ADAPTIVE CHANGES IN HIV-1 SUBTYPE C PROTEINS DURING EARLY INFECTION AND THEIR EFFECT ON DISEASE PROGRESSION

FLORETTE KATHLEEN TREURNICHT

This thesis is presented in fulfillment of the requirements for the degree of Doctor of Philosophy in the Division of Medical Virology, Department of Clinical Laboratory Sciences in the Faculty of Health Sciences and the Institute of Infectious Diseases and Molecular Medicine, University of Cape Town

STUDY LEADER

Prof. Carolyn Williamson

JUNE 2010

CONTENTS

ABSTRACT

High HIV viral diversity, even within a particular subtype, makes vaccine design strategies an enormously challenging task. Furthermore, host responses that define correlates of protection are poorly understood. Through the characterization of full-length viral genomes from individuals recently infected with subtype C, this thesis investigates the potential effect of viral diversity on vaccine effectiveness. In addition, through characterizing the transmitted virus and its early evolution we shed light on early events associated with control of viral replication.

HIV-1 subtype C accounts for 50% of all HIV-1 infections world-wide and for approximately 95% of infections in South Africa. In this study we characterized samples from the CAPRISA acute infection study which recruits individuals within 3 months of infection and monitors their disease progression. Characterisation of near full-length subtype C genomes from 23 individuals with primary infection showed that there was no evidence of a primary infection transmission network, however a comprehensive analysis of all sequences available from South Africa showed for the first time a number of supported clusters indicative of epidemiological linkage. We determined the potential of candidate vaccines tested in South Africa (Merck/Ad5, SAAVI-C and the mosaic candidate immunogens) to induce cytotoxic T cell responses against primary infection strains using a tool which determines the epitope coverage. As expected, the subtype C consensus-based SAAVI immunogen provided 4 – 20% better epitope coverage than the subtype B derived Merck immunogen. However the recombination-derived mosaic immunogens were markedly better compared to the other vaccine candidates providing 20-30% increased coverage.

It is unresolved whether recently transmitted HIV has genetic features that specifically favour transmissibility. To identify potential "transmission signatures", we compared 20 full-length HIV-1 subtype C genomes from primary infections, with 66 sampled from ethnically and geographically matched individuals with chronic infections. Controlling for recombination and phylogenetic relatedness, we identified 39 sites at which amino acid frequency spectra differed significantly between the groups. These

sites were predominantly located within Env, Pol and Gag (14/39, 9/39 and 6/39 respectively) and were significantly clustered (33/39) within known immunoreactive peptides. Within 6 months of infection we detected reversion-to-consensus mutations at 14 of the sites and potential CTL escape mutations at seven sites. Here we provide evidence that frequent reversion mutations probably allows the virus to recover replicative fitness which, together with immune escape driven by the HLA alleles of the new hosts, differentiate sequences from chronic infections from those sampled shortly after transmission.

Finally, we investigated the role of the adaptive immune responses on viral evolution during early infection. For 11 participants we had at least three full-length genomes derived longitudinally within the first six months of infection. We showed that early CTL escape across the proteome is not associated disease progression. However, viral diversification ($p = 0.0239$), reversion ($p = 0.0084$) and APOBEC-induced ($p = 0.0011$) mutations were significantly associated with lower CD4+ T cell counts. A more detailed analysis of a slow and rapid progressor showed that viruses from the rapid progressor experienced more positive selective pressure, and more frequent early CTL escape compared to the slow progressor. In the rapid progressor, escape was associated with a complex or shuffling escape patterns in 50% of epitopes.

In conclusion, this study elucidated characteristics of subtype C viruses from the primary infection phase and the role of early viral evolution and adaptive immune responses in the initial viral control. We show that genetic features of viruses in early infection are defined by HLA-associated immune adaptation and that regaining of replicative fitness through reversion to wild-type amino acids is the key driver of early viral evolution. We also showed that the candidate SAAVI-C immunogen are predicted to provide better CTL epitope coverage against subtype C infections than the subtype-mismatched subtype B Merck/Ad5 immunogen. However, the mosaic immunogens in preclinical development is superior in predicted epitope coverage to both the aforementioned vaccine candidates.

This study has contributed to the following publications:

(i) *Adaptive changes in HIV-1 subtype C proteins during early infection are driven by changes in HLA-associated immune pressure.* **F.K. Treurnicht**, C. Seoighe, D.P. Martin, N. Wood, M-R. Abrahams, D. de Assis Rosa, H. Bredell, Z. Woodman, W. Hide, K. Mlisana, S Abdool Karim, C.M. Gray, C. Williamson. Virology. 2010, 396:213-225.

(ii) *Conserved positive selection signals in gp41 across multiple subtypes and difference in selection signals detectable in gp41 sequences sampled during acute and chronic HIV-1 subtype C infection.* Bandawe GP, Martin DP, **Treurnicht F**, Mlisana K, Karim SS, Williamson C, CAPRISA 002 Acute Infection Study Team. Virol J. 2008 Nov 24; 5:141.

(ii) *Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage.* Chopera DR, Woodman Z, Mlisana K, Mlotshwa M, Martin DP, Seoighe C, **Treurnicht F**, de Rosa DA, Hide W, Karim SA, Gray CM, Williamson C; CAPRISA 002 Study Team. PLoS Pathog. 2008 Mar 21; 4(3):e1000033.

(iii) *Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection.* Gray ES, Moore PL, Choge IA, Decker JM, Bibollet-Ruche F, Li H, Leseka N, **Treurnicht F**, Mlisana K, Shaw GM, Karim SS, Williamson C, Morris L; CAPRISA 002 Study Team. J Virol. 2007; 81(12):6187-96.

ABBREVIATIONS

ADCC	Antibody-dependent cellular cytotoxicity
Ad5	Adenovirus type 5
AIDS	Acquired Immune Deficiency Syndrome
APOBEC3G	Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G
CAPRISA	Centre for the AIDS Programme of Research in South Africa
CCR5	Chemokine coreceptor 5
cDNA	copy DNA
CRF	Circulating recombinant form
CRFs	Circulating recombinant forms
CTL	cytotoxic T lymphocytes
COT	Center of the tree
CXCR4	CXC chemokine receptor 4
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanidine triphosphate
dNA	deoxyribonucleic acid
dH2O	distilled water
dN	number of non-synonymous substitutions
dS	number of synonymous substitutions
dTTP	deoxythymidine triphosphate
ds	double stranded
EIA	enzyme immunoassay
env	envelope
Gp120	Glycoprotein 120
Gp41	Glycoprotein 41
IAVI	International AIDS Vaccine Initiative

ICTV	International Committee on Taxonomy of Viruses
IN	integrase
IFN-γ	interferon gamma
HIV	human immunodeficiency virus
HIV-1	human immunodeficiency virus type 1
HLA	human leucocyte antigen
HTA	heteroduplex tracking assay
HTLV-III	human T cell lymphotropic virus type III
HVTN	Human Vaccine Trial Network
LAV	Lymphadenopathy-associated virus
LTNPs	long-term nonprogressors
LTRs	Long Terminal Repeats
PCR	polymerase chain reaction
PNGs	N-linked glycosylation sites
PR	protease
RDP	Recombination Detection Program
RT	reverse transcriptase
RNA	ribonucleic acid
SAAVI	South African AIDS Vaccine Initiative
SI	syncitium inducing
SIV	simian immunodeficiency virus
SIVcpz	simian immunodeficiency virus from chimpanzees
ss	single stranded
SSIII	SuperScript III reverse transcriptase
SU	surface unit
TAE	Tris-acetic acid-EDTA (Tris-acetate)
TM	transmembrane
TNF-α	tumour necrosis factor alpha
Tris	2-amino-2-(hydroxymethyl)-1,3-propandiol
mosB	mosaic proteins derived from subtype B viruses
mosC	mosaic proteins derived from subtype C viruses

mosM	mosaic proteins derived from M group viruses
MPER	membrane proximal region
MVA	modified Vaccinia Ankara
NK cells	Natural killer cells
nm	nanometers
NPs	nonprogressors
NSI	non-syncitium inducing
Pol	polymerase
QNE	quaternary epitopes
µl	micro liters
UCT	University of Cape Town
UK	United Kingdom
UNAIDS	Joint United Nations Programme on HIV/ AIDS
USA	United States of America
UV	ultraviolet
V1V2	variable regions 1 and 2

DECLARATION

I, Florette Kathleen Treurnicht hereby declare that the work contained in this thesis is my own original work and has not previously in its entirety or in part been submitted at any university for a degree. Contributions of colleagues, collaborators or use of published data have been duly acknowledged in the text. This work was done in the Division of Medical Virology, Department of Clinical Laboratory Sciences in the Faculty of Health Sciences and the Institute of Infectious Diseases and Molecular Medicine, University of Cape Town.

Florette Kathleen Treurnicht

Signature: Signed by candidate

Date: 10-05-2010

ACKNOWLEDGEMENTS

- Firstly, I would like to thank my heavenly Father for opportunities granted to me and for health, strength and perseverance, as well as the prayers, interest and support from my Bearers of blessing for without them this moment would not have dawned.

- A big thank you to Professor Carolyn Williamson, my Ph.D. supervisor and employer, through whom various opportunities came my way for her excellent mentorship especially in preparing this thesis and continuous encouragement. I learned to know her as someone who are very accommodating, fair and who would do everything to keep the harmony within her research group.

- Drs. Wendy Burgers and Tich Mangwende I would like to thank for scientific input and reviewing of Chapter 2, and Debbie Steward for proofreading Chapters 1 and 2.

- Thank you Drs. Darren Martin and Cathal Seoighe, as well as Natasha Wood for help with statistical analyses, helpful discussions preparation of data for publication.

- To everybody in the HIV Diversity and Pathogenesis Group I enjoyed working with you and for listening when I was discussing scientific questions that I needed to work through, although I did not always expect an answer. So thank you to Melissa-Rose Abrahams, Denis Chopera, Gama Bandawe, Dr. Zenda Woodman, Ruwayda Thebus, Craig Adams. I did not forget those who left for greener pastures also thank you to them for their help and encouragement especially when I was new to the lab: Helba Bredell, Dr. Baraka (Lucky) Malaza, Dr. Jandre Grobler, Celia Rademeyer, Dr. Joanne van Harmelen and Molefe Machaba.

❖ I would also like to thank Charlene Williams, Manorah Edwards, Debbie Steward and Nicky Balfour for help with administrative issues.

❖ My sincere gratitude and thankfulness to the following individuals and programs without whom this study would not have been possible: Participants from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) Acute Infection Study and the Study Team, Professor Salim Abdool Karim the director of CAPRISA for a Ph.D. student fellowship (2003 – 2007), the Columbia University-Southern Africa Fogarty AIDS Research Training Programme (TWO-02) for training in the amplification of full-length HIV-1 genomes (2004), Dr. Jim Mullins, Department of Microbiology, University of Washington, WA, USA for hosting me during my Fogarty training period (2004). Also thank you to our collaborators and their personnel at the National Institute of Communicable Diseases under the leadership of Drs. Lynn Morris and Clive Gray.

❖ Thank you to my family and friends for their continued interest and support.

❖ Last but not least, thank you to my husband Wilfred Noble for love and support, the cooking, cleaning and washing.

CHAPTER 1

LITERATURE REVIEW

1.1 HISTORY ON THE ORIGIN OF HIV/AIDS

Acquired Immunodeficiency Syndrome (AIDS) was first described in 1981 after the identification of *Pneumocystis carinii* pneumonia in previously healthy homosexual men in the United States of America (USA) (MMWR 30, 1981; MMWR 31, 1981; Gottlieb *et al*., 1981; Masur *et al*., 1982). Human immunodeficiency Virus type 1 (HIV-1), initially called Lymphadenopathy-associated virus (LAV) or Human T lymphotropic virus type III (HTLV-III), was subsequently identified as the aetiologic agent of AIDS by two independent research groups in France and the USA in 1983 and 1984 respectively (Barre-Sinoussi *et al*., 1983; Groopman *et al*., 1984).

The earliest clinical evidence of HIV-1 infection comes from samples collected in 1959 (Kinshasa male, Zhu *et al*., 1998) and 1971 to 1976 (Norwegian sailor, Froland *et al*., 1988). The identification of the second HIV-1 positive sample from Kinshasa collected in 1960 whose viruses differed by as much as 11.7% in the *env* gene from the 1959 strain suggested that HIV had already been diversifying in humans for a number of years before 1960 (Gao *et al*., 2001; Worobey *et al*., 2008). Based on this data, the latest introduction of HIV into humans was dated to around 1900 (1902 to 1921), 20 years earlier than previously estimated (Korber *et al*., 2000; Worobey *et al*., 2008). Although the origin of HIV-1 has been a topic of extensive debate in the past (Gao *et al*., 1999; Poinar *et al*., 2001; Worobey *et al*., 2004), recent studies have provided substantial evidence to show that HIV originated in chimpanzees (Keele *et al*., 2006). The zoonotic transfer of a SIVcpz strain is thought to have occurred through blood from non-human primates during hunting, butchering and handling of uncooked contaminated meat (Gao *et al*., 1999; Courgnaud *et al*., 2002; Peeters *et al*., 2002; Bibollet-Ruche *et al*., 2004). Phylogenetic evidence show that at least three zoonotic transmissions occurred resulting

in the HIV-1 M, N and O groups (Figure 1.1) (Keele *et al*., 2006, Sharp *et al*., 2001; Yang *et al*. 2000; Roques *et al*., 2004). Interestingly, the N group virus YBF30 displays characteristics of recombination between SIVcpz and HIV-1 which further support a common ancestral origin (Yang *et al*., 2000; Roques *et al*., 2004). Investigators have found evidence of an O-like virus in gorillas (SIVgor) suggesting that either gorillas were the origin of Group O in humans or that both humans or gorillas were infected from a common chimpanzee source (van Heuverswyn *et al*., 2006; Takehisa *et al*., 2009). However, recently a virus from a Cameroonian women was identified which grouped with the SIVgor sequences suggesting a fourth zoonotic transmission. It is proposed that this fourth lineage is called the HIV-1 P group, pending identification of further evidence of transmission between humans (Plantier *et al*., 2009).

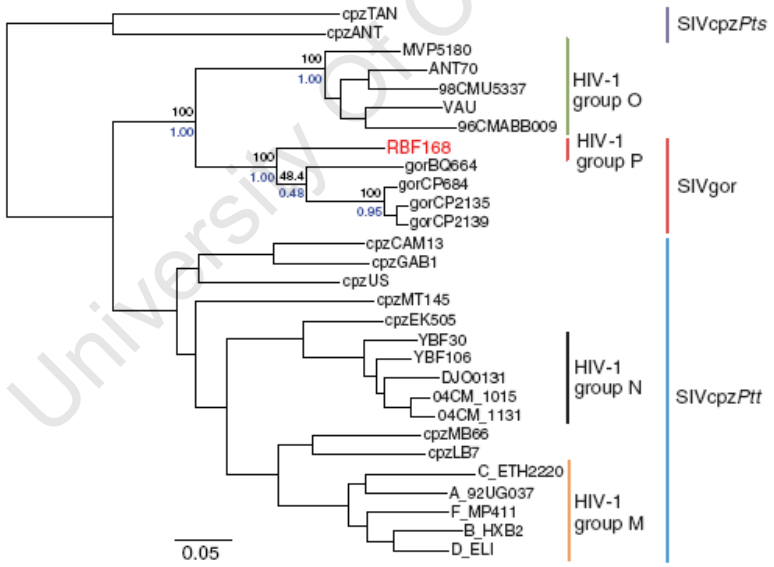


Figure 1.1 A phylogenetic tree of partial Pol amino acid sequences showing evolutionary relationships between M, N, O group HIV-1 viruses infecting humans and SIV strains from chimpanzees and gorillas, as well as the proposed new P group identified humans (Plantier *et al*., 2009).

1.2 HIV-1 VIRION STRUCTURE AND GENOMIC ORGANIZATION

The HIV virion is about 110-145 nanometers (nm) in diameter and has a roughly spherical host cell-derived membrane studded with the transmembrane (TM, gp41) and surface units (SU, gp120) of the envelope glycoprotein translated as the gp160 precursor. The envelope also contains other cellular proteins derived from the host during budding (Haseltine *et al*., 1991; Briggs *et al*., 2003). Below the host-derived membrane is a dense protein layer formed by the matrix protein (Gag p17) which surrounds the complex capsid (p24) that are either cone-shaped or cylindrical with the ends in close contact with the membrane (Briggs *et al*., 2003). Inside the capsid the nucleocapsid (p7) enfolds 2 copies of the viral single stranded (ss) positive sense RNA genome (Haseltine *et al*., 1991; Cohen *et al*., 2008). Viral enzymes encoded by the *pol* gene include integrase (IN) and reverse transcriptase (RT) and protease (PR) is encoded by the *pro* gene (http://phene.cpmc.columbia.edu/7thReport/7th Report of ICTV/ sites/descriptions/ Retroviridae/lentivirus.htm). The positive strand RNA genome has untranslated regions called R-U5 and U3-R at the 5'- and 3'-ends respectively. The complete HIV-1 provirus genome organization is shown in Figure 1.2 and a comprehensive list of the encoded viral proteins and their functions is given in Table 1.1 (Cohen *et al*., 2008).

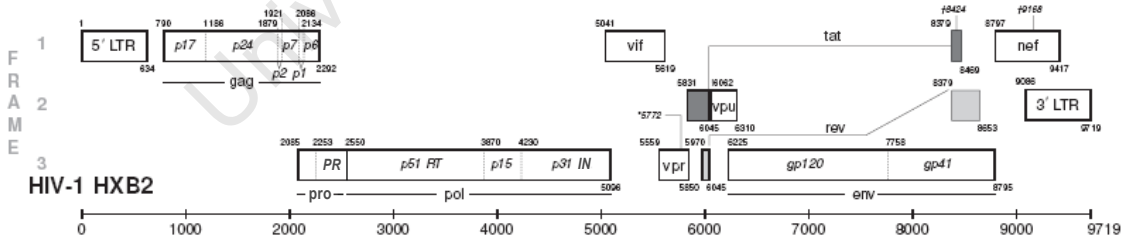


Figure 1.2 Genomic organization of HIV-1 with genes and the open reading frames (Frame 1, 2 and 3) from which they are expressed indicated as well as their protein products (adapted from http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/compendium. html).

Table 1.1 HIV-1 encoded proteins and their functions

Protein	Size (kDa)	Function and properties
Gag	p24	Structural capsid protein
Gag	p17	Myristoylated matrix protein
Gag	p7	Nucleocapsid protein, helps in reverse transcription
Gag	p6	L domain involved in budding
Pro	p10 (PR)	Posttranslational processing of viral proteins, Gag/Pol cleavage and maturation
Pol	p66, p51 (RT)	Reverse transcription
Pol	p15 (RNase H)	RNase H activity, removes RNA from DNA-RNA hybrids
Pol	p32 (IN)	DNA provirus integration requires integrase (IN)
Env	gp120	Envelope surface protein (SU : surface unit)
Env	gp41	Envelope transmembrane (TM) protein
Tat	p14	Transactivation, not virion-associated
Rev	p19	Regulation of viral mRNA expression, not virion-associated
Nef	p27	Can increase or decrease virus replication, MHC and CD4 down regulation
Vif	p23	Increases virus infectivity and cell-to-cell transmission, helps in proviral DNA synthesis and/or in virion assembly
Vpr	p15	Helps in virus replication, transactivation
Vpu	p16	Helps in virus release, disrupts gp160-CD4 complexes, not virion-associated

Adapted from Cohen *et al*., 2008

1.3 GENETIC DIVERSITY OF HIV-1

High genetic diversity is a well recognized characteristic of HIV with intra-person variation in long-term nonprogressors with chronic infection differing by as much as 15% in *env* (Shankarappa *et al*., 1999). This diversity of viral populations within an individual is described by the term quasispecies. A quasispecies describes a large and complex population of genetically diverse but related genomes that act as a whole (Goodenow *et al*., 1989). Each individual genome only has a short-lived existence which is subject to natural selection. A quasispecies (swarm) is inherently unstable and although variant genomes may at any one moment be at equilibrium with respect to selective pressures, it can easily shift if more advantageous mutants appear (Goodenow *et al*., 1989; Novak *et al*., 1990). However, the most-fit variant is not always favored but is subjected to the properties of the rest of the swarm and the host environment (Nowak *et al*., 1990).

Reverse transcription, mediated by the viral reverse transcriptase, results in the generation of double stranded (ds) DNA from the viral RNA genome. Characteristic of the reverse transcription process s (Figure 1.3) is the first and second DNA template

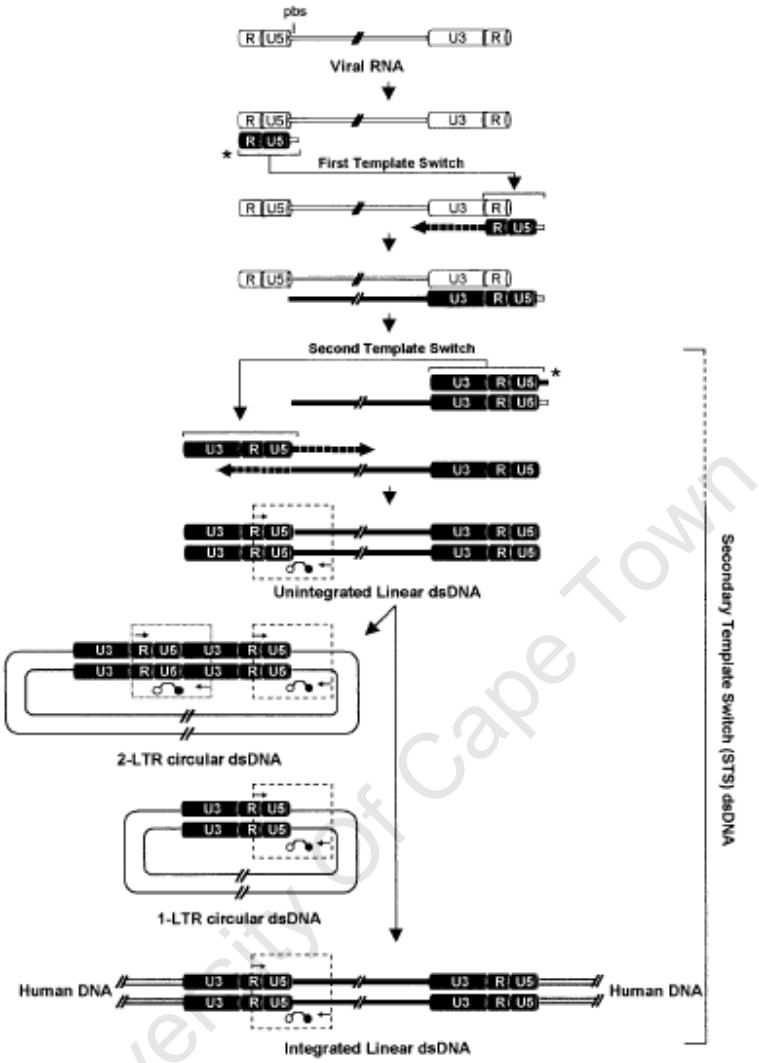


Figure 1.3 Schematic diagrams for reverse transcription of HIV-1 genomic RNA for integration into cellular DNA (proviruses) (Kostrikis *et al*., 2002).

switching steps that ultimately result in a DNA provirus with two direct repeats called the long terminal repeat (LTR) regions (van Wamel & Berkhout, 1998). The first copy DNA (cDNA) strand transfer process can involve the first strand from the duplicate RNA copy (termed intra-molecular strand transfer) or the first strand from another viral RNA genome (inter-molecular strand transfer) which would result in a recombinant genome during dual infections (van Wamel & Berkhout, 1998). The high degree of genetic diversity seen in HIV-1 infections is largely the result of its RT lacking 3'-5' exonuclease

proofreading ability and its ability to extend a mismatched primer on a DNA or RNA template. Reverse transcriptase has an estimated error rate of 2.16 x 10^{-5} (Mansky & Temin, 1995). Given that the estimated HIV generation time is approximately 2 days (Markowitz *et al*., 2003), and with a reproductive ratio of 6 (Stafford *et al*., 2000), the virus has a great ability to diversify over time. Thus high replication rates coupled with the error prone nature of the RT results in massive viral diversification over time. HIV-1 *env* gene diversifies at an estimated rate of approximately 1% per year (Shankarappa *et al*., 1999; Herbeck *et al*., 2006).

A second major mechanism that the virus uses to diversify is through recombination (Figure 1.4). A key feature of HIV-1 replication is strand-switching (Figure 1.3) which results in recombination between the two viral RNA copies from different variants infecting the same cell (Figure 1.4) (Haseltine, 1991; van Wamel & Berkhout, 1998; Katzenstein, 2006). Recombination allows the virus to make shifts in viral sequence by acquiring large pieces of sequences from other viruses co-infecting the same cell. HIV recombination occurs frequently during replication and some studies have suggested that there can be as many as 9 recombination events per virus replication cycle in T cells and up to ~30 in macrophages (Levy *et al*., 2004). The error-prone nature of reverse transcriptase together with recombination enables the virus to rapidly respond to selective pressures within the host.

Cellular apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) activity was identified as another mechanism that increases viral diversity and has also been implicated in the development of antiretroviral drug resistance mutations (Jern *et al*., 2009) as well as CTL escape (Wood *et al*., 2009). APOBEC3G is a cytidine deaminase which is effective at limiting retroviral infections (Harris *et al*., 2003; Zhang *et al*., 2003). Through this mechanism cytosines (C) from newly synthesized minus-strand viral DNA are modified to uracils (U) and presents as G>A transitions present in high frequency in the viral genome (G>A hypermutation) (Harris *et al*., 2003; Zhang *et al*., 2003, Yu *et al*., 2003). However HIV-1 like all other primate lentiviruses has evolved to overcome this host defense system as the viral Vif protein

forms an ubiquitin-ligase complex with cellular proteins which then bind to APOBEC3G and thereby target it for degradation (Sheehy *et al*., 2003; Yu *et al*., 2003; 2004).

A recent study showed that APOBEC3G–induced mutations are found at a frequency of 73% in proviral DNA, 17% in intracellular viral RNA and 5% in viral RNA. The low frequency in viral RNA is thought to be the result of strong selection against the incorporation of defective genomes (hypermutated genomes) into virus progeny (Russell *et al*., 2009). Proviral DNA hypermutation was recently shown to be associated with higher CD4 counts (Land *et al*., 2008).

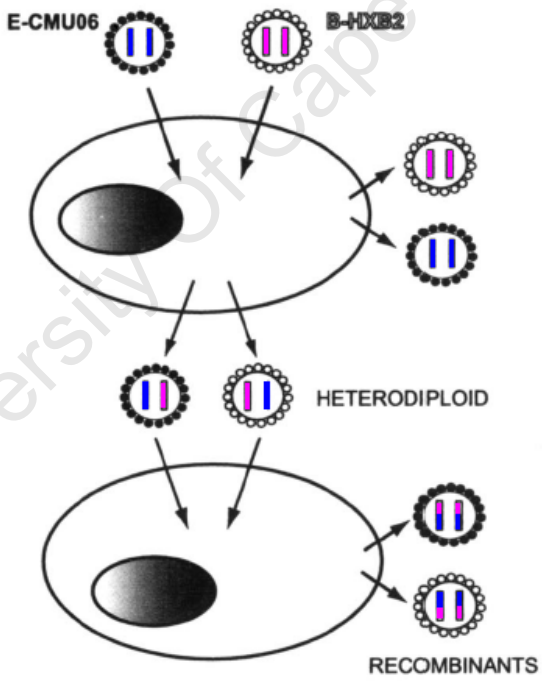


Figure 1.4 Schematic diagram showing dual infection of a single cell with 2 distinct HIV-1 variants followed by viral replication and packaging of 2 distinct vRNA copies per virion and subsequent replication with recombination (adapted from Quiñones-Mateu *et al*., 2002).

1.4 THE GLOBAL DISTRIBUTION OF HIV-1 STRAINS

HIV-1 is classified into Group M, N and O with the Group M viruses responsible for the global pandemic (Figure 1.5) as well as the recently described group P (see 1.1). Based on phylogenetic relationships Group M sequences have been classified into 9 subtypes with 6 sub-subtypes (subtypes A-D, F-H, J and K; sub-subtypes A1-A4, F1 and F2) (McCutchan 2006; Taylor *et al*., 2008). In some regions of the world some of the

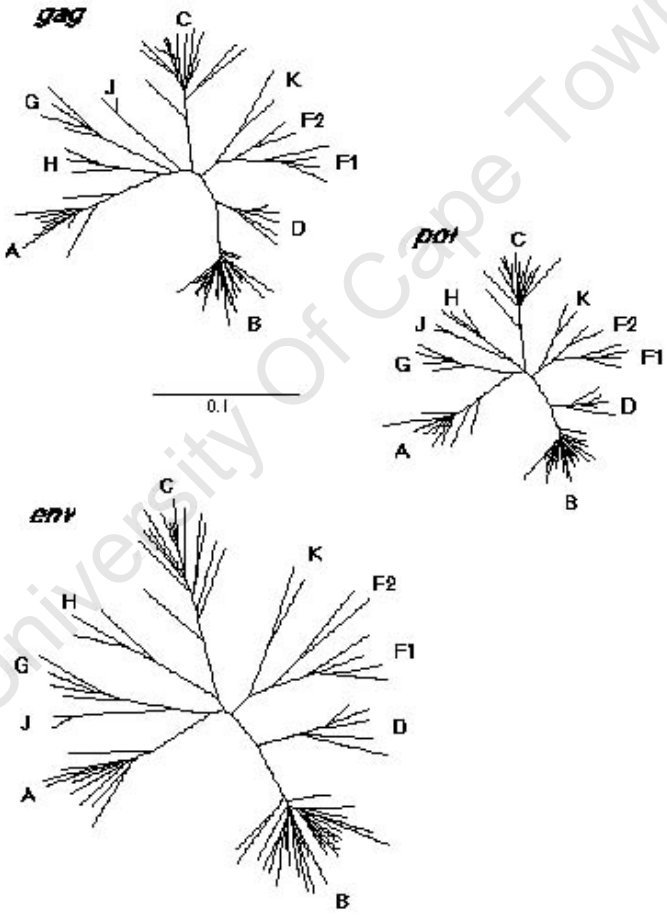


Figure 1.5 Unrooted Maximum Likelihood trees for the *gag, pol* and *env* genes showing the phylogenetic relationships between representative strains of the HIV-1 M group (from http://www.hiv.lanl.gov/content/sequence/HelpDocs/subtypes.html).

strains have recombined to form mosaic viruses that are successfully transmitted at high frequency and those reported in at least three epidemiologically unlinked cases are called circulating recombinant forms (CRFs).

There has been a dramatic increase in the number of CRFs described from 16 in 2005/ 2006 (Lal *et al*., 2005; McCutchan, 2006) to 43 in 2009 (Table 1.2) (http:// www.hiv.lanl.gov/content/sequence/HIV/REVIEWS/nomenclature/Nomen.html, Taylor *et al*., 2008; http://www.hiv.lanl.gov/content/sequence/HelpDocs/subtypes.html). The importance of CRFs in the global HIV-1 epidemic is highlighted by its association with epidemics in intraveneous drug users (IDUs) in Asia (CRF01_AE, CRF07_BC and CRF08_BC), the former Soviet Union (CRF03_AB), Spain and Portugal (CRF14_BG) and in Afghanistan and Iran (CRF35_AD) (Thomson *et al*., 2002; Taylor *et al*., 2008). In addition, CRF02_AG is responsible for the majority of infections in Nigeria, and dominates in the West African region (Figure 1.6).

All the different HIV-1 subtypes (M group) and CRFs (excluding those associated with sentinel IDU epidemics) as well as N and O group viruses are found in Africa. At least six strains account for the majority of HIV-1 infections world-wide namely subtypes A, B, C, D, CRF01_AE and CRF02_AG; and in most geographic regions outside Africa only one or two HIV-1 subtypes co-circulates with one usually dominating the epidemic (Figure 1.6) (Thomson *et al*., 2002; Lal *et al*., 2005; McCutchan, 2006). The most prevalent subtypes circulating in West Africa are subtypes A, G, CRF02_AG as well as other recombinant forms. Subtypes A and D, as well as multiple recombinant forms are also found at high frequency in East Africa.

Subtype B is responsible for most of the epidemics in the developed world including the Americas, Western Europe and Australia, while HIV-1 subtype A is highly prevalent in former Soviet Republics. The CRF01_AE circulates predominantly in Southeast Asia as well as subtype C which accounts for more than 50% of all HIV-1 infections world-wide. This subtype is found at high prevalence in Southern Africa (Taylor *et al*., 2008; Bredell *et al*., 2007; Hemelaar *et al*., 2006; McCutchan, 2006).

Table 1.2 List of circulating recombinant forms (CRFs) documented in June 2009 (from http://www.hiv.lanl.gov/content/sequence/HelpDocs/subtypes.html).

Name	Reference strain	Subtypes	Author
CRF01_AE	CM240	A, E	J.K. Carr
CRF02_AG	IbNG	A, G	J.K. Carr
CRF03_AB	Kal153	A, B	K. Liitsola
CRF04_cpx	94CY032	A, G, H, K, U	D. Paraskevis
CRF05_DF	VI1310	D, F	T. Laukkanen
CRF06_cpx	BFP90	A, G, J, K	R. B. Oelrichs
CRF07_BC	CN54	B', C	R. Wagner
CRF08_BC	GX-6F	B', C	F.E. McCutchan
CRF09_cpx	96GH2911	A, G, U	F.E. McCutchan
CRF10_CD	TZBF061	C, D	I.N. Koulinska
CRF11_cpx	GR17	A, CRF01, G, J	M. Peeters
CRF12_BF	ARMA159	B, F	J.K. Carr
CRF13_cpx	96CM-1849	A, CRF01, G, J, U	K. Wilbe
CRF14_BG	X397	B, G	R. Najera
CRF15_01B	99TH.MU2079	CRF01, B	F.E. McCutchan
CRF16_A2D	97KR004	A2, D	U. Visawapoka
CRF17_BF	ARMA038	B, F	J.K. Carr
CRF18_cpx	CU76	A1, F, G, H, K, U	M. Thomson
CRF19_cpx	CU7	A1, D, G	M. Thomson
CRF20_BG	Cu103	B, G	M. Thomson
CRF21_A2D	99KE_KER2003	A2, D	F.E. McCutchan
CRF22_01A1	CM001BBY	CRF01, A1	J.K. Carr
CRF23_BG	CB118	B, G	M. Thomson
CRF24_BG	CB378	B, G	M. Thomson
CRF25_cpx	02CM_1918LE	A, G, U	J.K. Carr
CRF26_AU	MBTB047	A, U	M. Peeters
CRF27_cpx	04FR-KZS	A, E, G, H, J, K, U	M. Peeters
CRF28_BF	BREPM12609	B, F	R. Diaz
CRF29_BF	BREPM16704	B, F	R. Diaz
CRF30_0206	NE36	CRF02, CRF06	M. Peeters
CRF31_BC	04BR142	B, C	M. Soares
CRF32_06A1	EE0369	CRF06, A1	M. Adojaan
CRF33_01B	05MYKL007	CRF01, B	K.P. Ng & K.K. Tee
CRF34_01B	OUR2275P	CRF01, B	F.E. McCutchan
CRF35_AD	AF095	A, D	F.E. McCutchan
CRF36_cpx	NYU830	A, G, CRF01, CRF02	R. Powell
CRF37_cpx	NYU926	A, G, CRF01, CRF02, U	R. Powell
CRF38_BF	GDJE	B, F	C. Lopez-Galindez
CRF39_BF	03BRRJ103	B, F	M.G. Morgado
CRF40_BF	05BRRJ055	B, F	M.G. Morgado
CRF41_CD	CO6650V1	C, D	S. Tovanabutra
CRF42_BF	luBF_13_05	B, F1	J-C. Schmit
CRF43_02G	J11223	CRF02, G	C. Brennan

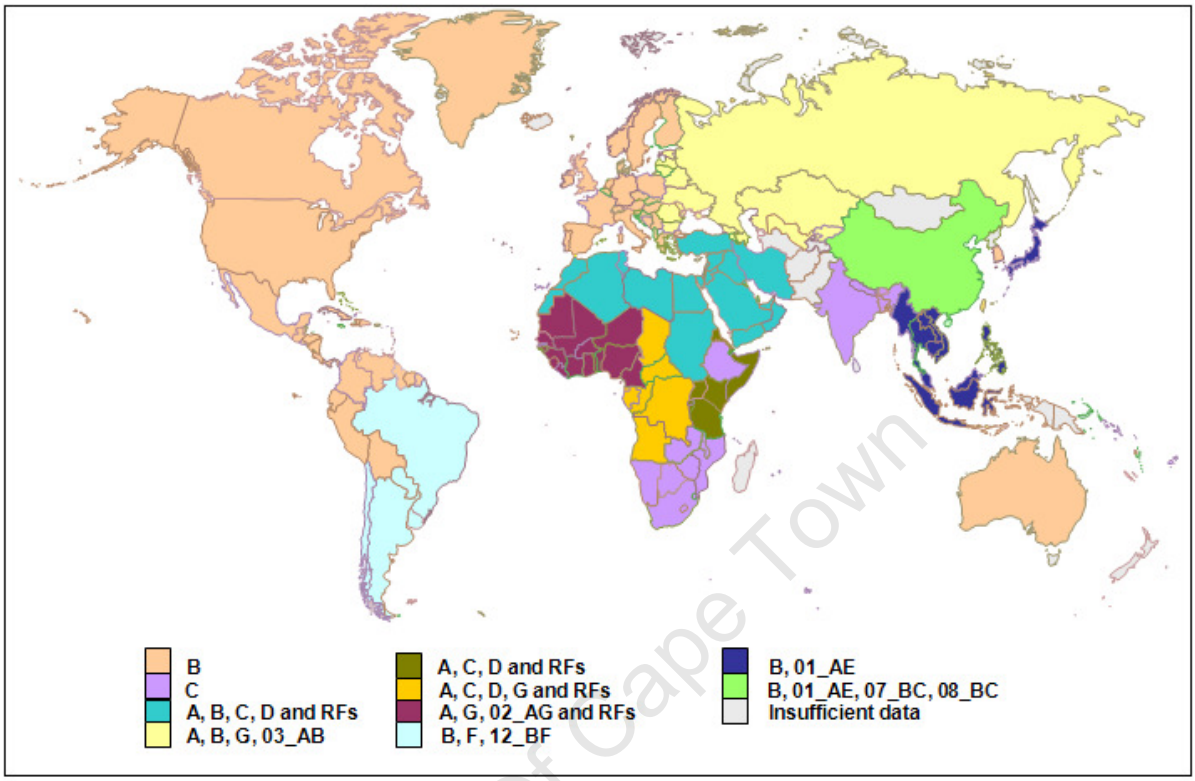


Figure 1.6 The global distribution of HIV-1 subtypes, circulating recombinant forms (CRFs, are numbered) and other recombinant forms (RFs) (Woodman & Williamson, 2009).

1.5 THE HIV-1 EPIDEMIC IN SOUTH AFRICA

The first cases of AIDS in South Africa were reported in 1983 and involved two homosexual male flight attendants who visited the USA during the course of their duty (Ras *et al*., 1983). In the early 1990s, HIV started to increase in prevalence in the heterosexual population. In 1990 there was a prevalence of 0.7% in women attending antenatal clinics which increased to 29.4% in 2007, making South Africa one of the hardest hit countries globally (National HIV and Syphilis Prevalence survey, South Africa, 2007; Dorrington & Bourne, 2008). Studies showed that the early epidemic associated with men who have sex with men was associated with subtypes B and D,

while the later heterosexual epidemic was associated with subtype C (Dietrich *et al*., 1993; Engelbrecht *et al*., 1995; Williamson *et al*., 1995; van Harmelen *et al*., 1997; Bredell *et al*., 1998; van Harmelen *et al*., 1999). HIV-1 subtype C is responsible for over 95% of infections in South Africa (van Harmelen *et al*., 1999; Bredell *et al*. 2007). A low frequency of subtypes like A, D, B, G and F1 and even some recombinants has also been reported (van Harmelen *et al*., 1997; Papathanasopoulos *et al*., 2002; Bredell *et al*., 2002; Engelbrecht *et al*., 2005; Loxton *et al*., 2005; Wilkinson & Engelbrecht, 2009).

1.6 HIV-1 INFECTION AND PATHOGENESIS

1.6.1 Natural History of Disease

Early studies based on large HIV-1 subtype B infected cohorts showed that CD4 T cell counts were predictive of disease progression (Mellors *et al*., 1997; de Wolf *et al*., 1997). They subsequently showed that viral loads were also highly predictive of disease outcome (Vlahov *et al*., 1998; Lyles *et al*., 2000; Arduino *et al*. 2001, Giorgi *et al*., 2002). The disease profile of HIV-1 infection can be divided into three stages namely: (i) primary infection, (ii) asymptomatic infection, and (iii) AIDS. The nomenclature of the early stage following infection has not been standardized. For the purpose of this thesis we will be use the term primary infection to describe the first 3 months following infection, and acute infection phase will be used to refer to the earliest stage of primary infection prior to the development of virus specific antibody responses.

HIV primary infection is often accompanied by a flu-like illness however it may also be completely asymptomatic. The severity of symptomatic primary infection has been associated with higher viral loads both during primary infection and at 12 months post infection (Schacker *et al*., 1998; Kelley *et al*., 2007). High plasma viral loads and a decline in CD4 T cells are characteristic of this stage. During primary infection, the decline in viral load occurs at the same time as the increase in HIV-specific cellular immune responses. After primary infection, viral load reaches a so-called steady state or

set-point. This decrease in viral load heralds the beginning of the asymptomatic or "clinical latency" stage which can last for as long as ten years or more. AIDS refers to the stage whereby there is an appearance of symptomatic disease and is associated with a loss in CD4 T lymphocytes to a critical level of <200 cells/mm^3 with a concomitant increase in viral load levels (Gange *et al*., 2001).

There is great variation in the rate of disease progression between individuals (Figure 1.7). Factors affecting the rate of disease progression include Gag-specific T helper and cytotoxic T cell responses, the virus inoculum, viral fitness as well as host genetic factors (Saah *et al*., 1998; Ioannidis *et al*., 1999; Edwards *et al*., 2002; Troyer *et al*., 2005, Brumme *et al*., 2008; Muira *et al*., 2009[a]; Muira *et al*., 2009[b]). About 10% of adults develop AIDS within 2 to 3 years of infection and are referred to as rapid progressors (Wang *et al*., 2000). A small proportion of HIV-1 infected adults (5%-10%) remains asymptomatic for more than ten years and is called nonprogressors (NPs) or long-term nonprogressors (LTNPs). These LTNPs have stable CD4 T cell counts and low (controllers) or undetectable (elite controllers) viral loads (Figure 1.7) (Wang *et al*., 2000; Kaur & Mehra, 2009). The majority of HIV-infected individuals have a clinical course that is highly variable. Most studies describing the rates of disease progression were determined before highly active antiretroviral treatment (HAART) became readily available (Kaur & Mehra, 2009). Anti-retroviral treatment is now widely available across the world to interrupt or slow disease progression. At the time of this study, the recommended treatment criteria in South Africa requires CD4 counts below 200 cells/mm^3 or AIDS defining illness for treatment initiation.

1.6.2 Transmission

Successful infection of a new host through sexual contact is estimated to occur at a frequency of approximately 1 in 1000 exposures although risk factors such as ulcerative genital disease and HIV disease stage can increase this to as high as 3 in 100 and in one case 1 in 10 (Cameron *et al*., 1989; Rakwar *et al*., 1999; Powers *et al*., 2008).

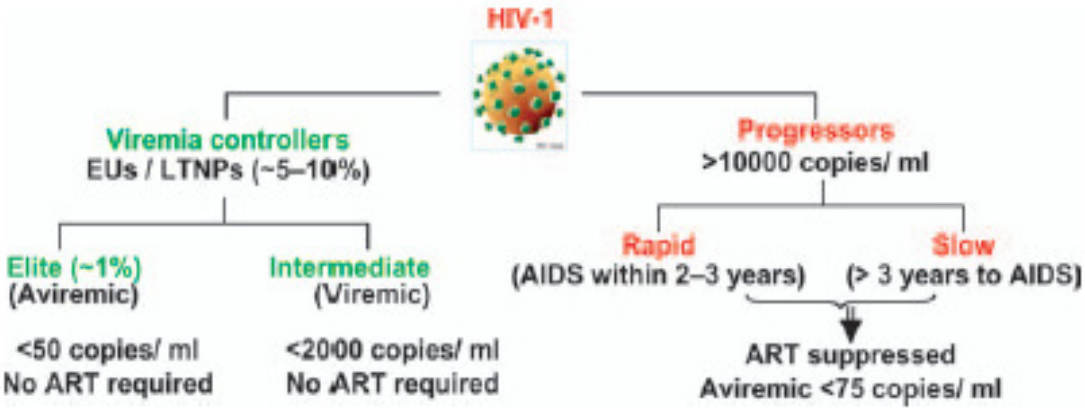


Figure 1.7 Shows variability in disease progression after HIV-1 infection (from Kaur & Mehra, 2009).

Transmission has been associated with a viral load in the donor of over 1500 viral RNA copies/ml of blood (Quinn *et al*., 2000).

The mucosal barrier is an important mechanical barrier that the virus needs to bypass to establish infection. It has long been known that there is a viral population bottleneck associated with transmission with recently infected individuals usually harbouring a highly homogeneous viral population (McNearney *et al*., 1992; Wolfs *et al*., 1992; Delwart *et al*., 2002). Recent studies using single genome amplification (SGA) have more accurately characterized this bottleneck through predicting the sequence of the founder virus(es) and quantitating the number of infecting variants (Keele *et al*., 2008; Abrahams *et al*., 2009; Haaland *et al*., 2009). SGA enables representative sampling, as well as reducing the chances of *in vitro* recombination through template switching. Direct sequencing instead of cloning also reduces the sequence errors due to *Taq* polymerase induced substitutions. Sequences generated from SGAs thus give a true reflection of quasispecies present *in vivo* (Salazar-Gonzalez *et al*., 2008).

Studies based predominantly on subtype B and subtype C sexual transmissions estimated that approximately 80%-90% of individuals are infected by a single virus

(homogeneous infection) (Keele *et al*., 2008; Abrahams *et al*., 2009; Haaland *et al*., 2009). However questions remain on the possible clonal outgrowth of a more fit viral variant during the first few days post infection as animal studies had shown the establishment of infection by multiple genetic variants and subsequent outgrowth of a single variant (Miller et al., 2005; Zhang et al., 1999). Prior infection with other pathogens, especially those causing ulcerative lesions, has been associated with a higher frequency of multiple HIV-1 variant, suggesting that mucosal inflammation compromises the mucosal barrier which in turns affects the genetic bottleneck (Kampinga *et al*., 1997; Long *et al*., 2000; Sagar *et al*., 2003; Haaland *et al*., 2009). This implies that the mucosal barrier is in part responsible for this bottleneck. In addition, it is clinically important as high diversity following transmission has been associated with more rapid disease progression (Sagar *et al*., 2003; Grobler *et al*., 2004; Gottlieb *et al*., 2004).

HIV-1 is thought to cross the mucosal barrier in several ways: particles may be captured by dendritic cells (DCs); may directly cross by transcytosis; or by infecting intraepithelial lymphocytes (Figure 1.8) (Haase & Pope, 2003). After crossing the mucosal barrier, HIV-1 infects resting and activated CD4 T cells, DCs and macrophages in the submucosa. Following local replication, it is thought that the virus seeds the lymphatic system from where it migrates to the gut-associated T lymphocytes (GALT) where it rapidly replicates and destroy CD4 T cells (Brenchley *et al*., 2004; Mehandru *et al*., 2004). From the GALT, the virus disseminates to the lymphoid organs which also serve as viral reservoirs during clinical latency (Guadalupe *et al*., 2003; Mehandru *et al*., 2004; Haase & Pope, 2003).

During this time of viral dissemination the latent viral reservoirs are established. The virus enters cells by firstly binding the CD4 receptor with its envelope glycoprotein which induces a structural transformation in the viral envelope to expose the binding site for the second receptor. At the mucosal membranes HIV-1 variants that preferentially use the CCR5 chemokine coreceptor (R5 tropic viruses) infect DCs, T lymphocytes and macrophages. These cells express both the CD4 receptor and the CCR5 coreceptor on its surface. However HIV can utilize a number of other coreceptors, with CXCR4

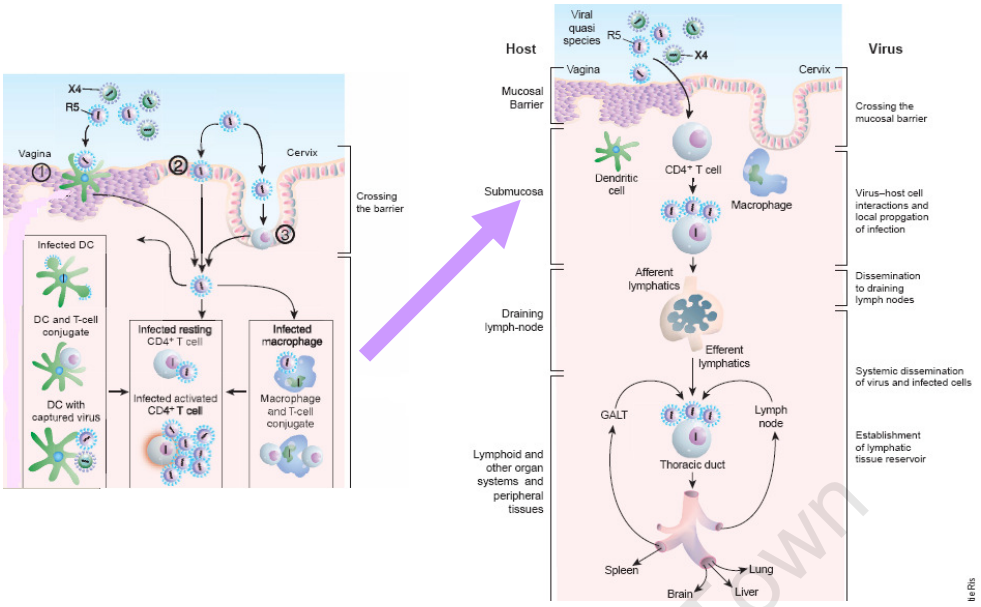


Figure 1.8 Diagram showing HIV-1 infection through the mucosal membrane using (1) dendritic cells, (2) transcytosis and (3) intraepithelial lymphocytes followed by local replication, systemic dissemination and establishing the lymphatic tissue reservoir (adapted from Haase & Pope, 2003).

chemokine receptor being the major alternative receptor which is often associated with viruses in late disease stage (X4 tropic viruses). Binding to the second receptor is needed to complete envelope transformation and membrane fusion to internalize the virus (Kozak *et al*., 1997; Rizzuto *et al*., 1998).

1.7 THE TRANSMITTED VIRUS

The primate model of SIV infection demonstrated that the virus, after localized replication in the submucosa, spread to the GALT where it establishes productive infection within days after transmission to a new host (Veazey *et al*., 1998). Thus the best opportunity to prevent infection through vaccination or other prevention technologies will be within the first few days post infection before spread to the lymphoid tissue and establishment of the viral reservoirs (Figure 1.9) (Haase & Pope, 2003; Haynes

& Shattock, 2008).

As vaccines and microbicides need to protect against the transmitted virus, there is great interest by researchers to identify and characterize the earliest transmitted HIV-1 variants. Some of the first reports suggested that there are specific signatures associated with transmission. The first characteristic is the predominant selection of CCR5-tropic variants (Deng *et al*., 1996); these are almost always associated with transmission, even if the donor harbored predominantly X4 viruses (Wolinsky *et al*., 1992; van't Wout *et al*., 1994). Further evidence of the role of CCR5 in transmission is that individuals who do not have functional CCR5 receptors are largely resistant to infection (Liu *et al*., 1996, Huang *et al*., 1996; Lucotte, 2002). A study of primary infection viral envelope glycoproteins sampled from a subtype C acute infection cohort showed different selection pressures in acute infection compared to those from chronic infection viruses (Bandawe *et al*., 2008). Furthermore, studies of donor recipient transmission pairs have shown that biological properties differ: subtype C viruses in recently infected individuals were shown to have a shorter variable loops in envelope and were more neutralization sensitive relative to the donor (Derdeyn *et al*., 2004; Rong *et al*., 2007). A cross sectional study of

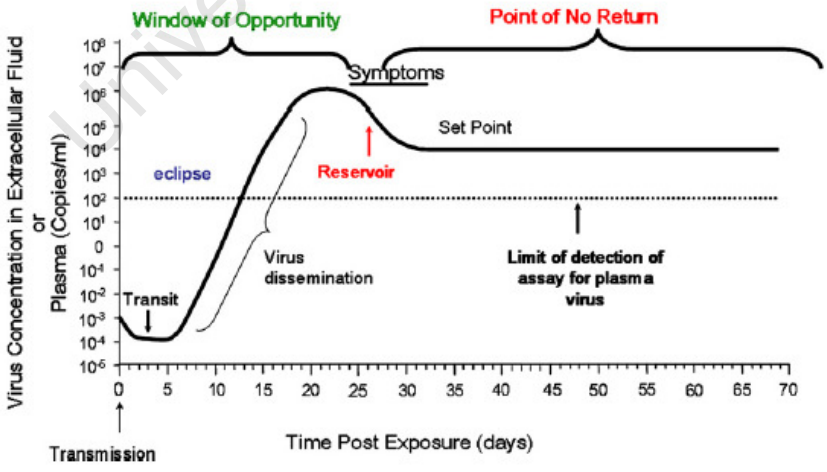


Figure 1.9 This diagram shows the time course of HIV-1 infection and indicating the window of opportunity for therapeutic or vaccine intervention (from Haynes & Shattock, 2008).

subtype A envelope sequences also suggested that subtype A viruses from early infection had shorter V1V2loops and fewer N-linked glycosylation sites than those from chronic infection (Chohan *et al*., 2005). Furthermore, envelope glycoprotein variable loop lengths have been shown to increase over time and this increase has been associated with escape from neutralizing antibody responses (Sagar *et al*., 2006; Rong *et al*., 2007). Potent autologous neutralizing responses were reported in early infection with HIV-1 subtype C and the magnitude of these responses were dependent on variable loop length and number of potential N-linked glycosylation sites (Derdeyn *et al*., 2004; Gray *et al*., 2007). Thus understanding the functional characteristics of the envelope glycoproteins is important in the design of vaccines that aim to induce neutralizing antibody responses.

1.8 HOST IMMUNE RESPONSES AND DISEASE PROGRESSION

1.8.1 Innate immune responses

The innate immune system is the first line of defense and exerts control of HIV-1 replication through cellular, extracellular and intracellular components. The cellular component involve DCs, natural killer (NK) cells, γδ T cells and regulatory T cells; whereas the extracellular component are made up of various cytokines and chemokines (Figure 1.10); and lastly the intracellular component represent the antiviral cellular proteins namely APOBEC3G, TRIM5α and the defensins (Lehner *et al*., 2008).

Evidence for the role of cellular components of innate immunity are shown in a study of highly exposed persistently seronegative sexworkers from Kenya where resistance to HIV-1 infection was associated with the genotype A1*010301 of the human leucocyte antigen-DP, that presents antigen to CD4 T cells (Hardie *et al*., 2008). Furthermore, natural killer cells that express the NK cell receptor, the killer immunoglobulin-like receptor 3DS1 (KIR3DS1) and its ligand (HLA-B Bw4-801) can effectively control HIV-1 replication and has been associated with slower disease progression (Alter *et al*., 2007).

Co-expression of protective HLA alleles (B*57) and NK cells with the KIR3DS1 phenotype are strongly associated with slower progression to AIDS (Boulet *et al*., 2008). Recently, strong HIV-1 specific NK cell responses have been associated with protection against HIV-1 infection in infants from mother-to-child transmission pairs (Tiemessen *et al*., 2009). A large cohort study on host determinants for control of HIV-1 infection, identified polymorphisms in host genes like HLA (B*5701, HLA-C), ring finger protein 39 (RNF39) and zinc ribbon domain-containing1 (ZNRD1) (Fellay *et al*., 2007).

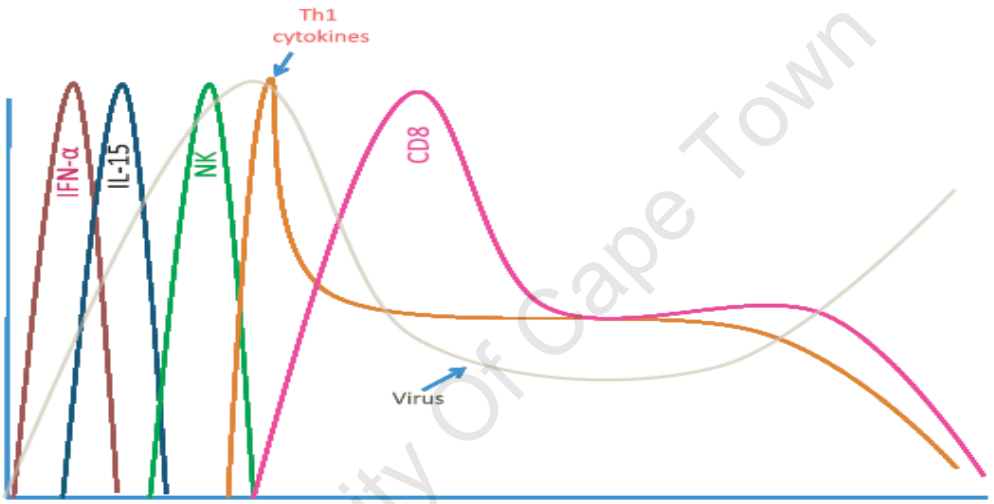


Figure 1.10 Diagram shows the kinetics of early immune responses to viral infections which begin with the release of IFN-α followed by the production of IL-15 that induce rapid proliferation of NK cells which also secrete Th1 cytokines which stimulates virus-specific CD8 T cell responses (from Alter & Altfeld, 2009).

First evidence of the role of the extracellular component in innate defenses was the demonstration that the β-chemokines macrophage inhibitory protein (MIP)-1α, MIP-1β and RANTES which is the natural ligand for the CCR5 coreceptor, inhibit entry of R5-tropic viruses (Cocchi *et al*., 1995). HIV acute infection is associated with a "cytokine storm" where rapid increases in IL-15 and IFN-α at peak viraemia may be associated with initial viral clearance but may also have negative immunopathological

effects (Stacey *et al.*, 2009).

Two important intracellular factors that restrict HIV-1 entry into target cells are APOBEC3G and TRIM5α (Harris *et al*., 2003; Zhang *et al*., 2003; Stremlau *et al*., 2004). TRIM5α, a tripartite motif bearing protein, is induced by IFN-α and affects the uncoating of the HIV-1 capsid during infection (Stremlau *et al*., 2004; reviewed by Towers, 2007). Increased levels of TRIM5α mRNA expression have been associated with reduced risk of infection (Sewram *et al*., 2009).

