The transmitted HIV-1 subtype C: Characterization of the transmitted/founder full-length virus genome and the influence of early immune selective pressure on virus replication

Melissa-Rose Hilda Abrahams

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Clinical Laboratory Sciences

UNIVERSITY OF CAPE TOWN

October 2014

Table of Contents

Declaration

I, **Melissa-Rose Hilda Abrahams**, hereby declare that the work on which this thesis/dissertation is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorise the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature:……Date: …….

Abstract

The identification of targets of early immune responses associated with control of HIV-1 infection will inform immunogen design for vaccine interventions. The early evolution of transmitted/founder subtype C virus sequences was investigated to determine the location and frequency of immune selection, and the impact of early immune escape mutations on viral replicative capacity.

Single-genome amplified *env* sequences from 26 acutely-infected women were evaluated for conformance to a model of random evolution to elucidate multiplicity of infection. Near full-length genome sequences from the first six months of infection were generated for five women and sites evolving under immune selection were mapped. CD8+ cytotoxic T-lymphocyte escape mutations in HLA-B-restricted epitopes were introduced into infectious molecular clones of cognate transmitted/founder viruses by site-directed mutagenesis and their impact on viral replicative fitness was evaluated using parallel replication assays.

In 77% of women (n=20) a single transmitted/founder variant established infection and two to five variants in the remaining 23% (n=6). Near full-length genome sequencing in five women confirmed single variant/low-diversity transmission and identified fifty-five genome regions evolving under immune selection, 40% of which was attributed to CD8+ cytotoxic T-lymphocyte pressure, 35% to antibody-mediated pressure, 16% to reversion and 9% could not be classified. The rate of sequence diversification and number of sites evolving under immune selection was highest in *nef*. The majority of evolving CD8+ cytotoxic T-lymphocyte epitopes (82%) contained shuffling/toggling mutations. A novel B*15:10-associated mutation, A164T, combined with a V85A Pol mutation reduced viral replication capacity in one individual. In a second individual, the attenuating HLA-B*58:01-associated mutation, T242N, enhanced viral replication capacity due to pre-existing compensatory polymorphisms in the transmitted/founder virus. A third individual, who had extremely rapid disease progression, was infected with the virus with the highest replication capacity.

This thesis describes the complex nature of early immune selection and escape in transmitted/founder viruses. Although attenuating escape mutations were identified in viruses from two individuals, this was not associated with clinical benefit. The extensive variability of epitopes evolving under early selection may implicate many early immune targets as poor candidates for vaccine immunogens; however some early targets may be useful if clinical benefit is conferred through attenuating escape mutations.

Acknowledgements

I am very grateful to the following people for their support and contributions toward this thesis:

- My parents, Lorrimer and Rosemarie Abrahams, and my sister, Leslie-Anne Boer, to whom I am immensely grateful for ongoing support and encouragement.

- My supervisor and mentor Prof Carolyn Williamson for her excellent supervision, enthusiasm for HIV research, her belief in me and her consistent investment in my development.

- Dr Florette Treurnicht for her mentorship, training and example as a researcher.

- International hosts and trainers: Dr Beatrice Hahn, Dr Jesus Salazar-Gonzalez and Dr Maria Salazar for training in single-genome amplification of full-length genomes; Dr John Kappes, Dr Christina Ochsenbauer, Dr Haitao Ding and Miss Corinne Zaragoza for training in the culturing of viruses from infectious molecular clones; Dr Feng Gao, Dr Bhavna Hora and Dr Amit Kumar for training in infectious molecular clone mutagenesis and replication kinetics assays.

- The HIV diversity group, which I'm privileged to be a member of, for their encouragement, assistance and support.

- The staff of the HIV diversity group and Institute of Infectious Disease and Molecular Medicine at UCT whose work and guidance contributed to this thesis: Miss Ruwayhida Thebus, Miss Jinny Marias and Miss Helba Bredell for their contribution to the generation of *env*, full-length and half-genome sequences; Dr Florette Treurnicht for training and assistance with the generation, sequencing and analysis of *env* and full-length genomes as well as for the generation of four of the infectious molecular clones used in this thesis along with Dr Haitao Ding; Miss Sarah Goodier for assistance with design of the infectious molecular clone generated for this thesis; Dr Philippe Selhorst for assistance and guidance with replication kinetics assay optimization; Dr Nobubelo Ngandu for assistance with positive selection analyses; Dr Cathal Seoighe for guidance with statistical and mutational analyses; Dr Clive Gray for immunology consultation and Miss Debra de Assis Rosa for HLA typing of CAPRISA participants.

- The funders of the work and training towards this thesis: the Center for HIV/AIDS Vaccine Immunology (CHAVI) and the Columbia University Fogarty AIDS International Training Program (AITRP).

List of Abbreviations

Ab	Antibody
ADCC	Antibody-dependant cellular cytotoxicity
AI	Acute infection
AIDS	Acquired immune deficiency syndrome
ANOVA	Analysis of variance
APOBEC	Apolipoprotein B mRNA editing enzyme
ART	Antiretroviral therapy
BCN	Broadly cross-neutralizing
BEAST	Bayesian evolutionary analysis sampling trees
CAPRISA	Centre for the AIDS program of research in South Africa
CDHR3	Cadhedrin-related family member 3
CHAVI	Center for HIV/AIDS vaccine immunology
COT	Centre-of-tree
CTL	Cytotoxic T-lymphocyte
D-MEM	Dulbecco's minimal essential medium
ELISA/EIA	Enzyme-linked immunosorbent assay
ELISPOT	Enzyme-linked immunospot
GWAS	Genome-wide association study
HEK293T	Human embryonic kidney 293T
HIV	Human immunodeficiency syndrome
HLA	Human leukocyte antigen
HSV2	Herpes Simplex Virus 2
HVTN	HIV vaccine trial network

IDU	Intravenous drug user
IMC	Infectious molecular clone
Indel	Insertion/deletion
KIR	Killer cell immunoglobulin-like receptor
min	minutes
ml	millilitre
MPER	Membrane proximal exterior region
nAb	neutralizing antibody
ng	nanogram
NK	Natural killer
ORF	Open reading frame
PBMCs	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PNGS	Potential N-linked glycosylation site
RC	Replication capacity
RPMI	Roswell Park Memorial Institute medium
RT	Reverse transcriptase
sec	seconds
SEVI	Semen-derived enhancer of virus infection
STI	Sexually transmitted infection
TB	Tuberculosis
t/f	transmitted/founder
MRCA	time to most recent common ancestor
U	Units

µg	microgram
µl	microliter
V-loop	Variable-loop

Chapter 1: Literature Review

Chapter 1: Literature Review

1.1 Current perspective of the global HIV-1 pandemic and the progress of vaccine development

Human Immunodeficiency Virus (HIV) continues to be a global health threat. Although the WHO reported stabilization in the global HIV incidence in 2010, the disease burden continues to increase in sub-Saharan Africa (World Health Organization, 2011). In 2012, there were 35 million infected individuals worldwide of which 25 million were in sub-Saharan Africa (http://www.who.int/gho/hiv/en/) constituting greater than two-thirds of the pandemic. South Africa carries the highest burden of any country globally (**Figure 1,** http://www.unaids.org/en/dataanalysis/datatools/aidsinfo/), and in 2013 it was estimated that 5.26 million people in South Africa were living with HIV (Statistics South Africa, 2013).

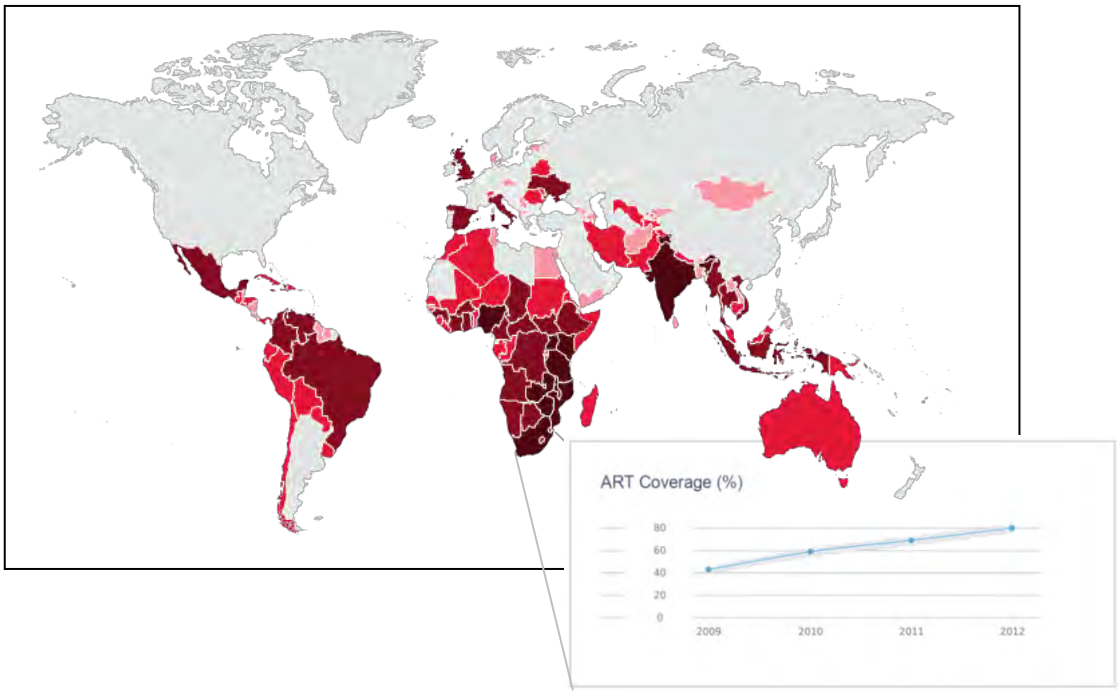


Figure 1. **Global HIV disease burden in 2012**. Disease burden is illustrated at country level, and the coverage of antiretroviral therapy (ART) is illustrated graphically for South Africa (inset) (modified from http://www.unaids.org/en/dataanalysis/datatools/aidsinfo/). The number of individuals living with HIV is indicated by colour coding: dark brown = >1 million people, pale brown = 100 000 to 1 million people, dark red = 10 000 to 100 000 people, pale red = < 10 000 people, and grey = no data available. South Africa remains the country with the highest number of HIV cases despite the roll-out of antiretroviral therapy (ART) (Pillay-van Wyk et al., 2013).

Furthermore, in South Africa HIV-related mortality is exacerbated by the burden of tuberculosis (TB) and HIV-TB co-infection (Karim et al., 2009). While antiretroviral (ARV) therapy has provided relief with respect prolonging life, and the lowering of HIV incidence, (Tanser et al., 2013; Bor et al., 2013), treatment is still not accessible to all those who need it. In the absence of a cure, a vaccine remains urgently needed.

To date, only one HIV vaccine, tested in the Thai RV144 trial, has provided protection, preventing infection in 31% of recipients (Rerks-Ngarm et al., 2009). The trial tested a combination of a canarypox viral vector carrying HIV genes (*gag*, *pol* and *env*) for stimulating cytotoxic T-lymphocyte (CTL) responses, with a viral envelope (Env) protein immunogen for stimulating antibody (Ab) responses. Prior to the RV144 trial, only five phase IIb/III vaccine efficacy trials were conducted, none of which protected from infection, nor lowered viral loads in recipients (**Figure 2**). Two of these trials used only viral protein immunogens (VaxGen trials VAX003 and VAX004), two used adenovirus viral vectors expressing HIV genes with no protein immunogens (HVTN502/Merck rAd5 Step and HVTN503/Phambili trial), and the last used a DNA prime followed by an adenovirus viral vector boost (HVTN 505) (Esparza, 2013).

A multi-pronged approach may thus be needed to prevent HIV infection, wherein vaccines will need to elicit a combination of Ab and CTL responses. The success of the RV144 trial in providing, albeit modest, protection against HIV infection demonstrates that an effective vaccine is an attainable goal. A similar vaccine trial approach to RV144, but using subytpe C immunogens which are more relevant to the pandemic in southern Africa, is planned for South Africa (**Figure 2**).

With the slow implementation of new vaccine trials, it is imperative that additional research is undertaken to elucidate the correlates of protection against the establishment of infection, and correlates of disease attenuation. Investigation of HIV-infected individuals to gain an understanding of the interplay between the virus and the host immune system following viral transmission can provide important clues regarding early responses associated with viral control. This would involve characterising transmitted viruses and determining whether they possess common properties that can be exploited by a vaccine or if mutational pathways can be identified that would attenuate the virus resulting in modification or control of disease.

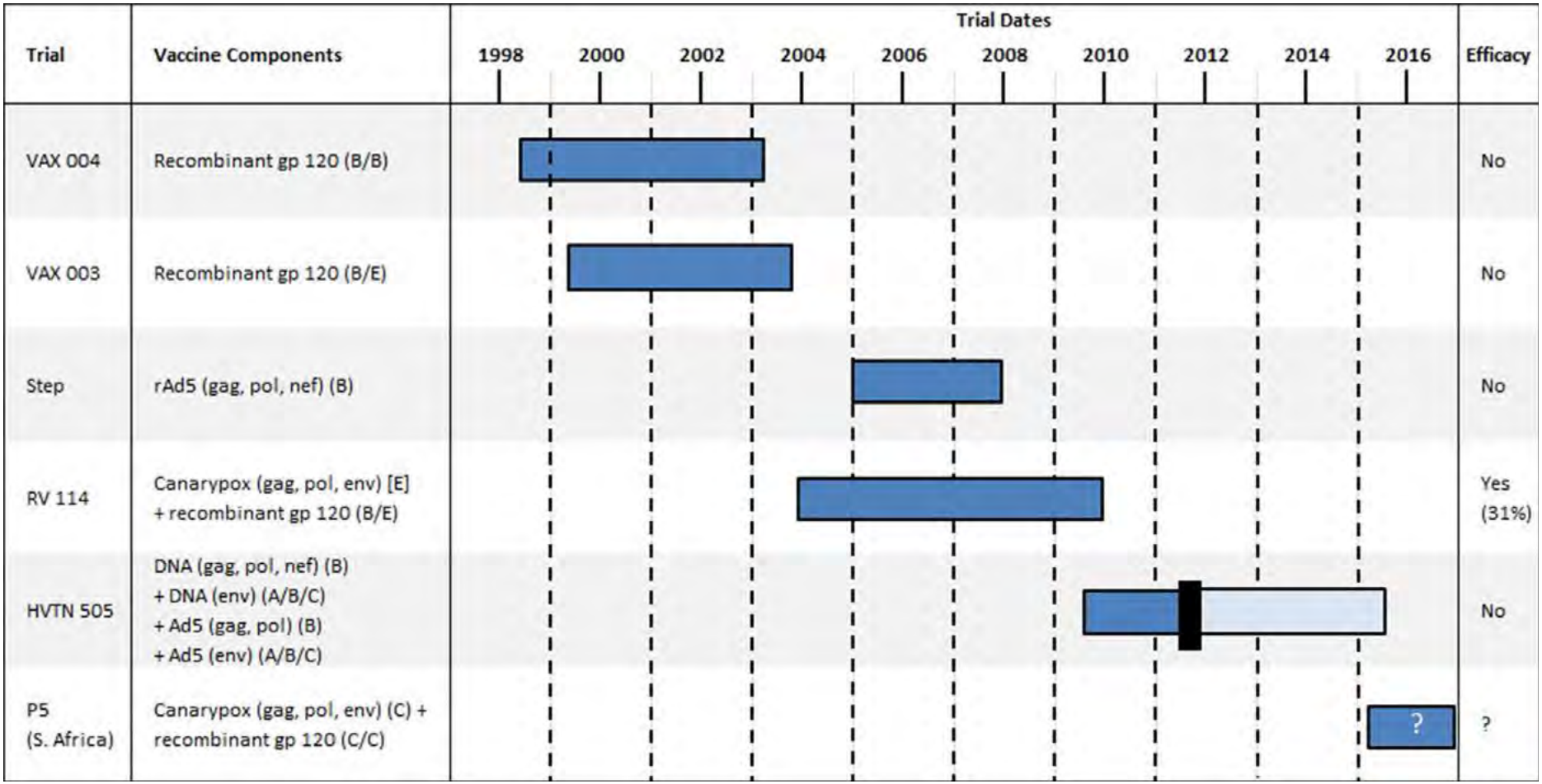


Figure 2. **HIV-1 vaccine efficacy trials conducted to date or in progress** (obtained from Esparza, 2013). The HVTN503 Phambili trial, which tested the same vaccines as the HVTN502 Step trial, is not illustrated here as it was stopped early due to lack of protection seen in HVTN502 vaccine recipients, and also for safety reasons (Gray et al., 2014).

The purpose of this literature review is to summarize the current understanding of HIV transmission and subsequent interaction with the host immune system, for informing the design of effective vaccines. The following areas are discussed: (i) HIV diversity, (ii) the transmitted virus, (iii) early immune pressure on the virus, and (iv) challenges to vaccine design.

1.2 Viral diversity

HIV-1 exists as a diversity of subtypes (reviewed in Ndung'u and Weiss, 2012), and development of a vaccine requires a detailed characterization of this diversity as viral subtype may impact vaccine efficacy. HIV-1 is classified into nine subtypes (A, B, C, D, F, G, H, J and K). A number of viral chimeras having genetic material from two different subtypes also exist and are referred to as circulating (if associated with spread) and unique recombinant forms (CRFs and URFs). Most vaccine trials to date have tested vaccine constructs containing subtype B immunogens (reviewed in Esparza 2013). However, global statistics reveal that subtype C is by far the most dominant strain (Hemelaar et al., 2011).

In an analysis of global HIV-1 subtype distribution between the years 2000 and 2007, subtype C was found to account for almost half (48%) of global infections, and is the focus of this thesis. Recombinant viruses are the second most common strain (21%), followed by subtype A (12%) and B (11%) (Hemelaar et al., 2011) (**Figure 3A**). Subtype C predominates in Southern Africa, Ethiopia and India, and appears to be on the increase in the southern regions of Brazil (Raboni et al., 2010, Almeida et al., 2012). Most research has been performed on subtype B as this is the dominant strain in the Americas, Europe and Australia (Hemelaar et al., 2011) (**Figure 3B**).

Subtypes may impact the efficacy of vaccines designed to elicit both B- and T-cell responses. Recently, in a study investigating a large panel of *env* pseudoviruses from a range of different subtypes, viruses were shown to be more sensitive to neutralization by plasma from individuals infected with a matched subtype compared to that from mismatched subtypes (Hraber et al., 2014). Similarly, CTL responses are more likely to recognise viruses from matched subtypes (Dorrell et al., 1999, Currier et al., 2003).

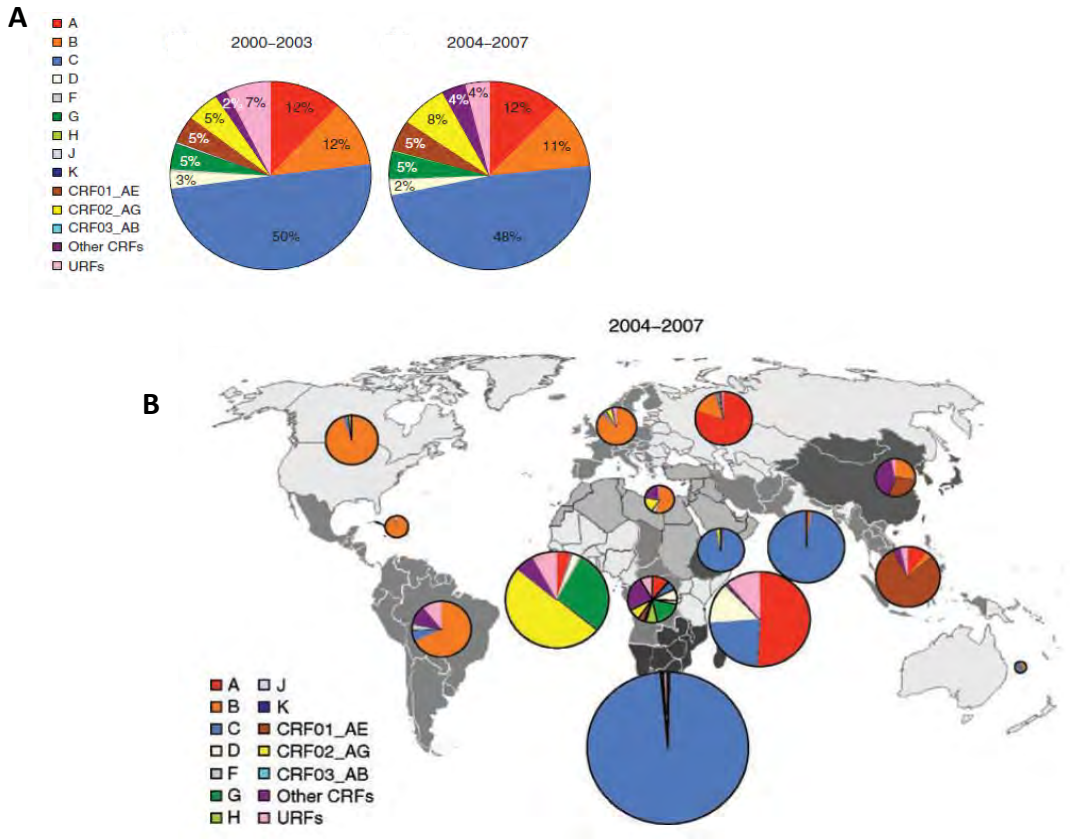


Figure 3. **Global burden and distribution of HIV-1 subtypes.** The contribution of the different subtypes, circulating and unique recombinant forms (CRFs and URFs) to (**A**) the global, and (**B**) the regional HIV-1 disease burden. In addition, the global distribution of subtypes and recombinant forms is illustrated in B, where the sizes of the circles reflect the extent of the pandemic in each region. The dominance of subtype C in Southern Africa and the contribution of this subtype to global infections are evident. Modified from Hemelaar et al., 2011.

