Characterisation of HIV Superinfection: Genetic Evolution and Adaptive Immune Responses

HAYLEY JANET HARVEY

Presented for the degree of Master of Science in Medicine in the Division of Medical Virology, Institute of Infectious Diseases and Molecular Medicine

Faculty of Health Sciences

University of Cape Town

March 2011

Supervisor: Prof. Carolyn Williamson

Co-Supervisor: Dr Zenda Woodman

This work is dedicated firstly to both my parents for enabling me to accomplish what I have and for their support every step of the way. Secondly to my boyfriend Shaun for always being my shoulder to lean on when I needed one and for pushing me to finish.

“The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.” – **Sir William Bragg**

Contents

Abstract

The occurrence of HIV-1 superinfection suggests that infection with one strain of HIV-1 is not always protective against infection with a different strain. This has important implications for the design of an effective HIV vaccine. In this thesis we aimed to determine the timing and frequency of intra-subtype C superinfection, and to determine if the reason for superinfection was a greater genetic distance within epitopes of the superinfecting virus compared to those of circulating strains from the same cohort.

Heteroduplex mobility assay was used together with sequencing and phylogenetic analysis of the C2-C3 region of *envelope* to determine the frequency of superinfection in a cohort of high risk women recruited during acute infection and followed for twelve months. Of the 19 women screened, two were found to be superinfected corresponding to an incidence of 10% per person year (CAP237 and CAP256). A third individual with superinfection was identified through another project (CAP281) (R. Ntale pers. Comm.). These individuals were superinfected at two, three and eleven months post primary infection respectively, with the date of superinfection calculated as midpoint between last detection of single infection and first detection of superinfection. Superinfection was always associated with a greater than 0.5 log increase in viral load.

The individuals identified as superinfected had significantly different disease profiles: CAP256 had a high viral load after primary infection (>100 000 copies /ml), and did not control viral replication after superinfection (>750 000 copies/ml). In contrast the disease progression of both CAP237 and CAP281 appeared to be unaffected by the second virus. CAP237 was a typical progressor with a viral load at set point of 10 000 copies per ml and CAP281 was a controller with viral loads maintained below 2000 copies per ml with occasional blips above this level.

To compare potential CTL epitopes in initial virus compared to the superinfecting virus, we generated whole genome sequencing from two of the women, from time points before and after superinfection. HLA restricted epitopes were mapped in both viruses to determine if epitope mismatch between the initial and superinfecting viruses may have resulted in lack of protection against superinfection. The genetic distance between the initial and superinfecting viruses of CAP256 was substantially greater (6%) than that of the superinfecting and initial viruses of CAP281 (3%). For CAP256 there was a significant difference in coverage of the superinfecting virus by the initial virus when compared to other circulating CAPRISA sequences ($p=<0.0001$). When the HLA associated epitopes of the initial and superinfecting viruses were compared in both CAP256 and CAP281 putative escape mutations were detected in both superinfecting viruses, however this was more pronounced in CAP256. It is proposed that sequence differences between the initial and superinfecting viruses in CAP256 may have resulted in an absence of immune recognition of the superinfecting virus and allowed superinfection to occur.

Finally, the pattern of recombination in this virus was examined in CAP256 to determine if recombination following superinfection could be driven by immune pressure. The recombinant virus present at six months post infection in CAP256 was predominantly composed of the superinfecting virus which was carrying several escape mutations. The regions where the primary virus was incorporated were found to contain epitopes which carried escape mutations.

The results of this study have shown that superinfection is occurring frequently. Unlike some previous studies, superinfection was not always associated with rapid disease progression. Finally, while the genetic distance between the primary and superinfecting strain may play a role in superinfection these signals were only detected in one of the two superinfected individuals examined. Further work is needed to expand the numbers to get a more accurate estimation of frequency and timing of superinfection, and to complement this work with detailed immunology to shed light on the correlates of protection from superinfection.

Abbreviations

AIDS	acquired immune deficiency syndrome
bp	bp
°C	degrees Celsius
cDNA	copy DNA
CRF	circulating recombinant form
CTL	cytotoxic T lymphocytes
DNA	deoxyribonucleic acid
EDTA	ethylenediamine tetraacetic acid
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HMA	heteroduplex mobility assay
IFN	interferon
kb	kilobases
M	Molar
µg	micrograms
µl	microlitres
mg	milligrams
MHC	major histocompatibility complex
ml	millilitres
mM	millimolar
NICD	National Institute for Communicable Diseases
PCR	polymerase chain reaction

SIV	simian immunodeficiency virus
UKZN	University of Kwa-Zulu Natal
UV	ultraviolet
Vpu	viral protein u
Vpr	viral protein r
UCT	University of Cape Town

Declaration

This thesis is composed of work carried out at the Division of Medical Virology, Institute for Infectious Disease and Molecular Medicine at the University of Cape Town (UCT), under the supervision of Professor Carolyn Williamson. This work is original and all my own, however where work of others has been used their contribution has been acknowledged in the text.

Hayley Janet Harvey Signed by candidate

March 2011

Acknowledgements

The guidance and encouragement provided by Professor Carolyn Williamson over the past three years that I have been a member of her laboratory have been greatly appreciated and I would sincerely like to thank her for this. My thanks also go to my co-supervisor Dr Zenda Woodman has helped to trouble shoot any problems encountered during the course of this work and has been a wonderful friend when things have not been going according to plan. In addition I would like to thank Professor Clive Gray, Mandla Mlotshwa and Debra Assis de Rosa for providing the immunology and HLA data as well as Dr Koleka Mlisana, Professor Salim Abdool-Karim and the CAPRISA 002 study team who manage the cohort used in this study. I would also like to thank Florette Treunicht for her help in generating the whole genomes from individual CAP256. I would also like to thank Roman for identifying CAP281 as a superinfected individual.

To my parents I would like to say that I hope I have made you proud! Thank you for providing me with the amazing opportunities that I have had and for always believing in me. You have always supported every decision I have made and guided me towards a wonderful life. I could not be where I am today without you both. To Shaun, I would like to thank you for always bringing a smile to my face when the weight of the world is bearing down on me. You are an amazing person, utterly selfless and such an inspiration. I am so grateful to have you in my life.

To my fellow colleagues in the HIV Diversity Research Group thank you for all your helpful advice and for always being able to lend a helping hand. I would especially like to thank Daniel Sheward for always being able to help when I had a problem.

Finally I would like to thank the Poliomyelitis Research Foundation, CAPRISA, the National Research Foundation and the University of Cape Town for funding my studies and giving me this opportunity.

Chapter 1: Introduction and Literature Review

1.0 Introduction

In 2009, HIV infected more than 33 million people and caused 1.8 million AIDS-related deaths (114). The most effective approach to controlling this epidemic is through vaccination. One of the major challenges facing the development of an effective vaccine is the highly diverse global viral population comprising of 9 subtypes and 48 circulating recombinant forms (CRFs) (1). The HIV-1 epidemic is becoming increasingly heterogeneous (38) and this diversity is driven, in part, by recombination between different strains due to dual infection with two independent HIV strains (43, 49). Dual infection can occur through co-infection where an individual is infected with a second strain at or close to infection with the first virus; or superinfection when an HIV-1 seropositive individual becomes infected with a second HIV strain(49).

The first cases of superinfection in humans were described in 2002 (5, 56, 86). Since then, more than 50 cases have been reported. It is not yet known what factors predispose individuals to superinfection but it has been proposed that high risk behaviour, the nature of the superinfecting virus, as well as the immune status of the host all play a role (16, 42, 118). Superinfection has implications on the development of an effective vaccine as it indicates that infection with one strain does not always protect against infection by a second strain. Consequently studies on superinfection may provide insight into the correlates of immune protection. Higher viral loads and faster disease progression have also been associated with its occurrence implying that it has pathogenic consequences for the individual (42, 50).

The aim of this thesis is to identify cases of HIV-1 superinfection in the CAPRISA 002 cohort, and to ascertain whether genetic differences between the two strains allowed superinfection to occur. The thesis also aimed to determine whether CTL responses were a driving factor behind recombination patterns between the initial and superinfecting strains.

1.1 HIV Diversity and Recombination

HIV is a rapidly evolving virus due to: firstly the error prone nature of the reverse transcriptase which results in approximately one nucleotide substitution per genome per replication cycle (82); and secondly to recombination between two different RNA templates which are co-packaged into each virion (129). In addition, substantial diversity is generated in a relatively short period of time due to the high viral turnover rate with an average generation time of 2.2 days (77, 107). The generation time refers to the period of time between release of a new virion from an infected cell and it infecting a new cell and producing progeny virions(77). Recombination is much more efficient at introducing changes compared to point mutations as through recombination numerous changes can be introduced simultaneously. In addition, the rate of recombination in HIV is estimated to be 2.8 times per genome per viral replication cycle (129) compared to one nucleotide substitution per genome per replication cycle for point mutations (82). Thus, superinfection with two phylogenetically distinct viruses will potentially result in more rapid viral evolution compared to infection with a single virus.

Table 1.1 Reported cases of HIV-1 superinfection which resulted in recombination

Citation	Time to Superinfection (Years)	HIV Subtypes	Risk Factor	Time to Recombination after Superinfection (Years)
Fang et al. (2004)(33)	>6	C after A	WSM	>1
McCutchan et al. (2005)(71)	0.5-0.75	AC after ACD recombinant	WSM	0.25
Pernas et al. (2006)(79)	~12	B after B	IDU	<1
Piantadosi et al. (2007)(80)	4.6-5.6	A1 after D/A2 recombinant	WSM	Not reported
Piantadosi et al. (2007)(80)	0.1	D after A	WSM	Not reported
Piantadosi et al. (2007)(80)	5-5.4	C after A	WSM	Not reported
Piantadosi et al. (2007)(80)	2-2.8	A after A	WSM	Not reported
Piantadosi et al. (2007)(80)	0.8-4	A after A	WSM	Not reported
Piantadosi et al. (2007)(80)	0.2-0.4	A after C/A recombinant	WSM	Not reported
Streeck et al. (2008)(102)	~1.4	B after B	Not reported	0.1
Templeton et al. (2009)(108)	0-5	B after B	IDU/WSM	Not reported

MSM, Men who have sex with men; WSM, Women who have sex with men; IDU, Intravenous Drug Users

The pathogenesis of HIV-1 may be extremely reliant on the virus's ability to maintain its diversity as it provides a means for the virus to escape selective pressures such as host immune responses and antiretroviral therapy (27). It may also prove useful for escaping evolutionary bottlenecks during later stages of the disease thus preserving the longevity of the virus (27).

While recombination is beneficial to pathogens enabling greater flexibility to respond to selective pressure, in dual infection it can also result in reduced diversity through the generation of a single recombinant strain. In an individual who was initially infected with an ACD recombinant virus and then became superinfected with an AC recombinant, there was initially an explosion of viral diversity with seven different molecular forms of the virus detected within three months of superinfection; however a recombinant form of the two infecting viruses became the dominant viral population later in infection (Table 1.1)(71).

1.2 Frequency of Superinfection

In regions where multiple HIV subtypes circulate, recombination between different subtypes can result in mosaic HIV genomes (24, 76, 103). Novel recombinant strains of the virus, generated by dual infection, are referred to as circulating recombinant forms (CRFs) and there are 48 which have been recorded. The high number of CRFs as well as Unique recombinant forms (URFs) in regions where two subtypes circulate suggests that co-infection and superinfection occurs frequently. Many attempts have been made to determine the frequency with which superinfection occurs and contrasting results have been reported with some studies finding high incidences while others were unable to detect superinfection at all (Table 1.2) (26, 28, 50, 52, 54, 80, 97, 113, 128).

Several studies have looked for the occurrence of dual infection in general without considering if the individuals are co- or superinfected. These studies are usually cross-sectional and have reported relatively high incidences of dual infection from women-who-have-sex-with-men

(WSM) cohorts of high risk individuals in Africa. In Tanzania a cross sectional study examined 600 female bar workers and found 19% of them to be dual infected (52). However, the shortcoming of this study was that analysis of a single time point post-infection could not distinguish whether the dual infections were co-infections or superinfections.

More recently in order to distinguish between co- and superinfection longitudinal studies have been carried out, these studies are the only way to means of determining accurate incidences of superinfection. In Kenya the reported incidence of superinfection was 4.3% per person year when a cohort of 21 female bar workers was followed for five years post-infection (28) and 3.7% of a cohort of 36 women over a period of six years (80). A more recent study by Templeton et al. (2009) found 17% of a female cohort in the United States of America to be superinfected over a period of six years which is equivalent to an incidence of approximately 2.8% per year (108). The results of these studies suggest high incidences of re-infection which are approaching the incidence of primary infection (Table 1.2).

Table 1.2 Studies which have examined the frequency of HIV Re-infection

Citation	Size of Cohort (No of individuals)	Longitudinal/ Cross section	Route of infection	Frequency (per person year)	Location	Incidence of Primary Infection (per person year)
Yerly et al. 2004(127)	156	Longitudinal	IDU	0.57%	Switzerland	NA
Chakraborty et al. 2004(26)	14*	Longitudinal	US/IDU	0%	Spain	NA
Smith et al. 2004(99)	78	Longitudinal	NA	5%	USA	~5%
Tsui et al. 2004(113)	37	Longitudinal	IDU	0%	USA	~0.02%
Chohan et al. 2005(28)	21	Longitudinal	WSM	4.3%	Kenya	7.8%
Herbinger et al. 2006(52)	600	Cross section	WSM	19%#	Tanzania	14%
Piantadosi et al. 2007(80)	36	Longitudinal	WSM	3.7%	Kenya	~8%
Sidat et al. 2008(97)	145	Longitudinal	MSM	0.036%	Australia	NA
Templeton et al. 2009(108)	58	Longitudinal	IDU/WSM	2.8%	USA	NA
Campbell et al. 2009(21)	8	Longitudinal	MSM	0.03%	USA	NA
Rachinger et al. 2010(84)	89	Longitudinal	MSM	0%	Netherlands	6.8%

MSM, Men who have sex with men; WSM, Women who have sex with men; IDU, Intravenous Drug Users; US, Unprotected Sex

*Couples not individuals

#Of the cohort not per person years

Compared to studies on women who have sex with men, lower frequencies of re-infection have been reported in cohorts of intravenous drug users (IDU). In the Netherlands an IDU cohort was found to have two superinfections out of the 156 individuals screened longitudinally, corresponding to an incidence of approximately 0.57% per person years (128). When intravenous drug users from San Francisco were screened for superinfection, 37 individuals were examined using samples which ranged from one to twelve years after initial infection and no dual infections were detected (112).

Similarly, three men-who-have-sex-with men (MSM) cohorts followed longitudinally in Australia, USA and Netherlands had low incidence of superinfection, 0.036%, 0.03% and 0% respectively (21, 84, 97). The MSM cohort in the USA comprised of seroconcordant couples but after 71 months of longitudinal follow up, only two couples became superinfected (21). Furthermore, a cohort of 14 HIV seroconcordant couples were followed for four years all of whom were at high risk of re-exposure to the virus and no cases of superinfection were detected (26). Therefore, even though studies have indicated that the incidence of superinfection amongst WSM approaches that of primary HIV infection (Table 1.2), re-infection within seroconcordant couples is low even though they are also considered to be at high risk of re-infection due to a number of factors such as unprotected sex and high viral loads.

The chance of detecting superinfection is influenced by the risk of infection. In Tanzania they compared the incidence of dual infection in individuals at high risk to individuals with normal risk exposure and found the incidence in the second group to be much lower at 9% compared to the high risk group at 19% (52). Furthermore the lack of superinfection in a cohort of 89 homosexual men in Amsterdam was attributed to a decrease in unsafe sexual behaviour (84). Several of the cohorts described in Table 1.2 comprise female sex workers with a high risk of infection, a likely factor driving the high incidence of primary infection which results in the observed frequency of superinfection.

1.3 Timing of HIV Superinfection in Humans

The timing of superinfection is important as it provides some clues as to the immune correlates of protection. If, for example, it occurs predominantly in the first few weeks of infection then it is occurring before the development of any neutralizing antibody responses but in the presence of a developing CTL response. However in early infection, the CTL response is initially narrow, recognizing only a limited number of epitopes and thus probably has limited cross reactivity (48, 61). Neutralizing antibodies only develop after three months post infection and to begin with are predominantly autologous, recognising only their own virus (47, 87). Neutralizing responses that recognise a diverse number of strains (broadly cross-neutralizing antibodies) usually only evolve two to three years following infection (47). Thus if superinfection occurs during the chronic stage of the disease, prior to the development of AIDS, then it is occurring in spite of relatively strong CD8+ T cell responses as well as neutralizing antibody responses. The AIDS stage of infection is associated with a compromised immune system making individuals more susceptible to opportunistic infections and thus infection with a second HIV-1 strain (34)(Figure 1.1).

Cases of superinfection in humans have been reported to occur from the first month to more than twelve years after the initial infection (Table 1.3) (10, 25, 29, 79, 80). However, there have been a limited number of studies in large cohorts which have focused on the timing of superinfection. It was initially proposed that superinfection only occurred before host immune responses were fully developed (100), however with cases of superinfection being reported during the chronic phase of infection this does not seem to be the case (5, 16, 29, 56, 80). Piantadosi et al (2007) who identified seven cases of superinfection in a high risk cohort of 36 women by longitudinal follow up for 5 years demonstrated that the timing of superinfection ranged from 53 days to 5 years after the initial infection. This illustrated that superinfection could occur at any stage of disease (80). Table 1.3 corroborates this finding as numerous studies have shown that superinfection is not restricted to a specific window of infection.

Table 1.3 Superinfection studies which have timed the occurrence of re-infection

Citation	Time to Superinfection (Years)	HIV Subtypes	Risk Factor
Jost et al. (2002)(56)	2.4	B after A/E recombinant	MSM
Altfield et al. (2002)(5)	2.8	B after B	IDU
Ramos et al. (2002)(86)	~ 0.1	B after CRF01_AE	IDU
Ramos et al. (2002)(86)	1.25-1.7	CRF01_AE after B	IDU
Smith et al. (2004)(99)	~0.5	B after B	NA
van der Kuyl et al. (2005)(119)	1.3-1.6	B after B	MSM
van der Kuyl et al. (2005)(119)	2.3	CRF01_AE after B	MSM
Yang et al. (2005)(124)	0.4	B after B	MSM
Smith et al. (2005)(101)	0.7	B after B	MSM
Chohan et al. (2005)(28)	0.7-1.1	A after D	WSM
Chohan et al. (2005)(28)	0.8-1.6	C/A recombinant after C	WSM
Chohan et al. (2005)(28)	0.3-1.3	A after D	WSM
Pernas et al. (2006)(79)	12	B after B	IDU
Piantadosi et al. (2007)(80)	4.6-5.6	A1 after D/A2 recombinant	WSM
Piantadosi et al. (2007)(80)	0.1-0.2	D after A	WSM
Piantadosi et al. (2007)(80)	5-5.4	C after A	WSM
Piantadosi et al. (2007)(80)	2-2.8	A after A	WSM
Piantadosi et al. (2007)(80)	0.8-4	A after A	WSM
Piantadosi et al. (2007)(80)	0.2-0.4	A after C/A recombinant	WSM
Templeton et al. (2009)(108)	1-5	B after B	IDU/WSM
Clerc et al. (2010)(29)	6	CRF11_cpx after B	IDU
Clerc et al. (2010)(29)	3.6	CRF11_cpx after B	IDU
Braibant et al. (2010)(16)	10	B after B	MSM

MSM, Men who have sex with men; WSM, Women who have sex with men; IDU, Intravenous Drug Users

Lack of large follow up studies over long periods of time and the wide range of differences in the timing (0.1-5.4 years) of reported cases of superinfection makes it difficult to predict when superinfection is most likely to occur. Viral and cellular kinetics may be one of the factors affecting the timing due to the availability of target cells. If the HIV-1 infected individual was exposed to a second virus in the later stages of infection after CD4+ T cells have been depleted (Figure 1.1) they may be less likely to become superinfected (118).

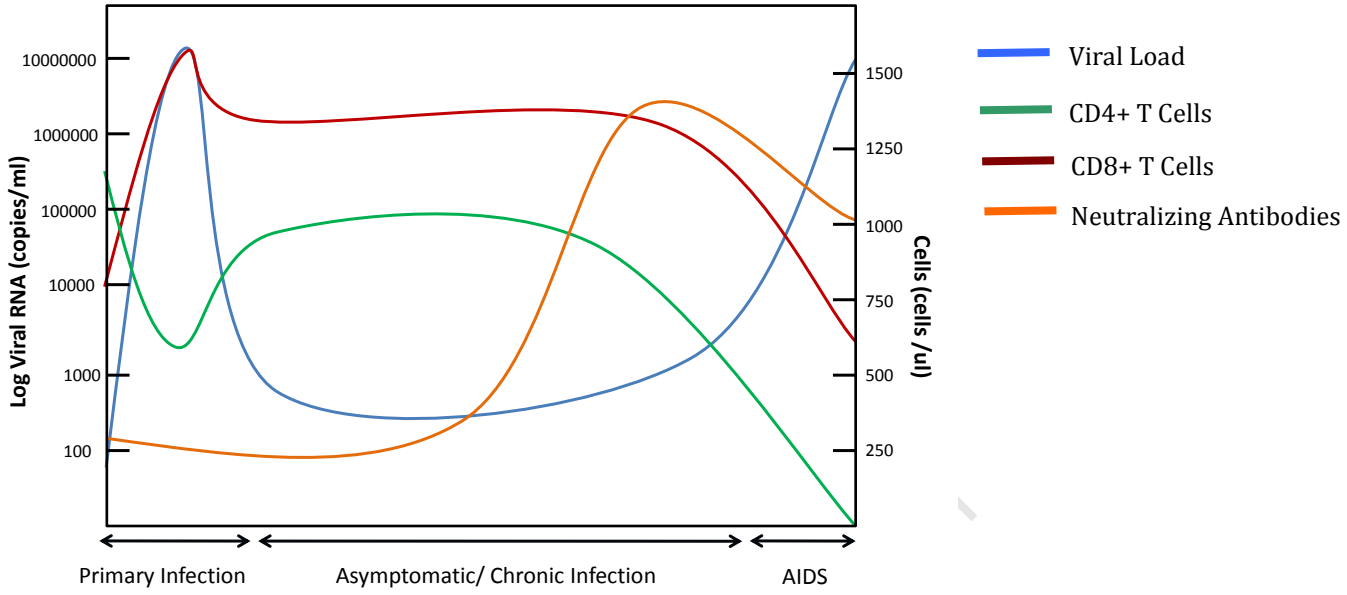


Figure 1.1 The Natural History of HIV infection, depicting the viral load, CD4 cell numbers CD8 T cell and neutralising antibody responses over the course of infection (94)(Adapted from Sewell. 2000).

It has been suggested that the early initiation of Highly Active Anti Retroviral therapy (HAART) may limit the HIV specific immune responses in HIV infected individuals due to lack of exposure to HIV-1 antigens (56). The waning immune responses may then allow superinfection of the individual. However ongoing use of antiretroviral therapy should prevent HIV superinfection if the superinfecting strain is still susceptible to that type of drug (40).

1.4 HIV Diversity and Superinfection

Some studies have shown that the genetic distance between the initial and the superinfecting virus affect the incidence of superinfection as responses to the first virus may not offer protection against the second virus due to mismatched epitopes (5, 86, 100, 124). This suggests that the greater the genetic diversity between the different strains the less likely that immune responses generated by the first virus will recognise and thus protect against re-infection.

If immune responses play a role in the occurrence of superinfection then one would expect the risk of intersubtype superinfection (ie infection with strains from different subtypes that may differ by 30% at the amino acid level in the envelope gene) (63) to be greater than intrasubtype superinfection (ie infection with strains from the same subtype which may only differ by 10% or less at the amino acid level in the envelope gene) as immune responses would be more likely to protect against infection by related viruses from the same subtype that may share antigenic properties (20, 64, 88). However, comparing the incidence of intra and intersubtype superinfection is difficult as these studies ideally need to be done in cohorts where multiple subtypes are circulating. In addition, difficulties in detecting intrasubtype superinfection may influence the reported incidence as intersubtype superinfection is easier to detect (33). In Kenya, an area where multiple subtypes circulate, of the seven superinfections detected equal incidences of intra [4 cases] and intersubtype superinfections [3 cases] were found indicating that intrasubtype superinfections may occur just as readily as intersubtype superinfections (80).

Several studies investigated whether HIV-2 infection prevented HIV-1 infection even though the sequence similarity was as low as 50% (74). One study in Senegal suggested that HIV-2 infection was protective due to the lower incidence of HIV-1 infection in HIV-2 infected individuals compared to uninfected individuals (110). However, this was not supported by a later study in Guinea Bissau which showed that there was in fact a higher risk of HIV-1 infection after HIV-2 infection in all of the women in the study however this correlation did not extend to men (116).

Further studies need to be carried out to determine whether a lack of cross recognition by host immune responses make individuals susceptible to re-infection by different subtypes as well as different types of HIV. These factors are of great importance in the design of an effective vaccine as they will provide insights into how specific the vaccine needs to be in order to be protective.

1.5 Immunology

1.5.1 Cellular Immune Response

Knowledge of what controls primary infection will give us some clues on correlates of protection from superinfection. During primary HIV infection strong Human Leukocyte Antigen (HLA) class I mediated CD8+ Cytotoxic T Lymphocyte (CTL) responses are triggered (83, 121). These CTL immune responses have been shown to play a role in the control of HIV replication (14, 41, 45, 46, 61). A very select number of individuals (viremic controllers with viral loads 50< and >2000 RNA copies/ml and elite controllers with viral loads below 50 RNA copies/ml) are able to exhibit almost complete control over viral replication (6, 78) usually due to persistent *gag* CTL responses. The loss of these *gag* responses has been associated with progression to AIDS (39, 60, 78, 91).

During the course of infection host immune pressures select for changes in CTL epitopes thereby enabling viral escape from CTL responses (7, 18, 31). CTL escape mutations can have two outcomes: 1) loss of control of viral replication and progression to AIDS (45) or 2) the escape mutations impact protein function thereby reducing viral replicative capacity leading to delayed disease progression (8, 13, 18, 18, 31). Escape from three immune-dominant HLA B*57/B*5801 associated epitopes located on the HIV-1 capsid was recently shown to come with a fitness cost as viruses carrying these CTL escape mutations had significantly reduced replicative capacity. The accumulation of additional escape mutations in the epitope reduced viral fitness further and the authors suggest that this likely impacts *in vivo* replication of the virus (15)(68).

Infection with one subtype was may be able to elicit CTL responses protective against a variety of subtypes, although intrasubtype CTL recognition is generally stronger than recognition of viruses of different subtypes (9, 22, 23). However, the mere existence of HIV superinfection raises the question as to the effectiveness of CD8+ T cell responses especially as several studies

have reported an absence of CTL recognition of the superinfecting strain (5, 102, 124). One of the first cases of superinfection was associated with an absence of CD8+ T cell responses to seven of the 25 previously recognised epitopes. This lack of response was due to sequence differences between the two viruses which differed by 12% at the amino acid level. However, eight epitopes were conserved in both viruses suggesting some degree of cross-recognition occurred although this was insufficient for protection (5). In a similar study a patient with an established cytotoxic T lymphocyte (CTL) response effectively controlling his primary infection was unable to protect against superinfection even though the CTL response was broad and stable. On further examination of their CD8 T cell responses it appeared that subsequent to superinfection the initial responses were reduced and there was some targeting of novel epitopes. The loss of initial responses was contributed to differences in sequence of the recognized epitopes in the superinfecting virus which subsequently outgrew the initial strain (124). It is evident from these studies that the specificity of the initial CTL immune response is unable to recognize and respond to the superinfecting strain due to significant differences in their sequences (5, 124).

Superinfection has been proposed to increase the rate at which CTL escape occurs through recombination. A study found that an individual who was effectively controlling his virus, lost control following superinfection (102). Examination of the *gag* and *env* genes showed that this loss of immune control coincided with recombination events in two immunodominant CD8 epitopes in these two genes. The recombination events in these epitopes (B27-KK10 in Gag and Cw1-CL9 in Envelope) rapidly facilitated escape from both dominant wild type-specific and variant-specific responses. Novel CD8+ T cell responses were then detected in these two regions of recombination which replaced the initial responses present before superinfection suggesting that cellular immune pressure may have played a role in driving recombination in this individual as a means of rapidly escaping immune recognition (102).

1.5.2 The Humoral Immune Response

The first anti-HIV antibodies are seen within the first week of infection and comprise of binding antibodies (109) followed by the production of autologous neutralising antibodies between 12 and 20 weeks after infection (47, 66). These early neutralising antibodies are usually specific to the infecting virus; however from around two years post infection these responses may broaden and recognise other variants as well (cross-neutralizing responses)(47, 72, 87, 122).

One of the first studies to look at the neutralising antibody responses in superinfected individuals found that, of the 14 individuals with superinfection, 3 were superinfected between 0 and 6 months post infection. A correlation was observed between the level of cross neutralizing antibodies and superinfection as individuals who became superinfected had lower titres of cross-protective neutralizing antibodies before superinfection than those who did not become superinfected. Furthermore, the antibodies of the non-superinfected individuals were also able to neutralise more HIV variants than those of the superinfected individuals (100). Ramos *et al.* (2002) investigated two intravenous drug users, who even though they had mature humoral immune responses to their initial viruses became superinfected with different viral subtypes due to limited cross-recognition (86). These studies suggest that broadly cross-reactive antibodies may play a role in protection from superinfection.

Conversely, it has recently been suggested that a diminished neutralizing antibody response is not responsible for the occurrence of superinfection. The neutralising antibody responses of six women who had become superinfected between one and five years post infection were compared to 18 women who were at a similar risk of re-infection and who did not become superinfected. There were no significant differences in the breadth or potency of responses prior to superinfection between the superinfected women and the control group. It was also shown that after superinfection most of the women were able to mount an autologous

neutralising antibody response against the second virus (11). Further studies are needed to elucidate the role of neutralizing antibodies in the preventing superinfection.

1.6 Animal Models

Animal models have been used to evaluate the correlates of protection from superinfection. Although these studies have yielded conflicting results due to differences in study design, methods of inoculation, animal models, as well as the viral strains, they nevertheless do provide some insights into understanding correlates of protection.

Early studies demonstrated that Chimpanzees could be re-infected by an HIV stain from a different subtype six months after infection with the first strain (36). However, using different viruses and much higher infectious doses a later study showed that chimpanzees which had been infected with HIV-1 for between 22 and 101 months were resistant to repeated and escalating inoculations of a heterologous intraclade strain of the virus (96). In the Simian Immunodeficiency Virus (SIV) Rhesus macaque model, a recent study showed infection with one strain of SIV did not protect against re-infection with a heterologous SIV challenge as 12 out of the 14 monkeys became superinfected. The timing of superinfection in the monkeys ranged from three to 20 months post infection. However this study also showed that the monkeys were able to control the replication of the second virus more efficiently than they controlled the first virus. When they compared the immune parameters of the twelve animals that became superinfected to the two animals that did not, they found that neither the CD4+ T-cell numbers, CD4+ memory T cell subsets, cytokine production by virus specific CD8+ T cells, nor neutralizing antibodies were associated with control of the second virus (126). However they found that the two SIV-infected macaques had protective HLA alleles and despite the challenge viruses containing CTL escape mutations in epitopes associated with these HLAs, they remained protected from superinfection. This suggests a role of CD8+ T cells in the prevention of infection (123).