1.8.2 Cytotoxic T cell responses

Cytotoxic immune responses are induced in response to the presentation of processed antigen by the HLA type 1 molecules on the surface of infected cells. CD8 (cytotoxic) T cell-mediated immune responses are thought to be important in the initial control of peak viraemia in primary HIV-1 infection (Koup *et al*., 1994; Borrow *et al*., 1994; Pontesilli *et al*., 1998; Goulder & Watkins, 2004). CD8 depletion experiments in SIV infection in the macaque animal model confirmed the essential role of CD8 T cell (CTL) responses in the control of virus replication (Schmitz *et al*., 1999; Jin *et al*., 1999).

Further evidence of the importance of CTL in controlling disease progression is the association between HLA and disease progression. HLA class I alleles differ in their ability to control HIV-1 replication and the protective HLA types B*13, B*27, B*51, B*57, B*5801, and B*8101 have been associated with better prognosis (slow or long-term nonprogression) (Saah *et al*., 1998; Ioannidis *et al*., 1999; Costello *et al*., 1999; Fellay *et al*., 2007; Goulder & Watkins, 2008). These protective HLA types target CTL epitopes in functionally constrained regions of especially the Gag protein and highly conserved residues in the HIV-1 proteome (Martinez-Picado *et al*., 2006; Boutwell *et al*., 2009; Wang *et al*., 2009). Overall increased numbers of predicted HLA-associated CTL epitopes were associated with lower viral loads and higher CD4 counts whereas increased numbers of predicted CTL epitopes in Gag and Nef was associated with lower viral loads

in individuals carrying protective HLA-I alleles (Rousseau *et al*., 2008; Rolland *et al*., 2008).

Studies on HIV-1 infected individuals from various ethnic origins shows that predominant responses against the Gag and Nef proteins are localized in specific immunodominant regions in Gag and Nef (Frahm *et al*., 2004; Masemola *et al*., 2004; Geldmacher *et al*., 2007; Gray *et al*., 2009). Gag-specific cytotoxic immune responses against an increased number of epitopes (increased breadth), together with CTL escape mutations in functionally constrained regions of Gag which reduced fitness, have both been associated with decreased viral loads (Martinez-Picado *et al*., 2006; Kiepiela *et al*., 2007; Boutwell *et al*., 2009; Miura *et al*., 2009; Wang *et al*., 2009).

However, T cell-mediated immune targeting of Env has been associated with increased viral loads (Kiepiela *et al*., 2007; Ngumbela *et al*., 2008; Troyer *et al*., 2009). Recently, Gray *et al*. (2009) reported that in a HIV-1 subtype C acute infection cohort, the HIV-1-specific T cell responses during primary infection were a poor predictor of viral set point and disease progression. Contradictory to this observation, Streeck *et al*. (2009) found that increased early immunodominant HIV-1-specific CD8 T cell responses are associated with lower viral set point and slower CD4 T cell loss in a large cohort of individuals infected with subtype B.

1.8.2.1 CTL escape and reversion to wild-type

HIV-1 overcomes control by CTLs through a mechanism of variation in amino acid residues that affect CTL epitope processing, binding or recognition by HLA alleles, resulting in non-recognition (Couillin *et al*., 1994; Jamieson *et al*., 2003; Yokomaku *et al*., 2004). Viral escape from cytotoxic T cell mediated responses (CTL escape) has been associated with loss of immune control and subsequent disease progression (Kelleher *et al*., 2001; Goulder *et al*., 1997; Karlsson *et al*., 2007; Kemal *et al*., 2008). However, CTL escape mutations in functionally constrained positions have also been shown to result in

virus progeny with decreased fitness and have been associated with control of viral replication and slow or long-term nonprogression (Wagner *et al*., 1999; Leslie *et al*., 2004; Martinez-Picado *et al*., 2006; Boutwell *et al*., 2009; Miura *et al*., 2009[a]; Miura *et al*., 2009[c]).

Reversion of transmitted CTL escape mutations to wild-type are often seen after transmission and may result in restoration of replicative fitness (Friedrich *et al*., 2004; Leslie *et al*., 2004; Li *et al*., 2007). Some of these attenuating mutations do not revert after transmission to an HLA-mismatched recipient because the virus evolved to acquire additional mutations, called compensatory mutations which alleviates or restores the fitness cost induced by the original escape mutation (Brockman *et al*., 2007; Crawford *et al*., 2007; Schneidewind *et al*., 2009).

1.8.3 Neutralizing antibody responses

Antibody responses are usually the first line of defense against viral infections as they target viral entry processes. However in HIV infection, although HIV-1-specific IgG and IgM antibody responses are detectable from 8 days post infection, it is clear that the majority of envelope-specific antibody responses are non-neutralizing and do not reduce the viral load in acute infection (Tomaras *et al*., 2008). Some non-neutralizing antibodies were found to be effective in inducing antibody dependent cellular cytotoxicity (ADCC), an immune mechanism that had been associated with virus control by Natural Killer cells (NK cells, Ahmad & Menezes, 1995; Weber *et al*., 2000, Stratov *et al*., 2008). Other mechanisms of antibody-mediated viral clearance are through antibody-virus aggregation and phagocytosis followed by virus killing or through complement-mediated ADCC (Srivastava *et al*., 2005).

Neutralizing antibodies only emerge some weeks or months after control of acute phase viraemia (Wei *et al*., 2003; Gray *et al*., 2007). Neutralization sensitivity of HIV-1 had been mapped to the V1V2 variable regions of the envelope glycoprotein and the virus

can evade neutralizing antibody responses by changing the variable loop length or its glycosylation profile (Wei *et al*., 2003; Derdeyn *et al*., 2004; Rong *et al*., 2007; Gray *et al*., 2007). HIV-1 can mutate to change the variable loop lengths without affecting the functional envelope structure and it had been shown that deletions in the variable loops are associated with positive selection pressures (Wood *et al*., 2009). A few neutralizing antibodies that effectively neutralize a broad range of HIV-1 strains have been identified from individuals with HIV-1 subtype B infections namely: 2F5, 4E10, IgG1b12 and 2G12 (Trkola *et al*., 1996; Zwick *et al*., 2001; Binley *et al*., 2004). IgG1b12 targets the CD4 binding site whereas 2G12 targets a cluster of oligomannose sugars on gp120 (Scanlan *et al*., 2002; Srivastava *et al*., 2005). The 2F5 and 4E10 broadly neutralizing antibodies target the membrane-proximal external region of gp41 where it interferes with membrane fusion.

1.9 VACCINE DESIGN STRATEGIES AND HIV-1 DIVERSITY

Current vaccine design strategies against HIV-1 aim to either prevent productive infection or to reduce the viral load set point following infection with good predicted clinical outcome (Figure 1.11) (reviewed by Barouch, 2008).

1.9.1 Antibody-based vaccine design strategies

A vaccine that can induce neutralizing antibodies remains one of the biggest challenges of HIV vaccine research. In the early 1980's the identification of HIV-1 as the eatiologic agent for AIDS raised the hopes that an effective vaccine based on the surface or envelope glycoprotein would provide protection (Srivastava *et al*., 2005). However, the complex trimeric structure of HIV makes eliciting relevant antibodies through vaccines difficult: for example the HIV-1 V3 region of the surface unit of the envelope glycoprotein (Gp120) was found to be a potent inducer of neutralizing antibodies against tissue culture adapted strains of HIV-1 but not primary isolates (Mascola *et al*., 1996). Following the failure of the human vaccine efficacy trials of

monomeric gp120 subunit vaccines (Flynn *et al*., 2005; Pitisuttithum *et al*., 2006) the field focused on the development of a more natural (trimeric) envelope vaccine that elicits broadly neutralizing antibodies (Srivastava *et al*., 2003; 2005). However developing such a vaccine immunogen is a major challenge. HIV-1 infected individuals take weeks but usually months to develop neutralizing antibodies against their own virus and only a minority of them develops antibodies that can neutralize other HIV strains.

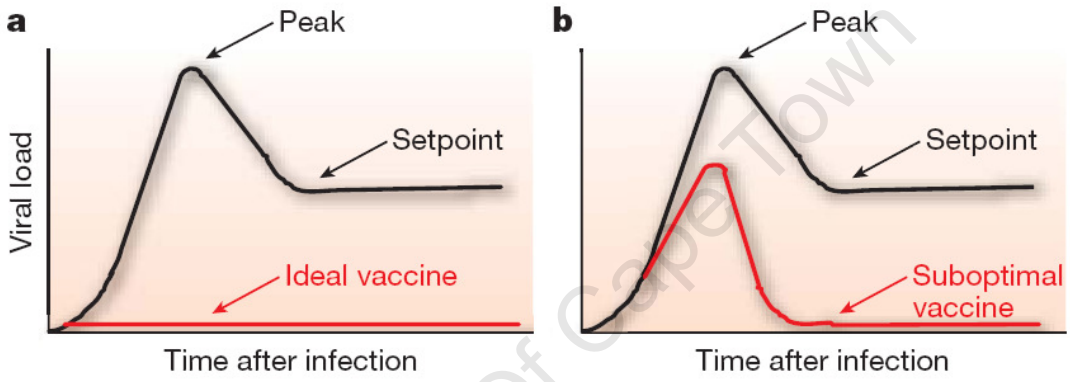


Figure 1.11 The expected effect of HIV-1 vaccines on potential infection indicated by longitudinal viral load trajectories. The black lines on the graphs show the normal longitudinal viral load in the absence of intervention. The red lines show the expected longitudinal viral load (a) with a successful (gold standard) vaccine inducing sterilizing immunity and (b) with a sub-optimal or therapeutic vaccine with a reduced viral load peak and lower set point after break-through infection (from Barouch, 2008).

The membrane proximal region (MPER) of gp41 also serves as a target for therapeutic and vaccine-based interventions. Massanella *et al*. (2009) demonstrated that anti-gp41 specific neutralizing antibodies (2F5, 4E10) prevent cell to cell spread of HIV-1. However mimics of the 4E10 epitope for vaccine design failed to induce the expected immune responses. More recently Li *et al*. (2008) found that deletion of the gp41 fusion peptide or destabilization of the fusion core enhanced the antigenicity and immunogenicity of 4E10 and this strategy may improve vaccine designs aimed at the induction of 4E10-like antibodies.

Newer targets for inducing neutralizing antibodies with improved antiviral properties include inducing antibodies against tertiary structures also called quarternary neutralizing epitopes (QNE). These include the cross-neutralizing antibody, 2G12 (Scanlan *et al*., 2002; Binley *et al*., 2004; Stamatatos *et al*., 2009). Recently the first non-subtype B induced broadly neutralizing antibodies (PG9 and PG16) were isolated from an African donor infected with subtype A (Walker *et al*., 2009). These two monoclonal antibodies seems to also target the QNEs on HIV-1 virus neutralization is dependent on the V1, V2 and V3 loops and the presence of glycans (Walker *et al*., 2009).

Strategies to design effective antibody-based vaccines to induce effective neutralizing antibodies with sufficient breath focus on (i) the CD4 binding site (including CD4 inducible conformational epitopes), (ii) the coreceptor binding site and (iii) the membrane proximal external region (MPER) of gp41 (Zwick *et al*., 2001; Binley *et al*., 2004; Stamatatos *et al*., 2009). However, Korber & Gnanakaran (2009) showed by using a structure-based evaluation of sequence variation in gp120 from subtype B that only 20% of potential antibody epitopes will be recognized by a subtype B strain and only 12% by a subtype C strain which demonstrates the low level of potential cross-neutralizing antibody responses predicted to be induced by potential envelope-based vaccines (Figure 1.12). Similarly Haynes *et al*. (2006) showed that subtype B consensus like V3 loop immunogens induced neutralizing antibodies in less than 50% of subtype B primary isolates with limited cross-neutralization potential against unrelated subtypes.

1.9.2 T cell based vaccine design strategies

Following the failure of the first Env-based vaccines to induce protective neutralizing antibody responses, focus had shifted to induce cytotoxic T cell responses induced against HIV-1 by using naked DNA or vectors that express whole protein (Moore & Burton, 2004; Duerr *et al.*, 2006). As CTL based vaccines target linear epitopes it is likely that these types of vaccines will be particularly susceptible to the effect of genetic diversity. Consensus, ancestral or COT-derived protein antigens (also

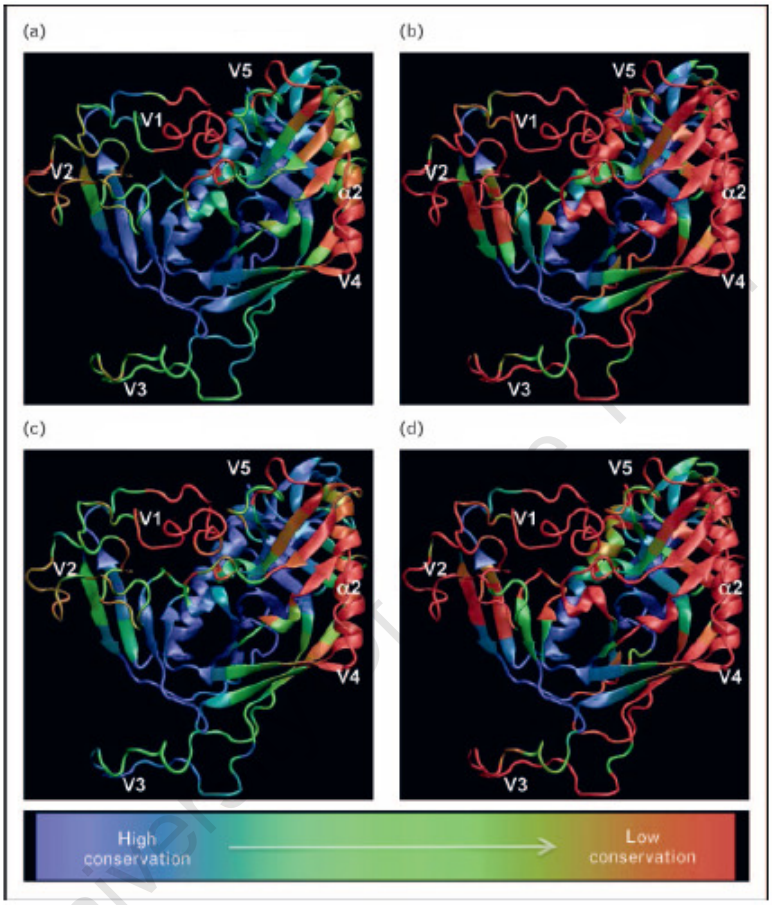


Figure 1.12 Showing epitope mismatches in clusters comparing a consencus B envelope against (a) acute subtype B viruses and (b) against acute subtype C viruses, whereas epitope coverage of a subtype C consensus envelope is shown in (c) against acute subtype C viruses and (d) against acute subtype B viruses (adapted from Korber & Gnanakaran, 2009).

called centralized immunogens) reduce diversity between the vaccines and the circulating viruses (further details provided in Chapter 2).

More recently an immunogen design strategy that also strives to reduce diversity as well as account for as much of the high frequency mutations as possible were used to design T cell-based vaccines. This strategy involves the design of mosaic or recombinant proteins generated through artificial recombination of HIV-1 genes with each recombinant selected retaining all the required functional properties of native proteins (Figure 1.13) (Fischer *et al*., 2007). The final immunogen is comprised of multiple recombinant proteins or peptides which contain an optimal number of ninemers (9-mers/ linear epitopes) as targets for a wide range of HLA alleles (Fischer *et al*., 2007; Kong *et al*., 2009).

The Merck Adenovirus vector type 5 (Ad5) Gag/Pol/Nef vaccines were the first promising vaccines aimed at the induction of cellular immune responses against HIV-1. These vaccines successfully induced HIV-1 specific cell-mediated immune responses

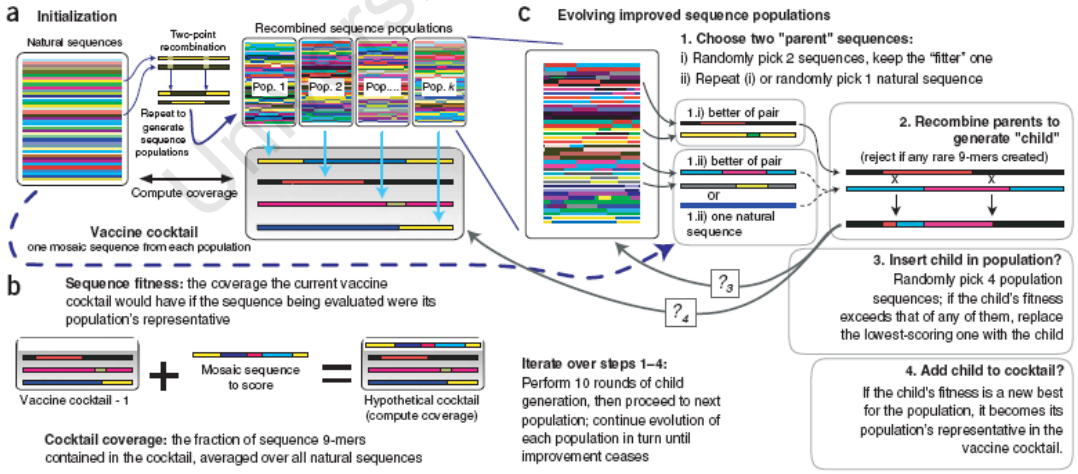


Figure 1.13 Diagram showing the processes of virtual HIV-1 recombination using computer-based algorithms to obtain "functional" mosaic proteins (Fischer *et al*., 2007).

during safety and immunogenicity trials, but failed to protect against infection and also did not lower the viral loads following break-through infection in a proof-of-concept trial (Priddy *et al*., 2008; McElrath *et al*, 2008; Buchbinder *et al*., 2008). The vaccines will also be discussed in chapter 2.

1.10 STUDY SCOPE AND RATIONALE

The Centre for the AIDS Programme of Research in South Africa (CAPRISA) was established by five institutions namely University of KwaZulu-Natal, University of Cape Town, University of the Western Cape, National Institute for Communicable Diseases and Columbia University in New York, USA in 2002 (http://www.caprisa.org/joomla/). The large CAPRISA 002 Acute Infection study funded by the Comprehensive International Programme of Research on AIDS (CIPRA) through the National Institute of Allergies and Infectious Diseases (NIAID) of the National Institute of Health, USA (http://www3.niaid.nih.gov/news/newsreleases/2002/sacipraawrd.htm; http://www.caprisa.org/joomla/) with Dr. Salim Abdool Karim as principal investigator and Drs. Carolyn Williamson and Koleka Mlisana as protocol chairs was established in 2002. The study aimed to define prognostic markers of disease progression in HIV-1 subtype C infection, and to investigate the impact of cellular and antibody immune responses in acute and early infection on disease progression. More specifically the study was entitled: "Viral set point and clinical progression in HIV-1 subtype C infection: The role of immunological and viral factors during acute and early infection". Primary study objectives were: (i) to determine whether viral load at 12 months post infection is related to the subsequent progression of HIV disease as defined by CD4+T count below 350 cells/mm^3 or AIDS defining illnesses or initiation of antiretroviral therapy, (ii) to determine whether the magnitude and breadth of HIV-1 specific CD8+ T cell responses at three months post infection correlates with viral load at 12 months post infection.

Our research under leadership of Dr. Carolyn Williamson focused on the following specific objectives of CAPRISA 002: (i) to describe the dynamics of viral evolution in subtype C infection and to determine if viral genetic changes associated with neutralizing antibody and cellular immune responses are associated with increased viral load, (ii) to determine if dual infection with two genotypically distinct HIV strains is associated with changes in viral load and disease progression.

The purpose of this study was to characterize full-length HIV-1 genomes from HIV-1 subtype C primary infection to gain further insights into HIV transmission and early viral evolution.

The aims of this study were:

(i) To characterize near full-length HIV-1 subtype C viruses from recent primary infections

(ii) To identify genetic features characteristic of recently transmitted HIV-1 subtype C viruses that distinguish them from viruses from chronic infection

(iii) To describe the effects of immune selective pressures in primary infection on viral evolution

South Africa has the most severe HIV epidemics globally and the development of a vaccine effective against variants circulating in South Africa is a national priority. HIV is continually evolving within individuals and at a population level. While there are more than 400 near full-length HIV-1 subtype C virus sequences in the HIV sequence database, only ~14 were obtained from individuals during primary infection (http://www.hiv.lanl.gov/; Lole *et al*., 1999; van Harmelen *et al*., 2001; Salazar-Gonzalez *et al*., 2009; Fernãndez-Garcia *et al*., 2009), with 3 from South Africa. As diversity may affect vaccine efficacy we were interested in characterizing viruses from primary infection as these represent currently circulating strains. Furthermore, we wanted to determine the predictive effectiveness of candidate HIV vaccines that are tested in South Africa against HIV-1 subtype C.

Biomedical interventions such as vaccines and microbicides target the transmitted virus and studies have shown that these viruses may have features different from viruses in chronic infection. It is therefore of interest to further elucidate genotypic features of viruses from acute infection which may differentiate them from those in chronic infection. Furthermore, as events in acute infection are thought to play an important role in determining subsequent disease progression, it is also important to understand the immune selective pressures responsible for initial control of virus replication in early infection and how these mould viral evolution.

CHAPTER 2

FULL-LENGTH SOUTH AFRICAN PRIMARY INFECTION SUBTYPE C STRAINS: COMPARISON TO VACCINE CONSTRUCTS

2.1 INTRODUCTION

High genetic diversity is a characteristic of HIV-1 and is one of the hurdles to developing an effective vaccine. The genetic distance in the envelope (Env) glycoprotein, for example, ranges from 25% to 35% between subtypes or clades, and up to 20% within a subtype (Gao *et al*., 2001; Gaschen *et al.*, 2002). Most vaccines currently in development focus on the induction of T cell immunity (Moore & Burton, 2004; Duerr *et al.*, 2006). As T cell-based vaccines target linear epitopes, it is likely that these vaccines will be vulnerable to the effects of genetic diversity.

To combat this problem of viral diversity, vaccine developers have devised different strategies to reduce the distance between the vaccine immunogen and the circulating viruses. As one approach, immunogens were developed based on sequences from primary infection and selected to be closest to the consensus of the dominant circulating subtype (Wagner *et al*., 1992; Williamson *et al.*, 2003; Kothe *et al*., 2006). Such immunogens were included in the modified vaccinia Ankara (MVA) and subtype C DNA vaccines developed by the University of Cape Town and the South African AIDS Vaccine Initiative (SAAVI) (Burgers *et al*., 2006; 2008; 2009; Shephard *et al*., 2008).

A second approach involved the synthesis of immunogens based on the consensus (Wagner *et al*., 1992; Kothe *et al*., 2007; Santra *et al*., 2008), ancestral (Kothe *et al*., 2006) or closest to the center of the tree (COT, Nickle *et al.*, 2007; Rolland *et al*., 2007; Frahm *et al*., 2008) viral sequences. This approach is estimated to reduce antigen

diversity by half for p17, p24 and gp160 (Gaschen *et al.*, 2002, reviewed by Korber *et al*., 2009). Consensus sequences represent the most common amino acid at each position whereas the ancestral state includes mutational changes associated with recent evolutionary events and lastly the COT algorithm combines the benefits of both the aforementioned (Nickle *et al.*, 2007; Rolland *et al*., 2007).

A third strategy is to include immunogens representative of different subtypes. This approach has been used by the Vaccine Research Centre (NIAID, NIH) who developed a DNA prime, adenovirus type 5 (Ad5) boost vaccine which included representative *env* genes from subtypes A, B and C (Catanzaro *et al*., 2006; 2007). A fourth approach, not yet in clinical trials, is mosaic immunogens designed to closely mimic natural sequences. Each immunogen consists of computationally derived multiple recombinant proteins that together constitute an optimal number of ninemers (9-mers/ linear epitopes) as targets for a wide range of HLA alleles (Fischer *et al*., 2007; Kong *et al*., 2009) (as discussed in Chapter 1, section 1.9.2; see Figure 1.13).

We were interested in characterizing full-length HIV genomes from recently infected individuals from South Africa. These genomes reflect the viral strains that are currently responsible for new infections and spread – therefore representing the viruses against which candidate vaccines need to show effectiveness. Although amplification of full-length genomes directly from plasma viral RNA was described more than ten years ago (Fang *et al*., 1996), this method was not widely used due to limited reproducibility. The development of reverse transcriptases capable of generating longer cDNAs and new heat-stable high fidelity polymerases produced amplicons of up to 40 kilobases (kb), leading to reliable amplification of full-length HIV-1 genomes from plasma viral RNA (Rousseau *et al.*, 2006). In this study we used the method developed by Rousseau *et al.*, (2006) to generate full-length sequences of HIV-1 subtype C strains from 23 participants recruited during primary infection.

The first vaccines tested in South Africa were the subtype C-based vaccine called AVX101 from AlphaVax (sponsored by HVTN) and HIVA.MVA; a subtype A based

vaccine sponsored by the International AIDS Vaccine Initiative (IAVI) (http://www.iavireport.org/trials-db/). More recently South Africa was involved in a phase IIb vaccine trial (HVTN 503/ Phambili trial) to evaluate the Merck, MRKAd5 HIV-1 Gag/Pol/Nef replication-defective trivalent vaccine (Priddy *et al.*, 2008, http://www.iavireport.org/trials-db/). This vaccine was based on HIV-1 subtype B consensus sequences.

The Phambili trial was prematurely terminated when the sister trial, STEP-HVTN 502 test-of-concept trial failed to prevent infection. One possible reason for the failure of the Merck trial includes the induction of insufficient breath of T cell responses, since each vaccine recipient showed responses to only one or two vaccine-derived epitopes (Korber *et al*., 2009), that may easily have been mismatched with respect to the infecting virus. Studies are currently underway to determine the reactive epitopes and the sequences of the infecting viral sequences in the vaccinees to investigate this hypothesis.

HIV-1 subtype C DNA and MVA candidate vaccines are currently being tested in phase I clinical trials in the USA and South Africa (SAAVI102/HVTN073). This vaccine incorporates South African HIV-1 subtype C consensus sequences for the *gag*, *pol* [RT], *tat* and *nef* genes, and *env* gene fragments from South African subtype C isolates sampled in 1998, chosen on the basis of their proximity to consensus sequences circulating at the time (Williamson *et al*., 2003). The construction of these candidate vaccines has been described (Burgers *et al.*, 2006; 2008) and the constructs were shown to be immunogenic in animal models (Shephard *et al*., 2008; Burgers *et al*., 2009).

In this chapter we characterized the full-length genomes of HIV-1 amplified from 23 recently infected women. To determine the potential of candidate vaccines to induce CD8 T cell responses against primary infection strains, we compared these sequences to three vaccine immunogens namely; Merck, SAAVI and the candidate mosaic immunogens.

2.2 MATERIALS AND METHODS

2.2.1 Cohorts

2.2.1.1 CAPRISA 002

The CAPRISA 002 acute infection study participants (n = 23, Table 2.1.) were recruited as recent seroconverters from a prospective cohort of mainly high-risk HIV-1 negative women established in 2004 (as well as other seroincidence cohorts) in Durban, KwaZulu-Natal, South Africa. Previous studies focusing on heterosexual HIV-1 transmissions in South African cities including Durban showed that the subtype C viruses from South Africa are genetically spread throughout the global distribution of subtype C viruses and follows a star-like phylogeny (Gordon *et al*., 2003; Williamson *et al*., 2003). Thus sampling from Durban only would represent the overall distribution of subtype C strains in South Africa. The CAPRISA 002 acute infection study cohort has been described in detail by van Loggerenberg *et al.*, 2008. Recent seroconverters were defined as those participants who had a reactive HIV-1 antibody test within 3 months of a previous negative test or had a positive HIV-1 RNA test in the absence of HIV-1 specific antibodies. Clinical assessment and sample collection were done at enrolment and weekly for the first month, then fortnightly for up to 3 months, then monthly until 12 months and thereafter quarterly.

Classification of HIV-1 infection stages (Table 2.1) was carried out using viral and antibody markers as previously reported (Fiebig *et al*., 2003; Keele *et al.* 2008). Briefly, individuals in HIV stages; I, were HIV RNA positive but p24 antigen negative, II, were HIV RNA and p24 antigen positive, III, were antibody-enzyme immuno assay (EIA) positive but Western blot negative, IV, were antibody-EIA positive with an indeterminate Western blot, V, were Western blot positive but with no p31 band and VI, were Western blot positive with a p31 band (Ruwayhida Thebus and Melissa-Rose Abrahams, unpublished data, University of Cape Town). From here on, all the sequences that were generated from the CAPRISA cohort have the prefix CAP. Viral strains reflect

transmissions that occurred between October 2004 and March 2006.

The study was approved by the ethics review boards of the Nelson R. Mandela School of Medicine at the University of KwaZulu-Natal and University of Cape Town (REC 025/2004). Informed consent was obtained from all the participants. Participants were terminated from the acute infection study on initiation of antiretroviral therapy.

2.2.1.2 Durban sex worker cohort

HIV negative sex workers were recruited with informed consent from truck stops along the main route connecting two major South African cities, in different provinces, namely Durban (in KwaZulu Natal) and Johannesburg (in Gauteng). Plasma samples (Table 2.1) were drawn from 12 participants that were originally recruited and monitored monthly as part of a UNAIDS-funded phase III vaginal microbicide (COL-1492) trial (van Damme *et al.*, 2002). Sequences generated from this cohort are identified by the prefix Du and reflect transmissions that occurred in 1998. The study was approved by the ethics review boards of the Nelson R. Mandela School of Medicine at the University of KwaZulu-Natal and the University of Cape Town (REC 137/95).

2.2.2 Diagnostic test and criteria

2.2.2.1 CAPRISA 002

HIV infection was detected monthly by screening using two rapid antibody tests Determine (Abbott Laboratories, Tokyo, Japan) and Capillus (Trinity Biotech, Jamestown, NY, USA). Unclotted blood was collected in EDTA from antibody positive individuals and the plasma stored at -70^{o}C. All the antibody negative samples were also screened by pooled PCR testing for HIV-1 RNA using the Ampliscreen v1.5 assay (Roche Diagnostics, Rotkreuz, Switzerland) and all identified positive samples were

Table 2.1 Study participants' sample and clinical data and 9 kb amplicon heterogeneity at time of sampling.

Participants	[a]Time post infection	Viral load (c/ml)	CD4 count (cells/mm^3)	[b]Laboratory stage	[c]Number of bands on HTA gel
CAP8	23 d	373000	360	V	1
CAP30	35 d	10200	989	V	1
CAP45	35 d	236000	974	V	ND
CAP61	57 d	610	389	VI	1
CAP63	34 d	202000	584	[d]V	2
CAP65	42 d	90800	243	VI	2
CAP69	14 d	15300	ND	I/II	1
CAP84	22 d	9140	636	V	2
CAP85	23 d	621000	419	V	2
CAP88	36 d	29400	963	VI	1
CAP174	28 d	474000	353	VI	1
CAP206	41 d	368000	365	VI	1
CAP210	36 d	468000	461	V	1
CAP217	67 d	75600	375	VI	1
CAP221	14 d	24300	ND	I/II	1
CAP228	53 d	2360	851	[d]VI	1
CAP229	48 d	126000	558	ND	1
CAP239	36 d	95800	845	V	2
CAP244	58 d	19200	557	VI	1
CAP248	62 d	55000	420	V	1
CAP255	54 d	196000	693	VI	1
CAP256	42 d	56500	689	VI	1
CAP257	49 d	276000	450	V	2
Du114	5 mo	14100	ND	ND	ND
Du123	2 yrs	6018	ND	ND	ND
Du172	10 mo	1042	ND	ND	ND
Du258	5 yrs	17774	601	ND	ND
Du301	10.5 mo	11797	ND	ND	ND
Du368	12 mo	2875	ND	ND	ND
Du442	4.5 yrs	1762	727	ND	ND
Du457	12.5 mo	7327	ND	ND	ND
Du467	2 yrs	85262	ND	ND	ND

CAP = CAPRISA 002 acute infection study participants, Du = Durban sex worker cohort study participants, ND = not done, d = days, mo = months, yrs = years, [a]Infection date was estimated as the midpoint between the last negative and first positive antibody test or as 14 days if the sample was PCR positive, antibody negative sample. ND = not done; [b]Fiebig *et al*., 2003; [c]No. of bands on heteroduplex tracking assay gel; [d]determined on samples from a week before.

confirmed using a quantitative RNA test (Roche Amplicor v1.5, Rotkreuz, Switzerland) and the BEP 2000 HIV enzyme immunoassay (Dade Behring, Marburg, Germany). HIV-1 viral loads were monitored using the COBAS AMPLICOR HIV-1 monitor v1.5 test (Roche Diagnostics, Rotkreuz, Switzerland). The viral loads and CD4 counts (Table 2.1) were done by the CAPRISA Acute Infection study team in Durban, University of KwaZulu-Natal (data kindly provided by Dr. Koleka Mlisana). HIV-1 specific antibody reactivity profiles in the early samples were determined using the GS Western blot assay following the manufacturer's protocol (Bio-Rad Laboratories, Redmond, WA, USA). Participants were referred to the CAPRISA antiretroviral treatment programme (van Loggerenberg *et al*., 2008) when their CD4 counts dropped below 350 cells/mm^3 at two consecutive timepoints and started treatment according to national guidelines when their CD4 counts were ≤ 200 cells/mm^3.

2.2.2.2 Du cohort

HIV-1 serostatus, CD4 counts and viral loads (Table 2.1) were determined as described previously (Grobler *et al.*, 2004). The recombinant HIV-1/HIV-2 ELIZAQ assay (Abbott Laboratories, Chicago, USA) was used for initial HIV diagnosis and positive tests were confirmed using the Vironostika HIV Uniform II micro-ELISA 4 system (Omnimed, United Kingdom; Grobler *et al*., 2004). Viral loads were determined by bDNA (version 3.0, Bayer, Emeryville, CA, USA). Diagnostic tests were performed at the National Institute for Communicable Diseases (Dr. Adrian Puren, NICD, Sandringham, Johannesburg, RSA) and CAPRISA, University of KwaZulu-Natal, RSA (data kindly provided by Dr. Koleka Mlisana).

2.2.3 RNA extraction and cDNA synthesis

Viral RNA was extracted manually from 140 to 500 µl plasma using the QIAamp® Viral RNA extraction kit and protocol (QIAGEN, Valencia, CA, USA;

Appendix C). RNA was eluted into 50 µl AVE buffer. The RNA was immediately used as template for complementary DNA (cDNA) synthesis using the Invitrogen SuperScript™ III Reverse Transcriptase (SSIII) System (Invitrogen, GmbH, Karlsruhe, Germany). Reaction master mixes (1 and 2, Table 2.2) were prepared as shown in Tables 2.2.1 and 2.2.2. Master mix 1 was incubated at 65°C for 5 minutes, cooled to 45°C and master mix 2 pre-warmed to 45°C was added. The reaction was incubated at 45°C for 1.5 hours before spiking with an additional µl of SSIII enzyme and continued incubation for another 1.5 hours.

Tables 2.2 cDNA synthesis reagent mixes for longer (~9 kb) transcripts

Table 2.2.1: Master mix 1

RNA	50 µl
dNTPs (20 mM)	2 µl
Oligo-dT (50 mM)	2 µl
Total volume	**54 µl per reaction**

Table 2.2.2: Master mix 2

5X buffer	20 µl
100 mM DTT	5 µl
H_2O	15 µl
RNase Inhibitor	2 µl
Superscript III RT	4 µl
Total volume	**46 µl per reaction**

Following reverse transcription the enzyme was inactivated at 70°C for 15 minutes. The RNA template was removed from the newly synthesized cDNA by incubation at 37°C for 20 minutes after addition of 2U *E. coli* RNase H. The cDNA was used as template for first round PCR amplification.

2.2.4 First and second round PCR amplification of ~9kb HIV-1 genome fragment

Near full-length HIV-1 genomes were amplified as a single fragment using a modified limiting dilution reverse transcription mediated nested PCR (Rousseau *et al.*, 2006) to obtained a third of positives for each replicate sample. Where necessary, more

replicates were tested at the determined cDNA dilution. Briefly, the Expand long template PCR system (Roche Diagnostics, Rotkreuz, Switzerland) used in the first and second round PCRs is composed of a unique mixture of thermostable *Taq* DNA polymerase and a proofreading polymerase. The primers used in the first and second round PCR are shown in Appendix A and Table 2.3. Two reaction mixtures (Table 2.3.1 and 2.3.2) were prepared, separated in reaction tubes with a thin wax layer which melt at 70°C thus allowing the reagents to mix and hot-starting the PCR reaction.

Tables 2.3 Near full-length HIV-1 genome (~9 kb) PCR reaction mixtures for 1st and 2nd round amplifications

Table 2.3.1: 1ST round PCR master mixes per reaction

Lower mix		**Upper mix**	
H2O	18.2 µl	H2O	x µl
20mM dNTPs	0.9 µl	10X long template buffer 1	5 µl
50pmoles/µl 1.U5C	0.3 µl	Expand long template enzyme	0.75 µl
50pmoles/µl 1.U5Cb	0.3 µl	cDNA	y µl
50pmoles/µl 1.3.3pIC	0.3 µl		
Total volume	**20 µl**	**Total volume**	**30 µl**

Table 2.3.2: 2nd round PCR master mixes per reaction

Lower mix		**Upper mix**	
H2O	18.5 µl	H2O	23.25 µl
20mM dNTPs	0.9 µl	10X buffer 1	5 µl
50pmoles/ul 2.U5C	0.3 µl	Expand enzyme	0.75 µl
50pmoles/ul 2.3.3pIC	0.3 µl	1st round product	1 µl
Total volume	**20 µl**	**Total volume**	**30 µl**

All samples were first screened in triplicate (or up to 5 replicates) using 1 µl of cDNA in the first round PCR. Two to 5 µl of second round PCR product was electrophoresed through a 1% agarose/ 1x TAE gel containing the DNA intercalating dye ethidium bromide (0.5 µg/ml, Appendix B). A 1kb DNA ladder (or a known 9kb amplicon) was used as a size marker. After electrophoresis the amplicons (9kb expected

size) are visualized under ultraviolet light (UV illumination).

2.2.5 Cloning of ~9kb fragments

Amplified full-length genomes were gel purified and cloned into the TOPO® XL PCR rapid ligation vector (Invitrogen, GmbH, Karlsruhe, Germany) following the manufacturer's instructions with minor modifications. Prior to cloning, the purified 9 kb PCR products were incubated at 70-72°C for 10 minutes in the presence of deoxynucleosides (dATP or dNTPs) and *Taq* enzyme to add 3'-end A-tails. The full transformation reaction was plated on a single LBagar plate containing kanamycin for selection (Appendix B). Transformation plates and liquid bacterial cultures were incubated at 30-32°C. Plasmid DNA was extracted from overnight bacterial cultures of putative clones using the QIAprep miniprep kit (QIAGEN, Valencia, CA, USA). Restriction enzyme *Mlu* I and *Xho* I double or *Not* I single digestions were used to identify positive clones carrying the desired insert size. Bacterial colonies containing the selected plasmids were cultured for midi-scale plasmid DNA isolation. The cloned HIV-1 genomes were sequenced in both directions using primer-walking.

We also controlled for potential sample mix-up or contamination during RNA preparation, PCR or cloning by inclusion of appropriate controls and physical separation of different steps as well as using phylogenetic analysis with tree drawing to identify any clustering of unrelated samples by including all sequences available in the laboratory as well a the appropriate reference sequences (http://www.hiv.lanl.gov/content/sequence/TUTORIALS/CONTAM/). In the event of suspected sample mix-up or contamination all samples and sequences in question were discarded.

2.2.6 Sequencing

DNA sequencing reactions were performed on 800 to 1600 ng of plasmid DNA

using the ABI PRISM Dye Terminator Cycle sequencing kit V3.1 (Applied Biosystems, Foster City, CA, USA). The reactions were slightly modified by using half the amount of prescribed sequencing enzyme mix in the presence of additional 5x Sequencing buffer. Whole genome sequencing primers (Appendix A) consisted of a combination described by Sanders-Buell *et al.* (1995) (some were re-optimised for subtype C), Rousseau *et al.* (2006) and in-house designed primers.

2.2.7 Complete genome assembly and quality of DNA sequence data

The CAPRISA sequence assembly pipeline tool (www.tools.caprisa.org) that employs the Phred, Phrap and Cross_Match software packages (Ewing & Green, 1998; Ewing *et al.*, 1998) was used to assemble full-length genome sequences. The assembled sequences and chromatograms were viewed and edited using Consed (Gordon *et al.*, 1998). The BushMan (Beta Version br1, developed by Ruby van Rooyen and Tulio de Oliveira, South African Bioinformatics Institute, University of the Western Cape, RSA, www.tools.caprisa.org) sequence/chromatogram assembly tool, a modification of the CAPRISA sequence assembly tool, was used to produce graphical output of the quality scores for single chromatograms and the assembled contig.

The Phred algorithm assigns the quality value for chromatograms where the average quality score for a chromatogram is the sum of the quality scores at each position divided by the sequence length. Each chromatogram was evaluated and sequence fragments categorized as either edge, bad or good read (Figure 2.1). Edge indicate average quality scores less than 35 at the start and end of a chromatogram, bad reads is a quality score of less than 35 in the middle of a chromatogram and good reads have a quality score of more than 35.

A quality evaluation of each chromatogram was done during assembly and a report generated with an average quality score of 35-40% being good or acceptable. The average quality score obtained for each assembled contig is the sum of the quality scores

obtained at each alignment position divided by the number of sequences that overlap at that position. The assembled contig depth coverage was also obtained and is the average number of overlapping reads at any alignment position. Sequence assemblies in Bushman were done with the help of Ruby van Rooyen (South African National Bioinformatics Institute [SANBI], University of the Western Cape, RSA).

2.2.8 Heteroduplex Tracking Assay (HTA)

First round whole genome PCR products corresponding to second round products selected for whole genome cloning were used as templates to amplify the V1V2 env gene fragment to asses the diversity of the viral population. A V1V2 *env* gene (HIV-1 subtype C, Du151) probe radioactively labeled with ^{35}S dATP through nick-translation with Klenow enzyme was used to identify complementary heteroduplex bands. Heteroduplex formation reactions were performed as follows: each reaction mixture consisted of 1 μl 10x HTA annealing buffer (Appendix B), 7 µl V1V2 PCR product and 2 μl labeled probe in a 10 μl final volume. Reaction mixtures were denatured at 95°C for 2 minutes and allowed to re-anneal for 5 minutes at room temperature for heteroduplex formation before addition of 2 μl loading dye and separation on a 5% non-denaturing polyacrylamide gel (Appendix B) (Kitrinos *et al.*, 2002, http://ubik.microbiol.washington.edu/HMA_/). Assays were preformed and data were kindly provided by H. Bredell, UCT.

2.2.9 Phylogenetic analysis

Reference whole genome HIV-1 subtype C and B sequences were downloaded from the HIV sequence database (http://hiv-web.lanl.gov). Multiple sequence alignment was done with Clustal X (Thompson *et al.*, 1994) and manual sequence editing and translation was done using BioEdit (Hall, 1999). Near full-length HIV-1 genome sequences (n = 419) were used for Neighbor-Joining phylogenetic tree construction in MEGA4 utilizing the Maximum Composite Likelihood model of nucleotide substitution

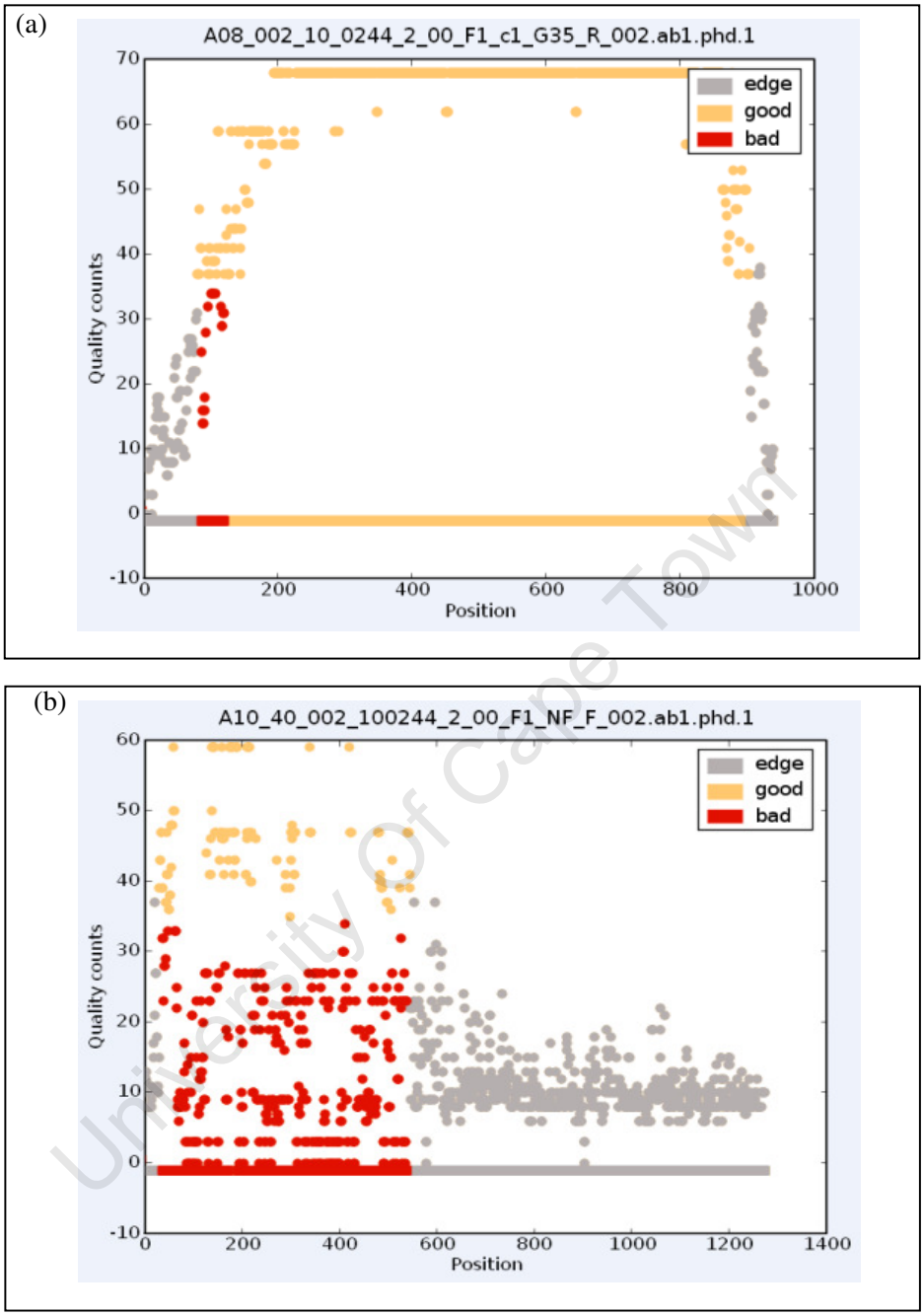


Figure 2.1 Graphical displays of the quality classifications for chromatograms (reads). Each dot represents one nucleotide position in a read. For overlapping reads the final quality score classification (solid line at bottom) for each nucleotide position in the contig is determined by the average score at that position. (a) good sequence data with ~25% of assembled reads classified as edge (average quality scores less than 35 at the ends of reads) or (b) bad data with >75% of the reads being classified as bad (average quality scores less than 35 in the middle of reads) or edge.

with gamma distribution ($\alpha = 1$) and 500 bootstrap replicates as a measure of confidence (Kumar *et al.*, 2004; Tamura *et al.*, 2007). The HXB2 subtype B strain was used as root. Pairwise DNA distances were also determined in MEGA4. The program, RDP3.34 (RDP, Martin & Rybicki, 2000) was also used to detect recombination in the full-length genome sequences. The automated settings for RDP, GENECONV and SISCAN previously described for the RDP2 program (Martin *et al.*, 2005) were used with default settings. The Bonferroni corrected p-value cutoff was 0.05.

Full-length genome sequences were split into the individual genes for gene-specific comparisons between primary infection subtype C viruses and the SAAVI/HVTN, Merck and mosaic protein immunogens. Pairwise amino acid distances were determined using the PROTDIST program based on the Dayhoff PAM 001 matrix, Kimura's distance and the Catagories distance models (Felsenstein, 1993) as described in BioEdit (Hall, 1999).

2.2.10 Vaccine constructs

The SAAVI DNA and MVA vaccine constructs express *gag* (Du422), reverse transcriptase (RT, Du422), *tat* (a consensus of Du151 and Du422), *nef* (Du151) and *env* (Du151) genes as described by Burgers *et al.* (2006; 2008). The HIV-1 subtype C strains (Du422 and Du151) on which the vaccine constructs are based were selected because of their high degree of similarity to the South African consensus subtype C at the time. Merck's Ad5 trivalent *gag/pol/nef* vaccine construct carry gene fragments from three subtype B strains [*gag* from CAM-1, RT and integrase (IN) from IIIB, and *nef* from JRFL, Priddy *et al.*, 2008]. All the genes included in vaccines were optimized for expression in humans.

Epitope coverage by test vaccines and newly derived consensus subtype C sequences was assessed using the epicover tool (http://www.hiv.lanl.gov/content/sequence/MOSAIC/epicover.html; Thurmond *et al.*,

2008). The epicover tool calculates the proportion of 9-mer peptide epitopes matched in the "vaccine" when compared to the background alignment (all naturally occurring subtype C proteins). The positional epitope coverage assessment tool (Posicover; http://www.hiv.lanl.gov/content/sequence/MOSAIC/posicover.html; Fischer *et al.*, 2007; Thurmond *et al.*, 2008) show position-by-position how well the "vaccine" 9-mer peptide epitopes map to the background of naturally occurring subtype C proteins. The average entropy was used to estimate the variability of amino acid residues at each alignment position using BioEdit (Hall, 1999).

2.2.11 Statistical analyses

The Kruskal-Wallis non-parametric test was used to identify differences between median amino acid distances for all vaccine immunogens versus primary infection virus datasets, with Dunn's correction for multiple comparisons. A p-value <0.05 was taken as significant. These statistical tests were performed using GraphPad Prism® 5.0 (GraphPad Software Inc., CA, USA).

2.3 RESULTS

2.3.1 Participant data

The participants from the CAPRISA cohort with primary infection (CAP, n = 23) were infected for an estimated median of 36 days (range 14 to 67 days), with date of infection estimated as the midpoint between the last seronegative and first seropositive sample, or 14 days if diagnostic tests were antibody negative but RNA-positive. Most participants had high vireamia typical of primary infection with a median of 90 800 viral RNA copies/ml (range 610 to 621 000 copies/ml). The median CD4 counts of these individuals were 557 cells/mm^3 (range 243 to 989 cells/mm^3). The Durban sex worker cohort samples (Du, n = 9) were from five individuals with early infection (5 to 12.5

months post infection), and 4 with chronic infection, having been infected for between 2 to 5 years. The median viral load in the Durban cohort was 7 327 copies/ml (range 1 042 to 85 262 copies/ml). CD4 counts were available for only two participants (Table 2.l).

2.3.2 HIV-1 near full-length genome amplification, cloning and quality of sequence data generated

Our strategy for amplification of near full-length HIV-1 genomes was as follows. Firstly, samples were amplified in triplicate using 1 μl of cDNA template. If all or most of the replicate samples showed a sharp ~1.7 kb band (Fig 2.2a; lanes 5, 6, 7), a mixture of 9 and 1.7 kb bands (Figure 2.2b, lanes 8, 9, 10) or a 9 kb band size for all replicates (Figure 2.2), limiting dilution of cDNA samples was performed. If all replicate samples were negative (Figure 2.2a, lanes 20, 21, 23), the amount of template was increased and the replicates of 2, 3 and 5 µl cDNA was amplified.

A total of 46 plasmid clones containing near the full-length genomes were generated from 23 individuals from the CAPRISA cohort with primary infection (n = 38); and from eight Du cohort participants with varying times post infection (n = 8). For a single Du participant (Du457), only *gag-pol* and *env* gene fragments were generated. The heterogeneity of full-length amplicons generated from CAPRISA primary infection samples was assessed by heteroduplex tracking assay (performed by H. Bredell, UCT).

Genetic homogeneity in V1V2 of the templates, indicative of amplification from a single genome, was confirmed for 16 of the 23 full-length genome amplicons. Two variants were identified from the remaining seven samples (Table 2.1). Thus, in summary the full-length genomes were generated from a homogenous viral population in the majority of participants, with a heterogeneous inoculum in seven.

Assessment of the quality of assembled whole genome sequences was performed for CAPRISA 002 cohort sequences. An acceptable average quality score of between

79.8 and 89.6 was obtained for all full-length genome sequences evaluated with an average depth coverage from 3.5 to 7.1 (Appendix D1). Thus the sequence data generated for this study was of high quality with good overlapping sequence reads on both strands.

In the following sections sequences generated from the eight Du cohort participants infected in 1998 were added to the global pool of subtype C viruses obtained from the HIV Sequence database, whereas the CAPRISA sequences represent the primary infection viruses.

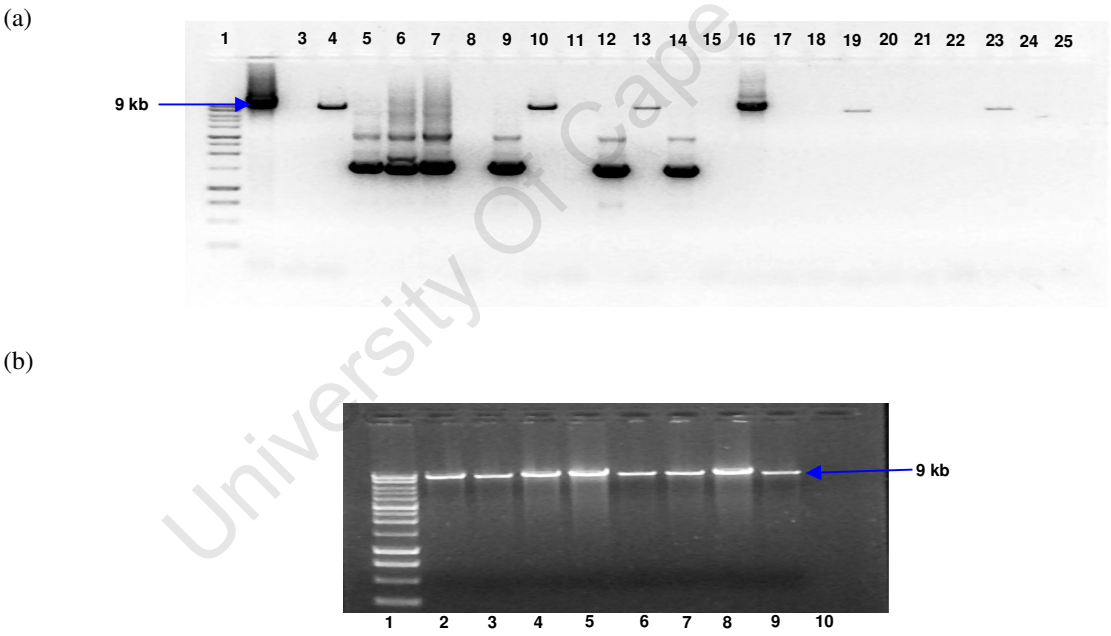


Figure 2.2 Ethidium bromide stained and UV illuminated 1% agarose gels:
(a) 2nd round whole genome PCR products showing in lane 1: 1 kb DNA ladder where top band (double band) = 8-10 kb, lanes 2-4: CAP8 , lanes 5-7: CAP30 , lanes 8-10: CAP45 , lanes 11-13: CAP61 , lanes 14-16: CAP63 , lanes 17-19: CAP84 , lanes 20-22: CAP88 and lanes 23-25: CAP210. (b) Gel-purified whole genome PCR products showing in lane 1: 1 kb DNA ladder, lanes 2 ; 4 ; 6 ; 8: Du258, lanes 3 ; 5 ; 7 ; 9 : Du123 and lane 10 : no sample

2.3.3 Phylogenetic analysis

All sequences were classified as HIV-1 subtype C throughout the genome with no evidence of inter-subtype recombination. The phylogenetic relationship between subtype C sequences generated from this study was compared to other full-length subtype C sequences from South Africa and sequences representing the global subtype C epidemic (Figure 2.3). South African viruses (all unlabelled tips and open red boxes) are spread throughout the starburst-like phylogeny. The primary infection strains from this study (filled red boxes) did not cluster together, indicating that we were not sampling from a single transmission network. One sequence from the CAPRISA cohort, CAP210 (gray line), was an outlier of the South African sequences and grouped with the Brazilian (pink lines) subcluster with 81% bootstrap support. This sequence was also closer in the median pairwise DNA distance to the Brazilian sequences than to the South Africa primary infection strains (7.6% compared to 9.6% respectively).

HIV only became apparent in the heterosexual population in South Africa in the late 1980s/early 1990s, and the high diversity star-like phylogeny of South African sequences represents the multiple introductions and subsequent spread of the virus in South Africa. This is in contrast to countries such as India (turquoise) and Brazil (pink), where there is evidence of founder effects, with sequences from these regions each forming a distinct monophyletic cluster. This pattern suggests a single introduction of subtype C into these countries. As a consequence, the subtype C epidemic in South Africa has higher overall diversity compared to the subtype C epidemics of India and Brazil, with median pairwise distances of 7.8% for South Africa sequences compared to 6.1 % for both India and Brazil (Figure 2.4). However the median pairwise DNA distances for the South African sequences (7.8%) was less than the median pairwise distances between virus sampled from Botswana (8.5%), Tanzania (8.7%), and Zambia (9.2%) as well as the CAPRISA primary infection viruses (ZA-PI, 9.5%) from this study. This lower median pairwise DNA distance within the South African (ZA) viruses may be the result of recent epidemiologically linked transmissions within the data, indicated by nine well supported clusters (Figure 2.3, red branches). The CAPRISA subtype C primary

infection viruses from this study represent the diversity generally seen in subtype C virus epidemics without significant evidence of founder effects.