Concepts for overcoming subtype diversity for CTL-based vaccines have been developed and include, among others, universal vaccines presenting subtype consensus sequences of the most conserved regions of the HIV genome (Letourneau et al., 2007, Rolland et al., 2007, Borthwick et al., 2014) or vaccines presenting COT (centre of tree) immunogens which are designed using phylogenetic methods and are more representative of diversity of viral gene regions (Nickle et al., 2007). Another concept is that of mosaic vaccines, which are recombinant vaccines designed and synthesized to represent a diversity of HIV strains. These immunogens, when tested in rhesus monkeys, were shown to stimulate broader cellular immune responses compared to control vaccine constructs harbouring naturally circulating and

consensus antigens (Barouch et al., 2010, Santra et al., 2010). There has been no vaccine construct to date that has elicited broadly neutralizing antibodies.

1.3 The transmitted virus

It has long been known that a bottleneck exists at the point of HIV-1 transmission wherein the genetic diversity of viruses in a newly infected individual is usually homogeneous despite the high diversity of viral quasispecies in the infecting donor (Wolinsky et al., 1992, Derdyn et al., 2004).

Keele et al. (2008) were the first group to quantify this bottleneck where, in a subtype B cohort, they estimated that approximately 80% of infections were the result of a single transmitted/founder (t/f) viral variant establishing clinical infection. This was subsequently confirmed in a number of studies including a subtype C study in South Africa and Malawi (Abrahams et al., 2009), a subtype A and C study in Zambia and Rwanda (Halaand et al., 2009), and in a study of Tanzanian samples which included a variety of different subtypes as well as recombinant forms (Nofemela et al., 2011). This thesis expands the characterization of the reported t/f viruses from South African women (Abrahams et al., 2009).

The transmission bottleneck is less pronounced in cohorts of men who have sex with men (MSM) (Li et al., 2010), and conflicting reports exist on the multiplicity of infection in intravenous drug users (IDUs) (Masharsky et al., 2010; Li et al., 2010). While approximately 70% of IDUs from a cohort from St Petersburg, Russia were reported to be infected with a single variant (Masharsky et al., 2010), single variant infection was reported in only 40% of IDUs in a Montreal cohort (Bar et al., 2010). The restriction in the number of viruses that establish clinical infection in heterosexual transmission is likely a reflection of the effectiveness of the mucosal barrier that the virus has to cross. Supporting this, in macaque studies the number of transmitted variants establishing infection was seen to increase with increasing inoculum dose and intravenous inoculation resulted in both a higher efficiency of transmission as well as a higher number of transmitted variants compared to intrarectal inoculation (Keele et al., 2009, Liu et al., 2010).

The different scenarios which could result in this transmission bottleneck are depicted in **Figure 4.** It is possible that during transmission multiple viral variants traverse the mucosal barrier but that only one or a few successfully infect target cells (Cohen et al., 2011), in which case these viruses may be referred to as ‘transmitted’ viruses. Alternatively, during transmission, multiple

viral variants may successfully infect target cells but infection may be aborted by all but one or a few which go on to expand (Cohen et al., 2011). In this scenario the virus which persists and expands would be referred to as the 'founder' virus.

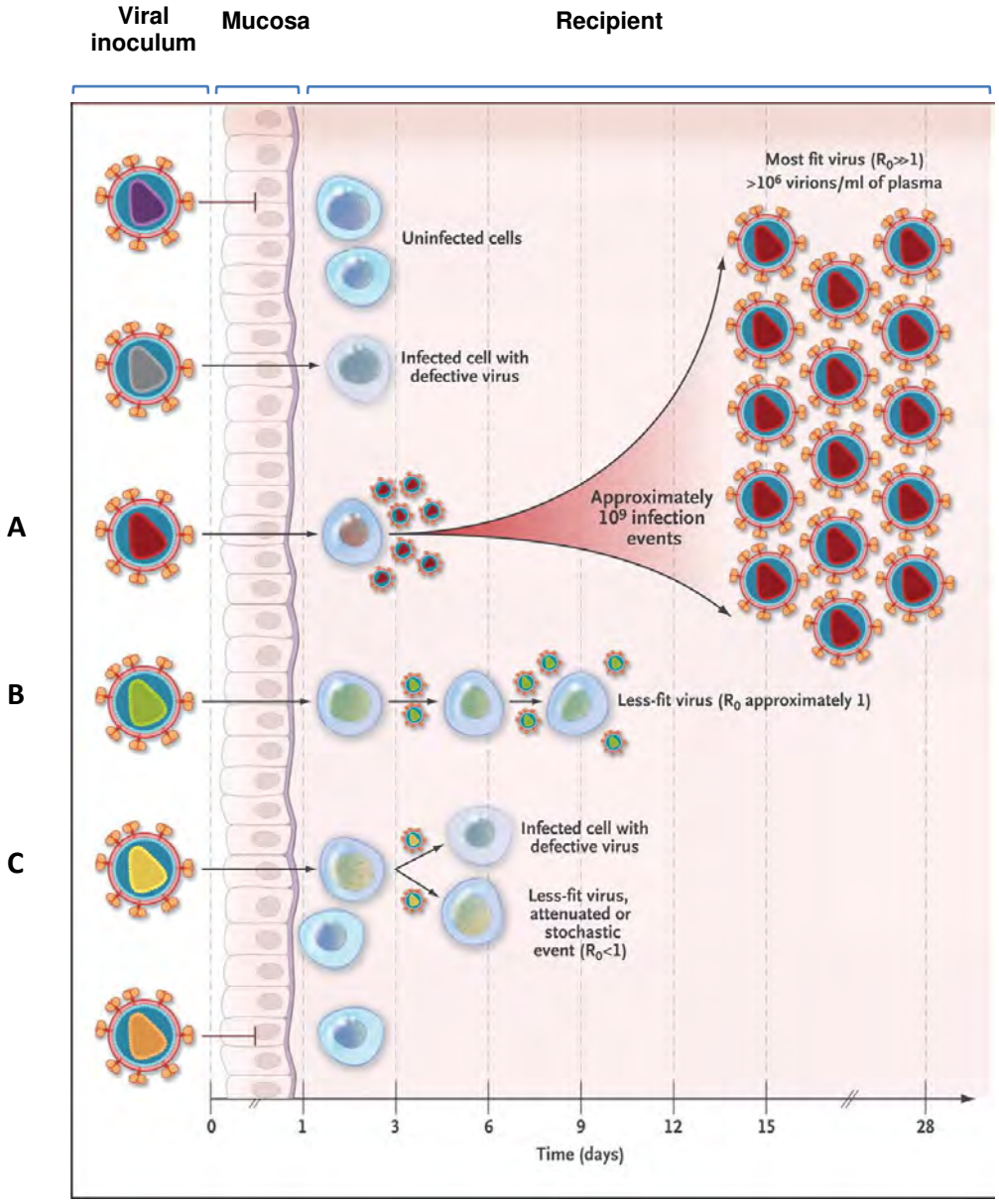


Figure 4. **The HIV transmission bottleneck.** The scenarios depicted are for productive infection (**A**), infection with a virus of lower fitness which propagates but less efficiently (**B**), and abortive infection (**C**) of viruses crossing the mucosal barrier during infection. R_o indicates the reproductive ratio of the transmitted virus, where R_o <1 is indicative of aborted infection. Illustration from Cohen et al. (2011).

This thesis examines a subset of women from the study which enumerated the multiplicity of infection in heterosexual subtype C-infected individuals from South Africa and Malawi, which found that the frequency of infection with two and greater than two variants was too high to be accounted for by random, chance events, suggesting that additional factors were involved (Abrahams et al., 2009). An association between the presence of other sexually transmitted infections (STIs) at transmission and infection with multiple viral variants was identified as one potential mechanism for overcoming the bottleneck (Halaand et al., 2009). Existing STIs may create physical breakages in the mucosal barrier thus allowing easier passage of virus particles, or may stimulate production of inflammatory cytokines which recruit activated HIV target cells to the site of infection. Also, infection with a cell containing multiple virus particles or possibly even host factors such as the semen-derived enhancer of virus infection (SEVI) could be speculated to facilitate establishment of infection with multiple variants (Abrahams et al., 2009, Munch et al., 2007).

Previous studies have proposed that viruses identified in recipients with early infection corresponded to minor variant populations in the blood (Zhu et al., 1993) and genital tract (Boeras et al., 2011) of the donor, and were more closely related to the common ancestor of the donor and recipient sequences combined in each transmitting pair (Sagar et al., 2009). Recently, in a study of Ugandan transmission pairs using Sanger sequencing, cloning and deep sequencing methods, it was shown that viral populations in the recipient more closely resembled those from earlier in infection, and not the virus present close to the time of transmission in the donor (Redd et al., 2012). These studies potentially indicate that properties of transmitted or ancestral viruses exist which make them more transmissible. Most recently, Carlson and associates described, in heterosexual transmission pairs, a biasing of transmission towards viral variants with amino acids which confer enhanced fitness (Carlson et al., 2014). Thus, despite the overwhelming diversity of HIV strains, there may be properties specific to transmitted viruses which need to be taken into consideration for vaccine design.

Distinct phenotypic properties of transmitted viruses reported in multiple studies include shorter Env V1-V2 loop lengths that are less glycosylated, which may facilitate more efficient binding of and entry into target cells but may also confer greater sensitivity to antibody neutralization (Derdyn et al., 2004, Chohan et al., 2005, Liu et al., 2008, Sagar et al., 2009, Wilen et al., 2011). Transmitted viruses have consistently been shown to have CCR5 co-receptor tropism (Keele et al., 2008, Ping et al., 2013); however they infect macrophages poorly due to low levels of CD4 receptor expression on these cells (Salazar-Gonzalez et al., 2009, Ochsenbauer et al., 2012;

Ping et al., 2013). It has also recently been shown that transmitted/founder viruses are more sensitive to inhibition of CCR5 binding using CCR5 antagonists (Parker et al., 2013, Ping et al., 2013). While certain studies have shown transmitted or early viruses to be more sensitive to neutralization than donor viruses (Derdyn et al., 2004), or chronic viruses (Wilen et al., 2011), others have not shown this to be true (Parrish et al., 2012, Ping et al., 2013). Similarly, compared to chronic viruses, transmitted viruses were previously reported to have higher reactivity to the α4β7 integrin on the surface of CD4+ T-cells (Nawaz et al., 2011), however subsequent studies failed to show this difference in binding affinity (Parrish et al., 2012, Perez et al., 2014).

Genetic signatures associated with transmission have been identified near the CCR5 and CD4 receptor binding sites, at amino acid position 12 of the signal peptide and within a single potential N-linked glycosylation site (PNGS) (positions 413-415) of Env in subtype B transmitted viruses (Gnanakaran et al., 2011). The transmission signature within the signal peptide was demonstrated to enhance protein expression, trafficking and incorporation into virus particles, and also proposed to enhance infectivity (Asmal et al., 2011). In keeping with these findings, Parrish et al. (2013) demonstrated, using participant-specific full-length genome infectious molecular clones, that transmitted/founder viruses were 1.7-fold more infectious in TZMbl cells (a HeLa cell-line expressing high levels of CD4 and CCR5) than viruses derived from chronic infection, that they packaged more Env and were more readily taken up by dendritic cells.

Whether or not the genotype of the transmitted virus is a determinant of the rate of disease progression is still uncertain, but convincing evidence exists to suggest that it is including that the replication capacity of transmitted viruses correlated with viral load and that viral load set-point was shown to be a heritable trait (reviewed in Fraser et al., 2014). Alizon et al showed, using phylogenetic methods, that disease progression is a heritable trait strongly associated with viral genotype (Alizon et al., 2010). This is supported by a number of studies on transmission pairs, where viral load set-point in the recipient was associated with viral load in the source partner (Yue et al., 2013; Lingappa et al., 2013; Tang et al., 2004, Hecht et al., 2010, Hollingsworth et al., 2010). Furthermore, the replication capacity of Gag chimera viruses derived from a Zambian transmission pair cohort correlated with viral load in both the recipient and the donor close to the time of transmission (Prince et al., 2012).

Deletions within the viral genome may facilitate viral control, for example, an elite suppressor who was infected with a virus containing a deletion in *nef* became viremic only when the viral

sequence evolved to incorporate a mutation which restored *nef* function (Salgado et al., 2014). Thus genetic properties of transmitted/founder viruses or the very early diversifying virus/es may very likely play a part in determining the severity of disease.

1.4 High viral mutability

HIV sequence diversification following transmission is driven by high viral mutability. HIV-1 is a retrovirus from the genus *Lentivirus*, which reproduces its RNA genome via a DNA intermediate generated by the reverse transcriptase (RT) enzyme. During each cycle of viral replication, the RT enzyme first generates single-stranded cDNA from the RNA template and thereafter double-stranded DNA while misincorporating nucleic acid residues resulting in mutation at an estimated frequency of 3.4 x 10^{-5} mutations per base pair per cycle (Mansky et al., 1995) prior to integration into the host cellular DNA in the nucleus (**Figure 5**).

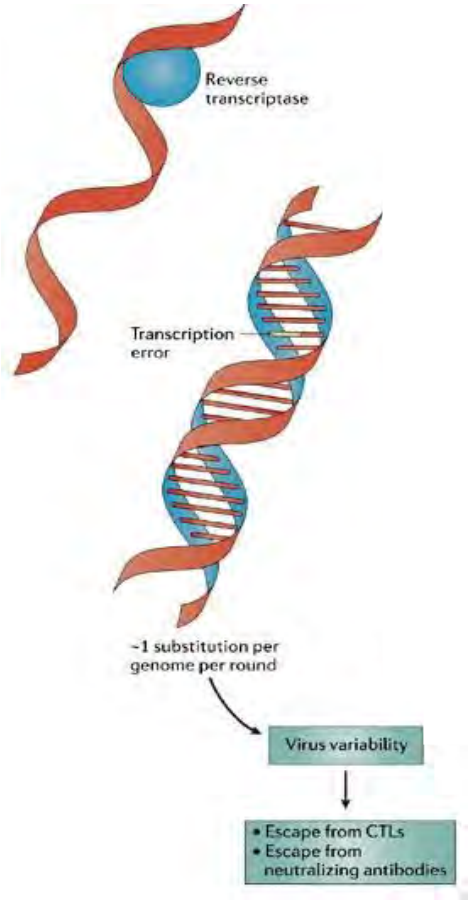


Figure 5. **The error-prone reverse transcriptase of HIV-1**. . Misincroporation of nucleotide base pairs occurs with each cycle of HIV-1 replication as single-stranded RNA is reverse transcribed to cDNA. This introduces variability within participant viral populations and facilitates immune evasion of neutralizing antibody and cytotoxic T-lymphocyte responses. Image taken from Letvin (2006).

The integrated DNA is then transcribed and translated by host cellular enzymes and viral packaging and assembly ensues. In an infected individual, over time, a diverse swarm of closely related but distinct viral populations develops, referred to as a quasispecies (Nowak et al., 1992). The virus is thus able to rapidly alter its genotypic and phenotypic properties through mutation and evade immune responses due to loss of antigen recognition and binding.

This high mutability remains one of the biggest challenges to vaccine design. The rapid genetic divergence and population diversification of the *env* gene over the course of infection was elucidated in a study by Shankarappa et al. (1999). These authors described an initial period of restricted genetic evolution, followed by a linear increase in both divergence from founder and population diversity and finally a plateau of both as the person progressed to AIDS (Shankarappa et al., 1999). A later study examining near full-length genome sequences from acute infection reported a contraction of viral diversity in the first month of infection before an increase in diversity (Herbeck et al., 2011).

Studies approximate an overall 0.55-1% increase in sequence diversity per year within an infected individual (Herbeck et al., 2006, Li et al., 2007, Herbeck et al., 2011). Individual genes within the viral genome, however, diversify at different rates. Of the nine genes encoded by the HIV-1 genome (**Figure 6A**), structural and enzymatic genes *gag* and *pol* are more conserved and diversify slower, compared to *env* and *nef* which diversify much faster (Yusim et al., 2002, Li et al., 2007, Herbeck et al, 2011) (Bansal et al., 2005) (**Figure 6B**). Functional constraints in proteins such as the Gag p24 capsid, a major structural protein, play a role in this conservation (Schneidewind et al., 2008).

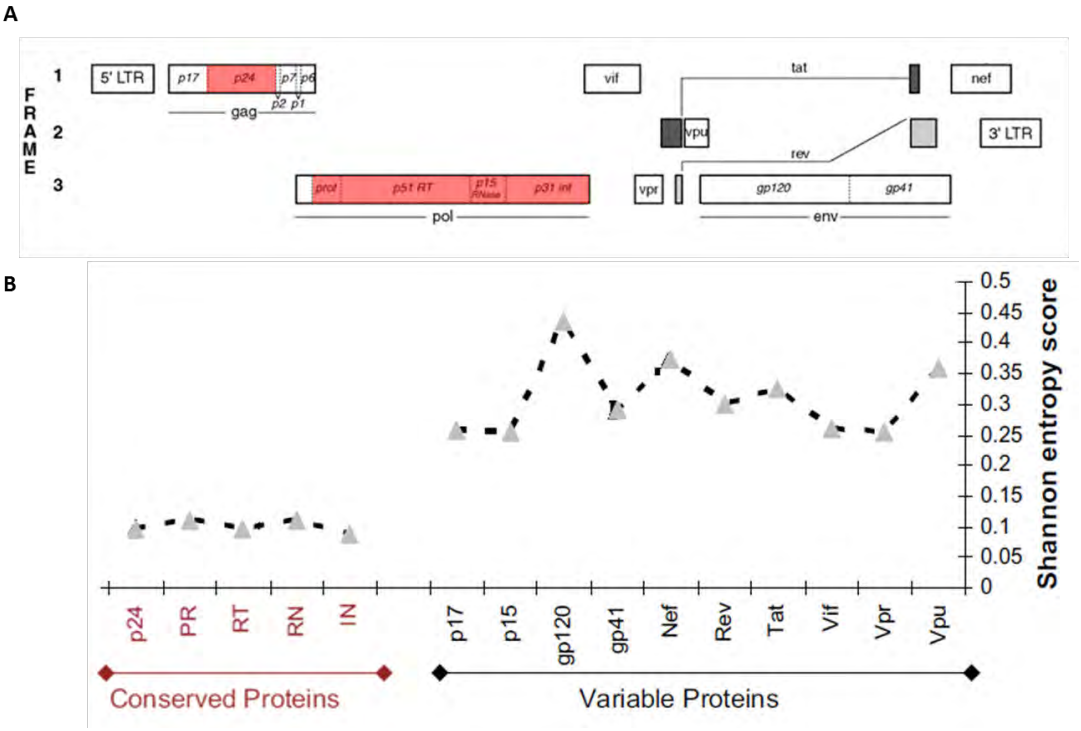


Figure 6. **Conserved and variable regions in the HIV-1 genome**. All nine genes of the viral genome in their respective reading frames are illustrated above (modified from http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html). Where certain genes encode multiple proteins, these are indicated. A graph of Shannon entropy scores for each gene/gene region was obtained from Bansal et al., 2005. Conserved genes (those having Shannon Entropy values below 0.15 (Bansal et al., 2005)) are indicated in red. PR/prot = protease, RT/p51 RT = reverse transcriptase, RN/p15 RNase = RNase and IN/p31 int = integrase.

1.5 Immune responses to HIV-1 infection

Vaccines need to stimulate responses which neutralize the transmitted virus, or control early replication, and clues as to which responses are needed can be gleaned from studies of natural infection. The innate immune responses modulated by dendritic cells and natural killer (NK) cells are the first responses detected following HIV infection, with CD8+ CTLs and non-neutralizing Ab responses evolving as viral load increases, and later autologous neutralizing antibody (nAb) responses develop (**Figure 7**) (reviewed in McMichael et al., 2010). Recently, a subtype B study reported a role for HIV-specific CD4+ T-cell responses in acute infection in slow disease progression (Frater et al., 2014).

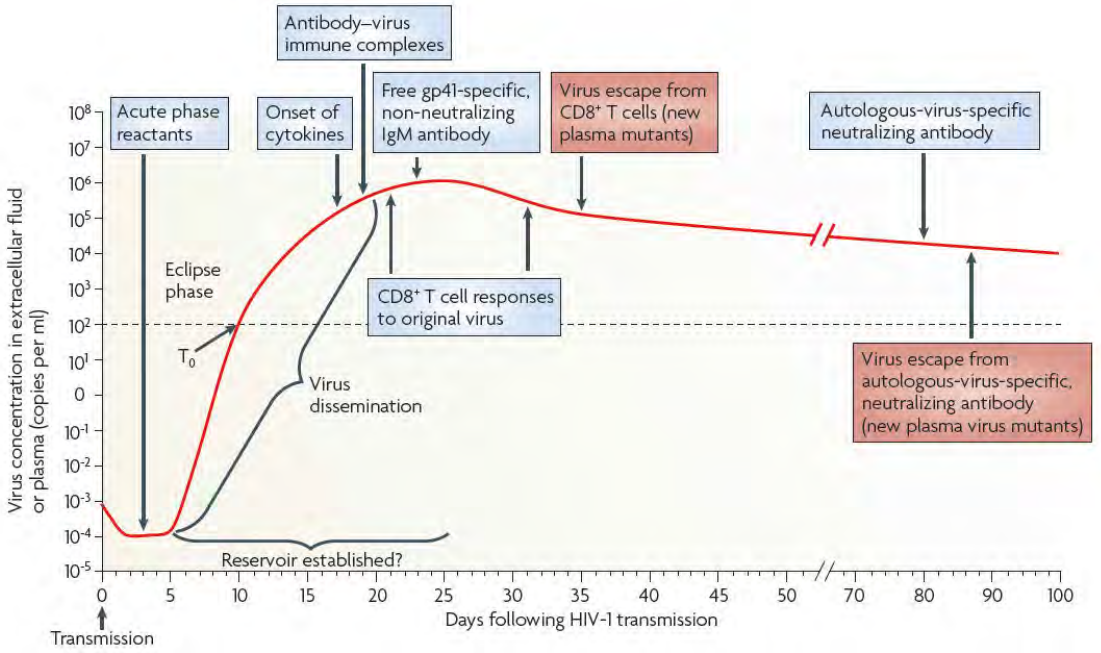


Figure 7. **Immune responses to HIV-1 infection**. Acute phase reactants refer to host antiviral proteins and cytokines. The eclipse phase refers to a period during which no markers of HIV infection are detectable (McMichael et al., 2010). Illustration taken from McMichael et al. (2010).