These studies suggest that superinfection is possible in most animal models however further studies are required to determine whether any protection is provided by the first infection against re-infection and what immune parameters are associated with this protection.

1.7 Detecting HIV superinfection

Superinfection can be difficult to detect which may have some bearing on differences in reported incidences. Many methods have been explored including the heteroduplex mobility assay, the multi-region hybridization assay, restriction fragment length polymorphism and the use of ambiguity codes present in population based sequences (28, 30, 53, 84, 118). Most of the methods are time consuming and accurate detection of superinfection requires longitudinal follow up (28, 30). As several cases of superinfection have been accompanied by increases in viral load, as well as changes in drug resistance profiles, these parameters can also be used to identify individuals with possible superinfection (17, 33, 44, 71, 99, 124). However, there have been cases of superinfection where viral load does not increase, and therefore superinfection may be missed if this is used as the sole means of screening for its occurrence (80, 100). Recombinant viral populations are also often used to infer the occurrence of HIV-1 superinfection as they can only be generated as a result of a dual infection (105) however recombination between the initial and superinfecting strains can also prevent detection of superinfection if only a single region of the genome is studied (80)(81).

In an individual infected with multiple viral variants the ratio of the infecting viruses fluctuates (50). The proportion of the multiple viral populations infecting an individual can conceal the occurrence of superinfection because if the superinfecting virus is a minor population then it may not be detected (71). Similarly, in some individuals the initial virus is lost through competition with the superinfecting virus (28, 80, 108, 120) which may result in the presence of only one strain at the sampled time point, preventing detection of multiple viral populations

(33, 118). It would thus appear that even if superinfection occurred frequently, the difficulty to detect the viral populations may lead to an underestimation of its incidence.

1.8 The Nature of the Superinfecting Virus

The susceptibility of host immune cells to reinfection, particularly CD4+ T cells, is a critical rate limiting feature of superinfection. Susceptibility is a measure of the viral and cellular factors which make an infected cell resistant to re-infection. If less target cells are susceptible to re-infection then competition between viral strains will result in the exclusion of one of the viruses. Conversely if there are many cells susceptible to re-infection then co-existence of the two viruses may occur particularly if the viruses are of equal fitness (37)

The fitness of the infecting viruses also impacts the dominance of viral populations due to competition for target cells. Low replicative fitness of the first virus has been associated with superinfection and it was suggested that individuals infected with poorly replicating variants have higher CD4 counts, making target cells available to the superinfecting strain (16, 17, 62, 99, 100, 115). However, superinfection has also been shown to occur when both initial and superinfecting viruses are of equal fitness (28, 101, 124). This suggests that although in some cases initial infection with a weaker virus may make the individual more at risk of superinfection fitness is only one possible factor which may influence their susceptibility.

1.9 Pathogenic Consequences of HIV Superinfection

Superinfection can impact an individual in a number of ways, either by accelerating disease progression or through the generation of multi-drug resistant strains (50, 71, 79).

There are several studies which have shown that superinfection has been associated with increased viral load and faster disease progression (16, 42, 50, 62). One study found that superinfection was associated with progression to AIDS within 3.4 years as opposed to non-

superinfected individuals who usually only progress to AIDS within 8 to 10 years (42). Similarly, superinfection of an HLA-B57+ long term non-progressor resulted in an increase in viral load and a corresponding decline in CD4+ cell numbers (16). However superinfection is not always associated with accelerated disease progression. In two long term non-progressors it appeared that their immune responses were able to control the superinfecting strain such that there was no acceleration of disease progression (25). The potential for superinfection to rapidly increase the diversity within an infected individual has consequences for the individual with respect to their disease profile as well as their immune responses.

1.10 Study Rationale and Specific Objectives

A vaccine remains our best hope in controlling the HIV-1 epidemic. Studies on superinfection may provide a better insight into what immune correlates of protection are required for an effective vaccine. This study investigates superinfection in a cohort of high risk women from KwaZulu-Natal to determine the incidence and timing of superinfection, and to ascertain whether genetic differences between the two strains were responsible for the occurrence of superinfection. In particular differences in Cytotoxic T cell epitopes between the initial and superinfecting viruses will be examined to determine whether the inability to recognise epitopes on the superinfecting strain is a determinant of superinfection. Furthermore, this study will also determine the role of recombination between the initial and superinfecting viruses in selecting specific CTL epitopes that confers escape from the immune system.

The research performed is part of the Centre for AIDS Program of Research in South Africa (CAPRISA) 002 study.

The main objectives of this research project were:

- To determine the frequency of HIV-1 superinfection over the first year of infection in the CAPRISA 002 Acute Infection cohort.

- To characterize the evolution of the initial and superinfecting strains within the full length genome of two superinfected individuals.
- To establish whether genetic differences between the two strains may have prevented recognition by the host immune system.
- To determine whether CTL epitopes drive recombination between the initial and superinfecting strains.

2.1 Introduction

Superinfection has implications for the development of effective vaccine strategies as it provides evidence that infection with one strain of HIV does not protect against a second infection (28, 80). It also has public health considerations as superinfection has been linked to faster disease progression (16, 42). If the incidence of superinfection approaches that of primary HIV infection, as was previously found in the study in Kenya, this implies that little or no protection is provided by the first infection (80).

Several methods have been used to determine the frequency of superinfection however differences in sensitivity of detection, study design, cohorts, circulating subtypes, and stage of infection have contributed to inconsistent findings (28, 30, 118). Some of the methods used to detect HIV-1 superinfection include restriction fragment length polymorphism (RFLP) (55), heteroduplex mobility assay (HMA) (67), multiregion hybridization assay (MHA) (53) and population based sequencing to determine the number of degenerate base codes (30). The RFLP and MHA methods can detect inter-subtype dual infections, whereas the HMA and population based sequencing method are able to discriminate between both inter and intra-subtype dual infections.

Subtype C is responsible for over 95% of infections in South Africa and we were therefore interested in a method that could identify intra-subtype superinfection. Population based sequencing uses the number of degenerate base codes detected during sequencing as an indicator of multiple viral populations where the presence of 45 or more degenerate base codes within *reverse transcriptase* (~ 1320 bp) were shown to correctly identify 73% of dual infected individuals (30). The heteroduplex mobility assay is based on the principle that heteroduplexes are formed when partially complementary DNA strands are denatured then allowed to re-anneal. The extent of variation in sequence can then be derived from the reduced electrophoretic mobility of these heteroduplexes. The slower mobility is due to the structural

distortions resulting from mismatched and unpaired nucleotides (32). The sensitivity of this assay allows the detection of viruses that differ by 0.86% in the C2-C3 region of *envelope* (50). The HMA was found to be more sensitive than the genotyping method suggesting it is the ideal method for detecting intra-subtype superinfection (84), thus this method was selected for our study.

One of the criteria used to target individuals for superinfection screening is an unexpected increase in viral load which often accompanies superinfection with a second strain (44, 71, 100). However, as an increase in viral load is not always present (57), using this as a screening tool may result in superinfection being missed. In addition, it is also important for longitudinal follow up as fluctuating viral populations and recombination may also result in superinfection not being detected (28, 30). In this study we therefore screened all participants for superinfection, over four time points within the first year of infection.

The CAPRISA 002 cohort recruited participants during acute infection and has followed them over time to assess virological and immunological factors impacting disease progression. The cohort is based in Kwa-Zulu Natal and consists primarily of high-risk women with a mean age of 34.3 years (range 18 to 58). The prevalence of HIV within the screening population was exceptionally high at 59.6% and the sero-incidence within the cohort was 7.2% per 100 person years (117). This cohort provides a means of determining the frequency of HIV superinfection as, despite extensive counselling, many of the HIV positive women continued to engage in high risk behaviour which provides the opportunity for re-infection (117).

In this chapter we aimed to determine the frequency of superinfection in 20 participants in the CAPRISA 002 cohort within 12 months post-infection. We used HMA of the C2-C3 region of *envelope* together with sequencing and phylogenetic analysis to detect cases of HIV superinfection. The effect of superinfection on viral load and subsequent disease progression was investigated.

2.2 Materials and Methods

2.2.1 Participants

The Centre for the AIDS Programme of Research in South Africa (CAPRISA) 002 Acute Infection Study is a natural history study of subtype C HIV infection. It monitored HIV negative high-risk individuals monthly (117) who are screened for HIV infection using PCR (Determine; Abbott Laboratories, Tokyo, Japan) and two rapid antibody tests. It also recruited from individuals in rural areas of KwaZulu Natal who were participating in prevention research. On obtaining a negative antibody test the individual was enrolled in the HIV-negative cohort however if it was positive a second rapid antibody test (Capillus; Trinity Biotech, Jamestown, NY, USA) was administered. The time of infection was taken to be either the midpoint between the first antibody positive test and the last HIV-1 negative test or as the first 14 days when individuals are PCR positive and antibody negative. On enrolment individuals had their CD4+ T cell counts and viral load monitored on a weekly basis for the first month, then monthly until a year post-infection and quarterly thereafter. Viral loads were measured using a COBAS AMPLICOR™ HIV-1 Monitor Test v1.5 (Roche Diagnostics, Basel Switzerland) and CD4+ T cell counts were measured using a FACSCalibur flow cytometer. Plasma was collected from each individual in EDTA and stored at -70°C. The clinical data was provided by K. Mlisana (University of KwaZulu-Natal, South Africa). Ethics approval was given by the research ethics committees of the University of Kwa-Zulu Natal (E013/04), the University of Cape Town (025/2004) and the University of the Witwatersrand (MM040202).

Twenty women were screened for superinfection at enrolment, 3, 6 and 12 months post-infection.

2.2.2 RNA Extraction

Viral RNA was extracted from 200 ul aliquots of plasma using the MagnaPure RNA Isolation Kit (Roche) and the MagnaPure Compact Nucleic Extractor (Roche). RNA was stored at -80°C.

2.2.3 Reverse Transcription

To generate cDNA the Promega Reverse Transcription Kit (Promega, Madison USA) was used. RNA (4 ul) was added to the Env N (0.2 uM) primer (Table 1) and incubated at 65°C for 5 minutes before adding Buffer (1X), RNasin (20 units), $MgCl_2$ (3.75 mM), deoxyribonucleoside triphosphates (dNTPS) (0.5 mM of each dNTP), reverse transcriptase (1 ul) and DEPC- treated water to a total volume of 15 ul. The reaction was carried out by heating to 25°C for 5 minutes, followed by 42°C for 1 hour and then 70°C for 15 minutes using the Applied Biosystems 2720 Thermal Cycler (Applied Biosystems, Carlsbad, CA USA).

2.2.4 PCR Amplification of C2-C3 region of the Envelope gene

A nested population PCR amplification protocol was followed for amplification of the C2-C3 *envelope* fragment. The first round reaction consisted of dNTPs (0.2 mM each)(Invitrogen, Carlsbad USA), 1X Buffer B (Kapabiosystems, Cape Town South Africa), the ED33 primer (0.2 uM)(Integrated DNA Technologies, Coralville USA), the ED31 primer (0.2 uM)(Integrated DNA Technologies, Coralville USA), 1 unit of *Kapataq* polymerase (Kapabiosystems, Cape Town South Africa) and 5 ul of the template cDNA. The cycling profile consisted of an initial denaturation step at 94°C for 2 minutes, followed by 35 cycles of denaturation at 94°C for 15 seconds, annealing at 50°C for 45 seconds and elongation at 72°C for 1 minute. The second round reaction or the inner reaction consisted of dNTPs (0.2 mM each), 1X Buffer B, the BF primer (0.2 uM)(Integrated DNA Technologies, Coralville USA), the BR primer (0.2 uM)(Integrated DNA Technologies, Coralville USA), 1 unit of the *Kapataq* polymerase, 5 ul of the product of the first round reaction and then made up to a final volume of 50 ul with distilled water. The cycling

profile used was the similar to that used for the first reaction except that 45 cycles of amplification were used and the annealing temperature was increased to 55°C to increase the specificity of the amplification. All amplifications were carried out on the Applied Biosystems 2720 Thermal Cycler (Applied Biosystems, Carlsbad, CA USA).

Table 2.1: PCR Primers used for Amplification of C2-C3

Name	Sequence	HXB2 Position	Forward/Reverse
EnvN	5'- CTGCCAATCAGGGAAGTAGCCTTGTGT	60-86	Reverse
ED31	5'-CCTCAGCCATTACAGGCCTGTCCAAAG	7359-7380	Forward
ED33	5'-TTACAGTAGAAAAATTCCCCTC	6816-6844	Reverse
envBF	5'-TAACACAAGCCTGTCCAAAGG	6826-6847	Forward
envBR	5'-AATTCTAGGTCCCCTCCTGA	7317-7337	Reverse

2.2.5 Agarose Gel Electrophoresis

To check the amplification of the C2-C3 region (~500 bp),PCR products were separated on a 1% agarose gel made up in Tris Borate EDTA (TBE) (Appendix C5) with a molecular weight marker (100bp DNA ladder Plus; Promega, Madison USA). Loading dye (Appendix C3) containing Gel Red (Biotum, Hayward USA) was added to each sample. The gel was then visualized using ultra violet light (UVIpro Silver; UVItec, Cambridge United Kingdom).

2.2.6 Heteroduplex Mobility Assay (HMA)

HMA was used to screen individuals for HIV superinfection (32).

In order to identify the presence of more than one variant at each time point the C2-C3 PCR product was heat denatured at 94°C for 2 minutes and allowed to anneal at ambient temperature. To determine whether there were differences in quasispecies between time points the PCR product from two different time points were combined before denaturing and re-annealing. Denaturation and annealing was carried out in HMA annealing buffer (Appendix C).

To check for heteroduplex formation samples were run on a polyacrylamide gel (5%)(Appendix C) at 250 volts for 150 minutes using the Scie-Plas V20 vertical gel apparatus with glass plates 20cm x 19.5cm and spacers 1.5mm thick were used (Scie-Plas Ltd, Southam Warwickshire England). Once the gel had run to completion it was stained with Ethidium Bromide for 5 minutes and then viewed using the UVIpro Silver UV box (UVItec, Cambridge United Kingdom). A positive control, included with each experiment comprised of enrolment samples amplified from two different CAPRISA individuals which were combined before denaturing and annealing.

2.2.7 Sequencing of C2-C3 PCR Product

To confirm the presence of different variants the C2-C3 PCR product was sequenced using the second round primers *env*BF and *env*BR. To clean up the PCR product and remove any components which could produce unwanted background in the sequencing results, the PCR product was treated with Exosap (1 unit)(USB, Cleveland USA) and placed at 37°C for 15 minutes followed by inactivation at 80°C for 15 minutes. For each individual sample two sequencing reactions were performed using the ABI PRISM dye terminator cycle-sequencing kit (Applied Biosystems, Carlsbad, CA USA) The sequencing reaction of buffer, the forward or the reverse primer (0.1 pM) and distilled water to a total volume of 20 ul were placed in a 0.25 ml PCR tube. The cycling conditions included an initial denaturation at 96°C for 5 minutes, followed by 25 cycles of denaturation at 96°C for 30 seconds, annealing at 50°C for 15 seconds and elongation at 60°C for 4 minutes. The sequencing reactions were sent to the Stellenbosch University DNA Sequencing Unit to be run on their genetic analyzer.

2.2.8 Sequence analysis

The sequences were assembled using ChromasPro Version 1.42 (Technelysium, Australia), the resulting contigs were then aligned using Bioedit Sequence Alignment Editor Version 7.0.9.0 (51). The alignment files were then imported into Mega version 4.0 (104) to generate the phylogenetic trees and determine pairwise DNA distances. The neighbour-joining method was

used to draw the phylogenetic trees, this method is based on the principal that the taxon are paired such that the smallest total branch length is generated in the phylogenetic tree (93).

Sequence alignment files for each individual, containing the sequences for each of the time points amplified, were uploaded into the Highlighter tool on the Los Alamos HIV Sequence Database (www.hiv.lanl.gov). The function of this tool is to highlight any mismatches, matches, transition and transversion mutations as well as silent and non-silent mutations in a set of aligned sequences which are in frame.

2.2.9 Graphs

The pairwise DNA distances of the individuals screened were plotted on graphs using GraphPad Prism 4.0 (GraphPad Software Inc. USA).

2.3 Results

To identify superinfection, the C2-C3 region of *envelope* was amplified in 20 individuals using plasma samples from enrolment (acute HIV infection), three, six and twelve months post infection. Of these 20 individuals 17 were screened for superinfection using the HMA, sequencing and phylogenetic analysis. The remaining three individuals were screened using only sequencing and phylogenetic analysis. All 20 individuals were infected with subtype C HIV throughout envelope as shown by M. Abrahams (2).

The median log viral load and CD4+ count for these individuals at twelve months post infection was 4.77 copies/ml (range= 2.66 - 6.09 copies/ml) and 351 cells/ul (range= 202 – 1016 cells/ul) respectively. Out of the 20 individuals screened 8 had a spike in viral load of greater than 0.5 log after acute infection and within the first year of infection (CAP65, CAP210, CAP228, CAP237, CAP239, CAP256, CAP258 and CAP281).

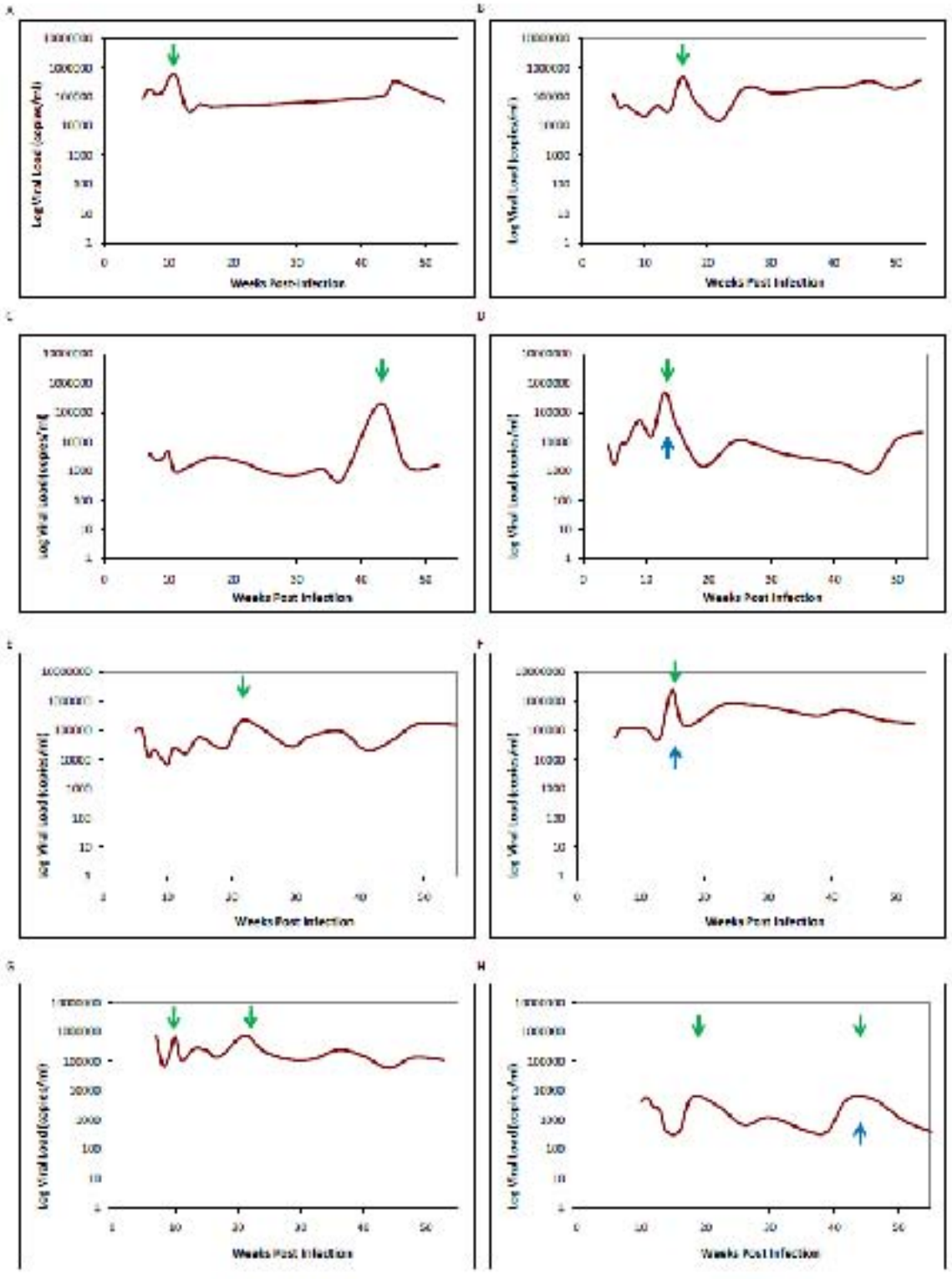


Figure 2.1 Viral Load Profiles of Individuals exhibiting a spike in viral load greater than 0.5 log, A) CAP65, B)CAP210, C) CAP228, D) CAP237, E) CAP239, F) CAP256, G) CAP258 and H) CAP281. The green arrows indicate the peaks in viral load and the blue arrows in the superinfected individuals indicate the timing of superinfection.

2.3.1 Viral Diversity

HMA was used to screen for diversity within a time point. In addition, each time point was also mixed with every other time point from that individual to determine if there was replacement of one strain with a second strain. The method had previously been used in this laboratory to screen for intrasubtype dual infection and was found to differentiate C2-C3 amplicons with a distance greater than 0.86% (50). The C2-C3 env region was selected as this region is variable enough to distinguish between different strains from the same subtype but contains the least number of insertions and deletions of all of the variable regions within envelope. It is important to have few insertions and deletions and they affect the migration of the DNA and thus may produce a false positive.

Three patterns of evolution were detected using HMA, together with sequencing: (i) women infected with a homogeneous viral population with no or limited diversification up to 12 months post infection, ii) women infected with a heterogeneous virus population that continued to evolve over time and iii) women infected with a homogenous viral population with evidence of other variants post infection.

Women infected with a homogenous viral population with no or limited diversification up to 12 months post infection

Thirteen women were identified to have a homogenous viral population with no or limited diversification over the first year of infection (CAP30, CAP61, CAP63, CAP65, CAP129, CAP174, CAP206, CAP210, CAP239, CAP244, CAP255, CAP257, CAP258). This is illustrated by the HMA of CAP30 which shows a discrete band at enrolment and the following time point indicative of a homogeneous viral population and when time points were mixed no heteroduplex bands were detected (Figure 2.2A). These results are reflected in the phylogenetic tree and highlighter plot (Figure 2.2B), which show that there is a gradual accumulation of mutations over time. Sequence divergence, apparent from the highlighter plot, between enrolment and twelve

months post infection was 0.014%. These results explain the absence of any heteroduplex bands as the assay is only able to pick up diversity greater than 0.86% (50).

Women infected with heterogeneous viral population which continued to evolve over the first year of infection

Four individuals were found to be infected with multiple variants which continued to diversify over the first year of infection (CAP69, CAP177, CAP222, CAP228). HMA analysis of CAP69 exhibited faint heteroduplexes at enrolment demonstrating that the individual was infected with highly diverse variants. In addition at twelve months post-infection (lane 6)(Figure 2.2C) the multiple heteroduplex bands were indistinct due to smearing. Smearing has been associated with a diverse quasispecies (50) suggesting that the variants transmitted to CAP69 continued to evolve over the course of infection. The highlighter plot indicated a number of mismatches between sequences from enrolment and three months post-infection (Figure 2.2D). These results suggest that one variant became dominant at three months and was then replaced by another virus at six months which then persisted. In support of these results individual CAP69 had been previously shown to be infected with five variants at transmission based on *envelope* single genome amplification with a maximum DNA distance in *envelope* of 3.5%(2) and 7.9 % within the C2-C3 region(2).

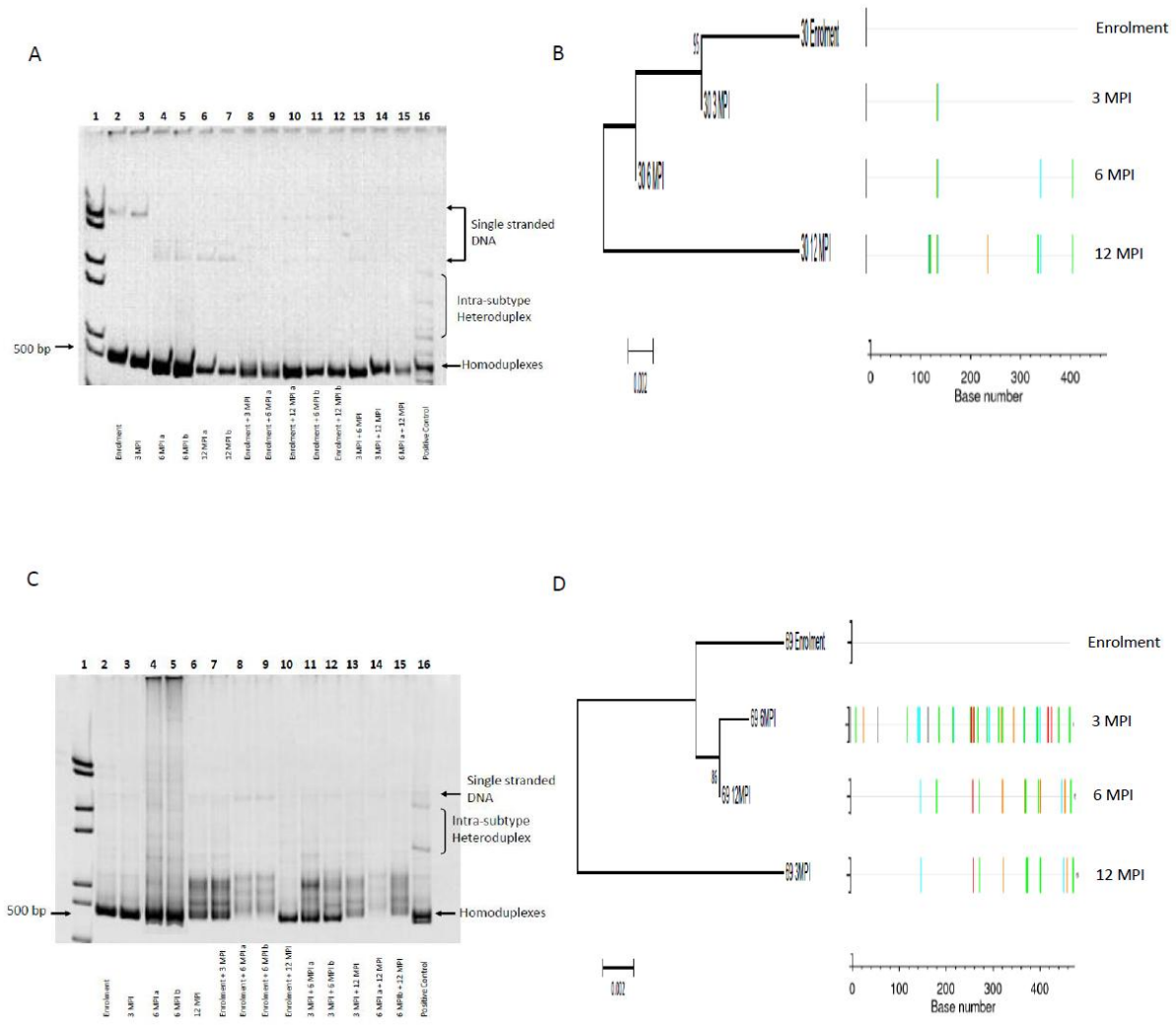


Figure 2.2 HMA screening and phylogenetic analysis of CAP30 and CAP69 at enrolment, 3, 6 and 12 months post-infection with a and b indicating duplicate PCR reactions. A) CAP30 HMA, B) CAP30 Neighbor-joining tree and a highlighter plot of the C2-C3 region of *envelope*. C) CAP69 HMA. D) CAP69 Neighbor-joining tree and a highlighter plot of the C2-C3 region of *envelope*. The colors in the highlighter plot represent changes to: A = Green, T = Red, G = Orange, C = Light Blue, IUPAC = Dark Blue and Gaps = Grey.

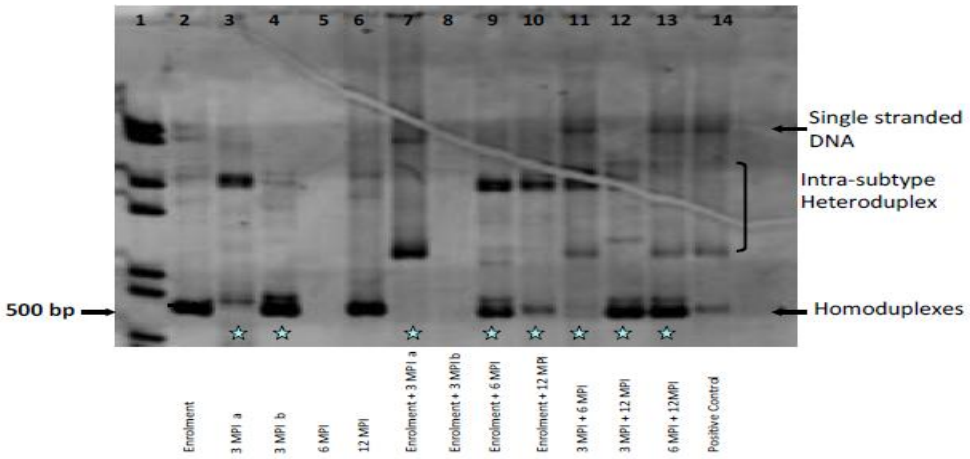
Women infected with homogeneous viral population with evidence of other variants post infection

Two women (CAP237 and CAP256) were identified as potentially superinfected using HMA, sequencing and phylogenetic analysis. These women were initially infected with a homogenous viral population and at a later time were re-infected with a different homogenous viral population.

CAP237

A single homogeneous population was detected at enrolment. The faint bands visible at enrolment (Fig 2.3A lane 2) co-migrated with the molecular weight marker and were thought to be due to overflow from the adjacent lane (lane 1). A slow migrating band was visualized at three months post infection, suggesting that multiple strains were present at this time point. When the enrolment and 3 months post-infection sample were mixed a new band was visualized with reduced mobility, suggesting a second virus infected the individual after enrolment. A similar banding pattern was observed when the enrolment sample was mixed with amplicons from six and twelve months post-infection (lanes 9 and 10) (Figure 2.3A), suggesting that the second virus that appeared at 3 months persisted at later time points. The twelve month sample (lane 6) contains a prominent homoduplex, suggesting that there was low diversity at 12 months post-infection. Population PCR followed by sequencing shows a high number of mismatches between the population sequences at enrolment and three months post infection (Figure 2.3B). Sequences at six and twelve months post-infection grouped with the three months post infection sequence supporting the HMA data that the variant that appeared at three months persisted until one year post-infection (Figure 2.3B).