2.3.4 Amino acid diversity of vaccine strains compared to subtype C primary infection strains

The high degree of genetic diversity found in HIV-1 is one of the major challenges to developing an effective vaccine. Our laboratory has been involved in the design and preclinical testing of the SAAVI/HVTN subtype C vaccine, and we are also investigating breakthrough infections in participants from the suspended HVTN503 (Phambili) phase IIb clinical trial. Here we compared the antigenic diversity of primary infection sequences to immunogen sequences contained in (i) the SAAVI DNA/MVA vaccines, (ii) the Merck Ad5 vaccine, and (iii) mosaic antigen vaccines described by Fischer *et al.* (2007). The SAAVI immunogen was derived from recently infected individuals (Du422 and Du151) and selection was based on closest to a subtype C consensus sequence at the time (Williamson *et al.,* 2003). This vaccine expresses a polyprotein comprising Gag, RT, Tat and Nef, as well as Env (Burgers *et al*., 2006). The Merck Ad5 vaccine expresses Gag, RT-IN and Nef from subtype B (Priddy *et al*., 2008). The mosaic immunogens were selected for comparison as they have been included in a number of vaccines in pre-clinical development. These proposed vaccines will include expression of four recombinant proteins (mosaic proteins) each for HIV-1 subtype B (mosB), C (mos C) or M group Gag, RT, Nef core (most conserved) region and Gp150 proteins (Fischer *et al*., 2007).

We initially investigated the potential coverage of candidate vaccines by determining the median amino acid distances between the CAPRISA subtype C primary infection viruses and the vaccine antigens. Gag, RT and Nef were compared for all vaccines, however as the Merck Ad5 vaccine did not contain *env*, this comparison was done only for the SAAVI and mosaic antigens. The SAAVI immunogen had the least median amino acid distance to the primary infection viruses with median amino acid

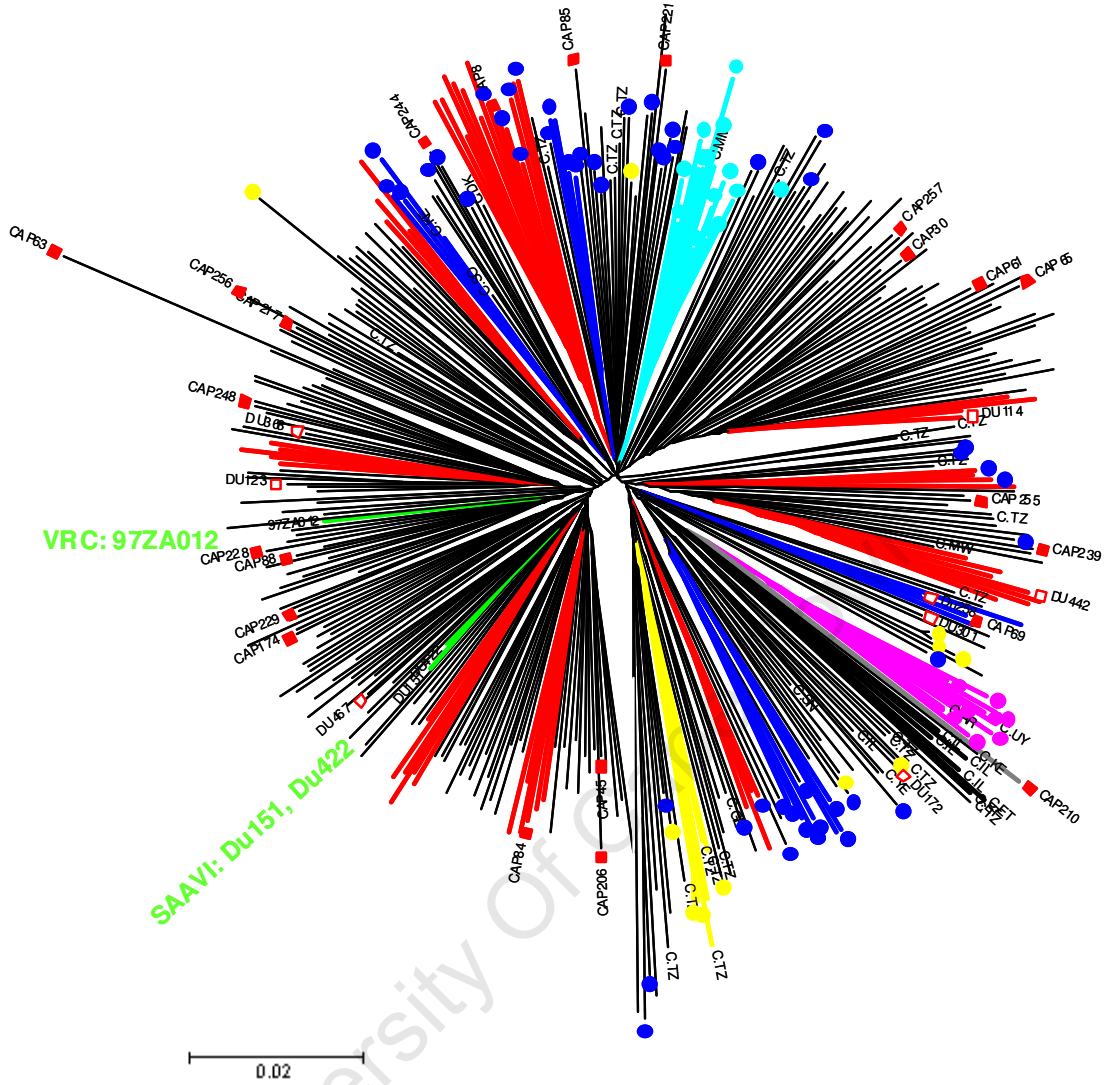


Figure 2.3 Radial display of a Neighbor-Joining tree of near full-length HIV-1 subtype C viruses (n = 418). Primary infection subtype C sequences (filled red boxes) and the Durban sexworker cohort sequences from this study (open red boxes) with subtype C viruses from the Los Alamos HIV sequence database. Thick colored branches = supported subclusters with >75% bootstrap value: Red: South Africa, Blue: Botswana, Pink: Brazil, Turquoise: India, Black: Israel-Ethiopia, Yellow: Zambia-Tanzania. The green lines = viruses used in vaccine strategies and all unlabelled tips also represent viruses from South Africa. Scale bar = 2%. The subtype B strain HXB2 was use as root and 500 bootstrap replicates was used as measure of confidence.

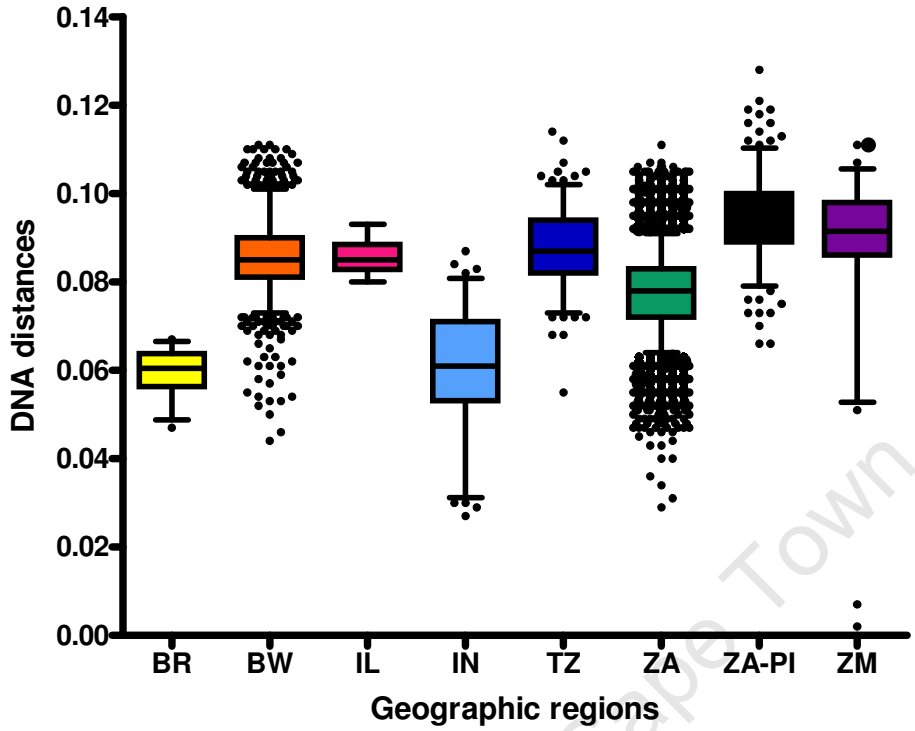


Figure 2.4 Median DNA distances within each group of HIV-1 subtype C viruses from various geographic regions. BR: Brazil, BW: Botswana, IL: Israel, IN: India, TZ: Tanzania, ZA: South Africa, ZA-PI: Primary Infection viruses from South Africa and ZM: Zambia. The 5-95% percentile indicated by Box and Wiskers plot.

distances of 6.7%, 6.2%, 8.6% and 17.5% for Gag, RT, Nef and Env respectively (Figure 2.5). The next best matched immunogen was mosC with median amino acid distance for Gag, RT, Nef and Env of 9.7%, 7.2%, 11.2% and 17.8%.

Although there was a significant difference in median amino acid distance for Gag and Nef immunogens included in the SAAVI compared to the mosC immunogens, this difference was not significant for RT and Env (Figure 2.5). As expected, the immunogens which were not matched to subtype C had the greatest amino acid distance from the primary infection viruses with mosM immunogen being a better match than the subtype B immunogens. The subtype C based immunogen had significantly lower amino

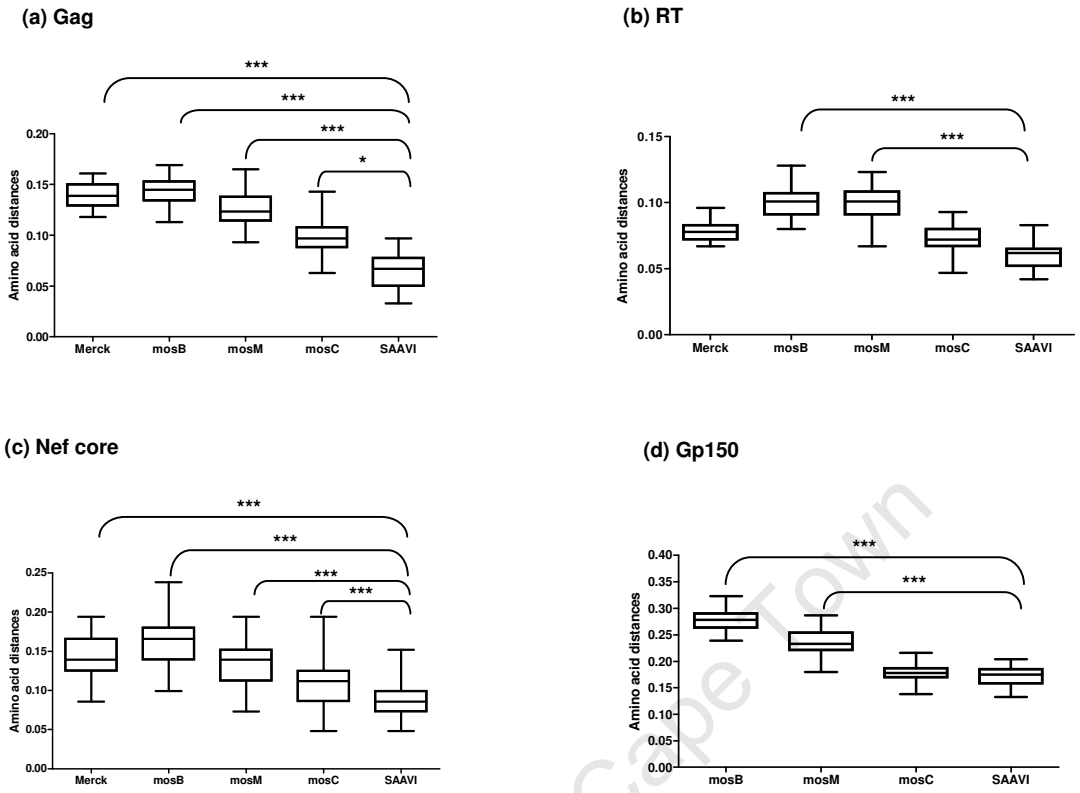


Figure 2.5 Amino acid distances between subtype C primary infection viruses and vaccine antigens from (a) Gag, (b) reverse transcriptase [RT], (c) Nef core and (d) Gp150 constructs. The box indicates median and 25th-75th percentiles, and the whiskers indicate minimum and maximum amino acid distances. Comparisons were made using the Kruskal-Wallis test with Dunn's correction for multiple comparisons. An asterisk (*) indicates $p<0.05$; *** indicates $p<0.001$.

acid distances for all genes compared to the non-subtype matched immunogens. The amino acid distances for example, of Merck Ad5 B immunogen tested in the Phase IIb clinical trial in South Africa was 13.9% (for Gag and Nef), and 7.8% (for RT) different from currently circulating viruses.

2.3.5 Comparison of CTL epitope coverage by vaccine strains

Amino acid similarity is a crude estimate of potential vaccine coverage within a population, and a more meaningful way of exploring vaccine coverage is a comparison of

9-mers, the typical length of a CD8 T cell epitope. Here we estimate, and compare, the proportion of 9-mer CD8 T cell epitope coverage provided by immunogen sequences included in the SAAVI DNA/MVA, Merck Ad5 and the mosaic candidate vaccines to both primary infection strains generated in this study, as well as subtype C viruses from the database. The analysis was restricted to antigens that are common in the candidate vaccines of interest, namely Gag, RT, Nef core region (all immunogens) and Gp150 proteins (SAAVI and mosaic immunogens). Utilizing tools provided on the HIV Sequence database (http://www.hiv.lanl.gov/content/sequence/HIV/HIVTools.html), we estimated the percentage of 9-mers that were perfectly matched between the immunogens and the viral sequences analysed. This should be considered a minimum value, as not all 9-mers will contain T cell epitopes, and single changes within an epitope will not always affect T cell recognition (Valentine *et al*., 2008).

The analysis included epitope coverage of the vaccine immunogens compared to the currently circulating subtype C viruses in South Africa (primary infection sequences, n = 23, except for Gag where n = 22), as well as to subtype C sequences from the database (subtype C database viruses, n = 407 for Gag, n = 407 for RT, n = 403 for Nef and n = 406 for Gp150). The database sequences are comprised predominantly of sequences from individuals with chronic infection from Durban, South Africa (Kiepiela *et al.,* 2004; Matthews *et al*., 2008). This analysis did not include coverage against other subtypes (reviewed by Korber *et al.*, 2009). Overall, similar degrees of 9-mer epitope coverage were obtained against primary infection subtype C viruses (CA) and the global subtype C virus pool (C, HIV database sequences) for all immunogens, with the exception of Nef (Figure 2.6).

For Gag immunogens, the best predicted coverage of perfectly matched 9-mers was obtained with the mosC mosaic immunogen (79% for C + CA, referred to as 'all Cs' in the text following), with the mosM (M group mosaic) performing nearly as well, at 76% coverage (Figure 2.6a). Although the SAAVI subtype C Gag immunogen was most similar in amino acid distance to the primary infection subtype C viruses, the predicted epitope coverage of 60% was markedly lower (19%) than for the mosC

immunogen. A similar degree of coverage against all Cs was estimated for the SAAVI Gag immunogen and the mosaic B Gag immunogen (59% and 58%, respectively). Epitope coverage by the Merck Gag immunogen was only 34% against all Cs. Thus it can be predicted that a mosM vaccine will perform better against rare non-C subtypes that are circulating in South Africa than a mosC immunogen.

To be expected, the conserved nature of HIV-1 RT resulted in less impact of on epitope coverage, with similar proportions of 9-mer epitope coverage obtained with the SAAVI (61%) and Merck (57%) immunogens (Figure 2.6b). However, the mosaic M group and subtype C derived RT immunogens still provided markedly better coverage (~90%) than the SAAVI subtype C RT immunogen. There was no difference in epitope coverage for the primary infection sequences compared to the database sequences.

The Nef core region was specifically selected for inclusion as a mosaic immunogen as this was the most conserved region of the Nef protein (Fischer *et al*., 2007; Korber *et al*., 2009). Here we found that the Nef core region of the SAAVI vaccine was predicted to provide only 48% coverage for perfectly matched 9-mers, whereas the Merck vaccine only had 31% of 9-mers perfectly conserved between the vaccine and the global subtype C pool. Again, the mosaic immunogens provided a marked improvement in predicted epitope coverage, with an increase to 60% by mosB, 78% by mosM and 85% by mosC. Although there was no significant difference between the median amino acid distances for SAAVI and mosC immunogens against primary infection viruses, the mosC immunogen provide 85% coverage compared to 48% coverage by the SAAVI Nef core immunogen. Here we also observed a difference in epitope coverage obtained against the global subtype C pool (C) and currently circulating viruses (CA) with the SAAVI (48% and 46% respectively), Merck (31% and 26% respectively) and mosB (60% and 57% respectively) Nef immunogens.

For the variable Gp150 envelope glycoprotein, the predicted epitope coverage for the SAAVI immunogen was poor (31%) when compared to the mosaic Gp150 immunogens for subtype C (58%) and the M group (50%). It is clear that the more

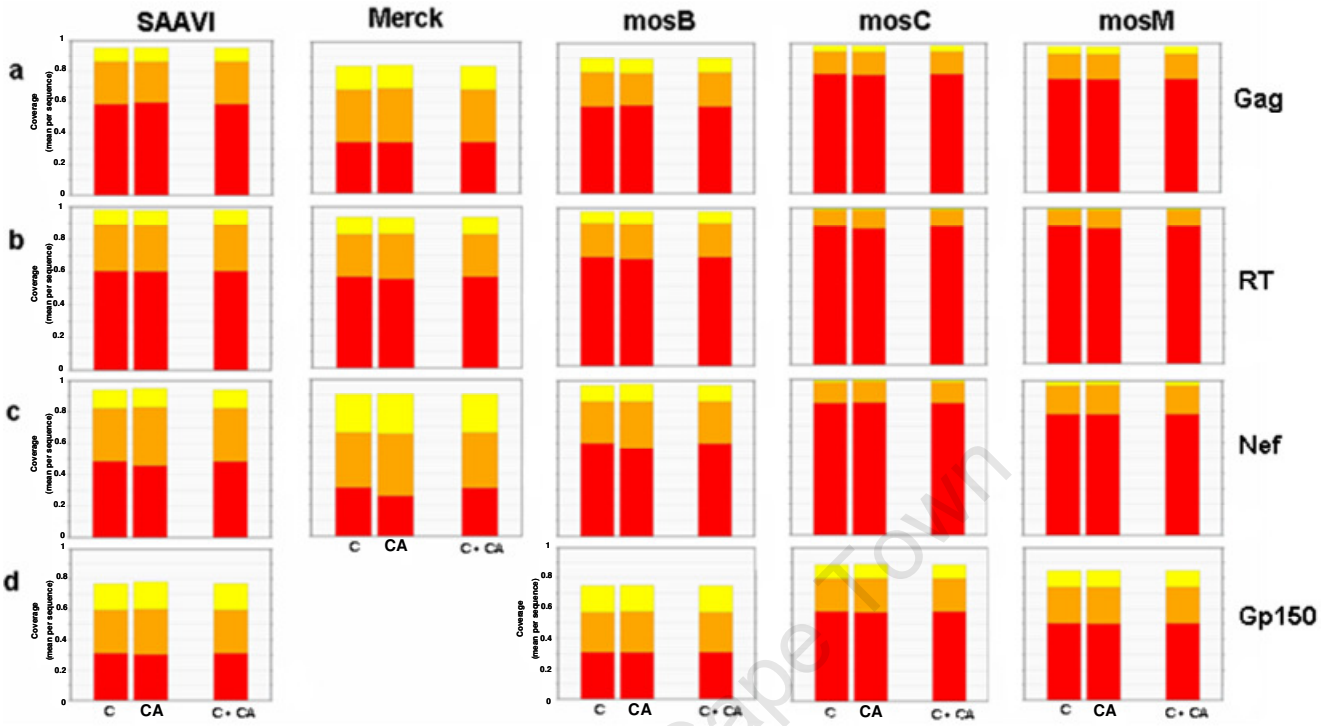


Figure 2.6 Predicted total epitope (9-mer) coverage of the global pool (C = HIV sequence database) and primary infection (CA) HIV-1 subtype C viruses by (a) Gag, (b) RT and (c) Nef core and (d) Gp150 immunogens from SAAVI, Merck and mosaic vaccine immunogens. The proportions of perfectly matched 9-mers (0 mismatches) are shown in red, 1 mismatch in orange and 2 mismatches in yellow.

variable HIV-1 proteins (Nef and Gp150) are potentially poorer immunogens compared with Gag and RT for the induction of cross-reactive cell mediated immune responses by vaccines, as immunogens derived from these variable proteins presents predicted CTL epitopes that are more frequently mismatched to the circulating viruses. This is especially pronounced for Gp150, where a lower predicted epitope coverage was also obtained for the mosaic immunogens, with a maximum coverage of 58% which is almost 20% lower than for Gag, Nef and RT.

To identify the regions associated with poor coverage, we assessed epitope coverage at each amino acid position for the variable Nef and Gp150 HIV-1 proteins

(Thurmond *et al.*, 2008). We found that for both the Gp150 and Nef core immunogens, regions of poor coverage map to high entropy (>0.2) regions of the HIV-1 subtype C proteins (Figure 2.7). A comparison between the epitope coverage across Nef and Gp150 show that the mosaic immunogens, particularly mosC, have increased coverage in these variable regions of Nef (Figure 2.7) and Gp150 (data not shown).

In summary, we observed that while subtype-matched vaccines (SAAVI DNA/MVA subtype C vaccine) have greater sequence and epitope coverage than subtype mismatched vaccines (Merck subtype B Ad5 vaccine), the increase in potential T cell coverage using the newer mosaic immunogen design strategy shows a marked increase in epitope coverage. However, it is clear that even with mosaic immunogens subtype-specific antigens for variable proteins (as shown here for subtype C) will be best suited in populations where a single subtype dominates the epidemic, followed by the M group specific mosaic immunogens.

We did not find any differences in estimated epitope coverage for global subtype C viruses vs. the primary infection viruses, except in the Nef core region. The primary infection viruses displayed overall increased entropy in the Nef core region compared to the global pool of subtype C viruses. The possible reason for differences in primary and chronic infection sequences will be explored further in chapter 3.

2.4 DISCUSSION

Genetic drift is a feature of the HIV-1 epidemic, and since vaccines need to target recently transmitted viruses, we were interested in characterizing full-length HIV genomes from individuals with primary HIV-1 infection. We sequenced a total of 32 full-length subtype C genomes, of which 23 were from primary infection, representing the largest available number of full-length genomes from primary infection. All viruses belonged to HIV-1 subtype C, with no evidence of inter-subtype recombination. While the South African subtype C epidemic largely follows a star-like phylogeny, a number of

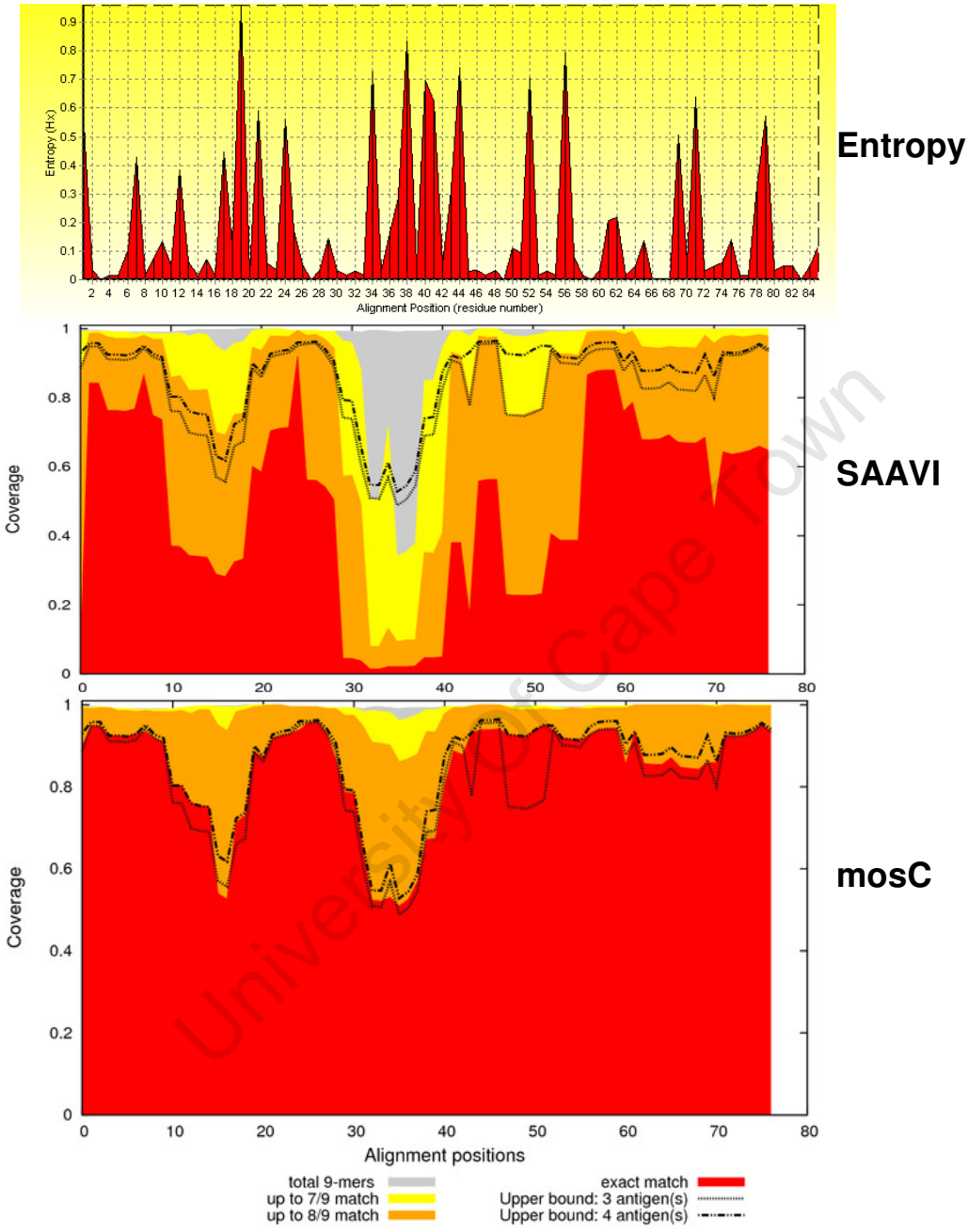


Figure 2.7 Comparison of the positional entropy for subtype C viruses with positional 9-mer epitope coverage against the highly variable Nef core from subtype C viruses by the SAAVI and mosaic mosC immunogens. The proportions of perfectly matched 9-mers are shown in red, 1 mismatch in orange and 2 mismatches in yellow.

subtype C supported clusters were identified, suggesting some recent historical linkage between viruses. This may be the reason that the overall genetic distance between the South Africa sequences, whilst greater than founder epidemics such as those occurring in India and Brazil, was lower than the median DNA distance for sequences from Zambia, for example.

However, the subtype C viruses from primary infection from this study showed no evidence of founder effects, indicating that we were not sampling from transmission networks, and they reflected the general diversity of the global subtype C epidemic. Interestingly, one sequence was genetically distinct from the other sequences from South Africa, grouping with the Brazilian subcluster.

Candidate HIV vaccines that are being tested in South Africa contain immunogens based on natural sequences with close homology to the subtype consensus sequence. The SAAVI subtype C immunogen, which is currently in Phase I clinical trial, was selected based on closest to the consensus C sequence. Our analysis shows that this candidate vaccine was predicted to provide better epitope coverage against subtype C viruses than the subtype B Merck immunogen that was recently tested in a Phase IIb trial in this country. The SAAVI subtype C immunogen provide predicted epitope coverage of 61%, 48% and 53% for RT, Nef and Gag, respectively compared to 57%, 31% and 30% respectively for RT, Nef and Gag immunogens from the Merck Ad5 subtype B based vaccine. We would therefore predict that the SAAVI subtype C vaccine would generate a greater number of matched T cell responses against subtype C viruses in South African vaccinees who would receive the strain-specific SAAVI vaccines compared to the Merck subtype B vaccine, which may translate to greater control of viral replication or protection.

This demonstrates clearly the advantages of a clade-matched vaccine in populations where a specific subtype predominantly circulates. However, the overall, epitope coverage of the SAAVI vaccine based on natural strains was relatively low, ranging from 31% for Gp150 and 61% for RT. Despite the extent of predicted epitope

mismatches, studies within subtype C-infected populations have shown extensive reactivity to peptides based on the SAAVI vaccine sequences (Masemola *et al*., 2004). They showed that most of the 46 participants from southern Africa recognized Nef (87%) followed by Gag (83%), Pol (75%), Env (63%), Vif (28%), Vpr (22%), Tat (17%), Rev (15%) and Vpu (2%).

A new approach to vaccine design using the mosaic immunogen strategy of Fischer *et al.* (2007) markedly increases epitope coverage compared to vaccines based on natural proteins. For example, a subtype C Gag vaccine against subtype C viruses (contained in the HIV sequence database) had a predicted coverage of 80% for perfectly matched 9-mers. Here we also show that even against the primary infection subtype C viruses, these mosaic immunogens (mosC) provide an 80% - 90% matched 9-mer epitope coverage for Gag, RT and the more conserved Nef core region. Epitope coverage for the more variable Gp150 protein was 60%.

Mosaic immunogens potentially provide superior immunological coverage against the high genetic diversity of HIV-1. In recent studies on non-human primates Barouch *et al*. (2010) showed that mosaic immunogens with a valency of two can provide 3 times higher numbers/magnitude of T cell responses, especially cytotoxic T cell responses than consensus or natural immunogens. Similarly, Santra *et al*. (2010) who used mosaic immunogens with a valency of four also showed superior T cell responses to those induced by consensus or natural immunogens. These mosaic designs were based on 9-mer optimal epitopes because it is most common length of MHC class I peptides. The mosaic peptides were generated through computer learning techniques to include virtual recombinants with high frequencies of optimal 9-mer coverage and to exclude rare 9-mers and those with unnatural amino acid sequence. An increase in the number of peptides (valency) in vaccine cocktails increased total epitope coverage (Fischer *et al*., 2007). However no real increase in benefit was observed in cocktails with a valency of 6 compared to 4, as the increased valency was predominantly associated with an increase in the number of very rare epitopes and there is concern that these rare epitope responses may reduce responses to more common CTL epitopes (Fischer *et al*., 2007).

One of the concerns with the failure of the Merck vaccine in regions where subtype B is the major circulating subtype is that the vaccine immunogens did not induce T cell responses with sufficient breadth to recognize the infecting variant and prevent or control infection (McElrath *et al.*, 2008); in fact, most vaccinees mounted only one or two epitope responses (Korber *et al*., 2009). The goal for vaccine development is to combine vaccination strategies that induce a large number of T cell responses (immunogenicity breadth) with immunogens that incorporate the diversity of HIV-1 (coverage breadth), maximising the chances that vaccine-induced T cell responses will match the infecting variants. Santra *et al.* (2008) showed that a consensus based M group immunogen induced broad cross-clade cell mediated immune responses in a rhesus macaque model. In a mouse model, a greater breadth of T cell responses was observed with a mosaic envelope-based DNA vaccine (Kong *et al.*, 2009). Thus mosaic immunogen design strategies in preclinical development are anticipated to induce responses of greater breadth, and also provide greater coverage against subtype or M group-specific antigens.

Addressing the high degree of genetic diversity of HIV-1 to develop an effective vaccine is one of the key questions facing HIV vaccinologists today. Here we show that the immunogens based on viral sequences are more likely to elicit cross reactive CTL responses if matched to the subtype circulating in the population. However we also illustrate that despite the subtype C SAAVI immunogens showing the closest amino acid similarity to primary infection subtype C viruses, the mosaic immunogens are predicted to provide much greater coverage than vaccines that have been in clinical trial. Fischer *et al*. (2007) showed that a single mosaic immunogen gives similar coverage compared to consensus-derived immunogens or the best natural strains (shown to be Du422 for subtype C; reviewed by Korber *et al*., 2009). Thus, the increased coverage with mosaic immunogens is in part linked to the valency (1 versus 4 mosaic proteins per immunogen), which increases the diversity to target additional high frequency or common T cell epitopes. We observed that for Gag, two mosaic proteins give the same degree of epitope coverage as the closest-to-consensus SAAVI immunogen (Appendix D2). Epitope coverage by COT-derived immunogens was increased with the addition of high

frequency 9-mer epitopes (Rolland *et al*., 2007; Nickle *et al*., 2007; Nickle *et al*., 2008). A similar strategy could be employed to increase the epitope coverage by immunogens in the SAAVI vaccine.

CHAPTER 3

ADAPTIVE CHANGES IN HIV-1 SUBTYPE C DURING EARLY INFECTION

3.1 INTRODUCTION

It is well established that HIV populations experience extreme bottlenecks during sexual transmission (Derdeyn *et al.*, 2004; Wolfs *et al.*, 1992) with approximately 80% to 90% of infections being a consequence of a single transmitted variant (Abrahams *et al.*, 2009; Haaland *et al.*, 2009; Keele *et al.*, 2008). The strongest evidence that "transmission sieves" have been a major factor in HIV evolution is that, relative to viruses sampled during chronic infections, recently transmitted HIV-1 subtype C genetic variants are in general more sensitive to neutralization and tend to have both shorter V1V2 and V1V4 loops and fewer glycosylation sites (Derdeyn *et al.*, 2004; Li *et al.*, 2006; Rong *et al.*, 2007). Differences in sites under selection have also been identified between the envelope glycoproteins (gp41) of viruses from the primary and chronic infection phases suggesting the existence of different selective pressures during these different infection phases (Bandawe *et al.*, 2008).

However, outside of studies on *env*, there is limited information on features which distinguish recently transmitted viruses from those found in chronic infections. Viruses from chronic infections have usually undergone strong cytotoxic T lymphocyte (CTL) driven selection pressures and are therefore expected to have accumulated immune escape mutations. These CTL escape mutations, while highly adaptive within the context of immune environments where hosts have the appropriate HLA-alleles (Brumme *et al.*, 2007; Brumme, *et al.*, 2008; Kelleher *et al.*, 2001; Rousseau *et al.*, 2008), can also seriously diminish viral replicative fitness (Allen *et al.*, 2005; Brockman *et al.*, 2007; Liu *et al.*, 2006; Liu *et al.*, 2007; Martinez-Picado *et al.*, 2006; Miura *et al.*, 2009). Although

there is some evidence from mother to child transmission pair studies that fitter virus variants are selectively transmitted (Kong *et al.*, 2008), other studies have shown that genetic variants that carry attenuating CTL escape mutations are also transmitted (Chopera *et al.*, 2008;Goepfert *et al.*, 2008).

Following transmission, viruses generally accumulate both immune evasion mutations and reversion mutations that recoup replicative fitness losses experienced due to deleterious escape mutations accrued in previous hosts (Brumme *et al.*, 2008; Leslie *et al.*, 2004; Liu *et al.*, 2007; Matthews *et al.*, 2008; Rousseau *et al.*, 2008). The rate at which such reversion mutations occur is most likely dependant on the magnitude of their effects on replicative fitness (Brumme *et al.*, 2008; Matthews *et al.*, 2008). The clinical importance of viruses carrying attenuating CTL escape mutations is that the recipients of such viruses will, in some cases at least, have lower set-point viral loads, higher CD4 counts and, possibly, better survival prospects (Chopera *et al.*, 2008;Goepfert *et al.*, 2008).

A successful HIV vaccine will need to effectively combat viruses during the earliest stages of infection. Identifying the specific genetic features that might predispose particular HIV variants to being more transmissible than others, and understanding the evolutionary processes at play during the early evolution of successfully transmitted variants, are therefore both important for defining potential targets for vaccine induced immunity. As events during acute HIV-1 infections are thought to have a disproportionately large influence on both long-term disease outcomes (de Wolf *et al.*, 1997; Lavreys *et al.*, 2006) and global HIV evolution in general (Rambaut *et al.*, 2004), understanding the transmission bottleneck and the subsequent evolution of successfully transmitted variants are probably key to identifying and understanding the viral and host determinants of HIV pathogenesis.

To identify genetic features that are characteristic of recently transmitted viruses, we developed a phylogeny and recombination aware method to compare amino acid mutation spectra between groups of sequences. We used this approach to identify amino

acid sites that differentiated between full-length HIV-1 subtype C genomes sampled during primary and chronic infections. We then examined longitudinally sampled sequences to infer the processes that might underlie the amino acid frequency differences observed in viruses from the different infection phases.

3.2 MATERIALS AND METHODS

3.2.1 Primary infection study subjects

Here we only included 20 of the 23 participants from the CAPRISA acute infection cohort described in chapter 2 (see 2.2.1) who were classified as Fiebig stage V and VI and whose full genome sequences covers the *gag* through to the *nef* gene.

3.2.2 Assembly of a chronic infection dataset

We assembled a reference HIV-1 subtype C chronic infection dataset consisting of 63 publicly available subtype C full-length sequences (Kiepiela *et al.*, 2004);http://hiv.lanl.gov/components/sequence/HIV/search/search.html) and we also generated an additional 3 full-length sequences from participants of in the Durban sex-worker cohort (Van Damme *et al.*, 2002). Similar to the sequences from primary infection, the chronic infection sequences were obtained from heterosexually infected women with the same ethnic background (Xhosa/ Zulu) and from the same geographic location (KwaZulu-Natal, South Africa).

Sequences from participants with AIDS as defined by CD4+ counts less than 200 cells per mm^3 were excluded. In addition, sequences from participants with viral loads >200 000 copies/ml were also excluded to minimize inadvertent inclusion of primary infection sequences in the chronic infection dataset.

3.2.3 Whole genome amplification and DNA sequencing

Full-length genome sequences were generated as described in chapter 2 (2.2.3, 2.2.4, 2.2.5 and 2.2.6) and only full-length genome sequences that covers *gag* to *nef* was included. Whole genome amplicon template diversity for each full-length genome sequence is indicated in Table 2.1 (Chapter 2).

3.2.4 Phylogenetic analysis

The primary and chronic infection full-length genome datasets were aligned using ClustalW as implemented in BioEdit with manual editing in BioEdit. Full-length genome sequences were split into individual gene fragments for gene-specific analyses. Maximum likelihood phylogenetic trees were inferred in PHYML as implemented in RDP3.26,using the General Time Reversible nucleotide substitution model with gamma correction for site-to-site rate variation (α = 2) selected by FindModel (http://www.hiv.lanl.gov).

3.2.5 Phylogeny-aware comparison of amino acid mutational spectra

As recombination can seriously confound phylogenetic analyses, we sought to account for potential recombination by performing separate analyses for different alignment partitions as defined by identified recombination breakpoints (Scheffler *et al*., 2006). Recombination breakpoints were identified in different HIV gene alignments using the RDP (Martin & Rybicki, 2000), GENECONV (Padidam *et al*., 1999), BOOTSCAN (Martin *et al*., 2005[a]), MAXCHI , CHIMAERA (Martin & Rybicki, 2000; Martin *et al*., 2005[b]) and SISCAN (Gibbs *et al*., 2000) methods implemented in RDP3. Default settings were used throughout and only potential recombination events detected by two or more of the above methods (with associate Bonferroni corrected P-values <0.05) coupled with phylogenetic evidence of recombination were considered significant.

The *gag*, *pol*, *env* and *nef* genes were partitioned at breakpoint positions. *Vif*, *vpr*, *vpu*, and *tat* genes were not detectably recombinant. Recombination could also not be detected in these genes using the GARD (Genetic Algorithm for Recombination Detection) method implemented on the Datamonkey webserver (Pond *et al*., 2006; http://www.datamonkey.org). Overlapping reading frames, variable regions in *env* as well as insertions or deletions were removed before genes and partition fragments were translated to amino acids. Neighbor-joining trees for protein alignments (without bootstrapping) were inferred with MEGA 4 (Tamura *et al*., 2007) using the Poisson correction distance model which assumes equal substitution rates and equal amino acid frequencies. Rev was not analysed as it is completely embedded in overlapping reading frames and was therefore unsuitable for analysis.

For each alignment partition defined by identified recombination breakpoints we inferred the sequences at the ancestral nodes of the corresponding tree (Edwards & Shields, 2004;Edwards & Shields, 2005) and, for each site, designated the amino acid at the root of the tree as the ancestral amino acid for that site.. Each terminal branch with a mutation towards the ancestral amino acid was assigned a score of +1; terminal branches with a mutation from the ancestral amino acid to any other amino acid were assigned a score of -1 and terminal branches for which no amino acid replacement was inferred were assigned a score of 0. We then compared the numbers of -1, 0 and +1 scores in terminal branches (Figure 3.1) leading to sequences sampled during primary infection to the corresponding numbers from chronic infection sequences using a two-tailed Wilcoxon rank-sum test with a p-value cut-off of 0.025. We then carried out a permutation test in order to investigate the impact of multiple hypotheses testing on our results. We randomly shuffled the sample labels (primary/ chronic infection) 1,000 times and for each randomization we repeated the test and evaluated the number of sites with a p-value below the significance threshold.

We then investigated each site in detail to identify possible biological processes responsible for these differences. The methodology described was developed by N. Wood and C. Seoighe, National Bioinformatics Network Node, University of Cape Town.

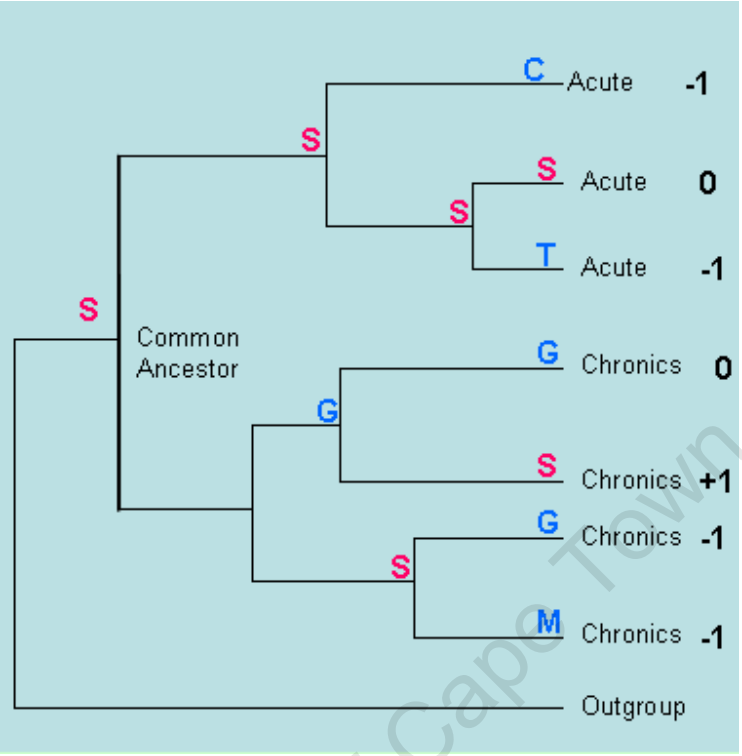


Figure 3.1 Diagram of a phylogenetic tree showing the relationship between evolution at the tip of the tree branches for primary infection and chronic infection sequences and the most recent common ancestor (MRCA) for each amino acid alignment position is used to demonstrate the methodology of phylogeny-aware comparisons of amino acid mutational spectra in 2 datasets. Changes towards and away from the ancestral amino acid are indicated by +1 and -1 respectively and no change between the MRCA and the tip of the tree equals 0 (drawn by Gama Bandawe and Cathal Seoighe, UCT).

3.2.6 Site-by-site Shannon entropy estimation

The average entropy was used to estimate the variability at amino acid sites at each alignment position of the primary and chronic datasets at signature positions (Yusim *et al*., 2002; Korber *et al*., 1994; http://www.hiv.lanl.gov/tmp/ENTROPY/). HIV-1 subtype C sequences available on the HIV sequence database were used to determine the database frequency of amino acids at alignment positions for gp41 (n = 508); Gp120 (n = 531); Gag (n=413); Nef (n= 586), Rev (n = 457 and n = 562 for exons 1 and 2

respectively), Vif (n = 409), Vpr (n = 401), and Pol (n = 412) (http://hiv.lanl.gov/components/ sequence/HIV/).

Liu *et al* (2007) determined that in most instances the wildtype CTL epitope mutations are found at database frequencies of >50% whereas mutant forms are mostly found at frequencies that are less than half the frequency of the wildtype mutation in the CTL epitope. They also suggested that low frequency mutant mutations are associated with a fitness cost.

3.2.7 Variable loop length and N-linked glycosylation sites (PNGs)

The length (number of amino acids) of *env* variable loops and the total number of PNGs were determined with the N-Glycosite tool on the HIV sequence database (http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html, (Zhang *et al.*, 2004).

3.2.8 Screening for possible CTL epitopes

Motif Scan (a program that uses known HLA-1 restricted CTL epitope binding motifs to predict HLA-peptide binding sites; http://www.hiv.lanl.gov/content/immunology/motif_scan/), the CTL epitope database (http://www.hiv.lanl.gov/content/immunology/ctl_search), the Los Alamos HIV Molecular Immunology Compendium 2006/2007 and the NetMHCpan tool (http://www.cbs.dtu.dk/services/NetMHCpan/) which use HLA and peptide sequence information to predict the affinity of peptide-HLA interactions (Nielsen *et al.*, 2007) were used to identify putative CTL epitopes predicted to be restricted by each study participant's particular HLA alleles. HLA-I A, B and C typing was carried out at high resolution by sequencing using the Atria AlleleSeqr (Abbott Diagnostics) and Assign-SBT 3.5 (Conexio Genomics) kits as described in Chopera *et al.* (2008). We calculated

the proportion of immunoreactive peptides with respect to the complete HIV-1 subtype C proteome, by mapping reported immunoreactive peptides in subtype C infections onto the viral proteins.

3.2.9 Statistical analyses

The non-parametric Wilcoxon rank-sum test was used to identify differences between the primary and chronic infection datasets with respect to both the numbers of N-linked glycosylation sites (PNGs) and the lengths of the variable loops in *env*. Differences in entropy scores between primary and chronic strains at each identified position were evaluated using the two-tailed Wilcoxon rank-sum test. These statistical tests were carried out using GraphPad Prism® 5.0 (GraphPad Software Inc., CA, USA).

The 2X2 Chi-square test (http://faculty.vassar.edu/lowry/tab2x2.html) was used to determine whether or not sites with significant allele frequency spectrum differences between viral isolates from primary and chronic infections, clustered within immunoreactive regions of the HIV-1 proteome (Fisher's exact one-tailed T test was used to measure significance).

3.3 RESULTS

3.3.1 Classification of infection stages

Twenty women experiencing primary HIV-1 infections who were recruited as part of the CAPRISA 002 acute infection study post seroconversion as described in chapter 2 was included (van Loggerenberg *et al.*, 2008). These women were estimated to have been infected for a median of 39 days (range 22 to 62 days) at enrolment and were in Fiebig stage V and VI. Most participants had high viraemia with a median viral load of 110 900 copies per ml (range from 610 to 621 000 copies/ml; Chapter 2, Table 2.1). We

excluded 3 participants who were seronegative or did not have a full-length genome sequence that covers *gag* to *nef* (CAP69, CAP217 and CAP221).

3.3.2 Characterization of full-length HIV-1 genomes

Full-length genomes were amplified and genetic homogeneity in V1V2 of the template, indicative of amplification from a single genome, was confirmed for 13 out of 20 amplicons from 20 participants and heterogeneity was identified in each of the remaining seven samples (Table 2.1). All 20 of the full length genome sequences clearly belonged to HIV-1 subtype C and none were detectably inter-subtype recombinants (Chapter 2, 2.3.3; Figure 2.3).

To identify polymorphisms associated with recently transmitted viruses, we compiled from public databases a dataset of subtype C chronic sequences which were closely matched to our acute infection dataset for geographical origin, host population and mode of transmission. As we were interested in identifying genetic features that differed between viruses sampled during primary and chronic infections, it was necessary to ensure that there were no obvious sampling biases. The mean genetic distances between the *env* genes of viruses within each dataset was similar: 11.5% (range 8.2% - 14.8%) in the primary infection dataset compared to 10.9% (range 6% - 15.1%) in the chronic infection dataset. In addition, there was no obvious evidence of close epidemiological linkage as the sequences were generally dispersed throughout a subtype C phylogenetic tree containing viruses sampled world-wide (Chapter 2, Figure 2.3).

A comparison of the 86 sequences used in this study showed limited structure in the phylogenetic tree (Figure 3.2) with only seven lineages displaying bootstrap support above 75%. Of these seven lineages, six consisted of only two sequences each. Most lineages contained a mixture of acute and chronic sequences. Thus, despite a common geographic origin, there was no obvious evidence of close genetic and phylogenetic relationships within or between primary and chronic sequences.

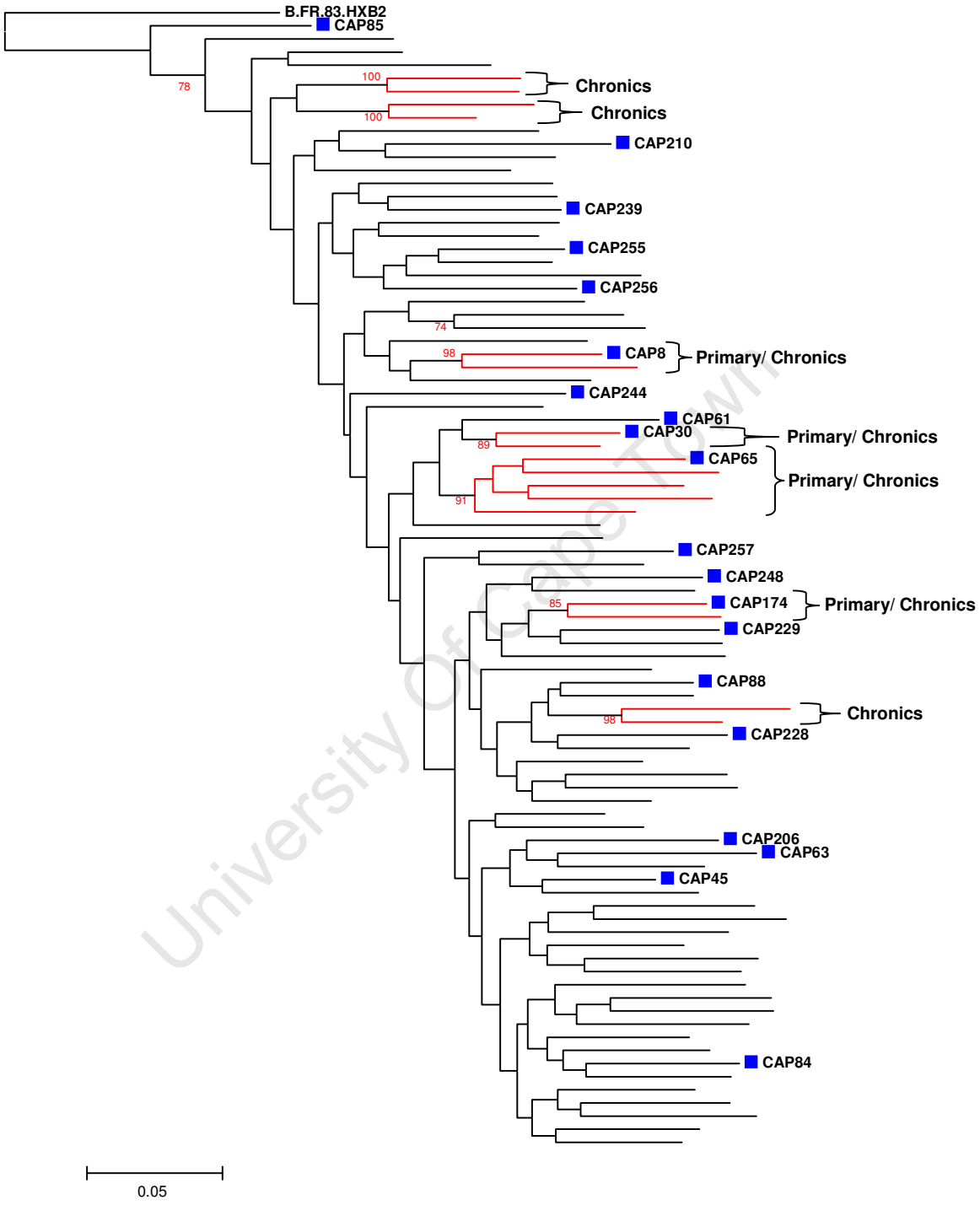


Figure 3.2 Maximum Likelihood tree of *env* gene sequences from primary (n= 20) and chronic (n= 66) infection HIV-1 subtype C strains. The HXB2 subtype B strain was used as root and 100 bootstrap replicates were done. Primary infection strains are indicated in blue squares and chronic strains as unlabelled tips. Subclusters indicated in red had bootstrap values ≥85%. Scale bar = 0.05.

3.3.3 Envelope glycoprotein variable loop length and N-linked glycosylation

Previous studies have shown statistically significant differences in both the lengths of variable loops and the numbers of N-linked glycosylation (PNGs) sites found in the envelope glycoproteins of viruses sampled during primary and chronic infections (Derdeyn *et al.*, 2004; Li *et al.*, 2006). Consistent with these studies we found significantly fewer PNGs in the V1V2 loop regions of the viral Env sequences sampled during primary infections ($p = 0.025$) (Figure 3.3). We did not, however, find any significant differences between the two datasets with respect to either the number of PNGs across the entire V1V4 region (median of 20 for both primary and chronic) or in the lengths of the V1V2 (median of 67 and 68 amino acids in the primary and chronic datasets, respectively) and V1V4 regions (median of 280.5 and 280 amino acids in the primary and chronic datasets respectively).

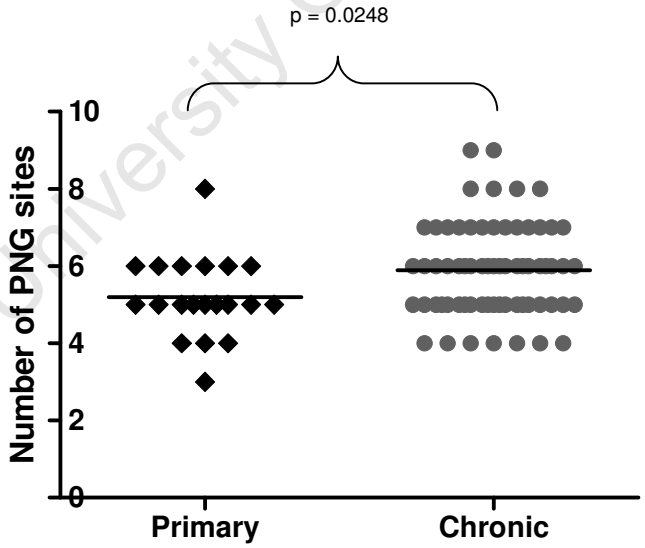


Figure 3.3 Number of potential N-linked glycosylation sites (PNGs) in the V1-V2 variable domains of gp120 from HIV-1 subtype C strains from primary and chronic infection ($p < 0.05$).

3.3.4 Site-specific differences in amino acid frequencies between the primary and chronic infection datasets

We used a phylogenetic approach to test for more subtle differences between the primary and chronic infection datasets. Our method accounts for detectable signals of recombination and controls for founder effects in the underlying evolution of these sequences (Bhattacharya *et al.*, 2007; Scheffler *et al.*, 2006). The method (developed by Natasha Wood and Cathal Seoighe, NBN node, UCT) infers the amino acid states of ancestral viruses, and evaluates the difference in the mutational patterns between two groups of sequences, at each site along a protein sequence alignment (See Materials and Methods for details). Intra-subtype recombination breakpoints were identified in *gag, pol, env* and *nef* genes. However, no recombination breakpoints were found in *vif, vpr, vpu* and *tat* using the GARD method (Pond *et al*., 2006; http://www.datamonkey.org).

The primary and chronic datasets differed at 39 amino acid sites (Table 3.1), using the phylogenetically corrected test and a p-value threshold of 0.025 (without correction for multiple comparisons). We used a permutation test to investigate the impact of multiple hypotheses testing on our results. We permuted the sample labels (i.e. primary versus chronic infection) randomly 1,000 times and counted the number of sites in each permuted dataset that differed significantly ($p < 0.025$) between the permuted primary and chronic groups).

While in the observed (unpermuted) data there were 39 sites with a p-value below 0.025, among the 1,000 permuted samples, the mean number of sites with a p-value below 0.025 was 9.3 and there were no permuted datasets with as many as 39 significant sites. This provides evidence that the significant sites are enriched for true positives and allows us to estimate a false discovery rate of approximately 24% (9.3/39) for the sites with a p-value less than 0.025. Fourteen of the 39 sites were within Env, nine in Pol, six in Gag, four in Nef, three in Vif, two in Vpr and one in Vpu (Figure 3.4). We then investigated each site in detail to identify possible biological processes responsible for these differences.

Table 3.1 Amino acid positions were the frequency of gain and loss of specific amino acids at terminal branches differ significantly between HIV-1 subtype C strains from primary and chronic infection.