In natural infection, the earliest Ab responses are constituted by binding Abs such as IgG and IgM which may arise as early as eight days post-infection but do little to reduce virus burden (Tomaras et al., 2008). Abs able to neutralize the virus were initially reported to arise only some weeks or months following infection (Wei et al., 2003; Gray et al., 2007; Yeh et al., 2010), and the virus is able to rapidly escape their responses. However, the presence of low-titre nAbs as early as two weeks following seroconversion has been evidenced by identification of nAb-escape mutations in acute infection (Bar et al., 2012). Antibodies that can neutralize a diversity of viruses, broadly cross-neutralizing (BCN) Abs, are the target responses for vaccines to elicit (see section 1.8.1). These types of responses develop in only a portion of individuals and are evidently too late to afford any benefit to the course of disease (reviewed in Mascola and Haynes 2013).

Immune responses that are proven to reduce viral load in natural infection are those mediated by CD8+ CTLs. CTLs have been shown to reduce peak viral load in acute infection thereby lowering the virus burden to a relatively steady state or set-point (Borrow et al., 1994, Koup et al., 1994, Ogg et al., 1998, Goonetilleke et al., 2009). Viral set-point is predictive of time to onset of AIDS (Lyles et al., 2000).

CTL responses are induced once a target cell becomes infected and the virus's proteins are processed by the cellular proteasome to produce short peptide fragments which are then trafficked via the endoplasmic reticulum (ER) to be presented on the surface of that cell via a human leukocyte antigen (HLA) (York and Rock, 1996). HLA molecules have the ability to recognize short viral peptides, typically nine to eleven amino acids in length and bind these targeted epitopes through specific interactions with amino acid anchor residues (reviewed in Goulder and Walker, 2012). HLA genotype determines which regions of the virus will be recognized and subsequently targeted, for example HLA-B*57/5801 recognizes the epitope TW10 (TSTLQEQIGW). CD8+ T-cell receptors on the surface of CTLs, which also recognize specific residues within the peptides displayed by the HLA, initiate killing of infected cells (**Figure 8**).

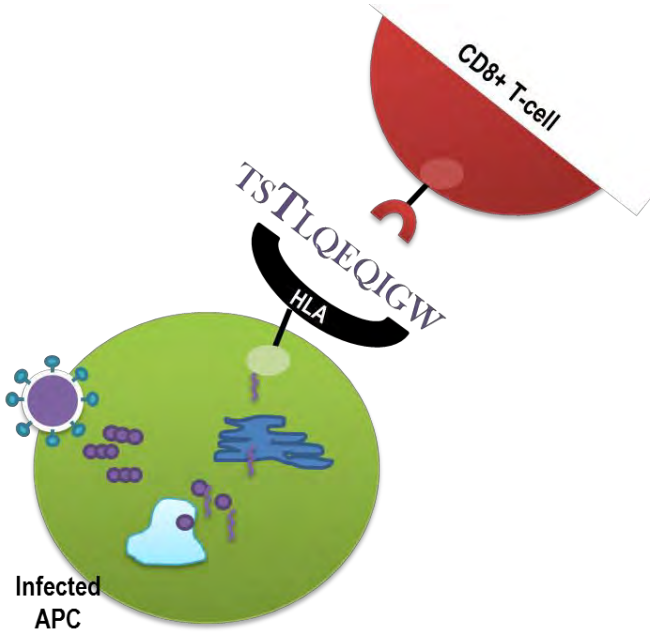


Figure 8. The cytotoxic CD8+ T-lymphocyte response. The presentation of HIV peptide fragments (in this instance the Gag TW10 epitope) to a CD8+ T-cell receptor via HLA binding on the surface of an infected antigen presenting cell (APC) is illustrated.

HIV escapes these immune responses largely due to the virus's mutability. The role of CTL immune selection of nucleotide mutations in viral escape will be discussed in the following section.

1.6 Immune selection and CTL escape

Host immune pressure drives the selection of nucleotide mutations (Moore et al., 2002, Richman et al., 2003, Wei et al., 2003, Rybarczik et al., 2004, Allen et al., 2005, Frost et al., 2005, Carlson et al., 2008, Treurnicht et la., 2010), which are introduced by the viruses error-prone RT enzyme (see section 1.2.4). Nucleotide substitutions, insertions or deletions introduced into or proximal to viral epitopes can result in loss of immune recognition. Mutations which alter amino acid translation are referred to as non-synonymous or non-silent, whereas those which do not are referred to as synonymous or silent. Not all mutations are maintained by the virus, but those which do become fixed within the genome and persist through subsequent replication cycles are referred to as evolving under positive selection.

Non-synonymous mutations identified early in infection have been reported to be more frequently associated with pressure from CTLs (Bernardin et al., 2005, Brumme et al., 2008a, Salazar-Gonzalez et al., 2009, Goonetilleke et al., 2009, Herbeck et al 2011). These mutations facilitate CTL escape as early as the first few weeks of infection (Borrow et al., 1997, Jones et al., 2004, Liu et al., 2006, Gray et al., 2009, Goonetilleke et al., 2009, Wood et al., 2009, Mlotshwa et al., 2010) and are generally characterized by transitions from more common (consensus) to less common (non-consensus) amino acids within or flanking the targeted epitope (Allen et al., 2004).

Mechanisms of evasion from CTL responses include alterations in either peptide binding at HLA anchor residues, peptide recognition by CD8+ T-cell receptors or antigen peptide processing within infected cells (Borrow et al., 1997, Goulder et al., 1997, Feeney et al., 2004, Draenert et al., 2004, Tenzer et al., 2009, Carlson et al., 2012a, Bronke et al., 2013). Continuing the example of the TW10 epitope targeting, these three mechanisms of CTL escape are illustrated in **Figure 9A-C**.

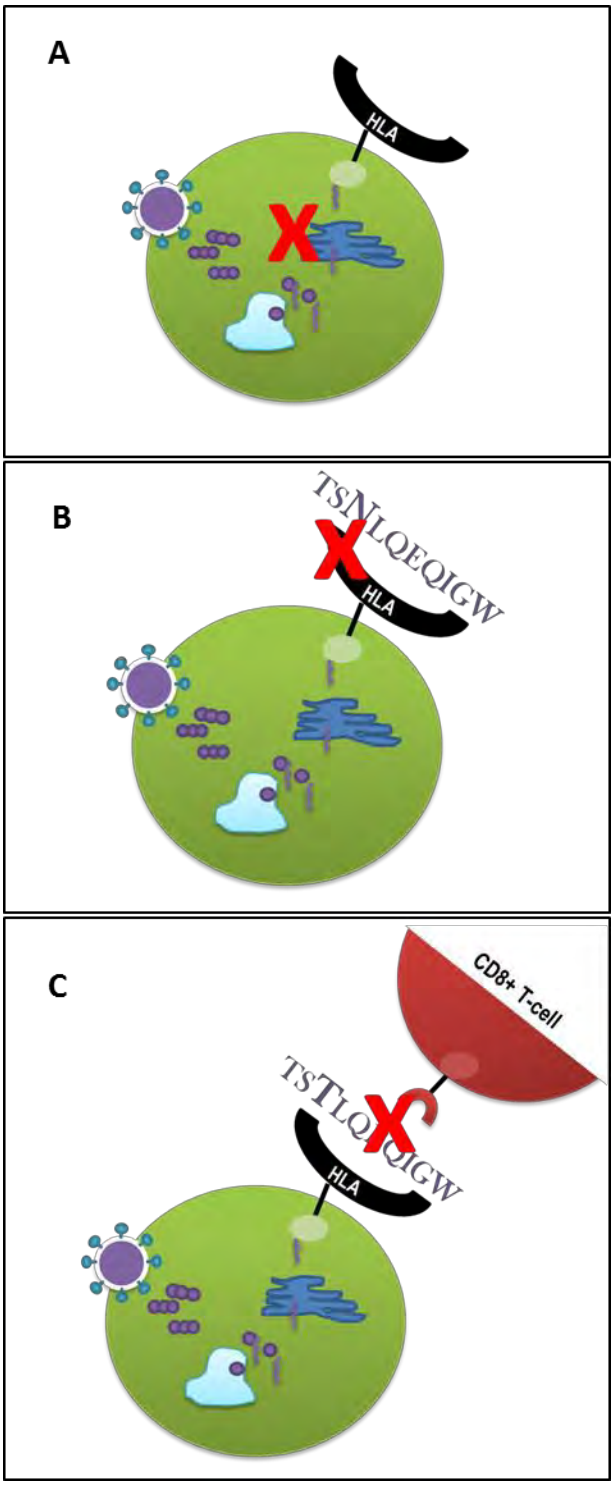


Figure 9. **Mechanisms of viral escape from CTL responses**. Following infection of an antigen presenting cell, CTL-mediated cell killing via HLA presentation of the Gag TW10 epitope to a CD8+ T-cell receptor is prevented by (**A**) altered antigen processing within the cell, (**B**) loss of HLA binding due to amino acid mutation, and (**C**) loss of peptide recognition by the CD8+ T-cell receptor due to amino acid mutation.

Escape mutations can become fixed within a demographic population wherein a specific HLA type is dominant, such as occurred in Japan where amino acid escape mutations within multiple B*51 epitopes have now become the consensus amino acids in circulating viruses that are transmitted within the population (Kawashima et al., 2009). Indeed, a more recent study by Chikata et al. (2014) reported extensive population HLA-restriction influence on the evolution of HIV in Japan, driven by the introduction of HLA polymorphisms. HLA adaption on a population level has also recently been reported for North America, although this was not as extensive (Cotton et al., 2014).

Studies by Liu et al. (2013) and Ferrari et al. (2011) demonstrated that CTL escape occurs more rapidly in high entropy epitopes, and that CTL responses to conserved epitopes arise later but are more effective in viral control. Escape from CTL pressure may come at a cost to the virus since some mutations impede viral protein function and replication (Leslie et al., 2004, Brockman et al., 2007, Troyer et al., 2009, Wright et al., 2011, Prince et al., 2012, Boutwell et al., 2013). This may comprise its ability to survive in a given environment, referred to as viral fitness (reviewed in Domingo et al., 1997). Summarised in **Figure 10** are the possible implications of CTL escape.

An example of this is the escape mutation T242N in the TW10 epitope used in previous illustrations, which results in a significant loss of viral replicative fitness due to its impact on function of the viruses Cyclophilin A binding loop (Martinez-Picardo et al., 2006, Brockman et al., 2007, Song et al., 2012). The virus may compensate for this loss by additional mutations within or adjacent to the TW10 epitope which restores the fitness loss incurred by the initial escape mutation to varying extents (Leslie et al., 2004; Brockman et al., 2007). It is advantageous to identify epitopes containing mutations which reduce viral fitness as these may make good vaccine targets.

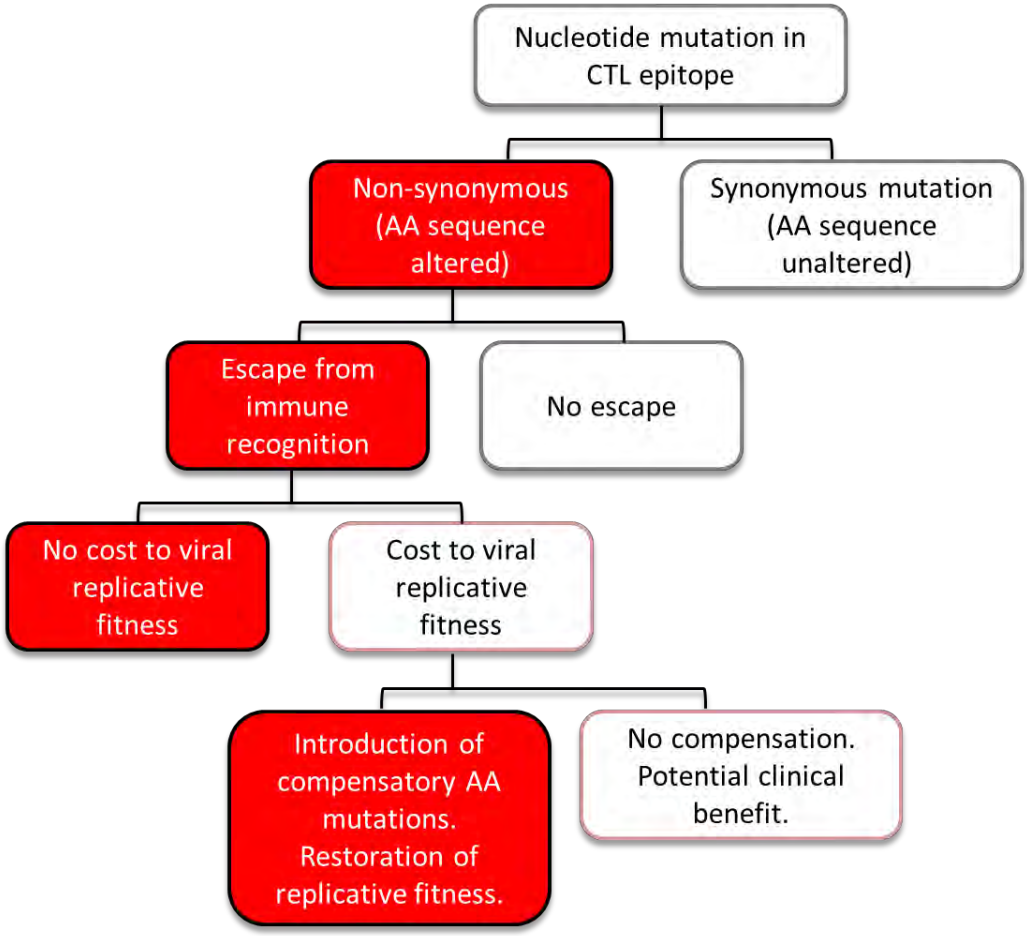


Figure 10. **Implications of immune selection and escape in CTL epitopes.** The eventual impact of CTL escape on control of virus is outlined in a flow diagram. Mutations which result in a cost to viral fitness and subsequent reduced viral replication are of interest for vaccine design.

1.7 CTL responses and viral control

A number of studies have postulated that stable responses to conserved, slow-to-escape epitopes would also be desirable for a vaccine to stimulate. For example, Mlotshwa and colleagues evaluated CTL responses in 53 acutely subtype C-infected individuals and reported that responses were temporal and fluid in nature within each individual, with responses to certain epitopes arising early and being lost early due to escape, and others arising later and being more stable in nature (Mlotshwa et al., 2010). Stable responses were less frequent in individuals with rapid disease progression (Mlotshwa et al., 2010).

It is well established that specific HLA alleles are associated with control of viremia or conversely with rapid progression. HLA alleles B*57, B*58:01, B*27, B*51, and B*81:01, commonly referred to as protective HLA alleles, have been associated with disease control and are enriched in individuals able to control viral loads (Kaslow et al., 1996, Carrington et al., 1999, Hendel et al., 1999, Migueles et al., 2000, Tang et al., 2002a, Tang et al., 2002b, Kiepiela et al., 2004, Pereyra et al., 2008, Leslie et al., 2010, Walker & Yu, 2013), whereas HLA alleles B*58:02 (Kiepiela et al., 2004, Ngumbela et al., 2008), B*35, Cw*04 and B*58:02 have been associated with rapid disease progression (Carrington et al., 1999; Lazaryan et al., 2006).

Identifying the immune targets associated with control of viremia following infection may inform the selection of vaccine immunogens. Protective HLA alleles more frequently target conserved regions of the virus resulting in disease control (Asquith, 2008). In addition, in the major structural protein Gag, escape mutations in epitopes restricted by protective HLA alleles were reported to reduce viral replication capacity (Brockman et al., 2010). Supporting the considerable replicative fitness cost of immune escape in Gag, a study in recently infected individuals showed that mutations in Gag epitopes targeted by protective HLA alleles reverted significantly more rapidly than mutations elsewhere in the gene (Brumme et al., 2008b). In fact, in general, CTL targeting of Gag has been associated with better control of disease regardless of the presence of protective HLA alleles (Zuniga et al., 2006; Brumme et al., 2008b; Pererya et al., 2008; Brennan et al., 2012).

Individuals infected with a virus carrying an immune escape mutation associated with reduced viral replicative capacity from a donor with a protective HLA, but who themselves do not have that protective HLA, was associated with reduced viral load or increased CD4+ counts in recipients (Chopera et al., 2008; Goepfert et al., 2008; Miura et al., 2010). This is not necessarily the case with transmission between individuals with matched protective HLAs since the virus transmitted would already have escaped the responses associated with viral control (Crawford et al., 2009). Wright et al. (2011) further demonstrated that the transmission of HLA-B polymorphisms in Gag to individuals with protective HLA alleles was associated with higher viral loads compared to individuals who did not have protective HLA alleles.

Thus the literature suggests that a vaccine presenting conserved epitopes (with Gag epitopes being the most promising candidates) known to be targeted by protective HLA alleles, to escape slowly and to mediate a loss of viral replication fitness in the event of escape may induce

protective responses in recipients. Indeed, vaccine immunogens representing conserved genome regions have been designed (Letourneau et al., 2007, Rolland et al., 2007) and one such vaccine, HIVconsv incorporating conserved elements of Gag, Pol, Vif and Env from subtypes A, B, C and D, has recently been tested in phase I clinical trials and elicited broad CD8+ T-cell responses which inhibited HIV replication *in vitro* (Borthwick et al., 2014). However, efficacy of such vaccines in preventing infection remains to be demonstrated, particularly since the HVTN502 Step trial, which included Gag as an immunogen, did not elicit protective responses against infection (Esparza et al., 2013). It is possible that the combination, number or sequence of epitopes presented was not sufficiently immunogenic.

1.8 Challenges to HIV-1 vaccine design

Imperative to the design of an HIV vaccine is the identification of immunogens that elicit protective responses. The study of natural infection has provided insights into responses associated with control of disease progression, while vaccine efficacy studies have provided insights into vaccine-induced responses associated with protection from infection. Combatting the extraordinary diversity and adaptability of the virus remains one of the greatest challenges to vaccine development.

1.8.1 Identifying immune correlates of protection: clues from natural infection

Host genetic determinants of disease control are most commonly associated with Class I HLA-mediated targeting of viral epitopes which results in stimulation of specific CD8+ CTL-mediated immune responses. HLA-B and C alleles have been shown to be the strongest determinants of disease progression in human genome-wide association studies (Fellay et al., 2007, Fellay et al., 2009, Pereyra et al., 2010, Bartha et al., 2013), implying a substantive role for CTL responses in controlling the virus (see section 1.6).

Understanding the differences between responses in individuals who are able to naturally control disease, such as elite controllers (these individuals have viral loads that cannot be detected by currently available assays) (reviewed in Walker & Yu, 2013), compared to individuals who do not control replication can provide insights into defining protective immune responses. Studies have shown CTLs from elite controllers to have greater polyfunctionality

(Betts et al., 2006), enhanced proliferative properties (Migueles et al., 2002) and unique activation phenotype which allow them to better destroy infected CD4+ T-cells (Saez-Cirion et al., 2007). Also in elite controllers, CTLs produced higher levels of granzymes and perforin that have cellular cytotoxic effects (Migueles et al., 2002; Migueles et al., 2008; Hersperger et al., 2010).

An important role for Abs in preventing HIV infection has been demonstrated. Monoclonal nAbs prevented infection when passively infused into macaques that were later challenged with virus (Mascola et al., 1999; Shibata et al., 1999, Mascola et al., 2000). Also in animal models, use of monoclonal nAbs has proven effective in immunotherapy to lower viral load (Barouch et al., 2013, Shingai et al., 2013). These antibodies recognize epitopes within the viral envelope which exists as a trimeric structure composed of gp120 and gp41 (transmembrane) subunit proteins (**Figure 11A**). The envelope protein is heavily glycosylated and has variable regions, referred to as V loops, as well as conserved domains. These serve to shield the virus from antibody binding and neutralization (Reitter et al., 1998, Mccaffery et al., 2004, Sagar et al., 2006). A detailed structure of an antibody-bound trimer in its soluble, cleaved form was recently elucidated using cryo-electron microscopy (Julien et al., 2013; Lyumkis et al., 2013) (**Figure 11B**).