A



B

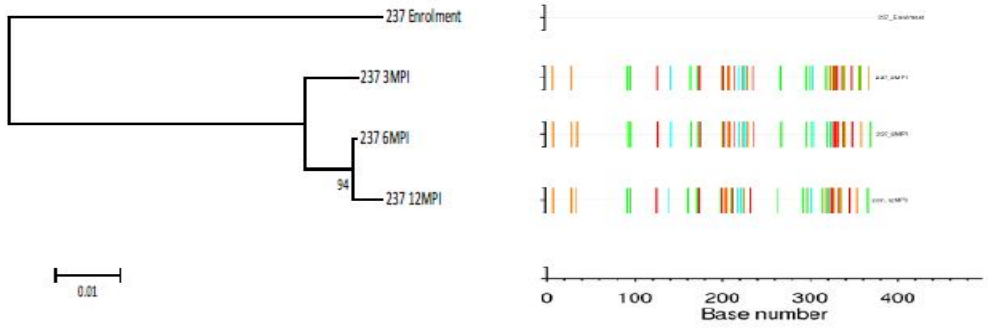


Figure 2.3 HMA screening and phylogenetic analysis of CAP237 A) HMA at enrolment, 3, 6 and 12 months post-infection with a and b indicating duplicate PCR reactions. B) Neighbor-joining tree and a highlighter plot of the C2-C3 region of *envelope*. The blue stars in the heteroduplex mobility assay indicate those lanes where heteroduplex bands were observed. The colors in the highlighter plot represent changes to: A = Green, T = Red, G = Orange, C = Light Blue, IUPAC = Dark Blue and Gaps = Grey. The six month sample and the mixed enrolment and one of the 3 months samples (lane 5 and 8) did not yield bands most likely due to pipetting error.

CAP256

For CAP256, a homogeneous population was found at enrolment with faint slow migrating bands visualized at four months post infection (Figure 2.4A). No heteroduplexes were present at six months post infection. Slow migrating heteroduplex bands indicative of intrasubtype dual infection were present when the enrolment time point was mixed with all time points after

three months post infection (lanes 10 to 13). There was also a heteroduplex band present when the PCR product from three months post infection was mixed with those from both four (lanes 14 and 15) and six months post infection (lanes 16 and 17), suggesting that the viral populations differed between these time points. The presence of multiple heteroduplex bands suggests that the individual was infected with multiple viral populations. Overall, HMA data suggests that CAP256 became infected with a second virus between enrolment and 4 months post-infection. The phylogenetic analysis of CAP256 supported the results of the heteroduplex mobility assay. The neighbor-joining tree suggests that the virus present at three months post-infection is identical to the transmitted variant and that these two viruses were different from those present at four, six and twelve months post-infection (Figure 2.4B).

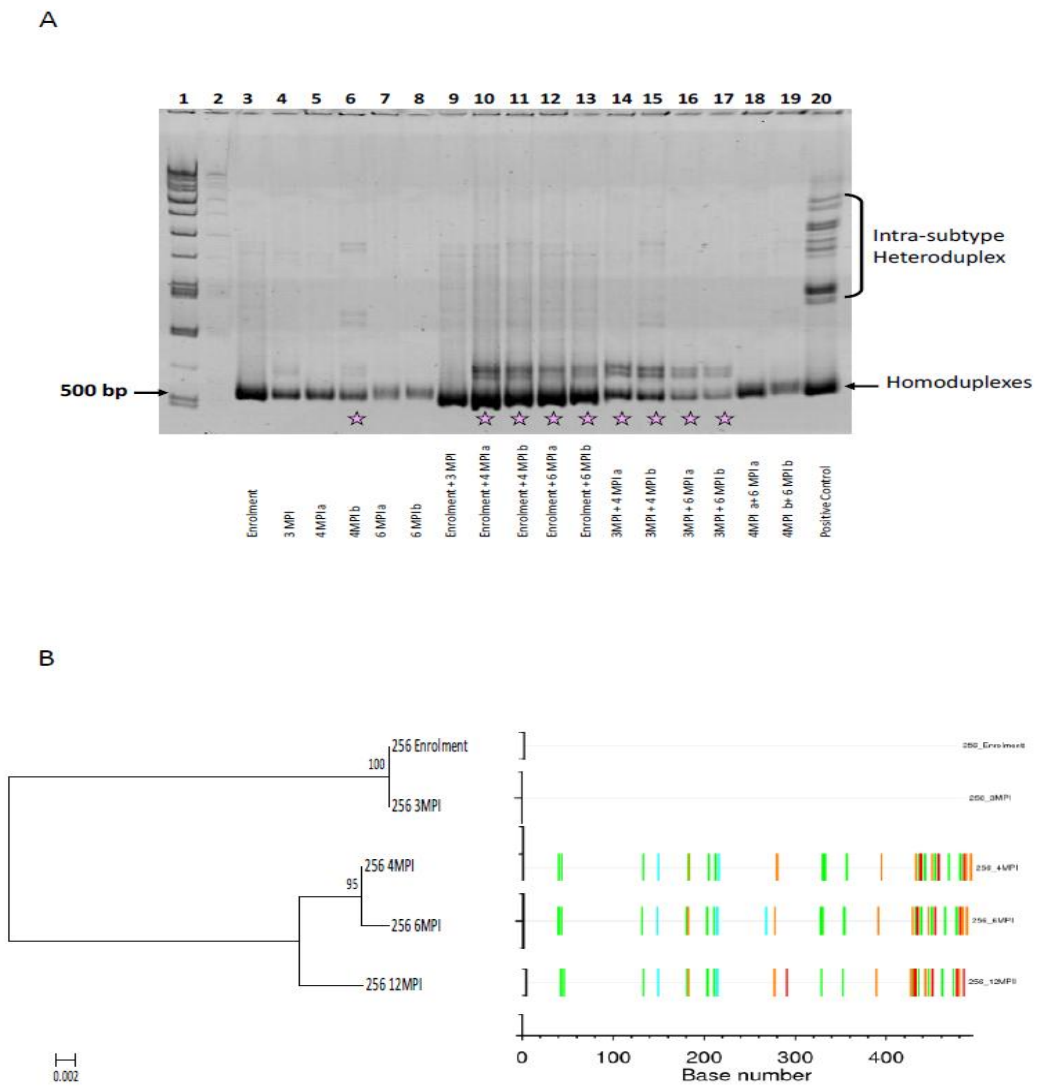


Figure 2.4 HMA screening and phylogenetic analysis of CAP256 A) HMA at enrolment, 3, 6 and 12 months post-infection with a and b indicating duplicate PCR reactions. B) Neighbor-joining tree and a highlighter plot of the C2-C3 region of *envelope*. The pink stars in the heteroduplex mobility assay indicate those lanes where heteroduplex bands were observed. The colors in the highlighter plot represent changes to: A = Green, T = Red, G = Orange, C = Light Blue, IUPAC = Dark Blue and Gaps = Grey.

2.3.2 Phylogenetic analysis of the individuals infected with homogeneous viral populations with evidence of other variants post infection

To determine if these individuals were superinfected with epidemiologically unlinked viruses, we compared the C2-C3 region of *envelope* from each individual over time (Figure 2.5). In addition to CAP256 and CAP237, CAP281 was included in this study after R. Ntale (University of Cape Town) identified this individual as a potential case of superinfection based on sequence

analysis of the *gag* region. As this individual had already been identified as a putative superinfection, she was not screened using HMA.

The individuals identified to be infected with homogeneous viral populations (n=13) had sequences over time which either exhibited low diversity with no structure on the phylogenetic tree or showed some structure in the tree over the first year of infection. Of the thirteen individuals identified as having been infected with homogenous viral populations in the HMA, two individuals (CAP61 and CAP244) had no evidence of diversification in C2-C3, while eleven individuals (CAP30, CAP63, CAP65, CAP129, CAP174, CAP206, CAP210, CAP239, CAP255, CAP257 and CAP258) exhibited some diversification. The sequences from those individuals that had heteroduplex banding patterns at enrolment on HMA clustered together with a high DNA distance (CAP69, CAP177, CAP222 and CAP228). The sequences of the three individuals with putative superinfection separated from one another by other epidemiologically unlinked viruses (superinfected individuals)(CAP237, CAP256 and CAP281)(Figure 2.5). The composite phylogenetic tree was therefore able to confirm the results of the screening.

Timing of Superinfection

The sequences from CAP256 separated into two clusters; enrolment and three months on one branch and four, six and twelve months on another. These sequences were separated on the tree by epidemiologically unlinked sequences confirming superinfection between three and four months post infection (Figure 2.5). The sequences of CAP237 generated at and after three months post infection clustered on a different branch to that of the enrolment sequence indicating superinfection between enrolment and three months post infection (Figure 2.5). For CAP281 one branch of the phylogenetic tree contained sequences from enrolment as well as time points leading up to and including ten months post infection, while the eleven months post infection sequence was located on a different branch of the tree, separated by an epidemiologically unlinked sequence confirming superinfection between ten and eleven months

post infection. These results were supported by Gag sequences generated by R. Ntale. For CAP237 and CAP256 although both the HMA and phylogenetic analysis support the occurrence of superinfection in these individuals strain specific was used to confirm the timing of superinfection (D. Sheward and R. Ntale, University of Cape Town. Pers. Comm).

Of the initial nineteen samples screened by HMA, two were confirmed as superinfected (10.5%) (Appendix A2). A third individual (CAP281), not included in the initial screen, was identified as superinfected through population sequencing of *gag*. It was interesting to note that the three individuals identified as superinfected were included in the eight individuals identified as having a spike in viral load (Figure 2.1).

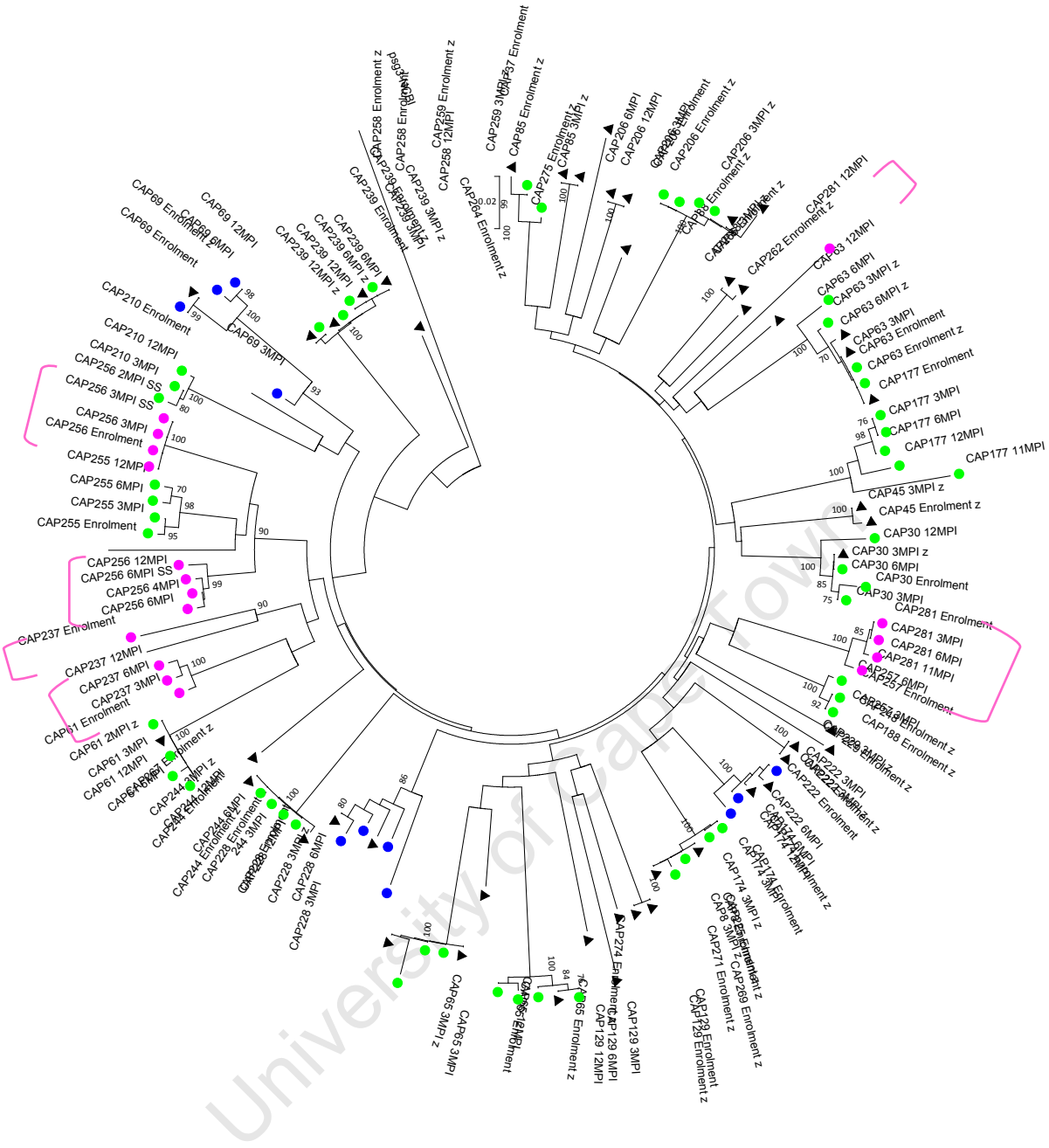


Figure 2.5 Neighbor-joining phylogenetic tree of longitudinal (enrolment to 12 months post infection) envelope C2C3 sequences. The tree includes sequences from CAPRISA individuals, both the individuals screened in this study and sequences generated by Z. Woodman and M. Abrahams, as well as reference sequences obtained from the Los Alamos HIV-1 Sequence Database (www.hiv.lanl.gov). The sequences from the individuals screened in this study are indicated by the green circles for homogenous transmission, the blue circles for heterogeneous transmission and the pink circles for superinfected individuals. The black triangles indicate sequences generated by Z. Woodman (University of Cape Town) and M. Abrahams (University of Cape Town). The branch nodes without any markers indicate those sequences obtained from the Los Alamos HIV-1 Sequence Database. The phylogenetic tree was generated based on 1000 bootstrap replicates.

2.3.3 DNA Distance as a indicator of Superinfection

We next wanted to determine whether the pairwise DNA distance over the first year of infection could be used as a means of detecting superinfection, whereby superinfected individuals would have a substantially higher DNA distance than non-superinfected individuals. The pairwise DNA distance between the *envelope* C2-C3 viral sequences for each individual at enrolment and at twelve months post infection was measured (Figure 2.6) to determine the difference between the viral populations. The three individuals who became superinfected had the greatest DNA distance with the highest seen in CAP281 (18.1%), followed by CAP237 (11.2%) and the lowest observed in CAP256 (7.1%) over the first year of infection. The observed difference in the timing of the superinfections may have influenced the observed genetic diversity. Both CAP256 and CAP237 were superinfected closer to the time of enrolment allowing greater time for recombination to occur which would reduce the distance between strains. CAP281 on the other hand was superinfected just prior to one year post infection, with limited time for recombination and this distance likely reflected by the distinct viral populations which co-exist at 12 months infection. From the C2-C3 region alone it is difficult to confidently distinguish whether recombination had occurred in both CAP256 and CAP237. However previous work by F. Treunicht (University of Cape Town) had determined the virus present at six months post infection in CAP256 to be a recombinant. As yet no further analysis has been carried out on CAP237.

However individuals CAP30 and CAP177 also exhibited higher DNA distances than the other individuals screened with distances above the 75 percentile. CAP177 is a potential dual infection as a different variant has been detected in the cervicovaginal cavity at enrolment (P. Moore. Pers. Comm). This may account for the higher DNA distance observed in this individual, however as dual infection has not been confirmed in this individual they are characterized here as a heterogenous infection. CAP30 was one of the individuals who had been infected with a

homogenous viral population which diversified over the first year of infection; the diversification over the first year of infection may have resulted in a higher DNA distance due to the extent of the genetic differences between the enrolment and twelve months viruses.

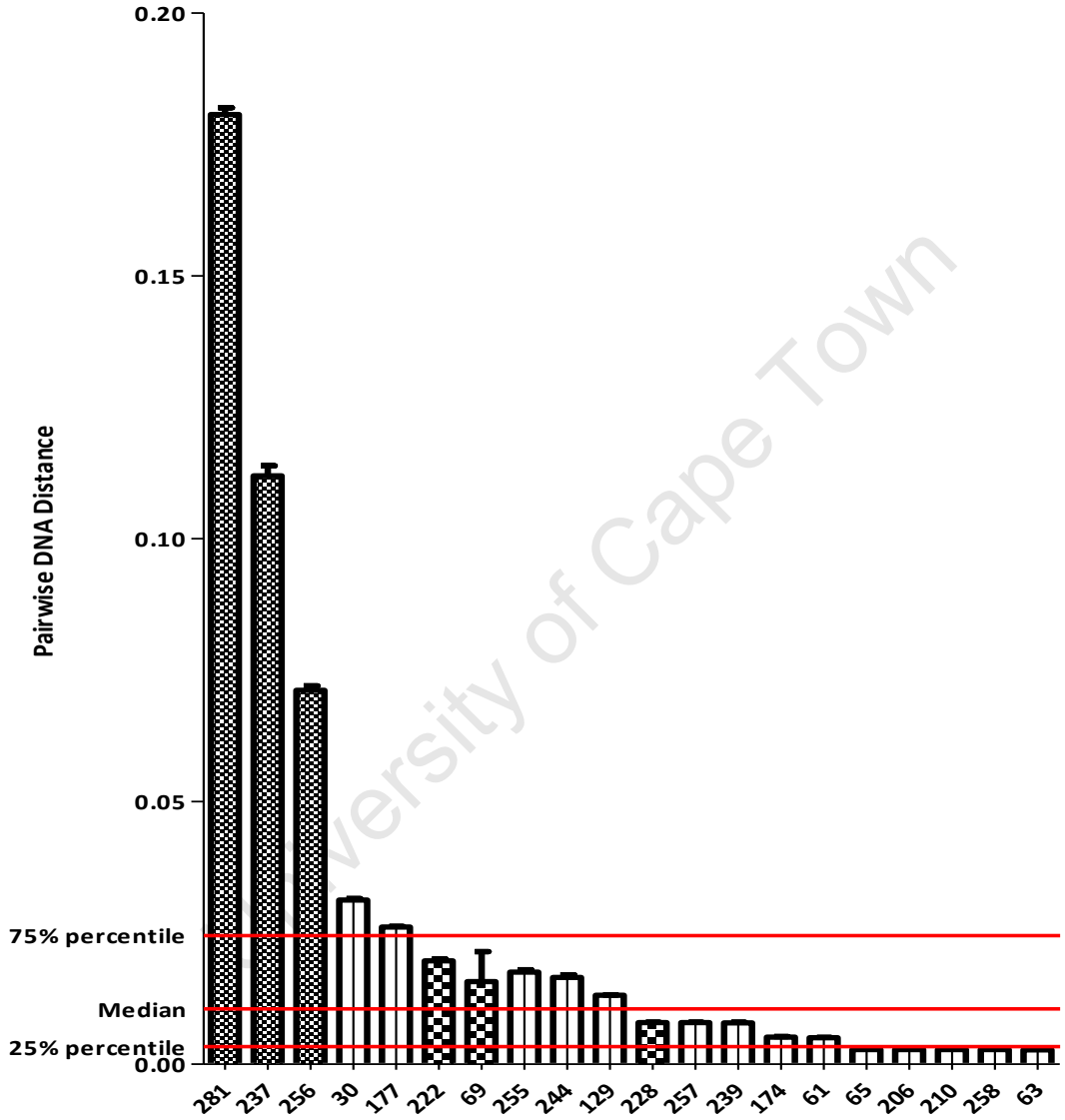


Figure 2.6 Pairwise DNA distance of each CAPRISA individual over the first year of infection. The bars with the small checkered pattern indicate those individuals identified as superinfected, the larger checkered bars indicate those individuals proposed to have heterogeneous primary infections and the striped bars indicate those individuals with homogenous primary infections. The red lines indicate the inter-quartile range and the median pairwise DNA distance of the all 20 individuals screened at twelve months post infection.

When the individuals were grouped according to their heteroduplex mobility assay results: homogenous, heterogeneous or superinfected, the median DNA distance of each group (0.5%, 2% and 11.2%, respectively) indicated that the superinfected individuals had significantly higher pairwise DNA distance than those infected with homogenous and heterogeneous viral populations (p= 0.005) (Figure 2.7A).

It would be expected that the genetic distance in a dual infected individual would be greater than or equivalent to the genetic distance between two epidemiologically unlinked viruses in the same cohort. Therefore the median pairwise intra-participant DNA distance was determined at twelve months post infection for all the individuals (n=20) screened in this study and was found to be 5.7%. All of the superinfected individuals had a DNA distance greater than the median intra-participant DNA distance, providing further evidence that these individuals were infected with two different viral populations.

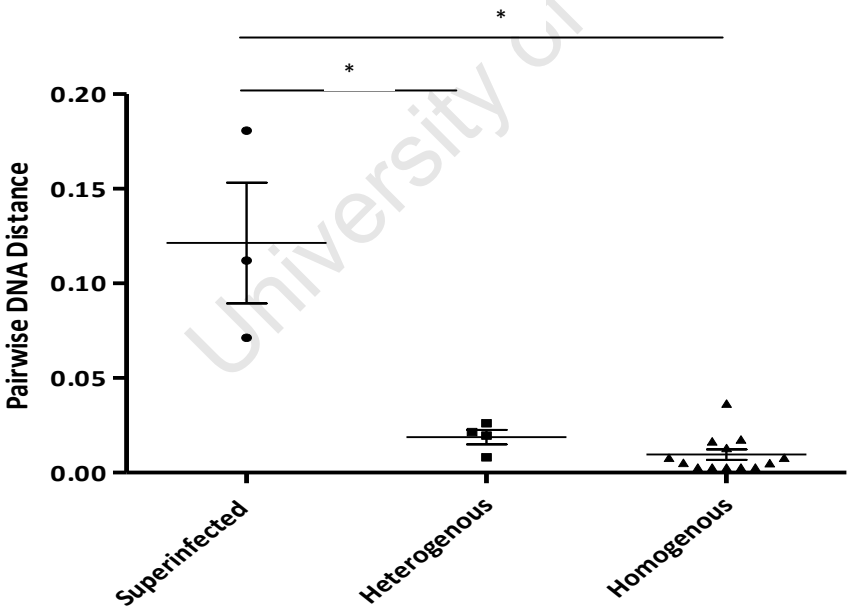


Figure 2.7 Comparison of the observed pairwise DNA distance over the first 12 months of infection in individuals with superinfections, heterogeneous primary infections and homogenous primary infections.

2.3.4 Effect of superinfection on disease progression

Several studies have shown faster disease progression to be associated with superinfection; therefore we wanted to determine if any of the superinfected individuals in this study exhibited an accelerated disease progression.

CAPRISA individual CAP237 (Figure 2.8) was superinfected sometime after enrolment and prior to three months post infection. There was a substantial change in viral load at the time of superinfection (>1 log) as well as a decline in CD4+ T-cell numbers from 620 cells/ul to 434 cells/ul. Some control of viremia was regained subsequent to superinfection however this control was interspersed with arbitrary spikes in viral load for the remainder of follow up. However, these spikes in viral load never exceeded 40 000 copies suggesting some degree of control over both viruses. Her CD4 cell counts fluctuated throughout the four years of follow up however her cell numbers never declined below 200 cells/ ul for longer than one sampling point. Thus for this individual although their control was not consistent their disease progression did not appear to be accelerated by the occurrence of superinfection.

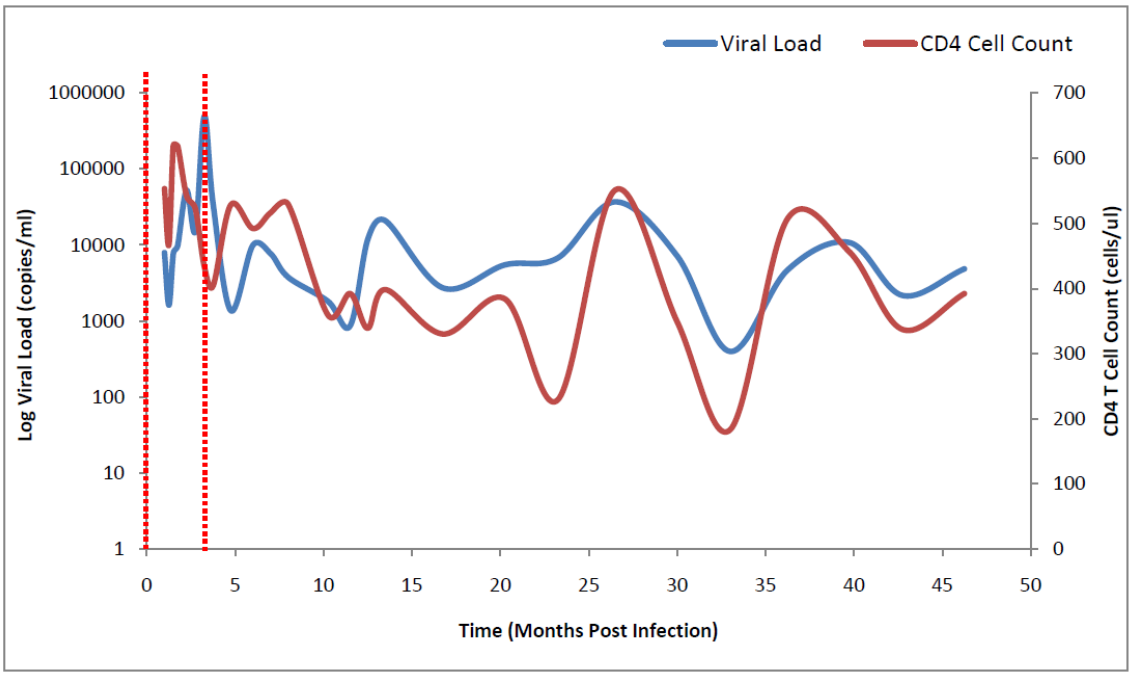


Figure 2.8 The viral load profile of CAPRISA Individual CAP237 over the first 50 months of infection, showing their HIV-1 plasma viral load and CD4+ T cell count. The viral load is indicated by the solid blue line and the CD4 cell count by the solid red line. The window period during which superinfection occurred is indicated on the figure by the dotted red lines.

For CAPRISA individual CAP256 (Figure 2.9), the superinfection occurred between three and four months post infection. Superinfection was associated with a greater than 1 log change in viral load, which was accompanied by a sharp decline in CD4+ T-cell numbers. Viral replication was reduced temporarily following superinfection suggesting some control of the superinfecting virus however this was not sustainable and this individual maintained high viral loads for duration of follow-up (6 years) (> 100 000 copies per ml). The CD4+ T cell numbers were above 500 cells/ul prior to superinfection and after superinfection continued to decline over the next twenty months to a low of 246 cells/ul. This individual had high viral loads and low CD4 counts, and initiated therapy after 4.5 years when she met the South African criteria (<200 cells/ul) for access to drugs.

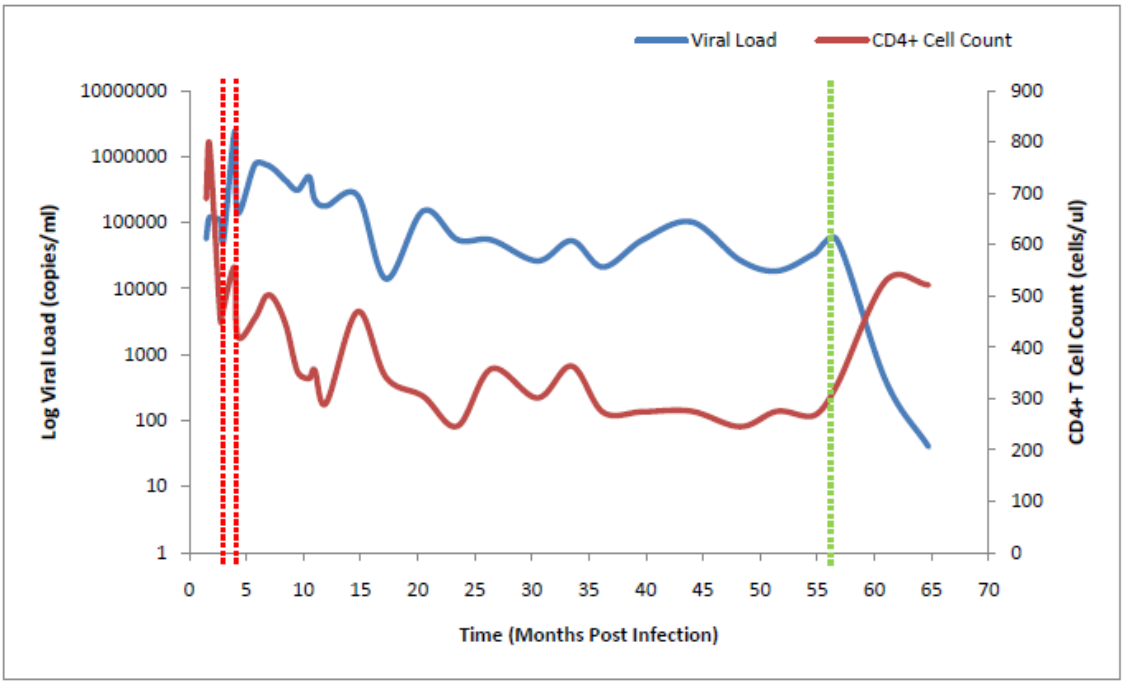


Figure 2.9 The viral load profile of CAPRISA Individual CAP256 over the first 70 months of infection, showing their HIV-1 plasma viral load and CD4+ T cell count. The viral load is indicated by the solid blue line and the CD4 cell count by the solid red line. The window period during which superinfection occurred is indicated on the figure by the dotted red lines and the initiation of antiretroviral therapy is indicated by the dotted green line.

The disease profile observed for CAPRISA individual CAP281 differed from CAP256 as prior to superinfection this individual had viral loads below the level of detection (<400 copies/ ml) (Figure 2.10). Superinfection occurred at around 11 months and was associated with an increase in viral load to levels similar to the initial infection (~6000 copies /ml). This individual regained control of her viral load after superinfection and maintained high CD4 counts of ~1000 cells/ul. These results suggest that this individual was better able to control the second virus than individual CAP256 and there was no evident acceleration of the disease as a result of the superinfection. The time to a CD4 count of 350 cells per ul was three years for CAP281 compared to two years for CAPRISA 256. Considering the control that this individual had shown previously it is interesting that after three years this individual started to lose control of their viremia. Further examination of sequences from later time points may provide some insight as to why this individual eventually loses control of their viruses.