Protein	[a]Amino acid position (HXB2)	P-value	[b]Subtype C CTL reactive peptide sequence (site in boldface and underlined)	[b]Known HLA restriction
Gag p17	69	0.0037373	EGCKQIMKQLQPALQTGT, QLQPALQTGTEELRSLY	B*0801, B*4006, A2, A*0101, B57
Gag p17	72	0.0103261	QLQPALQTGTEELRSLY	B*0801, B*4006, A2, A*0101, B57
Gag p17	105	0.0131367	EALDKIEEEQNK	A11
Gag p24	138	0.0097975	GKVSQNY/PIVQNLQGQMV	B13, A68, A*6802, A*2402
Gag p24	228	0.0163842	PVAPGQMREPRG	B35, B13
Gag p2	371	0.0192564	EAMSQANSVNIM	A2, A*0201, A2 supertype, B*4002, B*4501
Pol protease	113	0.0103261	GGIGGFIKVRQYDQIL	A2, B13, Cw6
Pol protease	128	0.0015348	QIPIEICGKKAIGTVLV, GKKAIGTVLVGPTPVNII	B*1503, B57, B58, B63
Pol protease	131	0.0103261	GKKAIGTVLVGPTPVNII	B*1503, B57, B58. B63, A*0201
Pol RT	276	0.0165847	DAYFSVPLDEGFRKYTAF	B*5702, B*5703, B35, B*3501, A11
Pol RT	447	0.0140015	AKALTDIVPLTEEA	B*0702, B*1501, B*3501, B*5101, B*5301, B35, B51, B7
Pol integrase	726	0.0193219	KAQEEHEKYHSNWR	B*4403
Pol integrase	756	0.0103261	EIVASCDKCQLKGE	B*8101
Pol integrase	813	0.0103261	PAETGQETAYYILKLAGR	A*6802, A*2601, B7, B56
Pol integrase	850	0.0131367	VKAACWWAGIQQEFGIPYNPQS	A2 supertype, B*1503
Vif	46	0.0103261	RHHYESRHPKVSSE	B*0702, B*4201, B7
Vif	78	0.0158979	D/WHLGHGVSI/, LQTGERDWHLGHGVSIEW	B*1510, B*5703, B35
Vif	137	0.0103261	HIVSPRCDYQAGHNKVGSLQYLAL	
Vpr	68	0.0015348	AIIRILQQL/L	A*0201, A2, A2 supertype
Vpr	81	0.0103261	GCQHSRIGILRQR	
Vpu	33	0.0097975	YIEYRKLVRQR, EYRKILRQR	A*3303
Env gp120	106	0.0103261	KNDMVDQMHEDIISLW	A*0201, B*3801, A2,
Env gp120	162	0.0015348	CSFNITTELRDKKQKVYA, NCSFNISTSI	Cw8, Antibody pressure
Env gp120	171	0.0007625	CSFNITTELRDKKQKVYA	
Env gp120	184	0.0099403	YALFYRLDIVPLNENNSSEY	
Env gp120	340	0.0192564	HCNISEAAWNKTLQQVR	A11, A*0201
Env gp120	352	0.0200731	QQVRKKLEEHFPNKTIIF	A*0201, A11
Env gp120	476	0.0197145	TFRPGGGDMRRNWRSELY, MRRNWRSELYKYKVVEI	A*2601
Env gp120	477	0.0003449	TFRPGGGDMRRNWRSELY, MRRNWRSELYKYKVVEI	A*2601
Env gp120	485	0.0103261	NWRSELYKYKVVEI	
Env gp41	535	0.0191913	GSTMGAASITLTVQARQ	A2
Env gp41	583	0.0015348	GIKQLQTRVLAIERYLK, RVLAIERYLKDQQLLGIW	B*5802, B14
Env gp41	668	0.0173911	EKDLIALDKW(Q/N)NLWNWFDIT	
Env gp41	687	0.0165847	WYIKIFIMIVGGLIGLR	A*2402, A2, A*0201
Env gp41	708	0.0103261	AVLSVVNRVRQGYSPLS	A*2501, A*3002, A30
Nef	5	0.0131367	MGGKWSKSSIV	A2, A*2501
Nef	65	0.0000520	WLRAQEEEEEVGFPVRPQV, EVGFPVRPQVPLRPMTFK	B*4501, B45, B7, A*0201, A1, B8, B35
Nef	88	0.0171700	KAAFDLSFF, GAFDLSFFL	B57/ B*5801, A*0205, B60, B62, A2, Cw8, Cw*0802
Nef	169	0.0103261	LLHPMSQHGMDDPER	B35

Sites identified with a p-value <0.025 are reported. [b]Reactive peptides of which some contains published CTL epitopes were obtained from the Los Alamos HIV database (http://www.hiv.lanl.gov/ content/ immunology/ hlatem/ study4/ index.html, http://www.hiv.lanl.gov/content/immunology/ctl_search)

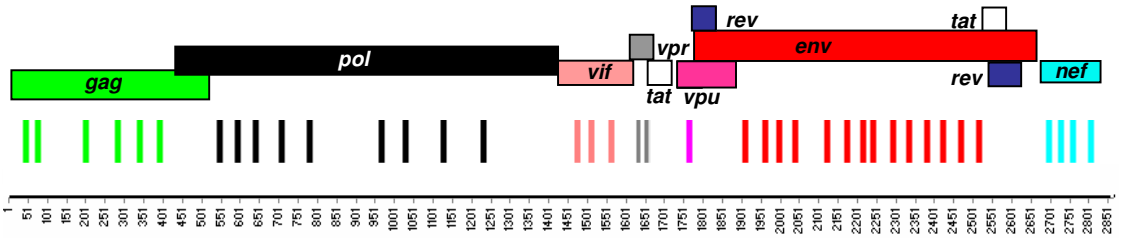


Figure 3.4 Amino acid positions that displayed a significant difference between primary and chronic infection subtype C sequences are showed graphically across the HIV-1 proteome ($p<0.025$).

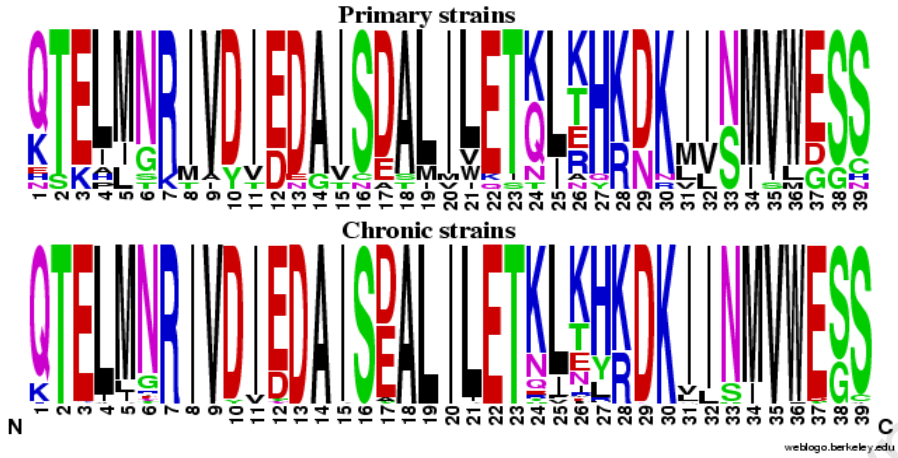
3.3.5 Sites differentiating the primary and chronic infection datasets have higher entropy in the primary infection dataset

To better explore the nature of the changes in amino acid mutational spectra between the primary and chronic infection datasets, we examined the relative entropies of the 39 identified sites (Figure 3.5a). On average, the site-specific entropy was higher in the primary infection dataset (median = 0.518) than it was in the chronic infection dataset (median = 0.263, $p < 0.0001$, two-tailed Wilcoxon rank-sum test) (Figure 3.5b). Based on analysis of HIV-1 protein sequences sampled from public databases, Bansal *et al.* (2005) defined high entropy sites as those with an entropy score greater than 0.25 and low entropy sites as those with an entropy score less than 0.15. Whereas in the primary infection dataset all 25/39 sites had high entropy, in the chronic infection dataset only twelve had high entropy. Seventeen sites with entropies from 0.325 to 0.588 in the primary infection dataset were either fully conserved or highly conserved, in the chronic infection dataset (Figure 3.5c).

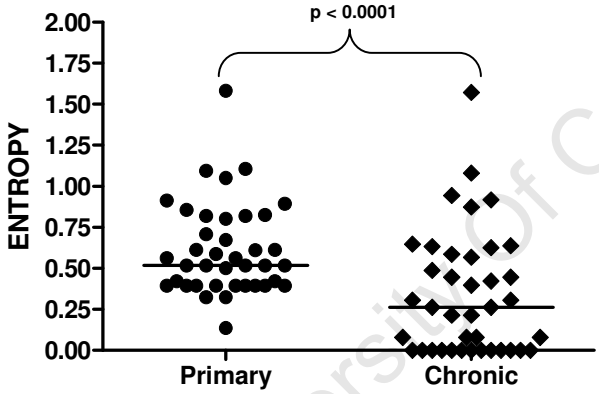
3.3.6 Sites with differential amino acid frequency spectra are significantly clustered within known CTL epitopes

It has been suggested that there is typically higher sequence entropy at amino acid positions where escape mutations occur (Liu *et al.*, 2007) and that CTL responses during

(a)



(b)



(c)

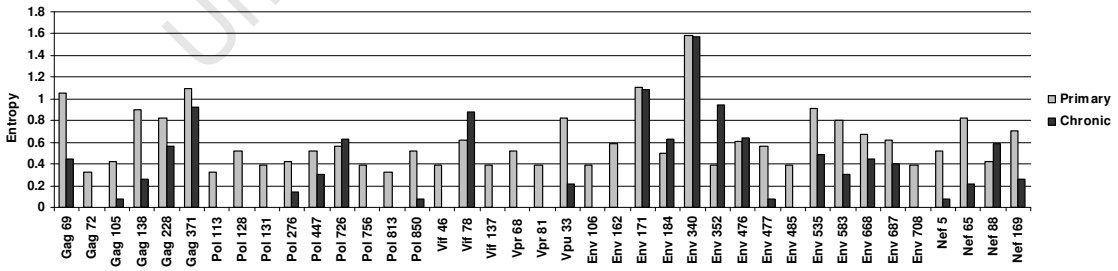


Figure 3.5 A comparison of the 39 positions identified with significant changes ($p<0.025$), in amino acid spectra between primary and chronic infection showing (a) amino acid variety and relative frequency (b) difference in median entropy ($p<0.05$) and (c) site-specific entropy at each position.

early infections mostly target peptides with high degrees of entropy (Bansal *et al.*, 2005). To investigate whether the sites identified by our analysis were associated with CTL responses we checked the sites against the genomic positions of peptides that were immuno-reactive in Elispot assays (Gray *et al.*, 2009; Kiepiela *et al.*, 2007; http://www.hiv.lanl.gov/content/immunology/hlatem/study4/; Matthews *et al.*, 2008). We found that 33 of the 39 sites were located within immunoreactive peptides (Table 3.1). Immunoreactivity has been mapped to approximately 48% of the HIV-1 subtype C proteome. We found that the 39 sites clustered more frequently within these immunoreactive regions than is expected by chance (p= 0.006). This implied that polymorphisms at the sites differentiating the primary and chronic infection datasets are most likely associated with CTL immune pressures.

3.3.7 Longitudinal monitoring of evolution at amino acid sites which differed between primary and chronic phases of infection

To more directly determine the nature of discordant amino acid mutation spectra in our primary and chronic infection datasets, we obtained longitudinal samples from 18 of the 20 study participants at between three and six months after our initial samples were taken. We were specifically interested in determining whether increased entropy at the sites identified in our analysis was due to (i) viruses sampled in primary infections carrying transient immune evasion mutations that they had carried over from former hosts (reversion), (ii) viruses accumulating novel immune evasion mutations in response to changes in the immune environment following transmission (escape), or (iii) a combination of both (i) and (ii).

Here we defined probable CTL escape mutations as amino acid substitutions within epitopes restricted by patient HLA alleles (or in immediately adjacent amino acids) where changes were from amino acids found in $\geq$50% in the population (i.e. consensus or wild-type states) to amino acids found in <50% of the population (i.e. mutant states; Allen *et al.*, 2005; Liu *et al.*, 2006; Li *et al.*, 2007). Conversely we defined

probable reversion mutations as being amino acid substitutions within known CTL epitopes that were not targeted by a patients HLA alleles in which low frequency amino acids were replaced with high frequency ones such as those corresponding with the subtype consensus sequence (Allen *et al*., 2005; Liu *et al*., 2006; Li *et al*., 2007).

Amino acid changes were seen in 13 participants at 20 sites, of which four sites were associated with both escape and reversion (Tables 3.2 and 3.3, HLA data are reported in chapter 4, table 4.1). Evolution from low to high frequency amino acids (putative reversion) occurred at 14 sites (Table 3.3), and evolution to low frequency amino acids (putative CTL escape) were seen at seven sites (Table 3.2). At three sites we saw escape followed by reversion to the original wild-type amino acid within six months (transient escape) post infection (Table 3.2). In total, eight putative escape events were identified at seven sites (Vif 78, Env 162, Env 352, and Nef 65) in six individuals with one individual (CAP256) showing escape at three sites (Gag 371, Pol 113, and Vif 78). Seven sites associated with escape evolved from high frequency (median frequency of 0.775) to low frequency (median frequency of 0.168) amino acids. At position 65 in Nef a change from the consensus E (frequency of 0.865) to a D is associated with CTL escape in HLA-B*45/ B*4501 positive individuals (Rousseau *et al.*, 2008). However in CAP63 position 65 in Nef evolved from a low frequency amino acid (G = 0.046) to another low frequency amino acid (D = 0.07). Although it is slightly more frequent, this new amino acid polymorphism was classified as an escape mutation. It is also possible that the original G polymorphism was itself also an early escape mutation as the first sample recorded for this patient was only obtained approximately 34 days post infection. Mutation to D at this site may have simply provided more selectively beneficial escape than was provided by the intermediate G state. In three participants the Vif and Nef sites were located in peptides restricted by the host HLA (B*1503 and HLA-B*45 respectively) providing further evidence that these sites were associated with evasion of CTL responses. The one putative escape in Env 162 reverted to consensus at 29 weeks with concomitant escape at an adjacent site. This oscillation of amino acids within nine–mer CTL epitopes is commonly observed in the early stages of escape prior to the selective expansion of viruses carrying in most cases just the single highest fitness escape

Table 3.2 Putative escape mutations within CTL epitopes

Site	PID	Weeks post infection	Putative epitopes aligned to matching test peptides	Amino acid frequency change	HLA restricted
Gag 371			^{364}AEAMSQANS-AIMMQR	77.48> 8.96	Affinity = 6.24
	CAP256	6	....._..TN.L...	N>N>G	B*1503
		13	....._..TN.L...		
		30	V...N..GQTS.....		
Pol 113			^{109}GGIGGFIKVRQYDQILI	97.68> 2.09	Affinity = 457.37
	CAP256	6	....._.....T.	R>R>K	B1503
		13	....._.....T.		
		30	.....K..E....		
Pol 756			^{748}AREIVASCDKCQLKGEAI	98.84> 0.46	No
	CAP45	2	.K....._.....	D>D>G	
		5	.K....._.....		
		12	.K.....G.....		
Vif 78			^{72}LQTGERDWHLGHGVSIEW	0.386 → 0.061	B*1503
	CAP256	6	.....E.....A....	E→A	
		13	.....E.....A....		
		30	.....A.....		
Vpr 81			^{71}HFRIGCQHSRIGILRQRR	100> 0	No
	CAP239	5	....._.....	I>I>M	
		11	....._.....		
		22	.....L...M.....		
Env 352			^{339}NKTLEEVRKKLQEHFPNK	0.727 → 0.168	No
	CAP244	8	....QQ.G...._...G.	E→K	
		12	....QQ.G.....G.		
		28	....QQ.GE...K...G.		
Nef 65			^{61}EEEPEVGFPVRPQVP	0.865 → 0.07	B *4501
	CAP85	5	...K_.....	E → D	
		13	...ED.....		
		29	...ED.....		
			^{61}EEEPEVGFPVRPQVP	0.046 → 0.07	B *4501
	CAP63	5	Q..EG.....	G → D	
		11	Q..EG.....		
		29	Q..ED..L.....		
Transient escape at sites within 6 months post infection					
Vif 46			^{41}RHHYESRHPKVSSEVHI	96.33> 0.73> 96.33	Affinity = 137.72,
	CAP45	2	....._.....	S>N>S	A*2902
		5	.....N.....		
		12	....._.....		
Env 162			^{156}NCSFNTTTEIRDKKQTVY	0.979 → 0.006 → 0.979	No
	CAP63	5	.....L.....K..	T→S→T	
		11	.....AS..L.....K..		
		29	.....A_.AL.....K..		
Nef 88			^{77}RPMTYKAAVDLSFFL	76.28> 23.21	Affinity = 183.39
	CAP257	7	.....G.F..._...	S>G>S	B*4202
		14	.....G.F..G...		
		30	.....G.F..._...		

mutation (Borrow *et al.*, 1997; Delport *et al.*, 2008; Iversen *et al.*, 2006). However as this site was located in an N-linked glycosylation motif, it is also possible that antibody pressures played some role in its selective value.

In total 17 potential reversion mutations were identified at 14 sites, within viruses sampled from nine of the study participants. Longitudinally sampled viruses from CAP256 showed putative reversion at 7/14 sites with CAP174 and CAP206 each having putative reversion mutations at two sites. Reversion mutations involved substitutions of low frequency amino acids (median population-wide frequencies = 0.058 at the site in question) with higher frequency amino acids (median population-wide frequencies = 0.930). These potential reversion mutations were distributed throughout the genome with six occurring in Env, three in Pol, two in Gag, and one each in Vif, Vpr and Nef (Table 3.3). There was no predicted HLA association for 14 out of the 17 reversion mutations providing further evidence that these sites were associated with reversion of CTL escape mutations that had occurred in former hosts which had different HLA alleles than the virus' current hosts.

The one exception was a probable reversion mutation located at Nef 65 within the CTL epitope restricted by one participant's HLA-B*4501 allele. Importantly, escape mutations were also seen at adjacent positions (63, 64) within this putative CTL epitope. The potential reversion at amino acid position 162 in Env is probably associated with regain-of-function as this site is almost invariably a threonine (T) residue (HIV-1 subtype C population-wide frequency = 0.979) and resides within a potential N-linked glycosylation site motif in the V2 loop.

In summary, a total of 28 evolutionary events were observed in 13 participants at 20 sites of which three events were associated with transient escape (10.7%), eight with putative escape (28.6%) and 17 with putative reversion (60.7%). Thus the longitudinal evolutionary changes observed at these sites were mainly associated with reversion to high frequency amino acids during primary infection with a minority of changes potentially being associated with CTL escape.

Table 3.3 Putative reversion mutations

Site	PID	Weeks post infection	Putative epitopes aligned to matching test peptides	Amino acid frequency change	[a]Fold frequency increase	HLA restricted
Gag 69			[69] QTGTEELRSLYNTVATLY	0.157 → 0.823	5.24	No
	CAP256	6	K.....F.....	K → Q		
		13	K.....F.....			
		30	.....V.F			
Gag 228			[223] IAPGQMREPRGSDIA	0.0024 → 0.981	409	No
	CAP256	6	.....I.....	I → I → M		
		13	.....I.....			
		30	...N.....			
Pol 131			[124] GKKAIGTVLVGPTPVNII	0 → 1.00		B*5802
	CAP174	4	.....A.....	A →V		
		28	.....			
Pol 447			[439] RGTKALTDIVPLTEEAEL	0.0534 → 0.944	17.7	No
	CAP255	8	..A.....V.....	V → I		
		13	..A.....			
Pol 850			[841] VKAACWWAGIQQEFGIPY	0.0139 → 0.981	70.6	No
	CAP61	8	.....V.....	V → I → I		
		11	.....			
		33	.....			
			[841] VKAACWWAGIQQEFGIPY	0.0046 → 0.981	213.3	No
	CAP174	4	.....T.....	T → I		
		28	.....I.....			
Vif 137			[128] IVIPRCDYQAGHNKVGSL	0.0073 → 0.993	136	A*2902
	CAP256	6	..S...E.LT..S.....	T →T → A		
		13	..S...E.LT.....			
		30	..S.....			
Vpr 81			[71] HFRIGCQHSRIGILRQRR	0 → 1.00		*No
	CAP206	8	.....V.....	V → I → I		
		15	.....			
		33	.....			
Env 106			[96] WKNDMVDQMHEDIINLW	0.0019 → 0.932	490.5	No
	CAP239	5	.....K...S..	K → K → E		
		11	.....K...S..			
		22	.....S..			
Env 162			[156] NCSFNTTTEIRDKKQTVY	0.004 → 0.979	245	No
	CAP206	8	.....A.....Q..	A → T		
		15	...X.....Q..			
		33	.....Q..			
Env 352			[339] NKTLEEVRKKLQEHFPNK	0.168 → 0.727	4.33	No
	CAP256	6	E...QR.SEE.RK.....	K → E		
		13	E...QR.SEE..K.....			
		30	E...QR.SE.....			
Env 477			[474] NMKDNWRSELYKYKVVEI	0.070 → 0.927	13.2	No
	CAP85	5	...N.....	N → D		
		13	.....			
		29	.T.....			
			[474] NMKDNWRSELYKYKVVEI	0.070 → 0.927	13.2	No
	CAP256	6	D.RN.....	N → D		
		13	D.RN.....			
		30	..R.....V			
Env 535			[524] GAAGSTMGAASITLTVQA	0.063 → 0.852	13.5	No
	CAP256	6	.....MA.....	M → M → I		
		13	.....MA.....			
		30	.....			
Env 668			[664] DKWQNLWSWFSITNWLWY	0.173 → 0.781	4.5	No
	CAP256	6	.S.NS..N....ST....	S → N		
		13	.S.NS..N....ST....			
		30	.S.N...N..D.ST....			
Nef 65			[61] EEEPEVGFPVRPQVP	0.070 → 0.865	12.4	B*4501
	CAP30	5	..GED.....K....	D → E		
		11	..GD......K....			
		29	..GD.....			
			[61] EEEPEVGFPVRPQVP	0.070 → 0.865	12.4	No
	CAP84	3	...GD.....	D → E		
		14	...GD.....			
		19	..AE.....E.			

[a]Fold frequency increase: the difference in frequency of an amino acid at an alignment position compared to the frequency of another amino acid at the same alignment position

3.3.8 GenBank accession numbers

Near full-length sequences from the primary infection stage were submitted to GenBank under accession numbers GQ999972 to GQ999991.

3.4 DISCUSSION

HIV transmission is associated with a severe virus population bottleneck and there is some evidence that viral envelopes with specific genotypic and phenotypic properties are selected during transmission (Derdeyn *et al.*, 2004; Rong *et al.*, 2007; Wolfs *et al.*, 1992). However, open questions remain as to whether this holds true both for genome regions other than the envelope, and for all HIV-1M subtypes. To further explore this concept, we generated full-length genomes from 20 recently HIV-1 subtype C infected individuals and compared these sequences to those sampled during chronic infections. Similar to Li *et al.* (2006) we also found fewer glycosylation sites but not shorter variable loop lengths within the envelopes of viruses sampled during primary infections. However, in an analysis corrected for founder effects and recombination, we found that site-specific amino acid mutational differences across the full-length proteome are almost exclusively associated with signals of virus adaptation to the new host rather than with signatures obviously associated with preferential transmission.

Our study describes full-length HIV-1 subtype C genomes sampled from individuals during the primary phase of infection. We identified thirty-nine sites within the proteomes of these viruses that differentiated them from viruses sampled during chronic infections. Through longitudinal analysis of amino acid frequency changes that occurred during the first six months of infection, together with data on host HLA alleles, we provide evidence that approximately 28.6% of site-specific differences in amino acid frequency spectra between primary and chronic infection proteomes are potential immune escape mutations. The remaining 60.7% of frequency differences between the two groups are probably due to defunct immune evasion substitutions reverting to consensus

amino acid states following transmission. This data provides further understanding of processes determining the genomic and immunogenic properties of viruses during early infections which is important if we are to understand HIV pathogenesis sufficiently well to design protective vaccines against the virus.

Almost all of the sites displaying substantially different amino acid frequency spectra between viruses sampled during primary and chronic infections were located within peptides that have known immuno-reactivity. Most of these sites were in Env followed by Pol and Gag containing more sites than any of the remaining proteins. We further investigated the nature of the immune selection operating on these sites through analysis of sequence sampled longitudinally from the study participants. Based on changes in amino acid frequencies relative to the global HIV database, we found that the amino acid frequency variations in 14/39 of the identified sites were consistent with high rates of reversion mutations being associated with either transmission or primary infection. The frequency spectra differences at 7/39 sites were consistent with early escape from CTL responses during primary infection. The timing of escape may well be crucial, as none of the individuals had IFNg responses to subtype C-based peptide pools containing the presumptive immunoreactive epitopes screened using the ELISPOT assay (Gray *et al.*, data not shown). It is also possible, however, that these assays may have missed responses due to mismatches between the peptide sequences used and the infecting virus.

Our observation that reversion mutations are potentially more common than CTL escape mutations during the early stages of HIV infections is broadly in agreement with that of Li *et al*. (2007) but is at odds with those of Goonetilleke *et al*. (2009) and Kearney *et al*. (2009). What differentiates ours and the Li *et al* study from those of Goonetilleke *et al*. and Kearney *et al*., is that the latter studies examined sequences sampled pre-seroconversion (Fiebig I/II and III). We and Li *et al.* sampled sequences post-seroconversion, in Fiebig V, VI and beyond. Although CTL escape mutations were only believed to be detectable 30 or more days after peak vireamia (Borrow *et al*., 1997; Liu *et al*., 2006), Goonetilleke *et al.* (2009) have recently described the appearance of CTL

escape mutations as early 14 days post infection. Thus we may potentially have underestimated the numbers of CTL escape mutations in viruses which were only sampled a median of 39 days post infection. This possibly indicates that despite a more rapid initial accumulation of novel CTL escape mutations during the first weeks of an infection, over the following months the rates at which successful CTL escape mutations emerge trails off to the point where their frequency is surpassed by that of reversion mutations.

Nevertheless our classification of reversion and escape mutations was supported by the fact that some of the sites predicted to be associated with escape (in Vif and Nef) were located in peptides reported to be restricted by the HLA-alleles of the relevant study participants, whereas 14/17 evolutionary events at fourteen sites we classified as being potentially associated with high frequency reversion mutations which were not restricted by the HLA-alleles of the relevant study participants. It must be pointed out, however, that certain HLA-epitope associations may have been missed as it has previously been shown that CTL responses are poorly predicted in subtype C sequences due, in part, to lack of detailed characterisation of HLA alleles in African populations (Ngandu *et al.*, 2007). This result, based on full-length genomes, supports previous studies based on Gag, Pol and Nef (Brumme *et al.*, 2008; Li *et al.*, 2007) which suggest that most of the high entropy sites displaying amino acid frequency differences between viruses sampled in primary and chronic HIV infections represent defunct escape mutations accumulated in former hosts that had different HLA alleles from the virus' current hosts.

During early infection many CTL evasion mutations accumulated within previous hosts revert to their consensus states because these "wild-type" polymorphisms provide a greater degree of replicative fitness (Brumme *et al.*, 2008; Martinez-Picado *et al.*, 2006). At the same time that defunct CTL evasion mutations are reverting, viruses are forced to escape the immune pressures exerted by the immune environment of their new hosts. CTL escape during early infections has been associated with oscillation of amino acids within CTL targeted epitopes prior to their convergence on more stable states (Borrow *et al.*, 1997; Delport *et al.*, 2008; Iversen *et al.*, 2006). This oscillation may either be due to

the negative replicative fitness effects of some CTL evasion mutations, or due to some mutations only providing partial escape from CTL responses due to, for example, their influencing epitope processing rather than recognition (Borrow *et al.*, 1997). By the chronic phase of infections many of these changes may have reached a degree of equilibrium. It is possible, therefore, that in our study we have detected different stages of this oscillation process in the different infection stages. The increased entropy within targeted CTL epitopes in early infections may be due to amino acid switching or shuffling within targeted epitopes as the viruses try to balance the survival benefits of CTL escape with the replicative fitness costs incurred by many CTL evasion mutations (Delport *et al.*, 2008; Goonetilleke *et al.*, 2009;Iversen *et al.*, 2006).

We provide evidence that the innate potential of particular genetic variants to mutationally respond to the selective constraints imposed by new hosts underlie virtually all detectable differences in amino acid frequency spectra between viruses sampled during primary and chronic infections. These data provide valuable insights into unique virological and immunological events during primary infection. We provide evidence which suggests that during the early stages of HIV infections adaptation to the immune environment of new hosts is perhaps secondary to the mutational recovery of replicative fitness losses incurred during CTL escape in former hosts. Our discovery that early infections are primarily characterised by reversion mutations adds to an accumulating body of evidence suggesting the transience of many immune evasion mutations during global population-wide HIV evolution. It is becoming increasingly apparent that CTL escape mutations often have complex evolutionary costs and benefits such that many are likely to have subtle and difficult to predict influences on long-term HIV pathology, epidemiology and evolution. Given that the mutational accessibility and fitness benefits of reversion mutations that occur during early infections should strongly impact the broader effects of CTL evasion mutations, our study emphasizes the importance of studying the evolutionary changes occurring in HIV during the very earliest stages of infection.

Although our data suggest that the majority of the amino acid frequency spectrum differences we have observed between viruses sampled during acute and chronic infections are rationally attributable to evolutionary processes at play post-transmission, it would nevertheless be of great interest to determine whether all such signals are generated *de novo* at the onset of infections. Evidence of even a small proportion of these signals having being generated prior to transmission would provide valuable support for the notion of an evolutionarily relevant "transmission sieve".

CHAPTER 4

EARLY VIRAL EVOLUTION IN PRIMARY HIV-1 SUBTYPE C INFECTION

4.1 INTRODUCTION

Elucidation of events in acute infection will make an important contribution to understanding pathogenesis, as well as informing vaccine design. Following transmission, changes in viral sequence will provide insights into selective forces during this early stage of infection. There are several host factors, such as HLA and host restriction genes (Lehner *et al*., 2008; Fellay *et al*., 2007), that are important in mediating HIV disease progression. These factors can directly impact by controlling virus replication; alternatively they may also act indirectly by inducing mutations that affect the replicative fitness of the virus. As events in acute infection are thought to be critical in defining long term disease outcome, it is of great interest to understand pressures driving early viral evolution and the consequences of these changes in the context of disease progression. There is limited information on the early evolution of the full-length HIV-1 genomes (Liu *et al*., 2006; Li *et al*., 2007; Salazar-Gonzalez *et al*., 2009), and there have been no studies on how these changes impact on subsequent disease progression. It is therefore of interest to define sequence changes occurring in acute infection, to identify selective pressures responsible for these changes, and to determine the impact on disease progression.

Cell-mediated immune pressure during primary HIV-1 infection drives the earliest viral evolutionary changes (Allen *et al*., 2005; Goonetilleke *et al*., 2009). CD8 cytotoxic T cells (CTLs) recognize HIV-1 antigens in association with human leukocyte antigen molecules (HLA). The cytotoxic T cell (CTL) mediated immune response abrogates viral replication through targeted killing of infected cells, whereas the virus evolves to escape

(CTL escape) these protective immune responses resulting in subsequent loss of immune control (Leslie *et al.*, 2004). Escape can occur by mutations in the HLA anchor residues which abrogate binding and presentation of the peptide, mutations or shuffling of amino acids involved in T cell receptor binding, as well as through mutations in epitope flanking regions that affect proteosomal cleavage (Kelleher *et al*., 2001; Iversen *et al*., 2006; Troyer *et al*., 2009). CTL escape mutations have been identified across the HIV proteome with the envelope glycoprotein as a major target of CTLs (Liu *et al*., 2006; Li *et al*., 2007; Goonetilleke *et al*., 2009).

In a cross-sectional study of chronic subtype C infection, seven amino acid residues across the viral proteome (4 in Gag and 1 each in Pol, Vif and Rev) were associated with viral escape mutations that resulted in lower plasma viral loads (Rousseau *et al*., 2008). The mechanism behind this is thought to be due to the fact that some CTL escape mutations (e.g. the T242N mutation in the TW10 epitope) result in impairment of the replicative fitness of the virus and have been associated with viral control and better prognosis (Leslie *et al*., 2004; Martinez-Picado *et al*., 2006; Chopera *et al*., 2008; [a]Miura *et al*., 2009).

Most of the escape mutations that impair viral fitness arise in the presence protective HLA alleles which are associated with better protection against symptomatic HIV disease (B*13, B*27, B*51, B*57, B*5801 and B*8101) (Saah *et al*., 1998; Ioannidis *et al*., 1999; Costello *et al*., 1999; Fellay *et al*., 2007; Goulder & Watkins, 2008, [a]Miura *et al*., 2009). Individuals carrying these protective HLA types generally have immune responses against peptides located in essential functional domains of the viral proteins which are under pressure to maintain wild-type features (Allen *et al*., 2005; Wang *et al*., 2009). A recent study found reduced fitness associated with Gag escape mutations whereas Env escape mutations did not carry a fitness cost (Troyer *et al*., 2009). Further evidence of the fitness cost of these escape mutations is illustrated by the fact that on transmission to HLA-mismatched recipients these CTL escape mutations often revert to wild-type (Leslie *et al*., 2004; Friedrich *et al*., 2004; Li *et al*., 2007; Matthews *et al*., 2008).

These 'attenuating' escape mutations can have benefit even in individuals who do not have these protective HLA alleles. Studies in the CAPRISA cohort, as well as transmission pairs from Zambia, have shown that transmission of HLA-B*57/B*5801-induced escape mutations to HLA-mismatched recipients was associated with a survival benefit although the duration of this benefit is not known (Chopera *et al*., 2008; Goepfert *et al*., 2009). Recipients of these transmitted Gag escape mutations have lower viral loads and higher CD4 counts during primary infection. It was shown recently that HLA-B*57/B*5801 positive individuals who select the Gag escape mutants with reduced replicative ability eventually lose viral control after acquiring compensatory mutations which overcome the defects of the original escape variants (Crawford *et al*., 2009).

APOBEC3G/F is cellular proteins that form part of the innate response which restricts HIV-1 infection by viral genome inactiviation through G-to-A hypermutation (Harris *et al*., 2003; Zhang *et al*., 2003). There is some conflicting information on the impact of hypermutation on disease progression. While one study found that higher APOBEC mRNA levels were associated with lower viral loads and higher CD4 counts (Jin *et al*., 2005), this was not confirmed in a second study (Cho *et al*., 2006). In more recent studies on levels of APOBEC associated mutations Ulenga *et al*. (2008) did not find any association between APOBEC-associated G-to-A mutations and viral load. However when assessing significantly hypermutated proviruses Land *et al*. (2008) found that the level of proviral hypermutation correlates with CD4 counts in infected Kenyan women. In addition to inactivating the virus, APOBEC may also be an important mechanism that the virus uses to avoid immune responses as a recent analysis linked APOBEC-induced mutations to rapidly evolving amino acid positions in CTL epitopes in HIV-1 envelope glycoproteins (Wood *et al*., 2009).

In this chapter we investigated the effect of early viral divergence, reversion mutations, and APOBEC-induced mutations across the HIV-1 proteome on markers of disease progression including viral load and CD4 count at 12 months post infection.

4.2 MATERIALS AND METHODS

4.2.1 Full-length HIV-1 genomes

Samples were collected from eighteen individuals monitored over time for CD4 counts and viral load as described in 2.3.2 (Table 2.1) and in Table 4.1. Near full-length genome sequences were generated using two methods: for 20 individuals, sequences were generated from full-length cloned genomes at two (n = 7) to three (n = 13) time points from 2 weeks to 33 weeks post infection as described in chapter 2 (2.3.3 to 2.3.6); and for two participants (CAP45 and CAP63) sequences were also generated from single genome amplicons at three time points (enrolment, 3 months and 6 months post infection. The method for the generation of near full-length HIV-1 genomes as described in chapter 2 (2.3.3 and 2.3.4) was modified to obtain single genome amplification (SGA) derived ~9 kb amplicons (Salazar-Gonzalez *et al*., 2009, Appendix C6). SGAs were generated using the approach described below for gp160 SGAs (4.2.4). For CAP45 and CAP63 a total number of 17 (2, 5, 12, 16 and 65 weeks post infection) and 24 (2, 5, 11, 29 and 37 weeks post infection) near full-length SGA derived HIV-1 genomes were generated respectively. Additional sequences were included in the analysis including full-length single genome amplification (SGA) derived *env* gene sequences from week 52 for CAP45 (n = 7) provided by Gama Bandawe (University of Cape Town); and CAP84 *gag* gene sequences (n = 1) at 3 months post infection was kindly provided by Denis Chopera from our laboratory.

4.2.2 Single genome amplicons (SGA) derived gp160 sequences

HIV-1 RNA was reverse transcribed to cDNA using Superscript III Reverse Transcriptase (RT) System (Invitrogen, CA) with oligo-dT as described in chapter 2 (2.2.3). Full-length gp160 envelope sequences were derived by single template based amplification and direct sequencing as described by Salazar-Gonzalez *et al.*, (2008) (Appendix C7). Briefly, this limiting dilution based PCR strategy aims to obtain less than

Table 4.1 Study participants' HLA, and viral loads and CD4 counts at sample time points.

Participants	[a]Time post infection	Viral load (c/ml)	CD4 count (cells/mm^3)	Viral load (c/ml) @ 12 mo PI	CD4 count (cells/mm^3) @ 12 mo PI	HLA
CAP8	23 d	373000	360	39300	332	A*2301/*2301,
	13 wks	188000	445			B*0801/*1510
	25 wks	98400	343			
CAP30	35 d	10200	989	204000	573	A*0201/ *3402,
	11 wks	616000	607			B*440301/ *4501
	29 wks	136000	572			
CAP45	14 d	12500	ND	556	1030	A*2301/ A*2902,
	35 d	236000	974			B*1510/*4501
	12 wks	6340	957			
CAP61	57 d	610	389	418	412	A*6602/ *6802,
	11 wks	17100	490			B*1401/ *4201
	33 wks	7380	537			
CAP63	34 d	202000	584	* 214000	* 235	A*0201/A*2301
	11 wks	305000	527			B*4501/B*4501
	29 wks	390000	336			
CAP84	22 d	9140	636	4350	479	A*2902/ *7401,
	14 wks	28500	611			B*1503/ *4407
	19 wks	2090	623			
CAP85	23 d	621000	419	< 400	682	A*3002/ *3002,
	13 wks	47300	469			B*0801/ *4501
	29 wks	6450	366			
CAP88	36 d	29400	963	38700	499	A*2902/ *6601,
	13 wks	38700	926			B*4501/ *5802
CAP174	28 d	474000	353	33600	303	A*0301/ *7401/
	28 wks	100000	457			B*4901/ *5802
CAP206	41 d	368000	365	315000	337	A*3204/ *741N,
	15 wks	113000	292			B*0702/ *440301
	33 wks	138000	325			
CAP210	14 d	468000	332	376000	344	A*6802/*6802,
	36 d	127000	461			B*1510/ *1510
	12 wks	50100	543			
CAP229	48 d	126000	558	22500	680	A*0101/ *0101,
	12 wks	8210	693			B*5801/ *5801
CAP239	36 d	95800	845	156000	793	A*0101/ *2902,
	11 wks	23900	1276			B*4201/*5801
	22 wks	224000	984			
CAP244	58 d	19200	557	14500	303	A*2301/ *3004,
	12 wks	39900	509			B*440301/ *5802
	28 wks	54300	298			
CAP248	62 d	55000	420	64600	313	A*0205/ *2902,
	12 wks	58600	346			B*1401/ *1503
CAP255	54 d	196000	693	18200	397	A*0301/ *8001,
	13 wks	96800	652			B*0801/ *1807
CAP256	42 d	56500	689	178000	291	A*2902/ *6601,
	13 wks	51600	475			B*1503/ *5802
	30 wks	626000	492			
CAP257	49 d	276000	450	10000	470	A*2301/*2902,
	14 wks	63400	591			B*4202/*4403
	30 wks	14900	489			

CAP = CAPRISA 002 acute infection study participants, ND = not done, d = days, mo = months, yrs = years.
Infection date was estimated as the midpoint between the last negative and first positive antibody test or as 14 days if the sample was PCR positive, antibody negative sample.
ND = not done; PI = post infection, c = copies, * final time point at endpoint = 9 mo PI.

30% positive amplicons from a set of replicate reactions and all sequences are screened for the absence of any heteroduplex peaks on the chromatograms. For CAP256, a total of 35 single genome amplicons derived gp160 envelope sequences at 1.5, 3, 6 and 24 months post infection. In addition, SGA derived gp160 sequences were generated from 2 weeks (n = 19) and 29 weeks (n = 26) post infection for CAP63 and 52 weeks post infection for CAP45 (n = 7).

This limiting dilution approach was initially described by Simmonds *et al*., (1990) and after several modifications (Edmonson and Mullins, 1992; Palmer *et al.,* 2005) it was finally modified by Salazar-Gonzalez *et al*. (2008).

4.2.3 Phylogenetic analysis

Near full-length genome sequences were aligned using ClustalW (Thompson *et al*., 1994) as implemented in BioEdit (Hall, 1999) with correction of indel placements being carried out by manual editing in BioEdit. The HXB2 subtype B strain was included as reference (root). Full-length genome sequences were split into individual gene fragments for translation to viral proteins and gene-specific analyses using Gene Cutter (http://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html).

A maximum likelihood (PHYML) phylogenetic tree depicting full genome evolution of HIV-1 subtype C was carried out using the F81 substitution model (Felsenstein, 1981) with gamma distribution ($\alpha = 4$) and ignoring recombination by using PHYML (Guindon and Gascuel, 2003) with automated best fit model selection and tree drawing in RDP3.34 (Martin *et al*., 2005a). The RDP3 beta 34 (http://darwin.uvigo.es/rdp/rdp.html) software programs, Recombination Analysis Program (RAP) on the HIV sequence Database (http://www.hiv.lanl.gov) and the Genetic Algorithm for Recombination Detection (GARD) in Datamonkey (http:// www. Datamonkey .org/) was used to identify recombination.

4.2.4 APOBEC signatures and hypermutation analysis

The Highlighter tool (http://www.hiv.lanl.gov/content/sequence/HighLighter/) was used to determine the total number of mutations in the full genome sequence with the APOBEC-associated and G>A mutations indicated. All sequences with APOBEC-associated signatures were further examined to identify any significantly hypermutated sequences with the hypermutation tool (http://www.hiv.lanl.gov/content/sequence/HYPERMUT/).

4.2.5 HLA data

HLA typing was performed on all 20 participants as described (Masemola *et al*., 2004). Briefly, DNA was extracted and high resolution HLA typing for HLA-A, -B, and –C was performed by sequencing exons 2, 3 and 4 using the Atria AlleleSeqr kits (Abbott Diagnostics) and Assign-SBT 3.5 (Conexio Genomics). Ambiguities were resolved with sequence-specific primers. Participants HLA data were kindly provided by Dr Clive Gray from the National Institute of Communicable Diseases (NICD, Sandringham, Johannesburg, Gauteng).

4.2.6 Identification of amino acid positions under positive selection

The Genetic Algorithm Recombination Detection (GARD) program in DATAMONKEY (Pond *et al*., 2006; http://www.datamonkey.org) was used to determine recombination breakpoints by searching for sequence fragment-specific phylogenetic trees and use fragment breakpoints to eventually assess goodness of fit for each segment by an information based-criterion. After breakpoint partitioning appropriate evolutionary models were used for each data set to identify sites under positive selection. The Internal Fixed Effects maximum Likelihood (IFEL) tool in Datamonkey (http://www.datamonkey.org) was used to identify codon positions subject to positive and

negative selective pressures in longitudinally sampled gene fragments (partitions). This method are able to distinguish between mutations that occur at the terminal ends or at the internal nodes of branches of the phylogenetic tree and can determine if selective pressures are different between two populations (Pond *et al*., 2006). Thus IFEL allows you to identify sites under selection that are associated with overall evolutionary change in longitudinally sampled data.

4.2.7 CTL epitope prediction and CTL escape

Putative CTL was defined as changes that occur within or flanking known or predicted HLA restricted epitopes. Known epitopes were defined using the list provided by the Los Alamos HIV Molecular Immunology Compendium 2006/2007. We also included significant HLA associations identified using a large database of subtype C sequences (Mathews *et al*., 2008; Rousseau *et al*., 2008). In addition MotifScan (http://www.hiv.lanl.gov/content/immunology/motif_scan/) and NetMHCpan (http://www.cbs.dtu.dk/services/NetMHCpan/) programs were used to identify putative CTL epitopes restricted by the participant's HLA-1 background. MotifScan uses known HLA-1 restricted CTL epitope binding motifs to predict HLA-peptide binding sites. NetMHCpan was used to supplement this analysis as N'gandu *et al*. (2007) reported that CTL responses against HIV-1 subtype C are poorly predicted by these known epitope-HLA binding motifs. NetMHCpan is a high-throughput computational bioinformatics tool which uses HLA and peptide sequence information to predict the affinity of peptide-HLA interactions (Nielsen *et al*., 2007). Validation of the NetMHCpan tool showed that more than 86% of peptides predicted to bind to the HLA molecule with a binding affinity <500 nM could experimentally be confirmed as HLA-binders (Nielsen *et al.*, 2007).

In this study we selected for each participant predicted peptides with a binding affinity of <500 nM. Changes in binding affinity in longitudinal sequences were also noted to identify escape (an increase in binding affinity over time) or reversion (a decrease in binding affinity over time, also from >500 to <500 nM). Known putative

CTL epitopes identified through HLA-peptide association algorithms were also used (Matthews *et al*., 2008). The number of peptides with putative CTL escape was enumerated for each viral protein and across the proteome.

4.2.8 Statistical analyses

Correlation of number of CTL epitopes, CTL escape mutations, reversion mutations and APOBEC induced mutations with CD4 counts and viral loads at 12 months post infection was done using Spearman nonparametric correlation with a two-tailed p-value ($p < 0.05$) as measure of significance. A two-tailed p-value ($p < 0.05$) was used as measure of significance. These statistical tests were carried out using GraphPad Prism® 5.0 (GraphPad Software Inc., CA, USA).

4.2.9 GenBank accession numbers

Envelope sequences derived by single genome amplification sequencing from the acute infection stage were submitted to GenBank under accession numbers FJ443149 to FJ443158 and FJ443292 to FJ443297 for CAP45 and accession numbers FJ443215 to FJ443220 and FJ443396 to FJ443408 for CAP63 (Abrahams *et al*., 2009).

4.3 RESULTS

4.3.1 HIV-1 complete genome analysis

A total of 52 near full-length subtype C HIV-1 genome sequences were generated from 20 individuals (Figure 4.1). Twenty-two near full-length HIV-1 subtype C genome sequences were generated from samples collected between 2 and 9 weeks post infection (Chapter 2), and a further 17 and 13 near full-length genome sequences were generated

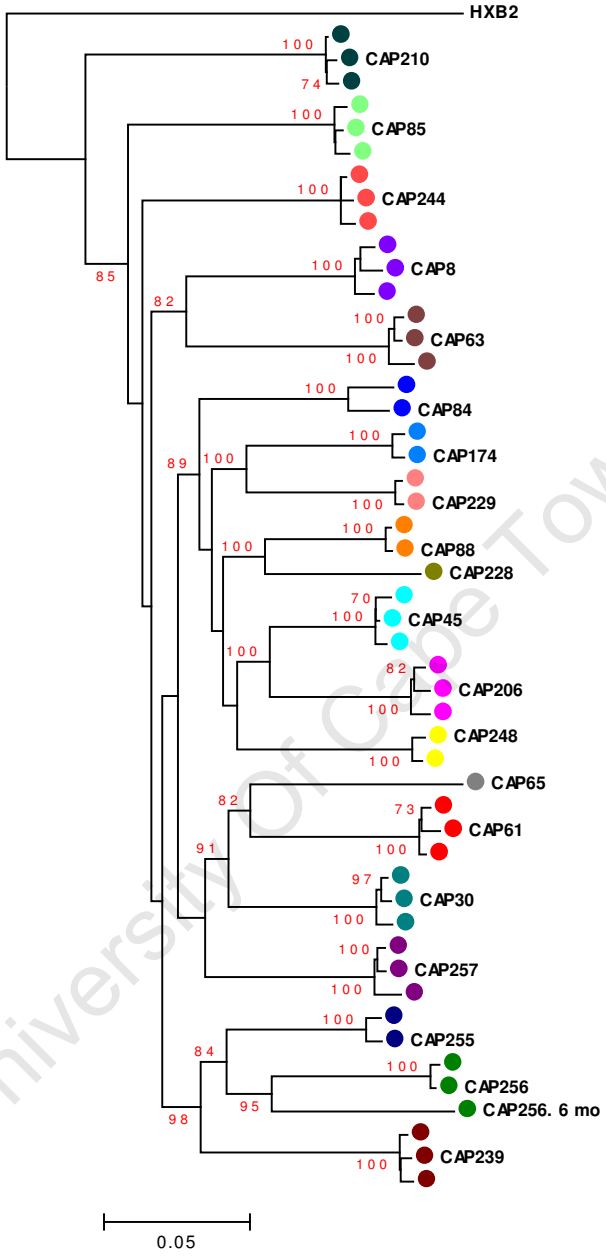


Figure 4.1 Maximum likelihood tree of 50 near full-length HIV-1 subtype C genomes from 20 individuals with primary infection sampled at 1 to 3 time points within 6 months post infection. The tree was rooted on the subtype B strain, HXB2 and 100 bootstrap replicates was used. Horizontal scale is 0.05.

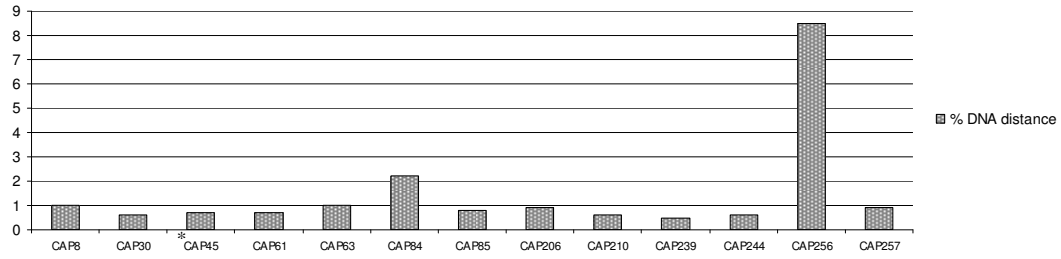


Figure 4.2 The % DNA distance between the earliest and 6 months post infection (PI) near full-length subtype C sequences for 13 participants. *CAP45 had viral loads below detection limit and we used the 3 months PI sequence as no 6 months PI whole genome amplicons were obtained.

from samples obtained longitudinally from about 3 and 6 months post infection respectively. Eighteen participants had complete genome sequences generated at more than one time points, and 11 of the 18 had HIV-1 complete genomes generated at 3 time points.

Phylogenetic analysis showed that sequences from sequential time points from 18 participants clustered with a percentage DNA distance of ≤1% indicating infection with a single strain (Figure 4.1 and Figure 4.2). However sequences from two participants (CAP84 and CAP256) had higher than expected diversity between longitudinal sampleswith a percentage DNA differences of 2.2 % and 8.5% respectively (Figure 4.2). Analysis of the *gag* gene showed that for CAP84, the sequences clustered together however the 3 and 6 month sequences were divergent from the 22 day post infection sequence (Figure 4.3). This individual was subsequently identified as infected with two epidemiologically unlinked HIV strains at three weeks post infection (co-infection with two strains) thus explaining the high diversity between these sequences (Zenda Woodman, University of Cape Town) and were excluded from further analysis. For CAP256, the 42 days and 3 months post infection time points clustered together, while the sequences generated from the 6 month time point clustered separately suggestive of superinfection. Interestingly, CAP256 had a more than one log increase (1.7 logs) in viral load at 4 months (15 weeks) post infection (Figure 4.4a), a feature commonly

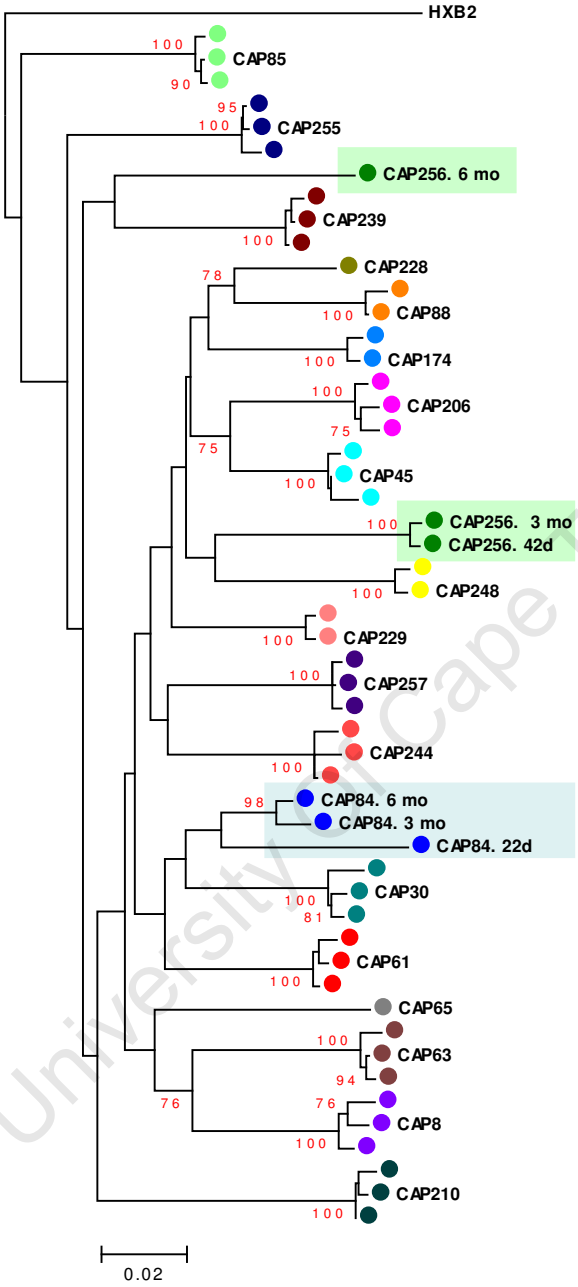
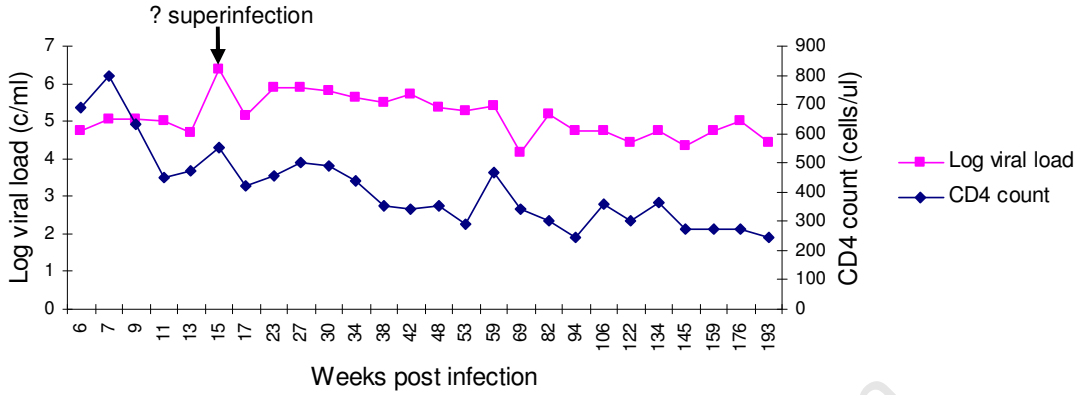


Figure 4.3 Maximum likelihood tree of full-length HIV-1 subtype C *gag* genes illustrating the relationship between sequences sampled at enrolment [2 to 8 weeks post infection (PI)], ~ 3 months, and 6 months PI. The subtype B strain, HXB2 was used as root and 100 bootstrap replicates were performed. Horizontal scale is 0.02.

(a)



(b)

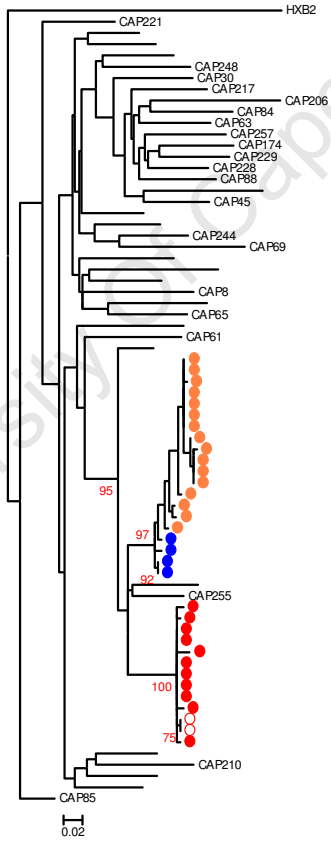


Figure 4.4 (a) Log viral load and CD4 T cell counts over time for CAP256. Arrow indicates the suspected time point of superinfection. (b) Maximum likelihood tree showing the relationship between partial *env* sequences (end C2 to beginning C3 region of gp120) sampled at the different time point indicated by colored dots. Red open circles: 6 weeks pi; red closed circles: 3 months (13 weeks); blue dots: 6 months (30 weeks), orange dots: 2 years (94 weeks) post infection. Horizontal scale is 0.02.

associated with superinfection.

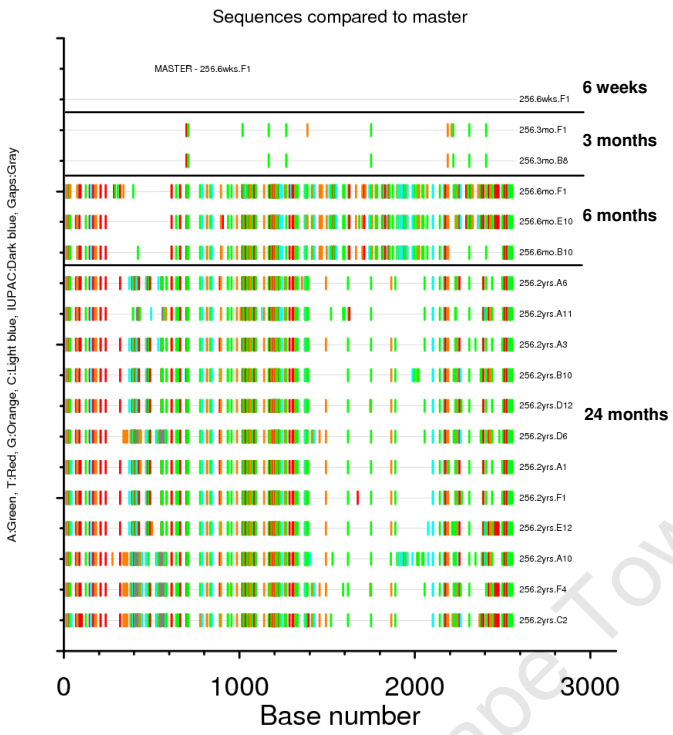
4.3.1.1 Recombination in CAP256 envelope glycoprotein.