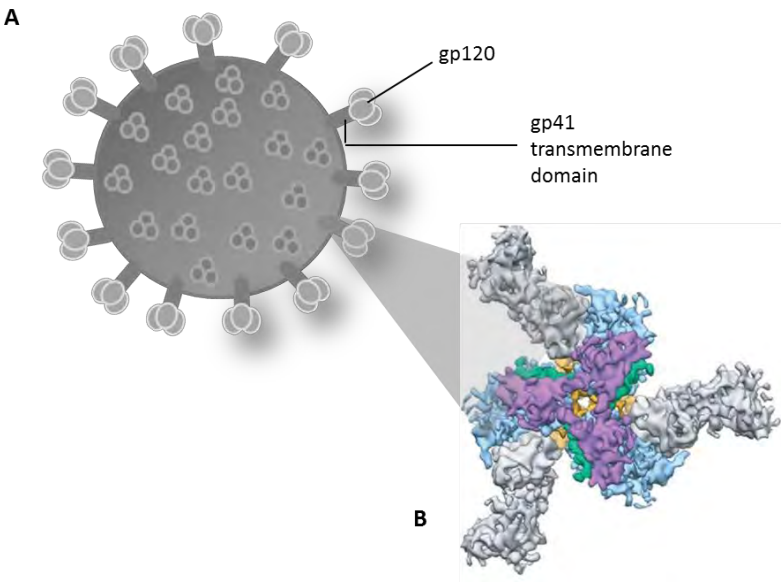


Figure 11. The structure of the HIV-1 envelope trimer. Illustrated are (**A**) an HIV-1 particle coated with *env* gp160 spike trimers with subunit proteins gp120 and gp41 indicated, and (**B**) a top view of the recently reconstructed cryo-EM of the BG505 SOSIP gp140 trimer as determined by Lyumkis et al. with gp120 shown in blue, gp41 in orange, V1V2 in purple, V3 in green, and the bound broadly neutralizing Ab PGV04 is shown in gray (cryo-EM image taken from Lyumkis et al., 2013).

Approximately 20 - 30 % of individuals produce nAbs that neutralize an extensive range of viruses, known as broadly cross-neutralizing (BCN) Abs, and which arise after approximately two years of infection (reviewed in Mascola and Haynes, 2013). Breakthroughs in the isolation of BCN Abs, insights into their ontogeny, as well as in identification of their targets, have been the cause of much excitement in the HIV research field in recent years (Walker et al., 2011, Huang et al., 2014, Moore et al., 2012, Liao et al., 2013, Doria-Rose et al., 2014). BCN targets that have been identified to date include the membrane proximal external region (MPER), the CD4 binding domain, the V1V2 and V3 loops (Mascola and Haynes, 2013).

Development of a vaccine that is able to stimulate the production of BCN Abs would make for an ideal barrier to infection. The challenge to BCN Ab production is that these Abs are unusual in their conformation and are challenging to induce since they require a high number of somatic mutations introduced over time (Mascola and Haynes, 2013). Some of these BCN Abs also have very long CDR H3 arms needed to penetrate the viral glycan shield, such as those isolated from broad-neutralizer CAP256 which target an epitope in the V2 loop (Doria-Rose et al., 2014). Maturation of BCN Abs may be driven by *env* evolution as the virus escapes sequential waves of antibody responses (Liao et al., 2013, Doria-Rose et al., 2014). This process would be extremely challenging to reproduce in the context of vaccination.

1.8.2 Identifying immune correlates of protection: the RV144 trial

Analyses of vaccine participant samples from the RV144 Thai trial provided insights into what types of immune responses HIV vaccines need to stimulate in order to prevent infection. Initial analyses identified high levels of IgG binding Abs targeting epitopes in the Env V1V2 and V3 loop regions as a correlate of protection (Haynes et al., 2012, Karasavvas et al., 2012, de Souza et al., 2012, Gottardo et al., 2013). Conversely, high levels of IgA Abs targeting the C1 region of Env were associated with lack of protection through inhibition of antibody-dependent cellular cytotoxicity (ADCC) (Haynes et al., 2012; Tomaras et al., 2013). These findings may implicate a role for ADCC in vaccine-mediated protection (Bonsignori et al., 2012).

In a sequence analysis study of RV144 breakthrough viruses, two amino acid signature sites in V2 were associated with vaccine protection, one of which is a broadly-neutralizing Ab target (amino acid 169) (Rolland et al., 2012). Abs targeting position 169 were isolated from vaccinees and closer examination of the target site revealed it to be variable in nature (Liao et al., 2013)

indicating a role for cross-reactive responses in protection. The cross-reactive nature of these Abs was confirmed when tested against V1V2 targets from multiple subtypes (Zolla-Pazner et al., 2014).

Taken together, results from the RV144 vaccine trial implicate a role for cross-reactive binding Abs and antibody-dependent cellular responses in mediation of protection against HIV infection. Since this vaccine elicited protection in only one third of vaccinees (Rerks Ngarm et al., 2009), much more is still to be learned about responses that would offer broader protection against infection.

1.9 Study rationale

Understanding the impact of immune pressure on the virus following transmission will inform the development of effective vaccines. This thesis aims to characterize transmitted/founder viruses, how they are moulded by early immune selection, and the effect of immune selection on viral replicative fitness. Viral sequencing was used as a sensitive, indirect method of characterizing immune pressure through the identification of mutations associated with immune selection. Furthermore, it allows the inference of the sequence of transmitted/founder viruses and the construction of research tools such as infectious molecular clones for evaluating the impact of immune escape on viral fitness.

We propose that mapping of virus sequence evolution provides a sensitive tool to detect immune pressure, and that immune escape will alter virus replication.

This is a focused study of five HIV-1 subtype C-infected South African women with differing disease progression, with the following objectives:

1. To derive the sequence of the near full-length transmitted/founder virus and to describe early viral genetic diversification
2. To identify immune selection in sequences from acute and early infection
3. To determine the effect of early CTL-mediated immune selection on replication fitness of transmitted/founder viruses to inform design of CTL-based vaccines

Chapter 2: Identification of near full-length HIV-1 transmitted/founder subtype C virus genomes and their early diversification

2.1 Abstract

An HIV vaccine would need to stimulate immune responses that block transmitted viruses from establishing (founding) clinical infection. Defining the sequences of transmitted/founder viruses and identifying sites evolving under immune pressures associated with early viral control, will aid in vaccine immunogen design.

The number of viral variants that established clinical infection in 26 acutely-subtype C-infected South African women was determined by comparing *env* sequence diversity to a model of random evolution. Early sequence diversification was evaluated for five women (one viremic controller, two rapid progressors and two intermediate progressors) by near full-length genome sequencing over the first six months of infection.

Infection was established by a single transmitted/founder variant in 77% of women, and by two to five variants in the remaining 23% percent. In the expanded study of five women, near full-length genome sequencing confirmed single variant infection in three women (CAP45, CAP210 and CAP239), while two women (CAP63 and CAP85) were found to be infected with a second closely related, minor variant differing at five and seven nucleotide positions respectively. Rates of sequence diversification were modestly higher in the first three months of infection compared to between three and six months. The transmitted/founder virus of a rapid progressor (CAP63), whose CD4+ T-cell counts declined to 200 cells/µl in the first year of infection, displayed the highest rate of diversification compared to the other women.

In conclusion, in line with other studies, we found nearly 80% of infections were a result of a single viral variant. We identified the near full-length transmitted/founder virus sequence of five women, and found that the woman who failed to control viral replication and rapidly progressed to AIDS within the first year of infection, also had the highest rate of viral diversification in early infection.

This study was performed as part of the Center for HIV/AIDS Vaccine Immunology (CHAVI) project, and was part of a larger study to characterize HIV transmission. A portion of this work has been published in Abrahams et al., 2009, Abrahams et al., 2013 and Liu et al., 2013.

2.2 Introduction

An HIV vaccine would need to stimulate immune responses to block the establishment and dissemination of viral infection. It is possible to derive the sequence of the virus that is the founder of clinical infection (transmitted/founder) as the consensus of sequences sampled close to the time of infection, prior to immune selection. Tracking early viral evolution from the transmitted/founder (t/f) virus allows identification of viral adaption to host immune responses and immune selection. This data can provide insights into responses associated with disease control, providing information which could be important to inform the types of responses to be elicited by a vaccine. This chapter utilizes viral sequencing to elucidate the sequence of viruses that established clinical infection in acutely subtype C-infected women and to evaluate early viral diversification and evolution.

In the last five years, HIV transmission studies have improved with respect to the identification of acute infections and the methodology used to characterize genetic properties of early diversifying viruses. These studies have used a combination of (i) early virus sampling, (ii) single genome amplification of viral RNA, and (iii) mathematical modelling to identify the genetic identity of t/f viruses and their early evolution (Salazar-Gonzalez et al., 2008; Keele et al., 2008) and have been validated in mucosal infection models in macaques (Keele et al., 2009; Liu et al., 2010). Each of these three approaches is briefly elaborated on below:

(i) Early virus sampling

Due to the rapid diversification of HIV following infection of a new host, to identify signatures of transmission and early immune selection, it is necessary to sample viruses from very early in infection.

Using plasma donor samples with intense follow-up, Fiebig et al. (2003) devised a system for staging early HIV infection by the sequential appearance of biological markers such as viral RNA and HIV-specific antibodies. The earliest stages of infection were defined as periods during which viral RNA only (stage I) and viral RNA and Gag p24 antigen but not HIV-specific antibodies (stage II) were detectable (Fiebig et al., 2003). Clinical samples from these early stages of infection are most desirable for transmission and early viral evolution studies, since there would be limited immune pressure on the virus and thus the viral sequences generated should be close to the transmitted genotype.

In order to obtain clinical samples from this early in infection, a prospective cohort of uninfected individuals is imperative. This thesis utilized samples from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) 002 Acute Infection (AI) study (see section 2.4), which recruited high-risk women from a prospective HIV negative cohort in Durban who were enrolled once identified as viral RNA or antibody positive following regular screening (van Loggerenberg et al., 2008). In this study infection was detected as early as an estimated two weeks post-transmission prior to the development of antibodies (van Loggerenberg et al., 2008).

(ii) Single genome amplification (SGA)

This method, originally developed by Simmonds et al. (1990) and Edmondson and Mullins (1992) was later modified by Palmer et al. for the screening of antiretroviral drug resistance mutations (Palmer et al., 2005), and was adapted by Salazar-Gonzalez et al. (2008) for application in HIV transmission studies. The approach dilutes cDNA templates such that $\leq$30% of subsequent PCR reactions are positive, which according to Poisson distribution would suggest that each amplicon would be from a single viral cDNA template 80% of the time (Salazar-Gonzalez et al., 2008). Amplification from a single genome template eliminates recombination during PCR, and direct sequencing of PCR products to generate a consensus sequence of the amplicon would control for PCR errors. Further validating this approach, cDNA products of reverse transcription of mixed RNA templates generated *in vitro* showed no evidence of recombination between templates (Salazar-Gonzalez et al., 2008). As a result, false signals of virus diversity are eliminated and sequence data are considered a true reflection of the sampled viral population. The SGA method has revolutionized HIV acute infection studies and has since been widely used in the field (Keele et al., 2008, Abrahams et al., 2009, Salazar-Gonzalez et al., 2009, Halaand et al., 2009, Bar et al., 2010, Li et al., 2010, Masharsky et al., 2010).

(iii) Mathematical modelling of early viral diversification

Utilizing sequences from early infection, the number of viruses causing clinical infection and the approximate period of infection for an individual can be determined. A mathematical model was developed by Keele et al. (2008) to identify the number of transmitted viruses in an infected individual. It incorporates the error rate of reverse transcriptase (2.16×10^{-5}) (Mansky et al., 1995) as well as the doubling time of HIV (approx 2 days) (Markowitz et al., 2003) and assumes no changes exist as a result of immune selection. If a single virus is transmitted, the resulting viral population following infection will be homogeneous/low diversity, having a star-like

phylogeny and a Poisson distribution of nucleic acid mutations (each mutation is a random and independent event) (Slatkin and Hudson, 1991). If this is violated, then more than one variant is thought to cause infection, or alternatively early immune selection may be driving sequence heterogeneity. Elucidation of multiplicity of infection is supported by phylogenetic methods which extrapolate time from infection based on derivation of a most recent common ancestral sequence (Drummond et al., 2006; Drummond and Rambaut, 2007).

This approach was initially applied to subtype B infection of men who have sex with men (MSM) and heterosexual populations in the USA, and subsequently to heterosexual subtype C populations in South Africa and Malawi (Abrahams et al., 2009), and subtype A- and C-infected populations in Zambia (Halaand et al., 2009). Together these studies showed that approximately 80-90% of infections result from a single t/f virus (Keele et al., 2008; Abrahams et al., 2009; Halaand et al., 2009).

Viral diversification is driven in part by immune selection and may thus occur at different rates within infected individuals. Early reports have described relationships between rate of viral diversification and the rate of progression of clinical disease (Wolfs et al., 1990; Lukashov et al., 1995; Wolinsky et al., 1996; Ganeshan et al., 1997; Liu et al., 1997; McDonald et al., 1997; Shioda et al., 1997; Markham et al., 1998), however there have only been a limited number of studies that have utilized cognate near full-length t/f virus sequences. This approach enables a more accurate approach to mapping early immune selection, and is of greater relevance to vaccine design.

This chapter describes the identification of single t/f viral infections in 26 women from the CAPRISA 002 AI cohort through *env* sequencing. Some of this work has been published as part of the larger South Africa/Malawi multiplicity of infection study (Abrahams et al., 2009). A subset of five women was selected for elucidation of near full-length genome t/f virus sequences and characterization of early viral evolution. These women were classified as having differential clinical disease progression: two women were rapid disease progressors, one woman was a viremic controller and finally two women were intermediate to these categories.

2.3 Research Aim and Objectives

Aim: To derive the sequence of the full-length t/f virus from acutely subtype C-infected women with differing disease progression and characterize early viral diversification.

Objective 1: To identify women from the CAPRISA 002 AI cohort with single variant infection and generate near full-length genomes from five women following infection (two to five weeks post-infection), at three and at six months post-infection

Objective 2: To determine if there are differences in early viral evolution and expansion of viral population diversity in individuals with different disease progression profiles

2.4 Rationale

This project formed part of a larger study by the Center for HIV/AIDS Vaccine Immunology (CHAVI) characterizing the HIV-1 transmission bottleneck and t/f viruses through *env* sequencing. At the time of this study, there were limited studies elucidating early diversification of full-length HIV-1 genomes, and no studies on HIV-1 subtype C viruses – which may differ from subtype B due to differences in host genetic and environment. Furthermore, South Africa and women in general carry the highest burden of HIV disease, this study will further our knowledge of acute and early infection in South African women.

2.5 Methods and Materials

A total of 515 *env* SGA sequences from 26 CAPRISA 002 AI participants used in the multiplicity of infection analysis in this chapter, were part of the larger CHAVI study published in Abrahams et al., 2009 (**Appendix 1**). PCR amplification of near full-length genomes was carried out by F. Truernicht (University of Cape Town), and sequencing was performed as part of this thesis. Of the 112 near-full length genome sequences generated 58 were generated as part of this thesis, and 54 were contributed by R. Thebus, J. Marais, F. Treurnicht and S. Goodier (University of Cape Town).

Clinical and STI data was provided by CAPRISA (K. Mlisana, University of KwaZulu-Natal), and HLA data was provided by C. Gray (University of Cape Town).

2.5.1 Study participants

Samples were obtained from the CAPRISA 002 AI study cohort, a prospective negative cohort of high-risk women recruited in Durban, South Africa (van Loggerenberg et al., 2008). Women were followed either monthly or three monthly, and were identified as HIV-infected based on two rapid HIV antibody tests, an HIV EIA and an HIV-1 RNA test (van Loggerenberg et al., 2008). Date of infection was estimated as the mid-point date between the last negative and first positive HIV antibody test. Where women were HIV antibody negative but viral RNA positive, the date of infection was taken as 14 days prior to the first viral RNA positive date. Four-digit high-resolution Human Leukocyte Antigen (HLA) typing was used to determine participant HLA-A, B and C types as described in Chopera et al. (Chopera et al., 2008). The study was approved by ethics committees from the University of Cape Town, Witwatersrand and KwaZulu-Natal. Informed consent was obtained from all participants.

Viral loads and CD4+ T-cell counts were determined using COBAS Amplicor v1.5 vRNA assay (Roche Diagnostics, Rotkreuz, Switzerland) and the BD FACSCalibur™ Flow Cytometer (BD Biosciences, San Jose, USA). Staging of infection, according to Fiebig et al (2003), was carried out on plasma using the COBAS Amplicor v1.5 vRNA assay (Roche Diagnostics, Rotkreuz, Switzerland), Determine AntiHIV-1/2 3rd Generation EIA (Abbott Laboratories, Illinois, USA), EIA BEP 2000 (Dade Behring, Marburg, Germany) and GS HIV-1 Western blot analysis kit (Bio-Rad, Washington, USA). The stages were as follows: (I) viral RNA positive only, (II) viral RNA and p24 antigen positive, (III) EIA antibody positive but Western blot negative, (IV) indeterminate Western blot, (V) Western blot positive without p31 (integrase) band, and (VI) Western blot positive with p31 band.

2.5.2 RNA extraction and cDNA synthesis

Viral RNA was extracted from a volume of 200 µl of plasma using the Qiagen Viral RNA Mini Kit (Qiagen, Valencia, CA, USA) according to manufacturers' instructions. For low viral load or difficult to amplify samples, a volume of 400 µl of plasma was used. Reverse transcription of viral RNA to cDNA was performed using the Superscript III reverse transcriptase and Oligo$(dT)_{20}$ (Invitrogen, GmbH, Karlsruhe, Germany) (or HIV-specific primer OFM-19 (5'-GCACTCAAGGCAAGCTTTATTGAGGCTTA-3') in the case of *env*) in a 100 µl final reaction

volume. Reaction reagents were prepared in two separate premixes. Per reaction volumes and reaction cycling parameters are below:

Premix 1:

Reagent	Volume (µl)
RNA	50
dNTPs (20 mM)	2
Oligo-dT or specific primer (50 µM)	2
Total	**54**

Premix 2:

Reagent	Volume (µl)
5X Buffer	20
100 mM DTT	5
H_2O	14-15
RNase OUT (40 U/µl)	2
Superscript III RT (200 U/µl)	4-5
Total	**46**

Premix 1 was incubated at 65°C for 5 minutes in a PCR cycler. The temperature of the cycler heating block was then reduced to 45°C at which point premix 2 was briefly warmed on the heating block before adding the 46 µl to premix 1. The final 100 µl reaction was incubated at 45°C for 2 hours to allow for reverse transcription of full-length product, followed by 70°C for 15 minutes to inactivate the RNase inhibitor enzyme RNase OUT (Invitrogen, GmbH, Karlsruhe, Germany). Thereafter a volume of 1 µl of RNase H (Invitrogen, GmbH, Karlsruhe, Germany) enzyme was added and incubated at 37 °C for 20 minutes to degrade all RNA-DNA complexes so that only single-stranded cDNA remained.

2.5.3 Single Genome Amplification (SGA) and sequencing of env

SGA of *env* was carried out as described by Salazar-Gonzalez et al (2008) and Keele et al (2008) using Platinum Taq DNA Polymerase High Fidelity (Invitrogen, GmbH, Karlsruhe, Germany). Amplification was performed by nested PCR using primers VIF-1 (5'-GGGTTTATTACAGGGACAGCAGAG-3') and OFM-19 (5'-GCACTCAAGGCAAGCTTTATTGAGGCTTA-3') (first round), and ENV-A (5'-GGCTTAGGCATCTCCTATGGCAGGAAGAA-3') and ENV-N (5'-CTGCCAATCAGGGAAGTAGCCTTGTGT-3') (second round). cDNA was serially diluted in order to determine at which dilution $\leq$30% of PCR replicate reactions were positive. Once this dilution was determined, the number of replicate PCR reactions was expanded in order to produce the desired number of SGA amplicons. A negative control reaction (no cDNA template)

was included as a contamination control. PCR reagent mixes (per reaction) and cycling parameters are indicated below.

1st round PCR reagent mix and cycling conditions:

Reagent	Volume (µl)
dH_2O	15.3
10 X High Fidelity Buffer	2
$MgSO_4$ (50 mM)	0.8
dNTPs (10 mM)	0.4
VIF-1 (20µM)	0.2
OFM-19 (20µM)	0.2
Platinum Taq High Fidelity	0.1
cDNA	1
Total	**20**

Temperature (°C)	Incubation Time	Number of Cycles
94	2 min	
94	15 sec	35
55	30 sec	
68	4 min	
68	20 min	
4	Hold	

2nd round PCR reagent mix and cycling conditions:

Reagent	Volume (µl)
dH_2O	15.3
10 X High Fidelity Buffer	2
$MgSO_4$ (50 mM)	0.8
dNTPs (10 mM)	0.4
ENV-A (20µM)	0.2
ENV-N (20µM)	0.2
Platinum Taq High Fidelity	0.1
1st round product	1
Total	**20**

Temperature (°C)	Incubation Time	Number of Cycles
94	2 min	
94	15 sec	45
55	30 sec	
68	4 min	
68	20 min	
4	Hold	

PCR products (3-5 µl) were electrophoresed on a 1% agarose gel alongside a 1 kb DNA ladder marker (Promega, Madison, USA). Full-length *env* amplicons were approximately 3 000 base pairs in size. PCR products were directly sequenced using the ABI 3000 genetic analyzer (Applied Biosystems, Foster City, CA, USA) and BigDye terminator reagents. Sequences with

chromatograms double peaks (indicating amplification of more than one template) were excluded. Likewise sequences with deletions that were > 100 base pairs when compared to the derived intraparticipant consensus were excluded.