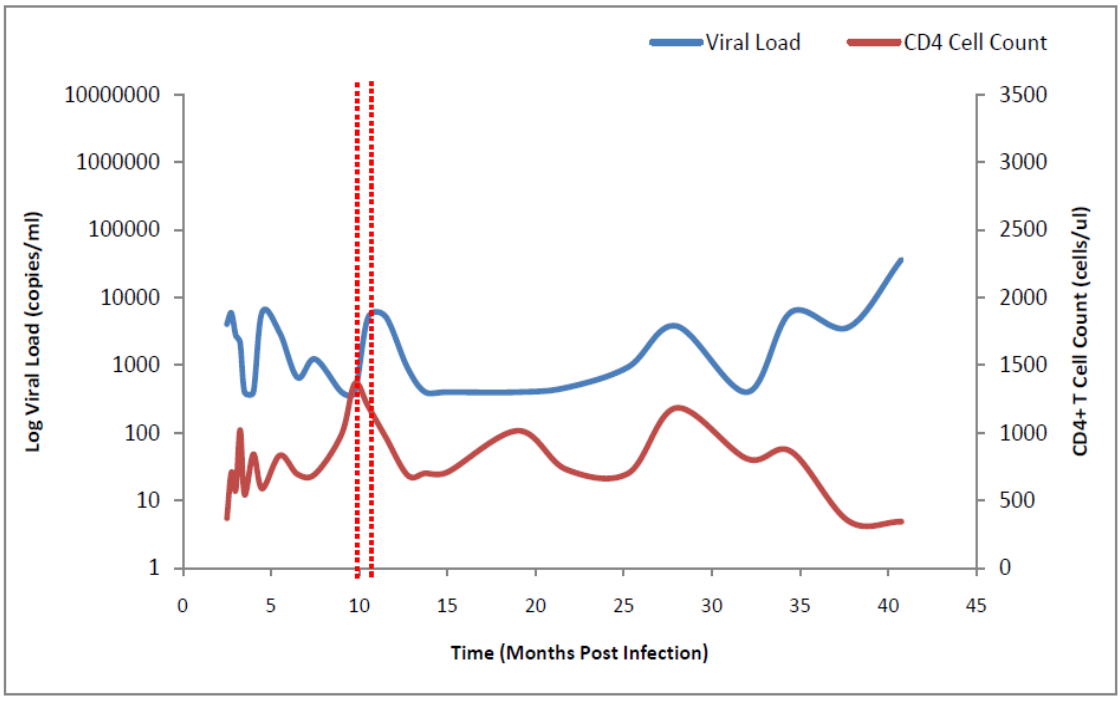


Figure 2.10 The viral load profile of CAPRISA Individual CAP281 over the first 27 months of infection, showing their HIV-1 plasma viral load and CD4+ T cell counts. The viral load is indicated by the solid blue line and the CD4 cell count by the solid red line. The window period during which superinfection occurred is indicated on the figure by the dotted red lines.

2.4 Discussion

Several studies have determined the frequency of HIV-1 superinfection. Most studies were carried out in regions where multiple subtypes circulate and have reported incidences of superinfection of 4% per year compared to an 8% per year incidence of primary infection (28). There are few studies in regions dominated by a single circulating subtype although one study found a 2.8% per person year incidence of superinfection in a subtype B infected cohort which was screened longitudinally for six years (108). Thus far there have been no incidence studies of HIV-1 superinfection in individuals infected with HIV-1 subtype C. In this study we found 10.5% of individuals superinfected (2/19), which is similar to the original incidence of infection of 7.2% (117). We found superinfection was not always associated with rapid disease progression with two of the superinfected individuals effectively controlling their viruses and only one individual progressing more rapidly to AIDS.

Cases of superinfection have often been identified as a result of a sudden unexpected increase in viral load (56, 79, 124). In this study all three superinfected individuals exhibited an increase in viral load which coincided with the estimated timing of HIV-1 superinfection. This spike in viral load was presumably as a result of the increased viral burden resulting from the presence of a second virus.

With respect to timing two individuals were re-infected within four months of initial infection (CAP237 and CAP256) whereas the other was re-infected at closer to one year post infection (CAP281). For individual CAP281 the superinfection occurs much later at closer to one year after the initial infection (Figure 2.10). These results suggest susceptibility to superinfection is not restricted to the early stage of infection where CTL immune responses are generally more focused and there is low or no autologous neutralizing antibody responses as superinfection of CAP281 occurred when both arms of the adaptive immune responses would have been mature.

There have been several reports of individuals presenting protective HLAs who following superinfection, progressed to AIDS (16, 102). In the first study the individual presented the B*27 protective HLA and was controlling their initial infection, after superinfection the individual lost control of their viral population through the development of an escape mutation in the KK10 epitope (102). In another study the B*57+ long-term non-progressor was proposed to have been initially infected with a weaker virus, this first virus was found to be nef-defective. After ten years of viral control this individual was superinfected with a fully competent virus and resulted in outgrowth of a recombinant virus which resulted in disease progression. In this individual, the neutralizing antibodies were unable to neutralize the second virus and only a very weak gag-specific CTL response had developed prior to superinfection (16). Therefore the presence of a protective HLA does not necessarily indicate that the individual will not progress to disease but rather the nature of the infecting virus, together with quantity and quality of the immune responses, is also important. CAP256 was B*5802 positive, an allele which has been associated with higher viral loads and faster disease progression (59, 75). She did not

effectively control the primary infecting virus, not the viruses after superinfection. This individual progressed to AIDS within 56 months of infection when her CD4+ counts reached 200 cells/ul and she initiated therapy. CAPRISA individual CAP281 on the other hand was B*4201 and B*4403 positive, both HLAs which have been associated with lower viral loads, higher CD4 cell numbers and slower disease progression (59, 65). She maintained a relatively low viral load up until the time of the superinfection, and after spike in viral load associated with superinfection regained control of the viruses reducing viral loads to the levels observed prior to superinfection (<400 copies/ml). These results suggest that HLA plays a dominant role in defining disease progression, even after superinfection. However at the time of this study, by 40 months post infection the viral load in CAP281 was increasing with a concomitant reduction in CD4 counts suggesting that this control may not be durable.

One of the consequences of HIV-1 superinfection is the explosion of diversity that results from two strains infecting a single individual. The typical intra-subtype distance among epidemiologically unrelated sequences is approximately 11%, while that expected of viruses from different subtypes or the inter-subtype distance is between 21 and 30% (73). Superinfected individuals had a median pairwise DNA distance of 11.2% when comparing viruses from enrolment and 12 months infection. The DNA distances observed over the first year of infection in superinfected individuals was substantially higher than individuals infected with homogenous or heterogeneous quasispecies. This indicates a much greater diversity within these individuals. Diversity is a crucial aspect of the virus's pathogenic ability as it enables evasion from immune recognition through mutagenesis and recombination (27).

HIV-1 evolves to adapt to the host environment, escape immune responses and resist antiretroviral therapy (106). A study by Shankarappa *et al.* (1999) showed that evolution within the envelope gene increases over the course of infection and peaks at the onset of AIDS (95) suggesting that it may play some role in disease progression. Viral diversity has been previously shown to be associated with increased disease progression and viral fitness (12, 92, 111). The

observed higher viral diversity seen in the superinfected individuals should provide their quasispecies with a substantial advantage by increasing the probability of variants with enhanced fitness.

The incidence of HIV-1 superinfection (10%) over the first year of infection suggests that it is occurring frequently in the CAPRISA 002 cohort. However it may not have pathogenic implications as seen in the case of CAP81 who is a slow progressor. However, superinfection does provide increased sequence diversity which can potentially facilitate escape. This will be investigated in the next chapter.

Chapter 3: Genetic characterization of superinfecting viruses

3.1 Introduction

HIV vaccine design is complicated by the genetic variability of the virus which results in genetic mismatch between the vaccine and the challenge virus. This mismatch may make the immune responses generated by the vaccine ineffective and it is proposed that this was a contributing factor to the failure of the Merck Ad5 vaccine trial (19). It is not yet understood what immune protection is required to protect against infection or re-infection with HIV however studying HIV superinfection and its correlates of protection may provide some insight. Understanding the selective forces which are acting on viral populations and the constraints on immune escape is central to developing an HIV-1 vaccine. Therefore examining the genetic evolution of two different viruses in the same individual, the role of recombination in driving HIV diversity, as well as the effect that host selective pressures have on these viruses will be important for the development of a vaccine which will elicit responses of considerable breadth.

Viruses within individuals with dual infection evolve both through HIV recombination between different virus strains, as well as through accumulation of point mutations. The accumulation of point mutations is a slower evolutionary route compared to recombination as an average only one substitution occurs per viral genome per generation (82), whereas a combination of multiple substitutions are exchanged through recombination in a single cycle of replication. The rate of recombination is estimated to be 2.8 times per genome per replication cycle (129) compared to point mutations which have a rate of one nucleotide substitution per genome per replication cycle (82). Recombination has the additional advantage in that polymorphisms which are gained during recombination have already proven to be viable in the parent virus therefore it is less likely for recombination to result in a deleterious mutation than it is during the accumulation of point mutations (85). In the scenario of superinfection, recombination between the initial and superinfecting viruses could lead to the generation of beneficial combinations of mutations, as well as the loss of potentially deleterious mutations. The resulting recombinant viruses may have a greater fitness than either of its parents (98). In this

way, recombination can accelerate disease progression directly by increasing the replicative fitness of the virus, or indirectly as it allows for faster escape from host immune pressures (90).

Examining full-length HIV-1 genomes provides information with more scientific power to characterise viral evolution, immune escape, patterns of recombination and viral pathogenesis. To date, only two superinfection studies have been published that have examined the full HIV-1 genome, both of which characterised viruses in a single individual (5, 102). Streeck et al., found that established and effective CTL responses to the initial infection were lost as a result of superinfection, large due to recombination between the two viruses. The superinfecting virus was carrying amino acid changes in epitopes which were associated with escape and following infection, the pattern of recombination that occurred between the two viruses appeared to be driven by selection for the amino acid residues associated with escape. This loss of CTL recognition subsequently resulted in loss of immune control in this individual and thus faster disease progression (102). Similarly, Altfeld et al., found that superinfected individuals who had substantial CTL responses to their initial infection were unable to protect against re-infection due to CTL escape mutations in the superinfecting virus (4).

This study examined the evolution of the full-length HIV-1 genome over time in two individuals, CAP256 and CAP 281, who had been superinfected at around 3 and 11 months post infection respectively. We aimed to determine if HLA class I restricted epitopes potentially recognised by cytotoxic T lymphocyte responses to the initial virus, were found in the superinfecting viruses. In addition, we aimed to determine if the superinfecting virus had a greater genetic distance from the circulating strain.

3.2 Materials and Methods

3.2.1 Participants

Two individuals from the CAPRISA002 Acute Infection cohort, who had been identified as superinfected, CAP256 and CAP281, (Chapter 2) were selected for characterisation of their evolving viral populations.

3.2.2 RNA Extraction

Plasma samples were obtained from before, after and at the estimated time of the superinfection. RNA was extracted the Qiagen Viral RNA Extraction Kit as per manufacturers' instructions (Qiagen, Valencia USA). The kit yielded 50ul of extracted RNA from a 200ul plasma sample, however for low viral loads a double extraction was performed where 400ul of plasma is used to generate 50ul of RNA. The RNA was then stored at -80°C until it was required.

3.2.3 Reverse Transcription

OligodT was used to generate HIV-1 full genome cDNA using the Superscript III reverse transcription kit (Roche, Basel Switzerland). This primer consists of a string of 20 deoxythymidylic acid residues and hybridizes to the poly (A) tail of mRNA thus initiating synthesis of a complementary strand. A mixture of 25ul of the extracted RNA, the primer OligoDT (1mM)(Integrated DNA Technologies, Coralville USA) and dNTPs (0.4uM) was made up. Following incubation at 65°C for 5 minutes, the solution was cooled to 4°C until 25ul of the mastermix was added to it. The mastermix was composed of the 5X First Strand Buffer, 2.5ul of DTT (40mM), 7.5 ul of DEPC treated water, 1 unit of RNase Inhibitor and 2ul of the enzyme Superscript III. All reagents excepting the primer, the dNTPs and the water were included in the reverse transcription kit. This tube was then heated at 45°C for 2 hours followed by denaturing of the enzyme at 70°C for 15 minutes, finally 1ul of RNase H (Invitrogen, Carlsbad USA) was added to the tube before it was heated to 37°C for 20 minutes. This amplification was carried

out using the Applied Biosystems 2720 Thermal Cycler (Applied Biosystems, Carlsbad, CA USA) and the cDNA was stored at -20°C until it was required.

3.2.4 Amplification of the HIV-1 Genome

Nested amplification was carried out which generated an approximately 9kb fragment covering the genome from the beginning of *gag* to the end of nef using the Expand Long Template PCR Amplification Kit (Roche, Basel Switzerland). Although the conditions of this reaction had been optimised previously in some cases the amount of cDNA added needed to be adjusted to get amplification.

The first round reaction typically consisted of 18.2ul of distilled water, dNTPs (0.4mM), 0.3ul of each primer (50pM/ul), 23.25ul of distilled water, 10X Buffer 1, 0.75ul of Expand enzyme and 1ul of cDNA. The primers used in the first reaction were 1.U5C, 1.U5Cb and 1.3.3pIC (Table 3.1)(Integrated DNA Technologies, Coralville USA). The cycling conditions consisted of an initial denaturation step at 94°C for 2 minutes, followed by 10 cycles of denaturation at 94°C for 10 seconds, annealing at 60°C for 30 seconds and elongation at 68°C for 8 minutes. A further 20 cycles of amplification were then carried out which consisted of denaturation at 94°C for 10 seconds, annealing at 68°C for 30 seconds and elongation at 68°C for 8 minutes with an additional 15 seconds every cycle. The amplification concluded with a final elongation step at 68°C for 10 minutes and finally cooling to 4°C. A negative control was run with each set of reactions, where the negative control had distilled water added instead of cDNA, to ensure no contaminants were present. All amplifications were carried out on either the Applied Biosystems 2720 Thermal Cycler (Applied Biosystems, Carlsbad, CA USA) or the Applied Biosystems GeneAmp PCR System 9700 (Applied Biosystems, Carlsbad, CA USA).

Table 3.1 PCR Primers used for the Amplification of HIV-1 Genome

Name	Sequence	HXB2 Position	Forward/ Reverse
1.U5C	5'-GGGTGAGTGCTCTAAGTAGTGTGTGCCCGTCTGT	538–571	Forward
1.U5Cb	5'-GGGTGAGTGCTCTAAGTAGTGTGTGCCCATCTGT	538–571	Forward
1.3.3pIC	5'-GGGACTTAGAGCACTCAAGGCAAGCTTTATTG	526–557	Reverse
2.U5C	5'-GGCCGCGGATCCAGTAGTGTGTGCCCGTCTGTTGTGTGACT	540-580	Forward
2.3.3pIC	5'-GGCCGCGCGGCCGCTAGAGCACTCAAGGCAAGCTTTATTGAGGCTTA	519–565	Reverse

The second round reaction or the inner reaction had 18.5ul of distilled water, 0.3ul of each primer (50pM/ul), 0.9ul of dNTPS (0.4mM), 23.25ul of distilled water, 10X Buffer 1, 0.75ul of the Expand enzyme and 1ul of the PCR product from the first round reaction. The primers used in the second round reaction were 2.U5C and 2.3.3pIC (Table 3.1)(Integrated DNA Technologies, Coralville USA). The cycling profile used was the same as that used for the first reaction. Once again a negative control was included to ensure there was no contamination and amplification was carried out on the Applied Biosystems 2720 Thermal Cycler or the Applied Biosystems GeneAmp PCR System 9700 (Applied Biosystems, Carlsbad, CA USA).

3.2.5 Visualization of DNA Product by Agarose Gel Electrophoresis

Following the second nested amplification, the PCR products were visualized on a 1% TAE agarose gel (Appendix C). The loading dye contained Gel Red (Biotum, Hayward CA USA) and was added to each sample as well as the molecular weight marker. Gels were viewed under UV light using the UVIpro Silver UV box (UVItec, Cambridge United Kingdom).

3.2.6 Purification and Sequencing of PCR Product

For the products which were successfully amplified, re-amplification of the first round product was carried out to generate sufficient DNA for sequencing. PCR products were purified using the Zymo DNA Clean and Concentrator 5 Kit (Zymo Research Corporation, Orange USA) generating about 60ul of DNA product with a concentration of between 100 and 300ng/ul. These PCR products were sequenced by Stellenbosch Sequencing Unit using the ABI PRISM dye

terminator cycle-sequencing kit (Applied Biosystems, Carlsbad, CA USA) and then run on a genetic analyzer. A table of the primers used are included in the appendix (F1).

3.2.7 Sequence Assembly and Phylogenetic Analysis

The chromatograms were then assembled using Sequencher (Gene Codes Corporation, Ann Arbor USA). Assembled sequences were aligned with other HIV-1 whole genome sequences from other CAPRISA individuals as well as a group of subtype C sequences obtained from the Los Alamos HIV Sequence Database (www.hiv.lanl.gov) using Bioedit Sequence Alignment Editor Version 7.0.9.0 (51) and these alignment files were then imported into Mega Version 4.0 (104) to generate a phylogenetic tree using the neighbour joining method.

Sequence alignment files for each individual, containing the sequences for each of the time points amplified, were uploaded into the Highlighter tool on the Los Alamos HIV Sequence Database (www.hiv.lanl.gov).

3.2.8 Recombination Analysis

Recombination Identification Programme (RIP) was used to identify recombination breakpoints (www.hiv.lanl.gov). To illustrate the regions of recombination from the perspective of the genes within the genome the results of the RIP scan were then used to draw the recombinant virus. This was done using the Recombinant Drawing Tool on the Los Alamos HIV Sequence Database (www.hiv.lanl.gov) which uses the recombination breakpoints from the RIP results and maps them onto the HIV-1 genome.

3.2.9 Epitope Coverage of Superinfecting Virus by Initial Virus

The Epicover tool from the Los Alamos HIV Immunology Database (www.hiv.lanl.gov) was used to determine the extent of coverage of all 9-mers in the whole genome sequence of primary virus infection compared to the superinfecting virus.

3.2.10 HLA Typing

The four digit high-resolution HLA typing was performed at the National Institute of Communicable Diseases; Johannesburg, South Africa by C. Grays' laboratory. The Pel-Freez DNA Isolation Kit (Pel-Freez, Arkansas USA) was used to extract DNA from either granulocytes or PBMCs. Exons 2, 3 and 4 were sequenced using the Atria AlleleSeqr kits (Abbott) and Assign-SBT 3.5 (Conexio Genomics, Applecross Western Australia) to type alleles A, B and C respectively. Sequence specific primers were used to resolve any ambiguities as a result of either identical heterozygote combinations or polymorphisms outside the sequenced exons.

3.2.11 Identification of CTL Escape Mutations

To determine where CTL escape mutations were present in epitopes of the superinfecting virus the sequences were studied for amino acid changes in either the putative HLA restricted epitope or in the flanking regions. Putative epitopes were identified using the HLA data for each individual as well as the list of published epitopes on the Los Alamos Immunology Database (www.hiv.lanl.gov). HLA-associated polymorphisms associated with subtype C chronic infection were also taken into consideration (69, 89). Substitutions within an epitope that resulted in change from a high frequency residue with a low frequency residue were classified as escape. These amino acid frequencies were determined using an alignment of 100 sequences from South African subtype C infected individuals obtained from the Los Alamos Sequence Database (www.hiv.lanl.gov). The relative frequencies of amino acids not associated with HLA restriction,

and where amino acid changed to consensus from low to high frequency amino acid were classified as reversions.

3.2.12 Statistical Analysis

Most of the figures in this chapter were generated using GraphPad Prism 4.0 (GraphPad Software Incorporated, USA). Both the Kruskal-Wallis and the Mann Whitney tests were used to determine whether the epitope coverage of the superinfecting virus by the initial virus was significantly better or worse than other CAPRISA sequences.

3.3 Results

To examine the genetic differences between the initial and superinfecting strain the full length genome of the viruses in superinfected individuals CAP256 and CAP281 were amplified. For CAP256 four sequences were generated from four different time points (Table 3.2). There were eight sequences generated from six different time points (Table 3.2) for CAP281. For both individuals both the initial and superinfecting viruses were amplified.

Table 3.2 Participant Data

Participant ID	Presenting HLA	Time Point	Viral Load (copies/ml)	Sequence Source
CAP256	A*2902; A*6601; B*1503;B*5802; Cw*0401;Cw*0602	Enrolment	56500	H. Harvey
		3 MPI	51600	F. Treunicht
		4 MPI	2390000	H. Harvey
		6 MPI	626000	F. Treunicht
CAP281	A*0214; A*3001; B*4201;B*4403; Cw*0401;Cw*1701	Enrolment	4060	H. Harvey
		3 MPI	2770	H. Harvey
		6 MPI	657	H. Harvey
		7 MPI	1230	H. Harvey
		10 MPI	5020	H. Harvey
		11 MPI	5220	H. Harvey

HLA*B*5802 is associated with rapid progression, HLA B*4201 is associated with lower viral loads and slower disease progression; HLA B*4403 has been associated with lower viral loads and high CD4 cell numbers.

3.3.1 Characterisation of full-length genome sequences over time

The full-length HIV-1 genome was amplified and sequenced at different intervals over time for individuals CAP256 and CAP281 (Table 2.2). Similar to the C2-C3 phylogenetic clustering (Figure 2.7), the FLG sequences from CAP256 separated out between three and four months post infection, while sequences from CAP281 separated out between ten and eleven months post infection (Figure 3.1). In CAP256 the four months post infection sequences, and the CAP281 eleven month post infection sequences, showed no evidence of recombination in any of the genes indicating recent infection.

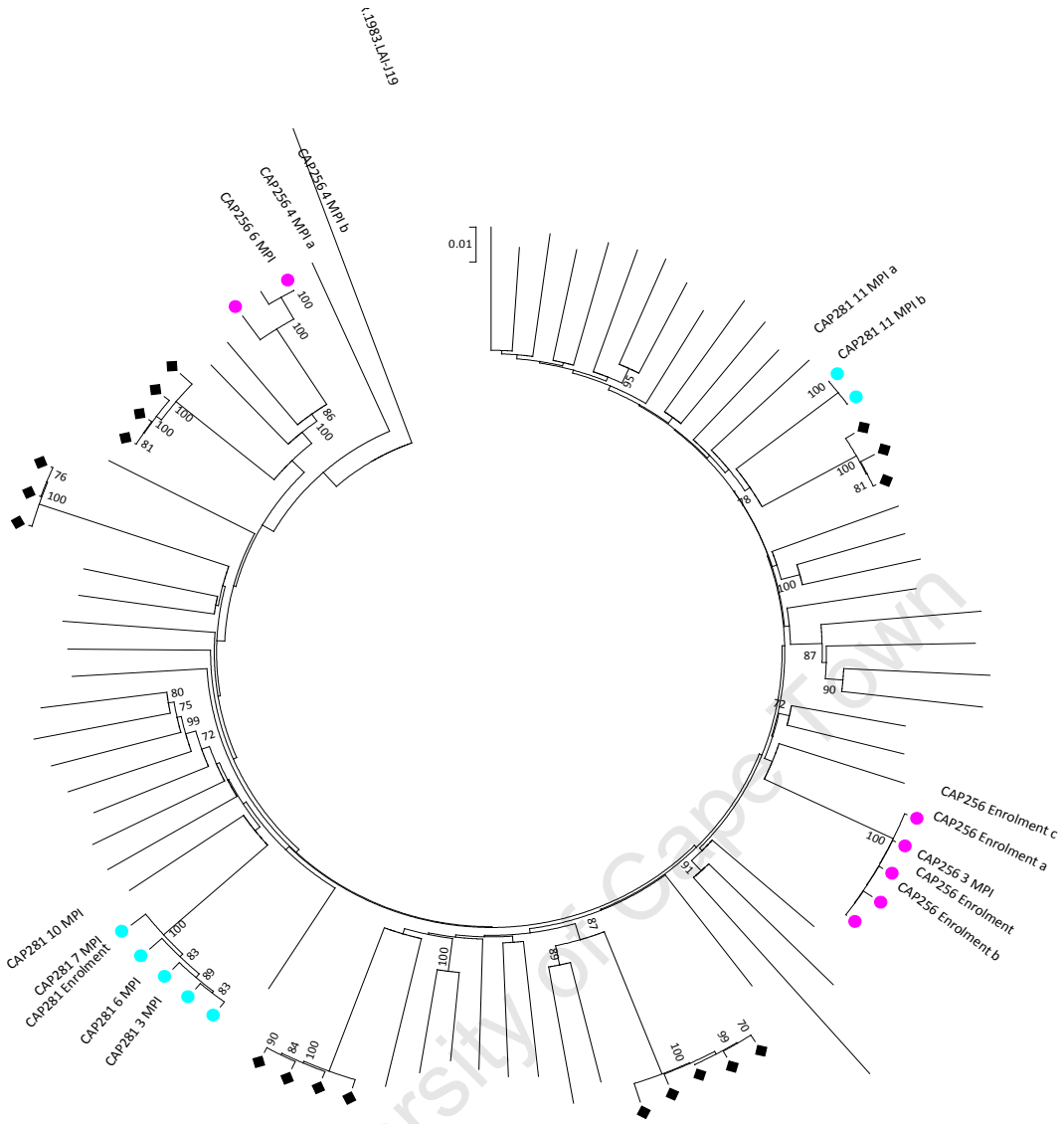


Figure 3.1 Neighbour joining phylogenetic tree of HIV-1 full length genome sequences (9000bp) from the CAPRISA cohort and reference sequences obtained from the Los Alamos HIV database (www.hiv.lanl.gov). The black diamonds indicate sequences generated from individuals in the CAPRISA cohort and the superinfected individuals CAP256 and CAP281 are indicated as the pink and turquoise circles respectively. Bootstrap values greater than 70% are indicated and the scale bar represents a DNA distance of 0.01%.

3.3.2 Recombination in CAPRISA Individual CAP256

The shift in viral populations can be visualised using highlighter plots which illustrate the number of nucleotide mismatches to the initial infecting virus (Figure 3.2A). At six months post infection, there is evidence of recombination between the primary infecting virus and the superinfecting virus, with the virus at this time point retaining a substantial portion of the sequence from the superinfecting virus (detailed below). For CAPRISA individual CAP281

(Figure 3.2B), evolution of the initial infecting strain can be visualised with the seven months post infection virus containing a substantial deletion at the start of envelope which may be an artefact of the PCR amplification.

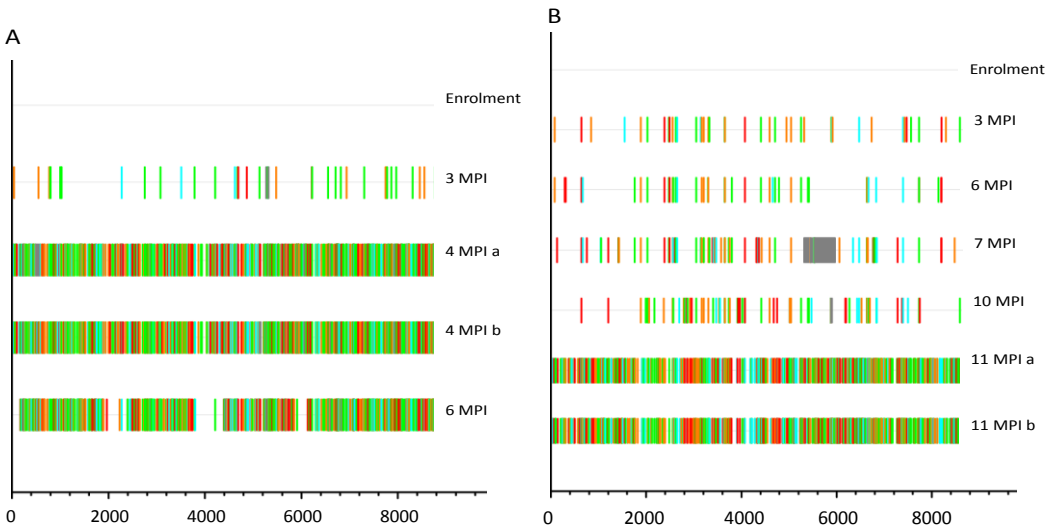


Figure 3.2 Highlighter plot of HIV-1 whole genome sequences at selected time points from A) CAPRISA individual CAP256 and B) CAPRISA individual CAP281 with a and b indicating different amplicons from the same time point. Each coloured line indicates a nucleotide mismatch with the primary infecting sequence. The colors in the highlighter plot represent changes to: A = Green, T = Red, G = Orange, C = Light Blue, IUPAC = Dark Blue and Gaps = Grey. The scale at the bottom refers to the number of base pairs.

The recombination patterns of viral populations in CAP256 over the three months following superinfection is more clearly illustrated using a recombination highlighter plot which distinguishes regions of the genome that originate from each parent (primary and superinfecting virus)(Fig 3.3). At 6 months post infection the recombinant is composed predominantly of the sequence from the superinfecting virus however there are four main regions where the sequence from the primary infecting virus was incorporated including regions of p17, a portion of reverse transcriptase, the majority of integrase, and V1/V2 *envelope* loops (Figure 3.3).

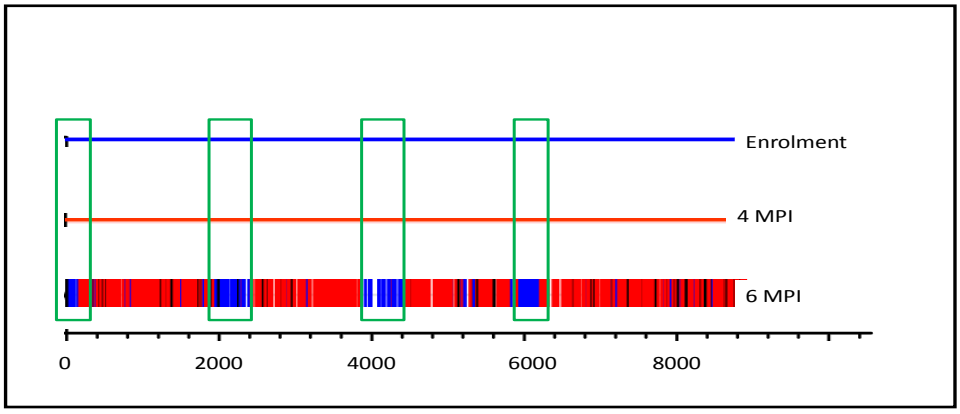


Figure 3.3 Recombination highlighter plot illustrating recombination in the 6 months post infection virus from CAPRISA individual CAP256. The plot highlights matches between the recombinant virus and its parent viruses. Therefore where the recombinant was most similar to either the primary infecting virus (blue) or the superinfecting virus (red) it is highlighted in the same colour as that parent. Regions where recombination occurred are indicated by the green squares.

To confirm the results of the recombination highlighter plot, the RIP recombination programme was used to query the six months post infection sequence (www.hiv.lanl.gov). The results of the similarity plot (Figure 3.4) reflects those seen in the highlighter plot, with the majority of the virus at six months post infection comprised of the superinfecting virus and four regions where the sequence originates from the primary infecting virus. These four regions in the RIP plots are in positions which are similar to the four regions indicated by the highlighter plot.