To further characterize viral evolution in CAP256, a total of 35 complete *env* sequences were derived from single genome amplicons generated from samples collected at 6 weeks, 3 months, 6 months and 24 months post infection. Similar to full-length genome results, HIV-1 sequences at 6 weeks to 3 months post infection clustered together with a mean gp160 DNA distance of 0.3%; and virus variants present at 6 months and 2 years post infection were phylogenetically distinct from early variants with a mean DNA distance of 9.1% (Figure 4.4b). A comparison of the sequences to the sequence from 6 weeks post infection using the highlighter plot (Figure 4.5a) shows the differences in sequence corresponding to emergence of a new variant (second infection) at 6 months post infection. This variant contained regions of similarity to viruses sampled at 6 weeks and 3 months post infection. This superinfecting virus became dominant across most of the genome at 6 months post infection, however by 24 months appears to have acquired the C2 region of gp120 and the V5 to beginning of gp41 (Figure 4.5b) from the first infecting virus. Two recombinants (B10, E10) were identified at 6 months post infection and 6 recombinants (A1, A3, A6, C2, D12 and F1) at 2 years post infection (Figures 4.5a and 4.5b). Thus CAP256 was excluded from further analysis

4.3.2 Number of escaping CTL epitopes is not associated with CD4 counts or viral load at set point

Very little is known about the impact of CTL escape across the full-length genome and disease progression (Liu *et al*., 2006; Li *et al*., 2007). Many studies investigating CTL escape have either focused on specific genes (Borrow *et al*., 1997; Allen *et al*., 2005; Brumme *et al*., 2007; 2008; Crawford *et al*., 2007; Karlsson *et al*., 2007; Brockman *et al*., 2008; Boutwell *et al*., 2009) or these studies have been done on

(a)



(b)

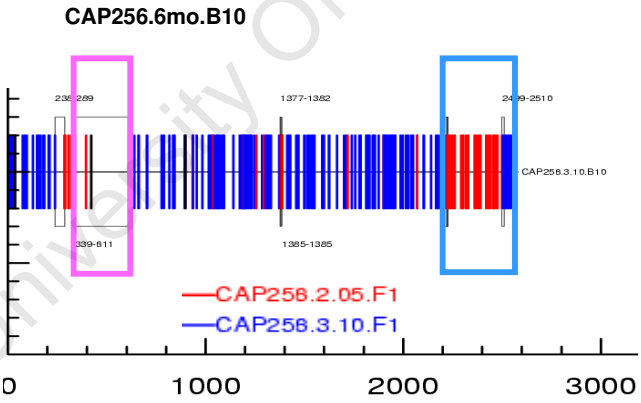


Figure 4.5 (a) Highlighter plot showing similarity in gp160 from CAP256 for longitudinal data (3, 6 and 24 months post infection) compared to the first sample collected at 6 weeks post infection. Each colored tick represents a nucleotide mismatch (A = green, T = red, G = orange, C = light blue, IUPAC = dark blue, gaps = grey) when compared to the 6 weeks (master) sequence. (b) Plot showing recombination between the "transmitted" (3 months post infection, red ticks) and a superinfecting variant (6 months post infection, blue ticks) detected at 6 months post infection. The C2 region is indicated by a pink box and the V5-gp41 region by a light blue box. The plot was generated using the RAP (Beta version) tool (http://www.hiv.lanl.gov).

samples collected during chronic infection. Here we identified putative escape mutations as amino acid changes in known or predicted epitopes associated with the participants HLA allele that evolved from high frequency to low frequency amino acids, and persisted between 3 and 6 months post infection. We also took into consideration HLA associated polymorphisms identified by Matthews *et al*. (2008). To account for escape associated with proteosomal cleavage sites we also investigated mutations within 8 amino acids of a known CTL epitope. Two individuals co-infected with two epidemiologically unlinked viruses as it is difficult to predict escape due to large shifts in viral sequences (CAP84 and CAP256). This analysis was performed only on the 11 individuals who were followed for 6 months, and for whom we have sequences from at least 3 time points. CTL escape mutations reported in this chapter are predicted and further work is required to confirm their role in cytotoxic T cell-mediated responses. This analysis excluded mutations or insertions and deletions in *env* that resulted in loss, gain or shift of N-linked glycosylation site motifs (NX[T/S]X, where X ≠ P) as these were classified as putative antibody escape.

Escape was most frequently detected in Env and Nef (n = 10), followed by Pol and Vif (both n = 8) and Gag (n = 7), with lower number of individuals with viruses that escape in Vpr, Vpu, Rev and Tat (n = 5, 3, 2 and 1 respectively) (Figure 4.6). Epitope regions associated with CTL escape in these individuals (Appendix D3) are summarized in Tables 4.2, 4.3 and 4.4. When this was adjusted for protein size (Figure 4.7), the

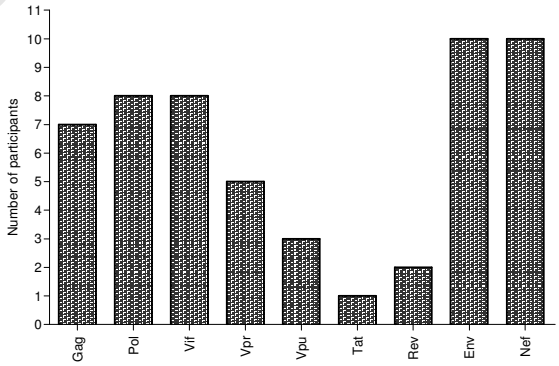


Figure 4.6 Number of participants with evidence of escape in each HIV-1 protein within 6 months post infection.

Table 4.2 Putative escaping CTL epitopes in Gag and Pol.

Protein	Epitope region	HLA	Comments
Gag	QY**M**LKHLVWAS	A*2301	Published
	KKHYMLKHL**V**W	A*2301	Published
	*DIYKRWIILGLN**K**IVRM*	B*0801	Published
	WIILGLNK**I**VRMYSPVSIL	A*0201	Published
	IILGLN/KIVRMYSPVSIL	A*0201	Published
	VRMYSPVSIL/D**I**	A*2304	Predicted (184.99 nM)
	ALNP/GLLET	A*0201	Published
	SLYNTVA*TL/YCVHE**D**T*	A*0201	Published
	GH**Q**AAMQMLKE	A*0201	Published
	F**L**QN**R**PEPT	A*0201	Published
	VQNAQGQ/MVHQPLSPR	A*6602	Predicted (60.3 nM), A*6603 published
	NIM**M**QRGNI	B*0801	Predicted (232.43 nM)
	MTNPPIPV/G**G**	A*6802	Published, A2
	D**I**AGTTSTLQEQI	A*6802	Published, A2
	S**T**LQEQVAM	B*5801	Published
Pol	ALTAICE**E**M	A*0201	Published
	SWTVNDIQ/KLVGKLNWA	A*0201	Published
	HTN/DVKQLTEAV	A*6802	Published
	DIVPLTEEA/EL**E**	A*6802	Predicted (272.92 nM)
	ELAENREIL/KEPV**H**	A*6802	Published, A2
	DIQ/KLVGKLNWA	A*0201	Published
	NWRAMAA**D**F	A*0201	Published
	IWQLDCTHL/EGKV**I**	A*2301	Predicted (200.47 nM)
	QAEHL/KTAVQ**M**AVF	A*0201	Published
	RGRQKVV**S**L	B*0801	Published
	GPKVKQWPL/TEE**K**	B*0702	Published
	WEFVNTP**P**L	B*4403	Predicted (123.76 nM)
	SESEIV**N**QI	B*0702	Published
	SESELVNQI/IEELI**K**	B*4403	Predicted (87.13 nM)
	LIKKE/KVYLSWVPA	B*0702	Published
	TVNDIQK/LVGKLNWASQIY	B*1510	Published, B15
	G/KLNWASQIYP	A*3204	Predicted (23.4 nM)
	PIVLP**E**KE**S**W	B*5801	Published
	WWAGIKQEF/GIP**Y**	A*2301	Predicted (83.62 nM)
	SAAVKAACW/WAGI**Q**QEF	B*5801	Published
	PYQIYQEPF	A*2301	Published
	VA**V**HVASGY	A*2902	Predicted (181.27 nM)
	MAVFIHNFK/R**K**	A*2902	Predicted (276.27 nM)
	TKELQ/KQITKIQNF	A*2301	Predicted (282.83 nM)

*Published = HIV Sequence Compendium, http://www.hiv.lanl.gov/content/immunology/; HLA-association = Matthews *et al*., 2008; Predicted = NetMHCpan prediction (affinities <500 nM).

Table 4.3 Putative escaping CTL epitopes in Vif, Vpr, Vpu, Tat, Rev and Nef.

Protein	Epitope region	HLA	Comments
Vif	EWRLRRY**S**TQVD	B*1510	Published
	WRLRR/YSTQVDPGL	A*6802	Predicted (442.54 nM)
	KPK**K**RKPPL	B*0801	Predicted (63.45 nM)
	GLADQLIHI/HY**F**	A*0201	Published
	NGWFYR**H**HY	B*5801	Published
	KVG/SLQYLALTA	A*0201	Published
	SLVKHH**M**YV	A*0201	Published
	RTWN**S**LVKH	A*3204	Predicted (378.02 nM)
	WNNPQKI/RGRRGNHTM	B*0702	Published
	WHLGHG**V**SIEWRLRKY	B*1510	Published
	TE**S**AIRQAI	B*4403	Predicted (302.83 nM)
Vpr	SLGQY**I**YET	A*0201	Published
	FPRP**W**LHDL	B*4201	Published
	I/R**I**LQQLLFI	A*0201	Published
	QHIY/ETYGDTWTGV/EAI	A*6802	Published
	QHIYETYGDTW	A*2301	Published
	WLHS**L**GQYI	B*4202	Published
Vpu	MEQLRLLDV/**N**	B*0801	Predicted (270.27 nM)
	IIAVIL**A**IV	A*6802	Predicted (62.61 nM)
	AIIVWTI**T**YLEY	A*3002	Predicted (346.57 nM)
Tat	CS**Y**HCLACF	B*1510	Published
Rev	**E**ALL/LAVRTIKLL	A*0201	Predicted (43.85 nM)
	QI**H**SISERIL	A*0201	Published
	RHI/REISERILS	B*4501	Predicted (484.82 nM)
	RNRRRRWRAR**QR**	B*0801	Predicted (406.53 nM)
Nef	**E**GLI/**Y**S**K**KRQEIL	B*0801	Published
	EVLRW**K**FDSQ/LARRH**I**A	B*0801	Published
	REVLQ/WKFDSSLAF	B*1510	Published
	QEEE**E**EVGFPV AQEEE**E**EVGFPV QEEE**G**VG**F**PV AQEEEK**E**VGFPV	B*4501	Published
	STN**A**D**M**AWL	A*6802	Predicted (114.25 nM)
	IL**D**LWVYHTQG	A*0201	Published
	G**Y**FPDWQNY	A*3002	Published
	AASQDLGKY	A*3002	Predicted (228.05 nM)
	MA**R**EKHPEF	B*5801	Predicted (53.59 nM)
	K**A**AVDLSFF	B*5801	Published
	YTPGPGVRYPL/TFG**W**C	B*5801	Published
	Q**E**ILDLWVY	B*4403	HLA association

*Published = HIV Sequence Compendium, http://www.hiv.lanl.gov/content/immunology/; HLA-association = Matthews *et al*., 2008; Predicted = NetMHCpan prediction (affinities <500 nM).

Table 4.4 Putative escaping CTL epitopes in Env.

Protein	Epitope region	HLA	Comments
Env	WWIWSSLGF/WM**L**	A*2301	Predicted (28.64 nM)
	LWVTVYYGV/PVW**K**	A*2301	Predicted (480.41 nM)
	KD/K**K**QKV**Y**ALF/	A*2301	Predicted (463.57 nM)
	K**Q**KEYALFY	A*3004	Predicted (137.19 nM)
	E**I**M**I**RSENL	B*0801	Predicted (184.12 nM)
	WYIK**I**FII	A*2301	Published
	RS**I**RLVNGF	B*0801	Predicted (250.03 nM)
	IR/LVSGFL**S**L**A**	A*6802	Predicted (62.65 nM)
	WMLM**I**YNVV	A*0201	Predicted (46.25 nM)
	CSFNATTER/RDK**K**	A*3402	Predicted (86.87 nM)
	CS**F**NTTTEI	A*6802	Predicted (142.21 nM)
	CSFN**A**TTEL	A*6802	Predicted (402.5 nM)
	VYEREVH/ NVWATHACV	A*0201	Published
	ALFYRPDIV/LLSQN**D**	A*0201	Predicted (89.3 nM)
	T/IIVHLNESV	A*0201	Predicted (460 nM)
	SGGD/LEITTHS/FNCR	A*0201	Published
	SEITGLLLT/RDGG**R**	B*4501	Predicted (193.6 nM)
	LLQYWS**R**EL	A*0201	Published
	SLFDTIAIA/VA**E**	A*6802	Published
	LLDSIAI**T**V	A*0201	Published
	ATAIAVAE**A/T**DRI	A*6802	Published
	W**K**NDMVDQM	B*4403	Predicted (123.08 nM)
	TEN/FNMWKNDMV	A*6802	Published
	NM/KDNWRSELY	A*3002	Predicted (126.89 nM)
	YTN**T**IYRLL	A*6802	Predicted (45.9 nM)
	HFPNK**T**ITF	A*2301	Predicted (45.18 nM)
	NY**T**QTIYQL	A*2301	Predicted (57.46 nM)
	RIIEV**I**QRF	A*0201	Published
	RYLG**S**LVQY	A*3002	Published
	KYLG**S**LVQY	A*2902/A*2301	Predicted (190.19/ 202.06 nM)
	KPCVKLTPLC/VTL**N**	B*0702	Published
	VTFDPIPIHY/**C**APA	A*2902	Published
	THI/FNCRGEFFY	A*2902	Published
	FSITNWLW**Y**	A*0101/ B*5801	Predicted (34.96/ 61.5 nM)
	GM**L**RNYQQW	A*2301	Predicted (394.97 nM)
	YE**K**EVHNVW	B*4403	Predicted (472.64 nM)
	NYTNSIYKL/LE**E**	A*2301	Predicted (100.12 nM)
	D/KWQNLWNWF	A*2301	Predicted (20.07 nM)

*Published = HIV Sequence Compendium, http://www.hiv.lanl.gov/content/immunology/; HLA-association = Matthews *et al.*, 2008; Predicted = NetMHCpan prediction (affinities <500 nM).

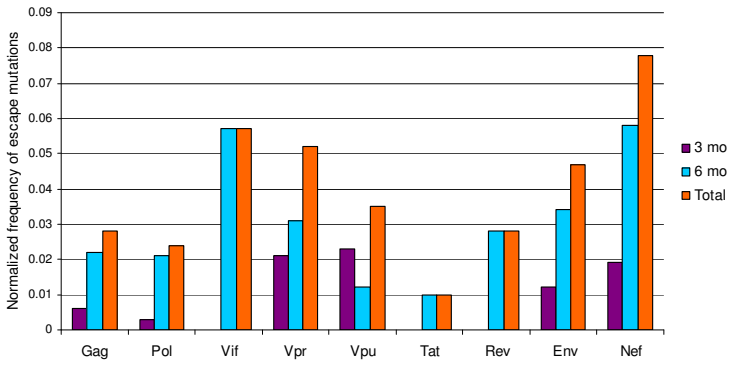


Figure 4.7 The normalized frequency of CTL escape mutations in HIV-1 proteins at 3 and 6 months post infection, as well as overall.

highest amount of escape per amino acid was found in Nef followed by Vif > Vpr > Env > Vpu > Gag/ Rev > Pol > Tat. The earliest escape was seen in Nef, Env, Gag, Pol, Vpu and Vpr with the highest density of escape in Nef, the protein which is also elicits the most immunodominant CTL response in acute infection (Gray *et al*., 2009). None of the participants developed CTL escape within the first 3 months in Vif, Tat and Rev.

In chronic subtype C infection, the number of HLA-B associated mutations in predicted CTL epitopes has been shown to be associated with lower viral loads and higher CD4 counts (Matthews *et al*., 2008; Rolland *et al*., 2008). However we find no association between CD4 or VLs and the total number of predicted CTL epitopes with escape associated with either HLA-A or B (Figure 4.8). We also found no association between CD4 and VLs and the total number of escaping CTL epitopes in Gag, Pol and Env.

4.3.3 Increase in reversion mutations are associated with lower CD4 counts at set point

While the virus tries to escape host immune responses through CTL escape mutations it also tries to restore or maintain replicative fitness through reversion of

(a) (b)

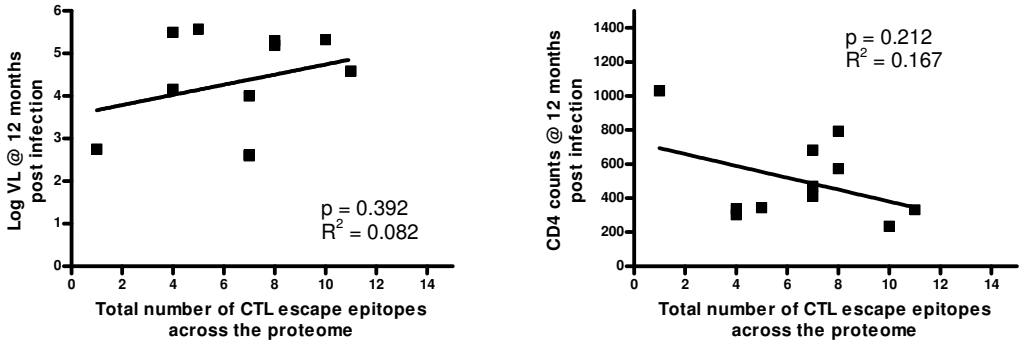
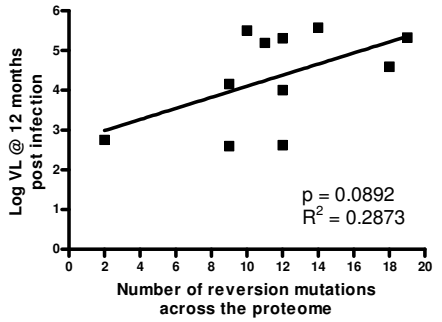


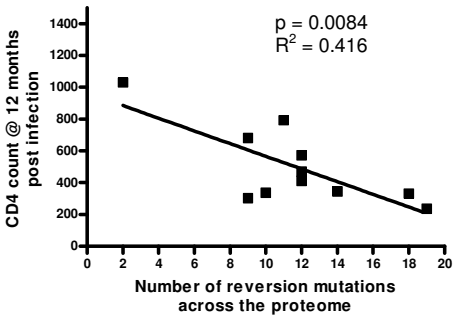
Figure 4.8 Total number of CTL escape epitopes compared to the (a) viral load (VL) and (b) CD4 counts at 12 months post infection.

deleterious mutations. Reversions appear during early primary infection (Li *et al*., 2007) and are thought to be associated with restoration of the fitness cost of transmitted escape mutations. Here we investigated the effect of reversions within six months of infection on the viral set point at 12 months post infection, with reversion defined as change of amino acid to consensus amino acid or to an amino acid present at a high frequency in the database.

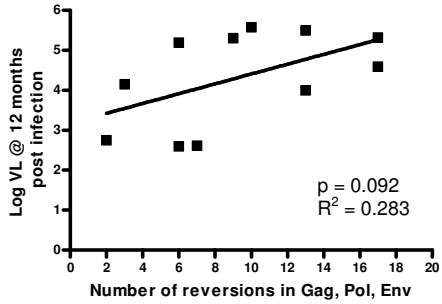
The total number of reversions across the HIV-1 proteome was inversely correlated with the CD4 counts at 12 months post infection ($p = 0.0084$, $R^2 = 0.5558$) where increase in the number of reversions was associated with lower CD4 counts (Figure 4.9a). However only a trend was seen between number of reversions and viral loads at 12 months post infection ($p = 0.0892$, $R^2 = 0.2873$). Reversions in the structural proteins (Gag, Pol and Env) were also inversely correlated with CD4 counts ($p = 0.032$, $R^2 = 0.416$) and were associated with increased viral loads at 12 months post infection however this was not significant ($p = 0.092$, $R^2 = 0.283$) (Figure 4.9b). Similar trends were seen when Gag and Env were assessed alone (Figure 4.9c and d).

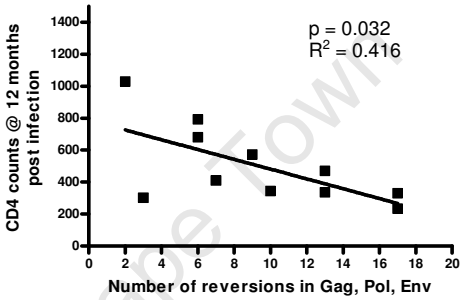
(a)



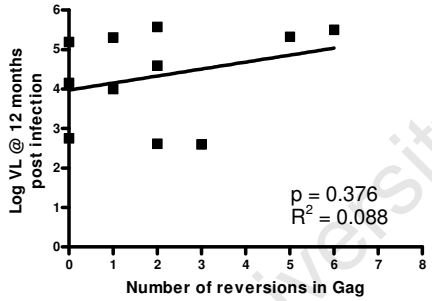


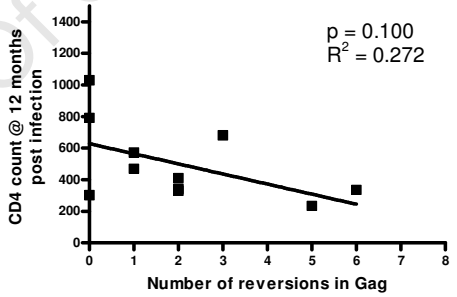
(b)



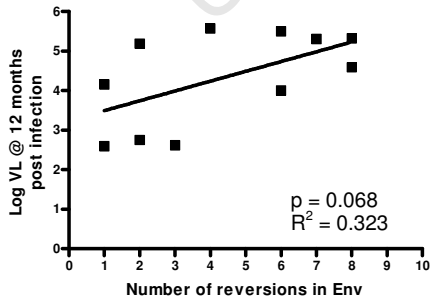


(c)





(d)



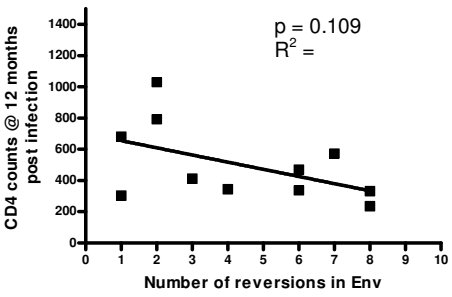
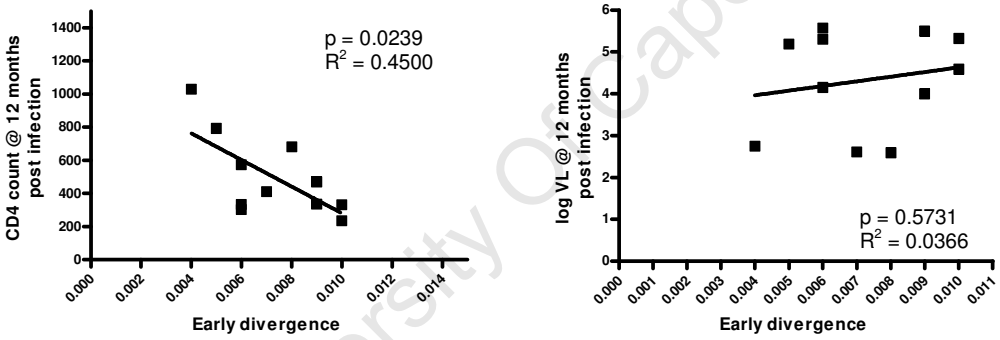


Figure 4.9 Correlation between the number of reversions across (a) the proteome, (b) structural proteins [Gag, Pol, and Env], (c) Gag and (d) Env with viral loads and CD4 counts at 12 months post infection.

4.3.4 Increased HIV-1 divergence is associated with lower CD4 counts at set point

Exploring the impact of viral divergence on viral set point at 12 months post infection we found that an increase in the genetic distance over time (divergence) was significantly associated with lower CD4 counts ($p = 0.0239$, $R^2 = 0.4500$) (Figure 4.10a). However, no association was found between viral divergence and viral load at set point ($p = 0.5731$, $R^2 = 0.0366$). Early HIV-1 divergence was also significantly associated with the total number of CTL escape epitopes ($p = 0.0373$, $R^2 = 0.3982$) and the total number of reversion mutations (p = 0.0162, R2 = 0.4918) suggesting that both escape and reversion were significant contributors to early viral evolution (Figure 4.10b).

(a)



(b)

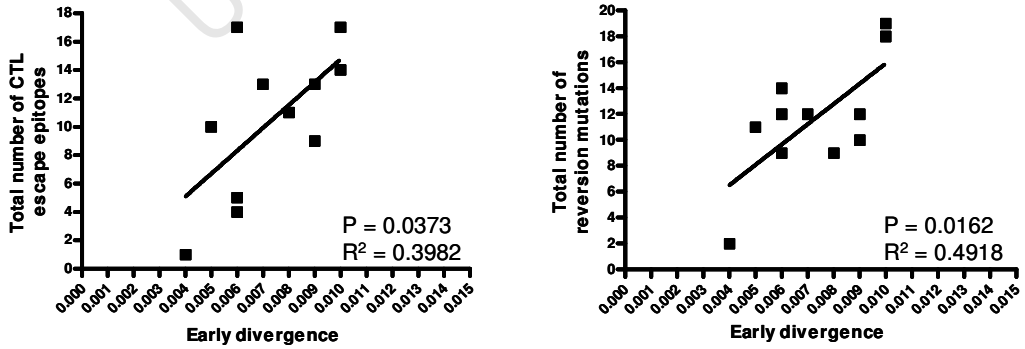


Figure 4.10 Early viral divergence compared to (a) CD4 count and VL, and (b) to the number of escaping CTL epitopes and reversion mutations.

4.3.5 APOBEC signature mutations are associated with decreased CD4 counts

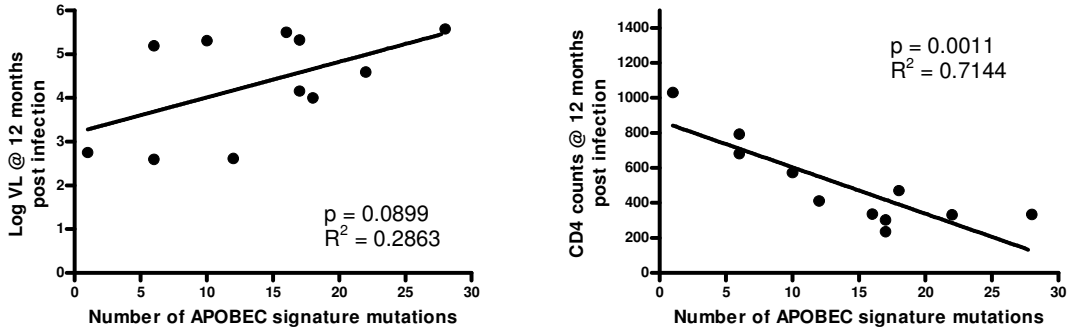
Studies have shown no association between the level of APOBEC3G associated mutations in viral RNA and viral load (Ulenga *et al*., 2008), however the degree of proviral hypermutation was significantly associated with CD4 counts in HIV-1 infected individuals (Land *et al*., 2008). Here we investigated the effect of APOBEC induced mutations in viral RNA on viral set point and CD4 count at 12 months post infection.

Interestingly, we found that an increased number of APOBEC signature mutations was significantly associated with lower CD4 counts ($p = 0.0011$, $R^2 = 0.7144$) and there was a trend towards increased viral loads ($p = 0.0899$, $R^2 = 0.2863$) (Figure 4.11a). In this study, none of the viral sequences were significantly hypermutated. The APOBEC associated mutations were found to contribute to viral diversity as increased numbers of APOBEC signature mutations was associated with an increase in viral divergence (Figure 4.11b) although this association was not significant ($p = 0.1240$, $R^2 = 0.2423$). In addition, the number of reversions was significantly associated with the number of APOBEC signature mutations ($p = 0.0206$, $R^2 = 0.4660$) (Figure 4.11c) suggesting that APOBEC may be a mechanism the virus uses to restore fitness which may account for the association between APOBEC signatures and lower CD4 counts. Although studies have suggested that the virus also uses APOBEC to generate CTL escape mutations (Wood *et al.,* 2009), we found no association between APOBEC signature mutations and CTL escape ($p = 0.7196$, $R^2 = 0.015$).

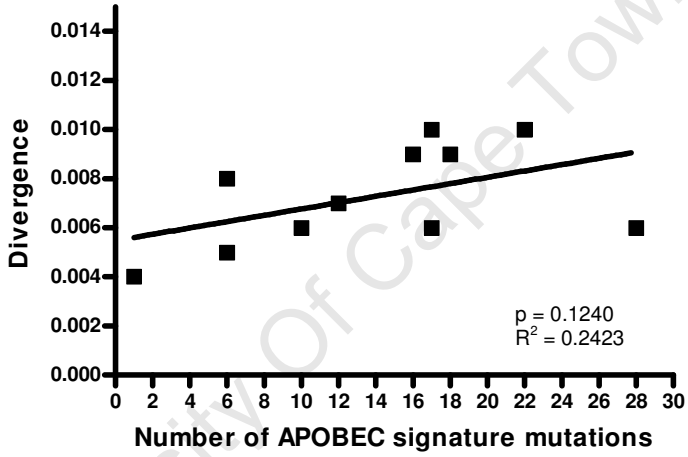
4.3.6 Longitudinal evolutionary changes in a slow (CAP45) and rapid progressor (CAP63)

Previous data reported on an analysis of single sequences generated through limiting diluting, cloning and sequencing. New approaches use the single genome amplification approach to eliminate errors introduced through *in vitro* recombination or bias sampling through cloning. Here we generated a total of 39 full-length genomes

(a)



(b)



(c)

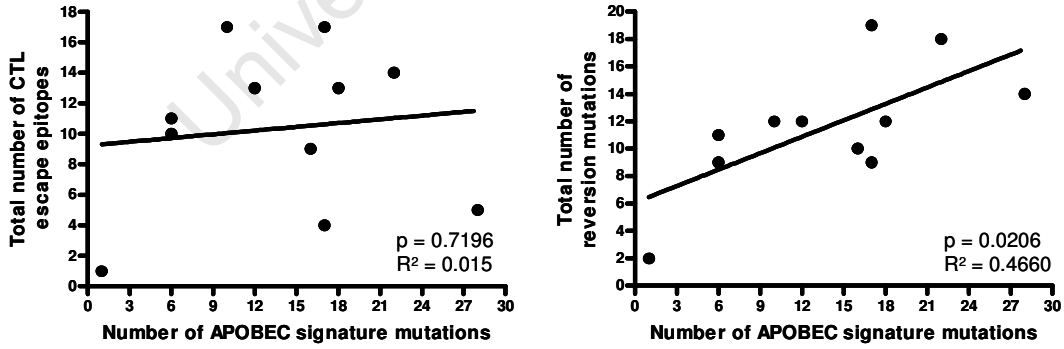


Figure 4.11 Correlation of APOBEC signature mutations with (a) viral load and CD4 count at 12 months post infection, (b) early divergence and (c) CTL escape epitopes and reversions.

directly sequenced from single genome or limiting dilution amplicons from two individuals, CAP45 and CAP63. Both individuals were recruited during acute infection with CAP45 were classified as Fiebig stage I/II (RNA positive, antibody negative) (Fiebig *et al*., 2003), whereas CAP63 was in Fiebig stage III [RNA positive, antibody-enzyme immuno assay (EIA) positive but Western blot negative]. Full-length genomes were generated for CAP45 at 2 (n = 3), 5 (n = 6), 12 (n = 3), 16 (n = 1) and 65 (n = 3) weeks post infection; and for CAP63 at 2 (n = 10), 11 (n = 7), 29 (n = 2) and 37 (n = 3) weeks post infection. SGA-derived full-length gp160 sequences from 2 weeks and 52 weeks post infection were available for CAP45 (n = 16 and n = 7 respectively), as well as for CAP63 at 2 (n = 19) and 29 (n = 26) weeks post infection. The generation of multiple (~10) sequences from two individuals in acute infection (antibody negative or indeterminate) enabled us to derive the transmitted consensus sequence which allowed us to more accurately investigate the effect of HIV-1 specific cytotoxic T cell mediated immunity on viral evolution over time.

CAP45 was defined as a slow progressor as she consistently controlled viral load to below 2000 copies per ml. CAP63 was defined as a rapid progressor and she was terminated from the study and initiated on anti-retroviral therapy at nine months post infection as her CD4 count reached below 200 cells/mm^3 (Table 4.1). CAP45 carries the following HLA alleles: A*2301or A*2304, A*2902 or A*2903 and B*1510/ B*4501; and CAP63 carries the HLA-A*0201/A*2301 and HLA-B*4501/B*4501 alleles. The envelope SGA sequences from 2 weeks post infection for these two individuals had been used to determine the viral heterogeneity at transmission and in both, a single variant was responsible for establishing infection (Abrahams *et al*., 2008).

Comparison of HIV-1 full-length genome sequences derived after cloning with directly sequenced SGA-derived genomes shows the presence of mutations in the clonal-derived sequences that are not present in any of the SGA derived sequences, as well as ambiguous bases shown in dark blue for directly sequenced amplicons in Figure 4.12. Salazar-Gonzalez *et al*. (2008) reported that bulk PCR and cloning of gp160 results in a significant increase in viral diversity compared to direct sequenced SGAs. *Taq*-induced

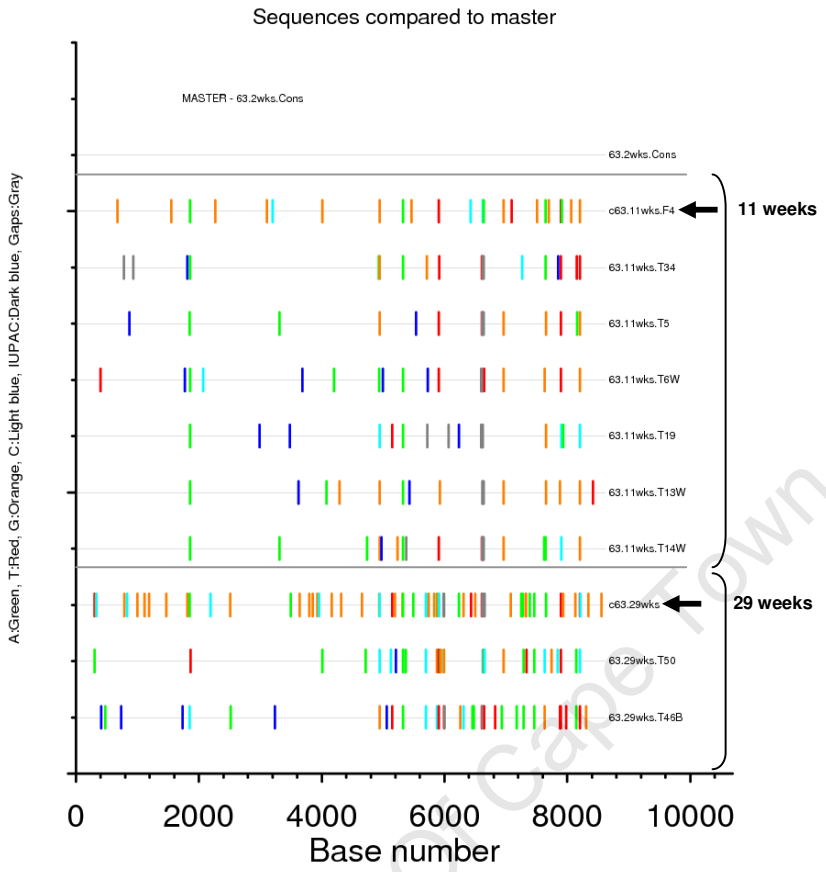


Figure 4.12 Comparison of directly sequenced HIV-1 full-length genomes and clone-derived sequences (indicated by black arrows) from CAP63 to the 2 weeks post infection consensus sequence.

errors (ambiguous bases) and recombination during bulk PCR was identified as the cause and found that the effect of *Taq*-induced errors was more pronounced in cloned bulk PCR sequences. Similarly we find increase numbers of guanosine (G) (shown as orange ticks) in the clonal-derived full-length sequences which may represent artifacts of *Taq*-induced G/A mixed bases. Of the 17 putative escape mutations that we identified in CAP63, we find only 14 were supported by the SGA analysis. However, in addition to the single CTL escape identified in the analysis of CAP45 clones being supported in SGA analysis, a further seven were identified. Using the expanded SGA derived dataset we identified CTL escape in CAP45 at one epitope in Nef within 12 weeks post infection, at one epitope in Rev at 16 weeks post infection; and one in Pol and two in Vif at 65 weeks post

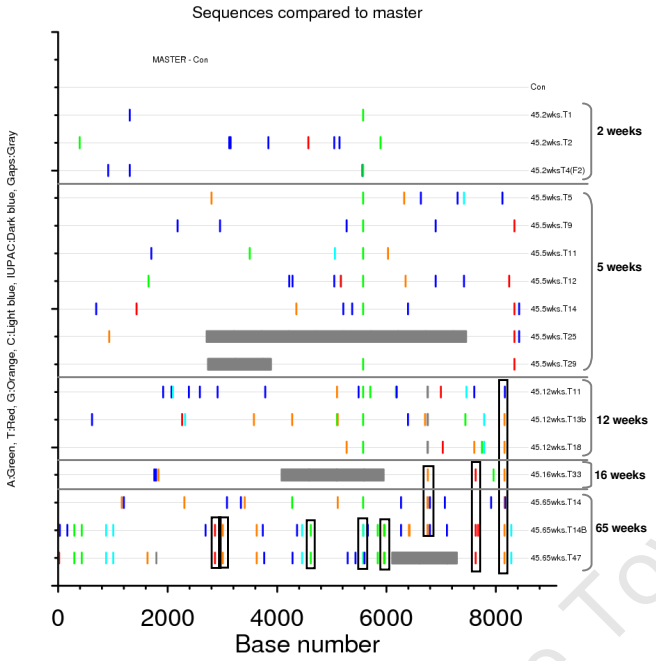
infection as well as in three epitopes in Gp120 at 52 – 65 weeks post infection (Figure 4.13a, Appendix D4). For CAP63 escape was seen in a total of only 14 CTL epitopes (Figure 4.13b, Appendix D5) namely: one each in Gag, Vpr, and Rev, two epitopes in Nef, three epitopes in Pol, in six epitopes for Env and no escape was identified in Vif.

Oscillation of amino acid residues within epitopes has been seen in both SIV (Friedrich *et al*., 2004) and HIV (Borrow *et al*., 1997; Goonetilleke *et al*., 2009) infection. Similarly shuffling escape was seen in CAP63 in putative CTL escape epitopes in Pol (n =1), Vpr (n =1), Rev (n = 1), and Env (n = 4) (Figure 4.14a). Here, shuffling escape also involved different amino acid positions ("clustered mutations") within the same putative CTL epitope or in adjoining regions (Figure 4.14b). Thus 7/14 (50%) putative CTL epitopes from CAP63 was experiencing shuffling escape. Only 1/8 CTL epitopes (Rev) from CAP45 showed shuffling between wild-type and escape residues.

We then identified which codons in the viral proteins were under positive selection pressure using a recombination aware Internal Fixed Effects maximum Likelihood (IFEL) method with p-value <0.2 regarded as significant. Three amino acid positions under positive selection were associated with APOBEC activity as shown here in Gp120 for CAP45 and four positions for CAP63 (Figure 4.15). The slow progressor, CAP45, had ten sites under positive selection, one site in Pol and nine in Gp120 (Figure 4.16a). Six of the nine sites in Gp120 were located in three known CTL epitopes associated with the participants HLA, and three were in two sites associated with glycosylation shifts and therefore thought to be associated with antibody pressure. The only positively selected site in Pol was also located in a predicted CTL epitope restricted by HLA-A*2902/ -B*4501 (LEGK**I**ILVAVHVASGY).

In contrast, the rapid progressor, CAP63 had twenty-one sites under positive selection. These were located in Gag (n = 2), Pol (n = 5 with 4/5 in RT), Vif (n = 1), Vpr (n = 1) and Env (n = 11 in Gp120, n = 1 in Gp41) (Figure 4.15b). Positively selected sites located in putative CTL epitopes were identified in one epitope in Gag (RSLYNTVAT**L**YCVH) and Vpr (IR**I**LQQLLF), in three epitopes in RT (ALTEIC**E**EM,

(a)



(b)

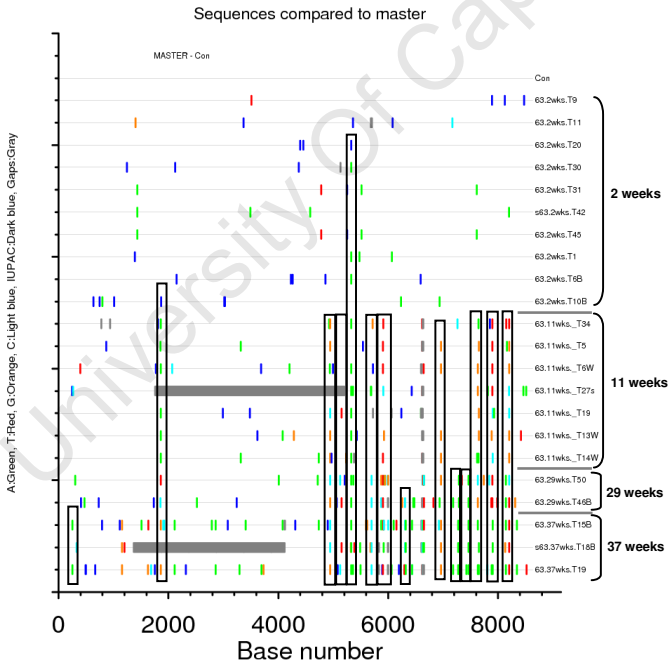


Figure 4.13 (a) Nucleotide polymorphisms in the full-length genome of CAP45 highlighting the putative escape mutations [boxed]. (b) Nucleotide polymorphisms in the full-length genome of CAP63 highlighting the putative escape mutations [boxed].

(a)

Pol		Vpr	
CONSENSUS_C	KIKALTAICEEMEKEGKI	CONSENSUS_C	TWTGVEAIIRILQQLLF
63.2wks.T9	.....E.....	63.2wks.T9	.....L.....
63.2wks.T11	.....E.....	63.2wks.T11	.....L.....
63.2wks.T20	.....E.....	63.2wks.T20	.....L.....
63.2wks.T30	.....E.....	63.2wks.T30	.....L.....
63.2wks.T31	.....E.....	63.2wks.T31	.....L.....
63.2wks.T31B	.....E.....	63.2wks.T31B	.....L.....
63.2wks.T32	.....E.....X..	63.2wks.T32	.....L.....
s63.2wks.T42	.....E.....	s63.2wks.T42	.....L.....
63.2wks.T45	.....E.....	63.2wks.T45	.....L.....
63.2wks.T1	.....E.....	63.2wks.T1	.....L.....
63.11wks.T34	.....E..K.....	63.11wks.T34	.....L..M.....
63.11wks.T5	.....K.....	63.11wks.T5	.....L..S.....
63.11wks.T6W	.....E..K.....	63.11wks.T6W	.....L.K.....
63.11wks.T19	.....E..K.....	63.11wks.T19	.....L..T.....
63.11wks.T13W	.....E..K.....	63.11wks.T13W	.....L..S.....
63.11wks.T14W	.....E..K.....	63.11wks.T14W	.....L..S.....
63.29wks.T50	.....E..D.....	63.29wks.T50	.....L..T.....
63.29wks.T46B	.....	63.29wks.T46B	.....L..M.....
63.37wks.T15B	.....E..G.....	63.37wks.T15B	.....L..T.....
63.37wks.T19	.....E..G.....	s63.37wks.T18B	.....L..T.....
		63.37wks.T19	.....L..M.....

(b)

Env		Rev	
CONSENSUS_C	GEIKNCSFNITTELRDKKQKVYALFYR	CONSENSUS_C	NRRRRWRARQRQIHSISERILS
63.2wks.T9	.........A.....	63.2wks.T9	.....Q...K.....
63.2wks.T11	.........A.....	63.2wks.T11	.....Q...K.....
63.2wks.T20	.........A.....	63.2wks.T20	.....Q...K.....
63.2wks.T30	.........A.....	63.2wks.T30	.....Q...K.....
63.2wks.T31	.........A.....	63.2wks.T31	.....Q...K.....
63.2wks.T31B	.........A.....#...	63.2wks.T31B	.....Q...K.....
63.2wks.T32	.........A.....	63.2wks.T32	.....Q...K.....
s63.2wks.T42	.........A.....	s63.2wks.T42	.....Q...K.....
63.2wks.T45	.........A.....	63.2wks.T45	.....Q...K.....
63.2wks.T1	.........A.....	63.2wks.T1	.....Q...K.....
		s63.2wks.4223	..K.....Q...K.....
63.11wks.T34	.........A..V.....	63.11wks.T34	.....Q...K.....
63.11wks.T5	.........AS.....	63.11wks.T5	.....Q..QK.....
63.11wks.T6W	.........AS.....	63.11wks.T6W	.....Q...K.....
63.11wks.T27s	.........A..A.....	63.11wks.T27s	.....Q...K.....
63.11wks.T19	.........A.....	63.11wks.T19	.....Q..QK.....
63.11wks.T13W	.........A.....E.....	63.11wks.T13W	.....Q..QK.....
63.11wks.T14W	.........AS.....	63.11wks.T14W	...K.....Q...K.....
63.29wks.T50	.........S.....	63.29wks.T50	.....Q...K.....
63.29wks.T46B	.........S.....	63.29wks.T46B	.....Q...K.....
s63.29wks.A2	.........S.....	s63.29wks.A2	.....Q...K.....
s63.29wks.A3	.........A...W.....	s63.29wks.A3	.....Q..NK.....
s63.29wks.A4	.........AS.....	s63.29wks.A4	.....Q...K.....
s63.29wks.A5	.........A..I...Q.....	s63.29wks.A5	.....E..Q...K.....
s63.29wks.A6	.........A...I.....	s63.29wks.A6	.....Q...K.....
s63.29wks.A7	.........A.....	s63.29wks.A7	.....Q...K.....
s63.29wks.A8	.........T...I.....	s63.29wks.A8	.....T..Q...K.....
s63.29wks.A9	.........A...I.....	s63.29wks.A9	.....Q..NK.....
s63.29wks.A10	.........A..A.....	s63.29wks.A10	.....Q..NK.....
s63.29wks.A11	.........S.....	s63.29wks.A11	.....G..Q...K.....
s63.29wks.A12	.........A...I.....	s63.29wks.A12	.....T..Q..NK.....
s63.29wks.A13	.........A...I.....	s63.29wks.A13	.....Q..NK.....
s63.29wks.A14	.........T...I.....	s63.29wks.A14	.....Q..YQ.....
s63.29wks.A15	.........A....W.....	s63.29wks.A15	.....Q..NK.....
s63.29wks.A16	.........A....W.....	s63.29wks.A16	.....E..Q...K.....
s63.29wks.B1	.........A....W.....	s63.29wks.B1	.....Q...K.....
s63.29wks.B2	.........A....W.....	s63.29wks.B2	.....Q..NK.....
s63.29wks.B3	.........A..V....E.....	s63.29wks.B3	.....Q..NK.....
s63.29wks.B4	E.........A..A.....	s63.29wks.B4	.....Q...K.....
s63.29wks.B5	.........S.....	s63.29wks.B5	.....Q...K.....
s63.29wks.B9	.........A...I.....	s63.29wks.B9	.....Q..NK.....
s63.29wks.C3	.........T...I.....	s63.29wks.C3	.....Q..NK.....
s63.29wks.C4	.........S.....	s63.29wks.C4	.....Q..NK.....
s63.29wks.TA1	.........A.....	s63.29wks.TA1	.....E..Q...K.....
s63.29wks.TA3	.........A...I.....	s63.29wks.TA3	.....G..Q...K.....
s63.29wks.TA4	.........A....W.....	s63.29wks.TA4	.....Q..NK.....
63.37wks.T15B	.........A..I....Q.....	63.37wks.T15B	.....E..Q...K.....
s63.37wks.T18B	.........A...I.....	s63.37wks.T18B	.....Q..NK.....
63.37wks.T19	....X....S.....	63.37wks.T19	.....T..Q..NK.....

Figure 4.14 Shuffling escape of amino acid residues at (a) a single alignment position in Pol and Vpr from 11 weeks post infection, (b) at multiple sites within the putative CTL epitope shown here for Env and Rev from CAP63.

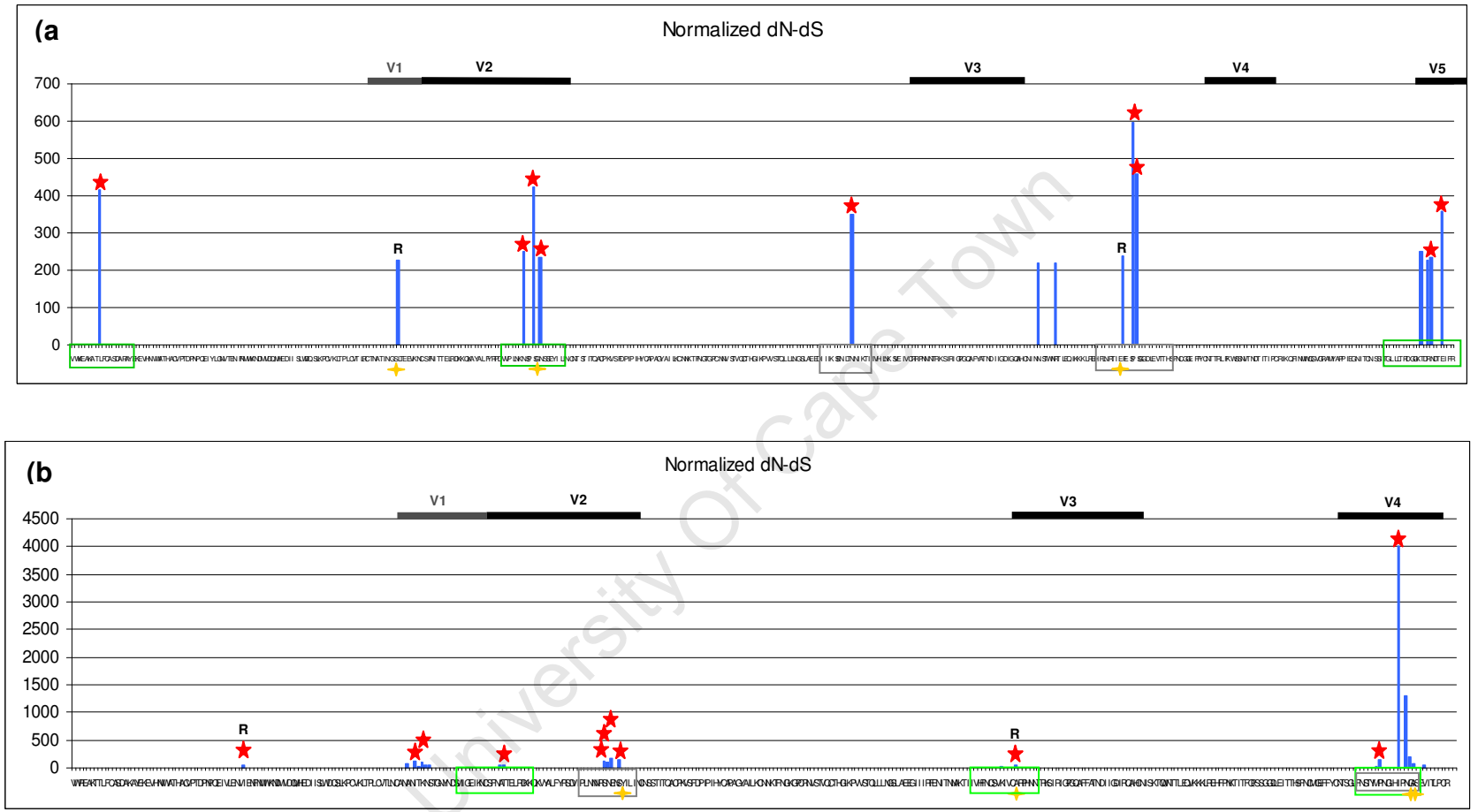


Figure 4.15 Positive selection as determined with IFEL (p <0.2) in Gp120 with significant sites indicated by red stars: (a) CAP45 and (b) CAP63. Green boxes = putative CTL epitopes, Grey boxes = putative Ab escape epitopes, Gold star = APOBEC induced mutations, R = reversion to consensus, V1 to V5 indicate positions of Gp120 variable loops.

DAYFSVPLDE**D**FRKYTAFTI, DVKQLTE**A**VHKIAIES) and in three epitopes in Env. One positively selected site in Env was located in an epitope predicted to be under both CTL and neutralizing antibody pressure and four positively selected sites were found in a single putative neutralizing antibody epitope.

Thus we observed escape in a total of eight CTL epitopes up to 65 weeks post infection in the slow progressor (CAP45) and 14 CTL epitopes by 37 weeks post infection in the rapid progressor (CAP63). CAP45, the slow progressor, experienced CTL escape at two epitopes at 16 weeks post infection, whereas the rapid progressor showed escape at seven epitopes by 11 weeks post infection (Figure 4.16). The rapid progressor (CAP63) showed increased escape early (at 11 weeks post infection) when compared to the later time points (Figure 4.16).

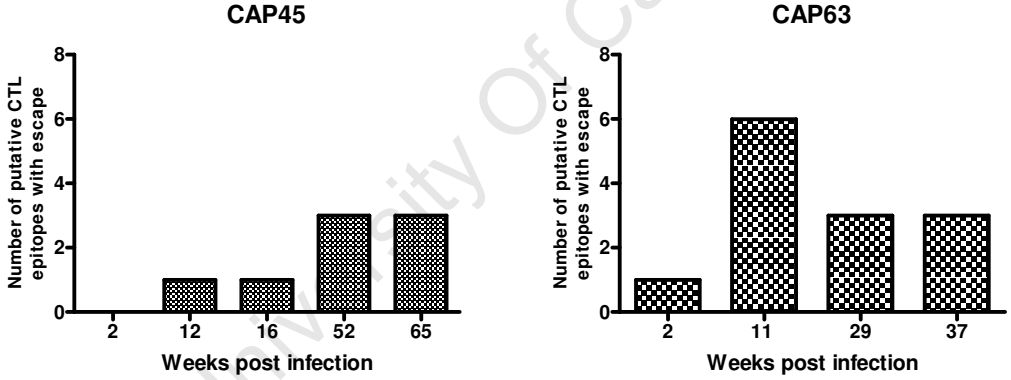


Figure 4.16 Frequency of escape during primary infection in a slow (CAP45) and rapid (CAP63) progressor over time.

4.4 DISCUSSION

Understanding the adaptive immune responses during early infection is important for the identification of determinants correlated with viral control during early infection. Here we show a high frequency of putative CTL escape mutations in longitudinally

sampled full-length viral genomes from eleven individuals recruited during primary infection. While the total number or location of putative escape was not associated with disease progression; viral diversification, reversion and APOBEC signature mutations was significantly associated with lower CD4 counts. In an in depth analysis of two individuals with discrepant disease profiles (rapid versus slow progression), we showed that the rapid progressor had an increased number of sites under positive selective pressure with a high percentage (~50%) of escape occurring through amino acid shuffling or complex escape patterns. This compared to the slow progressor whose virus showed very limited CTL escape.

Early CD8 T cell mediated immune responses were shown to contribute to early viral control after peak vireamia with rapid selection of escape mutations (Goonetilleke *et al*., 2009). Within the CAPRISA cohort, Nef is the major protein targeted by the immune response during primary infection followed by Pol, Gag, Env, Vif, Vpr, Rev, Vpu and Tat (Gray *et al*., 2009). In the same cohort, we find a similar hierarchy of putative CTL driven escape; with escape most frequently observed in Nef and Env, followed by Pol, Vif, Gag and Vpr. It is possible that in immune studies, the high diversity in Env results in an underestimation of responses to this region due to mismatch between reagents and infecting virus. In the CAPRISA cohort, immune responses were identified using a pooled peptide approach and further work is needed to confirm the specific peptides and identify optimal epitopes.

Escape from cell mediated immune responses is associated with loss of immune control and subsequent disease progression, except in individuals carrying protective HLA alleles where escape has been associated with slow disease progression (Leslie *et al*., 2004; Miura *et al*., 2008). In this study no association was found between the number of CTL epitopes and subsequent disease progression based on viral load and/or CD4 counts at 12 months post infection. It is likely that we did not see an effect due the conflicting impact of escape where escape may be beneficial, detrimental or neutral to the host. Strong selective pressure by protective HLA alleles has been identified in this cohort where infection with viruses containing HLA B*57/5801 induced Gag escape

mutations resulted in lower viral loads and higher CD4 counts in these individuals (Chopera *et al.,* 2008). It is likely that we will need larger numbers to differentiate between "good'" escape that results in fitness cost to the virus, and "bad" escape which results in loss of immune control of the virus and increased disease progression. This in depth analysis of viral evolution in early infection (within 6 months post infection) showed limited putative escape in a slow progressor (2/9 viral proteins) compared to the rapid progressor (5/9 viral proteins). We also found that shuffling or oscillation of amino acid residues was characteristic of escape in putative CTL epitopes in viruses from the rapid progressor.

In contrast to a previous study (Rousseau *et al*., 2008; Matthews *et al*., 2008) of chronic HIV-1 subtype C infection where the number of reversions in Gag and Pol had been associated with lower viral loads, we observed that an increased number of reversions to consensus C sequence during early infection are significantly associated with lower CD4 counts and show a trend of association with higher viral loads. This was driven by one point of low reversion and high CD4 count for participant CAP45. This individual shows good control of virus replication in early infection, but do not carry any of the known protective HLA alleles and no defects in the HIV-1 genes were observed. Recent studies also showed that APOBEC-3G induced mutations are associated with early viral evolutionary changes in HIV-1 infection (Wood *et al*., 2009; Goonetilleke *et al*., 2009). Goonetilleke *et al*. (2009) also reported that APOBEC induced mutations was associated with reversions in early infection, but that reversions did not play a major role in early viral evolution. Here, we firstly show a significant association between APOBEC induced mutations and reversions, as well with viral diversity during early infection. Secondly, both increased viral divergence and an increasing numbers of APOBEC induced mutations were significantly associated with lower CD4 counts at 12 months with a trend of association with higher viral loads. APOBEC inactivates lentivirus genomes by inducing hypermutation whereas the viral Vif protein facilitates the degradation of APOBEC although this mechanism isn't perfect and some effects of APOBEC activity can be seen in HIV-1 genomes. Our observations here suggest that the spurious effects of APOBEC in the absence of profound hypermutation contribute to the

overall increase in viral diversity. Some early reversion mutations are the result of APOBEC induced mutations and reversions are associated with restoration of viral fitness. Thus the direct correlation between the numbers of APOBEC induced and reversion mutations observed here in early infection suggests that more APOBEC mutations are associated with more reversions which will result in increased viral fitness and an accompanying decrease in CD4 target cells. How HIV-1 may potentially use selective APOBEC activity during early infection to its advantage as observed here need further investigation.

Collectively our data indicate that there is rapid viral evolution in CTL epitopes in early infection which implies that cellular immune responses are exerting a strong pressure. The hierarchy of escape is similar to the hierarchy of immune recognition. Screening eleven individuals over time showed no association between escape and disease progression however, in depth analysis showed that CTL escape is much more frequent in a rapid progressor compared to a slow progressor. Lastly, we find that APOBEC is playing a significant role in reversions suggesting that the virus is using this mechanism to restore fitness, and that APOBEC signatures were associated with lower CD4 counts.

CHAPTER 5

CONCLUSIONS

Since 1990 there was a steady increase in the prevalence of HIV-1 infection in women attending antenatal clinics from 0.7% to 29% making South Africa one of the hardest hit countries in the world (National HIV and Syphilis Prevalence survey, South Africa, 2007; Dorrington & Bourne, 2008). The development of an effective vaccine is thus a high priority for the region. Here we were interested in firstly, characterizing HIV diversity in full-length genomes as diversity remains one of the greatest challenges in vaccine development; and secondly in defining factors driving early viral evolution as events in early infection are thought to be predictive of long term disease outcome. This information would therefore be useful in informing correlates associated with control of viral replication.