2.5.4 Single Genome Amplification (SGA) and sequencing of near full-length genomes

SGA of near full-length genomes was carried out as described by Salazar-Gonzalez et al (2009) using the Expand Long Template Taq PCR system (Roche Diagnostics, Rotkreuz, Switzerland). Amplification was performed by nested PCR using primers described by Rousseau et al (2006). Primers 1U5C 5'-GGGTGAGTGCTCTAAGTAGTGTGTGCCCGTCTGT-3', 1U5Cb 5'-GGGTGAGTGCTCTAAGTAGTGTGTGCCCATCTGT-3' and 1.3.3pIC 5'-GGGACTTAGAGCACTCAAGGCAAGCTTTATTG-3' (first round) and primers 2U5C 5'-GGCCGCGGATCCAGTAGTGTGTGCCCGTCTGTTGTGTGACT-3' and 2.3.3pIC 5'-GGCCGCGCGGCCGCTAGAGCACTCAAGGCAAGCTTTATTGAGGCTTA-3' (second round) were used. cDNA was serially diluted in order to ascertain at which dilution $\leq$30% of PCR replicate reactions were positive. Once this dilution was determined, the number of replicate PCR reactions was expanded in order to produce the desired number of SGA amplicons. PCR reagent mixes (per reaction) and cycling parameters are indicated below.

1st round PCR reagent mix and cycling conditions:

Reagent	Volume (µl)
dH_2O	41.45
10 X Buffer 1 (with Mg^{++})	5
dNTPs (20 mM)	0.9
Primer 1U5C (50µM)	0.3
Primer 1U5Cb (50µM)	0.3
Primer 1.3.3pIC (50µM)	0.3
Expand High Fidelity polymerase	0.75
cDNA	1
Total	**50**

Temperature (°C)	Incubation Time	Number of Cycles
94	2 min	
94	10 sec	10
68	30 sec	
68	8 min	
94	10 sec	20
68	30 sec	
68	8 min*	
68	20 min	
4	Hold	

*An additional 20 sec of template elongation time was added on per cycle

2nd round PCR reagent mix and cycling conditions:

Reagent	Volume (µl)
dH_2O	41.75
10 X Buffer 1 (with Mg^{++})	5
dNTPs (20 mM)	0.9
Primer 2U5C (50µM)	0.3
Primer 2.3.3pIC (50µM)	0.3
Expand High Fidelity polymerase	0.75
1st round product	1
Total	**50**

Temperature (°C)	Incubation Time	Number of Cycles
94	2 min	
94	10 sec	10
68	30 sec	
68	8 min	
94	10 sec	20
68	30 sec	
68	8 min*	
68	20 min	
4	Hold	

*An additional 20 sec of template elongation time was added on per cycle

The PCR cycler heating block was heated to 80°C before inserting PCR tubes so as to allow for more specific primer binding. PCR products (3-5 µl) were electrophoresed on a 1% agarose gel alongside a 1 kb DNA ladder marker (Promega, Madison, USA). Near full-length amplicons were approximately 9 000 base pairs in size. PCR products were purified using ExoSAP-IT (USB, Cleveland, USA) according to manufacturer instructions and directly sequenced. All products were directly sequenced using the ABI 3000 genetic analyzer (Applied Biosystems, Foster City, CA, USA) and BigDye terminator reagents.

Amplicons with up to six ambiguous base pair positions were accepted. While the stipulated criteria for SGA sequences is zero base pair ambiguities for the *env* gene (Salazar-Gonzalez et al., 2008), near full-length genomes without any base ambiguities were rare. Salazar-Gonzalez et al. (2009) accepted up to five ambiguous base pair positions per near full-length genome when sequencing amplicons obtained at <20% PCR positivity, attributing these to polymerase Taq error when considering the fidelity of the Taq enzyme rather than a consequence of amplification of mixed template populations. Sequences with deletions that were >6 000 base pairs in length were accepted for inclusion in data analysis.

2.5.5 Limiting dilution amplification (LDA)

Due to difficulties with near full-length genome amplification which included low viral load plasma samples and limited volume of plasma samples (particularly for the viremic controller in

this study), sequence numbers were supplemented by including PCR amplicons obtained at >30% reaction positivity generated as per the protocol indicated in section 2.4.3 above. Seven percent of these contained more than six base pair ambiguities.

2.5.6 Derivation of the transmitted/founder (t/f) virus sequence

Sequences from the earliest time-point post-infection (2-5 weeks post-infection for the five women in this study) were aligned using BioEdit version 7.0.8.0 (Hall, 1999). A consensus of these sequences for each participant was generated by determining the majority nucleic acid base identity at each position along the genome. The t/f derivation for near full-length sequences was further dependent on all open reading frames (ORFs) of the consensus generated being intact (Salazar-Gonzalez et al., 2009) and on the *env* ORF matching that derived for t/f *env*-only data for that participant.

2.5.7 Sequence analyses

Sequences were subtyped using the REGA HIV Subtyping Tool (http://dbpartners.stanford.edu/RegaSubtyping/). Sequence diversity was evaluated using pair-wise DNA distance matrices and Neighbour-Joining phylogenies generated in MEGA 4 (Tamura et al., 2007), as well as pair-wise hamming distances. Branch lengths (the number of nucleotide substitutions per site) were estimated using Treerate (http://www.hiv.lanl.gov/content/sequence/TREERATEv2/treerate.html). Nucleotide base synonymous and non-synonymous mutations were visualized using *Highlighter* plots (http://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter). Sequences were assessed for the presence of host restriction enzyme APOBEC-driven G to A hypermutation using the *Hypermut* tool (http://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermnut.html). Potential N-linked glycosylation sites were identified using *N-GlycoSite* (http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html). CTL epitope prediction, based on participant HLA information and identification of anchor residues within *env* sequences were identified using Motif Scan (http://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan). Evidence of recombination in *env* sequences was identified using GARD (www.datamonkey.org/GARD/)

(Kosakovsky Pond et al., 2006), RAP Beta version (www.hiv.lanl.gov) and RDP version 3.27 (http://darwin.uvigo.es./rdp/rdp.html) (Martin et al., 2005).

The time to most recent common ancestor (tMRCA) of *env* sequences was determined by N. Wood (University of Cape Town) using BEAST (Bayesian Evolutionary Analysis Sampling Trees) v1.4.7 (Drummond et al., 2006; Drummond and Rambaut, 2007) with the following parameters: a relaxed, uncorrelated exponential molecular clock with general time-reversible substitution model, gamma distribution with four categories and a proportion of invariant sites, substitution rates unlinked across codon sites at 2.16 x 10^{-5} substitutions/site/generation (Mansky et al., 1995). Relative substitution rates were determined using HyPhy (Pond et al., 2005; Keele et al., 2008).

Distribution of *env* sequence hamming distances and conformance to a star-like phylogenetic tree structure was determined by E. Giorgi (Los Alamos National Laboratory, New Mexico).

2.5.8 Statistical testing

Student's T-tests and one-way ANOVA test were performed in GraphPad Prism version 5 (www.graphpad.com).

2.6 Results

2.6.1 Multiplicity of infection of women in the CAPRISA 002 AI cohort

In order to determine the number of t/f variants that established clinical infection in 26 women from the CAPRISA 002 AI cohort, Durban, South Africa, a total of 515 full-length *env* sequences were generated using single genome amplification and direct sequencing of amplicons. An average of 20 *env* SGA sequences was generated per participant from the earliest time point post-infection. These women were estimated to be infected for two to five weeks based on clinical testing (see Methods and Materials, section 2.5.1). Estimates of recent infection were confirmed by staging of plasma for the presence of biological markers as described by Fiebig et al. (2003), which revealed that seven women were in stage I/II (viral RNA positive only), three were in stage III (EIA antibody positive but Western blot negative), four in stage IV (indeterminate Western blot), ten in stage V (Western blot positive without p31 (integrase band)

and finally two in stage VI (Western blot positive with p31 band). Although peak viremia occurs in Fiebig I/II prior to detectable antibody responses, this is not evident in this cohort possibly because viral load was still increasing for some women at the time of sampling (**Figure 12**).

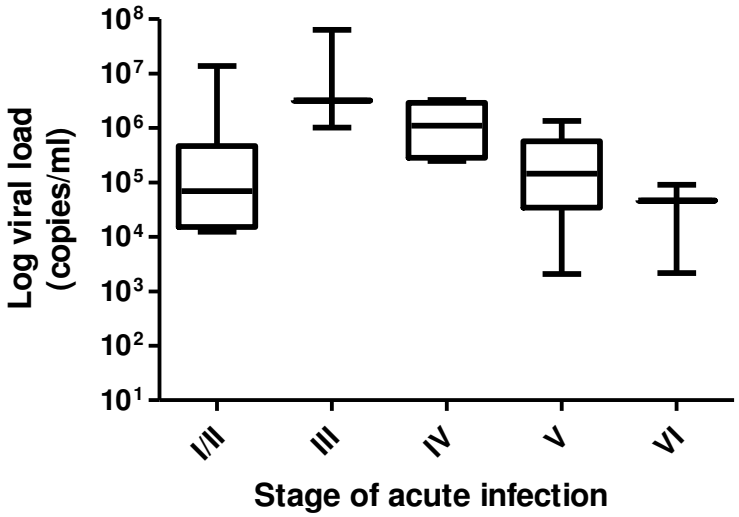


Figure 12. Staging of acute HIV infection and early viremia. The range of log viral load values for women in the six stages of acute infection defined by the appearance of early clinical markers (Fiebig et al., 2003).

In order to determine the number of t/f variants that established clinical, intraparticipant *env* sequences were analyzed for conformance to the following criteria: (i) sequence diversity based on pair-wise DNA distance and structure of a Neighbour-Joining phylogeny, (ii) the distribution of hamming distances between sequences, and (iii) the time to the most recent common ancestral sequence (tMRCA) based on a Bayesian analysis of sequence evolution. Sequences having high homogeneity, a star-like phylogeny, a Poisson distribution of hamming distances, and a tMRCA conferring with the clinical estimate of time of infection were classified as conforming to the model of random evolution from a single t/f virus (Keele et al., 2008).

Intraparticipant sequences from thirteen of the 26 women conformed to these criteria and were classified as infected with a single variant (**Table 1A, Figure 13A**). Sequences from a further seven women displayed low-level diversity (0.05-0.26% DNA distance) and violated one or more of the criteria for single t/f variant infection due to the presence of shared mutations or APOBEC hypermutation (**Table 1B, Figure 13B**). For these women, the mean tMRCA in days exceeded the clinically estimated days post-infection by a factor of two to eight. Five women

conformed to the model after removal of APOBEC mutations and for a further two women (CAP8 and CAP269) shared mutations could be explained by immune pressure. In the case of CAP8, these shared mutations were situated within a Class I HLA-B*08:01-restricted epitope and were classified as CTL-mediated immune pressure. In the case of CAP269, amino acid insertions resulted in the lengthening of the V1 loop and the introduction of multiple potential N-linked glycosylation sites and were therefore associated with Ab-mediated selection.

Sequences from the remaining six women displayed high-level diversity and did not conform to the model criteria. These women were classified as infected with multiple (two-five) t/f variants (**Table 1C**). Intraparticipant DNA distances of sequences from these women ranged from 0.4 to 1.29%, and the structure of the sequence phylogenies displayed clustering of unique variants (**Figure 12C**). Mean tMRCA in days exceeded the clinically estimated days post-infection by 33 to 400-fold. Recombination between variants was identified in four of the six women.

2.6.2 Sexually transmitted infections and multiplicity of infection

A relationship between sexually transmitted infections (STIs) close to the time of transmission and the number of variants transmitted was investigated. STI data from the time of study enrolment (two to nine weeks post-infection) were available from CAPRISA. Twenty-four of the 26 women had a current STI (**Table 2**). Only two women had genital ulcers at the time of examination.

All women were Herpes Simplex Virus 2 (HSV2) antibody positive, however only one woman, CAP220, who was infected with a single t/f variant, was viral PCR positive indicating active HSV2 infection. *Bacterial vaginosis* was the most common STI, detected in 73% percent of women. Fisher's categorical tests were used to investigate an association between having an STI and single or multivariant infection. No statistical difference was detected between the frequency of individuals with either *Neisseria gonorrhoea* (p=0.2184), *Bacterial vaginosis* (p=1), *Trichomonas vaginalis* (p=0.4154) or *Mycoplasma genitalium* (p=1) in the two groups. The average number of STIs in the single and multivariant-infected groups was equal (n=1).

Table 1A. **Women with single t/f *env* variant infection**

Participant ID	Stage of infection*	Sampling date	Viral load (copies/ml)	Total *env* sequences	Mean DNA distance (%)	BEAST mean number of days since MRCA (95% CI)	Clinical test date estimated time post-infection (days)
CAP45	I/II	20-Apr-05	12 500	16	0.02	18 (4-35)	14
CAP129	IV	13-Jun-06	1 800 000	19	0.03	22 (6-42)	17
CAP174	V	22-Sep-05	40 000	21	0.06	33 (15-54)	14
CAP177	I/II	07-Mar-06	359 000	20	0.03	17 (5-33)	14
CAP188	I/II	25-Jan-07	13 800 000	22	0.04	22 (7-40)	14
CAP200	IV	11-Oct-05	398 000	18	0.04	26 (13-45)	14
CAP206	V	15-Jun-05	196 000	21	0.05	29 (15-49)	28
CAP210	I/II	03-May-05	468 000	21	0.01	11 (3-25)	14
CAP217	IV	02-Dec-05	3 260 000	20	0.02	16 (4-38)	16
CAP220	V	15-Feb-07	2 070	15	0.08	45 (17-75)	23
CAP221	I/II	02-Mar-06	24 300	21	0.07	42 (23-73)	14
CAP237	III	06-Mar-07	1 020 000	22	0.04	26 (9-46)	13
CAP239	V	10-Aug-05	95 800	24	0.06	34 (11-58)	36

Table 1B. Women with low diversity, single t/f *env* variant infection not conforming to the model of random evolution

Participant ID	Stage of infection*	Sampling date	Viral load (copies/ml)	Total *env* sequences	Mean DNA distance (%)	BEAST mean no. days since MRCA (95% CI) w/ APOBEC	BEAST mean no. days since MRCA (95% CI) w/o APOBEC	Clinical test date estimated time post-infection (days)	Star-like phylogeny	Poisson Distribution	Proposed reasons for not conforming to the model
CAP8	V	11-May-05	207 000	19	0.18	147 (67-261)	N/A	18	No	No	CTL- selection
CAP40	VI	25-Apr-06	90 500	22	0.18	129 (68-208)	102 (42-168)	27	No	No	APOBEC-3G
CAP63	III	06-Jan-05	3 210 000	19	0.06	33 (17-55)	30 (10-53)	15	Yes	Yes	APOBEC-3G
CAP84	V	16-Feb-05	559 000	22	0.06	36 (17-64)	26 (8-49)	10	No	Yes	APOBEC-3G
CAP85	V	22-Jun-05	621 000	21	0.15	106 (50-181)	53 (20-101)	36	Yes	No	APOBEC-3G
CAP225	III	25-Oct-05	63 600 000	20	0.05	39 (20-68)	25 (12-46)	14	No	Yes	APOBEC-3G
CAP269	VI	27-Sep-06	2 170	18	0.26	205 (110-325)	N/A	25	No	Yes	Ab- selection

Table 1C. Women with high diversity, heterogeneous multi-variant *env* infection

Participant ID	Stage of infection*	Sampling date	Viral load (copies/ml)	Total *env* sequences	Number of transmitted variants	Recombination between transmitted variants	Mean DNA distance (%)	BEAST mean no. days since MRCA (95% CI)	Clinical test date estimated time post-infection (days)
CAP37	IV	11-May-06	248 000	20	3	Yes	1.18	5998 (2192-12321)	15
CAP69	I/II	31-Jan-06	15 300	20	5	Yes	1.29	2449 (1059-4622)	14
CAP136	V	29-Jun-06	85 300	16	2	Yes	0.37	605 (221-1208)	12
CAP222	I/II	05-Apr-06	69 700	21	3	No	0.79	1390 (512-2767)	14
CAP224	V	15-Nov-06	1 348 000	19	2	No	0.4	687 (257-1363)	21
CAP260	V	08-Feb-06	17 600	18	2	Yes	1.23	1702 (750-3083)	16

*Stage of infection according to presence of clinical markers described by Fiebig et al. (2003)

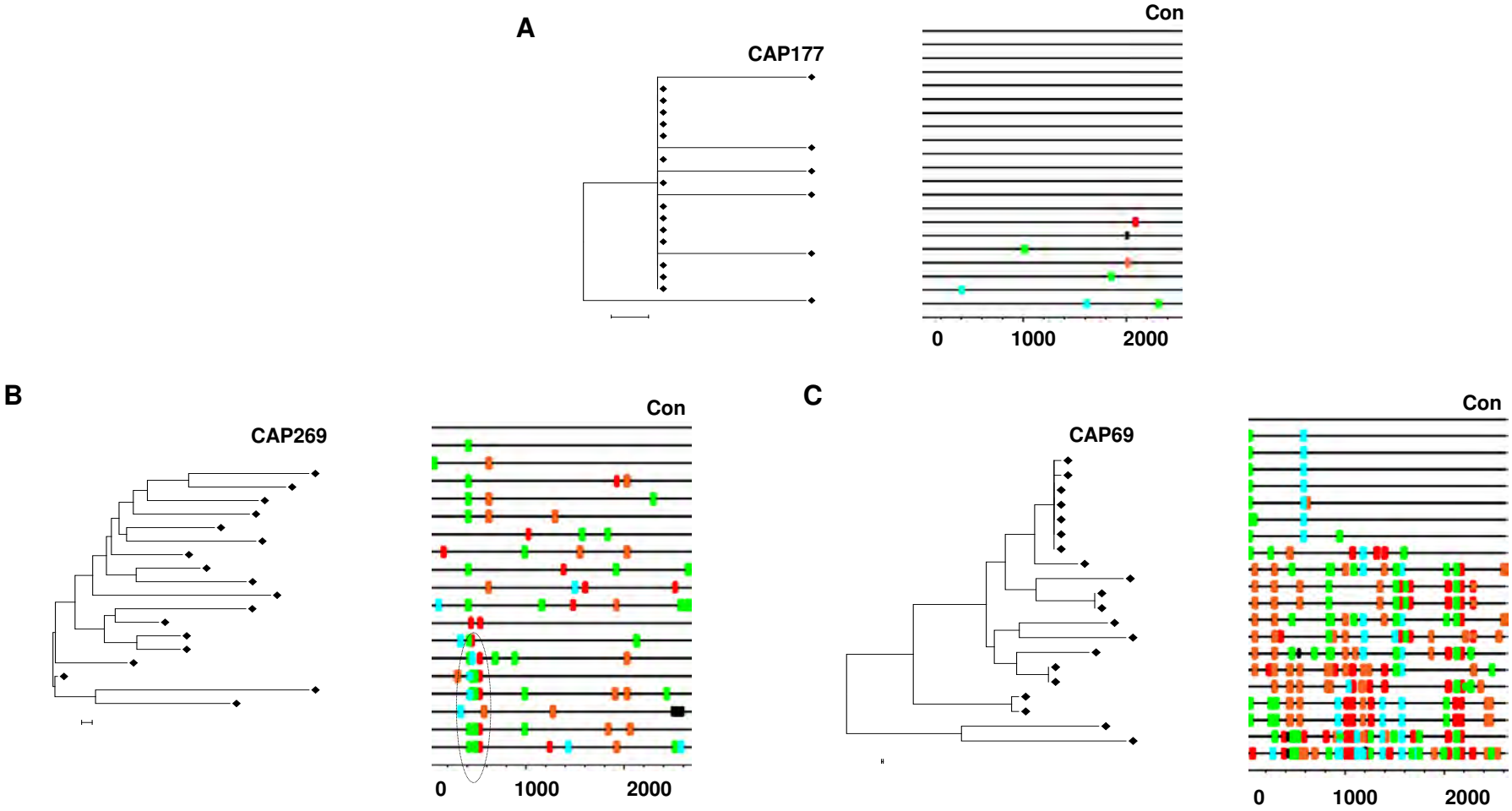


Figure 13. *Env* t/f multiplicity of infection. The diversity of *env* sequences sampled from two to five weeks post-infection for three women from the CAPRISA 002 AI cohort is illustrated by Neighbour-Joining phylogenies (left) and *Highlighter* plots of nucleotide differences to the derived t/f consensus sequence (www.hiv.lanl.gov) (right). Green=A, Red=T, Orange=G, Light Blue=C, Black=gaps. Represented are: (**A**) an homogeneous single t/f variant infection, (**B**) a single t/f variant infection with evidence of Ab-mediated immune selection in the form of amino acid insertions in the V1 loop (encircled), and (**C**) a multivariant t/f infection (five variants with recombination). The tree scale bar = 0.0001.