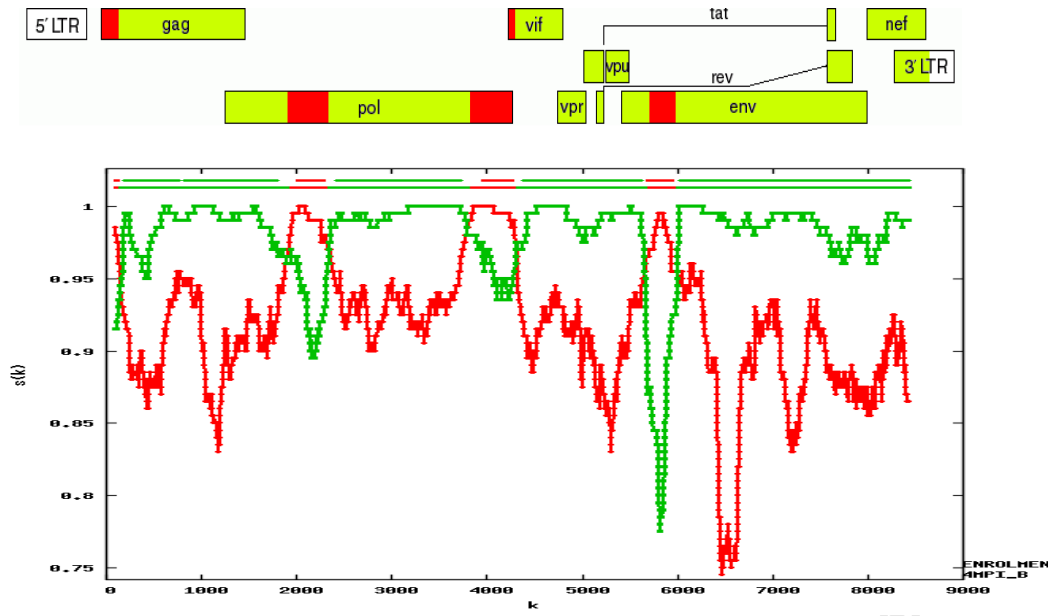


Figure 3.4 Illustration of the recombinant virus present at 6 MPI indicating regions of recombination between the primary infecting and superinfecting viruses. To determine the breakpoints of the recombinant virus a similarity plot was generated using a sliding window of 200 and a confidence interval of 99% and shows the similarity between the 6 MPI sequence and each of its parents, the primary infecting and superinfecting viruses. At the top of the plot the colour of the bottom line indicates the best match to the 6 MPI sequence with red indicating the primary infecting sequence and green the 4 MPI sequence. The presence of the top line indicates whether that match is significantly better than the other parent.

As we do not have sequences generated after eleven months post infection for CAP281, it was not possible to track evolving recombinant viruses in this individual.

3.3.3 Genetic Distance between the Initial and Superinfecting Viruses

To determine if the superinfecting virus was more different from the initial virus compared to viruses infecting other CAPRISA participants, a similarity plot was drawn. The overall similarity between the initial infecting virus in CAP256 and the superinfecting strain was relatively high (>72% across the genome) with the least similarity seen in *vpu* (82%) and *envelope* (88%), and the highest similarity seen in *gag* (92.5%) and *pol* (95%)(Figure 3.5). In the same way CAP281 had the least similarity between the initial and superinfecting viruses was observed in *vpu* (85%) followed by *envelope* (86.8%)(Figure 3.6). Once again *gag* (93%) and *pol* (95.5%) were the most conserved between the initial and superinfecting viruses (Figure 3.6).

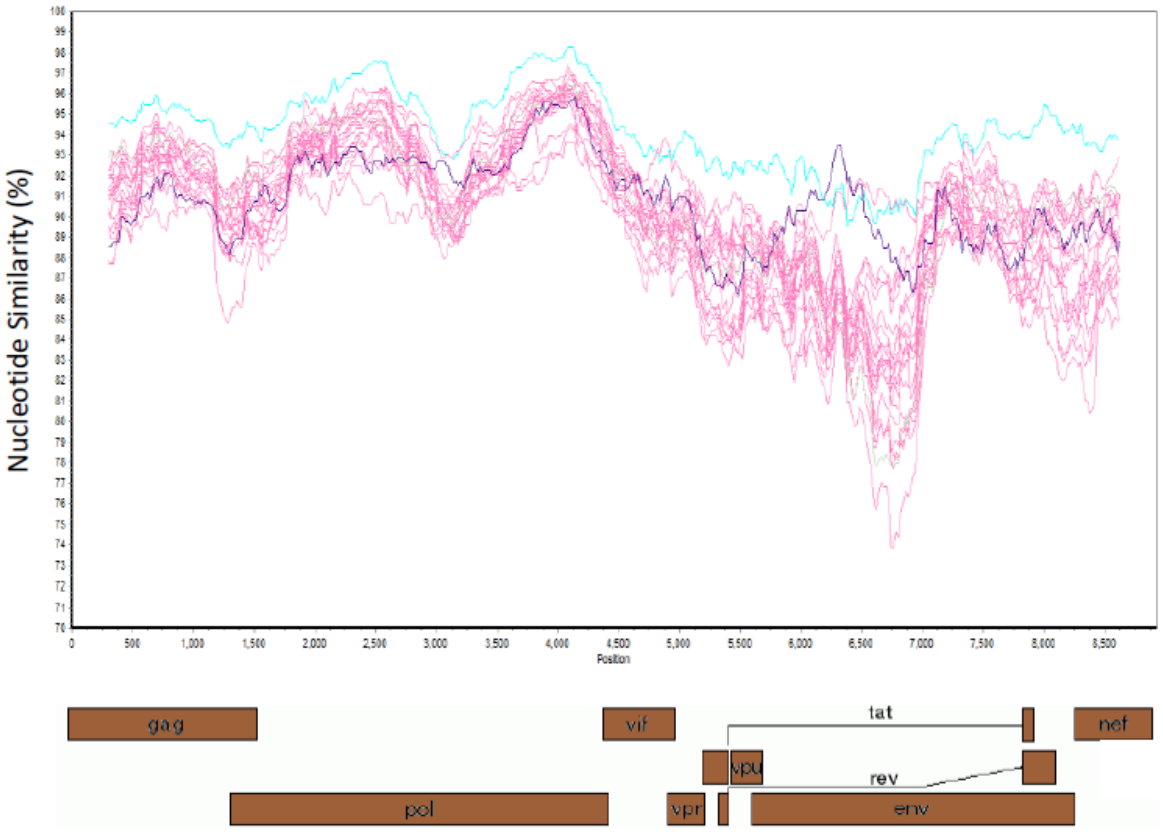


Figure 3.5 Similarity plot for CAP256 primary infecting sequence. This plot shows the percentage nucleotide similarity of the primary infecting sequence of CAP256 to the 4 MPI sequence (purple), to sequences from other CAPRISA participants (pink) and to the consensus C sequence from South Africa (turquoise). A sliding window of 600 nucleotides was used.

In CAP256 the consensus, generated from sequences acquired from CAPRISA002 women in acute infection, was closest to the primary infecting virus, as expected. The consensus C sequence was generated by aligning several whole genome sequences generated from participants in the CAPRISA 002 Acute Infection Study. A consensus was then generated from this alignment using the Consensus Maker tool on the Los Alamos HIV Sequencing Database (www.hiv.lanl.gov).The superinfecting virus appeared to have greater DNA distance in *gag* and *pol* compared to other CAPRISA sequences. These results suggest that, at least with *gag* and *pol*, greater genetic distance may be one factor enabling superinfection. This is similar to previous studies which have suggested that the genetic distance between the initial and superinfecting virus may be a factor in the occurrence of superinfection, where more distantly related viruses are likely to superinfect an individual (70).

A different trend is observed in CAP281 where the superinfecting virus lies in the middle of the other CAPRISA sequences when compared to the primary infecting virus. The similarity between the initial and superinfecting viruses is greatest in *gag* and *pol* with the superinfecting virus being considerably different in *envelope*, however these differences are no greater than those of other CAPRISA viruses. These results suggest that CAP281 is as different from the primary infecting virus as other circulating CAPRISA sequences and that for this individual the genetic distance may not have played a role in allowing superinfection to occur.

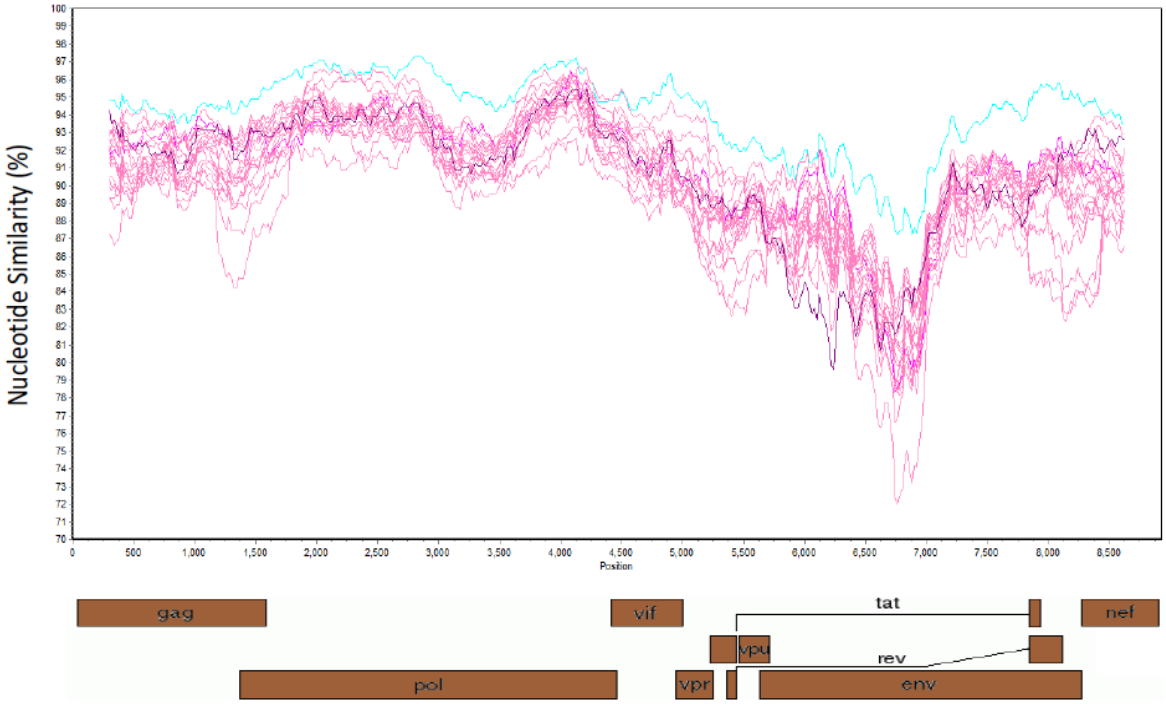


Figure 3.6 Similarity plot for CAP281 primary infecting sequence. This plot shows the percentage similarity of the primary infecting sequence of CAP281 to the 11 MPI sequence (purple), to sequences from other CAPRISA participants (pink) and to the consensus C sequence from South Africa (turquoise). A sliding window of 600 nucleotides was used.

3.3.4 Epitope Coverage of Superinfecting Virus by Initial Virus

To determine the effect of epitope diversity on superinfection, we estimated the epitope coverage by the primary infecting virus for both CAPRISA individual CAP256 and individual

CAP281. We compared the primary infecting virus to the superinfecting virus and then calculated the percentage of peptides of epitope-length in the superinfecting virus that match it. The primary infecting virus was also measured against sequences from other CAPRISA individuals to determine whether the coverage of the superinfecting virus by the primary infecting virus was any better than with any other viruses circulating in this region.

The results from CAPRISA individual CAP256 (Figure 3.7) suggest that there was less epitope coverage in the *gag*, *vpu* and *rev* proteins by the primary infecting virus. In the other proteins epitope coverage by the primary infecting virus was either similar, or in fact better, than other CAPRISA sequences. When we analyzed epitope coverage by the primary virus across the proteome, the difference in coverage across the genome was found to be highly significant (p= <0.0001)(Kruskal-Wallis Test). Further analysis of individual proteins showed that Gag from the superinfecting virus had significantly lower coverage by the primary virus compared to CAPRISA sequences (Mann-Whitney, p= 0.0353) as did Vpu (0.0156). For CAP256 the coverage of the CAPRISA viruses by the CAP256 primary infecting virus ranged from 18.2% to 58.5% with a mean of 22.6%. The coverage of the superinfecting virus ranged from 9.2% to 58.6% with a mean of 29.1%, suggesting that although the mean coverage was higher there were gene regions where the coverage was substantially lower. The lack of coverage in these regions may have contributed to a lack of CTL recognition which may have allowed superinfection to occur.

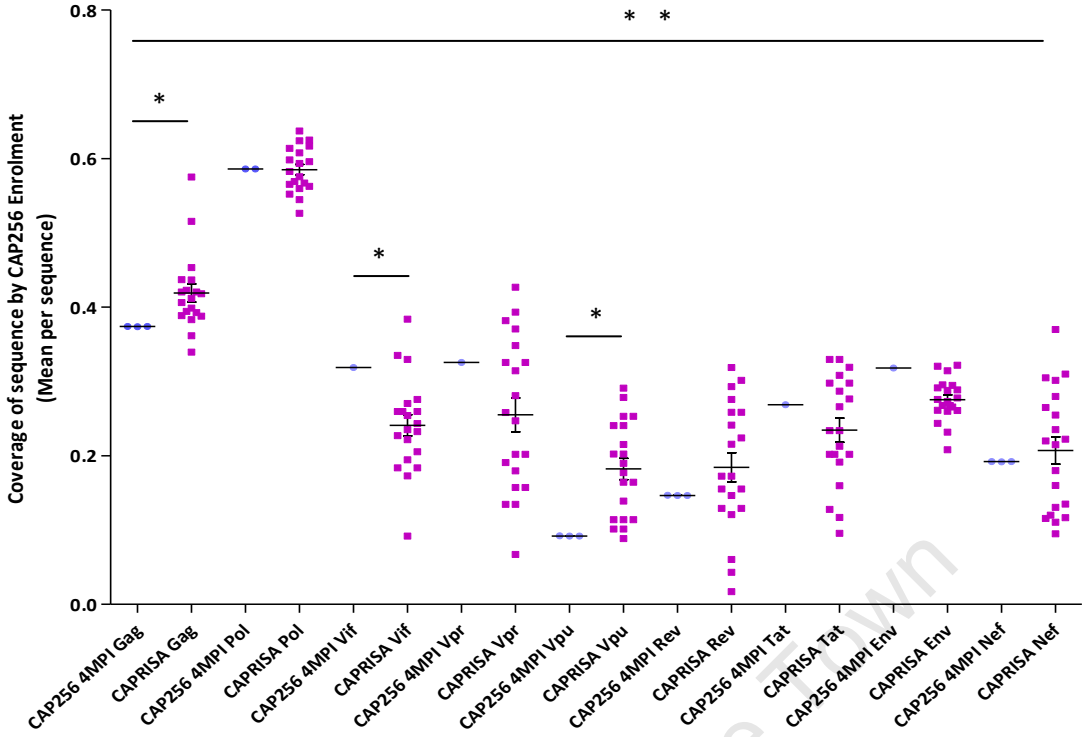


Figure 3.7 The epitope coverage by the CAP256 primary infecting sequence of the superinfecting virus and CAPRISA sequences (n=20) in each gene. For the analysis an epitope length of 9 amino acids was used and the coverage denotes the proportion of these 9-mers that were present in the primary infecting sequence averaged over all of the sequences analysed.

For CAPRISA individual CAP281 the trend was different to CAP256 as there were no regions of the genome in which the epitope coverage was lower than that observed in the other CAPRISA individuals (Figure 3.8) however the coverage of the superinfecting virus by the primary was found to be significantly different to that of the CAPRISA sequences (p= <0.0001). Interestingly in CAP281, the epitope coverage by the primary virus was significantly higher in Gag (p= 0.0071), Vpr (p= 0.0319) and Vpu (p= 0.0156) of the superinfecting virus than the other CAPRISA sequences. For the CAPRISA viruses the coverage by the CAP281 primary infecting virus ranged from 20.2% to 58.3% with a mean of 32.6%. This was only slightly lower than that of the superinfecting virus whose coverage of the primary infecting virus ranged from 20.2% to 60% with a mean of 39.93%.

However, this coverage never extends above 60% suggesting that the overall coverage by the primary infecting virus was relatively poor and that the CTL responses generated by the first infection may not be substantial enough to protect against re-infection. Together these results show that in CAP281, the superinfecting viruses were not more different from the primary infecting virus than the CAPRISA sequences; however differences were detected in epitope coverage in some CAP256 proteins from the superinfection strain. This suggests that for CAP281 the recognition of the second virus would have been the same as if they had been exposed to any of the other CAPRISA sequences, however for CAP256 the primary infecting virus appeared to differ from the CAPRISA sequences to a greater extent than the CAP281 virus suggesting that there were perhaps greater sequence differences between this virus and its superinfecting virus.

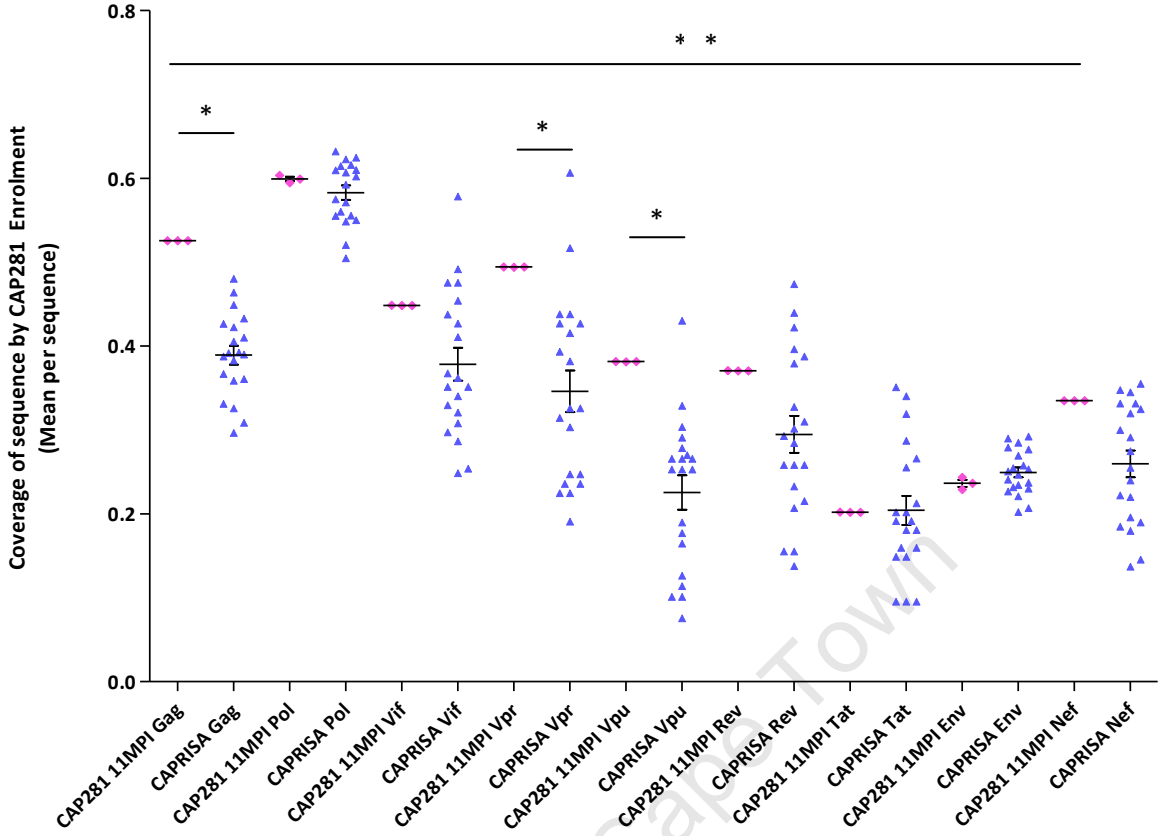


Figure 3.8 The epitope coverage by the CAP281 primary infecting sequence of the superinfecting virus and CAPRISA sequences (n=20) in each gene. For the analysis an epitope length of 9 amino acids was used and the coverage denotes the proportion of these 9-mers that were present in the primary infecting sequence averaged over all of the sequences analysed.

3.3.5 Mapping of HLA Restricted Epitopes in Primary and Superinfecting viruses

The epitope coverage tool does not take HLA into consideration. To further refine this analysis we identified putative epitopes associated with the HLA alleles of CAPRISA individuals CAP256 and CAP281, in both the primary infecting and superinfecting virus. We aimed to determine if CTL escape mutations in the superinfecting virus was the result of lack of immune recognition of the second infecting virus. We defined epitopes in two ways: (i) optimal epitopes, as defined by Frahm et al.(35), which have been experimentally shown to be restricted by the HLA in question; (ii) predicted epitopes are those which have been mapped to a peptide but where the optimal epitope has not yet been defined. The HLA alleles of CAP256 and CAP281 are described in Table 3.2. The optimal and predicted epitopes associated with these HLAs were identified

(Table 3.3 and Table 3.4). Any changes in epitopes proven in the literature to be associated with loss of recognition were classified as escape mutations, whereas putative escape were changes in amino acid residues within epitopes predicted to result in loss of HLA restriction using tools such as NetMHCpan and MotifScan. As escape has been associated with a change from high frequency to low frequency amino acids, we also took the amino acid frequency into consideration. A detailed summary of the change in sequence over time for each epitope is provided in the appendix for both the optimal and predicted epitopes.

Table 3.3 Summary of HLA associated Optimal and Predicted Epitopes across the HIV-1 Genome of CAPRISA individual CAP256 showing their location, their restricting HLA and their classification

Optimal Epitopes

Epitope Name	Gene	Restricting HLA	Epitope in Enrolment	Classification	Epitope in SI Virus	Classification
LY9	Gag (p17)	A*2902	LYNTVA**T**L**Y**	Wild type	LYNTVA**V**L**F**	Putative escape
VF9	Gag (p24)	B*1503	VKV**I**EEKAF	Wild type	VKV**V**EEKAF	Escape
RY9	Pol (Integrase)	B*1503	RKVKII**K**DY	Wild type	RK**A**KII**R**DY	Wild type
SY9	Gp160	A*2902	**S**FDPIPIHY	Wild type	**T**FDPIPIHY	Escape
SF9	Gp160	Cw*0401	SFNC**R**GEFF	Putative Escape	SFNC**G**GEFF	Escape
WF9	Nef	B*1503	W**M**FDS**S**LAR	Escape	W**K**FDS**L**LAR	Wild type

Predicted Epitopes

Epitope Name	Gene	Restricting HLA	Epitope in Enrolment	Classification	Epitope in SI Virus	Classification
HL9	Gag (p17)	Cw*0602	HIVWASREL	Escape	HL**V**WASREL	Wild type
QI9	Pol (Protease)	Cw*0401	QY**D**QI**T**I**D**I	Wild type	QY**E**QIL**I**E**I**	Escape
	Pol (Integrase)	B*1503	**T**KIQNFRVYY	Wild type	**IQI**QNFRVYY	Putative escape
	Vif	B*1503	SK**Q**A**K**GWFY	Putative Escape	SK**R**AN**C**WFY	Wild type
NY9	Tat	A*2902	KCYC**RH**C**S**Y	Wild type	KCYC**KR**C**C**Y	Escape

Epitopes were identified using the list of published epitopes on the Los Alamos Database (www.hiv.lanl.gov).

*Epitope sequence with respect to consensus subtype C

Bold residues indicate the residues at which amino acid changes were observed

For CAPRISA individual CAP256, eleven optimal epitopes were identified of which six were identified with differences in sequences between the primary infecting and superinfecting virus (Table 3.3). Of the six epitopes, four contained potential escape mutations in the superinfecting virus. Of the eleven predicted CTL epitopes restricted by the HLA alleles of CAP256, six were

identified with differences in amino acid residues between the primary infecting and superinfecting virus (Table 3.3). Three of these epitopes contained potential escape mutations in the superinfecting virus. The presence of escape mutations in these epitopes suggests that CTL recognition of these seven epitopes would have been diminished in the superinfecting virus.

For CAP281, of the six optimal epitopes identified in this individual there were five identified with differences in sequence between the primary infecting and superinfecting virus (Table 3.4). Of these five epitopes there were two which had amino acid residues associated with escape. Of the 21 predicted epitopes identified, ten contained differences in sequence between the primary infecting and superinfecting viruses (Table 3.4). Of these ten epitopes there were four which contained amino acid residues which were associated with escape.

Table 3.4 Summary of HLA associated Optimal and Predicted Epitopes across the HIV-1 Genome of CAPRISA individual CAP281 showing their location, their restricting HLA and their classification

Optimal Epitopes

Epitope Name	Gene	Restricting HLA	Epitope in Enrolment	Classification	Epitope in SI Virus	Classification
LY9	Gag (p17)	B*4403	L**Y**NTVATLY	Wild type	L**F**NTVATLY	Escape
YL9	Pol (RT)	B*4201	YP**R**IKV**RQ**L	Wild type	YP**G**IKV**KH**L	Escape
SF9	Gp160	Cw*0401	SFNC**G**GEFF	Escape	SFNC**R**GEFF	Wild type
RM9	Nef	B*4201	**K**PQVPLRPM	Escape	**R**PQVPLRPM	Wild type
TL10	Nef	B*4201	TPGPGVR**Y**PL	Wild type	TPGPGVR**L**PL	Escape

Predicted Epitopes

Epitope Name	Gene	Restricting HLA	Epitope in Enrolment	Classification	Epitope in SI Virus	Classification
	Gag (p17)	B*4201	RPGGKKQY**M**L	Wild type	RPGGKKQY**K**L	Escape
SL9	Gag (p17)	A*0214	SL**Y**NTVATL	Wild type	SL**F**NTVATL	Wild type
	Gag (p24)	Cw*0401	FFK**A**LRAEQ	Escape	FFK**T**LRAEQ	Wild type
	Gag (p24)	B*4403	AEQATQ**E**VKNW	Escape	AEQATQ**D**VKNW	Wild type
	Pol (Protease)	B*4403	E**E**INLPGKW	Wild type	E**D**INLPGKW	Wild type
	Pol (Integrase)	B*4403	**T**ETGQETAY**Y**	Escape	**A**ETGQETAY**F**	Wild type
HI10	Vif	B*4201	HPK**VS**SEVHI	Wild type	HPK**IA**SEVHI	Escape
	Vpr	B*4201	FPRWLH**S**L	Wild type	FPRWLH**G**L	Wild type
AY9	Gp160	B*4403	**WG**NL**W**VTVY	Wild type	**MG**NLWVTVY	Escape
	Gp160	B*4201	R**P**NNNTRKSI	Wild type	R**L**NNNTRKSI	Escape

Epitopes were identified using the list of published epitopes on the Los Alamos Database (www.hiv.lanl.gov)

*Epitope sequence with respect to consensus subtype C

Bold residues indicate the residues at which amino acid changes were observed

A summary of the total number of epitopes with putative escape, reversion and that were conserved between the initial and superinfecting viruses in CAP256 is presented in Figure 3.9 and for CAP281 in Figure 3.10.

For CAPRISA individual CAP256, of their optimal HLA associated epitopes, most were conserved between the initial and superinfecting strains (Figure 3.9A). A similar trend was observed with the predicted epitopes where most were conserved between the initial and superinfecting viruses (Figure 3.0B).Of the optimal epitopes of CAPRISA individual CAP281 there were equal numbers of epitopes which were conserved and which escaped (Figure 3.9B) however the

predicted epitopes contained a much larger number that were conserved between the initial and superinfecting viruses.

These results for both CAP256 and CAP281 suggest that even with more than half of the epitopes conserved between the two viruses this was not sufficient to protect against re-infection.

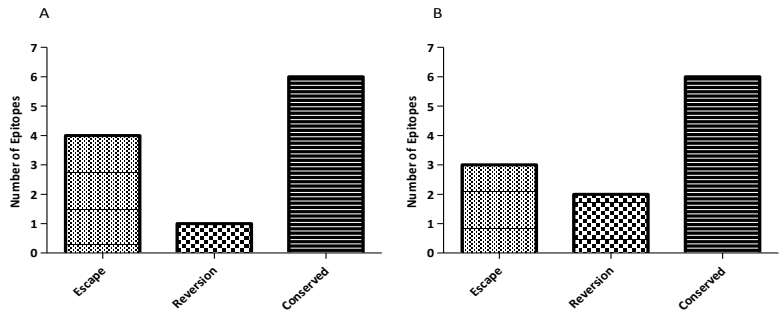


Figure 3.9 Frequencies of Escape, Reversion and Conservation of HLA associated epitopes within CAPRISA individual CAP256 for A) the optimal epitopes and B) the predicted epitopes across the whole HIV-1 genome.

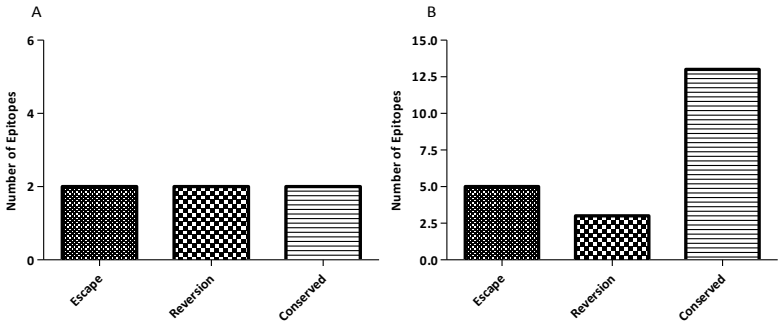


Figure 3.10 Frequencies of Escape, Reversion and Conservation of HLA associated epitopes within CAPRISA individual CAP281 for A) the optimal epitopes and B) the predicted epitopes across the whole HIV-1 genome.

3.3.6 Distribution of Escape Epitopes across the HIV-1 Recombinant Genome of CAP256

To determine whether selective pressure from the host immune response was driving the pattern of recombination we mapped the escape mutations in both the primary virus and the SI onto the HIV-1 genome (Figure 3.13 and Figure 3.14). The recombinant virus at 6 months predominantly contained regions from the superinfecting strain and thus the optimal and putative escape mutations in originating from the superinfecting strain were preserved. Interestingly, all of the regions of the primary infecting virus that were incorporated were those which carried escape mutations suggesting that for this individual recombination of the initial and superinfecting virus may have occurred so as to generate a virus with an advantage. This included the V1V2 regions which have been shown in a separated study to be associated with escape from neutralizing antibodies (P. Moore, NICD, Pers. Communication). These results indicate that superinfection may drive more rapid CTL escape through recombination between the initial and superinfecting strain which may have negative implications on the individual's disease progression.