Characterisation of full-length genomes from an acute infection cohort in Durban found no evidence of a transmission network, however for the first time we observed a number of supported clusters of viruses from chronic infections within the South African epidemic, suggesting a recent common ancestor for each of them. As expected the subtype matched vaccine, SAAVI subtype C consensus-based immunogen currently being evaluated in Phase I clinical trial, was predicted to provide better coverage against the primary infection than the subtype B Merck immunogen tested in the HVTN503/ Phambili trial. However the new generation mosaic immunogens based on group M or subtype C (valency of 4) are predicted to give the best coverage. However interestingly, the mosaic Gag immunogens with a valency of 2 gave similar coverage as the SAAVI consensus-based Gag (valency of 1) immunogen. Thus we propose that the inclusion of additional mosaic immunogens to the HIV-1 subtype C SAAVI vaccine may be sufficient to broaden epitope coverage of this candidate. However it is also possible that suboptimal vaccine candidates that do not induce sterilizing immunity may still be beneficial as they could serve as to prime early CTL responses to reduce viral loads during acute infection

and thus reduce viral setpoint. While this has been demonstrated in many non-human primate models, unfortunately this has not been shown in the HVTN502/503 vaccine trial.

HIV transmission is associated with a severe virus population bottleneck and there is some evidence that certain genotypic and phenotypic properties of the viral envelope are selectively advantageous during transmission (Derdeyn *et al.*, 2004; Rong *et al.*, 2007; Wolfs *et al.*, 1992). However, open questions remain as to whether this holds true both for genome regions other than the envelope, and for all HIV-1M subtypes. We identified thirty-nine sites within the proteomes of these viruses that differentiated them from viruses sampled during chronic infections. Through longitudinal analysis of amino acid frequency changes that occurred during the first six months of infection, together with data on host HLA alleles, we provide evidence that approximately 28.6% of site-specific differences in amino acid frequency spectra between primary and chronic infection proteomes are potential immune escape mutations. The remaining 60.7% of frequency differences between the two groups are probably due to defunct immune evasion substitutions reverting to consensus amino acid states following transmission. This data provides further understanding of processes determining the genomic and immunogenic properties of viruses during early infections which is important if we are to understand HIV pathogenesis sufficiently well to design protective vaccines against the virus.

Lastly, we investigated the role of the adaptive immune response on early viral evolution. In eleven individuals monitored over time, we found that viral diversification, reversion and APOBEC induced mutations were associated with disease progression. While CTL escape in these individuals was not associated with disease progression, an in depth analysis of multiple full-length viral genomes from a rapid compared to a slow progressor showed that viruses from the rapid progressor had an increased number of sites under positive selection pressure and an increased number of escape mutations. Fifty percent of escape in this individual occurred through amino acid shuffling or complex escape mutation patterns.

In summary this study contributes to our understanding of optimal immunogen design for South Africa. It shows that, within the first three months of infection, almost all changes in the viral genomes are associated with adaptation to the HLA of the new host, with a larger proportion of changes attributable to reversion as the virus presumably adapts to restore viral fitness. Host genetic factors such as APOBEC- induced mutations and HLA associated mutations are key factors driving viral evolution in early infection. However, larger numbers are needed to unravel the relative importance of the different viral factors affecting disease progression.

APPENDIX A

HIV-1 SUBTYPE C PCR AND COMPLETE GENOME SEQUENCING PRIMERS

A1. PCR primers for near full-length genome amplification

Primer	Direction	Location	Sequence (5'-3')
1.U5C	forward	5'	GGG TGA GTG CTC TAA GTA GTG TGT GCC C**G**T CTG T
1.U5Cb	forward	5'	GGG TGA GTG CTC TAA GTA GTG TGT GCC C**A**T CTG T
1.3.3pIC	reverse	3'	GGG ACT TAG AGC ACT CAA GGC AAG CTT TAT TG
2.U5C	forward	554-580, 5'	GGC CGC GGA TCC AGT AGT GTG TGC CCG TCT GTT GTG TGG ACT
2.3.3plC	reverse	3'	GGC CGC GCG GCC GCT AGA GCA CTC AAG GCA AGC TTT ATT GAG GCT TA

A2. *Gag* gene sequencing primers optimised for subtype C from Sanders-Buell et al. (1995).

Primer	Location	Sequence (5'-3')
G00	767→782	GAC TAG CGG AGG CTA GAA G (19mer)
GF20	895→912	GTA TGG GCA AGC AGG GAG CTG (21mer)
GF40	1078→1094	GAC ACC AAG GAA GCC TTA GA (20mer)
GF60	1176→1192	CAG TCA AAA TTA TCC TAT AG (20mer)
GF80	1477→1493	AGA GAA CCA AGG GGA AGT GA (20mer)
G100	1820→1834	TAG AAG AAA TGA TGA CAG (18mer)
G110	2081→2095	AGG CTA ATT TTT TAG GGA (18mer)
GF01	2264←2278	AGG GGT CGC TGC CAA AGA (18mer)
GF15	1960←1978	CCT TGC CAC AGT TGA AAC ATT T (22mer)
G35	1817←1835	CAT GCT GTC ATC ATT TCT TCT A (22mer)
GF45	1740←1758	TTG GAC CAA CAA GGT GTC TGT C (22mer)
GF65	1295←1307	GCT GTA AAC ATG GGT A (16mer)
GF85	1177←1193	TGC ACT ATA GGA TAA TTT TG (20mer)

A3. *Env* gene sequencing primers optimised for subtype C from Sanders-Buell et al. (1995).

Primer	Location	Sequence (5'-3')
EF0	5856→5883	TAG AGC CCT GGA ACC ATC CAG GAAGTCAGC C
EF00	6204→6228	GGG AAA GAG CAG AAG ACA GTG GCA ATG A (28 mer)
EF20	6433→6456	GGG CTA CAC ATG CCT GTG TAC CCA CAG (27mer)
E70	6562→6585	GGG ATC AAA GCC TAA AGC CAT GTG TAA (27mer)
E90	6956→6976	CAC AGT ACA ATG TAC ACA TGG AAT (24mer)
EF110	7005→7025	CTG TTA AAT GGT AGC CTA GCA GAA (24mer)
EF140	7669→7687	GTG AAT TAT ATA AAT ATA AAG T (22mer)
E170	7802→7816	AGC AGG AAG CAC TAT GGG (18mer)
E200	8095→8118	GGG ATA ACA TGA CCT GGA TGC AGT GGG (27mer)
E230	8276→8296	AAT ATT CAT AAT GAT AGT AGG AGG (24mer)
EF260	8523→8544	TTC AGC TAC CAC CGA TTG AGA GAC T (25mer)
E01	9064←9086	TCC AGT CCC CCC TTT TCT TTT AAA AA (26 mer)
EF05	8587←8605	TAC TTA AGG GCT TCC CAC CCC C (22mer)
EF15	8424←8442	CTT GCT CTC CAC CTT CTT CTT C (22mer)
EF35	8343←8362	GGT GAG TAT CCC TGC CTA ACT CT (23mer)
EF55	7914←7937	GCC CCA GAC CGT GAG TTG CAA CAT ATG (27mer)
E75	7793←7816	GCG CCC ATA GTG CTT CCT GCT GCT CCC (27mer)
EF95	7524←7538	AAT GGG AGG GGC ATA CAT (18mer)
EF115	7351←7371	AGA AAA ATT CTC CTC TAC AAT TAA (24mer)
EF135	7061←7084	AGA TGT ACT ATT ATT GTT TTG ACA TTG T (28mer)
EF155	6505←6530	CTG ATC CAC CAT GTC ATT TTT CCA CAT GT (29mer)
EF175	6378←6398	TTT AGC ATC TGA TGC ACA GAA TAG (24mer)

A4. *Pol* gene sequencing primers

Primer	Location	Sequence (5'-3')
P1	2255→2271	CCC TCA AAT CAC TCT TTG GC (20mer)
P2	2585→2595	ATT AAA GCC AGG AA (14)
P3	2885→2899	GGG GGA TGC ATA TTT TTC (18)
P4	3205→3220	ACA AGA AAC ATC AGA AAG A (19)
P5	3489→3505	CCA GTA CAT GGA GTA TAT TA (20)
P6	3765→3780	TAT TGG CAA GCC ACC TGG A (19)
P7	4025→4039	AGA AGT AAA CAT AGT AAC AG
P8	4325→4339	AGT AGC AAA AGA AAT AGT (18)
P9	4655→4669	TCC CTA CAA TCC CCA AAG (18)
P10	4954→4969	TAC TCT GGA AAG GTG AAG G (19)
PR	5199←5214	TAA TGG (T/G)AT GTG TAC TTC T (19)

A5. Select *pol* and accessory genes sequencing primers described by Rousseau et al. (2006)

Primer	Location	Sequence (5'-3')
SQ9FC(3)	5074→5099 , *pol-vif*	CAGGTAGACAGGATGAAGATCAGAAC
SQ10FC	5833→5857, *tat* exon 1 start	GGAGCCAGTAGATCCTAACCTAGAG
SQ8R	5963←5988, *tat* exon 1	CTCCGCTTCTTCCTGCCATAGGAGAT
SQ9'RC	5341←5366, *vif*	ATATGAATTAGTTGGTCTGCCAGGCC
SQ9RC(3)	5165←5188, *vif*	TTGGATGTCTGCTTTCATAATGAT
SQ16F	3' LTR	CCACACACAAGGCTACTTCC
SQ8R	5988, *tat* exon 1	CTCCGCTTCTTCCTGCCATAGGAGAT
SQ8RC	4660←4685, *pol*	TTCTACTACTCCCTGACTTTGGGGAT
SQ10RC	4555←4583, *pol*	GCCATTGTCTGTATGTATTACTTTGACTG
RT2 (reverse)	3301←3321, *pol*	GTATGTCATTGACAGTCCAGC
SQ11RC(3)	3687←3713, *pol*	TTTAGGAGTCTTTCCCCATATTACTAT
SQ11RC(4)	3960←3981, *pol*	CAGTCTTCTGATTTGTTGTTTC
SQ12R	3304←3322, *pol*	TGTATGTCATTGACAGTCC
HIP202 (reverse)	2328←2352, *pol*	CTAATACTGTATCATCTGCTCCTGT
13R(2)C(2)	2712←2735, *pol*	GGCAAATATTGGAGTGTTATATGG

A6. Additional full-length *env* gene sequencing primers

Primer	Location	Sequence (5'-3')
ENV A	5957 - 5982	GGC TTA GGC ATC TCC TAT GGC AGG AAG AA
ENV N	9544 - 9521	CTG CCA ATC AGG GAA GTA GCC TTG TGT
VIF1	4903 - 4923	GGG TTT ATT ACA GGG ACA GCA GAG
Env-CF	6208 - 6224	AAG AGC AGA AGA CAG TGG CA
Env-CR	6875 - 6859	ACA ATA ATG TAT AGG AAT TG
Env-AF	6206 - 6223	GAA AGA GCA GAA GAC AGT GGC
Env-AR	7799 - 7783	TGC TGC TCC CAA GAA CCC AA
Env-DF	7211 - 7226	AGC ACA TTG TAA CAT TAG T
Env-DR	7684 - 7665	ACT TTA TAT TTA TAT AAT TCA CT
Env-MF	7641 - 7661	GGA GGA GAT ATG AGG GAC AAT TGG
Env-BR	7334 - 7317	AAT TTC TAG GTC CCC TCC TGA
Env-NF	8104 - 8117	TGA CCT GGA TGC AGT GG
Env-NR	8362 - 8342	GGT GAG TAT CCC TGC CTA ACT CTA

APPENDIX B

REAGENTS AND BUFFERS

B1. 50X TAE buffer

121g Tris base

28.55 ml glacial acetic acid

50 ml 0.5M EDTA pH 8.0

Make up to 500 ml with deionised sterile H_2O (Sambrook *et al*., 1989).

B2. 0.5M EDTA (pH 8.0)

Add 186.1g disodium ethylenediaminetetraacetate.2H2O to ~800 ml deionised sterile water. Dissolve by stirring vigorously on magnetic stirrer while adjusting the H to 8.0 with ~20g NaOH pellets. **The disodium salt of EDTA will only dissolve if the pH of the solution is adjusted to ~8.0** (Sambrook *et al*., 1989)

B3. 6X loading buffer II (for agarose gel electrophoresis)

0.25% bromophenol blue

0.25% xylene cyanol FF

15% ficoll (type 400)

Make up in deionised sterile H_2O and store at room temperature (Sambrook *et al*., 1989).

B4. 10X TBE buffer

108g Tris-HCl

55g Boric acid

20 ml 1.5M EDTA

Make up to 1 liter with deionised sterile H_2O

B5. 6X Polyacrylamide gel loading dye

0.25% Bromophenol blue

0.25% Xylene cyanol FF

15% Ficoll in deionised sterile H_2O

B6. 30% Acrylamide stock solution (29:1)

29g Acrylamide

1g Bis-acrylamide

Dissolve in 60 ml deionised sterile H2O at 37°C. Make up to 100 ml with water and filter sterilize through a Nalgene 0.45 µm filter

Keep in dark bottle at 4°C for a month.

Caution! Acrylamide is a potential neurotoxin that is absorbed through the skin. (Sambrook *et al*., 1989)

B7. 5% Non-denaturing Polyacrylamide gel

8.3 ml 30% Acrylamide stock (29:1)

5 ml 10X TBE

10g urea

Add water to about 36.7 ml. Allow reagents to dissolve by stirring and when all powder is dissolved make up with water to a final volume of 50 ml.

When ready to pour gel add 200 µl 25% APS and 33 µl TEMED to initiate gel polymerization. Pour gel immediately after addition of APS and TEMED.

B8. 5% Denaturing Polyacrylamide gel

8.3 ml 30% Acrylamide stock (29:1)

5 ml 10X TBE

10g urea

B9. 25% Ammonium persulfate (APS)

2.5g APS

Make up to 10 ml with deionised sterile water. Store at 4°C.

B10. HTA annealing buffer

100 mM Tris-HCl, pH 7.5

1M NaCl

20 mM EDTA

B11. LB broth

10g NaCl

5g Yeast extract

10g Tryptone

Dissolve and make up to 1 liter with deionised sterile water. Autoclave immediately.

B12. LB agar selection plates containing Kanamycin

Prepare LB medium as above, but add 15 g/L agar before autoclaving.

Autoclave on a liquid cycle for 20 minutes at 15 psi.

After autoclaving, cool to ~55°C.

Add 50 µg/ml of kanamycin, and pour into 10 cm plates.

Let harden, then invert and store at +4°C, in the dark.

APPENDIX C

METHODS AND MANUFACTURERS' PROTOCOLS

C1: Purification of Viral RNA (Spin Protocol, QIAGEN)

This protocol is for purification of viral RNA from 140 µl plasma, serum, urine, cellculture media, or cell-free body fluids using a microcentrifuge. For automated purification of viral RNA using the QIAamp Viral RNA Mini Kit on the QIAcube, refer to the QIAcube User Manual and the relevant protocol sheet.
Larger starting volumes, up to 560 µl (in multiples of 140 µl), can be processed by increasing the initial volumes proportionally and loading the QIAamp Mini column multiple times, as described below in the protocol. Some samples with very low viral titers should be concentrated before the purification procedure; see "Protocol: Sample Concentration" (page 30).
Alternatively, larger sample volumes can be processed using one of the following kits, which provide simultaneous purification of viral DNA and RNA.

- QIAamp MinElute® Spin Kit* 200 µl
- QIAamp MinElute Vacuum Kit 500 µl
- QIAamp UltraSens® Virus Kit 1000 µl

Important points before starting

- Read "Important Notes" (pages 15–22) before starting the protocol.
- All centrifugation steps are carried out at room temperature (15–25°C).

Things to do before starting

- Equilibrate samples to room temperature (15–25°C).
- Equilibrate Buffer AVE to room temperature for elution in step 11.
- Check that Buffer AW1 and Buffer AW2 have been prepared according to the instructions on page 17.
- Add carrier RNA reconstituted in Buffer AVE to Buffer AVL according to instructions on page 15.

Procedure

1. Pipet 560 µl of prepared Buffer AVL containing carrier RNA into a 1.5 ml microcentrifuge tube.
If the sample volume is larger than 140 µl, increase the amount of Buffer AVL–carrier RNA proportionally (e.g., a 280 µl sample will require 1120 µl Buffer AVL–carrier RNA) and use a larger tube.

2. Add 140 µl plasma, serum, urine, cell-culture supernatant, or cell-free body fluid to the Buffer AVL–carrier RNA in the microcentrifuge tube. Mix by pulse-vortexing for 15 s.
To ensure efficient lysis, it is essential that the sample is mixed thoroughly with Buffer AVL to yield a homogeneous solution. Frozen samples that have only been thawed once can also be used.

3. Incubate at room temperature (15–25°C) for 10 min.
Viral particle lysis is complete after lysis for 10 min at room temperature. Longer incubation times have no effect on the yield or quality of the purified RNA. Potentially infectious agents and RNases are inactivated in Buffer AVL.

4. Briefly centrifuge the tube to remove drops from the inside of the lid.

5. Add 560 µl of ethanol (96–100%) to the sample, and mix by pulse-vortexing for 15 s. After mixing, briefly centrifuge the tube to remove drops from inside the lid.
Only ethanol should be used since other alcohols may result in reduced RNA yield and purity. Do not use denatured alcohol, which contains other substances such as methanol or methylethylketone. If the sample volume is greater than 140 µl, increase the amount of ethanol proportionally (e.g., a 280 µl sample will require 1120 µl of ethanol). In order to ensure efficient binding, it is essential that the sample is mixed thoroughly with the ethanol to yield a homogeneous solution.

6. Carefully apply 630 µl of the solution from step 5 to the QIAamp Mini column (in a 2 ml collection tube) without wetting the rim. Close the cap, and centrifuge

at 6000 x g (8000 rpm) for 1 min. Place the QIAamp Mini column into a clean 2 ml collection tube, and discard the tube containing the filtrate.
Close each spin column in order to avoid cross-contamination during centrifugation.
Centrifugation is performed at 6000 x g (8000 rpm) in order to limit microcentrifuge noise. Centrifugation at full speed will not affect the yield or purity of the viral RNA. If the solution has not completely passed through the membrane, centrifuge again at a higher speed until all of the solution has passed through.
7. Carefully open the QIAamp Mini column, and repeat step 6.
If the sample volume was greater than 140 µl, repeat this step until all of the lysate has been loaded onto the spin column.
8. Carefully open the QIAamp Mini column, and add 500 µl of Buffer AW1. Close the cap, and centrifuge at 6000 x g (8000 rpm) for 1 min. Place the QIAamp Mini column in a clean 2 ml collection tube (provided), and discard the tube containing the filtrate.
It is not necessary to increase the volume of Buffer AW1 even if the original sample volume was larger than 140 µl.
9. Carefully open the QIAamp Mini column, and add 500 µl of Buffer AW2. Close the cap and centrifuge at full speed (20,000 x g; 14,000 rpm) for 3 min. Continue directly with step 11, or to eliminate any chance of possible Buffer AW2 carryover, perform step 10, and then continue with step 11.
Note: Residual Buffer AW2 in the eluate may cause problems in downstream applications. Some centrifuge rotors may vibrate upon deceleration, resulting in flow-through, containing Buffer AW2, contacting the QIAamp Mini column. Removing the QIAamp Mini column and collection tube from the rotor may also cause flow-through to come into contact with the QIAamp Mini column. In these cases, the optional step 10 should be performed.
10. Recommended: Place the QIAamp Mini column in a new 2 ml collection tube (not provided), and discard the old collection tube with the filtrate. Centrifuge at full speed for 1 min.
11. Place the QIAamp Mini column in a clean 1.5 ml microcentrifuge tube (not provided). Discard the old collection tube containing the filtrate. Carefully open the QIAamp Mini column and add 60 µl of Buffer AVE equilibrated to room temperature. Close the cap, and incubate at room temperature for 1 min. Centrifuge at 6000 x g (8000 rpm) for 1 min.
A single elution with 60 µl of Buffer AVE is sufficient to elute at least 90% of the viral RNA from the QIAamp Mini column. Performing a double elution using 2 x 40 µl of Buffer AVE will increase yield by up to 10%. Elution with volumes of less than 30 µl will lead to reduced yields and will not increase the final concentration of RNA in the eluate.
Viral RNA is stable for up to one year when stored at –20°C or –70°C.

Preparation of reagents

Addition of carrier RNA to Buffer AVL*
Add 310 µl Buffer AVE to the tube containing 310 µg lyophilized carrier RNA to obtain a solution of 1 µg/µl. Dissolve the carrier RNA thoroughly, divide it into conveniently sized aliquots, and store it at –20°C. Do not freeze–thaw the aliquots of carrier RNA more than 3 times.
Check Buffer AVL for precipitate, and if necessary incubate at 80°C until the precipitate is dissolved. Calculate the volume of Buffer AVL–carrier RNA mix needed per batch of samples by selecting the number of samples to be **simultaneously** processed from Table 1. For larger numbers of samples, volumes can be calculated using the following sample calculation:
n x 0.56 ml = **y** ml
y ml x 10 µl/ml = **z** µl
* Contains chaotropic salt. Take appropriate laboratory safety measures, and wear gloves when handling. Not compatible with disinfecting agents that contain bleach. See page 6 for safety information.
where:
n = number of samples to be processed simultaneously
y = calculated volume of Buffer AVL
z = volume of carrier RNA–Buffer AVE to add to Buffer AVL

Gently mix by inverting the tube 10 times. To avoid foaming, do not vortex.

Note: The sample-preparation procedure is optimized for 5.6 µg of carrier RNA per sample. If less carrier RNA has been shown to be better for your amplification system, transfer only the required amount of dissolved carrier RNA to the tubes containing Buffer AVL. (Use of less than 5.6 µg carrier RNA per sample must be validated for each particular sample type and downstream assay.)
Buffer AVL–carrier RNA should be prepared fresh, and is stable at 2–8°C for up to 48 hours. This solution develops a precipitate when stored at 2–8°C that must be redissolved by warming at 80°C before use. Do not warm Buffer AVL–carrier RNA solution more than 6 times. Do not incubate at 80°C for more than 5 minutes. Frequent warming and extended incubation will cause degradation of carrier RNA, leading to reduced recovery of viral RNA and eventually false negative RT-PCR results. This is particularly the case with low-titer samples.

Buffer AW1*
Buffer AW1 is supplied as a concentrate. Before using for the first time, add the appropriate amount of ethanol (96–100%) as indicated on the bottle and in Table 2.
Buffer AW1 is stable for 1 year when stored closed at room temperature, but only until the kit expiration date.
* Contains chaotropic salt. Take appropriate laboratory safety measures, and wear gloves when handling. Not compatible with disinfecting agents that contain bleach. See page 6 for safety information.
† Contains sodium azide as a preservative.

Buffer AW2†
Buffer AW2 is supplied as a concentrate. Before using for the first time, add the appropriate amount of ethanol (96–100%) to Buffer AW2 concentrate as indicated on the bottle and in Table 3.
Buffer AW2 is stable for 1 year when stored closed at room temperature, but only until the kit expiration date.
(extracted from QIAamp Viral RNA Mini Handbook 12/2007)

C2: pCR-XL TOPO cloning of large (3 - 10 kb) DNA inserts (Invitrogen)

Gel-Purifying PCR Products

Before Starting
Bring the Sodium Iodide solution, Binding Buffer, 4X Final Wash, and TE Buffer to room temperature. **Mix well to re-dissolve salts before using. Store these solutions at room temperature after the first use.**
To prepare 1X Final Wash, take a sterile 15 ml bottle and transfer all of the 4X Final Wash solution to the bottle. Add 7.5 ml of 100% **ethanol** to the 4X Final Wash solution to prepare the 1X Final Wash solution. Store at room temperature.

Nuclease Control
It is very important to minimize the presence of nucleases to ensure the highest cloning efficiencies. Follow the guidelines listed below. While some guidelines may not appear as rigorous as others, they are sufficient for purifying long PCR fragments.
• Wear gloves at all times
• Use sterile plasticware and glassware
• Autoclave TAE to use as the running buffer
• Rinse agarose gel apparatus and comb with autoclaved water or TE buffer
• Use a new razor to excise gel slice*
• Use new plastic wrap (*e.g.* Saran® Wrap) if needed
*The same razor may be used to excise different bands in the same gel if you are careful not to bring over pieces from an earlier excision.

Preparing the Gel
Follow the instructions below to prepare a 0.8% agarose gel. The recipe will make one agarose gel with a volume of 50 ml.
1. Mix 0.4 g of general purpose agarose and 50 ml 1X TAE buffer in a clean glass flask.
2. Place flask in the microwave and heat until just boiling. Swirl to dissolve

agarose and continue to heat in this fashion for 3 minutes to destroy nucleases.
3. Remove from the microwave and allow to cool for 3 minutes.
4. Add 20-40 µl of the 2 mg/ml Crystal Violet solution to the agarose and swirl to mix. The agarose should be light to medium purple in color.
5. Rinse the gel box and comb with autoclaved water or TE buffer.
Note: Use a comb that will hold all of the PCR amplification (48 µl) in one well, if possible.
6. Pour the gel and set the comb in the gel.
7. When the gel has solidified, cover the gel with 1X TAE buffer. You do not have to add crystal violet to the running buffer. Proceed to the next section to prepare your sample.

Loading and Running the Gel
1. Add 8 µl of 6X Crystal Violet Loading Buffer to 40 µl of the PCR amplification and load onto the gel. Prepare the control PCR product in the same manner.
2. Run the gel at 80 volts until the crystal violet in the gel has run about a quarter of the way UP the gel (crystal violet appears to migrate towards the negative pole). You should also see the thin blue PCR product move down into the gel. If no PCR product is visible, insufficient DNA was loaded (<200 ng).
3. Compare your PCR product to the 7 kb control PCR fragment. If your PCR product is sufficiently resolved so that you can easily excise the fragment, turn off the power supply. Proceed directly to **Excising the PCR Product,** below.

Excising the PCR Product
Remember to excise the control fragment as well as your band.
1. Pour off the buffer (or transfer the gel to new Saran® Wrap).
Note: Placing the gel on a fluorescent light box may help visualize the fragments.
2. Using a new razor blade, carefully excise the desired band from the gel.
Note: Razor blade may be rinsed with autoclaved water or TE prior to cutting the next band.
3. Cut up the excised piece of agarose into small chunks and transfer to a sterile 1.5 ml microcentrifuge tube.
Note: Cutting up the agarose piece reduces the time and temperature required to melt the agarose.
4. Estimate the volume of the agarose pieces (generally this is around 100 µl). Alternatively, you can weigh the gel slice and assume that 1 mg ~ 1 µl.
5. Add 2.5 times its volume of 6.6 M sodium iodide (*i.e.* 250 µl) and mix by shaking vigorously by hand or vortexing.
6. Incubate at 42 to 50°C until the agarose is **completely** melted (~2 minutes). Mix the solution periodically by shaking vigorously.
7. Place the tube at room temperature and add 1.5 volumes of Binding Buffer (*i.e.* 525 µl) and mix well. Proceed directly to **Isolating the PCR Product,** next page.

Isolating the PCR Product
1. Assemble a S.N.A.P.™ purification column and collection vial and load all of the mixture from Step 7, previous page, onto the column (875 µl).
2. Centrifuge at 2,000 to 3,000 x g in a microcentrifuge for 30 seconds at room temperature.
3. Pour the liquid in the collection vial **back onto the column** and repeat Step 2.
4. Repeat Step 3 one more time to bind all the DNA to the column (*i.e.* load solution onto the column for a total of 3 times).
5. After the last centrifugation, discard the liquid in the collection tube.
6. Add 400 µl of 1X Final Wash to the S.N.A.P.™ column and centrifuge as in Step 2.
7. Repeat Step 6 and discard the liquid in the collection tube after the final centrifugation (800 µl).
8. Centrifuge the column again **at maximum speed (>10,000 x g) for at least 1 minute** to dry the column resin. Discard the collection vial.

9. Transfer the column to a new, sterile 1.5 ml microcentrifuge tube.
10. Add 40 µl of TE buffer directly to the column material and incubate for 1 minute at room temperature to let the buffer absorb into the column.
11. Centrifuge the column at maximum speed (>10,000 x g) for 1 minute to elute the DNA into the microcentrifuge tube.
12. Place the tube on ice and discard the column.
13. Assay 10 µl by ethidium bromide agarose gel electrophoresis to estimate the DNA concentration. Concentration should be between 2 and 40 ng/µl. In most cases, there is no need to concentrate the DNA further.
Store the DNA at -20°C for up to one week or proceed directly to **Setting Up the TOPO® Cloning Reaction**, next page. For the highest efficiencies, proceed directly to TOPO® Cloning.

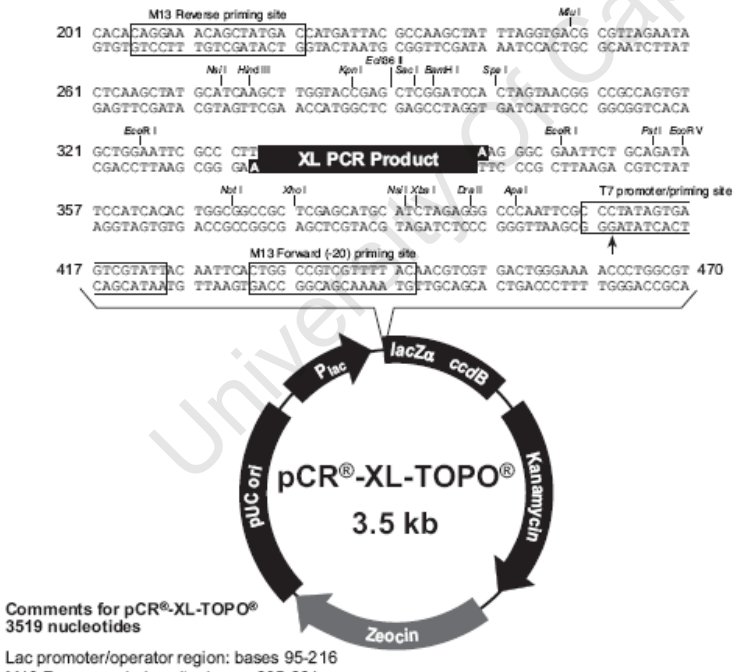
TOPO® Cloning Reaction
1. Set up the following 5 µl TOPO® Cloning reaction in a sterile microcentrifuge tube. Remember to set up a reaction for the control PCR product. You may also wish to set up a "vector-only" reaction as a negative control (use TE buffer instead of insert).
Gel-purified long PCR product from Step 12, page 8 4 µl
pCR®-XL-TOPO® vector 1 µl
Final Volume 5 µl
2. Mix gently and incubate for **5 minutes** at room temperature (~25°C). **For the best possible results, do not leave for more than 5 minutes at room temperature or the transformation efficiencies will decrease!**
3. After the 5 minute incubation, add 1 µl of the 6X TOPO® Cloning Stop Solution and mix for several seconds at room temperature.
Note: Addition of the Stop Solution increases the yield of transformants by an average of 2-fold for chemically competent cells and 10-fold for electrocompetent cells.
4. Briefly centrifuge the tube and place on ice.
Note: The TOPO® Cloning reaction may be stored on ice or frozen at -20°C for up to 24 hours. You may see a decrease in the transformation efficiency, but the cloning efficiency should remain high. We recommend that you proceed immediately to **Transforming Competent Cells**

Transforming Competent Cells
Introduction Once you have performed the TOPO® Cloning reaction, you will transform your pCR®-XL-TOPO® construct into competent *E. coli* provided with your kit.
Materials Supplied by the User
• LB plates containing 50 µg/ml kanamycin or Low Salt LB plates containing 25 µg/ml Zeocin™ (see page 18 for recipes)
Note: For the control reaction you will need an additional LB plate containing 50 µg/ml kanamycin **and** 50 µg/ml ampicillin
• 37°C shaking and non-shaking incubator
• 42°C water bath (for transformation of chemically competent cells)
Preparing for Transformation
For each transformation, you will need one vial of competent cells and two selective plates.
• Thaw the vial of S.O.C. medium from Box 2 and bring to room temperature.
• Warm selective plates at 37°C for 30 minutes.
• Thaw **on ice** 1 vial of One Shot® competent cells for each transformation.
One Shot® Chemical Transformation
Use this protocol only with One Shot® TOP10 chemically competent cells.
1. Add 2 µl of the TOPO® Cloning reaction into a vial of One Shot® cells and mix gently. **Do not mix by pipetting up and down.**
2. Incubate on ice for 30 minutes.
3. Heat-shock the cells for 30 seconds at 42°C without shaking.
4. Immediately transfer the tubes to ice and incubate for 2 minutes.
5. Add 250 µl of room temperature S.O.C. medium.
6. Cap the tube tightly and shake the tube horizontally at 37°C for 1 hour. Place on ice.
7. Spread 50-150 µl from each transformation on a prewarmed plate. For the control transformation, spread 50 µl **each** on a prewarmed LB plate

containing 50 µg/ml kanamycin **and** on an LB plate containing 50 µg/ml kanamycin and 50 µg/ml ampicillin.
8. Incubate plates overnight at 37°C.
9. An efficient TOPO® Cloning reaction will produce several hundred colonies.
Note: Larger plasmids will produce fewer colonies.

Analyzing Positive Clones
1. Pick ~10 colonies and culture them overnight in LB medium containing 50 µg/ml kanamycin. Alternatively, you may use Low Salt LB medium containing 25 µg/ml Zeocin™.
Note: If you transformed One Shot® Mach1™-T1R competent *E. coli*, you may inoculate overnight-grown colonies and culture them for 4 hours in **prewarmed** LB medium containing 50 µg/ml kanamycin or Low Salt LB medium containing 25 µg/ml Zeocin™ before isolating plasmid. For optimal results, we recommend inoculating as much of a single colony as possible.
2. Isolate plasmid DNA using your method of choice.
3. Analyze the plasmids by restriction analysis to confirm the presence and correct orientation of the insert. Use a restriction enzyme or a combination of enzymes that cut once in the vector and once in the insert.
Sequencing You may sequence your construct to confirm that your gene is cloned in the correct orientation.
(extracted from Invitrogen Life Technologies Instruction manual Version M, 14 June 2004, 25-0199)

TOPO XL PCR Cloning vector map



Comments for pCR®-XL-TOPO®
3519 nucleotides

Lac promoter/operator region: bases 95-216
M13 Reverse priming site: bases 205-221
Lac Za ORF: bases 217-576
Multiple Cloning Site: bases 248-399
TOPO® Cloning site: bases 336-337
T7 promoter priming site: bases 406-425
M13 Forward (-20) priming site: bases 433-448
Fusion joint: bases 577-585
*ccd*B lethal gene ORF: bases 586-888
Kanamycin resistance ORF: bases 1237-2031
Zeocin resistance ORF: bases 2238-2612
pUC origin: bases 2680-3393

C3: Plasmid DNA Purification Using the QIAprep Spin Miniprep Kit (QIAGEN)

This protocol is designed for purification of up to 20 µg of high-copy plasmid DNA from 1–5 ml overnight cultures of E. coli in LB (Luria-Bertani) medium. For purification of low-copy plasmids and cosmids, large plasmids (>10 kb), and DNA prepared using other methods, refer to the recommendations on page 44.

Please read "Important Notes" on pages 15–21 before starting.

Note: All protocol steps should be carried out at room temperature.

Procedure

1. Resuspend pelleted bacterial cells in 250 µl Buffer P1 and transfer to a microcentrifuge tube.

Ensure that RNase A has been added to Buffer P1. No cell clumps should be visible after resuspension of the pellet.

If LyseBlue reagent has been added to Buffer P1, vigorously shake the buffer bottle to ensure LyseBlue particles are completely dissolved. The bacteria should be resuspended completely by vortexing or pipetting up and down until no cell clumps remain.

2. Add 250 µl Buffer P2 and mix thoroughly by inverting the tube 4–6 times.

Mix gently by inverting the tube. Do not vortex, as this will result in shearing of genomic DNA. If necessary, continue inverting the tube until the solution becomes viscous and slightly clear. Do not allow the lysis reaction to proceed for more than 5 min.

If LyseBlue has been added to Buffer P1 the cell suspension will turn blue after addition of Buffer P2. Mixing should result in a homogeneously colored suspension. If the suspension contains localized colorless regions or if brownish cell clumps are still visible, continue mixing the solution until a homogeneously colored suspension is achieved.

3. Add 350 µl Buffer N3 and mix immediately and thoroughly by inverting the tube 4–6 times.

To avoid localized precipitation, mix the solution thoroughly, immediately after addition of Buffer N3. Large culture volumes (e.g. ≥5 ml) may require inverting up to 10 times. The solution should become cloudy.

If LyseBlue reagent has been used, the suspension should be mixed until all trace of blue has gone and the suspension is colorless. A homogeneous colorless suspension indicates that the SDS has been effectively precipitated.

4. Centrifuge for 10 min at 13,000 rpm (~17,900 x g) in a table-top microcentrifuge.

A compact white pellet will form.

5. Apply the supernatants from step 4 to the QIAprep spin column by decanting or pipetting.

6. Centrifuge for 30–60 s. Discard the flow-through.

7. Recommended: Wash the QIAprep spin column by adding 0.5 ml Buffer PB and centrifuging for 30–60 s. Discard the flow-through.

This step is necessary to remove trace nuclease activity when using endA+ strains such as the JM series, HB101 and its derivatives, or any wild-type strain, which have high levels of nuclease activity or high carbohydrate content. Host strains such as XL-1 Blue and DH5□™ do not require this additional wash step.

8. Wash QIAprep spin column by adding 0.75 ml Buffer PE and centrifuging for 30–60 s.

9. Discard the flow-through, and centrifuge for an additional 1 min to remove residual wash buffer.

Important: Residual wash buffer will not be completely removed unless the flow-through is discarded before this additional centrifugation. Residual ethanol from Buffer PE may inhibit subsequent enzymatic reactions.

10. Place the QIAprep column in a clean 1.5 ml microcentrifuge tube. To elute DNA, add 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) or water to the center of each QIAprep spin column, let stand for 1 min, and centrifuge for 1 min.

(+extracted from QIAprep Miniprep Handbook 12/2006)

C4: Plasmid DNA Purification Using QIAfilter Plasmid Midi and Maxi Kits (QIAGEN)

This protocol is designed for preparation of up to 100 µg of high- or low-copy plasmid or cosmid DNA using the QIAfilter Plasmid Midi Kit, or up to 500 µg using the QIAfilter Plasmid Maxi Kit. In this protocol, QIAfilter Cartridges are used instead of conventional centrifugation to clear bacterial lysates. For purification of double-stranded M13 replicative-form DNA, see the recommendations at www.qiagen.com/goto/plasmidinfo .
Low-copy plasmids which have been amplified in the presence of chloramphenicol should be treated as high-copy plasmids when choosing the appropriate culture volume.

Maximum Recommended Culture Volumes*

	QIAfilter Midi	**QIAfilter Maxi**
High-copy plasmids	25 ml	100 ml
Low-copy plasmids	50–100 ml	250 ml†

* For high-copy plasmids, expected yields are 75–100 µg for the QIAfilter Plasmid Midi Kit and 300–500 µg for the QIAfilter Plasmid Maxi Kit. For low-copy plasmids, expected yields are 20–100 µg for the QIAfilter Plasmid Midi Kit and 50–250 µg for the QIAfilter Plasmid Maxi Kit using these culture volumes.
† The maximum recommended culture volume applies to the capacity of the QIAfilter Maxi Cartridge. If higher yields of low-copy plasmids are desired, the lysates from two QIAfilter Maxi Cartridges can be loaded onto one QIAGEN-tip 500.

Important points before starting

■ New users are advised to familiarize themselves with the detailed protocol provided in this handbook. In addition, extensive background information is provided on our plasmid resource page www.qiagen.com/goto/plasmidinfo .

■ If working with low-copy vectors, it may be beneficial to increase the lysis buffer volumes in order to increase the efficiency of alkaline lysis, and thereby the DNA yield. In case additional Buffers P1, P2, and P3 are needed, their compositions are provided in Appendix B: Composition of Buffers, on page 37. Alternatively, the buffers and additional QIAfilter Cartridges may be purchased separately (see page 42).

■ Optional: Remove samples at the steps indicated with the symbol ☞ in order to monitor the procedure on an analytical gel (see page 34).

■ In contrast to the standard protocol, the lysate is not incubated on ice after addition of Buffer P3.

■ Blue (marked with a ▲) denotes values for the QIAfilter Plasmid Midi Kit; red (marked with a ●) denotes values for the QIAfilter Plasmid Maxi Kit.

Things to do before starting

■ Add the provided RNase A solution to Buffer P1 before use. Use one vial of RNase A (centrifuge briefly before use) per bottle of Buffer P1, to give a final concentration of 100 µg/ml.

■ Check Buffer P2 for SDS precipitation due to low storage temperatures. If necessary, dissolve the SDS by warming to 37°C.

■ Pre-chill Buffer P3 at 4°C.

■ Optional: Add the provided LyseBlue reagent to Buffer P1 and mix before use. Use one vial LyseBlue (centrifuge briefly before use) per bottle of Buffer P1 to achieve a 1:1000 dilution. LyseBlue provides visual identification of optimum buffer mixing thereby preventing the common handling errors that lead to inefficient cell lysis and incomplete precipitation of SDS, genomic DNA, and cell debris. For more details see “Using LyseBlue reagent” on page 15.

Procedure

1. Pick a single colony from a freshly streaked selective plate and inoculate a starter culture of 2–5 ml LB medium containing the appropriate selective antibiotic. Incubate for approx. 8 h at 37°C with vigorous shaking (approx. 300 rpm).
Use a tube or flask with a volume of at least 4 times the volume of the culture.
2. Dilute the starter culture 1/500 to 1/1000 into selective LB medium. For high-copy plasmids, inoculate ▲25 ml or ● 100 ml medium with ▲25–50 µl or ● 100–200 µl of starter culture. For low-copy plasmids, inoculate ▲ 50–100 ml or ● 250 ml medium with ▲ 100–200 µl or ● 250–500 µl of starter culture. Grow at 37°C for

12–16 h with vigorous shaking (approx. 300 rpm).
Use a flask or vessel with a volume of at least 4 times the volume of the culture. The culture should reach a cell density of approximately 3–4 x 10_9 cells per milliliter, which typically corresponds to a pellet wet weight of approximately 3 g/liter medium (see page 12).
3. Harvest the bacterial cells by centrifugation at 6000 x *g* for 15 min at 4°C.
f If you wish to stop the protocol and continue later, freeze the cell pellets at –20°C.
4. Resuspend the bacterial pellet in ▲ 4 ml or ● 10 ml Buffer P1.
For efficient lysis it is important to use a vessel that is large enough to allow complete mixing of the lysis buffers. Ensure that RNase A has been added to Buffer P1.
If LyseBlue reagent has been added to Buffer P1, vigorously shake the buffer bottle before use to ensure LyseBlue particles are completely resuspended. The bacteria should be resuspended completely by vortexing or pipetting up and down until no cell clumps remain.
5. Add ▲4 ml or ● 10 ml Buffer P2, mix thoroughly by vigorously inverting the sealed tube 4–6 times, and incubate at room temperature (15–25°C) for 5 min.
Do not vortex, as this will result in shearing of genomic DNA. The lysate should appear viscous. Do not allow the lysis reaction to proceed for more than 5 min. After use, the bottle containing Buffer P2 should be closed immediately to avoid acidification from CO_2 in the air.
If LyseBlue has been added to Buffer P1 the cell suspension will turn blue after addition of Buffer P2. Mixing should result in a homogeneously colored suspension. If the suspension contains localized colorless regions or if brownish cell clumps are still visible, continue mixing the solution until a homogeneously colored suspension is achieved.

During the incubation prepare the QIAfilter Cartridge:
Screw the cap onto the outlet nozzle of the QIAfilter Midi or QIAfilter Maxi Cartridge. Place the QIAfilter Cartridge in a convenient tube.
6. Add ▲ 4 ml or ● 10 ml chilled Buffer P3 to the lysate, and mix immediately and thoroughly by vigorously inverting 4–6 times.

Proceed directly to step 7. Do not incubate the lysate on ice.

Precipitation is enhanced by using chilled Buffer P3. After addition of Buffer P3, a fluffy white material forms and the lysate becomes less viscous. The precipitated material contains genomic DNA, proteins, cell debris, and KDS. The lysate should be mixed thoroughly to ensure even potassium dodecyl sulfate precipitation. If the mixture still appears viscous, more mixing is required to completely neutralize the solution. It is important to transfer the lysate into the QIAfilter Cartridge immediately in order to prevent later disruption of the precipitate layer.
If LyseBlue reagent has been used, the suspension should be mixed until all trace of blue has gone and the suspension is colorless. A homogeneous colorless suspension indicates that the SDS has been effectively precipitated.
7. Pour the lysate into the barrel of the QIAfilter Cartridge. Incubate at room temperature (15–25°C) for 10 min. Do not insert the plunger!
Important: This 10 min incubation at room temperature is essential for optimal performance of the QIAfilter Midi or QIAfilter Maxi Cartridge. Do not agitate the QIAfilter Cartridge during this time. A precipitate containing proteins, genomic DNA, and detergent will float and form a layer on top of the solution. This ensures convenient filtration without clogging. If, after the 10 min incubation, the precipitate has not floated to the top of the solution, carefully run a sterile pipet tip around the walls of the cartridge to dislodge it.
8. Equilibrate a ▲ QIAGEN-tip 100 or ● QIAGEN-tip 500 by applying ▲ 4 ml or ● 10 ml Buffer QBT, and allow the column to empty by gravity flow.
Flow of buffer will begin automatically by reduction in surface tension due to the presence of detergent in the equilibration buffer. Allow the QIAGEN-tip to drain completely. QIAGEN-tips can be left unattended, since the flow of buffer will stop when the meniscus reaches the upper frit in the column.
9. Remove the cap from the QIAfilter Cartridge outlet nozzle. Gently insert the plunger into the ▲ QIAfilter Midi or ● QIAfilter Maxi Cartridge and filter the cell lysate into the previously equilibrated QIAGEN-tip.

Filter until all of the lysate has passed through the QIAfilter Cartridge, but do not apply extreme force. Approximately ▲**10 ml** and ● **25 ml** of the lysate are generally recovered after filtration.

☞ Remove a ▲ 240 µl or ● 120 µl sample of the filtered lysate and save for an analytical gel (sample 1) in order to determine whether growth and lysis conditions were optimal.

10. Allow the cleared lysate to enter the resin by gravity flow.

☞ Remove a ▲ 240 µl or ● 120 µl sample from the flow-through and save for an analytical gel (sample 2) in order to determine the efficiency of DNA binding to the QIAGEN Resin.

11. Wash the QIAGEN-tip with ▲ 2 x 10 ml or ● 2 x 30 ml Buffer QC.

Allow Buffer QC to move through the QIAGEN-tip by gravity flow. The first wash is sufficient to remove all contaminants in the majority of plasmid DNA preparations. The second wash is especially necessary when large culture volumes or bacterial strains producing large amounts of carbohydrates are used.

☞ Remove a ▲ 400 µl or ● 240 µl sample from the combined wash fractions and save for an analytical gel (sample 3).

12. Elute DNA with ▲ 5 ml or ● 15 ml Buffer QF.

Collect the eluate in a 15 ml or 50 ml tube (not supplied). Use of polycarbonate centrifuge tubes is not recommended as polycarbonate is not resistant to the alcohol used in subsequent steps.

Note: For constructs larger than 45–50 kb, prewarming the elution buffer to 65°C may help to increase yield.

☞ Remove a ▲ 100 µl or ● 60 µl sample of the eluate and save for an analytical gel (sample 4).

ƒ If you wish to stop the protocol and continue later, store the eluate at 4°C. Storage periods longer than overnight are not recommended.

13. Precipitate DNA by adding ▲ 3.5 ml or ● 10.5 ml (0.7 volumes) room-temperature isopropanol to the eluted DNA. Mix and centrifuge immediately at ≥15,000 x *g* for 30 min at 4°C. Carefully decant the supernatant.

All solutions should be at room temperature in order to minimize salt precipitation, although centrifugation is carried out at 4°C to prevent overheating of the sample. Alternatively, disposable conical bottom centrifuge tubes can be used for centrifugation at 5000 x *g* for 60 min at 4°C. Isopropanol pellets have a glassy appearance and may be more difficult to see than the fluffy, salt-containing pellets that result from ethanol precipitation. Marking the outside of the tube before centrifugation allows the pellet to be more easily located. Isopropanol pellets are also more loosely attached to the side of the tube, and care should be taken when removing the supernatant.

14. Wash DNA pellet with ▲ 2 ml or ● 5 ml of room-temperature 70% ethanol, and centrifuge at ≥15,000 x *g* for 10 min. Carefully decant the supernatant without disturbing the pellet.

Alternatively, disposable conical-bottom centrifuge tubes can be used for centrifugation at 5000 x *g* for 60 min at 4°C. The 70% ethanol removes precipitated salt and replaces isopropanol with the more volatile ethanol, making the DNA easier to redissolve.

15. Air-dry the pellet for 5–10 min, and redissolve the DNA in a suitable volume of buffer (e.g., TE buffer, pH 8.0, or 10 mM Tris·Cl, pH 8.5).

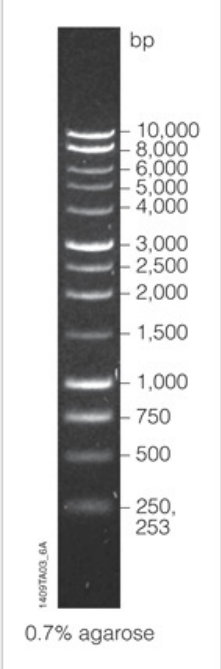
Redissolve the DNA pellet by rinsing the walls to recover all the DNA, especially if glass tubes have been used. Pipetting the DNA up and down to promote resuspension may cause shearing and should be avoided. Overdrying the pellet will make the DNA difficult to redissolve. DNA dissolves best under slightly alkaline conditions; it does not easily dissolve in acidic buffers.

Determination of yield

To determine the yield, DNA concentration should be determined by both UV spectrophotometry at 260 nm and quantitative analysis on an agarose gel. For reliable spectrophotometric DNA quantification, A_{260} readings should lie between 0.1 and 1.0.

Agarose gel analysis
We recommend removing and saving aliquots during the purification procedure (samples 1–4). If the plasmid DNA is of low yield or quality, the samples can be analyzed by agarose gel electrophoresis to determine at what stage of the purification procedure the problem occurred (see page 34).
(extracted from QIAfilter plasmid purification handbook 3rd edition 11/2005)

C5: 1 kb DNA ladder (Promega)



C6: Single genome amplification (SGA) of HIV-1 subtype C complete genomes

(based on SOP # CHAVI-MBSC-2 entitled "whole genome amplification of hiv-1 from a single RNA template", University of Alabama at Birmingham, Alabama, USA)

1. Purpose:

To PCR amplify the whole HIV-1 genome, minus the Long Terminal Repeats (LTRs), from a single viral RNA template (single genome).

2. Background and overview of procedures:

-This protocol involves the dilution of viral cDNA to a point that, upon setting up multiple replicate PCR reactions with the cDNA as template, the product generated has a high probability of having been amplified from a single genome copy.

-In this case, according to Poisson's distribution, no more than 30%, or 28 out of a 96-well reaction plate with 2 negative controls, should be positive for amplification product to ensure that 6 of 7 amplified DNAs arise from a single-copy template.

-For example, if viral load data is provided, a known amount of viral RNA is purified from plasma using the QIAamp Viral RNA Mini Kit and it is transcribed at a theoretical concentration of 200 copies per µL using a gene-specific primer with Invitrogen's Superscript III (a reverse transcriptase that has been genetically modified to reduce RNase H activity and to maintain cDNA synthesis activity at temperatures (ie. 50-55ºC) required to melt template secondary structure).

-The viral cDNA is then serially diluted to obtain single copy in a two-step "nested" PCR procedure.

-Due to variations in sample integrity, stability (ex. freeze/thaw), and serum components, etc., it may be necessary to repeat the PCR reactions at a lower or higher dilution to achieve the desired results.

-The resulting positive amplification products are initially subjected to direct DNA sequencing the entire amplicon, all of which must be free of any double peaks.

3. Viral RNA extraction

NOTE 1: perform in RNA-only clean room.

Use the QIAamp Viral RNA Mini Kit and protocol, (catalogue numbers 52904, 52906, or 52908);

4 cDNA synthesis

4.1 Reagents and Solutions

4.1.1 Superscript III Reverse Transcriptase, Cat. No. 18080-044, Invitrogen Corp. (components: SSIII RT (200 U/µL), 5X First-Strand Buffer, 0.1 M DTT)

4.1.2 Sterile, RNase-free water

4.1.3 10mM dNTP mix

4.1.4 RNA template (Sec. 6)

4.1.5 RNaseOUT, Cat. No. 10777-019, Invitrogen Corp.

4.1.6 20 uM Primer

4.1.6.1 oligodT20

4.1.7 RNase H, Cat. No. 18021-014

4.2 Procedure

NOTE 1: Perform in RNA-only clean room.

NOTE 2: Spin down condensate after heat incubation steps.

NOTE 3: Use viral RNA extract equivalent to 5,000 copies.

A. Pipette the following components into a 0.5 ml RNase-free tube:

µL/tube	[stock]	[final]
8.50 H2O		
0.50 primer	20 µM	0.25
2.00 dNTP mix	10 mM each	0.5
15.00 RNA template		
26.00 per tube		

B. Place tube in 65ºC heat block for 5', remove to ice for 1'.

C. Add the following components:

µL/tube	[stock]	[final]
8.00 5x Buffer	5x	1x
2.00 DTT	0.1 M	0.005 M
2.00 RNaseOUT	40 u/µL	2 u/µL
1.00 SSIII RT	200 u/µL	5 u/µL
13.00 per tube		

40.00 final volume

D. Mix and incubate at 45ºC for 1.5 hr.

E. Add 1 µL additional SSIII RT, and then increase to 45ºC for 1.5 hr.

F. Inactivate SSIII RT by heating at 70ºC for 15'.

G. To each tube, add 1 µL RNase H, incubate at 37ºC for 20'.

5. PCR amplification

Note: PCR reaction mixes should be made up and aliquoted in an area free of PCR amplified or plasmid DNA.

5.1 Establish appropriate titration

5.1.1 cDNA Dilutions

5.1.1.1 Perform serial dilutions of cDNA to obtain <30% positivity.

5.1.1.2 Proceed to PCR amplification. If too few/too many positive reactions, repeat PCR with lower/higher dilution.

5.2 First round amplification reaction

5.2.1 Prepare PCR reaction mix for 1-16 reactions (ex. 15 test + 1 negative control).

µL/rxn	x 16	[stock]	[final]
40.00	640 dH2O		
5.00	80 10X buffer 1 with Mg++		
1.75	28 dNTPs	10mM each	0.35 µM
0.75	12 Expand Long Template Enzyme Mix		
0.75	12 1.U5C	20µM	0.3 µM
0.75	12 1.U5Cb	20µM	0.3 µM
0.75	12 1.5.5pIC	20µM	0.3 µM
49.00	784		

5.5.2 If using master mix, add 49.00 µL + 1 µL H2O to the negative well. Add 1.0 µL of 2-fold serially diluted cDNA per well in triplicate wells; for example. 1:2, 1:4, 1:8. 1:16, and 1:32.

5.2.3 Once the correct cDNA dilution has been determined, add 15 µL of diluted cDNA to the master mix and then add 50 µL to each of 15 wells.

5.2.3 Place in thermal cycler.

5.2.4 Run with following parameters:

	ºC	Time
	94	2 min
10 cycles	94	15 sec
	60	30 sec
	68	8 min
25 cycles	94	15 sec
	68	30 sec
	68	8 min, add 20 sec each cycle
	68	7 min
	4	hold

5.2.5 From all positive wells, save remaining first round in 0.5ml tube at -80ºC for future full-length genome cloning.

6.1 Second round (Nested) PCR reaction

6.1.1 Prepare PCR reaction mix for 1-16 reactions (ex. 15 test + 1 negative control).

µL/rxn	x 16		[stock]	[final]
40.00	640	dH2O		
5.00	80	10X buffer 1 (Mg++)		
1.75	28	dNTPs	10mM each	0.35 µM
0.75	12	Expand Long Template Enzyme Mix		
0.75	12	2.U5C	20 µM	0.3 µM
0.75	12	2.5.5pIC	20µM	0.3 µM
49.00	784			

6.1.2 Pipette 49.0 µL to each well of a 96-well plate. Add 1.0 µL from each of the First Round PCR reaction samples to the corresponding well of the nested PCR plate. Mix by pipetting up and down.

6.1.3 Place in thermal cycler.

6.1.4 Run with following parameters:

	ºC	Time
	94	2 min
10 cycles	94	15 sec
	55	30 sec
	68	8 min
25 cycles	94	15 sec
	55	30 sec
	68	8 min, add 20 sec each cycle
	68	7 min
	4	hold

7 Proceed to reaction analysis.

7.1 Agarose Gel Electrophoresis, 2 µL sample. Select all positive wells.

7.2 Sequence DNA directly (0.5 – 1µL per sequencing reaction) after ExoSAP pretreatment.

7.3 Store remaining 2nd round product for future use.

C7: Single genome amplification of the HIV-1 *env* gene (gp160)

(adapted from CHAVI-MBSC-0.03, University of Alabama at Birmingham, Alabama, USA)

1. Purpose:

To amplify the HIV-1 envelope gene from a single viral RNA template (single genome).

2. Background and overview of procedures:

-This protocol involves the dilution of viral cDNA to a point that, upon setting up multiple replicate PCR reactions with the cDNA as template, the product generated has a high probability of having been amplified from a single genome copy.

-In this case, according to Poisson's distribution, no more than 30%, or 28 out of a 96-well reaction plate with 2 negative controls, should be positive for amplification product to ensure that 6 of 7 amplified DNAs arise from a single-copy template.

-For example, if viral load data is provided, a known amount of viral RNA is purified from plasma using the QIAamp Viral RNA Mini Kit and it is transcribed at a theoretical concentration of 200 copies per µL using a gene-specific primer with Invitrogen's Superscript III (a reverse transcriptase that has been genetically modified to reduce RNase H activity and to maintain cDNA synthesis activity at temperatures (ie. 50-55ºC) required to melt template secondary structure).
-The viral cDNA is then serially diluted to obtain single copy in a two-step "nested" PCR procedure.
-Due to variations in sample integrity, stability (ex. freeze/thaw), and serum components, etc., it may be necessary to repeat the PCR reactions at a lower or higher dilution to achieve the desired results.

-The resulting positive amplification products are initially subjected to direct DNA sequencing in the highly variable V1-V2 regions to pre-test for sequence homogeneity (lack of multiple superimposed peaks).
-To confirm single-genome amplification, the entire envelope gene is sequenced and cannot contain any double peaks.

3. Procedure for RNA extraction

Note: Work with plasma in sterile Biosafety cabinet.

