chain reaction
PI	protease inhibitor
Pol	polymerase
PR	protease
R	recombinant
R5 virus	HIV-1 with CCR5 tropism
RFLP	restriction fragment length polymorphism
RIP	recombination identification program
RNA	ribonucleic acid
RT	reverse transcriptase
sc	seroconverter
SHIV	HIV/SIV hybrid
SI	superinfection
SIV_{CPZ}	simian immunodeficiency virus (chimpanzee)
SIV_{SMM}	simian immunodeficiency virus (sooty mangabey)
ssRNA	single stranded ribonucleic acid
STIs	sexually transmitted infections
SU	surface glycoprotein
T cell	thymus originating lymphocyte
Tm	melting temperature
TM	transmembrane protein
U	unclassified
UNAIDS	Joint United Nations Programme on HIV/AIDS
UNICEF	United Nations Children's Fund
URF	unique recombinant forms
VL	viral load
WHO	World Health Organization
X^2	Chi-square
X4 virus	HIV-1 with CXCR4 tropism
xg	times gravity