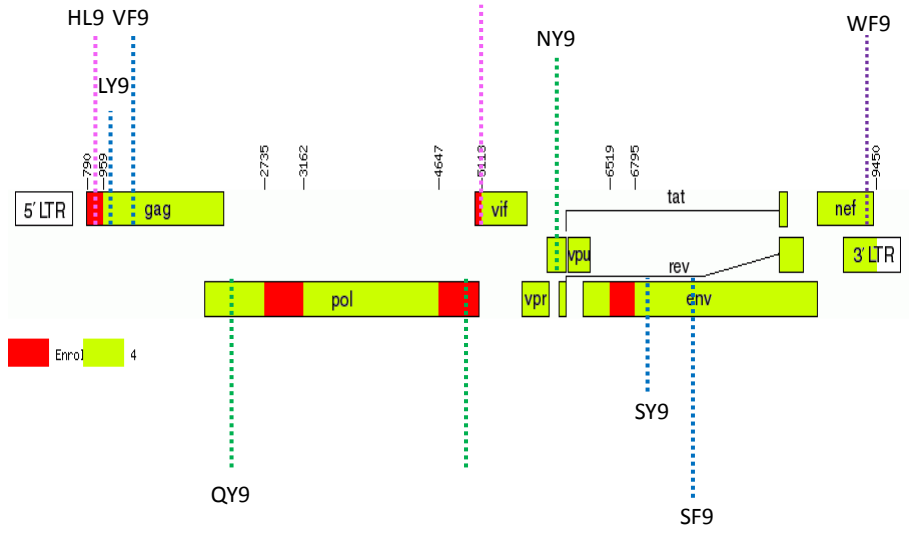


Figure 3.13 Distribution of HLA associated optimal and predicted epitopes plotted on the recombinant virus from 6 MPI of CAPRISA individual CAP256. The blue dotted lines indicate optimal epitopes with putative escape and the purple dotted lines indicate optimal epitopes with putative reversion. The pink dotted lines indicate predicted epitopes with putative reversion and the green dotted lines indicate predicted epitopes with putative escape.

3.3.7 CTL Data for CAP256

We compared our finding to predicted peptides responses based on deconvoluted pool analysis (Table 3.5)(data provided by C. Gray and Mandla Mlotshwa, NICD).

Table 3.5: CTL Responses in Gene Regions containing HLA restricted epitopes prior to, at and after Superinfection

Optimal Epitopes

Restricting HLA	Epitope in Enrolment	CTL Responses Prior to SI (SFU/10^6 PBMCs)	Epitope in SI Virus	CTL Responses At SI (SFU/10^6 PBMCs)	Classification	CTL Responses After SI (SFU/10^6 PBMCs)
A*2902	LYNTVATLY	0	LYNTVAVLF	472	Escape	93
B*1503	VKVIEEKAF	0	VKVVEEKAF	472	Escape	93
B*1503	RKVKIIKDY	178	RKAKIIRDY	1084	Conserved	1146
A*2902	SFDPIPIHY	3816	TFDPIPIHY	5446	Escape	936
Cw*0401	SFNCRGEFF	3816	SFNCGGEFF	5446	Escape	936
B*1503	WMFDSSLAR	494	WKFDSLLAR	844	Wild type	2809

Predicted Epitopes

Restricting HLA	Epitope in Enrolment	CTL Responses Prior to SI (SFU/10^6 PBMCs)	Epitope in SI Virus	CTL Responses At SI (SFU/10^6 PBMCs)	Classification	CTL Responses After SI (SFU/10^6 PBMCs)
Cw*0602	HIVWASREL	0	HLVWASREL	472	Escape	93
Cw*0401	QYDQITIDI	436	QYEQILIEI	2099	Wild type	2351
B*1503	TKIQNFRVYY	178	IQIQNFRVYY	1084	Wild type	1146
B*1503	SKQAKGWFY	0	SKRANGWFY	1948	Putative Escape	83
A*2902	KCYCRHCSY	148	KCYCKRCCY	118	Wild type	0

The IFN-gamma Elispot responses over time (Table 3.5) showed that prior to superinfection there was a narrow response, with only Env, Nef and Pol recognised out of the gene regions containing HLA restricted epitopes. There was a boost of both magnitude and breadth of responses with all of the proteins recognised at the time of superinfection. This increased breadth was partially sustained after superinfection and at the time at which the recombinant is dominant there are several genes recognised. However envelope, to which there had previously been substantial responses, now had much lower responses which may have been a result of the two escape mutations carried by the superinfecting virus. In addition Nef had gained in the number of responses which may have been due to its return to the wild type sequence which generated a greater number of responses. Thus it would seem that these responses are supportive of the proposal that host immune selection pressure shaped the pattern of

recombination in the virus at 6 months post infection to help the virus rapidly escape immune recognition.

3.4 Discussion

In this study we found evidence in CAP256 that genetic distance may play a role in enabling superinfection to occur presumably through escape from the immune response. For this individual there were also significant differences in epitope coverage between the initial and superinfecting viruses in several proteins. These differences may have contributed to the lack of recognition of the superinfecting virus by the immune response triggered by the first infection. To examine this further HLA restricted epitopes were studied for differences in sequence associated with escape. Several epitopes within the superinfecting virus were found to contain escape mutations. Finally a recombinant virus in CAP256 was studied; we found that the pattern of recombination in this virus appeared to be moulded by host immune selection as the recombinant was composed predominantly of the superinfecting virus which carried several escape mutations. The Elispot responses over time were supportive of this proposal as several of the responses to previously immunodominant epitopes were diminished at 6 months post infection when the recombinant virus was dominant.

However, genetic distance does not play such a large role in every case of superinfection as this was not seen in CAP281. For this individual the genetic distance between the superinfecting virus and the primary infecting virus was as great as that of other circulating viruses. With respect to epitope coverage by the primary infecting virus, the coverage was significantly better in several proteins within the superinfecting virus than in other circulating viruses. The lower genetic distance between the initial and superinfecting viruses may have allowed the immune response to have better immune control over the superinfecting virus. When HLA restricted epitopes were examined a similar result to CAP256 was seen in CAP281, with several escape mutations present in HLA restricted epitopes. It was proposed that these escape mutations may

have prevented the superinfecting virus from being detected by the host immune response prior to superinfection and thus allowed superinfection to occur. There was no CTL or antibody data available for CAP281.

The genetic distance between the initial and superinfecting viruses had been previously suggested to influence the occurrence of superinfection with more genetically distinct viruses being more likely to re-infect an individual (70). On examining the genetic distance between the initial and superinfecting viruses in CAP256 it was discovered that in both Gag and Pol the superinfecting virus was more genetically distinct from the initial virus than other circulating CAPRISA sequences. Responses targeted against Gag have been previously shown to be involved with immune control of infection (58) and for individual CAP256 the Gag region was found to have one of the lowest percentages of epitope coverage with less than 40% of the initial virus matched the superinfecting virus. In addition when the coverage within HLA associated epitopes was studied the overall coverage of the superinfecting viruses by the primary infecting viruses in CAP256 was found to be significantly different to that of the other CAPRISA sequences. This suggests that the epitopes to which immune responses were directed towards were less likely to be recognized in the superinfecting virus than those of other phylogenetically unrelated viruses. These results are similar to those reported by Yang et al. where superinfection of an individual was associated with a CTL response which was primarily targeted at the initial virus and was thus unable to recognize the second virus. Their study found that the superinfecting virus had changes in the amino acid sequence of the epitopes targeted by the CTL response; they proposed that these sequence changes may have resulted in a lack of recognition of these epitopes (125).

In individual CAP281, the genetic distance and epitope coverage between the initial and superinfecting strains was not as substantial as with CAP256. The regions in which the genetic distance was the lowest was Gag and Pol with the greatest distance observed in Envelope. These results suggest that for CAP281 the likelihood of the immune response generated by the

primary infecting virus recognising the superinfecting virus was the same as for any of the other CAPRISA sequences.

We examined the HLA associated epitopes to determine whether the superinfecting virus contained any mutations which may have prevented CTL recognition. Both superinfected individuals, CAP256 and CAP281, contained amino acid substitutions which were identified as potential escape mutations in several epitopes. Almost one third of the HLA associated epitopes in these individuals had changes in sequence which may have resulted in a lack of recognition of this virus. However, there was also a substantial amount of cross recognition between the initial and superinfecting virus within HLA associated epitopes. More than half of the HLA associated epitopes in CAP256 and just under half in CAP281 were conserved between the initial and superinfecting viruses. Thus while the superinfecting viruses in both individuals carried mutations which have prevented recognition, there was also potential to cross-recognition of the superinfecting strains. Altfield et al. also described a case of superinfection where there were epitopes and responses which were conserved between the two viruses. However, as was the case with this our study these epitopes were not sufficient to generate an immune response which was cross protective (4).

A previous study had reported a case of superinfection followed by recombination which had resulted in loss of immune control through transmission of escape mutations from the superinfecting strain (102). Similarly in CAPRISA individual CAP256 the recombinant virus at 6 months contained regions from both the primary and the superinfecting virus associated with escape. Superinfection thus provided the virus with the opportunity to rapidly escape host immune responses through recombination resulting in the generation of a virus harbouring an array of both CTL and antibody and CTL escape mutations. These results illustrate the role of adaptive CD8 T cell responses and antibody selective pressure in molding viral genetic changes (3). CAP256 was already struggling to control their viral load prior to superinfection and the

resulting loss of recognition after recombination may have resulted in the observed acceleration of disease progression.

To further evaluate the role of pre-existing immunity to protection from superinfection, we had available interferon-gamma Elispot data (C. Gray and M. Mlotshwa, NICD) and autologous neutralizing responses (P. Moore, NICD) from CAP256. This individual was superinfected between three and four months post primary infection. At this time there were no autologous neutralizing responses. However, after superinfection a neutralising antibody response began to develop (titres of approximately 62.5) which were directed predominantly against the superinfecting virus but which also exhibited some cross neutralisation of the primary infecting virus. There were CTL responses totalling 5200 SFU/106 PBMCs prior to superinfection compared to 8080 SFU/106 PBMCs after superinfection. Thus despite strong immune responses following superinfection the individual was still unable to control both viruses.

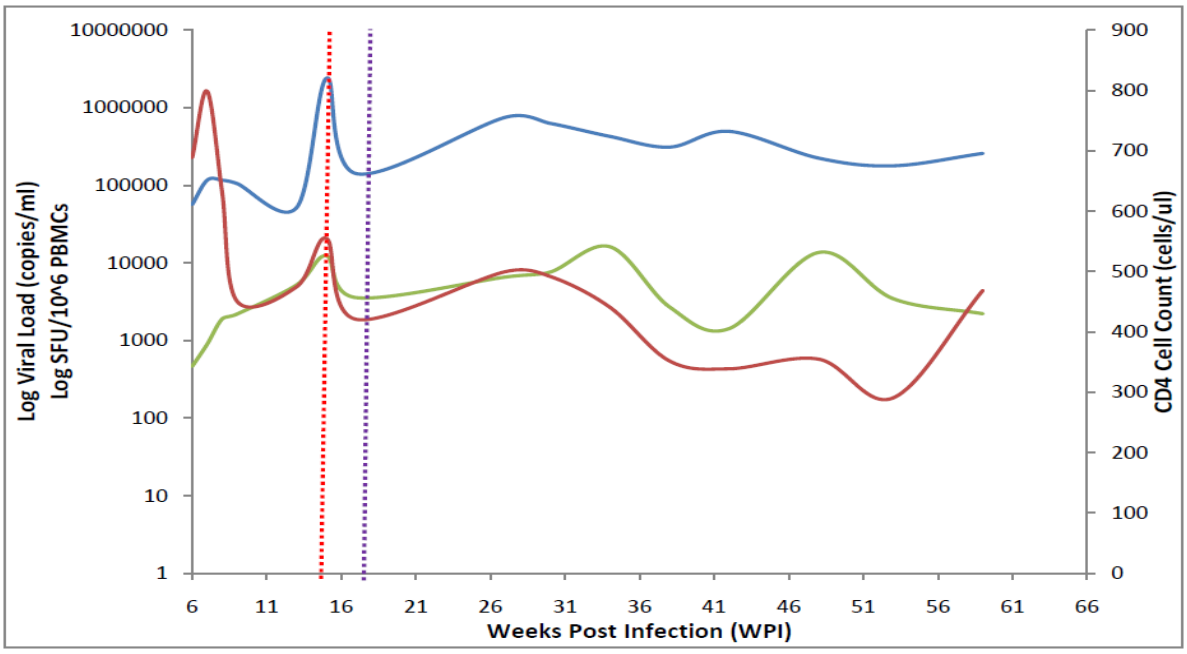


Figure 3.17 CAP256 Natural History of Infection. Disease progression profile for CAP256 showing their viral load (blue), CD4 cell count (red) and CTL responses (green) over time. The red dotted line indicates the occurrence of superinfection and the purple dotted line indicates the development of an autologous neutralising antibody response.

This chapter examined the genetic distance and epitope coverage between the initial and superinfecting viruses in two superinfected individuals. For one of these individuals (CAP256) the genetic distance between the initial and superinfecting virus was greater than the currently circulating viruses, particularly in gag and pol. Genetic distance may play a role in superinfection by preventing the host immune response from recognising the second virus. Support for this was provided by the epitope coverage across the genome which was significantly different to that of other CAPRISA sequences particularly in Gag, Vpr and Vpu. In addition the superinfecting virus was found to contain CTL escape mutations in several HLA restricted epitopes which may have contributed to the lack of recognition by the host immune responses. The host immune response also plays a role in shaping viral evolution with patterns of recombination appearing to be driven by the selection of escape mutations enabling the virus to rapidly escape immune pressure. This lack of immune recognition due to genetic differences between the initial and superinfecting viruses, together with rapid immune escape through recombination, may also have contributed to the inability of CAP256 to control both viruses and subsequently resulted in faster disease progression. CAP281 however showed a different pattern whereby there was a much lower DNA distance observed between the initial and superinfecting virus, as well as better epitope coverage which may have allowed host immune responses to control both primary and superinfecting virus. Futher research, with increased numbers, is required to definitively determine the role of different host and viral factors are in preventing re-infection.

Chapter 4: Summary and Conclusions

Vaccine strategies rely on the principle that pre-existing immunity to the pathogen will block infection or prevent disease. However, in HIV infection it is now established that individuals who are HIV positive are at risk of re-infection if exposed to another strain of HIV. Elucidation of the mechanism of protection from superinfection will provide important insights into the correlates of immune protection that an effective HIV vaccine would need to elicit.

The aim of this study was to determine the frequency and timing with which HIV superinfection was occurring within a cohort of high risk women from Kwa-Zulu Natal, South Africa. In addition we wanted to characterize the genome of the superinfecting viruses to determine whether the genetic distance between the initial and superinfecting viruses, and thus presumably the loss of immune cross-recognition, may have allowed superinfection to occur. Most previous studies have examined one or two regions of the HIV-1 genome of the superinfecting virus, this study looked at the entire HIV-1 genome to gain insight into the genetic diversity, patterns of recombination and immune escape.

The incidence of HIV-1 superinfection over the first year of infection was 10% per person year (2/19) which was similar to the reported incidence of primary infection in the cohort at 7.2% per 100 person year (117). This suggests that it is occurring as frequently as primary infection, although larger numbers are required to determine the significance.

It had been previously proposed that the genetic distance between the initial and superinfecting viruses may result in lack of recognition of the second virus by the host immune system thereby allowing re-infection. To determine if there were significant differences in sequence between the initial and superinfecting viruses the genetic distance across the genome between the two viruses was determined in two of the superinfected individuals. It was discovered that in one individual (CAP256) there were considerable differences in sequence between the initial and superinfecting viruses in *gag* and *pol*. In contrast, there was a much smaller genetic distance

between the primary and superinfecting viruses in CAP281 and for this individual the greatest distance was observed in *envelope*.

These differences in the CAP256 sequences were reflected in the 9-mer coverage of the superinfecting virus by the initial virus where coverage was found to be significantly lower in Gag, Vpu and Vpr in the superinfecting virus compared to other circulating CAPRISA viruses. Similar to the genetic distance results, for CAP281 the coverage was found to be significantly better in Gag, Vpu and Vpr than that of other CAPRISA sequences.

HLA associated epitopes were then examined in both the superinfecting and initial viruses to determine whether sequence changes may have prevented CTL recognition. Several of the HLA associated epitopes in the superinfecting viruses of both CAP256 and CAP281 contained amino acid differences associated with escape. Approximately one third of the HLA associated epitopes in both individuals had potential escape mutations which may have prevented CTL recognition however the sequence in more than half of the HLA associated epitopes was conserved between the initial and superinfecting viruses. These results suggest that substantial cross-reactive immune responses may be required to protect against re-infection.

We were able to track recombination patterns following superinfection for CAP256. We found that recombination facilitated immune escape and the recombinant virus at six months was predominantly composed of the superinfecting virus. The recombinant virus had gained several escape mutations from the superinfecting virus, however had retained regions of the genome of primary infection where the initial virus had already escaped CTL responses. Thus superinfection within this individual allowed the virus to rapidly escape host immune responses through recombination, as had been previously found by Streeck et al., and this may have contributed to an accelerated disease progression(102).

The three superinfected individuals differed in their disease progression profiles: CAP256 was evidently struggling to control her viral load before and after superinfection; while both CAP237

and CAP281 were effectively controlling their viral loads before and after superinfection. On studying their HLA alleles it was found that CAP256 presented the B*5802 HLA which was associated with high viral loads and faster disease progression which may explain the absence of control. CAP281 was found to present the HLA alleles B*4201 and B*4403 both of which are associated with low viral loads and higher CD4 cell counts which may account for the effective control of her viremia. No HLA data was available for CAP237 at the time of this study. The timing of the superinfection may also have played a role in disease outcome with both CAP237 and CAP256 superinfected within four months of initial infection when host immune responses were relatively immature and CAP281 was superinfected at close to one year post infection when most immune responses are developed. In addition the genetic distance between the initial and superinfecting viruses in CAP256 was far more pronounced than that of CAP281. This genetic distance may mean that in the case of CAP256 once the second virus was established host immune responses may have been unable to recognise this virus and thus been unable to control viral replication contributing to the loss of control seen in this individual. In the case of CAP281 the DNA distance is far less which may mean that once the individual had been infected by the superinfecting virus the immune responses were able to more rapidly gain control of the replication of this virus resulting in the steady control seen in this individual.

In conclusion this study has found that superinfection is occurring frequently with an incidence similar to that of initial infection. The occurrence of superinfection and the resulting recombination enables rapid evolution of the virus. The genetic distance between the primary and superinfecting viruses in some cases may play a role in enabling superinfection to occur as more distantly related viruses are not easily recognised by the established immune responses which is thus unable to prevent infection. Once infected the genetic distance may also play a role in how rapidly control of the viral replication of the superinfecting virus is achieved. This study also suggests that similar to single infection, cytotoxic T cell responses play an important role in determining the outcome of superinfection with individuals with protective HLAs appearing to

control their second viruses more effectively. Thus it would seem that the degree of cross recognition of the CTL responses, together with the maturity of the immune response, may be crucial in preventing infection through vaccination.

Appendices

Appendix A: Detailed summary of individuals who were screened for the occurrence of superinfection, including viral loads and CD4+ counts at twelve months post infection

Appendix A1: Viral Loads and CD4+ Cell Counts at 12 months post infection for the 20 individuals screened for superinfection

PID	Viral Load (copies/ml)	Log10 Viral Load	CD4+ Cell Count (cells/ul)
CAP30	57400	4.76	574
CAP61	<400	2.60	308
CAP63	214000	5.33	235
CAP65	71300	4.85	241
CAP69	1230000	6.09	202
CAP129	145000	5.16	634
CAP174	60700	4.78	363
CAP177	18000	4.26	381
CAP206	315000	5.50	337
CAP210	189000	5.28	288
CAP222	448	2.65	654
CAP228	1520	3.18	642
CAP237	11900	4.08	339
CAP239	166000	5.22	1016
CAP244	14500	4.16	303
CAP255	18200	4.26	397
CAP256	178000	5.25	291
CAP257	8260	3.92	634
CAP258	136000	5.13	233
CAP281	5220	3.72	970
Median	**66000**	**4.77**	**351**

Appendix A2: Detailed Summary of Results of Each Individual Screened including the DNA distance and heteroduplex mobility assay result

PID	Visit	Viral Load (Copies/ml)	HMA Result (Heteroduplex +/ 0)	DNA Distance
CAP30	Enrolment	10200	0	0
	3 MPI	616000	0	0.009
	6 MPI	136000	0	0.015
	12 MPI	57400	0	0.0369
	Mixed	NA	0	NA
CAP61	Enrolment	610	0	0
	3 MPI	4380	+	0
	6 MPI	7640	0	0
	12 MPI	<400	0	0.0051
	Mixed	NA	+	NA
CAP63	Enrolment	277000	NA	0
	3 MPI	215000	NA	0.0025
	6 MPI	186000	NA	0.0025
	12 MPI	214000	NA	NA
CAP65	Enrolment	90800	0	0
	3 MPI	33600	0	0.0536
	12 MPI	71300	0	0.0026
	Mixed	NA	0	NA
CAP69	Enrolment	5510000	0	0
	3 MPI	1890000	0	0.0646
	6 MPI	638000	+	0.0297
	12 MPI	1230000	0	0.0242
	Mixed	NA	0	NA
CAP129	Enrolment	55900	0	0
	3 MPI	20600	0	0.0026
	6 MPI	62600	0	0.0079
	12 MPI	145000	0	0.0132
	Mixed	NA	0	NA
CAP174	Enrolment	474000	0	0
	3 MPI	55900	0	0.0026
	6 MPI	28100	0	0.0053
	12 MPI	60700	0	0.0053
	Mixed	NA	0	NA

PID	Visit	Viral Load (Copies/ml)	HMA Result (Heteroduplex +/ 0)	DNA Distance
CAP177	Enrolment	698000	0	0
	3 MPI	60200	0	0.0052
	6 MPI	32300	0	0.0026
	12 MPI	18000	NA	0.0261
	Mixed	NA	+	NA
CAP206	Enrolment	368000	0	0
	3 MPI	113000	0	0
	6 MPI	156000	0	0.0026
	12 MPI	315000	0	0.0026
	Mixed	NA	0	NA
CAP210	Enrolment	127000	0	0
	3 MPI	50100	0	0.0161
	6 MPI	15700	0	NA
	12 MPI	189000	0	0.0026
	Mixed	NA	0	NA
CAP222	Enrolment	15800	+	0
	3 MPI	5770	0	0.0365
	6 MPI	2170	0	0.051
	12 MPI	448	NA	NA
	Mixed	NA	+	NA
CAP228	Enrolment	3840	0	0
	3 MPI	1330	0	0
	6 MPI	817	0	0
	12 MPI	1520	0	0.0079
	Mixed	NA	+	NA
CAP237	Enrolment	7950	0	0
	3 MPI	15400	+	0.1107
	6 MPI	9960	0	0.1111
	12 MPI	11900	NA	0.1139
	Mixed	NA	+	NA
CAP239	Enrolment	95800	NA	0
	3 MPI	23900	NA	0
	6 MPI	224000	NA	0.0026
	12 MPI	166000	NA	0.0077
CAP244	Enrolment	19200	0	0
	3 MPI	48900	0	0.0508
	6 MPI	54300	0	0.0333
	12 MPI	15100	0	0.0193
	Mixed	NA	0	NA

PID	Visit	Viral Load (Copies/ml)	HMA Result (Heteroduplex +/ 0)	DNA Distance
CAP255	Enrolment	196000	0	0
	3 MPI	96800	0	0.0026
	6 MPI	49200	0	0.0128
	12 MPI	74500	NA	0.018
	Mixed	NA	+	NA
CAP256	Enrolment	56500	0	0
	3 MPI	51600	+	0
	6 MPI	2390000	+	0.0724
	12 MPI	178000	0	0.0725
	Mixed	NA	+	NA
CAP257	Enrolment	276000	0	0
	3 MPI	173000	0	0
	6 MPI	16900	0	0.0081
	12 MPI	8260	0	NA
	Mixed	NA	0	NA
CAP258	Enrolment	735000	0	0
	3 MPI	268000	0	0
	6 MPI	228000	0	0.0026
	12 MPI	136000	0	0.0026
	Mixed	NA	0	NA
CAP281	Enrolment	4060	NA	0
	3 MPI	<400	NA	0.0027
	6 MPI	1230	NA	0.0054
	12 MPI	5220	NA	0.1795

Appendix B: Standard Molecular Biology Techniques

Appendix B1: Manual Extraction of Viral RNA

For amplification of the whole HIV-1 genome the RNA was extracted from plasma using the QIAamp® Viral RNA Mini Kit (Qiagen, Valencia CA). As instructed by the kit 800ul of AVL buffer, which contained carrier RNA, was added to a 200ul sample of plasma in a 2ml micro-centrifuge tube. The tube was mixed for 15 seconds by pulse vortexing and then incubated at room temperature for 10 minutes. Any remaining droplets on the lid of the tube were removed by a brief centrifugation after which 800ul of ethanol (96-100%) was added, once again the tube was mixed by pulse vortexing for 15 seconds and centrifuged to remove drops from the lid. 630ul of this solution was then carefully applied to a QIAamp spin column which was placed in a 2ml collection tube. The column was then centrifuged at 6000x g (8000rpm) for 1 minute, the collection tube containing the filtrate was then discarded and the spin column placed in a clean 2ml collection tube. Once more 630ul of the sample was applied to the spin column which was then centrifuged at 6000 x g for 1 minute. This step was repeated until all of the solution had been passed through the column. The spin column was then placed in a clean collection tube and 500ul of buffer AW1 was added to the column. The tube was centrifuged at 6000x g for 1 minute, the collection tube was then discarded and the spin column placed in a clean one. 500ul of buffer AW2 was added to the column and was centrifuged at 20 000x g (14 000rpm) for 3 minutes. To ensure there was no remaining A2 buffer on the column it was placed in a clean 2ml collection tube and centrifuged at full speed for 1 minute. To elute the RNA the spin column was placed in a clean 1.5ml micro-centrifuge tube and 50ul of buffer AVE was carefully applied to the column. The column was left to stand at room temperature for at least 1 minute and was then centrifuged at 6000x g for 1 minute. The eluted RNA was then either used directly or stored at -80°C until it was required.

Appendix B2: PCR Product Purification

To prepare amplified whole genome PCR products for sequencing the Zymo DNA Clean and Concentrator 5 kit (Zymo Research Corporation, Orange CA) was used. According to the manufacturer's instructions, 600ul of DNA binding buffer was added to 120ul of PCR product in a clean 1.5ml micro-centrifuge tube. The contents of the tube were mixed briefly by vortexing and then transferred to a Zymo-Spin Column in a collection tube. The column was centrifuged at ≥ 10 000x g for 30 seconds after which the flow through was discarded and 200ul of wash buffer was added to the column. The tube was again centrifuged at ≥ 10 000x g for 30 seconds. This wash step was then repeated. To elute the purified PCR product 60ul of distilled water was then added directly to the column matrix; the column was then placed in a clean 1.5ml micro-centrifuge tube and left at room temperature for at least one minute. The column was then centrifuged at ≥ 10 000x g for 30 seconds to elute the PCR product. The PCR product was now ready for sequencing.

Appendix C: Standard Buffers and Solutions

Appendix C1: 10X TBE (Tris-Boric acid EDTA) Buffer

108g Tris-HCl, 55g Boric acid, 20ml EDTA, made up to one litre with distilled water.

Appendix C2: 50X TAE (Tris-Acetic acid EDTA) Buffer

242g Tris Base, 57.1ml Glacial Acetic Acid, 20ml EDTA, made up to one litre with distilled water.

Appendix C3: 6X Agarose Gel Electrophoresis Loading Dye

0.25% bromophenol blue, 0.25% xylene cyanol FF 30% glycerol (In deionised water) with 1ul of Gel Red added per 10ml loading dye

Appendix C4: 10X HMA Annealing Buffer

1000mM NaCl, 100mM Tris-HCl, 20mM EDTA

Appendix C5: 1X Electrophoresis Buffer

100ml Buffer in 900 ml of distilled water to make up 1 litre (Diluted from either 10X TBE buffer or 10X TAE buffer).

Appendix C6: 1% Agarose gel

1% Agarose in either TBE or TAE Buffer

Appendix C7: 5% Non-Denaturing Polyacrylamide Gel

A 50 ml gel is prepared using 8.3ml acrylamide, 5ml of 10X TBE and 36.7ml distilled water. The gel is polymerized by adding 50mg Ammonium persulfate and 33ul of TEMED

Appendix D: Change in HLA associated CTL epitopes of CAPRISA individual CAP256

Appendix D1: Published Epitopes

MPI	Epitope	Restricting HLA	Gene	Classification	Comment	References
Enrolment	I Y N E W E L V					
3	. . . . .					
4	. . . . . V . E	A*2902	Gag p17	Escape	Y to E HLA associated	Brumme et al. 2008[2]; Kiepiela et al. 2006[10]; Masemola et al. 2004[11]
6	. . . . . V . E					
Enrolment	V K V I E E K A E					
3	. . . . .					
4	. . . V	B*1503	Gag p24	Escape	Subtype B cohort found the V substitution removed recognition	Zhai et al. 2008; Kiepiela et al. 2006[10]; Chopera et al. 2008[13]
6	. . . V					
Enrolment	R K V R I I K D Y					
3	. . . . .					
4	. . A . . . R . .	B*1503	Pol (Integrase)	No effect	The K to R substitution is returning the virus to wild type	Troyer et al. 2009[9]
6	. . . . .					
Enrolment	S F D F I P I I Y					
3	. . . . .					
4	I	A*2902	Envelope	Escape	T variant has been shown to have less than half the CTL recognition of the S variant	Jones et al. 2004[8]
6	T					
Enrolment	S I N C K G R I					
3	. . . . .					
4	. . . R	Cw*0401	Envelope	Escape	High to low frequency which suggests escape	
6	. . . G					
Enrolment	W N T D S S L A I					
3	. . . . .					
4	. D . . . E . . .	B*1503	Nef	Reversion	The G variant has been shown to elicit a potent CTL response as opposed to the R variant to which no responses are seen	Karlsson et al. 2007[9]
6	. D . . . E . . .					

Appendix D2: Predicted Epitopes

MPI	Epitope	Restricting HLA	Gene	Classification	Comment	References
Enrolment	H I V W A S K E L					
3	. . . . .	Cw*0602	Gag p17	Reversion	Novel epitope associated with this HLA	Thakar et al 2005(18)
4	. L					
6	. . . . .					
Enrolment	C Y D Q I T D I					
3	. . . . .	Cw*0401	Pol (protease)	Escape	Not reverts to wild type which is not recognized	Studies et al. 2005(17)
4	. . E . . L . E .					
6	. . I . . I . E .					
	I K I C N I K V Y					
Enrolment						
	. . . . .					
3	I Q	B*1503	Pol (integrase)	Escape	In subtype C infected individuals the I variant is only recognised by 4% of individuals with this HLA	Frahm et al 2008(19)
4						
	. . . . .					
6						
Enrolment	S K Q A K G W F Y					
3	. . . . .	B*1503	Vif	Reversion	Low to high frequency which suggests reversion	
4	. . R . V					
6	. . . . .					
Enrolment	E C Y C R H C S Y					
3	. . . K	A*2902	Tat	Escape	The R to K substitution is a reversion to wild type however the K to R substitution is associated with loss of CTL recognition	Jones et al. 2008(8)
4	. . . K R . C .					
6	. . . K R . C .					