3.1 Use QIAamp Viral RNA Mini Kit following protocol as written in Handbook with the following adaptations:

3.2 Normalize the 140 µL plasma sample to contain equivalent viral copies (i.e. 20,000 copies per 140 µL).

3.2.1.1.1 Increase volume to 140 µL with PBS.

3.2.1.1.2 Or, if viral load is too low, concentrate by centrifugation for 1 hr at 17k rpm. Remove supernatant down to 140 µL and resuspend viral pellet.

3.3 Elute RNA from spin column with 55 µL buffer (stock elution buffer is diluted 1:5 in sterile RNase/DNase free H_2O) to achieve approximately 50 µL net recovery.

4 cDNA synthesis

4.1 Reagents and Solutions

4.1.1 Superscript III Reverse Transcriptase, Cat. No. 18080-044, Invitrogen Corp. (components : SSIII RT (200 U/µL), 5X First-Strand Buffer, 0.1 M DTT)

4.1.2 Sterile, RNase-free water

4.1.3 10mM dNTP mix

4.1.4 RNA template (Sec. 6)

4.1.5 RNaseOUT, Cat. No. 10777-019, Invitrogen Corp.

4.1.6 20 uM Primer, subtype specific

4.1.7 RNase H, Cat. No. 18021-014

4.2 Procedure for cDNA synthesis

Note:perform in RNA-only clean room. Spin down condensate after heat incubation

A Pipette the following components into a 0.5 ml RNase-free tube:

µL/tube		[stock]	[final]
8.75	H2O		
1.25	primer (OFM19or oligo-dT_{20})	20 µM	0.25
5.00	dNTP mix	10 mM each	0.5
15.00	per tube		
50.00	RNA template		
65.00 total volume			

Place tube in 65ºC heat block for 3-5', remove to ice for 1'.

Add the following components:

µL/tube		[stock]	[final]
20.00	5x Buffer	5x	1x
5.00	DTT	100 mM	5
5.00	RNaseOUT	40 u/µL	2
5.00	SSIII RT	200 u/µL	10

35.00 per tube

100.00 final volume

Mix and incubate at 45ºC for 1 hr and then increase to 45ºC for 1 hr. Inactivate SSIII RT by heating at 70ºC for 15'.

To each tube, add 1 µL RNase H, incubate at 37ºC for 20'.

5 PCR amplification

Note: PCR reaction mixes should be made up and aliquoted in an area free of PCR amplified or plasmid DNA.

5.1 Reagents and Solutions

5.1.1 Sterile H2O, ultrapure

5.1.2 Subtype C Primers:

5.1.2.1 OFM19: 5'– GCACTCAAGGCAAGCTTTATTGAGGCTTA

5.1.2.2 VIF1: 5'– GGGTTTATTACAGGGACAGCAGAG

5.1.2.3 ENV A: 5'– GGCTTAGGCATCTCCTATGGCAGGAAGAA

5.1.2.4 ENV N: 5'– CTGCCAATCAGGGAAGTAGCCTTGTGT

5.1.3 High Fidelity Platinum Taq–Invitrogen 11304.

5.1.4 dNTPs (10mM each).

5.1.5 Diluted cDNA template (above)

5.2 Establish Appropriate Titration

5.2.1 cDNA Dilutions

5.2.1.1 Perform serial dilutions of cDNA to obtain <30% positivity.

5.2.1.2 Proceed to PCR amplification. If too few/too many positive reactions, repeat PCR with lower/higher dilution.

If viral load is not known, use the PCR conditions in sec. 8.5 to determine the dilution at which a single cDNA molecule is present per reaction by nested PCR. Do up to 16

replicates at each dilution of trial serial cDNA dilutions. Select dilution that gave less than 30% positives for further SGA study.

If viral load is known and 20,000 viral copies are extracted, determine the dilution at which cDNA will yield a single cDNA molecule per reaction by starting out with the following dilution:

For example:

viral copies	total cDNA vol	copies/ul	cDNA dilution	copies/ul	Estimated # PCR+/ total replicates	% PCR +
20,000	100	200	1:20	10	16/16	100%
			1:60	3.3	13/16	81%
			1:180	1.1	4/16	25%
			1:540	0.4	0/16	0%

*Select 1:180 dilution for further SGA study.

5.3 First Round Amplification Reaction

5.3.1 Prepare PCR reaction mix for 1-100 reactions (94 test + 2 negative controls).

µL/rxn	x 100		[stock]	[final]
15.3	1530	dH2O		
2	200	10X buffer (supplied)		
0.8	80	$MgS0_4$ (supplied)		
0.4	40	dNTPs (10mM each)		
0.1	10	Taq (High Fidelity Platinum)		
0.2	20	OFM19	20µM	0.2 µM
0.2	20	VIF	120µM	0.2 µM
19.0	1900			

Add 20 µL to the negative wells.

1µL cDNA per well. Once the correct cDNA dilution has been determined, add 100 µL of diluted cDNA to the master mix and then add 20 µL per well.

5.3.2 Place in thermal cycler. Run with following parameters:

	ºC	Time
	94	2 min
35 cycles	94	15 sec
	55	30 sec
	68	4 min
	68	10-20 min
	4	hold

5.3.3 From all positive wells, save remaining first round in 0.5ml tube at -80ºC for future envelope cloning.

5.4 Second Round (Nested) PCR Reaction

5.4.1 Prepare PCR reaction mix for 1-100 reactions (94 test + 2 negative controls).

µL/rxn	x 100		[stock]	[final]
15.3	1530	dH2O		
2	200	10X buffer (supplied)		
0.8	80	$MgS0_4$ (supplied)		
0.4	40	dNTPs (10mM each)		
0.1	10	Taq (High Fidelity Platinum)		
0.2	20	ENV A	20µM	0.2 µM
0.2	20	ENV N	20µM	0.2 µM
19.0	1900			

5.4.2 Pipette 19 µL to each well of a 96-well plate. Add 1µL from each of the First Round PCR reaction samples to the corresponding well of the nested PCR plate. Mix by pipetting up and down.

5.4.3 Place in thermal cycler. Run with following parameters.

	ºC	Time
	94	2 min
45 cycles	94	15 sec
	55	30 sec
	68	4 min
	68	10-20 min
	4	hold

6 Proceed to reaction analysis.

6.1.1 Agarose Gel Electrophoresis, 3-5 µL sample. Select all positive wells.

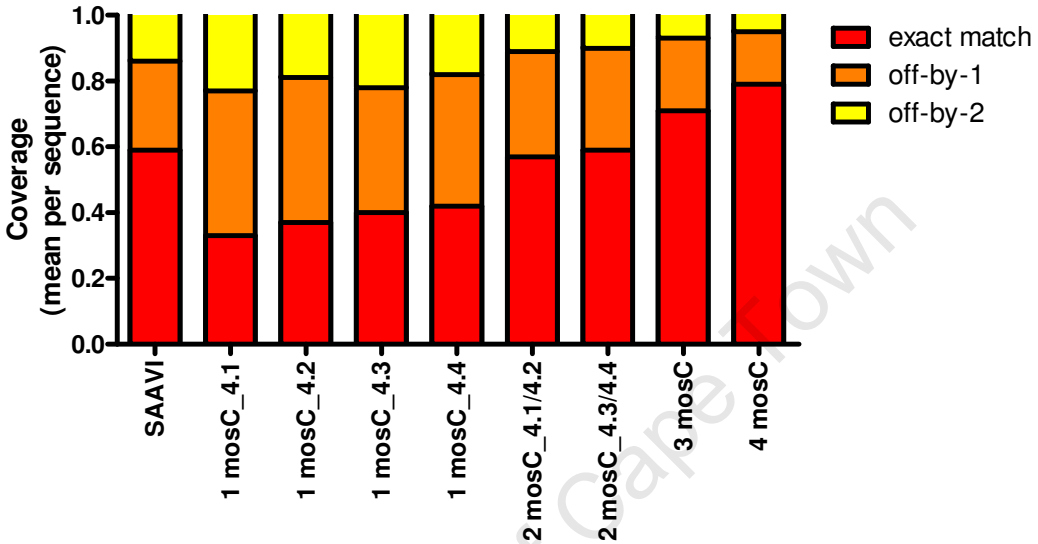
6.1.2 Sequence DNA directly (0.5 – 1µL per sequencing reaction). Store remaining 2nd round product for future use.

APPENDIX D

D1: Quality and depth coverage assessment of HIV-1 whole genomes assembled from sequences obtained from cloned genome amplicons.

Strain_clone	Time post infection	Average quality	Coverage
CAP63_F4	34 d	86	4.9
CAP65_F1	42 d	87.4	6.3
CAP69	*14 d*	89.4	7.2
CAP85_F1	23 d	84.1	5.1
CAP88_F2	36 d	87.1	5.3
CAP174	28 d	88.9	6.9
CAP206_F1	41 d	86.4	6
CAP210_F1	36 d	88.1	6
CAP221	*14 d*	87.8	5.4
CAP228_F2	53 d	86.5	4.5
CAP229_F1	48 d	81.6	3.8
CAP239_F1	36 d	87.8	7.1
CAP244_F1	58 d	87.7	6.6
CAP248_F1	62 d	89.7	5.5
CAP255	54 d	86.9	5.2
CAP256	42 d	86.6	6.2
CAP257	49 d	88	5.6

D2: This figure shows the predicted epitope coverage by the SAAVI consensus-based Gag immunogen and different valencies of the subtype C Gag mosaic proteins (Fischer *et al.*, 2007) against primary infection subtype C viruses



D3: Putative CTL epitopes with evidence of escape

Gag	Peptide	HLA	Comment
CAP8	QYMLKHLVWAS	A*2301	HIV Immunology Compendium
	DIYKRWIILGLNKIVRM	B*0801	HIV Immunology Compendium
CAP30	ALNP/GLLET	A*0201	A2, HIV Immunology Compendium
	SLYNTVA*TL/YCVHEDT*	A*0201	A2, HIV Immunology Compendium 2008
	GHQAAMQMLKE	A*0201	A2, HIV Immunology Compendium
	WIILGLNKIVRMYSPVSIL	A*0201	A2, HIV Immunology Compendium
	FLQNRPEPT	A*0201	A2, HIV Immunology Compendium
CAP61	VQNAQGQ/MVHQPLSPR	A*6602	A*6603, HIV Immunology Compendium
CAP63	IILGLN/KIVRMYSPVSIL	A*0201	HIV Immunology Compendium
CAP85	NIMMQRGNI	B*0801	
CAP239	STLQEQVAM	B*5801	HIV Immunology Compendium
CAP257	KKHYMLKHLVW	A*2301	HIV Immunology Compendium
	VRMYSPVSIL/DI	A*2301	HIV Immunology Compendium
Pol		HLA	Comment
CAP30	ALTAICEEM	A*0201	HIV Immunology Compendium
	SWTVNDIQ/KLVGKLNWA	A*0201	A2, HIV Immunology Compendium
CAP61	HTN/DVKQLTEAV	A*6802	HIV Immunology Compendium
	DIVPLTEEA/ELE	A*6802	
	ELAENREIL/KEPVH	A*6802	
CAP63	DIQ/KLVGKLNWA	A*0201	A2, HIV Immunology Compendium
	NWRAMAADF	A*0201	A2, HIV Immunology Compendium
	IWQLDCTHL/EGKVI	A*2301	
	WWAGIKQEF/GIPY	A*2301	
	QAEHL/KTAVQMAVF	A*0201	A2, HIV Immunology Compendium
CAP85	RGRQKVVSL	B*0801	B*0801
CAP206	GPKVKQWPL/TEEK	B*0702	B7, HIV Immunology Compendium
	G/KLNWASQIYP	A*3204	
	WEFVNTPPL	B*4403	
	SESEIVNQI	B*0702	B7, HIV Immunology Compendium
	LIKKE/KVYLSWVPA	A*0702	B7, HIV Immunology Compendium
CAP210	WTVNDIQKL	A*6802/ B*1510	B15, HIV Immunology Compendium
CAP239	PIVLPEKESW	B*5801	HIV Immunology Compendium
	SAAVKAACW/WAGIQQEF	B*5801	HIV Immunology Compendium
CAP257	PYQIYQEPF	A*2301	HIV Immunology Compendium
	SESELVNQI/IEELIK	B*4403	
	TKLEGKVIL/VAVHVASGY	?B*1503	B*1510, HIV Immunology Compendium
	MAVFIHNFK/RK	A*2902	
	TKELQ/KQITKIQNF	B*1503	HIV Immunology Compendium

D3: Continue

Env		HLA	Comment
CAP8	WWIWSSLGF/WML	A*2301	
	LWVTVYYGV/PVWK	A*2301	
	KD/KKQKVYALF/	A*2301	
	EIMIRSENL	B*0801	
	WYIKIFII	A*2301	
	RSIRLVNGF	B*0801	
CAP30	WMLMIYNVV	A*0201	
	CSFNATTER/RDKK	A*3402	
	VYEREVH/ NVWATHACV	A*0201	A2, HIV Immunology Compendium
	ALFYRPDIV/LLSQND	A*0201	
	T/IIVHLNESV	A*0201	
	SGGD/LEITTHS/FNCR	A*0201	A2, HIV Immunology Compendium
	SEITGLLLT/RDGGR	B*4501	
	LLQYWSREL	A*0201	HIV Immunology Compendium
CAP61	TEN/FNMWKNDMV	A*6802	A68, HIV Immunology Compendium
	CSFNTTTEI	A*6802	
	YTNTIYRLL	A*6802	
	IR/LVSGFLSLA	A*6802	
	SLFDTIAIA/VAE	A*6802	A68, HIV Immunology Compendium
CAP63	HFPNKTITF	A*2301	
	NYTQTIYQL	A*2301	
	LLDSIAITV	A*0201	HIV Immunology Compendium
	RIIEVIQRF	A*0201	HIV Immunology Compendium
CAP85	NM/KDNWRSELY	A*3002	
	RYLGSLVQY	A*3002	HIV Immunology Compendium
CAP206	KPCVKLTPLC/VTLN	B*0702	HIV Immunology Compendium
	WKNDMVDQM	B*4403	
CAP210	CSFNATTEL	A*6802	A2, HIV Immunology Compendium
	LLNATAIAV/AEAT	A*6802	HIV Immunology Compendium
CAP239	VTFDPIPIHY/CAPA	A*2902	A29, HIV Immunology Compendium
	THI/FNCRGEFFY	A*2902	A29, HIV Immunology Compendium
	FSITNWLWY	B*5801, A*0101	
CAP244	GMLRNYQQW	A*2301	
	KQKEYALFY	A*3004	
CAP257	YEKEVHNVW	B*4403	
	NYTNSIYKL/LEE	A*2301	
	D/KWQNLWNWF	A*2301	
	KYLGSLVQY	A*2902,A*2301	

D3: Continue

Vif		**HLA**	**Comment**
CAP8	EWRLRRYSTQVD KPKKRKPPL	B*1510 B*0801	HIV Immunology Compendium
CAP30	GLADQLIHI/HYF	A*0201	A2, HIV Immunology Compendium
CAP61	WRLRR/YSTQVDPGL	A*6802	
CAP63	SLVKHHMYV KVG/SLQYLALTA	A*0201 A*0201	A2, HIV Immunology Compendium A2, HIV Immunology Compendium
CAP206	RTWNSLVKH KPPLPSVTKL/TEDRWNN	A*3204 B*0702	 B7, HIV Immunology Compendium
CAP210	WHLGHGVSIEWRLRKY	B*1510	HIV Immunology Compendium
CAP239	VSGKARGWF/YRHHY	B*5801	HIV Immunology Compendium
CAP257	TESAIRQAI	B*4403	
Vpr		**HLA**	**Comment**
CAP30	SLGQYIYET	A*0201	A2, HIV Immunology Compendium
CAP61	FPRPWLHDL	B*4201	HIV Immunology Compendium
CAP63	I/RILQQLLFI	A*0201	HIV Immunology Compendium
CAP244	QHIYETYGDTW	A*2301	
CAP257	WLHSLGQYI	A*2902/ B*4202	B*4201, HIV Immunology Compendium
Vpu		**HLA**	**Comment**
CAP8	MEQLRLLDV/N	B*0801	
CAP61	IIAVILAIV	A*6802	A2, HIV Immunology Compendium
CAP85	AIIVWTITYLEY	A*3002	
Tat		**HLA**	**Comment**
CAP8	CSYHCLACF, YHCLAC/FITKALGISY	A*2301/ B*1510	B15, HIV Immunology Compendium
Rev		**HLA**	**Comment**
CAP63	EALL/LAVRTIKLL QIHSISERIL	A*0201 A*0201	 A2, HIV Immunology Compendium
CAP85	RNRRRRWRA/RQR RHI/REISERILS	B*0801 B*4501	

D3: Continue

Nef		**HLA**	**Comment**
CAP8	EGLI/YSKKRQEIL	B*0801	HIV Immunology Compendium
	KFD/SQLARRHIA	B*0801/ B*1510	HIV Immunology Compendium
CAP30	QEEEEEVGFPV	B*4501	HIV Immunology Compendium
CAP45	AQEEEEEVGFPV	B*4501	HIV Immunology Compendium
CAP61	STNADMAWL	A*6802	
CAP63	QEEEGVGF/PV	B*4501	HIV Immunology Compendium
	ILDLWVYHT	A*0201	HIV Immunology Compendium
CAP85	AQE/EEKEVGFPV	B*4501	HIV Immunology Compendium
	AASQDLGKY	A*3002	
	GYFPDWQNY	A*3002	HIV Immunology Compendium
	MAREKHPEF	B*0801	
CAP210	REVLQ/WKFDSSLAF/	B*1510	B15, HIV Immunology Compendium
CAP239	RPMTYKAAV / KAAVDLSFF	B*5801	HIV Immunology Compendium
	YTPGPGVRYPL/TFGWC	B*4201/ B*5801	B58, HIV Immunology Compendium
CAP244	QEILDLWVY	B*4403	Matthews *et al*., 2008
CAP257	QEILDLWVY	B*4403	Matthews *et al*., 2008

D4: Escape at putative CTL epitopes in a slow progressor – CAP45

Rev

Weeks post infection	Peptide (p49-61)	No. of sequences with mutation	HLA association	Escape
2 (P1V05)	QRQIRSISERILSTCLGR	20/20	B*1510	12-16 wks
5 (P2V00)	.....	7/7		
12 (P2V05)	.....	4/4		
16 (P2V07)	.....F.....	1/1		
52 (P3V10)	.....	1/7		
	.....F.....	6/7		
65 (P4V17)	.....	1/3		
	.....F.....	2/3		

Nef

Weeks post infection	Peptide (p58-75)	No. of sequences with mutation	HLA association	Escape
2	LQAQEEEEEVGFPVRPQV	4/4	B*4501	5-12 wks
5	.....	8/8		
12	.....G.....	3/4		
	.....X.....	1/4		
16	.....G.....	1/1		
65	.....G.....	3/3		

D5: Escape at putative CTL epitopes in a rapid progressor – CAP63

Pol

Weeks post infection	Peptide (p186-223)	No. of sequences with mutation	HLA association	Escape
2 (P1V00)	KIKALTEICEEMEKEGKI	10/10	A*0201	5-11 wks
5 (P2V01)	.....	3/3		
11 (P2V05)	.....K.....	6/7		
	.....K.....	1/7		
29 (P3V10)	.....K.....	1/3		
	.....D.....	1/3		
	.....A.....	1/3		
37 (P3V12)	.....K.....	1/4		
	.....G.....	3/4		
	Peptide (p522-539)			
2	KQLTEAVHKIAIESIVIW	11/11	A*0201	29- 37 wks
5	.....	1/1		
11	.....	7/7		
29	.....	3/3		
37	.....Q.....	1/4		
	.....V.....	1/4		
	.....T.....	2/4		

Vpr

Weeks post infection	Peptide (p57-74)	No. of sequences with mutation	HLA association	Escape
2	VEALIRILQQLLFVHFRI	11/11	A*0201	5-11 wks
5	.....	1/1		
11	.....S.....	4/7		
	.....M.....	1/7		
	.....T.....	1/7		
	.....K.....	1/7		
29	....V.T.....	1/3		
	.....M.....	1/3		
	.....T.....	1/3		
37	.....M.....	1/5		
	.....T.....	4/5		

Tat

Weeks post infection	Peptide (p81-101)	No. of sequences with mutation	HLA association	Escape
2	.....	28/30	No	2-11 wks
	.....K.....	2/30		
5	.....	1/1		
11	.....T....	3/8		
	.....G.....	2/8		
	.....E.	3/8		
29	.....Q..	10/29		
	.....G.....	6/29		
	.....A.....	1/29		
	.....R.....	3/29		
	.....N.....	1/29		
	S.....Q..	1/29		
	....K.....Q..	1/29		
	.....N.....Q..	1/29		
	.....R..A.....	2/29		
	.....A.....	1/29		
	.....G...T....	1/29		
	.....LS.	1/29		
37	.....R.....	1/5		
	.....Q..	3/5		
	.....N.....Q..	1/5		

Rev

Weeks post infection	Peptide (p41-58)	No. of sequences with mutation	HLA association	Escape
2	RRRRWRARQQQIHKISER	29/30	A*2301	2-11 wks
	.K.....	1/30		
5	.....	1/1		
11	.....	4/8		
	.....Q.....	3/8		
	..K.....	1/8		
29	.....	9/29		
	......E.....	3/29		
	......T.....N.....	1/29		
	.....YQ....	1/29		
	......G.....	2/29		
	......T.....	1/29		
	.....N.....	12/29		
37	.....N.....	2/4		
	......E.....	1/4		
	......T.....N.....	1/4		

Env

Weeks post infection	Peptide (p147-166)(gp120)	No. of sequences with mutation	HLA association	Escape
2	MIGEIKNCSFNATTELRDKK	30/30	A*2301	2-5 wks
5	..........T........	1/1		
11	..........S.......	4/8		
	..............V.....	1/8		
	....	1/8		
	..............A.....	1/8		
	....................E	1/8		
29	.TE..........A.....	1/29		
	.R.........S........	4/29		
	.T.........S........	1/29		
	.R..................	1/29		
	................W...	6/29		
	..........S.......	1/29		
	..............I....Q	1/29		
	..............I....	4/29		
	..........T...I....	3/29		
	..............A.....	1/29		
	I..............I....	2/29		
	..............V....E	1/29		
	..E..........A.....	1/29		
	..........S........	1/29		
	I...................	1/29		
37	I..............I....Q	1/5		
	I..............I....	1/5		
	..............I....	1/5		
	.R............I....	1/5		
	......X....S.......	1/5		
	Peptide (p325-342)(gp120)			
2	IRQAHCNISKTQWNTTLE	30/30	No associated HLA	11-29 wks
5	....	1/1		
11	........X.........	1/8		
29	....	16/29		
	.........N........	2/29		
	..........P.......	2/29		
	.........E........	2/29		
	....Y.....K.......	1/29		
	............D.N..	1/29		
	..............I..	1/29		
	..............A...	4/29		
37	....	1/4		
	..........A.......	1/4		
	..........K.......	1/4		
	.........QK.......	1/4		

Env

Weeks post infection	Peptide (p389-413)(gp120)	No. of sequences with mutation	HLA association	Escape
2	FNSTYMPNGIHIPNGASEVITLPCR	30/30	No associated HLA	2-5 wks
5	.......S..................	1/1		
11	..........T.....T........	1/8		
	.......S---------........	2/8		
	.......S...-----.........	1/8		
	.......S-------S.........	1/8		
	........-------S.........	1/8		
	----------L..............	1/8		
	---------.........D......	1/8		
29	.......S..T---------.....	6/29		
	..........T.....T---.....	2/29		
	.......S..T------..F.....	1/29		
	...N.------.....T..D.....	6/29		
	......------....T..D.....	3/29		
	.......S.------.-.........	1/29		
	....S-----......T..D.....	1/29		
	..........TQ....T........	1/29		
	.....D....T.---------....	1/29		
	.......S.----DASDN.......	1/29		
	.......S..--....T...-....	1/29		
	...N.------.....TN.D.....	2/29		
	...N..------....T..D.....	1/29		
	......LS..N-----...D.....	1/29		
	.......S..T------N.D.....	1/29		
37	...N.------.....T..F.....	1/4		
	...N.------.....TN.D.....	1/4		
	.......S..T.MS.DT..D.....	1/4		
	.......S..T---------.....	1/4		
	Peptide (p608-625)(gp41)			
2	SWSNKSEEDIWGNMTWMQ	30/30	No associated HLA	11-29 wks
5	....	1/1		
11	....	8/8		
29	....	6/29		
	......KK..........	1/29		
	...........D......	6/29		
	.......K...D......	2/29		
	.......K..........	14/29		
37	.......G...D......	1/5		
	.......K...D......	4/5		

Env

Weeks post infection	Peptide (p728-745)(gp41)	No. of sequences with mutation	HLA association	Escape
2	IEEEGGEQDNSRSIRLVS	29/30	A*2301/A*0201	2-11wks
	..K.....	1/30		
5	.....	1/1		
11	.....N.....	2/8	Gain of glycosylation site, putative Ab pressure	
	...K.....N.....	1/8		
	.....R....	3/8		
	.....G.....	2/8		
29	.....	16/29		
	.....G.....	6/29		
	.....G...N.....	1/29		
	.....K.....	3/29		
	.....E.S.....	2/29		
	.....S.....	1/29		
37	.....	4/5		
	.....K.....	1/5		
	Peptide (p816-833)(gp41)			
2	LLDSIAITVAEGTDRIIE	30/30	A*0201	2-5 wks
5	..N.....	1/1		
11	.....I.....	4/8		
	.....A.....	2/8		
	..G.....	1/8		
	.....I....R…..	1/8		
29	.....I.....	16/29		
	.F.....I.....	1/29		
	..N....A.....	4/29		
	.....AA.....	1/29		
	.....A.....	7/29		
37	.....I.....	3/5		
	.....A.....	1/5		
	..N.....I.....	1/5		

Nef

Weeks post infection	Peptide (p41-58)	No. of sequences with mutation	HLA association	Escape
2	GALTSSNTAATNADCAWL	10/11	A*0201	11 wks
	.....X....	1/11		
5	.....	1/1		
11	.....	7/8		
	.....D.....	1/8		
29	.....	1/3		
	....N.....	2/3		
37	.....	2/5		
	.....R.....	1/5		
	....N.....	2/5		
	Peptide (p57-73)			
2	WLEAQEEEEVGFPVRPQV	8/11	B*4501	2-5 wks
	.....X.....	1/11		
	.....X.....	1/11		
	.....K.....	1/11		
5	.....G.....	3/3		
11	.....G.....	5/8		
	.....D.....	2/8		
	.....D.....	1/8		
29	.....D..L.....	1/3		
	.....D.....	2/3		
37	.....D.....	5/5		

REFERENCES

A

Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping LH, Athreya GS, Treurnicht FK, Keele BF, Wood N, Salazar-Gonzalez JF, Bhattacharya T, Chu H, Hoffman I, Galvin S, Mapanje C, Kazembe P, Thebus R, Fiscus S, Hide W, Cohen MS, Karim SA, Haynes BF, Shaw GM, Hahn BH, Korber BT, Swanstrom R, Williamson C; CAPRISA Acute Infection Study Team; 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(8):3556-67. Erratum in: J Virol. 2009 Jul; 83(13):6974.

Ahmad A, Menezes J. 1995. **Positive correlation between the natural killer and gp 120/41-specific antibody-dependent cellular cytotoxic effector functions in HIV-infected individuals.** J Acquir Immune Defic Syndr Hum Retrovirol., 10(2):115-9.

Alam SM, Scearce RM, Parks RJ, Plonk K, Plonk SG, Sutherland LL, Gorny MK, Zolla-Pazner S, Vanleeuwen S, Moody MA, Xia SM, Montefiori DC, Tomaras GD, Weinhold KJ, Karim SA, Hicks CB, Liao HX, Robinson J, Shaw GM, Haynes BF. 2008. **Human immunodeficiency virus type 1 gp41 antibodies that mask membrane proximal region epitopes: antibody binding kinetics, induction, and potential for regulation in acute infection.** J Virol., 82(1):115-25.

Alam SM, McAdams M, Boren D, Rak M, Scearce RM, Gao F, Camacho ZT, Gewirth D, Kelsoe G, Chen P, Haynes BF. 2007. **The role of antibody polyspecificity and lipid reactivity in binding of broadly neutralizing anti-HIV-1 envelope human monoclonal antibodies 2F5 and 4E10 to glycoprotein 41 membrane proximal envelope epitopes.** J Immunol., 178(7):4424-35.

Allen TM, Altfeld M, Geer SC, Kalife ET, Moore C, O'sullivan KM, Desouza I, Feeney ME, Eldridge RL, Maier EL, Kaufmann DE, Lahaie MP, Reyor L, Tanzi G, Johnston MN, Brander C, Draenert R, Rockstroh JK, Jessen H, Rosenberg ES, Mallal SA, Walker BD. 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(21):13239-49.

Alter G, Altfeld M. 2009. **NK cells in HIV-1 infection: evidence for their role in the control of HIV-1 infection.** J Intern Med., 265(1):29-42.

Alter G, Teigen N, Ahern R, Streeck H, Meier A, Rosenberg ES, Altfeld M. 2007. **Evolution of innate and adaptive effector cell functions during acute HIV-1 infection.** J Infect Dis., 195(10):1452-60.

Ariën KK, Vanham G, Arts EJ. 2007. **Is HIV-1 evolving to a less virulent form in humans?** Nat Rev Microbiol., 5(2):141-51.

Arduino JM, Fischl MA, Stanley K, Collier AC, Spiegelman D. 2001. **Do HIV type 1 RNA levels provide additional prognostic value to CD4(+) T lymphocyte counts in patients with advanced HIV type 1 infection?** AIDS Res Hum Retroviruses, 17(12):1099-105.

B

Bandawe GP, Martin DP, Treurnicht F, Mlisana K, Karim SS, Williamson C; CAPRISA 002 Acute Infection Study Team. 2008. **Conserved positive selection signals in gp41 across multiple subtypes and difference in selection signals detectable in gp41 sequences sampled during acute and chronic HIV-1 subtype C infection.** Virol J., 5:141.

Bansal A, Gough E, Sabbaj S, Ritter D, Yusim K, Sfakianos G, Aldrovandi G, Kaslow RA, Wilson CM, Mulligan MJ, Kilby JM, Goepfert PA. 2005. **CD8 T-cell responses in early HIV-1 infection are skewed towards high entropy peptides.** AIDS, 19(3):241-50.

Barouch DH. 2008. **Challenges in the development of an HIV-1 vaccine.** Nature, 455(7213):613-9.

Barouch DH, O'Brien KL, Simmons NL, King SL, Abbink P, Maxfield LF, Sun YH, La Porte A, Riggs AM, Lynch DM, Clark SL, Backus K, Perry JR, Seaman MS, Carville A, Mansfield KG, Szinger JJ, Fischer W, Muldoon M, Korber B. 2010. **Mosaic HIV-1 vaccines expand the breadth and depth of cellular immune responses in rhesus monkeys.** Nat Med., 16(3):319-23.

Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vézinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L. 1983. **Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).** Science, 220(4599):868-71.

Bibollet-Ruche F, Gao F, Bailes E, Saragosti S, Delaporte E, Peeters M, Shaw GM, Hahn BH, Sharp PM. 2004. **Complete genome analysis of one of the earliest SIVcpzPtt strains from Gabon (SIVcpzGAB2).** AIDS Res Hum Retroviruses., 20(12):1377-81.

Binley JM, Lybarger EA, Crooks ET, Seaman MS, Gray E, Davis KL, Decker JM, Wycuff D, Harris L, Hawkins N, Wood B, Nathe C, Richman D, Tomaras GD, Bibollet-Ruche F, Robinson JE, Morris L, Shaw GM, Montefiori DC, Mascola JR. 2008. **Profiling the specificity of neutralizing antibodies in a large panel of plasmas from patients chronically infected with human immunodeficiency virus type 1 subtypes B and C.** J Virol., 82(23):11651-68.

Binley JM, Wrin T, Korber B, Zwick MB, Wang M, Chappey C, Stiegler G, Kunert R, Zolla-Pazner S, Katinger H, Petropoulos CJ, Burton DR. 2004. **Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies.** J Virol., 78(23):13232-52.

Bhattacharya T, Daniels M, Heckerman D, Foley B, Frahm N, Kadie C, Carlson J, Yusim K, McMahon B, Gaschen B, Mallal S, Mullins JI, Nickle DC, Herbeck J, Rousseau C, Learn GH, Miura T, Brander C, Walker B, Korber B. 2007. **Founder effects in the assessment of HIV polymorphisms and HLA allele associations.** Science, 315(5818):1583-6.

Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, Nelson JA, Gairin JE, Hahn BH, Oldstone MB, Shaw GM. 1997. **Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus.** Nat Med., 3(2):205-11.

Borrow P, Bhardwaj N. 2008. **Innate immune responses in primary HIV-1 infection.** Curr Opin HIV AIDS, 3(1):36-44.

Boulet S, Kleyman M, Kim JY, Kamya P, Sharafi S, Simic N, Bruneau J, Routy JP, Tsoukas CM, Bernard NF. 2008. **A combined genotype of KIR3DL1 high expressing alleles and HLA-B*57 is associated with a reduced risk of HIV infection.** AIDS, 22(12):1487-91.

Boutwell CL, Rowley CF, Essex M. 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(6):2460-8.

Bredell H, Williamson C, Sonnenberg P, Martin DJ, Morris L. 1998. **Genetic characterization of HIV type 1 from migrant workers in three South African gold mines.** AIDS Res Hum Retroviruses, 14(8):677-84.

Bredell H, Hunt G, Casteling A, Cilliers T, Rademeyer C, Coetzer M, Miller S, Johnson D, Tiemessen CT, Martin DJ, Williamson C, Morris L. 2002. **HIV-1 Subtype A, D, G, AG and unclassified sequences identified in South Africa.** AIDS Res Hum Retroviruses, 18(9):681-3.

Bredell H, Martin DP, Van Harmelen J, Varsani A, Sheppard HW, Donovan R, Gray CM; HIVNET028 Study Team, Williamson C. 2007. **HIV type 1 subtype C gag and nef diversity in Southern Africa.** AIDS Res Hum Retroviruses, 23(3):477-81.

Brenchley JM, Hill BJ, Ambrozak DR, Price DA, Guenaga FJ, Casazza JP, Kuruppu J, Yazdani J, Migueles SA, Connors M, Roederer M, Douek DC, Koup RA. 2004. **T-cell subsets that harbor human immunodeficiency virus (HIV) in vivo: implications for HIV pathogenesis.** J Virol., 78(3):1160-8.

Briggs JA, Wilk T, Welker R, Kräusslich HG, Fuller SD. 2003. **Structural organization of authentic, mature HIV-1 virions and cores.** EMBO J., 22(7):1707-15.

Brockman MA, Schneidewind A, Lahaie M, Schmidt A, Miura T, Desouza I, Ryvkin F, Derdeyn CA, Allen S, Hunter E, Mulenga J, Goepfert PA, Walker BD, Allen TM. 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(22):12608-18.

Brumme ZL, Brumme CJ, Heckerman D, Korber BT, Daniels M, Carlson J, Kadie C, Bhattacharya T, Chui C, Szinger J, Mo T, Hogg RS, Montaner JS, Frahm N, Brander C, Walker BD, Harrigan PR. 2007. **Evidence of differential HLA class I-mediated viral evolution in functional and accessory/regulatory genes of HIV-1.** PLoS Pathog., 3(7):e94. Erratum in: PLoS Pathog. 2007 Aug 10;3(8):e121.

Brumme ZL, Tao I, Szeto S, Brumme CJ, Carlson JM, Chan D, Kadie C, Frahm N, Brander C, Walker B, Heckerman D, Harrigan PR. 2008. **Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection.** AIDS, 22(11):1277-86.

Buchbinder SP, Mehrotra DV, Duerr A, Fitzgerald DW, Mogg R, Li D, Gilbert PB, Lama JR, Marmor M, Del Rio C, McElrath MJ, Casimiro DR, Gottesdiener KM, Chodakewitz JA, Corey L, Robertson MN; 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(9653):1881-93.

Burgers WA, Riou C, Mlotshwa M, Maenetje P, de Assis Rosa D, Brenchley J, Mlisana K, Douek DC, Koup R, Roederer M, de Bruyn G, Karim SA, Williamson C, Gray CM; and the CAPRISA 002 Acute Infection Study Team. 2009. **Association of HIV-specific and total CD8+ T memory phenotypes in subtype C HIV-1 infection with viral set point.** J Immunol., 182(8):4751-61.

Burgers WA, Chege GK, Müller TL, van Harmelen JH, Khoury G, Shephard EG, Gray CM, Williamson C, Williamson AL. 2009. **Broad, high-magnitude and multifunctional CD4+ and CD8+ T-cell responses elicited by a DNA and modified vaccinia Ankara vaccine containing human immunodeficiency virus type 1 subtype C genes in baboons.** J Gen Virol., 90(Pt 2):468-80.

Burgers WA, Shephard E, Monroe JE, Greenhalgh T, Binder A, Hurter E, Van Harmelen JH, Williamson C, Williamson AL. 2008. **Construction, characterization, and immunogenicity of a multigene modified vaccinia Ankara (MVA) vaccine based on HIV type 1 subtype C.** AIDS Res Hum Retroviruses, 24(2):195-206.

Burgers WA, van Harmelen JH, Shephard E, Adams C, Mgwebi T, Bourn W, Hanke T, Williamson AL, Williamson C. 2006. **Design and preclinical evaluation of a multigene human**

immunodeficiency virus type 1 subtype C DNA vaccine for clinical trial. J Gen Virol., 87(Pt 2):399-410.

Burgers WA, Williamson C. 2005. **The challenges of HIV vaccine development and testing.** Best Pract Res Clin Obstet Gynaecol., 19(2):277-91.

Burton DR, Desrosiers RC, Doms RW, Koff WC, Kwong PD, Moore JP, Nabel GJ, Sodroski J, Wilson IA, Wyatt RT. 2004. **HIV vaccine design and the neutralizing antibody problem.** Nat Immunol., 5(3):233-6.

C

Cameron DW, Simonsen JN, D'Costa LJ, Ronald AR, Maitha GM, Gakinya MN, Cheang M, Ndinya-Achola JO, Piot P, Brunham RC, et al. 1989. **Female to male transmission of human immunodeficiency virus type 1: risk factors for seroconversion in men.** Lancet, 2(8660):403-7.

Catanzaro AT, Roederer M, Koup RA, Bailer RT, Enama ME, Nason MC, Martin JE, Rucker S, Andrews CA, Gomez PL, Mascola JR, Nabel GJ, Graham BS; VRC 007 Study Team. 2007. **Phase I clinical evaluation of a six-plasmid multiclade HIV-1 DNA candidate vaccine.** Vaccine, 25(20):4085-92. Epub 2007 Mar 7.

Catanzaro AT, Koup RA, Roederer M, Bailer RT, Enama ME, Moodie Z, Gu L, Martin JE, Novik L, Chakrabarti BK, Butman BT, Gall JG, King CR, Andrews CA, Sheets R, Gomez PL, Mascola JR, Nabel GJ, Graham BS; Vaccine Research Center 006 Study Team. 2006. **Phase 1 safety and immunogenicity evaluation of a multiclade HIV-1 candidate vaccine delivered by a replication-defective recombinant adenovirus vector.** J Infect Dis., 194(12):1638-49.

Chakrabarti BK, Ling X, Yang ZY, Montefiori DC, Panet A, Kong WP, Welcher B, Louder MK, Mascola JR, Nabel GJ. 2005. **Expanded breadth of virus neutralization after immunization with a multiclade envelope HIV vaccine candidate.** Vaccine, 23(26):3434-45.

Charpentier C, Nora T, Tenaillon O, Clavel F, Hance AJ. 2006. **Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients.** J Virol., 80(5):2472-82.

Chopera DR, Woodman Z, Mlisana K, Mlotshwa M, Martin DP, Seoighe C, Treurnicht F, de Rosa DA, Hide W, Karim SA, Gray CM, Williamson C; CAPRISA 002 Study Team. 2008. **Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage.** PLoS Pathog., 4(3):e1000033.

Crawford H, Prado JG, Leslie A, Hué S, Honeyborne I, Reddy S, van der Stok M, Mncube Z, Brander C, Rousseau C, Mullins JI, Kaslow R, Goepfert P, Allen S, Hunter E, Mulenga J, Kiepiela P, Walker BD, Goulder PJ. 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(15):8346-51.

Cocchi F, DeVico AL, Garzino-Demo A, Arya SK, Gallo RC, Lusso P. 1995. **Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells.** Science, 270(5243):1811-5.

Cohen MS, Hellmann N, Levy JA, DeCock K, Lange J. 2008. **The spread, treatment, and prevention of HIV-1: evolution of a global pandemic.** J Clin Invest., 118(4):1244-54. Review.

Couillin I, Culmann-Penciolelli B, Gomard E, Choppin J, Levy JP, Guillet JG, Saragosti S. 1994. **Impaired cytotoxic T lymphocyte recognition due to genetic variations in the main immunogenic region of the human immunodeficiency virus 1 NEF protein.** J Exp Med., 180(3):1129-34.

Courgnaud V, Salemi M, Pourrut X, Mpoudi-Ngole E, Abela B, Auzel P, Bibollet-Ruche F, Hahn B, Vandamme AM, Delaporte E, Peeters M. 2002. **Characterization of a novel simian immunodeficiency virus with a vpu gene from greater spot-nosed monkeys (Cercopithecus nictitans) provides new insights into simian/human immunodeficiency virus phylogeny.** J Virol., 76(16):8298-309.

Costello C, Tang J, Rivers C, Karita E, Meizen-Derr J, Allen S, Kaslow RA. 1999. **HLA-B*5703 independently associated with slower HIV-1 disease progression in Rwandan women.** AIDS, 13(14):1990-1.

D

Delport W, Scheffler K, Seoighe C. 2008. **Frequent toggling between alternative amino acids is driven by selection in HIV-1.** PLoS Pathog., 4(12):e1000242.

Delwart E, Magierowska M, Royz M, Foley B, Peddada L, Smith R, Heldebrant C, Conrad A, Busch M. 2002. **Homogeneous quasispecies in 16 out of 17 individuals during very early HIV-1 primary infection.** AIDS, 16(2):189-95.

Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di Marzio P, Marmon S, Sutton RE, Hill CM, Davis CB, Peiper SC, Schall TJ, Littman DR, Landau NR. 1996. **Identification of a major co-receptor for primary isolates of HIV-1.** Nature, 381(6584):661-6.

Derdeyn CA, Decker JM, Bibollet-Ruche F, Mokili JL, Muldoon M, Denham SA, Heil ML, Kasolo F, Musonda R, Hahn BH, Shaw GM, Korber BT, Allen S, Hunter E. 2004. **Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission.** Science, 303(5666):2019-22.

de Wolf F, Spijkerman I, Schellekens PT, Langendam M, Kuiken C, Bakker M, Roos M, Coutinho R, Miedema F, Goudsmit J. 1997. **AIDS prognosis based on HIV-1 RNA, CD4+ T-cell count and function: markers with reciprocal predictive value over time after seroconversion.** AIDS, 11(15):1799-806.

Dietrich U, Grez M, von Briesen H, Panhans B, Geissendörfer M, Kühnel H, Maniar J, Mahambre G, Becker WB, Becker ML, et al. 1993. **HIV-1 strains from India are highly divergent from prototypic African and US/European strains, but are linked to a South African isolate.** AIDS, 7(1):23-7.

Dorrington R, Bourne D. 2008. **Re-estimated provincial HIV antenatal survey prevalence for 2007 and a reinterpretation of the national trend.** S Afr Med J., 98(12):940-1.

Duerr A, Wasserheit JN, Corey L. 2006. **HIV vaccines: new frontiers in vaccine development.** Clin Infect Dis., 43(4):500-11.

E

Edmonson PF, Mullins JI. 1992. **Efficient amplification of HIV half-genomes from tissue DNA.** Nucleic Acids Res., 20(18):4933.

Edwards BH, Bansal A, Sabbaj S, Bakari J, Mulligan MJ, Goepfert PA. 2002. **Magnitude of functional CD8+ T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma.** J Virol., 76(5):2298-305.

Edwards RJ, Shields DC. 2004. **GASP: Gapped Ancestral Sequence Prediction for proteins.** BMC Bioinformatics, 5:123.

Edwards RJ, Shields DC. 2005. **BADASP: predicting functional specificity in protein families using ancestral sequences.** Bioinformatics, 21(22):4190-1.

Eggink D, Melchers M, Sanders RW. 2007. **Antibodies to HIV-1: aiming at the right target.** Trends Microbiol., 15(7):291-4.

Engelbrecht S, Laten JD, Smith TL, van Rensburg EJ. 1995. **Identification of env subtypes in fourteen HIV type 1 isolates from south Africa.** AIDS Res Hum Retroviruses, 11(10):1269-71.

Ewing B, Green P. 1998. **Base-calling of automated sequencer traces using phred. II. Error probabilities.** Genome Res., 8(3):186-94.

Ewing B, Hillier L, Wendl MC, Green P. 1998. **Base-calling of automated sequencer traces using phred. I. Accuracy assessment.** Genome Res., 8(3):175-85.

F

Fang G, Zhu G, Burger H, Keithly JS, Weiser B. 1998. **Minimizing DNA recombination during long RT-PCR.** J Virol Methods., 76(1-2):139-48.

Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, Cozzi-Lepri A, De Luca A, Easterbrook P, Francioli P, Mallal S, Martinez-Picado J, Miro JM, Obel N, Smith JP, Wyniger J, Descombes P, Antonarakis SE, Letvin NL,

McMichael AJ, Haynes BF, Telenti A, Goldstein DB. 2007. **A whole-genome association study of major determinants for host control of HIV-1.** Science, 317(5840):944-7.

Fernández-García A, Cuevas MT, Muñoz-Nieto M, Ocampo A, Pinilla M, García V, Serrano-Bengoechea E, Lezaun MJ, Delgado E, Thomson M, González-Galeano M, Contreras G, Nájera R, Pérez-Alvarez L. 2009. **Development of a panel of well-characterized human immunodeficiency virus type 1 isolates from newly diagnosed patients including acute and recent infections.** AIDS Res Hum Retroviruses, 25(1):93-102.

Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, Heldebrant C, Smith R, Conrad A, Kleinman SH, Busch MP. 2003. **Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection.** AIDS, 17(13):1871-9.

Fischer W, Perkins S, Theiler J, Bhattacharya T, Yusim K, Funkhouser R, Kuiken C, Haynes B, Letvin NL, Walker BD, Hahn BH, Korber BT. 2007. **Polyvalent vaccines for optimal coverage of potential T-cell epitopes in global HIV-1 variants.** Nat Med., 13(1):100-6.

Flynn NM, Forthal DN, Harro CD, Judson FN, Mayer KH, Para MF; rgp120 HIV Vaccine Study Group. 2005. **Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection.** J Infect Dis., 191(5):654-65.

Frahm N, Kaufmann DE, Yusim K, Muldoon M, Kesmir C, Linde CH, Fischer W, Allen TM, Li B, McMahon BH, Faircloth KL, Hewitt HS, Mackey EW, Miura T, Khatri A, Wolinsky S, McMichael A, Funkhouser RK, Walker BD, Brander C, Korber BT. 2007. **Increased sequence diversity coverage improves detection of HIV-specific T cell responses.** J Immunol., 179(10):6638-50.

Frahm N, Yusim K, Suscovich TJ, Adams S, Sidney J, Hraber P, Hewitt HS, Linde CH, Kavanagh DG, Woodberry T, Henry LM, Faircloth K, Listgarten J, Kadie C, Jojic N, Sango K, Brown NV, Pae E, Zaman MT, Bihl F, Khatri A, John M, Mallal S, Marincola FM, Walker BD, Sette A, Heckerman D, Korber BT, Brander C. 2007. **Extensive HLA class I allele promiscuity among viral CTL epitopes.** Eur J Immunol., 37(9):2419-33.

Frahm N, Nickle DC, Linde CH, Cohen DE, Zuñiga R, Lucchetti A, Roach T, Walker BD, Allen TM, Korber BT, Mullins JI, Brander C. 2008. **Increased detection of HIV-specific T cell responses by combination of central sequences with comparable immunogenicity.** AIDS, 22(4):447-56.

Frahm N, Kiepiela P, Adams S, Linde CH, Hewitt HS, Sango K, Feeney ME, Addo MM, Lichterfeld M, Lahaie MP, Pae E, Wurcel AG, Roach T, St John MA, Altfeld M, Marincola FM, Moore C, Mallal S, Carrington M, Heckerman D, Allen TM, Mullins JI, Korber BT, Goulder PJ, Walker BD, Brander C. 2006. **Control of human immunodeficiency virus replication by cytotoxic T lymphocytes targeting subdominant epitopes.** Nat Immunol., 7(2):173-8.

Frahm N, Adams S, Kiepiela P, Linde CH, Hewitt HS, Lichterfeld M, Sango K, Brown NV, Pae E, Wurcel AG, Altfeld M, Feeney ME, Allen TM, Roach T, St John MA, Daar ES, Rosenberg E, Korber B, Marincola F, Walker BD, Goulder PJ, Brander C. 2005. **HLA-B63 presents HLA-B57/B58-restricted cytotoxic T-lymphocyte epitopes and is associated with low human immunodeficiency virus load.** J Virol., 79(16):10218-25.

Frahm N, Korber BT, Adams CM, Szinger JJ, Draenert R, Addo MM, Feeney ME, Yusim K, Sango K, Brown NV, SenGupta D, Piechocka-Trocha A, Simonis T, Marincola FM, Wurcel AG, Stone DR, Russell CJ, Adolf P, Cohen D, Roach T, StJohn A, Khatri A, Davis K, Mullins J, Goulder PJ, Walker BD, Brander C. 2004. **Consistent cytotoxic-T-lymphocyte targeting of immunodominant regions in human immunodeficiency virus across multiple ethnicities.** J Virol., 78(5):2187-200.

Frater AJ, Brown H, Oxenius A, Günthard HF, Hirschel B, Robinson N, Leslie AJ, Payne R, Crawford H, Prendergast A, Brander C, Kiepiela P, Walker BD, Goulder PJ, McLean A, Phillips RE. 2007. **Effective T-cell responses select human immunodeficiency virus mutants and slow disease progression.** J Virol., 81(12):6742-51.

Friedrich TC, Dodds EJ, Yant LJ, Vojnov L, Rudersdorf R, Cullen C, Evans DT, Desrosiers RC, Mothé BR, Sidney J, Sette A, Kunstman K, Wolinsky S, Piatak M, Lifson J, Hughes AL, Wilson N, O'Connor DH, Watkins DI. 2004. **Reversion of CTL escape-variant immunodeficiency viruses in vivo.** Nat Med., 10(3):275-81..

Frøland SS, Jenum P, Lindboe CF, Wefring KW, Linnestad PJ, Böhmer T. 1988. **HIV-1 infection in Norwegian family before 1970.** Lancet, 1(8598):1344-5.

G

Gange SJ, Mellors JW, Lau B, Detels R, Phair JP, Muñoz A, Margolick JB. 2001. **Longitudinal patterns of HIV type 1 RNA among individuals with late disease progression.** AIDS Res Hum Retroviruses, 17(13):1223-9.

Garrido C, Roulet V, Chueca N, Poveda E, Aguilera A, Skrabal K, Zahonero N, Carlos S, García F, Faudon JL, Soriano V, de Mendoza C. 2008. **Evaluation of eight different bioinformatics tools to predict viral tropism in different human immunodeficiency virus type 1 subtypes.** J Clin Microbiol., 46(3):887-91.

Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH. 1999. **Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes.** Nature, 397(6718):436-41.

Gao F, Trask SA, Hui H, Mamaeva O, Chen Y, Theodore TS, Foley BT, Korber BT, Shaw GM, Hahn BH. 2001. **Molecular characterization of a highly divergent HIV type 1 isolate obtained early in the AIDS epidemic from the Democratic Republic of Congo.** AIDS Res Hum Retroviruses, 17(12):1217-22.

Gaschen B, Taylor J, Yusim K, Foley B, Gao F, Lang D, Novitsky V, Haynes B, Hahn BH, Bhattacharya T, Korber B. 2002. **Diversity considerations in HIV-1 vaccine selection.** Science, 296(5577):2354-60.

Geldmacher C, Currier JR, Herrmann E, Haule A, Kuta E, McCutchan F, Njovu L, Geis S, Hoffmann O, Maboko L, Williamson C, Birx D, Meyerhans A, Cox J, Hoelscher M. 2007. **CD8 T-cell recognition of multiple epitopes within specific Gag regions is associated with maintenance of a low steady-state viremia in human immunodeficiency virus type 1-seropositive patients.** J Virol., 81(5):2440-8.

Geldmacher C, Gray C, Nason M, Currier JR, Haule A, Njovu L, Geis S, Hoffmann O, Maboko L, Meyerhans A, Cox J, Hoelscher M. 2007. **A high viral burden predicts the loss of CD8 T-**

cell responses specific for subdominant gag epitopes during chronic human immunodeficiency virus infection. J Virol., 81(24):13809-15.

Gibbs, M. J., Armstrong, J. S., Gibbs, A. J., 2000. **Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences.** Bioinformatics 16, 573-582.

Giorgi JV, Lyles RH, Matud JL, Yamashita TE, Mellors JW, Hultin LE, Jamieson BD, Margolick JB, Rinaldo CR Jr, Phair JP, Detels R; Multicenter AIDS Cohort Study. 2002. **Predictive value of immunologic and virologic markers after long or short duration of HIV-1 infection.** J Acquir Immune Defic Syndr., 29(4):346-55.

Gray ES, Taylor N, Wycuff D, Moore PL, Tomaras GD, Wibmer CK, Puren A, Decamp A, Gilbert PB, Wood B, Montefiori DC, Binley JM, Shaw GM, Haynes BF, Mascola JR, Morris L. 2009. **Antibody specificities associated with neutralization breadth in plasma from HIV-1 subtype C infected blood donors.** J Virol., Jun 24.

Gray CM, Mlotshwa M, Riou C, Mathebula T, de Assis Rosa D, Mashishi T, Seoighe C, Ngandu N, van Loggerenberg F, Morris L, Mlisana K, Williamson C, Karim SA; CAPRISA 002 Acute Infection Study Team. 2009. **Human immunodeficiency virus-specific gamma interferon enzyme-linked immunospot assay responses targeting specific regions of the proteome during primary subtype C infection are poor predictors of the course of viremia and set point.** J Virol., 83(1):470-8.

Gray ES, Moore PL, Bibollet-Ruche F, Li H, Decker JM, Meyers T, Shaw GM, Morris L. 2008. **4E10-resistant variants in a human immunodeficiency virus type 1 subtype C-infected individual with an anti-membrane-proximal external region-neutralizing antibody response.** J Virol., 82(5):2367-75.

Gray ES, Moore PL, Pantophlet RA, Morris L. 2007. **N-linked glycan modifications in gp120 of human immunodeficiency virus type 1 subtype C render partial sensitivity to 2G12 antibody neutralization.** J Virol., 81(19):10769-76.

Gray ES, Moore PL, Choge IA, Decker JM, Bibollet-Ruche F, Li H, Leseka N, Treurnicht F, Mlisana K, Shaw GM, Karim SS, Williamson C, Morris L; CAPRISA 002 Study Team. 2007.

Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection. J Virol., 81(12):6187-96.

Gray ES, Meyers T, Gray G, Montefiori DC, Morris L. 2006. **Insensitivity of paediatric HIV-1 subtype C viruses to broadly neutralising monoclonal antibodies raised against subtype B.** PLoS Med., 3(7):e255.

Goepfert PA, Lumm W, Farmer P, Matthews P, Prendergast A, Carlson JM, Derdeyn CA, Tang J, Kaslow RA, Bansal A, Yusim K, Heckerman D, Mulenga J, Allen S, Goulder PJ, Hunter E. 2008. **Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients.** J Exp Med., 205(5):1009-17.

Gordon D, Abajian C, Green P. 1998. **Consed: a graphical tool for sequence finishing.** Genome Res., 8(3):195-202.

Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, Saxon A. 1981. **Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency.** N Engl J Med., 10; 305(24):1425-31.

Gottlieb GS, Nickle DC, Jensen MA, Wong KG, Grobler J, Li F, Liu SL, Rademeyer C, Learn GH, Karim SS, Williamson C, Corey L, Margolick JB, Mullins JI. 2004. **Dual HIV-1 infection associated with rapid disease progression.** Lancet, 363(9409):619-22. Erratum in: Lancet; 363(9415):1166.

Goodenow M, Huet T, Saurin W, Kwok S, Sninsky J, Wain-Hobson S. 1989. **HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions.** J Acquir Immune Defic Syndr., 2(4):344-52.

Goonetilleke N, Liu MK, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Ganusov VV, Keele BF, Learn GH, Turnbull EL, Salazar MG, Weinhold KJ, Moore S; CHAVI Clinical Core B, Letvin N, Haynes BF, Cohen MS, Hraber P, Bhattacharya T, Borrow P, Perelson AS, Hahn BH, Shaw GM, Korber BT, McMichael AJ. 2009. **The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection.** J Exp Med., 206(6):1253-72.

Gordon M, De Oliveira T, Bishop K, Coovadia HM, Madurai L, Engelbrecht S, Janse van Rensburg E, Mosam A, Smith A, Cassol S. 2003. **Molecular characteristics of human immunodeficiency virus type 1 subtype C viruses from KwaZulu-Natal, South Africa: implications for vaccine and antiretroviral control strategies.** J Virol.; 77(4):2587-99.

Goulder PJ, Watkins DI. 2004. **HIV and SIV CTL escape: implications for vaccine design.** Nat Rev Immunol., 4(8):630-40.

Goulder PJ, Sewell AK, Lalloo DG, Price DA, Whelan JA, Evans J, Taylor GP, Luzzi G, Giangrande P, Phillips RE, McMichael AJ. 1997. **Patterns of immunodominance in HIV-1-specific cytotoxic T lymphocyte responses in two human histocompatibility leukocyte antigens (HLA)-identical siblings with HLA-A*0201 are influenced by epitope mutation.** J Exp Med., 185(8):1423-33.

Grobler J, Gray CM, Rademeyer C, Seoighe C, Ramjee G, Karim SA, Morris L, Williamson C. 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(7):1355-9.

Groopman JE, Salahuddin SZ, Sarngadharan MG, Markham PD, Gonda M, Sliski A, Gallo RC. 1984. **HTLV-III in saliva of people with AIDS-related complex and healthy homosexual men at risk for AIDS.** Science, 226(4673):447-9.

Guadalupe M, Reay E, Sankaran S, Prindiville T, Flamm J, McNeil A, Dandekar S. 2003. **Severe CD4+ T-cell depletion in gut lymphoid tissue during primary human immunodeficiency virus type 1 infection and substantial delay in restoration following highly active antiretroviral therapy.** J Virol., 77(21):11708-17.