Table 2. Sexually transmitted infections in 26 women from the CAPRISA 002 AI cohort

Participant ID	Estimated time post-infection (weeks)	HIV-1 t/f variants	Sexually transmitted infection*
CAP8	3	Single	*C. trachomatis*
CAP37	6	Multiple	*B. vaginosis, N. gonorrhoea*
CAP40	3	Single	*T. vaginalis, M. genitalium*
CAP45	5	Single	*B. vaginosis*
CAP63	4	Single	*B. vaginosis*
CAP69	5	Multiple	*N. gonorrhoea*
CAP84	3	Single	*B. vaginosis, C. trachomatis*
CAP85	5	Single	*B. vaginosis*
CAP129	4	Single	*B. vaginosis*
CAP136	2	Multiple	*B. vaginosis*
CAP174	4	Single	*B. vaginosis*
CAP177	4	Single	*B. vaginosis, N. gonorrhoea*
CAP188	4	Single	*B. vaginosis*
CAP200	6	Single	*B. vaginosis*
CAP206	8	Single	*B. vaginosis*
CAP210	5	Single	
CAP217	9	Single	*B. vaginosis, M. genitalium*
CAP220	3	Single	HSV2 PCR+
CAP221	5	Single	*B. vaginosis, N. gonorrhoea*
CAP222	6	Multiple	*T. vaginalis*
CAP224	5	Multiple	*B. vaginosis*
CAP225	3	Single	*B. vaginosis*
CAP237	4	Single	
CAP239	5	Single	*B. vaginosis*
CAP260	2	Multiple	*B. vaginosis, M. genitalium*
CAP269	4	Single	*B. vaginosis*

**C. trachomatis* = *Chlamydia trachomatis*; *B. vaginosis* = *Bacterial vaginosis*; *N. gonorrhoea* = *Neisseria gonorrhoea*; *T. vaginalis* = *Trichomonas vaginalis*; *M. genitalium* = *Mycoplasma genitalium*; HSV2 = Herpes simplex virus

2.6.3 Selection of five women for near full-length genome characterization

Five participants CAP45, CAP63, CAP85, CAP210, and CAP239 were selected for further characterization based on their recent infection with a single t/f variant and differential disease progression profiles. Demographic information for each of the women is provided in **Table 3**. Participants CAP45 and CAP210 were Ab negative, viral RNA positive (stage I/II) at the first sampling date, while participants CAP63 and CAP85 were Ab positive but did not yet have the full spectrum of reactive HIV-specific Abs (stage III and V). For participant CAP239, due to limited sample availability, staging of infection was carried out on the sample taken three weeks subsequent to the earliest sampled time-point. However, since the stage of infection for the later sample was classified as stage V, we can extrapolate that this participant was likely to be in the very early stages of infection at the first sampled time-point. All participants harboured viruses with low *env* sequence diversity (mean DNA distance 0.06%) and a calculated mean tMRCA of 34 days (see Table 1A).

Disease progression was defined based on viral loads and CD4+ T-cell trajectory in the first year of infection. Participant CAP45 was a viremic controller with viral loads consistently below 2 000 copies/ml and CD4 counts above 350 cells/µl (Pereyra et al., 2008), while participants CAP63 and CAP210 were rapid progressors based on CD4 counts below 350 cells/µl and viral loads of >100 000 copies/ml on two consecutive visits in the first year of infection (Mlotshwa et al., 2010, Moore et al., 2009) (see **Figure 14**).

For participant CAP63 disease progression was particularly rapid with referral to anti-retroviral therapy (based on prevailing guidelines of therapy initiation at counts <200 cells/µl) (National Department of Health South Africa, 2004) within nine months post-infection. Participants CAP85 and CAP239 did not fit criteria for either rapid progression or viremic control and were classified as intermediate progressors.

Table 3. Demographic and viral sequencing information for five selected women from the CAPRISA 002 AI cohort

Participant ID	Age	Disease Progression[1]	Fiebig Stage[2]	HLA Type	Sample date	Weeks post-infection[4]	Whole (half) genome sequences[5]
CAP45	41	Viremic controller	I/II	A*23:01, 29:02	20-Apr-05	2	3
				B*15:10, 45:01	11-May-05	5	6
				Cw*06:02, 16:01	28-Jun-05	12	3
					27-Jul-05	16	1
CAP63	32	Rapid	III	A*02:01, 23:01	06-Jan-05	2	11
				B*45:01	09-Mar-05	11	7
				Cw*04:01, 16:01	13-Jul-05	29	10
CAP85	24	Intermediate	V	A*30:02	22-Jun-05	5	8
				B*08:01, 45:01	18-Aug-05	13	9
				Cw*07:01, 16:01	07-Dec-05	29	7
CAP210	43	Rapid	I/II	A*68:02	03-May-05	2	9
				B*15:10	13-Jun-05	12	7
				Cw*03:04	19-Oct-05	26	11
CAP239	44	Intermediate	nd[3]	A*01:01, 29:02	19-Jul-05	2	2 (3)
			V*	B*42:01, 58:01	10-Aug-05	5	8
				Cw*06:02, 17:01	21-Sep-05	11	2
					07-Dec-05	22	9

[1]According to CD4 count/viral load in the first year of infection

[2]Fiebig stages (Fiebig et al., 2003) determined for first sequenced time-point: (I/II) HIV RNA positive but antibody negative; (III) ELISA positive but non-reactive Western blot; (V) reactive Western blot without p31 band

[3]Not determined

[4]Time from mid-point date between first positive and last negative HIV EIA test to date of sampling, or time since first HIV RNA positive date plus 14 days in antibody negative individuals

[5]Whole genome sequence numbers include smaller genomes no less than 6 000 bp in size

*Fiebig stage determined for five weeks post-infection sample.

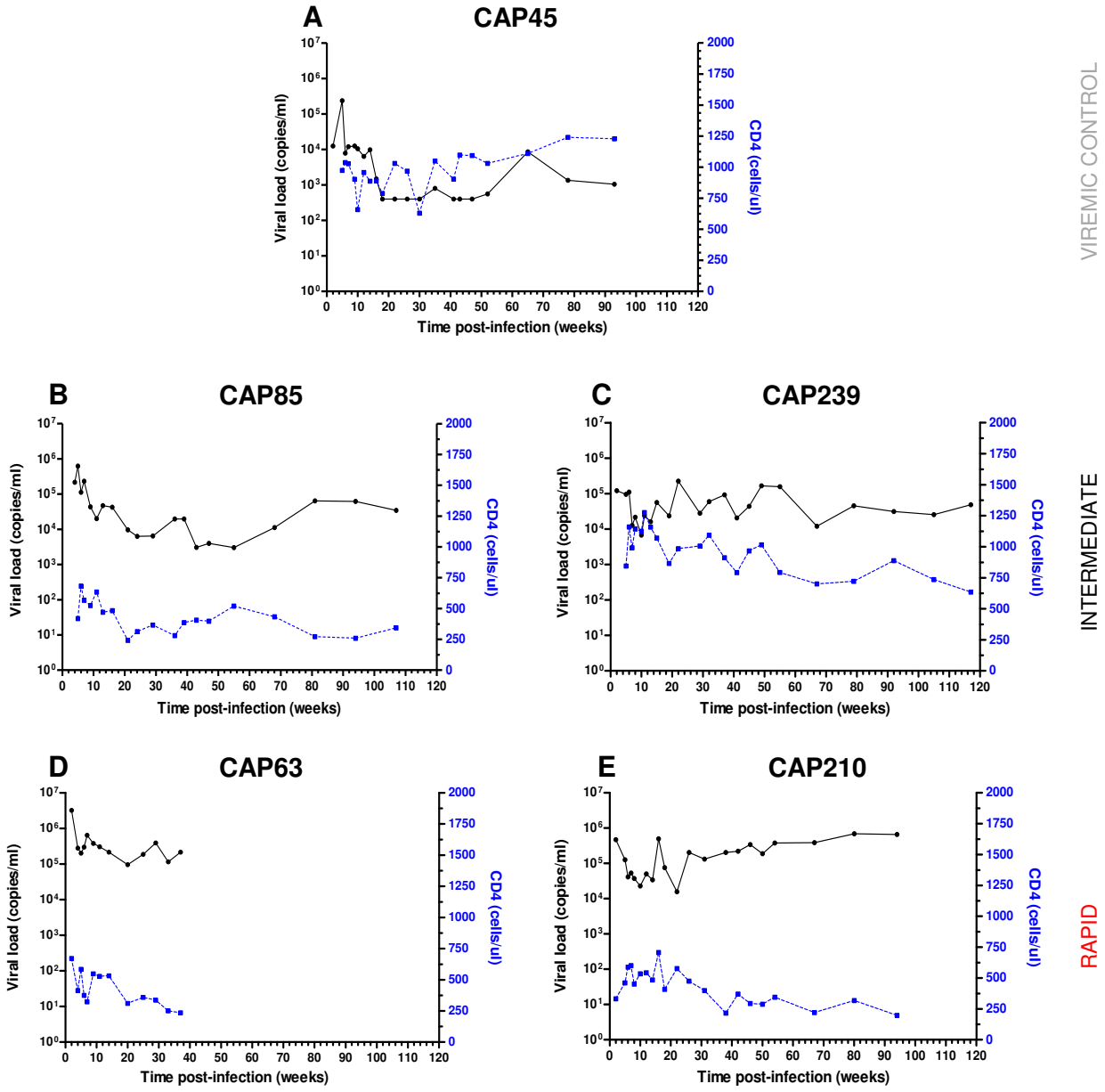


Figure 14. **Clinical disease progression in five selected women.** Viral load (black lines) and CD4+ T-cell count (blue lines) trajectory over 120 weeks of infection for five women from the CAPRISA 002 AI cohort, participants CAP45 (**A**), CAP85 (**B**), CAP239 (**C**), CAP63 (**D**) and CAP210 (**E**). Disease progression profile categories are indicated on the right.

2.6.4 Derivation of near full-length genome t/f sequences

To derive the sequence of the t/f virus and characterize early viral diversity, a total of 113 near full-length genomes were generated from time-points corresponding to CAPRISA 002 AI study screening/enrolment (two to five weeks post-infection; average n=9 sequences per participant), three months (11-13 week post-infection; average n=6 sequences per participant) and six months (22-29 weeks post-infection; average n=9 sequences per participant) (see Table 3). Due to low viral loads for viremic controller participant CAP45, no sequences could be generated at six months post-infection for this individual.

Sequences were all HIV-1 subtype C across the genome and intra-participant sequences grouped together on a phylogenetic tree (**Figure 15**). The median intra-participant viral population diversity at the earliest sampled time-point was 0.03% (range 0.008 to 0.257%). As expected, participants sampled earlier in infection (Fiebig stage I/II, CAP45 and CAP210), displayed lower intra-participant pair-wise DNA distances (0.016% and 0.008% respectively) compared to participants sampled later in infection (Fiebig stage III and V for CAP 63 and 85 respectively), whose intra-participant population diversity was 0.032% and 0.257% respectively. Finally, for participant CAP239 who was classified as stage V at five weeks post-infection, population diversity was 0.047% based on the only two sequences sampled at the earliest time-point (two weeks post-infection). Low numbers of sequences sampled at the earliest time-point for participants CAP45 (n=3) and CAP239 (n=2) may have resulted in an under-estimation of diversity.

Using methodology described by Salazar-Gonzalez et al. (2009), a consensus of sequences sampled at the earliest time-point was extrapolated and assumed to be the t/f viruses for each of the five women. A limited number of non-SGA sequences, obtained from limiting dilution PCR, were included in derivation of the t/f consensus since any nucleotide ambiguities in these sequences would represent a minority at a given position and thus their impact would be negated through consensus generation. For example, in the case of viremic controller CAP45 for whom a limited number of sequences (n=3) could be generated, shared ambiguity between two sequences at one nucleotide position was resolved by the third sequence which had no ambiguity at this site. Thus all positions in the derived t/f sequences were determined unambiguously.

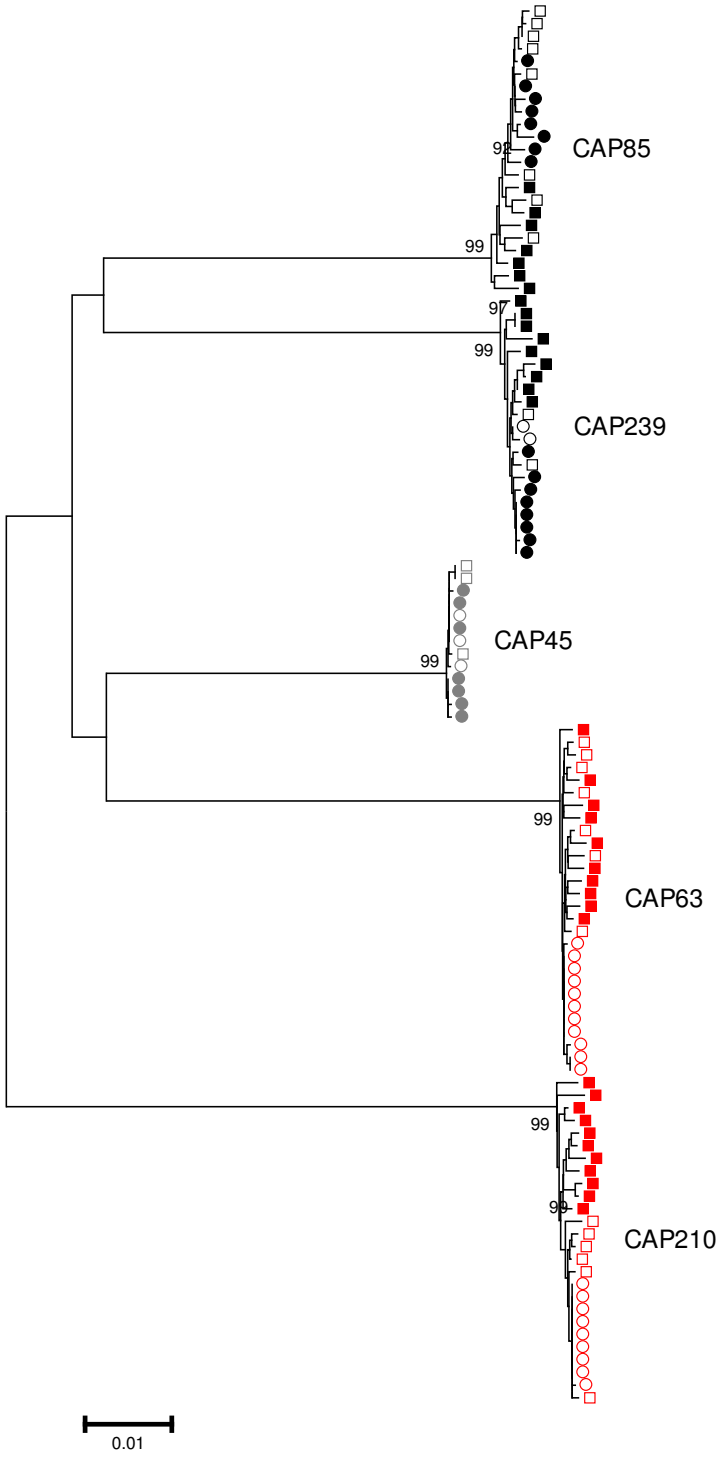


Figure 15. Neighbour-Joining tree of HIV-1 near full-length genome sequences from five subtype C-infected women from the CAPRISA 002 AI cohort. Clinical disease progression is indicated by colour: viremic control (grey), intermediate progression (black) and rapid progression (red). Sequences from two weeks post-infection are indicated by open circles, five weeks post-infection by solid circles, 11-13 weeks post-infection by open squares, and 22-29 weeks post-infection by solid squares.

As with the *env*-only analyses, near full-length genome sequences from CAP45, CAP210 from the earliest time-point and CAP239 from five weeks post-infection (the time-point used for the *env*-only study) displayed a star-like phylogeny (**Figure 16-18**). No shared mutations were present in earliest sequences from CAP45 and CAP210; however shared mutations potentially associated with immune selection were identified in CAP239 (**Figure 18**). These participants were thus confirmed to be infected with a single t/f full-length genome viral variant. For participants CAP63 and CAP85, although *env*-only sequencing indicated infection with a single t/f variant, analysis of near full-length genome sequences revealed that each of these individuals was infected with a second very closely related variant (**Figure 17**). In the case of CAP63, the two variants differed at five nucleotide positions, none of which were in *env*. In the case of CAP85, the two variants differed at seven nucleotide positions, of which three were situated in *env*. These minor variants were not detected at later sequenced time-points for either participant.

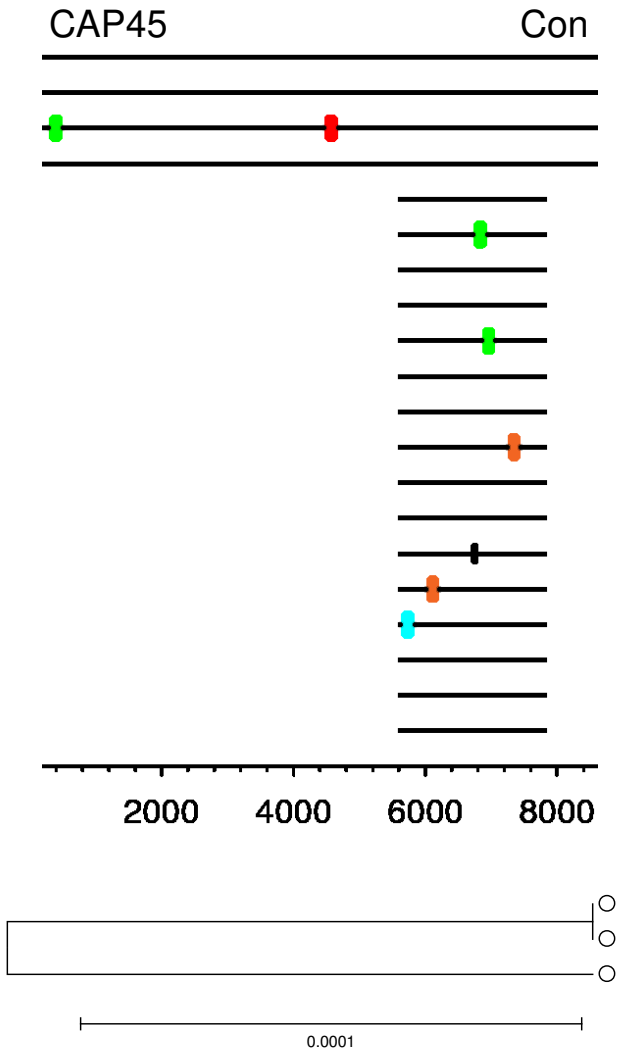


Figure 16. Near full-length genome t/f virus derivation for CAP45. The diversity of near full-length genome sequences sampled from two weeks post-infection for participant CAP45 is illustrated by a Neighbour-Joining phylogeny (below) and a *Highlighter* plot of nucleotide differences to the derived t/f consensus sequence (www.hiv.lanl.gov) (top). Green=A, Red=T, Orange=G, Light Blue=C, Black=gaps. *Env* SGA sequences from two weeks post-infection, used for identification of multiplicity of infection are included. The tree scale bar = 0.0001.

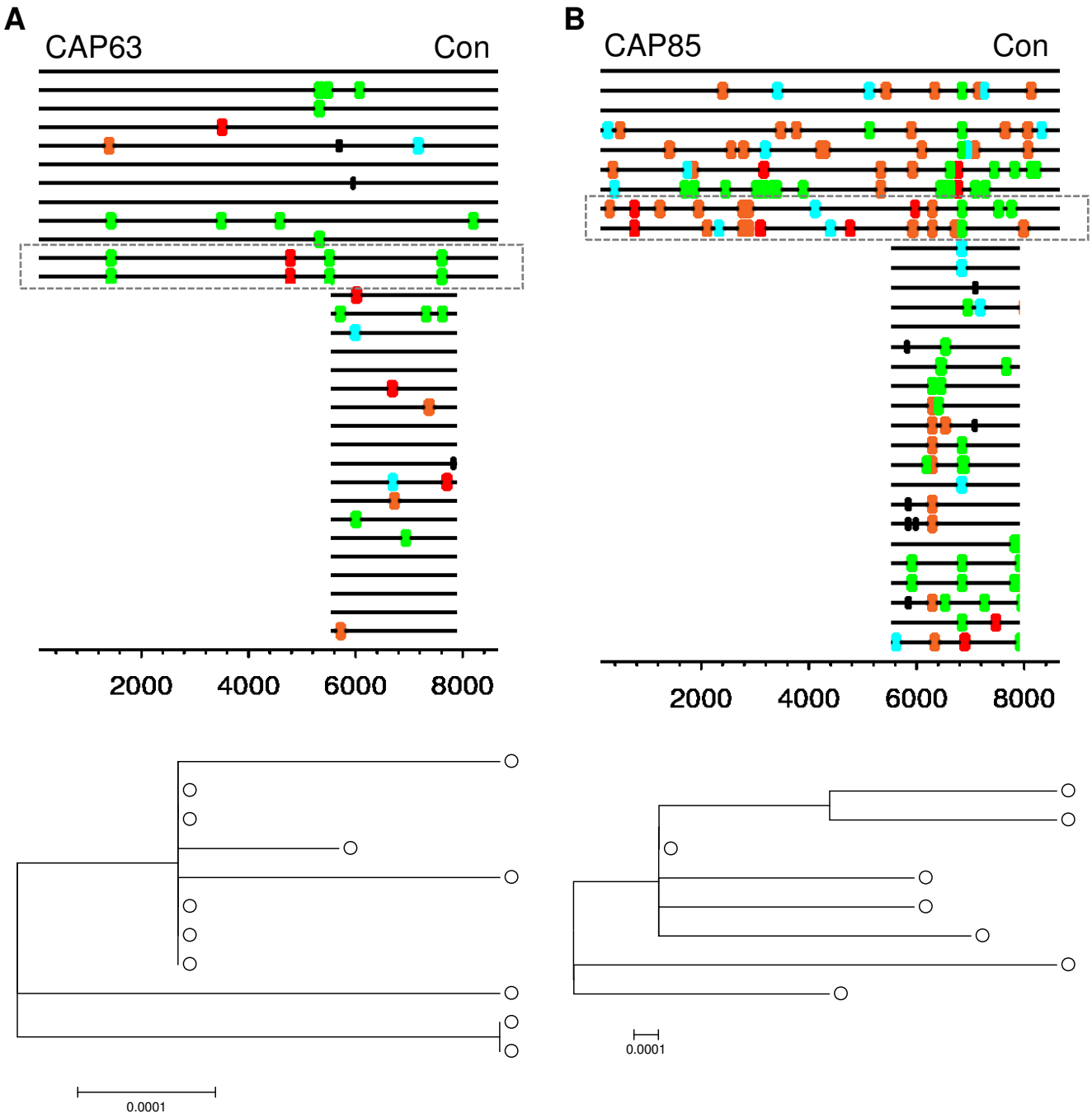


Figure 17. Near full-length genome t/f virus derivation for CAP63 and CAP85. The diversity of near full-length genome sequences sampled from two weeks post-infection for participant CAP63 (**A**) and five weeks post-infection for participant CAP85 (**B**) is illustrated by Neighbour-Joining phylogenies (below) and *Highlighter* plots of nucleotide differences to the derived t/f consensus sequence (www.hiv.lanl.gov) (top). Green=A, Red=T, Orange=G, Light Blue=C, Black=gaps. The second, closely related t/f variant identified in each of these individuals is indicated in a grey box. *Env* SGA sequences used for identification of multiplicity of infection are included. The tree scale bar = 0.0001.