Appendix E: Change in HLA associated CTL epitopes of CAPRISA individual CAP281

Appendix E1: Published Epitopes

MPI	Epitope	Restricting HLA	Gene	Classification	Comment	References
Enrolment	L Y N T V A T L Y					
3	. . . . .					
6	. . . . .	B*4403	Gag p17	No effect	The substitution from Y to F is equally recognised	Masemola et al. 2004(14), Honeyborne et al 2009
7	. . . . .					
10	. . . . .					
11	. F					
Enrolment	V P R I K V R Q I					
3	. . G					
6	. . R . . K . .	B*4201	Pol (RT)	Escape	Cytotoxic from higher to low frequency suggesting escape. P residue is an HLA associated reversion	Bourwell et al. 2007(1)
7	. . G					
10	. . G . . K . .					
11	. . R . . K H .					
Enrolment	S F N S G I F F					
3	. . . . .					
6	. . . . .	Cw*0401	Envelope	Reversion	Epitope allocated in a site free from positive selection. Cytotoxic from low to high frequency suggesting reversion	Long et al. 2008
7	. . . . .					
10	. . . . .					
11	. . . . R					
Enrolment	R P Q V P L R P M					
3	. . . . .					
6	. . . . .	B*4201	Nef	Reversion	The K variant is less common than the R variant	Leslie et al 2006(19)
7	. . . . .					
10	. . . . .					
11	R					
Enrolment	T P G P G V R Y P L					
3	. . . . .					
6	. . . . .	B*4201	Nef	Escape	Substitution from high to low frequency suggesting escape	
7	. . . . .					
10	. . . . .					
11	. . . . . L . .					

Appendix E2: Predicted Epitopes

MPI	Epitope	Restricting HLA	Gene	Classification	Comment	References
Enrolment	R P C G K K Q Y N L					
3	. . . . .					
6	. . . . .	B*4201	Gag p17	Escape	Epitope is associated with low viral loads when restricted by the HLA. Mutation is highly variable however frequency of E variant is much lower	Geldmacher et al. 2007[5]
7	. . . . .					
10	. . . . .					
11	. . . . . K .					
Enrolment	S L Y N T V A T L					
3	. . . . .					
6	. . . . .	A*0214	Gag p17	No Effect	Both amino acid residues are present at a similar frequency in subtype C infected individuals	[illegible] 2010[16]
7	. . . . .					
10	. . . . .					
11	. . F					
Enrolment	F F R A E R A C D					
3	. . . . .					
6	. . . . .	Cw*0401	Gag p24	Reversion	A to T is from low to high frequency suggesting reversion	Goonetilleke et al. 2009[6]
7	. . . . .					
10	. . . . .					
11	. . . T					
Enrolment	R E Q A T Q E V K N W					
3	. . . . .					
6	. . . . .	B*4403	Gag p24	Reversion	The E to D substitution is a known B*4403 reversion	Matthews et al. 2008[15]
7	. . . . .					
10	. . . . .					
11	. . . . . D . . .					
Enrolment	E E I N L P G K W					
3	. D					
6	. . . . .	B*4403	Pol (protease)	Escape	The E to D substitution is HLA associated. Escape occurs as recognition is mutually exclusive either the wild type or the variant is recognised	Matthews et al. 2008[15], Moore et al. 2007
7	. . . . .					
10	. . . . .					
11	. D					
Enrolment	T E T G G E E A Y Y					
3	. . . . .					
6	. . . . .	B*4403	Pol (Integrase)	Reversion	The T to A substitution is a reversion and the Y to I substitution does not have any effect	Matthews et al. 2008[15]
7	. . . . .					
10	. . . . .					
11	A . . . . . I .					
Enrolment	E A E V S S E V E I					
3	. . . . .					
6	. . . . .	B*4701	Vif	Escape	Epitope is seen in rapid progressors. Both substitutions move from high frequency to low frequency suggesting escape	Gray et al. 2009[7]
7	. . . . .					
10	. . . . .					
11	. . . I A					

MPI	Epitope	Restricting HLA	Gene	Classification	Comment	References
Enrolment	I P E P W L H S L					
3	. . . . .					
6	. . . . .	B*4201	Vpr	No Effect	Both amino acid residues are present at a similar frequency in subtype C individuals	Kiepiela et al 2004(12), Kiepiela et al 2006(10), Gray et al. 2009(7)
7	. . . . .					
10	. . . . .					
11	. . . . . R .					
Enrolment	W G N L W V T V Y					
3	. . . . .					
6	. . . . .	B*4002	Envelope	Escape	From high frequency residue to a low frequency residue suggesting escape	Jones et al. 2004(8)
7	. . . . .					
10	. . . . .					
11	M					
Enrolment	R P N N N T R R S I					
3	. . . . .					
6	. . . . .	B*4201	Envelope	Escape	From high frequency residue to a low frequency residue suggesting escape	Kiepiela et al 2004(11)
7	. . . . .					
10	. . . . .					
11	. L					

Appendix F: Primers used for Sequencing

Name	Sequence	HXB2 Position	Forward/Reverse
13R2C2	5' -GGCAAATATTGGAGTGTTATATGG	2712-2735	Reverse
2U5C	5'-GGCCGCGGATCCAGTAGTGTGTGCCCGTCTGTTGTGTGACT	540-580	Forward
5R3C	5' -CATTGCTCGTCCTACCCCCTGCCAC	7502-7526	Reverse
E230	5' -AATATTCATAATGATAGTAGGAGG	8276-8296	Forward
EF00	5' -GGGAAAGAGCAGAAGACAGTGGCAATGA	6204-6228	Forward
EF110	5' -CTGTTAAATGGTAGCCTAGCAGAA	7005-7025	Forward
EF15	5' -CTTGCTCTCCACCTTCTTCTTC	8424-8442	Reverse
EF55	5' -GCCCCAGACCGTGAGTTGCAACATATG	7914-7937	Reverse
ENVCR	5' -TTATATAATTCACTTCTCCAAT	7657-7678	Reverse
For9	5' -AAAATTAGCAGGAAGATGGCCAGT	4549-4572	Forward
G00	5' -GACTAGCGGAGGCTAGAAG	767-782	Forward
G35	5' -CATGCTGTCATCATTTCTTCTA	1817-1835	Reverse
GF80	5' -AGAGAACCAAGGGGAAGTGA	1477-1493	Forward
GF85	5' -TGCACTATAGGATAATTTTG	1177-1193	Reverse
H1P202	5' -CTAATACTGTATCATCTGCTCCTGT	2328-2352	Reverse
P10	5' -TACTCTGGAAAGGTGAAGG	5199-5214	Reverse
Rev12	5' -AGAGATCCTACCTTGTTATGTCCT	5488-5465	Reverse
SQ10FC	5' -GGAGCCAGTAGATCCTAACCTAGAG	5833-5857	Forward
SQ11RC3	5' -TTTAGGAGTCTTTCCCCATATTACTAT	3687-3713	Reverse
SQ15FC	5' -GAGAGCGGTGGAACTTCTGG	8561-8580	Forward
SQ16F	5' -CCACACACAAGGCTACTTCC	56-75	Forward
SQ3R2C	5' -GCTATGGTATCAAGCAGACTAATAGCACTC	8651-8680	Reverse
SQ4F	5' -ACAGGCTAATTTTTTAGGGA	2076-2095	Forward
SQ5F	5' -AAACAATGGCCATTAACAGAAGAGA	2613-2637	Forward
SQ6.5FC	5' -GCAGAGTTAGAATTAGCAGAGAACAG	3444-3469	Forward
SQ8R	5' -CTCCGCTTCTTCCTGCCATAGGAGAT	5963-5988	Reverse
SQ8RC	5' -TTCTACTACTCCCTGACTTTGGGGAT	4660-4685	Reverse
SQ9RC	5' -ATATGAATTAGTTGGTCTGCCAGGCC	5341-5366	Reverse

REFERENCES FOR APPENDIX

1. **Boutwell, C. L., C. F. Rowley, and M. Essex.** 2009. **Reduced Viral Replication Capacity of Human Immunodeficiency Virus Type 1 Subtype C Caused by Cytotoxic-T-Lymphocyte Escape Mutations in HLA-B57 Epitopes of Capsid Protein**. J. Virol. **83:**2460-2468. doi: 10.1128/JVI.01970-08.

2. **Brumme, Z. L., C. J. Brumme, J. Carlson, H. Streeck, M. John, Q. Eichbaum, B. L. Block, B. Baker, C. Kadie, M. Markowitz, H. Jessen, A. D. Kelleher, E. Rosenberg, J. Kaldor, Y. Yuki, C. Kadie, M. Carrington, T. M. Allen, S. Mallal, M. Altfield, and Heckerman, David. and Walker, Bruce. D.** 2008. Marked Epitope- and Allele-Specific Differences in Rates of Mutation in Human Immunodeficiency Type 1 (HIV-1) Gag, Pol, and Nef Cytotoxic T-Lymphocyte Epitopes in Acute/Early HIV-1 Infection. J. Virol. **82:**9216-9227. doi: 10.1128/JVI.01041-08.

3. **Chopera, D. R., Z. Woodman, K. Mlisana, M. Mlotshwa, D. P. Martin, C. Seoighe, F. Treurnicht, D. Assis de Rosa, W. Hide, S. Abdool Karim, C. M. Gray, C. Williamson, and and the CAPRISA 002 Study Team.** 2008. Transmission of HIV-1 CTL Escape Variants Provides HLA-Mismatched Recipients with a Survival Advantage. PLoS Pathog. **4:**. doi: 10.1371/journal.ppat.1000033.

4. **Frahm, N., C. Linde, and C. Brander.** 2006. Identification of HIV-Derived, HLA Class I Restricted CTL Epitopes: Insights into TCR Repertoire, CTL Escape and Viral Fitness. HIV Molecular Immunology. .

5. **Geldmacher, C., J. Currier, M. Gerhardt, A. Haule, L. Maboko, D. Birx, C. Gray, A. Meyerhans, J. Cox, and M. Hoelscher.** 2007. **In a mixed subtype epidemic, the HIV-1 Gag-specific T-cell response is biased towards the infecting subtype**. AIDS. **21:**135-143.

6. **Goonetilleke, N., M. K. P. Liu, J. F. Salazar-Gonzalez, G. Ferrari, E. Giorgi, V. V. Ganusov, B. F. Keele, G. H. Learn, E. L. Turnbull, M. G. Salazar, K. J. Weinhold, S. Moore, CHAVI Clinical Core B, N. Letvin, B. F. Haynes, M. S. Cohen, P. Hraber, T. Bhattacharya, P. Borrow, A. S. Perelson, B. H. Hahn, G. M. Shaw, and Korber, B. T. and McMichael, A. J.** 2009. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection Journal of Experimental Medicine. **206:**1253-1272. doi: 10.1084/jem.20090365.

7. **Gray, E. S., N. Taylor, D. Wycuff, P. L. Moore, G. D. Tomaras, C. K. Wibmer, A. Puren, A. DeCamp, P. B. Gilbert, B. Wood, D. C. Montefiori, J. M. Binley, G. M. Shaw, B. F. Haynes, and**

,Mascola, J. R. and Morris, L. 2009. Antibody Specificities Associated with Neutralization Breadth in Plasma from Human Immunodeficiency Virus Type 1 Subtype C-Infected Blood Donors. J. Virol. **83:**8925-8937. doi: 10.1128/JVI.00758-09.

8. **Jones, N. A., X. Wei, D. R. Flower, M. Wong, Michor., Franziska., M. S. Saag, B. H. Hahn, M. A. Nowak, and Shaw, George M. and Borrow, Persephone.** 2004. Determinants of Human Immunodeficiency Virus Type 1 Escape from the Primary $CD8^+$ Cytotoxic T Lymphocyte Response Journal of Experimental Medicine. **200:**1243-1256. doi: 10.1084/jem.20040511.

9. **Karlsson AC, Iversen AK, Chapman JM, de Oliveira T, Spotts G, et al.** 2007. Sequential Broadening of CTL Responses in Early HIV-1 Infection Is Associated with Viral Escape. Plos One. **2:**. doi: 10.1371/journal.pone.0000225.

10. **Kiepiela, P., K. Ngumbela, C. Thobakgale, D. Ramduth, I. Honeyborne, E. Moodley, S. Reddy, C. de Pierres, Z. Mncube, N. Mkhwanazi, K. Bishop, M. van der Stok, K. Nair, N. Khan, H. Crawford, R. Payne, A. Leslie, J. Prado, A. Prendergast, J. Frater, N. McCarthy, C. Brander, G. H. Learn, D. Nickle, C. Rousseau, H. Coovadia, J. I. Mullins, D. Heckerman, and Walker, Bruce. D. and Goulder, Philip.** 2006. **$CD8^+$ T-cell responses to different HIV proteins have discordant associations with viral load**. Nature Medicine. **13:**46-53. doi: 10.1038/nm1520.

11. **Kiepiela, P., A. J. Leslie, I. Honeyborne, D. Ramduth, C. Thobakgale, S. Chetty, P. Rathnavalu, C. Moore, K. J. Pfafferott, L. Hilton, P. Zimbwa, S. Moore, T. Allen, C. Brander, M. M. Addo, M. Altfeld, I. James, S. Mallal, M. Bunce, and L. D. Barber.** 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature. **432:**769. doi: 10.1038/nature03113.

12. **Kiepiela, P., A. J. Leslie, I. Honeyborne, D. Ramduth, C. Thobakgale, S. Chetty, P. Rathnavalu, C. Moore, K. J. Pfafferott, L. Hilton, P. Zimbwa, S. Moore, T. Allen, C. Brander, M. M. Addo, M. Altfeld, I. James, S. Mallal, M. Bunce, and L. D. Barber.** 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature. **432:**769. doi: 10.1038/nature03113.

13. **Leslie, A., D. A. Price, P. Mkhize, K. Bishop, A. Rathod, C. Day, H. Crawford, I. Honeyborne, T. E. Asher, G. Luzzi, A. Edwards, C. M. Rosseau, J. I. Mullins, G. Tudor-Williams, V. Novelli, C. Brander, D. Daniel. C., K. Photini., and , Walker, Bruce. D. and Goulder, Philip. J.** 2006. **Differential Selection Pressure Exerted on HIV by CTL Targeting**

Identical Epitopes but Restricted by Distinct HLA Alleles from the Same HLA Supertype. Journal of Immunology. **177:**4699-4708.

14. **Masemola, A. M., T. N. Mashishi, G. Khoury, H. Bredell, M. Paximadis, T. Mathebula, D. Barkhan, A. Puren, E. Vardas, M. Colvin, L. Zijenah, D. Katzenstein, R. Musonda, S. Allen, N. Kumwenda, T. Taha, G. Gray, J. McIntyre, S. Abdool Karim, and Sheppard, Haynes. W. and Gray, Clive. M.** 2004. **Novel and Promiscuous CTL Epitopes in Conserved Regions of Gag Targeted by Individuals with Early Subtype C HIV Type 1 Infection from Southern Africa**. Journal of Immunology. **173:**4607-4617.

15. **Matthews, P. C., A. Prendergast, A. Leslie, H. Crawford, R. Payne, C. Rousseau, M. Rolland, I. Honeyborne, J. Carlson, C. Kadie, C. Brander, K. Bishop, N. Mlotshwa, J. I. Mullins, H. Coovadia, T.**

REFERENCES

1. **Anonymous** 2009. Los Alamos HIV Sequence Database. **2010:**.

2. **Abrahams, M. -., J. A. Anderson, E. E. Giorgi, C. Seoighe, K. Mlisana, L. -. Ping, G. S. Athreya, F. K. Treurnicht, B. F. Keele, N. Wood, J. F. Salazar-Gonzalez, T. Bhattacharya, H. Chu, I. Hoffman, S. Galvin, C. Mapanje, P. Kazembe, R. Thebus, S. Fiscus, W. Hide, M. S. Cohen, S. A. Karim, B. F. Haynes, G. M. Shaw, B. H. Hahn, B. T. Korber, R. Swanstrom, C. Williamson, and for the CAPRISA Acute Infection Study Team and the Center for HIV-AIDS Vaccine Immunology Consortium,.** 2009. Quantitating the Multiplicity of Infection with Human Immunodeficiency Virus Type 1 Subtype C Reveals a Non-Poisson Distribution of Transmitted Variants. J. Virol. **83:**3556-3567. doi: 10.1128/JVI.02132-08.

3. **Allen, T. M., M. Altfeld, S. C. Geer, E. T. Kalife, C. Moore, K. M. O'Sullivan, I. DeSouza, M. E. Feeney, R. L. Eldridge, E. L. Maier, D. E. Kaufmann, M. P. Lahaie, L. Reyor, G. Tanzi, M. N. Johnston, C. Brander, R. Draenert, J. K. Rockstroh, H. Jessen, E. S. Rosenberg, S. A. Mallal, and B. D. Walker.** 2005. Selective Escape from CD8+ T-Cell Responses Represents a Major Driving Force of Human Immunodeficiency Virus Type 1 (HIV-1) Sequence Diversity and Reveals Constraints on HIV-1 Evolution. J. Virol. **79:**13239-13249. doi: 10.1128/JVI.79.21.13239-13249.2005.

4. **Altfeld, M., T. M. Allen, X. G. Yu, M. N. Johnston, D. Agrawal, B. T. Korber, D. C. Montefiori, D. O'Connor, B. T. Davis, P. K. Lee, E. L. Maier, J. Harlow, P. J. Goulder, C. Brander, E. S. Rosenberg, and B. D. Walker.** 2002. HIV-1 superinfection despite broad CD8+ T-cell responses containing replication of the primary virus. Nature. **420:**434-439.

6. **Bailey, J. R., T. M. Williams, R. F. Siliciano, and J. N. Blankson.** 2006. Maintenance of viral suppression in HIV-1–infected HLA-B*57+ elite suppressors despite CTL escape mutations. The Journal of Experimental Medicine. **203:**1357-1369. doi: 10.1084/jem.20052319.

7. **Barouch, D. H., J. Kunstman, J. Glowczwskie, K. J. Kunstman, M. A. Egan, F. W. Peyerl, S. Santra, M. J. Kuroda, J. E. Schmitz, K. Beaudry, G. R. Krivulka, M. A. Lifton, D. A. Gorgone, S. M. Wolinsky, and N. L. Letvin.** 2003. Viral Escape from Dominant Simian Immunodeficiency Virus Epitope-Specific Cytotoxic T Lymphocytes in DNA-Vaccinated Rhesus Monkeys. J. Virol. **77:**7367-7375. doi: 10.1128/JVI.77.13.7367-7375.2003.

8. **Barouch, D. H., J. Kunstman, M. J. Kuroda, J. Schmitz, S. Santra, F. W. Peyerl, G. R. Krivulka, K. Beaudry, M. A. Lifton, D. A. Gorgone, D. C. Montefiori, M. G. Lewis, S. M. Wolinsky, and N. L. Letvin.** 2002. Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. Nature. **415:**335.

9. **Betts, M., J. Krowka, C. Santamaria, K. Balsamo, F. Gao, G. Mulundu, C. Luo, N. N'Gandu, H. Sheppard, B. Hahn, S. Allen, and J. Frelinger.** 1997. Cross-clade human immunodeficiency virus (HIV)-specific cytotoxic T- lymphocyte responses in HIV-infected Zambians. J. Virol. **71:**8908-8911.

10. **Blackard, J., D. Cohen, and K. Mayer.** 2002. Human Immunodeficiency Virus Superinfection and Recombination: Current State of Knowledge and Potential Clinical Consequences. Clinical Infectious Diseases. **34:**1108-1114.

11. **Blish, C. A., O. C. Dogan, N. R. Derby, M. Nguyen, B. Chohan, B. A. Richardson, and J. Overbaugh.** 2008. Human Immunodeficiency Virus Type 1 Superinfection Occurs despite Relatively Robust Neutralizing Antibody Responses. J. Virol. **82:**12094-12103. doi: 10.1128/JVI.01730-08.

12. **Bordería, A. V., R. Lorenzo-Redondo, M. Pernas, C. Casado, T. Alvaro, E. Domingo, and C. Lopez-Galindez.** 2010. Initial Fitness Recovery of HIV-1 Is Associated with Quasispecies Heterogeneity and Can Occur without Modifications in the Consensus Sequence. Plos One. **5:**15 November 2010. doi: 10.1371/journal.pone.0010319.

13. **Borrow, P., H. Lewicki, X. Wei, M. S. Horwitz, N. Peffer, H. Meyers, J. A. Nelson, J. E. Gairin, B. H. Hahn, and Oldstone, Michael. B. A. and Shaw, George. M.** 1997. **Antiviral pressure exerted by HIV-l-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus**. Nature Medicine. **3:**205-211. doi: 10.1038/nm0297-205.

14. **Borrow, P., H. Lewicki, B. H. Hahn, G. M. Shaw, and M. B. Oldstone.** 1994. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. J. Virol. **68:**6103-6110.

15. **Boutwell, C. L., C. F. Rowley, and M. Essex.** 2009. **Reduced Viral Replication Capacity of Human Immunodeficiency Virus Type 1 Subtype C Caused by Cytotoxic-T-Lymphocyte Escape Mutations in HLA-B57 Epitopes of Capsid Protein**. J. Virol. **83:**2460-2468. doi: 10.1128/JVI.01970-08.

16. **Braibant, M., J. Xie, A. Samri, H. Agut, B. Autran, and F. Barin.** 2010. Disease progression due to dual infection in an HLA-B57-positive asymptomatic long-term nonprogressor infected with a nef-defective HIV-1 strain. Virology. **405:**81-92. doi: DOI: 10.1016/j.virol.2010.05.026.

17. **Brenner, B., J. Routy, Y. Quan, D. Moisi, M. Oliveira, D. Turner, M. A. Wainberg, and co-investigators of the Quebec Primary Infection Study.** 2004. **Persistence of multidrug-resistant HIV-1 in primary infection leading to superinfection**. AIDS. **18:**1653-1660. doi: 10.1097/01.aids.0000131377.28694.04.

18. **Brockman, M. A., A. Schneidewind, M. Lahaie, A. Schmidt, T. Miura, I. DeSouza, F. Ryvkin, C. A. Derdeyn, S. Allen, E. Hunter, J. Mulenga, P. A. Goepfert, B. D. Walker, and T. M. Allen.** 2007. Escape and Compensation from Early HLA-B57-Mediated Cytotoxic T-Lymphocyte Pressure on Human Immunodeficiency Virus Type 1 Gag Alter Capsid Interactions with Cyclophilin A. J. Virol. **81:**12608-12618. doi: 10.1128/JVI.01369-07.

19. **Buchbinder, S. P., D. V. Mehrotra, A. Duerr, D. W. Fitzgerald, R. Mogg, D. Li, P. B. Gilbert, J. R. Lama, M. Marmor, C. del Rio, M. J. McElrath, D. R. Casimiro, K. M. Gottesdiener, J. A. Chodakewitz, L. Corey, Corey, Lawrence. and Robertson, Michael. N., and the Step Study Protocol Team.** 2008. Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. Lancet. **372:**1881-1893. doi: 10.1016/S0140-6736(08)61591-3.

20. **Bures, R., L. Morris, C. Williamson, G. Ramjee, M. Deers, S. A. Fiscus, and Abdool-Karim, S. and Montefiori, D. C.** 2002. Regional Clustering of Shared Neutralization Determinants on Primary Isolates of Clade C Human Immunodeficiency Virus Type 1 from South Africa. J. Virol. **76:**2233-2244.

21. **Campbell, M. S., G. S. Gottlieb, S. E. Hawes, D. C. Nickle, K. G. Wong, and et al.** 2009. HIV-1 Superinfection in the Antiretroviral Therapy Era: Are Seroconcordant Sexual Partners at Risk? Plos One. **4:**. doi: 10.1371/journal.pone.0005690.

22. **Cao, H., P. Kanki, J. Sankale, A. Dieng-Sarr, G. Mazzara, S. Kalams, B. Korber, S. Mboup, and B. Walker.** 1997. Cytotoxic T-lymphocyte cross-reactivity among different human immunodeficiency virus type 1 clades: implications for vaccine development. J. Virol. **71:**8615-8623.

23. **Cao, H., I. Mani, R. Vincent, R. Mugerwa, P. Mugyenyi, P. Kanki, J. Ellner, and B. D. Walker.** 2000. Cellular Immunity to Human Immunodeficiency Virus Type 1 (HIV-1) Clades: Relevance to HIV-1.. J. Infect. Dis. **182:**1350.

24. **Casado, C., M. Pernas, T. Alvaro, V. Sandonis, S. Garcia, C. Rodriguez, J. D. Romero, E. Grau, L. Ruiz, and C. Lopez-Galindez.** 2007. Coinfection and Superinfection in Patients with Long-Term, Nonprogressive HIV-1 Disease. J. Infect. Dis. **196:**895-899.

26. **Chakraborty, B., L. Valer, C. De Mendoza, V. Soriano, and M. E. and Quiñones-mateu.** 2004. Failure to Detect Human Immunodeficiency Virus Type 1 Superinfection in 28 HIV-Seroconcordant Individuals with High Risk of Reexposure to the Virus. AIDS Research and Human Retroviruses. **20:**1026. doi: 10.1089/aid.2004.20.1026.

27. **Charpentier, C., T. Nora, O. Tenaillon, F. Clavel, and A. J. Hance.** 2006. Extensive Recombination among Human Immunodeficiency Virus Type 1 Quasispecies Makes an Important Contribution to Viral Diversity in Individual Patients. J. Virol. **80:**2472-2482. doi: 10.1128/JVI.80.5.2472-2482.2006.

28. **Chohan, B., L. Lavreys, S. M. J. Rainwater, and J. Overbaugh.** 2005. Evidence for Frequent Reinfection with Human Immunodeficiency Virus Type 1 of a Different Subtype. J. Virol. **79:**10701-10708. doi: 10.1128/JVI.79.16.10701-10708.2005.

29. **Clerc, O., S. Colombo, S. Yerly, A. Telenti, and M. Cavassini.** 2010. HIV-1 elite controllers: Beware of super-infections. Journal of Clinical Virology. **47:**376-378. doi: DOI: 10.1016/j.jcv.2010.01.013.

30. **Cornelissen, M., S. Jurriaans, K. Kozaczynska, J. M. Prins, R. A. Hamidjaja, F. Zorgdrager, M. Bakker, N. Back, and d. K. van.** 2007. Routine HIV-1 genotyping as a tool to identify dual infections. AIDS. **21:**807-811.

31. **Crawford, H., J. G. Prado, A. Leslie, S. Hue, I. Honeyborne, S. Reddy, M. van der Stok, Z. Mncube, C. Brander, C. Rousseau, J. I. Mullins, R. Kaslow, P. Goepfert, S. Allen, E. Hunter, J. Mulenga, P. Kiepiela, B. D. Walker, and P. J. R. Goulder.** 2007. Compensatory Mutation Partially Restores Fitness and Delays Reversion of Escape Mutation within the Immunodominant HLA-B*5703-Restricted Gag Epitope in Chronic Human Immunodeficiency Virus Type 1 Infection. J. Virol. **81:**8346-8351. doi: 10.1128/JVI.00465-07.

32. **Delwart, E. L., E. W. Hanley, B. Herring, G. H. Learn Jr, F. Lyagoba, and Rodrigo, Allen. G. and Shankarappa, Raj.**
Heteroduplex Mobility Analysis. NIH AIDS Research & Reference Reagent Program. **Protocol Version 5:**.

33. **Fang, G., B. Weiser, C. Kuiken, S. M. Philpott, S. RowlandJones, F. Plummer, J. Kimani, B. Shi, R. Kaul, J. Bwayo, O. Anzala, and H. Burger.** 2004. Recombination following superinfection by HIV-1. AIDS. **18:**153-159.

34. **Fauci, A. S.** 1986. Current issues in developing a strategy for dealing with the acquired immunodeficiency syndrome. Proc. Natl. Acad. Sci. USA. **83:**9278-9283.

35. **Frahm, N., B. Baker, and C. Brander.** 2008. Identification and Optimal Definition of HIV-Derived Cytotoxic T Lymphocyte (CTL) Epitopes for the Study of CTL Escape, Functional Avidity and Viral Evolution. HIV Molecular Immunology. .

36. **Fultz, P. N., A. Srinivasan, C. R. Greene, D. Butler, R. B. Swenson, and H. M. McClure.** 1987. Superinfection of a chimpanzee with a second strain of human immunodeficiency virus. J. Virol. **61:**4026-4029.

37. **Fung, I. C., M. Gambhir, A. van Sighem, F. de Wolf, and G. P. Garnett.** 2010. Superinfection with a heterologous HIV strain per se does not lead to faster progression. Math. Biosci. **224:**1-9. doi: DOI: 10.1016/j.mbs.2009.11.007.

38. **Gifford, R. J., T. de Oliveira, A. Rambaut, O. G. Pybus, D. Dunn, A. M. Vandamme, P. Kellam, D. Pillay, and UK Collaborative Group on HIV Drug Resistance.** 2007. Phylogenetic surveillance of viral genetic diversity and the evolving molecular epidemiology of human immunodeficiency virus type 1. J. Virol. **81:**13050-13056. doi: 10.1128/JVI.00889-07.

39. **Goepfert, P. A., W. Lumm, P. Farmer, P. Matthews, A. Pendergast, J. M. Carlson, C. A. Derdeyn, J. Tang, R. A. Kaslow, A. Bansal, K. Yusim, D. Heckerman, J. Mulenga, S. Allen, and Goulder, Philip. J. R. and Hunter, Eric.** 2008. Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. Journal of Experimental Medicine. **205:**1009-1017. doi: 10.1084/jem.20072457.

40. **Gonzales, M. J., E. Delwart, S. Y. Rhee, R. Tsui, A. R. Zolopa, J. Taylor, and R. W. Shafer.** 2003. Lack of detectable human immunodeficiency virus type 1 superinfection during 1072 person-years of observation. J. Infect. Dis. **188:**397-405.

41. **Goonetilleke, N., M. K. P. Liu, J. F. Salazar-Gonzalez, G. Ferrari, E. Giorgi, V. V. Ganusov, B. F. Keele, G. H. Learn, E. L. Turnbull, M. G. Salazar, K. J. Weinhold, S. Moore, CHAVI Clinical Core B, N. Letvin, B. F. Haynes, M. S. Cohen, P. Hraber, T. Bhattacharya, P. Borrow, A. S. Perelson, B. H. Hahn, G. M. Shaw, and Korber, B. T. and McMichael, A. J.** 2009. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection Journal of Experimental Medicine. **206:**1253-1272. doi: 10.1084/jem.20090365.

42. **Gottlieb, G. S., D. C. Nickle, M. Jensen, K. Wong, J. Grobler, F. Li, S. Liu, C. Rademeyer, G. H. Learn, S. A. Karim, C. Williamson, L. Corey, J. Margolick, and J. I. Mullins.** 2004. Dual HIV-1 infection associated with rapid disease progression. Lancet. **363:**619-622.

44. **Gottlieb, G., D. Nickle, M. Jensen, K. Wong, R. Kaslow, J. Shepherd, J. Margolick, and J. Mullins.** 2007. HIV Type 1 Superinfection with a Dual-Tropic Virus and Rapid Progression to AIDS: A Case Report. Clinical Infectious Diseases. **45:**501-509.