H

Haaland RE, Hawkins PA, Salazar-Gonzalez J, Johnson A, Tichacek A, Karita E, Manigart O, Mulenga J, Keele BF, Shaw GM, Hahn BH, Allen SA, Derdeyn CA, Hunter E. 2009. **Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1.** PLoS Pathog., 5(1):e1000274.

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

Hardie RA, Knight E, Bruneau B, Semeniuk C, Gill K, Nagelkerke N, Kimani J, Wachihi C, Ngugi E, Luo M, Plummer FA. 2008. **A common human leucocyte antigen-DP genotype is associated with resistance to HIV-1 infection in Kenyan sex workers.** AIDS, 22(15):2038-42.

Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH. 2003. **DNA deamination mediates innate immunity to retroviral infection.** Cell, 113(6):803-9. Erratum in: Cell. 2004 Feb 20; 116(4):629.

Haseltine WA. 1991. **Molecular biology of the human immunodeficiency virus type 1.** FASEB J., 5(10):2349-60.

Haynes BF, Ma B, Montefiori DC, Wrin T, Petropoulos CJ, Sutherland LL, Scearce RM, Denton C, Xia SM, Korber BT, Liao HX. 2006. **Analysis of HIV-1 subtype B third variable region peptide motifs for induction of neutralizing antibodies against HIV-1 primary isolates.** Virology, 345(1):44-55.

Haynes BF, Shattock RJ. 2008. **Critical issues in mucosal immunity for HIV-1 vaccine development.** J Allergy Clin Immunol., 122(1):3-9.

Hemelaar J, Gouws E, Ghys PD, Osmanov S. 2006. **Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004.** AIDS, 20(16):W13-23.

Herbeck JT, Nickle DC, Learn GH, Gottlieb GS, Curlin ME, Heath L, Mullins JI. 2006. **Human immunodeficiency virus type 1 env evolves toward ancestral states upon transmission to a new host.** J Virol., 80(4):1637-44.

Hu DJ, Subbarao S, Vanichseni S, Mock PA, Ramos A, Nguyen L, Chaowanachan T, Griensven F, Choopanya K, Mastro TD, Tappero JW. 2005. **Frequency of HIV-1 dual subtype infections, including intersubtype superinfections, among injection drug users in Bangkok, Thailand.** AIDS, 19(3):303-8.

Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, He T, Kang S, Ceradini D, Jin Z, Yazdanbakhsh K, Kunstman K, Erickson D, Dragon E, Landau NR, Phair J, Ho DD, Koup RA. 1996. **The role of a mutant CCR5 allele in HIV-1 transmission and disease progression.** Nat Med., 2(11):1240-3.

Hutnick NA, Carnathan DG, Dubey SA, Cox KS, Kierstead L, Ratcliffe SJ, Robertson MN, Casimiro DR, Ertl HC, Betts MR. 2009. **Baseline Ad5 serostatus does not predict Ad5 HIV vaccine-induced expansion of adenovirus-specific CD4+ T cells.** Nat Med., 15(8):876-8.

I

Ioannidis JP, Goedert JJ, McQueen PG, Enger C, Kaslow RA. 1999. **Comparison of viral load and human leukocyte antigen statistical and neural network predictive models for the rate of HIV-1 disease progression across two cohorts of homosexual men.** J Acquir Immune Defic Syndr Hum Retrovirol., 20(2):129-36.

Iversen AK, Stewart-Jones G, Learn GH, Christie N, Sylvester-Hviid C, Armitage AE, Kaul R, Beattie T, Lee JK, Li Y, Chotiyarnwong P, Dong T, Xu X, Luscher MA, MacDonald K, Ullum H, Klarlund-Pedersen B, Skinhøj P, Fugger L, Buus S, Mullins JI, Jones EY, van der Merwe PA, McMichael AJ. 2006. **Conflicting selective forces affect T cell receptor contacts in an immunodominant human immunodeficiency virus epitope.** Nat Immunol., 7(2):179-89.

J

Jacobs GB, de Beer C, Fincham JE, Adams V, Dhansay MA, van Rensburg EJ, Engelbrecht S. 2006. **Serotyping and genotyping of HIV-1 infection in residents of Khayelitsha, Cape Town, South Africa.** J Med Virol., 78(12):1529-36.

Jamieson BD, Yang OO, Hultin L, Hausner MA, Hultin P, Matud J, Kunstman K, Killian S, Altman J, Kommander K, Korber B, Giorgi J, Wolinsky S. 2003. **Epitope escape mutation and decay of human immunodeficiency virus type 1-specific CTL responses.** J Immunol., 171(10):5372-9.

Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, He HX, McLaughlin S, Shankarappa R, Margolick JB, Mullins JI. 2003. **Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences.** J Virol., 77(24):13376-88.

Jern P, Russell RA, Pathak VK, Coffin JM. 2009. **Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance.** PLoS Pathog., 5(4):e1000367.

Jin X, Bauer DE, Tuttleton SE, Lewin S, Gettie A, Blanchard J, Irwin CE, Safrit JT, Mittler J, Weinberger L, Kostrikis LG, Zhang L, Perelson AS, Ho DD. 1999. **Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques.** J Exp Med., 189(6):991-8.

K

Karlsson AC, Iversen AK, Chapman JM, de Oliviera T, Spotts G, McMichael AJ, Davenport MP, Hecht FM, Nixon DF. 2007. **Sequential broadening of CTL responses in early HIV-1 infection is associated with viral escape.** PLoS One, 2(2):e225.

Kampinga GA, Simonon A, Van de Perre P, Karita E, Msellati P, Goudsmit J. 1997. **Primary infections with HIV-1 of women and their offspring in Rwanda: findings of heterogeneity at seroconversion, coinfection, and recombinants of HIV-1 subtypes A and C.** Virology, 227(1):63-76.

Kaur G, Mehra N. 2009. **Genetic determinants of HIV-1 infection and progression to AIDS: susceptibility to HIV infection.** Tissue Antigens, 73(4):289-301.

Katzenstein D. 2006. **Diversity, drug resistance, and the epidemic of subtype C HIV-1 in Africa.** J Infect Dis., 194 Suppl 1:S45-50.

Kearney M, Maldarelli F, Shao W, Margolick JB, Daar ES, Mellors JW, Rao V, Coffin JM, Palmer S. 2009. **Human immunodeficiency virus type 1 population genetics and adaptation in newly infected individuals.** J Virol., 83(6):2715-27.

Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping LH, Swanstrom R, Tomaras GD, Blattner WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH, Shaw GM. 2008. **Identification and**

characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A., 105(21):7552-7.

Keele BF, Li H, Learn GH, Hraber P, Giorgi EE, Grayson T, Sun C, Chen Y, Yeh WW, Letvin NL, Mascola JR, Nabel GJ, Haynes BF, Bhattacharya T, Perelson AS, Korber BT, Hahn BH, Shaw GM. 2009. **Low-dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1.** J Exp Med., 206(5):1117-34.

Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, Bibollet-Ruche F, Chen Y, Wain LV, Liegeois F, Loul S, Ngole EM, Bienvenue Y, Delaporte E, Brookfield JF, Sharp PM, Shaw GM, Peeters M, Hahn BH. 2006. **Chimpanzee reservoirs of pandemic and nonpandemic HIV-1.** Science, 313(5786):523-6.

Kelleher AD, Long C, Holmes EC, Allen RL, Wilson J, Conlon C, Workman C, Shaunak S, Olson K, Goulder P, Brander C, Ogg G, Sullivan JS, Dyer W, Jones I, McMichael AJ, Rowland-Jones S, Phillips RE. 2001. **Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses.** J Exp Med., 193(3):375-86.

Kemal KS, Beattie T, Dong T, Weiser B, Kaul R, Kuiken C, Sutton J, Lang D, Yang H, Peng YC, Collman R, Philpott S, Rowland-Jones S, Burger H. 2008. **Transition from long-term nonprogression to HIV-1 disease associated with escape from cellular immune control.** J Acquir Immune Defic Syndr., 48(2):119-26.

Kelley CF, Barbour JD, Hecht FM. 2007. **The relation between symptoms, viral load, and viral load set point in primary HIV infection.** J Acquir Immune Defic Syndr., 45(4):445-8.

Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferott KJ, Hilton L, Zimbwa P, Moore S, Allen T, Brander C, Addo MM, Altfeld M, James I, Mallal S, Bunce M, Barber LD, Szinger J, Day C, Klenerman P, Mullins J, Korber B, Coovadia HM, Walker BD, Goulder PJ. 2004. **Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA.** Nature, 432(7018):769-75.

Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, Reddy S, de Pierres C, Mncube Z, Mkhwanazi N, Bishop K, van der Stok M, Nair K, Khan N, Crawford H, Payne R, Leslie A, Prado J, Prendergast A, Frater J, McCarthy N, Brander C, Learn GH, Nickle D, Rousseau C, Coovadia H, Mullins JI, Heckerman D, Walker BD, Goulder P. 2007. **CD8+ T-cell responses to different HIV proteins have discordant associations with viral load.** Nat Med., 13(1):46-53.

Kitrinos KM, Hoffman NG, Nelson JA, Swanstrom R. 2003. **Turnover of env variable region 1 and 2 genotypes in subjects with late-stage human immunodeficiency virus type 1 infection.** J Virol., 77(12):6811-22.

Kong WP, Wu L, Wallstrom TC, Fischer W, Yang ZY, Ko SY, Letvin NL, Haynes BF, Hahn BH, Korber B, Nabel GJ. 2009. **Expanded breadth of the T-cell response to mosaic human immunodeficiency virus type 1 envelope DNA vaccination.** J Virol., 83(5):2201-15.

Kong X, West JT, Zhang H, Shea DM, M'soka TJ, Wood C. 2008. **The human immunodeficiency virus type 1 envelope confers higher rates of replicative fitness to perinatally transmitted viruses than to nontransmitted viruses.** J Virol., 82(23):11609-18.

Kostrikis LG, Touloumi G, Karanicolas R, Pantazis N, Anastassopoulou C, Karafoulidou A, Goedert JJ, Hatzakis A; Multicenter Hemophilia Cohort Study Group. 2002. **Quantitation of human immunodeficiency virus type 1 DNA forms with the second template switch in peripheral blood cells predicts disease progression independently of plasma RNA load.** J Virol., 76(20):10099-108.

Korber B & Gnanakaran S. 2009. **The implications of patterns in HIV diversity for neutralizing antibody induction and susceptibility.** Current Opinion in HIV and AIDS, 4.

Korber BT, Letvin NL, Haynes BF. 2009. **T-cell vaccine strategies for human immunodeficiency virus, the virus with a thousand faces.** J Virol., 83(17):8300-14.

Kothe DL, Decker JM, Li Y, Weng Z, Bibollet-Ruche F, Zammit KP, Salazar MG, Chen Y, Salazar-Gonzalez JF, Moldoveanu Z, Mestecky J, Gao F, Haynes BF, Shaw GM, Muldoon M,

Korber BT, Hahn BH. 2007. **Antigenicity and immunogenicity of HIV-1 consensus subtype B envelope glycoproteins.** Virology, 360(1):218-34.

Kothe DL, Li Y, Decker JM, Bibollet-Ruche F, Zammit KP, Salazar MG, Chen Y, Weng Z, Weaver EA, Gao F, Haynes BF, Shaw GM, Korber BT, Hahn BH. 2006. **Ancestral and consensus envelope immunogens for HIV-1 subtype C.** Virology, 352(2):438-49.

Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, Farthing C, Ho DD. 1994. **Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome.** J Virol., 68(7):4650-5.

Kozak SL, Platt EJ, Madani N, Ferro FE Jr, Peden K, Kabat D. 1997. **CD4, CXCR-4, and CCR-5 dependencies for infections by primary patient and laboratory-adapted isolates of human immunodeficiency virus type 1.** J Virol., 71(2):873-82.

Kumar S, Tamura K, Nei M. 2004. **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** Brief Bioinform., 5(2):150-63.

L

Lal RB, Chakrabarti S, Yang C. 2005. **Impact of genetic diversity of HIV-1 on diagnosis, antiretroviral therapy & vaccine development.** Indian J Med Res., 121(4):287-314.

Lambotte O, Ferrari G, Moog C, Yates NL, Liao HX, Parks RJ, Hicks CB, Owzar K, Tomaras GD, Montefiori DC, Haynes BF, Delfraissy JF. 2009. **Heterogeneous neutralizing antibody and antibody-dependent cell cytotoxicity responses in HIV-1 elite controllers.** AIDS, 23(8):897-906.

Land AM, Ball TB, Luo M, Pilon R, Sandstrom P, Embree JE, Wachihi C, Kimani J, Plummer FA. 2008. **Human immunodeficiency virus (HIV) type 1 proviral hypermutation correlates with CD4 count in HIV-infected women from Kenya.** J Virol., 82(16):8172-82.

Lasaro MO, Ertl HC. 2009. **New insights on adenovirus as vaccine vectors.** Mol Ther., 17(8):1333-9.

Lavreys L, Baeten JM, Chohan V, McClelland RS, Hassan WM, Richardson BA, Mandaliya K, Ndinya-Achola JO, Overbaugh J. 2006. **Higher set point plasma viral load and more-severe acute HIV type 1 (HIV-1) illness predict mortality among high-risk HIV-1-infected African women.** Clin Infect Dis., 42(9):1333-9.

Lehner T, Wang Y, Pido-Lopez J, Whittall T, Bergmeier LA, Babaahmady K. 2008. **The emerging role of innate immunity in protection against HIV-1 infection.** Vaccine, 26(24):2997-3001.

Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, Feeney M, Tang Y, Holmes EC, Allen T, Prado JG, Altfeld M, Brander C, Dixon C, Ramduth D, Jeena P, Thomas SA, St John A, Roach TA, Kupfer B, Luzzi G, Edwards A, Taylor G, Lyall H, Tudor-Williams G, Novelli V, Martinez-Picado J, Kiepiela P, Walker BD, Goulder PJ. 2004. **HIV evolution: CTL escape mutation and reversion after transmission.** Nat Med., 10(3):282-9.

Levy DN, Aldrovandi GM, Kutsch O, Shaw GM. 2004. **Dynamics of HIV-1 recombination in its natural target cells.** Proc Natl Acad Sci U S A., 101(12):4204-9. Erratum in: Proc Natl Acad Sci U S A. 2005 Feb 1; 102(5):1808.

Li B, Gladden AD, Altfeld M, Kaldor JM, Cooper DA, Kelleher AD, Allen TM. 2007. **Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution.** J Virol., 81(1):193-201.

Li B, Decker JM, Johnson RW, Bibollet-Ruche F, Wei X, Mulenga J, Allen S, Hunter E, Hahn BH, Shaw GM, Blackwell JL, Derdeyn CA. 2006. **Evidence for potent autologous neutralizing antibody titers and compact envelopes in early infection with subtype C human immunodeficiency virus type 1.** J Virol., 80(11):5211-8. Erratum in: J Virol. 2007 Nov; 81(22):12716.

Li J, Chen X, Jiang S, Chen YH. 2008. **Deletion of fusion peptide or destabilization of fusion core of HIV gp41 enhances antigenicity and immunogenicity of 4E10 epitope.** Biochem Biophys Res Commun., 376(1):60-4.

Liao HX, Sutherland LL, Xia SM, Brock ME, Scearce RM, Vanleeuwen S, Alam SM, McAdams M, Weaver EA, Camacho Z, Ma BJ, Li Y, Decker JM, Nabel GJ, Montefiori DC, Hahn BH, Korber BT, Gao F, Haynes BF. 2006. **A group M consensus envelope glycoprotein induces antibodies that neutralize subsets of subtype B and C HIV-1 primary viruses.** Virology, 353(2):268-82.

Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, MacDonald ME, Stuhlmann H, Koup RA, Landau NR. 1996. **Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection.** Cell, 86(3):367-77.

Liu Y, McNevin J, Cao J, Zhao H, Genowati I, Wong K, McLaughlin S, McSweyn MD, Diem K, Stevens CE, Maenza J, He H, Nickle DC, Shriner D, Holte SE, Collier AC, Corey L, McElrath MJ, Mullins JI. 2006. **Selection on the human immunodeficiency virus type 1 proteome following primary infection.** J Virol., 80(19):9519-29.

Liu Y, McNevin J, Zhao H, Tebit DM, Troyer RM, McSweyn M, Ghosh AK, Shriner D, Arts EJ, McElrath MJ, Mullins JI. 2007. **Evolution of human immunodeficiency virus type 1 cytotoxic T-lymphocyte epitopes: fitness-balanced escape.** J Virol., 81(22):12179-88.

Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. 1999. **Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination.** J Virol., 73(1):152-60.

Loxton AG, Treurnicht F, Laten A, van Rensburg EJ, Engelbrecht S. 2005. **Sequence analysis of near full-length HIV type 1 subtype D primary strains isolated in Cape Town, South Africa, from 1984 to 1986.** AIDS Res Hum Retroviruses, 21(5):410-3.

Lucotte G. 2002. **Frequencies of 32 base pair deletion of the (Delta 32) allele of the CCR5 HIV-1 co-receptor gene in Caucasians: a comparative analysis.** Infect Genet Evol., 1(3):201-5.

Lyles RH, Muñoz A, Yamashita TE, Bazmi H, Detels R, Rinaldo CR, Margolick JB, Phair JP, Mellors JW. 2000. **Natural history of human immunodeficiency virus type 1 viremia after**

seroconversion and proximal to AIDS in a large cohort of homosexual men. Multicenter AIDS Cohort Study. J Infect Dis., 181(3):872-80.

Lyles CM, Graham NM, Astemborski J, Vlahov D, Margolick JB, Saah AJ, Farzadegan H. 1999. **Cell-associated infectious HIV-1 viral load as a predictor of clinical progression and survival among HIV-1 infected injection drug users and homosexual men.** Eur J Epidemiol., 15(2):99-108.

M

Mahalanabis M, Jayaraman P, Miura T, Pereyra F, Chester EM, Richardson B, Walker B, Haigwood NL. 2009. **Continuous viral escape and selection by autologous neutralizing antibodies in drug-naive human immunodeficiency virus controllers.** J Virol., 83(2):662-72.

Mansky LM, Temin HM. 1995. **Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase.** J Virol., 69(8):5087-94.

Markham RB, Wang WC, Weisstein AE, Wang Z, Munoz A, Templeton A, Margolick J, Vlahov D, Quinn T, Farzadegan H, Yu XF. 1998. **Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline.** Proc Natl Acad Sci U S A., 95(21):12568-73.

Markowitz M, Louie M, Hurley A, Sun E, Di Mascio M, Perelson AS, Ho DD. 2003. **A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and T-cell decay in vivo.** J Virol., 77(8):5037-8.

Martin D, Rybicki E. 2000. **RDP: detection of recombination amongst aligned sequences.** Bioinformatics, 16(6):562-3.

Martin DP, Williamson C, Posada D. 2005[a]. **RDP2: recombination detection and analysis from sequence alignments.** Bioinformatics, 21(2):260-2.

Martin DP, Posada D, Crandall KA, Williamson C. 2005[b]. **A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints.** AIDS Res Hum Retroviruses, 21(1):98-102.

Martinez-Picado J, Prado JG, Fry EE, Pfafferott K, Leslie A, Chetty S, Thobakgale C, Honeyborne I, Crawford H, Matthews P, Pillay T, Rousseau C, Mullins JI, Brander C, Walker BD, Stuart DI, Kiepiela P, Goulder P. 2006. **Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1.** J Virol., 80(7):3617-23.

Mascola JR, Snyder SW, Weislow OS, Belay SM, Belshe RB, Schwartz DH, Clements ML, Dolin R, Graham BS, Gorse GJ, Keefer MC, McElrath MJ, Walker MC, Wagner KF, McNeil JG, McCutchan FE, Burke DS. 1996. **Immunization with envelope subunit vaccine products elicits neutralizing antibodies against laboratory-adapted but not primary isolates of human immunodeficiency virus type 1. The National Institute of Allergy and Infectious Diseases AIDS Vaccine Evaluation Group.** J Infect Dis., 173(2):340-8.

Masemola A, Mashishi T, Khoury G, Mohube P, Mokgotho P, Vardas E, Colvin M, Zijenah L, Katzenstein D, Musonda R, Allen S, Kumwenda N, Taha T, Gray G, McIntyre J, Karim SA, Sheppard HW, Gray CM; HIVNET 028 Study Team. 2004. **Hierarchical targeting of subtype C human immunodeficiency virus type 1 proteins by CD8+ T cells: correlation with viral load.** J Virol., 78(7):3233-43.

Masemola AM, Mashishi TN, Khoury G, Bredell H, Paximadis M, Mathebula T, Barkhan D, Puren A, Vardas E, Colvin M, Zijenah L, Katzenstein D, Musonda R, Allen S, Kumwenda N, Taha T, Gray G, McIntyre J, Karim SA, Sheppard HW, Gray CM; HIVNET 028 Study Team. 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.** J Immunol., 173(7):4607-17.

Massanella M, Puigdomènech I, Cabrera C, Fernandez-Figueras MT, Aucher A, Gaibelet G, Hudrisier D, García E, Bofill M, Clotet B, Blanco J. 2009. **Antigp41 antibodies fail to block early events of virological synapses but inhibit HIV spread between T cells.** AIDS, 23(2):183-8.

Masur H, Michelis MA, Wormser GP, Lewin S, Gold J, Tapper ML, Giron J, Lerner CW, Armstrong D, Setia U, Sender JA, Siebken RS, Nicholas P, Arlen Z, Maayan S, Ernst JA, Siegal FP, Cunningham-Rundles S. 1982. **Opportunistic infection in previously healthy women. Initial manifestations of a community-acquired cellular immunodeficiency.** Ann Intern Med., 97(4):533-9.

Matthews PC, Prendergast A, Leslie A, Crawford H, Payne R, Rousseau C, Rolland M, Honeyborne I, Carlson J, Kadie C, Brander C, Bishop K, Mlotshwa N, Mullins JI, Coovadia H, Ndung'u T, Walker BD, Heckerman D, Goulder PJ. 2008. **Central role of reverting mutations in HLA associations with human immunodeficiency virus set point.** J Virol., 82(17):8548-59.

McCutchan FE. 2006. **Global epidemiology of HIV.** J Med Virol., 78 Suppl 1:S7-S12.

McElrath MJ, De Rosa SC, Moodie Z, Dubey S, Kierstead L, Janes H, Defawe OD, Carter DK, Hural J, Akondy R, Buchbinder SP, Robertson MN, Mehrotra DV, Self SG, Corey L, Shiver JW, Casimiro DR; Step Study Protocol Team. 2008. **HIV-1 vaccine-induced immunity in the test-of-concept Step Study: a case-cohort analysis.** Lancet, 372(9653):1894-905.

McNearney T, Hornickova Z, Markham R, Birdwell A, Arens M, Saah A, Ratner L. 1992. **Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease.** Proc Natl Acad Sci U S A., 89(21):10247-51.

Mehandru S, Poles MA, Tenner-Racz K, Horowitz A, Hurley A, Hogan C, Boden D, Racz P, Markowitz M. 2004. **Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract**. J Exp Med., 200(6):761-70.

Mehandru S, Poles MA, Tenner-Racz K, Manuelli V, Jean-Pierre P, Lopez P, Shet A, Low A, Mohri H, Boden D, Racz P, Markowitz M. 2007. **Mechanisms of gastrointestinal CD4+ T-cell depletion during acute and early human immunodeficiency virus type 1 infection.** J Virol., 81(2):599-612.

Mellors JW, Muñoz A, Giorgi JV, Margolick JB, Tassoni CJ, Gupta P, Kingsley LA, Todd JA, Saah AJ, Detels R, Phair JP, Rinaldo CR Jr. 1997. **Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection.** Ann Intern Med., 126(12):946-54.

Miller CJ, Li Q, Abel K, Kim EY, Ma ZM, Wietgrefe S, La Franco-Scheuch L, Compton L, Duan L, Shore MD, Zupancic M, Busch M, Carlis J, Wolinsky S, Haase AT. 2005. **Propagation and dissemination of infection after vaginal transmission of simian immunodeficiency virus.** J Virol., 79(14):9217-27. Erratum in: J Virol. 2005 Sep; 79(17)11552.

Miura T, Brockman MA, Schneidewind A, Lobritz M, Pereyra F, Rathod A, Block BL, Brumme ZL, Brumme CJ, Baker B, Rothchild AC, Li B, Trocha A, Cutrell E, Frahm N, Brander C, Toth I, Arts EJ, Allen TM, Walker BD. 2009[a]. **HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphocyte [corrected] recognition.** J Virol., 83(6):2743-55. Erratum in: J Virol. 2009 Jun; 83(11):5961.

Miura T, Brumme CJ, Brockman MA, Brumme ZL, Pereyra F, Block BL, Trocha A, John M, Mallal S, Harrigan PR, Walker BD. 2009[b]. **HLA-associated viral mutations are common in human immunodeficiency virus type 1 elite controllers.** J Virol., 83(7):3407-12.

[c]Miura T, Brockman MA, Brumme ZL, Brumme CJ, Pereyra F, Trocha A, Block BL, Schneidewind A, Allen TM, Heckerman D, Walker BD. 2009. **HLA-associated alterations in replication capacity of chimeric NL4-3 viruses carrying gag-protease from elite controllers of human immunodeficiency virus type 1.** J Virol., 83(1):140-9.

Montefiori D, Sattentau Q, Flores J, Esparza J, Mascola J; Working Group convened by the Global HIV Vaccine Enterprise. 2007. **Antibody-based HIV-1 vaccines: recent developments and future directions.** PLoS Med., 4(12):e348.

Moore JP, Burton DR. 2004. **Urgently needed: a filter for the HIV-1 vaccine pipeline.** Nat Med., 10(8):769-71.

Moore PL, Gray ES, Choge IA, Ranchobe N, Mlisana K, Abdool Karim SS, Williamson C, Morris L; CAPRISA 002 Study Team. 2008. **The c3-v4 region is a major target of autologous neutralizing antibodies in human immunodeficiency virus type 1 subtype C infection.** J Virol., 82(4):1860-9.

Mortality and Morbidity Weekly Report (MMWR). 1981. ***Pneumocystis* pneumonia - Los Angeles.** MMWR, 30: 250 – 252.

Mortality and Morbidity Weekly Report (MMWR). 1981. **Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men - New York City and California.** MMWR, 31: 305 – 308.

Ngandu NG, Bredell H, Gray CM, Williamson C, Seoighe C; HIVNET028 Study Team. 2007. **CTL response to HIV type 1 subtype C is poorly predicted by known epitope motifs.** AIDS Res Hum Retroviruses, 23(8):1033-41.

Ngumbela KC, Day CL, Mncube Z, Nair K, Ramduth D, Thobakgale C, Moodley E, Reddy S, de Pierres C, Mkhwanazi N, Bishop K, van der Stok M, Ismail N, Honeyborne I, Crawford H, Kavanagh DG, Rousseau C, Nickle D, Mullins J, Heckerman D, Korber B, Coovadia H, Kiepiela P, Goulder PJ, Walker BD. 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(1):72-82.

Nickle DC, Rolland M, Jensen MA, Pond SL, Deng W, Seligman M, Heckerman D, Mullins JI, Jojic N. 2007. **Coping with viral diversity in HIV vaccine design.** PLoS Comput Biol., 3(4):e75.

Nickle DC, Jojic N, Heckerman D, Jojic V, Kirovski D, Rolland M, Kosakovsky Pond S, Mullins JI. 2008. **Comparison of immunogen designs that optimize peptide coverage: reply to Fischer et al.** PLoS Comput Biol., 4(1):e25. Erratum in: PLoS Comput Biol. 2008 Mar;4(3). doi: 10.1371/annotation/

Nickle DC, Rolland M, Jensen MA, Pond SL, Deng W, Seligman M, Heckerman D, Mullins JI, Jojic N. 2007. **Coping with viral diversity in HIV vaccine design.** PLoS Comput Biol., 3(4):e75.

Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, Røder G, Peters B, Sette A, Lund O, Buus S. 2007. **NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence.** PLoS One, 2(8):e796.

Nowak MA, May RM, Anderson RM. 1990. **The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease.** AIDS, 4(11):1095-103.

Novitsky V, Smith UR, Gilbert P, McLane MF, Chigwedere P, Williamson C, Ndung'u T, Klein I, Chang SY, Peter T, Thior I, Foley BT, Gaolekwe S, Rybak N, Gaseitsiwe S, Vannberg F, Marlink R, Lee TH, Essex M. 2002. **Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design?** J Virol., 76(11):5435-51.

O

O'Brien KL, Liu J, King SL, Sun YH, Schmitz JE, Lifton MA, Hutnick NA, Betts MR, Dubey SA, Goudsmit J, Shiver JW, Robertson MN, Casimiro DR, Barouch DH. 2009. **Adenovirus-specific immunity after immunization with an Ad5 HIV-1 vaccine candidate in humans.** Nat Med., 15(8):873-5.

P

Papathanasopoulos MA, Cilliers T, Morris L, Mokili JL, Dowling W, Birx DL, McCutchan FE. 2002. **Full-length genome analysis of HIV-1 subtype C utilizing CXCR4 and intersubtype recombinants isolated in South Africa.** AIDS Res Hum Retroviruses, 18(12):879-86.

Peeters M, Courgnaud V, Abela B, Auzel P, Pourrut X, Bibollet-Ruche F, Loul S, Liegeois F, Butel C, Koulagna D, Mpoudi-Ngole E, Shaw GM, Hahn BH, Delaporte E. 2002. **Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat.** Emerg Infect Dis., 8(5):451-7.

Pitisuttithum P, Gilbert P, Gurwith M, Heyward W, Martin M, van Griensven F, Hu D, Tappero JW, Choopanya K; Bangkok Vaccine Evaluation Group. 2006. **Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand.** J Infect Dis., 194(12):1661-71.

Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemée V, Damond F, Robertson DL, Simon F. 2009. **A new human immunodeficiency virus derived from gorillas.** Nat Med., 15(8):871-2.

Poinar H, Kuch M, Pääbo S. 2001. **Molecular analyses of oral polio vaccine samples.** Science, 292(5517):743-4.

Pope M, Haase AT. 2003. **Transmission, acute HIV-1 infection and the quest for strategies to prevent infection.** Nat Med., 9(7):847-52.

Pond, K. S.L., Posada, D., Gravenor, M. B., Woelk, C. H., Frost, S. D., 2006. **GARD: a genetic algorithm for recombination detection.** Bioinformatics. 22(24):3096-8.

Pontesilli O, Klein MR, Kerkhof-Garde SR, Pakker NG, de Wolf F, Schuitemaker H, Miedema F. 1998. **Longitudinal analysis of human immunodeficiency virus type 1-specific cytotoxic T lymphocyte responses: a predominant gag-specific response is associated with nonprogressive infection.** J Infect Dis., 178(4):1008-18.

Powers KA, Poole C, Pettifor AE, Cohen MS. 2008. **Rethinking the heterosexual infectivity of HIV-1: a systematic review and meta-analysis.** Lancet Infect Dis., 8(9):553-63.

Priddy FH, Brown D, Kublin J, Monahan K, Wright DP, Lalezari J, Santiago S, Marmor M, Lally M, Novak RM, Brown SJ, Kulkarni P, Dubey SA, Kierstead LS, Casimiro DR, Mogg R, DiNubile MJ, Shiver JW, Leavitt RY, Robertson MN, Mehrotra DV, Quirk E; Merck V520-016 Study Group. 2008. **Safety and immunogenicity of a replication-incompetent adenovirus type 5 HIV-1 clade B gag/pol/nef vaccine in healthy adults.** Clin Infect Dis., 46(11):1769-81.

Q

Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, Meehan MO, Lutalo T, Gray RH. 2000. **Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group.** N Engl J Med., 342(13):921-9.

Quiñones-Mateu ME, Gao Y, Ball SC, Marozsan AJ, Abraha A, Arts EJ. 2002. **In vitro intersubtype recombinants of human immunodeficiency virus type 1: comparison to recent and circulating in vivo recombinant forms.** J Virol., 76(19):9600-13.

R

Rademeyer C, Moore PL, Taylor N, Martin DP, Choge IA, Gray ES, Sheppard HW, Gray C, Morris L, Williamson C; HIVNET 028 study team. 2007. **Genetic characteristics of HIV-1 subtype C envelopes inducing cross-neutralizing antibodies.** Virology, 368(1):172-81.

Rakwar J, Lavreys L, Thompson ML, Jackson D, Bwayo J, Hassanali S, Mandaliya K, Ndinya-Achola J, Kreiss J. 1999. **Cofactors for the acquisition of HIV-1 among heterosexual men: prospective cohort study of trucking company workers in Kenya.** AIDS, 13(5):607-14.

Rambaut A, Posada D, Crandall KA, Holmes EC. 2004. **The causes and consequences of HIV evolution.** Nat Rev Genet., 5(1):52-61.

Ras GJ, Simson IW, Anderson R, Prozesky OW, Hamersma T. 1983. **Acquired immunodeficiency syndrome. A report of 2 South African cases.** S Afr Med J., 64(4):140-2.

Repits J, Sterjovski J, Badia-Martinez D, Mild M, Gray L, Churchill MJ, Purcell DF, Karlsson A, Albert J, Fenyö EM, Achour A, Gorry PR, Jansson M. 2008. **Primary HIV-1 R5 isolates from end-stage disease display enhanced viral fitness in parallel with increased gp120 net charge.** Virology, 379(1):125-34.

Rizzuto CD, Wyatt R, Hernández-Ramos N, Sun Y, Kwong PD, Hendrickson WA, Sodroski J. 1998. **A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding.** Science, 280(5371):1949-53.

Rolland M, Heckerman D, Deng W, Rousseau CM, Coovadia H, Bishop K, Goulder PJ, Walker BD, Brander C, Mullins JI. 2008. **Broad and Gag-biased HIV-1 epitope repertoires are associated with lower viral loads.** PLoS One, 3(1):e1424.

Rolland M, Jensen MA, Nickle DC, Yan J, Learn GH, Heath L, Weiner D, Mullins JI. 2007. **Reconstruction and function of ancestral center-of-tree human immunodeficiency virus type 1 proteins.** J Virol., 81(16):8507-14.

Rong R, Bibollet-Ruche F, Mulenga J, Allen S, Blackwell JL, Derdeyn CA. 2007. **Role of V1V2 and other human immunodeficiency virus type 1 envelope domains in resistance to autologous neutralization during clade C infection.** J Virol., 81(3):1350-9. Erratum in: J Virol. 2007 Nov; 81(22):12715.

Roques P, Robertson DL, Souquière S, Apetrei C, Nerrienet E, Barré-Sinoussi F, Müller-Trutwin M, Simon F. 2004. **Phylogenetic characteristics of three new HIV-1 N strains and implications for the origin of group N.** AIDS, 18(10):1371-81.

Rousseau CM, Daniels MG, Carlson JM, Kadie C, Crawford H, Prendergast A, Matthews P, Payne R, Rolland M, Raugi DN, Maust BS, Learn GH, Nickle DC, Coovadia H, Ndung'u T, Frahm N, Brander C, Walker BD, Goulder PJ, Bhattacharya T, Heckerman DE, Korber BT, Mullins JI. 2008. **HLA class I-driven evolution of human immunodeficiency virus type 1 subtype c proteome: immune escape and viral load.** J Virol., 82(13):6434-46.

Rousseau CM, Birditt BA, McKay AR, Stoddard JN, Lee TC, McLaughlin S, Moore SW, Shindo N, Learn GH, Korber BT, Brander C, Goulder PJ, Kiepiela P, Walker BD, Mullins JI. 2006. **Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes.** J Virol Methods, 136(1-2):118-25.

Russell RA, Moore MD, Hu WS, Pathak VK. 2009. **APOBEC3G induces a hypermutation gradient: purifying selection at multiple steps during HIV-1 replication results in levels of G-to-A mutations that are high in DNA, intermediate in cellular viral RNA, and low in virion RNA.** Retrovirology, 6:16.

S

Saah AJ, Hoover DR, Weng S, Carrington M, Mellors J, Rinaldo CR Jr, Mann D, Apple R, Phair JP, Detels R, O'Brien S, Enger C, Johnson P, Kaslow RA. 1998. **Association of HLA profiles with early plasma viral load, CD4+ cell count and rate of progression to AIDS following acute HIV-1 infection. Multicenter AIDS Cohort Study.** AIDS, 12(16):2107-13.

Sagar M, Laeyendecker O, Lee S, Gamiel J, Wawer MJ, Gray RH, Serwadda D, Sewankambo NK, Shepherd JC, Toma J, Huang W, Quinn TC. 2009. **Selection of HIV variants with signature genotypic characteristics during heterosexual transmission.** J Infect Dis., 199(4):580-9.

Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, Decker JM, Wang S, Baalwa J, Kraus MH, Parrish NF, Shaw KS, Guffey MB, Bar KJ, Davis KL, Ochsenbauer-Jambor C, Kappes JC, Saag MS, Cohen MS, Mulenga J, Derdeyn CA, Allen S, Hunter E, Markowitz M, Hraber P, Perelson AS, Bhattacharya T, Haynes BF, Korber BT, Hahn BH, Shaw

GM. 2009. **Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection.** J Exp Med., 206(6):1273-89.

Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, Derdeyn CA, Farmer P, Hunter E, Allen S, Manigart O, Mulenga J, Anderson JA, Swanstrom R, Haynes BF, Athreya GS, Korber BT, Sharp PM, Shaw GM, Hahn BH. 2008. **Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing.** J Virol., 82(8):3952-70.

Sanders-Buell E, Salminen M.O, McCuthan F.E. Sequencing primers for HIV-1. 1995. **Human Retroviruses and AIDS Compendium 1995. http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/95compendium.html.**

Santra S, Korber BT, Muldoon M, Barouch DH, Nabel GJ, Gao F, Hahn BH, Haynes BF, Letvin NL. 2008. **A centralized gene-based HIV-1 vaccine elicits broad cross-clade cellular immune responses in rhesus monkeys.** Proc Natl Acad Sci U S A., 105(30):10489-94.

Santra S, Liao HX, Zhang R, Muldoon M, Watson S, Fischer W, Theiler J, Szinger J, Balachandran H, Buzby A, Quinn D, Parks RJ, Tsao CY, Carville A, Mansfield KG, Pavlakis GN, Felber BK, Haynes BF, Korber BT, Letvin NL. 2010. **Mosaic vaccines elicit CD8+ T lymphocyte responses that confer enhanced immune coverage of diverse HIV strains in monkeys.** Nat Med., 16(3):324-8.

Scanlan CN, Pantophlet R, Wormald MR, Ollmann Saphire E, Stanfield R, Wilson IA, Katinger H, Dwek RA, Rudd PM, Burton DR. 2002. **The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of alpha1-->2 mannose residues on the outer face of gp120.** J Virol., 76(14):7306-21.

Schacker TW, Hughes JP, Shea T, Coombs RW, Corey L. 1998. **Biological and virologic characteristics of primary HIV infection.** Ann Intern Med., 128(8):613-20.

Scheffler K, Martin DP, Seoighe C. 2006. **Robust inference of positive selection from recombining coding sequences.** Bioinformatics, 15;22(20):2493-9.

Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA, Racz P, Tenner-Racz K, Dalesandro M, Scallon BJ, Ghrayeb J, Forman MA, Montefiori DC, Rieber EP, Letvin NL, Reimann KA. 1999. **Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes.** Science, 283(5403):857-60.

Schneidewind A, Brockman MA, Sidney J, Wang YE, Chen H, Suscovich TJ, Li B, Adam RI, Allgaier RL, Mothé BR, Kuntzen T, Oniangue-Ndza C, Trocha A, Yu XG, Brander C, Sette A, Walker BD, Allen TM. 2008. **Structural and functional constraints limit options for cytotoxic T-lymphocyte escape in the immunodominant HLA-B27-restricted epitope in human immunodeficiency virus type 1 capsid.** J Virol., 82(11):5594-605.

Sewram S, Singh R, Kormuth E, Werner L, Mlisana K, Karim SS, Ndung'u T; CAPRISA Acute Infection Study Team. 2009. **Human TRIM5alpha expression levels and reduced susceptibility to HIV-1 infection.** J Infect Dis., 199(11):1657-63.

Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, Gupta P, Rinaldo CR, Learn GH, He X, Huang XL, Mullins JI. 1999. **Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection.** J Virol., 73(12):10489-502.

Sharp PM, Bailes E, Chaudhuri RR, Rodenburg CM, Santiago MO, Hahn BH. 2001. **The origins of acquired immune deficiency syndrome viruses: where and when?** Philos Trans R Soc Lond B Biol Sci., 356(1410):867-76.

Sheehy AM, Gaddis NC, Malim MH. 2003. **The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif.** Nat Med., 9(11):1404-7.

Shephard E, Burgers WA, Van Harmelen JH, Monroe JE, Greenhalgh T, Williamson C, Williamson AL. 2008. **A multigene HIV type 1 subtype C modified vaccinia Ankara (MVA) vaccine efficiently boosts immune responses to a DNA vaccine in mice.** AIDS Res Hum Retroviruses, 24(2):207-17.

Shiver JW, Fu TM, Chen L, Casimiro DR, Davies ME, Evans RK, Zhang ZQ, Simon AJ, Trigona WL, Dubey SA, Huang L, Harris VA, Long RS, Liang X, Handt L, Schleif WA, Zhu L, Freed

DC, Persaud NV, Guan L, Punt KS, Tang A, Chen M, Wilson KA, Collins KB, Heidecker GJ, Fernandez VR, Perry HC, Joyce JG, Grimm KM, Cook JC, Keller PM, Kresock DS, Mach H, Troutman RD, Isopi LA, Williams DM, Xu Z, Bohannon KE, Volkin DB, Montefiori DC, Miura A, Krivulka GR, Lifton MA, Kuroda MJ, Schmitz JE, Letvin NL, Caulfield MJ, Bett AJ, Youil R, Kaslow DC, Emini EA. 2002. **Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity.** Nature, 415(6869):331-5.

Srivastava IK, Ulmer JB, Barnett SW. 2005. **Role of neutralizing antibodies in protective immunity against HIV.** Hum Vaccin., 1(2):45-60. Epub 2005 Mar 17.

Srivastava IK, Stamatatos L, Kan E, Vajdy M, Lian Y, Hilt S, Martin L, Vita C, Zhu P, Roux KH, Vojtech L, C Montefiori D, Donnelly J, Ulmer JB, Barnett SW. 2003. **Purification, characterization, and immunogenicity of a soluble trimeric envelope protein containing a partial deletion of the V2 loop derived from SF162, an R5-tropic human immunodeficiency virus type 1 isolate.** J Virol., 77(20):11244-59.

Stacey AR, Norris PJ, Qin L, Haygreen EA, Taylor E, Heitman J, Lebedeva M, DeCamp A, Li D, Grove D, Self SG, Borrow P. 2009. **Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections.** J Virol., 83(8):3719-33.

Stafford MA, Corey L, Cao Y, Daar ES, Ho DD, Perelson AS. 2000. **Modeling plasma virus concentration during primary HIV infection.** J Theor Biol., 203(3):285-301.

Stamatatos L, Morris L, Burton DR, Mascola JR. 2009. **Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine?** Nat Med., 15(8):866-70.

Stratov I, Chung A, Kent SJ. 2008. **Robust NK cell-mediated human immunodeficiency virus (HIV)-specific antibody-dependent responses in HIV-infected subjects.** J Virol., 82(11):5450-9.

Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. 2004. **The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys.** Nature, 427(6977):848-53.

Sheehy AM, Gaddis NC, Choi JD, Malim MH. 2002. **Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein.** Nature, 418(6898):646-50.

Spracklen FH, Whittaker RG, Becker WB, Becker ML, Holmes CM, Potter PC. 1985. **The acquired immune deficiency syndrome and related complex. A report of 2 confirmed cases in Cape Town with comments on human T-cell lymphotropic virus type III infections.** S Afr Med J., 68(3):139-43.

Streeck H, Lichterfeld M, Alter G, Meier A, Teigen N, Yassine-Diab B, Sidhu HK, Little S, Kelleher A, Routy JP, Rosenberg ES, Sekaly RP, Walker BD, Altfeld M. 2007. **Recognition of a defined region within p24 gag by CD8+ T cells during primary human immunodeficiency virus type 1 infection in individuals expressing protective HLA class I alleles.** J Virol., 81(14):7725-31.

Streeck H, Jolin JS, Qi Y, Yassine-Diab B, Johnson RC, Kwon DS, Addo MM, Brumme C, Routy JP, Little S, Jessen HK, Kelleher AD, Hecht FM, Sekaly RP, Rosenberg ES, Walker BD, Carrington M, Altfeld M. 2009. **Human immunodeficiency virus type 1-specific CD8+ T-cell responses during primary infection are major determinants of the viral set point and loss of CD4+ T cells.** J Virol., 83(15):7641-8.

Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. 2004. **The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys.** Nature, 427(6977):848-53.

T

Takehisa J, Kraus MH, Ayouba A, Bailes E, Van Heuverswyn F, Decker JM, Li Y, Rudicell RS, Learn GH, Neel C, Ngole EM, Shaw GM, Peeters M, Sharp PM, Hahn BH. 2009. **Origin and biology of simian immunodeficiency virus in wild-living western gorillas.** J Virol., 83(4):1635-48.

Tamura K, Dudley J, Nei M, Kumar S. 2007. **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** Mol Biol Evol., 24(8):1596-9.

Tang J, Costello C, Keet IP, Rivers C, Leblanc S, Karita E, Allen S, Kaslow RA. 1999. **HLA class I homozygosity accelerates disease progression in human immunodeficiency virus type 1 infection.** AIDS Res Hum Retroviruses, 15(4):317-24.

Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. 2008. **The challenge of HIV-1 subtype diversity.** N Engl J Med., 358(15):1590-602. *Erratum in*: N Engl J Med. 2008 Oct 30;359(18):1965-6; 1972.

Tiemessen CT, Shalekoff S, Meddows-Taylor S, Schramm DB, Papathanasopoulos MA, Gray GE, Sherman GG, Coovadia AH, Kuhn L. 2009. **Cutting Edge: Unusual NK cell responses to HIV-1 peptides are associated with protection against maternal-infant transmission of HIV-1.** J Immunol., 182(10):5914-8.

Thompson JD, Higgins DG, Gibson TJ. 1994. **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** Nucleic Acids Res., 22(22):4673-80.

Thomson MM, Pérez-Alvarez L, Nájera R. 2002. **Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy.** Lancet Infect Dis., 2(8):461-71.

Thurmond J, Yoon H, Kuiken C, Yusim K, Perkins S, Theiler J, Bhattacharya T, Korber B, Fischer W. 2008. **Web-based design and evaluation of T-cell vaccine candidates.** Bioinformatics, 24(14):1639-40.

Tomaras GD, Yates NL, Liu P, Qin L, Fouda GG, Chavez LL, Decamp AC, Parks RJ, Ashley VC, Lucas JT, Cohen M, Eron J, Hicks CB, Liao HX, Self SG, Landucci G, Forthal DN, Weinhold KJ, Keele BF, Hahn BH, Greenberg ML, Morris L, Karim SS, Blattner WA, Montefiori DC, Shaw GM, Perelson AS, Haynes BF. 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(24):12449-63.

Trkola A, Purtscher M, Muster T, Ballaun C, Buchacher A, Sullivan N, Srinivasan K, Sodroski J, Moore JP, Katinger H. 1996. **Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1.** J Virol., 70(2):1100-8.

Troyer RM, Collins KR, Abraha A, Fraundorf E, Moore DM, Krizan RW, Toossi Z, Colebunders RL, Jensen MA, Mullins JI, Vanham G, Arts EJ. 2005. **Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression.** J Virol., 79(14):9006-18.

Troyer RM, McNevin J, Liu Y, Zhang SC, Krizan RW, Abraha A, Tebit DM, Zhao H, Avila S, Lobritz MA, McElrath MJ, Le Gall S, Mullins JI, Arts EJ. 2009. **Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response.** PLoS Pathog., 5(4):e1000365.

Towers GJ. 2007. **The control of viral infection by tripartite motif proteins and cyclophilin A.** Retrovirology, 4:40.

V

Valentine LE, Piaskowski SM, Rakasz EG, Henry NL, Wilson NA, Watkins DI. 2008. **Recognition of escape variants in ELISPOT does not always predict CD8+ T-cell recognition of simian immunodeficiency virus-infected cells expressing the same variant sequences.** J Virol.; 82(1):575-81.

Van Damme L, Ramjee G, Alary M, Vuylsteke B, Chandeying V, Rees H, Sirivongrangson P, Mukenge-Tshibaka L, Ettiègne-Traoré V, Uaheowitchai C, Karim SS, Mâsse B, Perriëns J, Laga M; COL-1492 Study Group. 2002. **Effectiveness of COL-1492, a nonoxynol-9 vaginal gel, on HIV-1 transmission in female sex workers: a randomised controlled trial.** Lancet, 360(9338):971-7. Erratum in: Lancet 2002 Dec 7; 360(9348):1892.

van Harmelen J, Wood R, Lambrick M, Rybicki EP, Williamson AL, Williamson C. 1997. **An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa.** AIDS, 11(1):81-7.

Van Harmelen JH, Van der Ryst E, Loubser AS, York D, Madurai S, Lyons S, Wood R, Williamson C. 1999. **A predominantly HIV type 1 subtype C-restricted epidemic in South African urban populations.** AIDS Res Hum Retroviruses, 15(4):395-8.

Van Heuverswyn F, Li Y, Neel C, Bailes E, Keele BF, Liu W, Loul S, Butel C, Liegeois F, Bienvenue Y, Ngolle EM, Sharp PM, Shaw GM, Delaporte E, Hahn BH, Peeters M. 2006. **Human immunodeficiency viruses: SIV infection in wild gorillas.** Nature, 444(7116):164.

Van Heuverswyn F, Li Y, Bailes E, Neel C, Lafay B, Keele BF, Shaw KS, Takehisa J, Kraus MH, Loul S, Butel C, Liegeois F, Yangda B, Sharp PM, Mpoudi-Ngole E, Delaporte E, Hahn BH, Peeters M. 2007. **Genetic diversity and phylogeographic clustering of SIVcpzPtt in wild chimpanzees in Cameroon.** Virology, 368(1):155-71.

van Loggerenberg F, Mlisana K, Williamson C, Auld SC, Morris L, Gray CM, Abdool Karim Q, Grobler A, Barnabas N, Iriogbe I, Abdool Karim SS; 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(4):e1954.

van't Wout AB, Kootstra NA, Mulder-Kampinga GA, Albrecht-van Lent N, Scherpbier HJ, Veenstra J, Boer K, Coutinho RA, Miedema F, Schuitemaker H. 1994. **Macrophage-tropic variants initiate human immunodeficiency virus type 1 infection after sexual, parenteral, and vertical transmission.** J Clin Invest., 94(5):2060-7.

van Wamel JL, Berkhout B. 1998. **The first strand transfer during HIV-1 reverse transcription can occur either intramolecularly or intermolecularly.** Virology, 244(2):245-51.

Veazey RS, DeMaria M, Chalifoux LV, Shvetz DE, Pauley DR, Knight HL, Rosenzweig M, Johnson RP, Desrosiers RC, Lackner AA. 1998. **Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection.** Science, 280(5362):427-31.

Vlahov D, Graham N, Hoover D, Flynn C, Bartlett JG, Margolick JB, Lyles CM, Nelson KE, Smith D, Holmberg S, Farzadegan H. 1998. **Prognostic indicators for AIDS and infectious**

disease death in HIV-infected injection drug users: plasma viral load and CD4+ cell count. JAMA, 279(1):35-40.

W

Wagner R, Leschonsky B, Harrer E, Paulus C, Weber C, Walker BD, Buchbinder S, Wolf H, Kalden JR, Harrer T. 1999. **Molecular and functional analysis of a conserved CTL epitope in HIV-1 p24 recognized from a long-term nonprogressor: constraints on immune escape associated with targeting a sequence essential for viral replication.** J Immunol., 162(6):3727-34.

Wagner R, Modrow S, Böltz T, Fliessbach H, Niedrig M, von Brunn A, Wolf H. 1992. **Immunological reactivity of a human immunodeficiency virus type I derived peptide representing a consensus sequence of the GP120 major neutralizing region V3.** Arch Virol., 127(1-4):139-52.

Wahren B, Morfeldt-Månsson L, Biberfeld G, Moberg L, Sönnerborg A, Ljungman P, Werner A, Kurth R, Gallo R, Bolognesi D. 1987. **Characteristics of the specific cell-mediated immune response in human immunodeficiency virus infection.** J Virol., 61(6):2017-23.

Wang B, Spira TJ, Owen S, Lal RB, Saksena NK. 2000. **HIV-1 strains from a cohort of American subjects reveal the presence of a V2 region extension unique to slow progressors and non-progressors.** AIDS, 14(3):213-23.

Wang YE, Li B, Carlson JM, Streeck H, Gladden AD, Goodman R, Schneidewind A, Power KA, Toth I, Frahm N, Alter G, Brander C, Carrington M, Walker BD, Altfeld M, Heckerman D, Allen TM. 2009. **Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in highly conserved regions of human immunodeficiency virus type 1.** J Virol., 83(4):1845-55.

Watkins DI, Burton DR, Kallas EG, Moore JP, Koff WC. 2008. **Nonhuman primate models and the failure of the Merck HIV-1 vaccine in humans.** Nat Med., 14(6):617-21.

Weaver EA, Lu Z, Camacho ZT, Moukdar F, Liao HX, Ma BJ, Muldoon M, Theiler J, Nabel GJ, Letvin NL, Korber BT, Hahn BH, Haynes BF, Gao F. 2006. **Cross-subtype T-cell immune**

responses induced by a human immunodeficiency virus type 1 group m consensus env immunogen. J Virol., 80(14):6745-56.

Weber K, Meyer D, Grosse V, Stoll M, Schmidt RE, Heiken H. 2000. **Reconstitution of NK cell activity in HIV-1 infected individuals receiving antiretroviral therapy.** Immunobiology, 202(2):172-8.

Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS, Komarova NL, Nowak MA, Hahn BH, Kwong PD, Shaw GM. 2003. **Antibody neutralization and escape by HIV-1.** Nature, 422(6929):307-12.

Wilkinson E, Engelbrecht S. 2009. **Molecular characterization of non-subtype C and recombinant HIV-1 viruses from Cape Town, South Africa.** Infect Genet Evol., 9(5):840-6.

Williamson C, Engelbrecht S, Lambrick M, van Rensburg EJ, Wood R, Bredell W, Williamson AL. 1995. **HIV-1 subtypes in different risk groups in South Africa.** Lancet, 346(8977):782.

Williamson C, Morris L, Maughan MF, Ping LH, Dryga SA, Thomas R, Reap EA, Cilliers T, van Harmelen J, Pascual A, Ramjee G, Gray G, Johnston R, Karim SA, Swanstrom R. 2003. **Characterization and selection of HIV-1 subtype C isolates for use in vaccine development.** AIDS Res Hum Retroviruses, 19(2):133-44.

Wolfs TF, Zwart G, Bakker M, Goudsmit J. 1992. **HIV-1 genomic RNA diversification following sexual and parenteral virus transmission.** Virology, 189(1):103-10.

Wolinsky SM, Wike CM, Korber BT, Hutto C, Parks WP, Rosenblum LL, Kunstman KJ, Furtado MR, Muñoz JL. 1992. **Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants.** Science, 255(5048):1134-7.

Wood N, Bhattacharya T, Keele BF, Giorgi E, Liu M, Gaschen B, Daniels M, Ferrari G, Haynes BF, McMichael A, Shaw GM, Hahn BH, Korber B, Seoighe C. 2009. **HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC.** PLoS Pathog., 5(5):e1000414.

Woodman Z, Williamson C. 2009. **HIV molecular epidemiology: transmission and adaptation to human populations.** Curr Opin HIV AIDS, 4(4):247-52.

Worobey M, Santiago ML, Keele BF, Ndjango JB, Joy JB, Labama BL, Dhed'A BD, Rambaut A, Sharp PM, Shaw GM, Hahn BH. 2004. **Origin of AIDS: contaminated polio vaccine theory refuted.** Nature, 428(6985):820.

Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, Muyembe JJ, Kabongo JM, Kalengayi RM, Van Marck E, Gilbert MT, Wolinsky SM. 2008. **Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960.** Nature, 455(7213):661-4.

Y

Yang C, Dash BC, Simon F, van der Groen G, Pieniazek D, Gao F, Hahn BH, Lal RB. 2000. **Detection of diverse variants of human immunodeficiency virus-1 groups M, N, and O and simian immunodeficiency viruses from chimpanzees by using generic pol and env primer pairs.** J Infect Dis., 181(5):1791-5.

Yokomaku Y, Miura H, Tomiyama H, Kawana-Tachikawa A, Takiguchi M, Kojima A, Nagai Y, Iwamoto A, Matsuda Z, Ariyoshi K. 2004. **Impaired processing and presentation of cytotoxic-T-lymphocyte (CTL) epitopes are major escape mechanisms from CTL immune pressure in human immunodeficiency virus type 1 infection.** J Virol., 78(3):1324-32.

Yu X, Yu Y, Liu B, Luo K, Kong W, Mao P, Yu XF. 2003. **Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex.** Science, 302(5647):1056-60.

Yu Y, Xiao Z, Ehrlich ES, Yu X, Yu XF. 2004. **Selective assembly of HIV-1 Vif-Cul5-ElonginB-ElonginC E3 ubiquitin ligase complex through a novel SOCS box and upstream cysteines.** Genes Dev., 18(23):2867-72.

Z

Zhang C, Cornette JL, Berzofsky JA, DeLisi C. 1997. **The organization of human leucocyte antigen class I epitopes in HIV genome products: implications for HIV evolution and vaccine design.** Vaccine, 15(12-13):1291-302.

Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. 2003. **The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA.** Nature, 424(6944):94-8.

Zhang M, Gaschen B, Blay W, Foley B, Haigwood N, Kuiken C, Korber B. 2004. **Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin.** Glycobiology, 14(12):1229-46.

Zhang Z, Schuler T, Zupancic M, Wietgrefe S, Staskus KA, Reimann KA, Reinhart TA, Rogan M, Cavert W, Miller CJ, Veazey RS, Notermans D, Little S, Danner SA, Richman DD, Havlir D, Wong J, Jordan HL, Schacker TW, Racz P, Tenner-Racz K, Letvin NL, Wolinsky S, Haase AT. 1999. **Sexual transmission and propagation of SIV and HIV in resting and activated CD4+ T cells.** Science; 286(5443):1353-7. Erratum in: Science 1999 Dec 17; 286(5448):2273.

Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD. 1998. **An African HIV-1 sequence from 1959 and implications for the origin of the epidemic.** Nature, 391(6667):594-7.

Zwick MB, Labrijn AF, Wang M, Spenlehauer C, Saphire EO, Binley JM, Moore JP, Stiegler G, Katinger H, Burton DR, Parren PW. 2001. **Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41.** J Virol., 75(22):10892-905.