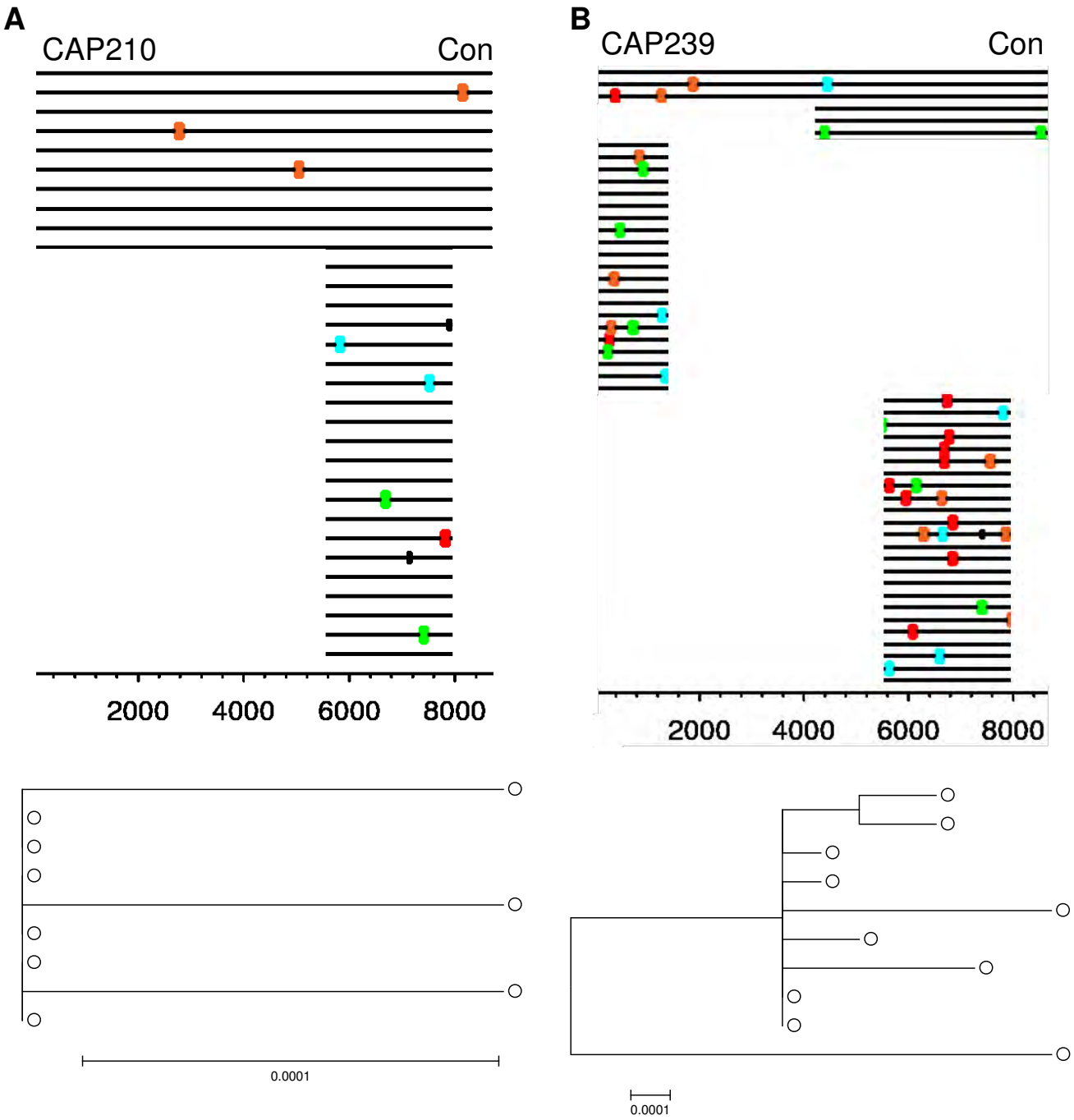


Figure 18. Near full-length genome t/f virus derivation for CAP210 and CAP239. The diversity of near full-length genome sequences sampled from two weeks post-infection for participant CAP210 (**A**) and CAP239 (**B**) is illustrated by Neighbour-Joining phylogenies (below) and *Highlighter* plots of nucleotide differences to the derived t/f consensus sequence (www.hiv.lanl.gov) (top). Green=A, Red=T, Orange=G, Light Blue=C, Black=gaps. *Env* SGA sequences used for identification of multiplicity of infection are included. *Gag* sequences used in derivation of the near full-length t/f virus sequence in CAP239 are also included. The tree scale bar = 0.0001.

2.6.5 Viral diversification in acute and early infection

Longitudinal near full-length genomes from each woman were analyzed for the rate and extent of nucleotide sequence diversification in early infection. Diversity was measured as intraparticipant pair-wise DNA distances at screening/enrolment, three months (acute infection) and six months (early infection) post-infection.

The mean DNA distance across the genome increased by an average of 0.15% between the earliest sampled time-point (two to five weeks post-infection) and three months for four of the women, and decreased by 0.017% in one woman (CAP85) (**Figure 19A**). In the subsequent three months, mean DNA distances increased in all women analyzed by an average of 0.13% (no sequence data was available for CAP45 for this analysis).

When considering the rate of sequence diversification, on average, mean intraparticipant DNA distances increased by 0.016% per week (range 0.013 – 0.024%) across the genome for all women between screening/enrolment and three months post-infection with the exception of CAP85 for whom diversity decreased, by 0.008% per week (range 0.003 – 0.012%) between three and six months. The rate of sequence diversification for rapid progressor CAP63 was highest compared to other women from both screening/enrolment to three months and three to six months with mean DNA distance increases of 0.024 and 0.012% per week respectively. Rates of diversification across the genome from the earliest time-point to three months were modestly higher than in the subsequent three months, even when including the participant for whom diversity initially decreased, however values were not significantly different (**Figure 19B**).

The range of pair-wise DNA distances for acute and early infection was compared between participants using one-way ANOVA with Bonferroni correction for multiple comparisons. Significant differences in diversity between participants were observed at both three and six months ($p<0.0001$) (**Figure 19C and D**). Rapid progressor CAP63 and intermediate progressor CAP85 stood out as having greater diversity. DNA distances for these women were significantly higher than that of CAP210 at three months and than that of both CAP210 and CAP239 at six months.

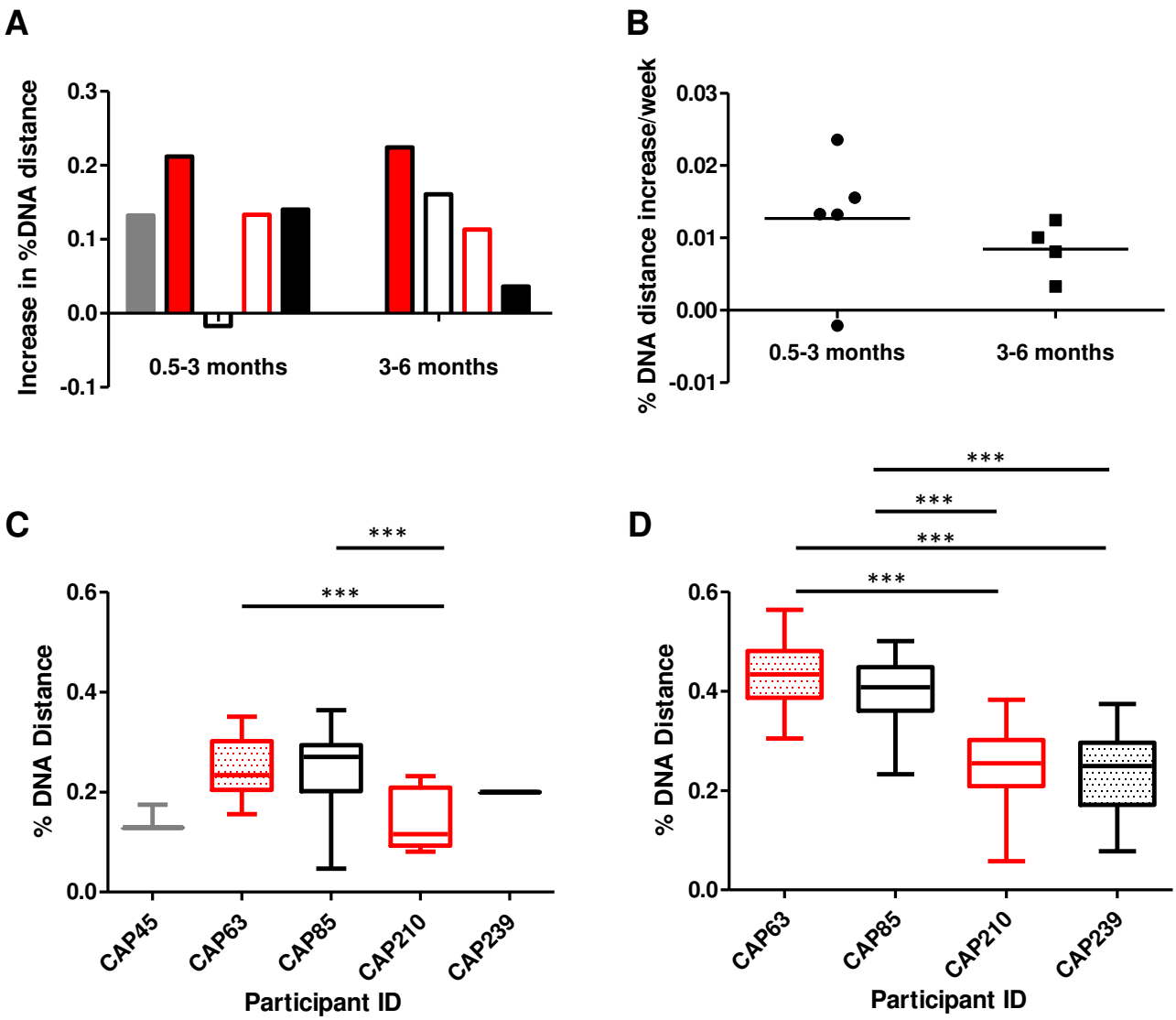


Figure 19. **Sequence diversity of near full-length genome sequences in early infection.** The rates of sequence diversification in the first three months and between three and six months of infection are illustrated in (**A**) and compared in (**B**). Pair-wise DNA distance ranges for the five women at three (**C**) and six (**D**) months of infection are illustrated by box and whisker (minimum to maximum) plots. Data for CAP45 was only available for the first three months. Significant differences in data sets are indicated by horizontal bars (* = $p<0.05$ but >0.01, ** = $p<0.01$ but >0.001, *** = $p<0.001$).

Although viremic controller CAP45 had the lowest range of DNA distances at three months, this did not differ significantly from other participants, possibly due to low sequence numbers. For both participants CAP45 and CAP239 at three months, low sequence numbers (n=3 and n=2 respectively) may have resulted in an under-estimation of sequence diversity.

2.6.6 Viral evolution of the t/f sequence

In order to investigate whether the rate of evolution from t/f virus sequences in early infection differed between individuals with different disease progression, branch lengths of near full-length genome and gene-specific phylogenies over the first three and six months of infection were calculated. Branch lengths represent the number of substitutions per nucleotide site over time.

In a comparison of evolution of near full-length genomes between participants, significant differences in branch lengths were identified at both three (p=0.0009) (**Figure 20A**) and six months post-infection (p<0.0001) (**Figure 20B**) [using one-way ANOVA with Bonferroni correction for multiple comparisons]. Sequences from participants CAP63 and CAP85 had the longest branch lengths at both time-points. Differences in branch lengths between these two women and participants CAP210 and CAP239 were more pronounced at six months than at three months. Viremic controller CAP45 had the shortest branch lengths, however this may be attributed to low sequence numbers (n=3).

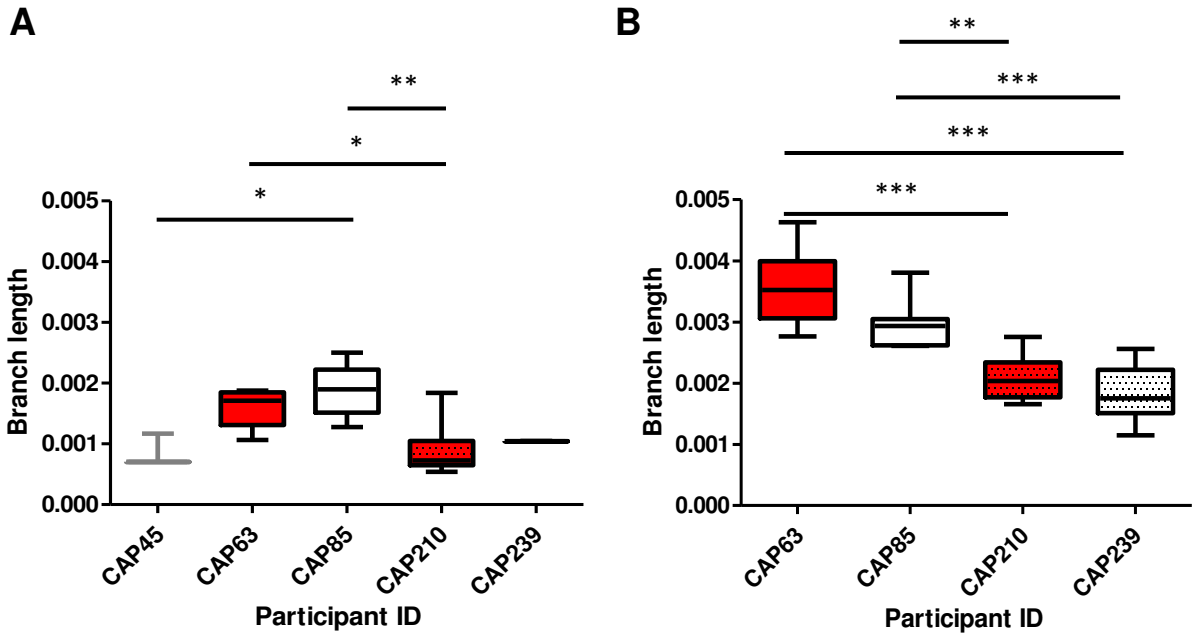


Figure 20. **Near full-length genome branch length comparisons between women.** The range of branch lengths of phylogenetic trees from each of the five women are illustrated by minimum-to-maximum box and whisker plots for (**A**) three months and (**B**) six months post-infection. Significant differences in branch lengths between participants are indicated by horizontal bars (* = p<0.05 but >0.01, ** = p<0.01 but >0.001, *** = p<0.001).

The evolution of sequences was next investigated on a per-gene level and compared between women. Smaller genes *tat*, *rev* and *vpu* were excluded since evolution in these genes on a nucleotide level would overlap with the *env* open reading frame and would conflict comparisons between genes. Average per-gene branch lengths were longest in *env* and *nef* at three and six months compared to those in *gag*, *pol* and *vif* (**Figure 21**).

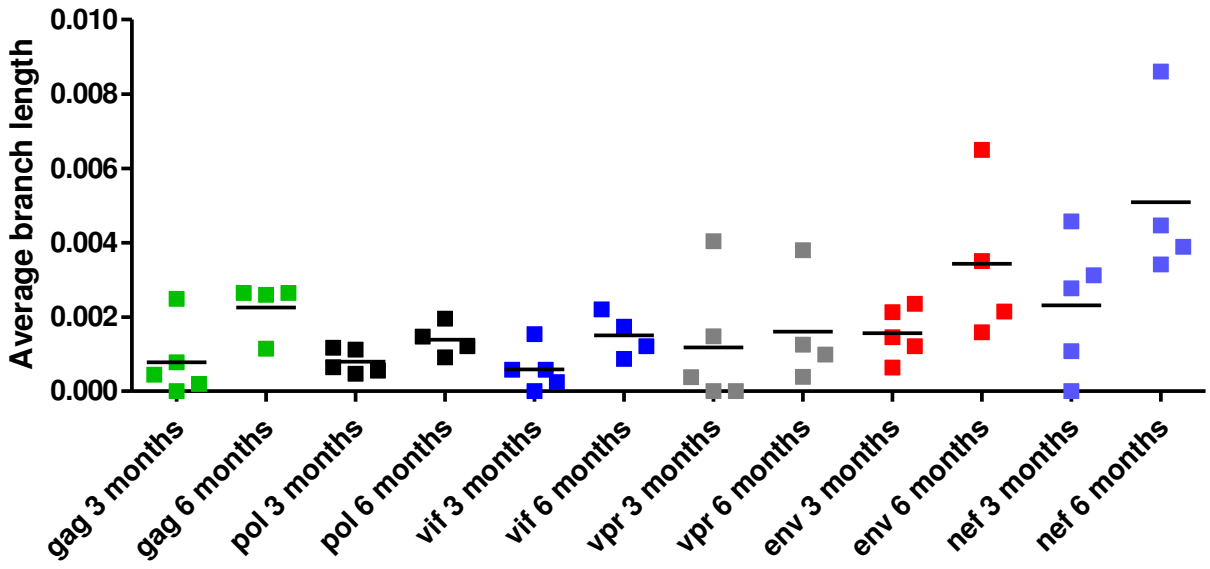


Figure 21. Average per-gene branch lengths in early infection. An average per-gene branch length of phylogenetic trees from each of the five women is plotted at three and six months post-infection. Horizontal bars indicate mean values.

When comparing between women, significant differences in branch lengths were identified at three months post-infection in four of the six genes analyzed, *gag* (p=0.0011), *vpr* (p=0.0044), *env* (p<0.0001) and *nef* (p=0.0001) [one-way ANOVA with Bonferroni correction for multiple comparisons] (**Figure 22A**). Significant differences in branch lengths between women were identified in the same four genes at six months post-infection (*gag* p=0.0101; *vpr* p<0.0001; *env* p<0.0001, *nef* p<0.0001), but were more pronounced (**Figure 22B**).

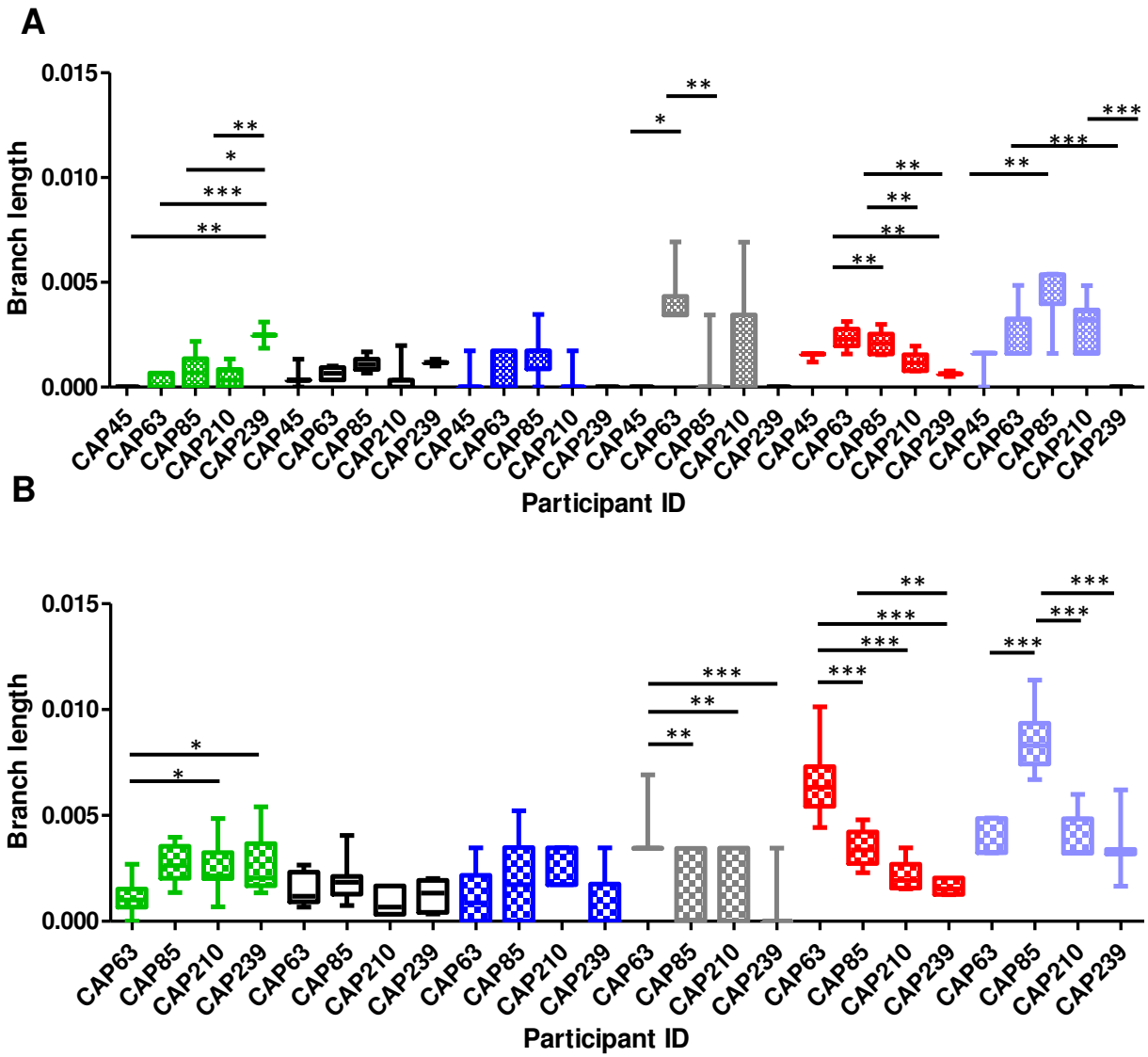


Figure 22. **Per-gene branch length comparisons between women.** The range of branch lengths of phylogenetic trees at three (**A**) and six (**B**) months post-infection from each of the five women for *gag* (green), *pol* (black), *vif* (blue), *vpr* (grey), *env* (red) and *nef* (purple) are illustrated by minimum-to-maximum box and whisker plots. Significant differences in branch lengths between participants are indicated by horizontal bars (* = p<0.05 but >0.01, ** = p<0.01 but >0.001, *** = p<0.001).