45. **Goulder, P. J. R., R. E. Phillips, R. A. Colbert, S. McAdam, G. Ogg, M. A. Nowak, P. Giangrande, G. Luzzi, B. Morgan, A. Edwards, and McMichael, Andrew. J. and Rowland-Jones, Sarah.** 1997.
Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nature Medicine. **3:**212.

46. **Goulder, P. J. R., and D. I. Watkins.** 2004. HIV and SIV CTL escape: implications for vaccine design. Nature Reviews Immunology. **4:**630-640. doi: 10.1038/nri1417.

47. **Gray, E. S., P. L. Moore, I. A. Choge, J. M. Decker, F. Bibollet-Ruche, H. Li, N. Leseka, F. Treurnicht, K. Mlisana, G. M. Shaw, S. S. A. Karim, C. Williamson, L. Morris, and and the CAPRISA 002 Study Team.** 2007. **Neutralizing Antibody Responses in Acute Human Immunodeficiency Virus Type 1 Subtype C Infection**. J. Virol. **81:**6187-6196. doi: 10.1128/JVI.00239-07.

48. **Gray, E. S., N. Taylor, D. Wycuff, P. L. Moore, G. D. Tomaras, C. K. Wibmer, A. Puren, A. DeCamp, P. B. Gilbert, B. Wood, D. C. Montefiori, J. M. Binley, G. M. Shaw, B. F. Haynes, and ,Mascola, J. R. and Morris, L.** 2009. Antibody Specificities Associated with Neutralization Breadth in Plasma from Human Immunodeficiency Virus Type 1 Subtype C-Infected Blood Donors. J. Virol. **83:**8925-8937. doi: 10.1128/JVI.00758-09.

49. **Grobler, J., C. M. Gray, C. Rademeyer, C. Seoighe, G. Ramjee, S. A. Karim, L. Morris, and C. Williamson.** 2004. Incidence of HIV-1 dual infection and its association with increased viral load set point in a cohort of HIV-1 subtype C-infected female sex workers. J. Infect. Dis. **190:**1355-1359.

51. **Hall, T. A.** 1999. BioEdit:a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. **41:**95-98.

52. **Herbinger, K., M. Gerhardt, P. Piyasirisilp, D. Mloka, M. A. Arroyo, O. Hoffmann, L. Maboko, D. L. Birx, D. Mmbando, and Mccutchan, F. E. and Hoelscher, M.** 2006. Frequency of HIV Type 1 Dual Infection and HIV Diversity: Analysis of Low- and High-Risk Populations in Mbeya Region, Tanzania. AIDS Research and Human Retroviruses. **22:**599. doi: 10.1089/aid.2006.22.599.

53. **Hoelscher, M., W. E. Dowling, E. Sanders-Buell, J. K. Carr, M. E. Harris, A. Thomschke, M. L. Robb, D. L. Birx, and F. E. McCutchan.** 2002. Detection of HIV-1 subtypes, recombinants, and dual infections in east Africa by a multi-region hybridization assay. AIDS. **16:**2055-2064.

54. **Hu, D. J., S. Subbarao, S. Vanichseni, P. A. Mock, A. Ramos, L. Nguyen, T. Chaowanachan, F. Griensven, K. Choopanya, T. D. Mastro, and J. W. Tappero.** 2005. Frequency of HIV-1 dual subtype infections, including intersubtype superinfections, among injection drug users in Bangkok, Thailand. AIDS. **19:**303-308.

55. **Janini, L. M., D. Pieniazek, J. M. Peralta, M. Schechter, A. Tanuri, A. C. P. Vicente, N. J. Pieniazek, C. Luo, M. L. Kalish, and Schochetman, Gerald. and Rayfield, Mark A..** 1996. Identification of single and dual infections with distinct subtypes of human immunodeficiency virus type 1 by using restriction fragment length polymorphism analysis. Virus Genes. **13:**69-81. doi: 10.1007/BF00576981.

56. **Jost, S., M. Bernard, L. Kaiser, S. Yerly, B. Hirschel, A. Samri, B. Autran, L. Goh, and L. Perrin.** 2002. A Patient with HIV-1 Superinfection. N. Engl. J. Med. **347:**731-736. doi: 10.1056/NEJMoa020263.

57. **Jurriaans, S., K. Kozaczynska, F. Zorgdrager, R. Steingrover, J. M. Prins, d. K. van, and M. Cornelissen.** 2007. A sudden rise in viral load is infrequently associated with HIV-1 superinfection. JAIDS. **in press:**.

58. **Kiepiela, P., K. Ngumbela, C. Thobakgale, D. Ramduth, I. Honeyborne, E. Moodley, S. Reddy, C. de Pierres, Z. Mncube, N. Mkhwanazi, K. Bishop, M. van der Stok, K. Nair, N. Khan, H. Crawford, R. Payne, A. Leslie, J. Prado, A. Prendergast, J. Frater, N. McCarthy, C. Brander, G. H. Learn, D. Nickle, C. Rousseau, H. Coovadia, J. I. Mullins, D. Heckerman, and Walker, Bruce. D. and Goulder, Philip.** 2006. **$CD8^+$ T-cell responses to different HIV proteins have discordant associations with viral load**. Nature Medicine. **13:**46-53. doi: 10.1038/nm1520.

59. **Kiepiela, P., A. J. Leslie, I. Honeyborne, D. Ramduth, C. Thobakgale, S. Chetty, P. Rathnavalu, C. Moore, K. J. Pfafferott, L. Hilton, P. Zimbwa, S. Moore, T. Allen, C. Brander, M. M. Addo, M. Altfeld, I. James, S. Mallal, M. Bunce, and L. D. Barber.** 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature. **432:**769. doi: 10.1038/nature03113.

60. **Klein, M. R., C. A. van Baalen, A. M. Holwerda, S. R. Kerkhof Garde, R. J. Bende, I. P. Keet, J. K. Eeftinck-Schattenkerk, A. D. Osterhaus, H. Schuitemaker, and F. Miedema.** 1995. Kinetics of Gag-specific cytotoxic T lymphocyte responses during the clinical course of HIV-1 infection: a longitudinal analysis of rapid progressors and long-term asymptomatics. The Journal of Experimental Medicine. **181:**1365-1372. doi: 10.1084/jem.181.4.1365.

61. **Koup, R. A., J. T. Safrit, Y. Cao, C. A. Andrews, G. McLeod, W. Borkowsky, C. Farthing, and D. D. Ho.** 1994. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. J. Virol. **68:**4650-4655.

62. **Kozaczynska, K., M. Cornelissen, P. Reiss, and Zorgdrager, F. and van der Kuyl, A.C.** 2007. HIV-1 sequence evolution in vivo after superinfection with three viral strains. Retrovirology. **4:**59. doi: 10.1186/1742-4690-4-59.

63. **Kuiken, C., T. Leitner, B. Foley, B. Hahn, P. Marx, F. McCutchan, and Wolinsky,Steven M. and Korber, Bette.** 2009. HIV Sequence Compendium 2009. Theoretical Biology and Biophysics, Los Alamos National Laboratory.

64. **Kulkarni, S. S., A. Lapedes, H. Tang, S. Gnanakaran, M. G. Daniels, M. Zhang, T. Bhattacharya, M. Li, V. R. Polonis, F. E. McCutchan, L. Morris, D. Ellenberger, and Butera, S. T.,Bollinger, R. C.,Korber, B. T.,Paranjape, R. S. and Montefiori, D. C.** 2009. Highly complex neutralization determinants on a monophyletic lineage of newly transmitted subtype C HIV-1 Env clones from India. Virology. **385:**505-520. doi: 10.1016/j.virol.2008.12.032.

65. **Leslie, A., P. C. Matthews, J. Listgarten, J. M. Carlson, C. Carl Kadie, T. Ndung'u, C. Brander, H. Coovadia, B. D. Walker, and Heckerman, David. and Goulder, Philip. J. R.** 2010. Additive Contribution of HLA Class I Alleles in the Immune Control of HIV-1 Infection. Journal of Virology. **84:**9879-9888. doi: 10.1128/JVI.00320-10.

66. **Li, M., J. F. Salazar-Gonzalez, C. A. Derdeyn, L. Morris, C. Williamson, J. E. Robinson, J. M. Decker, Y. Li, M. G. Salazar, V. R. Polonis, K. Mlisana, S. A. Karim, K. Hong, K. M. Greene, M. Bilska, J. Zhou, S. Allen, E. Chomba, J. Mulenga, C. Vwalika, F. Gao, M. Zhang, B. T. M. Korber, E. Hunter, B. H. Hahn, and D. C. Montefiori.** 2006. **Genetic and Neutralization Properties of Subtype C Human Immunodeficiency Virus Type 1 Molecular env Clones from Acute and Early Heterosexually Acquired Infections in Southern Africa**. J. Virol. **80:**11776-11790. doi: 10.1128/JVI.01730-06.

67. **Manigart, O., V. Courgnaud, O. Sanou, D. Valéa, N. Nagot, N. Meda, E. Delaporte, M. Peeters, and P. Van de Perre.** 2004. HIV-1 superinfections in a cohort of commercial sex workers in Burkina Faso as assessed by an autologous heteroduplex mobility procedure. AIDS. **18:**1645-1651.

68. **Martinez-Picado, J., J. G. Prado, E. E. Fry, K. Pfafferott, A. Leslie, S. Chetty, C. Thobakgale, I. Honeyborne, H. Crawford, P. Matthews, T. Pillay, C. Rousseau, J. I. Mullins, C. Brander, B. D. Walker, D. I. Stuart, and Kiepiela, Photini. and Goulder, Philip.** 2005. Fitness Cost of Escape Mutations in p24 Gag in Association with Control of Human Immunodeficiency Virus Type 1. J. Virol. **80:**3617-3623. doi: 10.1128/JVI.80.7.3617–3623.2006.

69. **Matthews, P. C., A. Prendergast, A. Leslie, H. Crawford, R. Payne, C. Rousseau, M. Rolland, I. Honeyborne, J. Carlson, C. Kadie, C. Brander, K. Bishop, N. Mlotshwa, J. I. Mullins, H. Coovadia, T. Ndung'u, B. D. Walker, and Heckerman, David. and Goulder, Philip. J. R.** 2008. Central Role of Reverting Mutations in HLA Associations with Human Immunodeficiency Virus Set Point. J. Virol. **82:**8548-8559. doi: 10.1128/JVI.00580-08.

71. **McCutchan, F. E., M. Hoelscher, S. Tovanabutra, S. Piyasirisilp, E. Sanders-Buell, G. Ramos, L. Jagodzinski, V. Polonis, L. Maboko, D. Mmbando, O. Hoffmann, G. Riedner, F. von Sonnenburg, M. Robb, and D. L. Birx.** 2005. 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. **79:**11693-11704. doi: 10.1128/JVI.79.18.11693-11704.2005.

72. **Montefiori, D. C., L. Morris, G. Ferrari, and J. R. Mascola.** 2007. Neutralizing and other antiviral antibodies in HIV-1 infection and vaccination. Current Opinion in HIV and AIDS. **2:**169-176. doi: 10.1097/COH.0b013e3280ef691e.

73. **Moore, R. D., and R. E. Chaisson.** 1996. Natural History of Opportunistic Disease in an HIV-Infected Urban Clinical Cohort. Annals of Internal Medicine. **124:**633-642. doi: 10.1059/0003-4819-124-7-199604010-00003.

74. **Motomura, K., and Chen, J. and Hu, W.** 2008. Genetic Recombination between Human Immunodeficiency Virus Type 1 (HIV-1) and HIV-2, Two Distinct Human Lentiviruses. J. Virol. **82:**1923-1933. doi: 10.1128/JVI.01937-07.

75. **Ngumbela, K. C., C. L. Day, Z. Mncube, K. Nair, D. Ramduth, C. Thobakgale, E. Moodley, S. Reddy, C. De Pierres, N. Mkhwanazi, K. Bishop, D. S. Van, N. Ismail, I. Honeyborne, H. Crawford, D. G. Kavanagh, C. Rousseau, D. Nickle, J. Mullins, D. Heckerman, B. Korber, H. Coovadia, P. Kiepiela, P. J. R. Goulder, and B. D. Walker.** 2008. Targeting of a CD8 T cell env epitope presented by HLA-B*5802 is associated with markers of HIV disease progression and lack of selection pressure. AIDS Res. Hum. Retroviruses. **24:**72(11).

76. **Osmanov, S., C. Pattou, N. Walker, B. Schwardländer, J. Esparza, and WHO-UNAIDS Network for HIV Isolation and Characterization.** 2002. Estimated Global Distribution and Regional Spread of HIV-1 Genetic Subtypes in the Year 2000. JAIDS Journal of Acquired Immune Deficiency Syndromes. **29:**184-190.

77. **Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho.** 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science. **271:**p1582(5).

78. **Pereyra, F., M. Addo, D. Kaufmann, Y. Liu, T. Miura, A. Rathod, B. Baker, A. Trocha, R. Rosenberg, E. Mackey, P. Ueda, Z. Lu, D. Cohen, T. Wrin, C. Petropoulos, E. Rosenberg, and B. Walker.** 2008. Genetic and Immunologic Heterogeneity among Persons Who Control HIV Infection in the Absence of Therapy. J. Infect. Dis. **197:**563-571.

79. **Pernas, M., C. Casado, R. Fuentes, M. J. PerezElias, and C. LopezGalindez.** 2006. A Dual Superinfection and Recombination Within HIV-1 Subtype B 12 Years After Primoinfection. JAIDS J. Acquired Immune Defic. Syndromes. **42:**12-18. doi: 10.1097/01.qai.0000214810.65292.73.

80. **Piantadosi, A., B. Chohan, V. Chohan, and McClelland, R. S. and Overbaugh, J.** 2007. Chronic HIV-1 Infection Frequently Fails to Protect against Superinfection. PLoS Pathog. **3 (11):**1745. doi: 10.1371/journal.ppat.0030177.

81. **Piantadosi, A., D. Panteleeff, C. A. Blish, J. M. Baeten, W. Jaoko, R. S. McClelland, and J. Overbaugh.** 2009. **Breadth of Neutralizing Antibody Response to Human Immunodeficiency Virus Type 1 Is Affected by Factors Early in Infection but Does Not Influence Disease Progression**. J. Virol. **83:**10269-10274. doi: 10.1128/JVI.01149-09.

82. **Pieniazek, D., L. M. Janini, A. Ramos, A. Tanuri, M. Schechter, J. M. Peralta, A. C. P. Vicente, N. J. Pieniazek, and Schochetman, Gerald. and Rayfield, Mark. A.** 1995. HIV-1 Patients May Harbor Viruses of Different Phylogenetic Subtypes: Implications for the Evolution of the HIV/AIDS Pandemic. Emerg Infect Dis. **1:**86-88.

83. **Plata, F., B. Autran, L. P. Martins, S. Wain-Hobson, M. Raphaël, C. Mayaud, M. Denis, and Guillon, Jean-Marc. and Debré, Patrice.** 1987. **AIDS virus-specific cytotoxic T lymphocytes in lung disorders**. Nature. **328:**348-351. doi: 10.1038/328348a0.

84. **Rachinger, A., I. G. Stolte, van de Ven, Tom Derks., J. A. Burger, M. Prins, and Schuitemaker, Hanneke. and van 't Wout, Angelique B.** 2010. **Absence of HIV-1 Superinfection 1 Year after Infection between 1985 and 1997 Coincides with a Reduction in Sexual Risk Behavior in the Seroincident Amsterdam Cohort of Homosexual Men**. Clinical Infectious Diseases. **50:**1309-1315. doi: 1058-4838/2010/5009-0017.

85. **Ramirez, B. C., E. Simon-Loriere, R. Galetto, and M. Negroni.** 2008. Implications of recombination for HIV diversity. Virus Res. **134:**64-73. doi: DOI: 10.1016/j.virusres.2008.01.007.

86. **Ramos, A., D. J. Hu, L. Nguyen, K. Phan, S. Vanichseni, N. Promadej, K. Choopanya, M. Callahan, N. Young L., J. .McNicholl, T. D. Mastro, F. Thomas M., and S. Shambavi.** 2002. **Intersubtype Human Immunodeficiency Virus Type 1 Superinfection following Seroconversion to Primary Infection in Two Injection Drug Users**. J. Virol. **76:**7444-7452. doi: 10.1128/JVI.76.15.7444-7452.2002.

87. **Richman, D. D., T. Wrin, S. J. Little, and C. J. Petropoulos.** 2003. Rapid Evolution of the Neutralizing Antibody Response to HIV Type 1 Infection. Proc. Natl. Acad. Sci. U. S. A. **100:**pp. 4144-4149.

88. **Robertson, D. L., J. P. Anderson, J. A. Bradac, J. K. Carr, B. Foley, R. K. Funkhouser, F. Gao, B. H. Hahn, C. Kuiken, G. H. Learn, T. Leitner, F. McCutchan, S. Osmanov, M. Peeters, D. Pieniazek, M. L. Kalish, M. Salminen, P. Sharp, and Wolinsky, S. and Korber, B.** 2000. HIV-1 Nomenclature Proposal. Science. **288:**55.

89. **Rousseau, C., M. G. Daniels, J. Carlson, C. Kadie, H. Crawford, A. Prendergast, P. Matthews, R. Payne, M. Rolland, D. N. Raugi, B. S. Maust, G. H. Learn, D. C. Nickle, H. Coovadia, T. Ndung'u, N. Frahm, C. Brander, B. D. Walker, P. J. Goulder, T. Bhattacharya, D. Heckerman, and Korber,B. T. and Mullins, James. I.** 2008. HLA Class I-Driven Evolution of Human Immunodeficiency Virus Type 1 Subtype C Proteome: Immune Escape and Viral Load. J Virol. **82:**6434-6446. doi: 10.1128/JVI.02455-07.

90. **Rousseau, C. M., G. H. Learn, T. Bhattacharya, D. C. Nickle, D. Heckerman, S. Chetty, C. Brander, P. J. R. Goulder, B. D. Walker, P. Kiepiela, B. T. Korber, and J. I. Mullins.** 2007. Extensive Intrasubtype Recombination in South African Human Immunodeficiency Virus Type 1 Subtype C Infections. J. Virol. **81:**4492-4500. doi: 10.1128/JVI.02050-06.

91. **Saez-Cirion, A., M. Sinet, S. Y. Shin, A. Urrutia, P. Versmisse, C. Lacabaratz, F. Boufassa, V. Avettand-Fenoel, C. Rouzioux, J. Delfraissy, F. Barre-Sinoussi, O. Lambotte, A. Venet, and Pancino, Gianfranco. for the ANRS EP36 HIV Controllers Study Group.** 2009. **Heterogeneity in HIV Suppression by CD8 T Cells from HIV Controllers: Association with Gag-Specific CD8 T Cell Responses**. Journal of Immunology. **182:**7828-7837. doi: 10.4049/jimmunol.0803928.

92. **Sagar, M., L. Lavreys, J. M. Baeten, B. A. Richardson, K. Mandaliya, B. H. Chohan, J. K. Kreiss, and J. Overbaugh.** 2003. **Infection with Multiple Human Immunodeficiency Virus Type 1 Variants Is Associated with Faster Disease Progression**. J. Virol. **77:**12921-12926. doi: 10.1128/JVI.77.23.12921-12926.2003.

93. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. **4:**406-425.

94. **Sewell, A. K.** 2000. Cytotoxic T lymphocyte responses to human immunodeficiency virus: control and escape. Stem Cells. **18:**230.

95. **Shankarappa, R., J. B. Margolick, S. J. Gange, A. G. Rodrigo, D. Upchurch, H. Farzadegan, P. Gupta, C. R. Rinaldo, G. H. Learn, X. He, X. Huang, and J. I. Mullins.** 1999. **Consistent Viral Evolutionary Changes Associated with the Progression of Human Immunodeficiency Virus Type 1 Infection**. J. Virol. **73:**10489-10502.

96. **Shibata, R., C. Siemon, M. W. Cho, L. O. Arthur, S. M. J. Nigida, T. Matthews, L. A. Sawyer, A. Schultz, K. K. Murthy, Z. Israel, A. Javadian, P. Frost, R. C. Kennedy, and Lane, H. Clifford. and Martin, Malcolm. A.** 1996. Resistance of Previously Infected Chimpanzees to Successive Challenges with a Heterologous Intraclade B Strain of Human Immunodeficiency Virus Type 1. J. Virol. **70:**4361-4369. doi: 0022-538X/96/.

97. **Sidat, M. M., A. M. Mijch, S. R. Lewin, J. F. Hoy, and Hocking, J. and Fairley, C. K.** 2008. Incidence of putative HIV superinfection and sexual practices among HIV-infected men who have sex with men. Sexual Health. **5:**61-67.

98. **Simon-Lorier, E., R. Galetto, M. Hamoudi, J. Archer, P. Lefeuvre, D. P. Martin, D. L. Robertson, and M. Negroni.** 2009. Molecular mechanisms of recombination restriction in the envelope gene of the human immunodeficiency virus. PLoS Pathog. **5:**. doi: 10.1371/journal.ppat.100D418.

99. **Smith, D. M., J. K. Wong, G. K. Hightower, C. C. Ignacio, K. K. Koelsch, E. S. Daar, D. D. Richman, and S. J. Little.** 2004. Incidence of HIV superinfection following primary infection. JAMA. **292:**1177-1178.

100. **Smith, D. M., M. C. Strain, S. D. W. Frost, S. K. Pillai, J. K. Wong, T. Wrin, Y. Liu, C. J. Petropolous, E. S. Daar, S. J. Little, and D. D. Richman.** 2006. Lack of neutralizing antibody response to HIV-1 predisposes to superinfection. Virology. **355:**1-5. doi: DOI: 10.1016/j.virol.2006.08.009.

101. **Smith, D. M., J. K. Wong, G. K. Hightower, C. C. Ignacio, K. K. Koelsch, C. J. Petropoulos, D. D. Richman, and S. J. Little.** 2005. HIV drug resistance acquired through superinfection. AIDS. **19:**1251-1256.

102. **Streeck, H., B. Li, A. F. Y. Poon, A. Schneidewind, A. A. Gladden, K. A. Power, D. Daskalakis, S. Bazner, R. Zuniga, C. Brander, E. S. Rosenberg, S. D. W. Frost, and Altfeld, M. and Allen, T. M.** 2008. Immune-driven recombination and loss of control after HIV superinfection Journal of Experimental Medicine. **205:**1789-1796. doi: 10.1084/jem.20080281.

103. **Takebe, Y., and Kusagawa, Shigeru. and Motomura, Kazushi. †.** 2004. Molecular epidemiology of HIV: Tracking AIDS pandemic. Pediatrics International. **46:**236-244.

104. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Molecular Biology and Evolution. **24:**1596-1599. doi: 10.1093/molbev/msm092.

105. **Taylor, J. E., and B. T. Korber.** 2005. HIV-1 intra-subtype superinfection rates: estimates using a structured coalescent with recombination. Infect Genet Evol. **5:**85-95.

106. **Tebit, D. M., I. Nankya, and Arts, Eric. J. and Gao, Yong.** 2007. **HIV Diversity, Recombination and Disease Progression: How Does Fitness "Fit" Into the Puzzle?**. AIDS Reviews. **9:**75-87.

107. **Tee, K. K., O. G. Pybus, J. Parker, K. P. Ng, A. Kamarulzaman, and Y. Takebe.** 2009. Estimating the date of origin of an HIV-1 circulating recombinant form. Virology. **387:**229-234. doi: DOI: 10.1016/j.virol.2009.02.020.

108. **Templeton, A. R., M. G. Kramer, J. Jarvis, J. Kowalski, S. Gange, M. F. Schneider, Q. Shao, G. W. Zhang, M. Yeh, H. Tsai, and Zhang, Hong. and Markham, Richard B.** 2009. **Multiple-infection and recombination in HIV-1 within a longitudinal cohort of women**. Retrovirology. **6:**. doi: 10.1186/1742-4690-6-54.

109. **Tomaras, G. D., N. L. Yates, P. Liu, L. Qin, G. G. Fouda, L. L. Chavez, A. C. Decamp, R. J. Parks, V. C. Ashley, J. T. Lucas, M. Cohen, J. Eron, C. B. Hicks, H. Liao, S. G. Self, G. Landucci, D. N. Forthal, K. J. Weinhold, B. F. Keele, B. H. Hahn, M. L. Greenberg, L. Morris, S. S. A. Karim, W. A. Blattner, D. C. Montefiori, G. M. Shaw, A. S. Perelson, and B. F. Haynes.** 2008. **Initial B-Cell Responses to Transmitted Human Immunodeficiency Virus Type 1: Virion-Binding Immunoglobulin M (IgM) and IgG Antibodies Followed by Plasma Anti-gp41 Antibodies with Ineffective Control of Initial Viremia** . J. Virol. **82:**12449-12463.

110. **Travers, K., S. Mboup, R. Marlink, A. Gueye-Nidaye, T. Siby, I. Thior, I. Traore, A. Dieng-Sarr, J. Sankale, C. Mullins, and a. et.** 1995. Natural protection against HIV-1 infection provided by HIV-2. Science. **268:**1612-1615. doi: 10.1126/science.7539936.

111. **Troyer, R. M., K. R. Collins, A. Abraha, E. Fraundorf, D. M. Moore, R. W. Krizan, Z. Toossi, R. L. Colebunders, M. A. Jensen, J. I. Mullins, G. Vanham, and E. J. Arts.** 2005. **Changes in Human Immunodeficiency Virus Type 1 Fitness and Genetic Diversity during Disease Progression**. J. Virol. **79:**9006-9018. doi: 10.1128/JVI.79.14.9006-9018.2005.

112. **Tsui, R., B. L. Herring, J. D. Barbour, R. M. Grant, P. Bacchetti, A. Kral, B. R. Edlin, and E. L. Delwart.** 2004. Human immunodeficiency virus type 1 superinfection was not detected following 215 years of injection drug user exposure. J. Virol. **78:**94-103.

114. **UNAIDS.** 2009. AIDS epidemic update : November 2009. **2010:**100. doi: UNAIDS/09.36E / JC1700E. http://www.unaids.org.

115. **van der Kuyl, A. C., K. Kozaczynska, K. K. Ariën, Y. Gali, V. R. Balázs, S. Dekker, F. Zorgdrager, G. Vanham, B. Berkhout, and M. Cornelissen.** 2010. Analysis of infectious virus clones from two HIV-1 superinfection cases suggests that the primary strains have lower fitness. Retrovirology. **7:**60.

116. **van der Loeff, Maarten. F. Schim., P. Aaby, K. Aryioshi, T. Vincent, A. A. Awasana, C. Da Costa, L. Pembrey, F. Dias, E. Harding, H. A. Weiss, and H. C. Whittle.** 2001. HIV-2 does not protect against HIV-1 infection in a rural community in Guinea-Bissau. AIDS. **15:**2303-2310.

117. **van Loggerenberg, F., K. Mlisana, C. Williamson, S. C. Auld, L. Morris, C. M. Gray, Q. A. Karim, A. Grobler, N. Barnabas, I. Iriogbe, S. S. Karim, and for the CAPRISA 002 Acute Infection Study Team.** 2008. **Establishing a Cohort at High Risk of HIV Infection in South Africa: Challenges and Experiences of the CAPRISA 002 Acute Infection Study** . Plos One. **3:**. doi: 10.1371/journal.pone.0001954.

118. **van, d. K., and M. Cornelissen.** 2007. Identifying HIV-1 dual infections. Retrovirology. **4:**67.

119. **van, d. K., K. Kozaczynska, d. B. Van, F. Zorgdrager, N. Back, S. Jurriaans, B. Berkhout, P. Reiss, and M. Cornelissen.** 2005. Triple HIV-1 infection. N. Engl. J. Med. **352:**2557-2559.

120. **van, d. K., F. Zorgdrager, S. Jurriaans, N. . Back, J. Prins, K. Brinkman, A. van Eeden, M. Bakker, and M. Cornelissen.** 2009. Incidence of Human Immunodeficiency Virus Type 1 Dual Infections in Amsterdam, The Netherlands, during 2003–2007. Clinical Infectious Diseases. **48:**973-978.

121. **Walker, B. D., S. Chakrabarti, B. Moss, T. J. Paradis, T. Flynn, A. G. Durno, R. S. Blumberg, J. C. Kaplan, and Hirsch, Martin. S. and Schooley, Robert. T.** 1987. **HIV-specific cytotoxic T lymphocytes in seropositive individuals**. Nature. **328:**345-348. doi: 10.1038/328345a0.

122. **Wei, X., J. M. Decker, S. Wang, H. Hui, J. C. Kappes, X. Wu, J. F. Salazar-Gonzalez, M. G. Salazar, J. M. Kilby, M. S. Saag, N. L. Komarova, M. A. Nowak, B. H. Hahn, P. D. Kwong, and G. M. Shaw.** 2003. Antibody neutralization and escape by HIV-1. Nature. **422:**307.

123. **Weinfurter, Jason T., May, Gemma E., Soma, Taeko, Hessell, Ann J., Leon, Enrique J., Mac Nair, Caitlin E., Piaskowski, Shari M., Weisgrau, Kim, Furlott, Jessica, Maness, Nicholas J., Reed, Jason, Wilson, Nancy A., Rakasz, Eva G., Burton, Dennis R., Friedrich, Thomas C.** 2010. **Macaque long-term nonprogressors resist superinfection with multiple CD8+ T cell escape variants of simian immunodeficiency virus**. J. Virol. . doi: 10.1128/JVI.01025-10.

124. **Yang, O. O., E. S. Daar, B. D. Jamieson, A. Balamurugan, D. M. Smith, J. A. Pitt, C. J. Petropoulos, D. D. Richman, S. J. Little, and A. J. L. Brown.** 2005. **Human Immunodeficiency Virus Type 1 Clade B Superinfection: Evidence for Differential Immune Containment of Distinct Clade B Strains**. J. Virol. **79:**860-868. doi: 10.1128/JVI.79.2.860-868.2005.

126. **Yeh, W. W., P. Jaru-ampornpan, D. Nevidomskyte, M. Asmal, S. S. Rao, A. P. Buzby, D. C. Montefiori, and Korber, Bette. T. and Letvin, Norman. L.** 2009. Partial Protection of Simian Immunodeficiency Virus (SIV)-Infected Rhesus Monkeys against Superinfection with a Heterologous SIV Isolate. J. Virol. **83:**2686-2696. doi: 0022-538X/09/.

127. **Yerly, S., S. Jost, M. Monnat, A. Telenti, M. Cavassini, J. Chave, L. Kaiser, P. Burgisser, and Perrin, Luc. and the Swiss HIV Cohort Study.** 2004. HIV-1 co/super-infection in intravenous drug users. AIDS. **18:**1413-1421. doi: 10.1097/01.aids.0000131330.28762.0c.

129. **Zhuang, J., A. E. Jetzt, G. Sun, H. Yu, G. Klarmann, Y. Ron, B. D. Preston, and J. P. Dougherty.** 2002. Human Immunodeficiency Virus Type 1 Recombination: Rate, Fidelity, and Putative Hot Spots. J. Virol. **76:**11273-11282. doi: 10.1128/JVI.76.22.11273-11282.2002.