Sequences from intermediate progressor CAP239 had significantly longer branch lengths in *gag* in the first three months post-infection compared to all other women, while viremic controller CAP45 and rapid progressor CAP63 had the shortest branch lengths for this gene. CAP63 and intermediate progressor CAP85 had the longest branch lengths in *env* at three and six months post-infection. In *nef*, at both time-points analyzed, participant CAP85 had the longest branch lengths and CAP239 had the shortest branch lengths.

2.7 Discussion

This study investigated the number of t/f viruses establishing clinical infection in acutely subtype C-infected women from the CAPRISA 002 AI cohort in Durban, South Africa, using *env* sequence data. In a focused analysis of five women with differing disease progression, and identified as infected with a single t/f variant by *env* sequence analyses, near full-length t/f virus genome sequences were derived and early viral diversification and evolution across the genome in each woman was described. To our knowledge, this was the first subtype C near full-length genome study which investigated early viral evolution of t/f virus sequences in the context of disease progression.

In 26 CAPRISA 002 women investigated, 77% of infections were found to be due a single t/f *env* variant based on conformance to a model of random evolution (Keele et al., 2008). This result is similar to that described for the larger CHAVI study, which these women formed part of, which reported 78% of 69 individuals with subtype C infection from Malawi and South Africa to be due to a single variant (Abrahams et al., 2009). Sequences from six women (23%) did not conform to the model of random evolution and these women were classified as infected with between two and five t/f variants. For seven women with single-variant infection, low-level sequence diversity was identified and attributed to host-restriction from APOBEC and adaptive immune responses. This diversity resulted in deviation from model conformance, since evolution was not random but rather influenced by selection.

Halaand et al. (2009) reported an association between multivariant t/f virus infection and the presence of STIs. The women in the present study had a high frequency of STIs. The frequency of infection with any STI (38%) for the 26 CAPRISA 002 women was slightly increased compared to the 31% reported for HIV-uninfected high-risk South African women (Mlisana et al., 2012). The percentage of women infected with *N. gonorrhoea*, the STI most strongly associated with risk of HIV infection (Mlisana et al., 2012), was however three-fold higher than that reported for uninfected women (5% compared to 15%) (Mlisana et al., 2012). No association between the number of STIs or specific STIs and multiplicity of infection was identified. Of interest, the woman with the highest multiplicity of infection (five variants) also had *N. gonorrhoea*, however two other women with single-variant infection were also infected with this STI.

Transmission of single-variant, low-diversity t/f viruses in a subset of five women was confirmed by near full-length genome sequencing. Two women (CAP63 and CAP85), although identified

as infected with a single variant by *env*-only sequence analyses (Abrahams et al., 2009), were each found to be infected with a second very closely related variant following near full-length genome analyses. These variants represented minority populations and were not detected at subsequent time-points. However, since half the number of genomes were sequenced at three and six months compared to the number sequenced at screening/enrolment, it is possible that these variants may have persisted at later time-points and were not detected due to small sampling size, yet they likely did not gain dominance in the overall population. Although these additional variants were not detected by *env*-only SGA, we do not believe that this disputes the use of *env*-only SGA for determining multiplicity of infection nor advocates the use of near full-length genome amplification for this purpose since: (i) within an individual, variants differed only by a few nucleotides across the genome and did not represent high diversity, multi-variant infection; (ii) amplification of near full-length genomes would not only be more costly, but achieving high sequence numbers needed for diversity analyses would be more challenging. The results of this study support *env*-only SGA as a sensitive technique for determining multiplicity of infection.

Early studies have reported a relationship between viral evolution or diversity of HIV-1 and clinical disease progression (Wolfs et al., 1990; Lukashov et al., 1995; Wolinsky et al., 1996; Ganeshan et al., 1997; Liu et al., 1997; McDonald et al., 1997; Shioda et al., 1997; Markham et al., 1998), with the majority investigating only the *env* gene or regions thereof, and finding an association between higher genetic diversity or rate of evolution and slow disease progression. Of the five women focused on in this study, viral diversification and evolution was more rapid and extensive in rapid progressor CAP63 and intermediate progressor CAP85 compared to the other three women. Interestingly, for participant CAP63, diversification was extensive in *env* and only minimal diversification was seen in the *gag* and *pol* genes. Sequences from viremic controller CAP45 displayed the lowest diversity and extent of evolution at three months post-infection, however this may have been due to low sample number.

In a more recent study, Herbeck et al. (2011) reported a contraction of viral diversity in near full-length genomes from 11 subtype B-infected MSM individuals within the first 20-40 days post-infection (taken as days after onset of clinical symptoms, thus the actual period of infection was likely longer), despite continuous viral outgrowth. The authors hypothesized this to be due to viral expansion from a single or limited number of variants, or due to purifying selection eliminating hypermutated viruses (Herbeck et al., 2011). In the current study, a comparison of viral diversification between participants showed an increase in sequence diversity in the first

three months for four women and a contraction in diversity for one woman. Furthermore, the rate of sequence diversification was higher in the first three months of infection than between three and six months, although this difference was not statistically significant. Interestingly, the Herbeck study also reported early selection to be dominant in *env* and *nef*. In the current study, evolution was more extensive in *env* and *nef* at both three and six months post-infection compared to *gag*, *pol*, *vif* and *vpr*. Immune selection for these women will be investigated in Chapter 3 of this thesis.

The use of near full-length genomes in this study allowed for comparison of rates of diversification and evolution of individual HIV-1 genes. However, the numbers of sequences used in these analyses were limiting for certain individuals/time-points due to challenges inherent in full-length genome amplification techniques including sample availability and cost. It is therefore important to note that the rates of diversity and evolution reported may have been underestimated for the individuals in this study. None-the-less, the data provides insight into which genes evolve more rapidly and are potentially preferred targets of early immune responses, and therefore form a basis for selection of individual genes for which greater depth of sequencing would be beneficial.

This chapter has described the multiplicity of infection and the early viral evolution of subtype C t/f viruses in the context of differing disease progression. Increased diversity and more rapid evolution were reported for virus from one rapid disease progressor. The impact of specific immune responses mediated by CD8+ T-lymphocytes and antibodies in acute/early infection will be investigated and discussed in the following chapter.

Chapter 3: Mapping immune selection in early subtype C infection

3.1 Abstract

Elucidation of HIV-1 immune escape mutations acquired following infection is a sensitive approach to identifying the earliest immune pressures associated with control of viral replication in primary infection. The purpose of this study was to characterize the earliest changes in near full-length genomes following infection to better understand early control of viremia.

Near full-length transmitted/founder and early (first six months of infection) viral sequences from five women recruited within two to five weeks of infection (Chapter 2) were analyzed for the presence of amino acid mutations associated with cytotoxic T-lymphocyte and neutralizing antibody immune selection.

Fifty-five genome regions were identified with amino acid mutations associated with viral adaption to the new host. The majority of mutations were attributed to CD8+ T-lymphocyte pressure (40%), followed by antibody-mediated pressure (35%) and reversion of transmitted mutations (16%). Nine percent of mutations could not be classified. CD8+ T-lymphocyte escape was most frequent in Nef, followed by Pol and Env, and occurred at a higher frequency in the first five weeks of infection followed by a decline as viral load declined. In the majority of mutating CD8+ T-lymphocyte epitopes (82%), amino acid shuffling and toggling was observed with only 9% reaching fixation in the first six months of infection.

In conclusion, CD8+ T-lymphocyte-mediated immune pressure was the predominant selective force in primary subtype C infection. Immune escape was rapid and frequently characterized by complex escape pathways.

3.2 Introduction

3.2.1 Immune selection in the HIV-1 genome

HIV replication is highly error prone (Mansky et al., 1995) which enables the virus to rapidly escape immune responses through the introduction of amino acid mutations. Mutations that are advantageous to the virus are maintained and become fixed over time within the host and therefore represent a form of positive selection (Price et al., 1997). Following transmission however, these mutations may not be advantageous in the new recipient and may revert to the wild-type form (reversion) (Leslie et al., 2004; Moore et al., 2012). If escape mutations are associated with a loss of viral fitness, they may be compensated by additional mutations at proximal or distant sites (Brockman et al., 2007, Crawford et al., 2007). Sequence changes associated with immune selection mediated by CD8+ T-lymphocytes (CTLs) and antibody (Ab) responses, as well as reversion and compensatory mutations, have been well described (**Figure 23**).

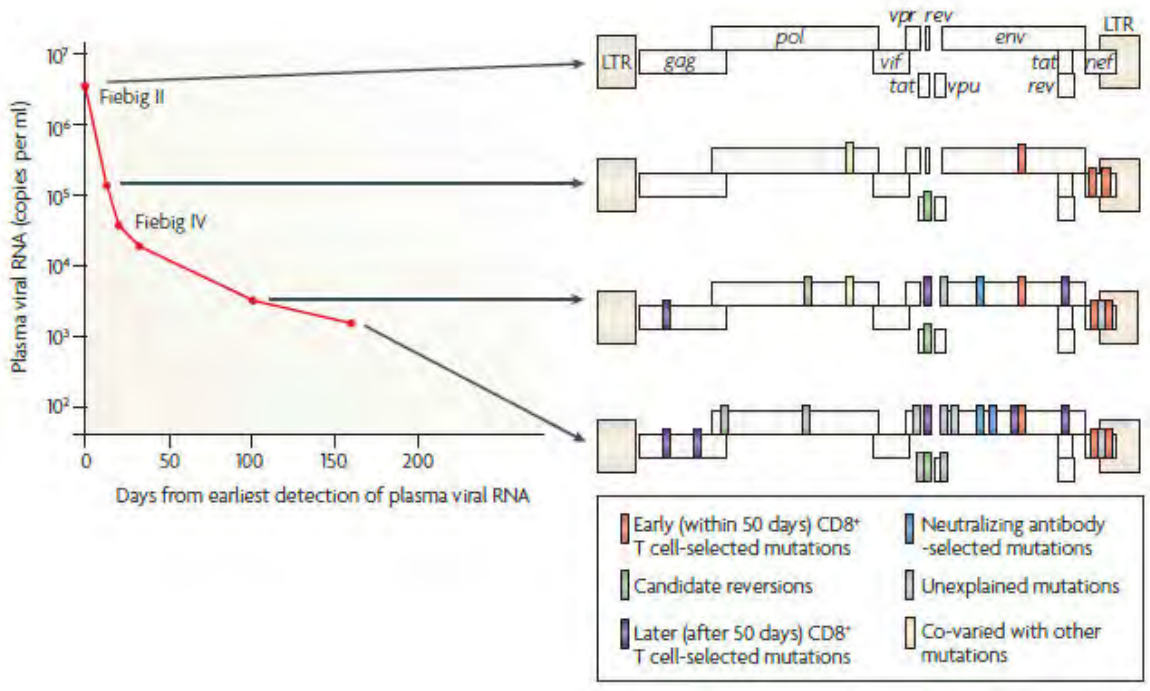


Figure 23. **The accumulation of mutations associated with immune selection across the genome over the course of acute and early HIV-1 infection** (illustration taken from McMichael et al., 2010). Stages II (viral RNA and p24 antigen positive, but Ab negative) and IV (ELISA positive but Western blot indeterminate) of acute infection as defined by Fiebig et al (2003) are indicated corresponding to peak viremia and viral load decrease following peak viremia respectively.

Selection patterns in viral sequences from other sources such as CD4+ T-cells and natural killer (NK) cells are not as well characterized, although polymorphisms associated with escape from NK cell pressure mediated by killer immunoglobulin-like (KIR) receptors has been described (Alter et al., 2011).

CTL responses are mediated via HLA Class I molecules which recognize sequences within the viral genome and present short peptides to CD8 receptors on the surface of CTLs. CTL escape occurs when mutations (i) occur in HLA-binding anchor residues and prevent peptide presentation; (ii) obscure recognition by T-cell receptors; or (iii) alter peptide processing within an infected cell thereby preventing peptide presentation (Borrow et al., 1997, Goulder et al., 1997, Feeney et al., 2004, Draenert et al., 2004; Tenzer et al., 2009). Mutations associated with CTL escape have been extensively characterized (Borrow et al., 1997, Goulder et al., 1997, Draenert et al., 2004, Allen et al., 2005, Brumme et al., 2007, Brumme et al., 2008, Brumme et al., 2008b, Brumme et al., 2009, Carlson et al., 2012a; Carlson et al., 2012b). This data, together with mapping of HLA Class I genotype associated mutations in population-based studies (HLA-associated polymorphisms) (Moore et al., 2002; Kiepiela et al., 2004; Brumme et al., 2007; Kiepiela et al., 2007; Brumme et al., 2008; Brumme et al., 2008b; Matthews et al., 2008; Rousseau et al., 2008; Brumme et al., 2009; Carlson et al., 2012a) have made it possible to predict CTL escape through viral sequencing when the HLA type of the individual is known.

Escape is typically characterized by transitions from high frequency/consensus to low frequency/nonconsensus amino acids and may fall within or flank targeted epitopes (Allen et al., 2004; Liu et al., 2006). A single amino acid change in a site evolving under positive selection or a cluster of mutations in short peptide stretches have both been used as proxies for identifying immune selection (de Oliveira et al., 2004; Liu et al., 2006; Kelleher et al., 2001; Jones et al., 2004; Wood et al., 2009). APOBEC is a host restriction factors which incorporates nucleic acid G-to-A mutations along the viral genome in order to restrict viral replication (Malim and Bieniasz, 2012). A study by Wood et al. (2009) showed that APOBEC facilitated immune escape in CTL epitopes.

Antibodies predominantly target the viral envelope glycoprotein spikes which are exposed on the outer surface of viral particles, and antibody selection typically presents as gains and losses of potential N-linked glycosylation sites (PNGSs) and nucleotide insertions and deletions (indels) in the variable loops (Wei et al., 2003; Richman et al., 2003, Rybarczik et al., 2004). Addition of PNGSs can result in steric hindrance and thus shielding the virus from antibody

binding (Wei et al., 2003). The variable loops also shield the virus from neutralization (Rusert et al., 2011) and the introduction of insertions resulting in lengthening of loops may obscure access to epitope targets and enhance shielding (Laakso et al., 2007).

3.2.2 Immune selection in acute infection

CTLs emerge very early in HIV-1 infection and are associated with early control of viremia (Koup et al., 1994; Borrow et al., 1994; Ogg et al., 1998). CTL escape has also been observed in the first weeks of infection (Borrow et al., 1997; Jones et al., 2004; Liu et al., 2006; Keele et al., 2008; Gray et al., 2009; Salazar-Gonzalez et al., 2009; Goonetilleke et al., 2009; Wood et al., 2009; Mlotshwa et al., 2010). The large majority of early selection identified in acute infection studies evaluating full-length genomes was attributed to CTL-mediated pressure, although Ab-associated selection and reversion has also been described (Goonetilleke et al., 2009; Henn et al., 2012). Studies such as these have been limited to subtype B-infected individuals (Goonetilleke et al., 2009; Herbeck et al., 2011; Henn et al., 2012).

In this thesis, a companion study performed in conjunction with a subtype B and C study by Liu et al (2013) is described, wherein early immune selection in full-length genome transmitted subtype C viruses was characterized. This required firstly identifying the genome sequence of the virus that was transmitted in five subtype C-infected women, followed by longitudinal sequencing over acute and early infection (up to six months post-infection) (see Chapter 2). The focus of this chapter is the identification of early immune selection in all nine HIV-1 genes.

Furthermore, the impact of early immune selection on the clinical course of disease remains unclear. Since the five women in this study include rapid and intermediate progressors as well as a viremic controller, early selection will also be evaluated in the context disease progression.

Results from this study were published in a manuscript entitled ‘Rapid, complex adaptation of transmitted HIV-1 full-length genomes in subtype C-infected individuals with differing disease progression’ (Abrahams et al., 2013) (**Appendix 2**).

3.3 Research Aim and Objectives

Aim: To identify immune selection in near full-length genome subtype C HIV-1 sequences during acute and early infection

Objective 1: Analyze longitudinal near full-length genome sequences from five acutely infected women for regions containing known HLA Class I-associated polymorphisms, clustered amino acid polymorphisms and/or sites evolving under positive selection

Objective 2: Classify evolving sites associated with CTL- and Ab-mediated immune pressure or reversion

Objective 3: Evaluate immune selection in the context of disease progression

3.4 Rationale

The correlates of protection from HIV-1 infection are unknown, however are likely to include neutralizing and non-neutralizing antibodies to block infection; and cellular immune responses to eliminate or contain the initial foci of infection. Studies of natural infection can provide clues on early responses and immune targets associated with control of viral infection.

3.5 Methods and Materials

Single genome amplification and sequencing of near full-length genomes, as well as derivation of t/f sequences, are described in Methods and Materials of Chapter 2. Supplemental sequence data used in analyses for this chapter included four half-genome sequences generated by H. Bredell (University of Cape Town); 11 near full-length genomes (amplified by F. Treurnicht, University of Cape Town); *gag* and *nef* sequences generated by limiting dilution PCR and cloning done by D. Chopera and R. Ntale (University of Cape Town) respectively (described in Chopera et al., 2011); and *env* sequences generated by single genome amplification as described in Abrahams et al., 2009 as well as additional *env* SGA sequences generated by G. Bandawe (University of Cape Town) (details of supplemental data are provided in **Appendix 3**).

Epitope sequences were generated by targeted direct sequencing of gene-specific amplicons generated using limiting dilution PCR. Of the 390 epitope sequences used, 321 were generated as part of this thesis and 69 sequences were contributed by J.Marais, S. Goodier and N. Ndabambi. Details for *gag*, *vif* and *nef* limiting dilution PCR are described in **Appendix 4**.

3.5.1 Sequence analyses

Sequence alignments and amino acid frequency plots were performed in BioEdit version 7.0.8.0 (Hall, 1999). Intra-participant sequence alignments were analyzed for the presence of silent and non-silent mutations away from the transmitted/founder virus using the *Highlighter* tool (http://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter). Subtype C amino acid frequencies were calculated based on per gene database alignments obtained from the Los Alamos HIV Database (www.hiv.lanl.gov).

The *Hypermut* tool (http://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html) was used for the detection of APOBEC-associated G-to-A hypermutation. Shannon entropies were determined using the *Entropy One* tool (http://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy_one). A high entropy score was taken as $\geq$ 0.25 (Bansal et al., 2005). Structural and functional motifs/regions within the genome were identified using the HIV landmark map from the Los Alamos Database (http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark).

3.5.2 Classification of immune selection

Mutations from high frequency/consensus to low frequency/non-consensus amino acids within known class I HLA-restricted CTL epitopes were classified as CTL-mediated selection (Allen et al., 2005, Liu et al., 2006, Li et al., 2007, Brumme et al., 2008). Known class I HLA-associated epitopes and polymorphisms associated with CTL-escape and compensation of escape were identified using the Los Alamos HIV Molecular Immunology Compendium (http://www.hiv.lanl.gov/content/immunology/compendium.html) and Matthews et al., 2008. Predicted CTL epitopes were identified using Epitope Location Finder (ELF) (http://www.hiv.lanl.gov/content/sequence/ELF/epitope_analyzer) and NetMHCpan 2.2 (http://www.cbs.dtu.dk/services/NetMHCpan).

Mutations from low frequency/non-consensus to high frequency/consensus amino acids were classified as reversion (Allen et al., 2005, Li et al., 2007, Brumme et al., 2008). Mutations resulting in gains or losses of potential N-linked glycosylation sites (PNGSs) or which altered the length of hypervariable loops within the envelope gene were classified as Ab-mediated selection (Wei et al., 2003). PNGSs were identified using the *N-GlycoSite* tool (http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html).

3.5.3 Positive selection analyses

The MG94xHKY85 codon model (Kosakovsky Pond et al., 2005) implemented in HyPhy (Pond et al., 2005) was used to calculate non-synonymous/synonymous substitution (dN/dS) rate ratios. Allowance was made for dS to vary across codon sites with use of the Dual model which accounts for dS varying independently of dN (Pond and Muse, 2005). For genomic regions separated by recombination breakpoints, the use of correct phylogenetic relationships was ensured (Scheffler et al., 2006). Supplemental clonal and SGA sub-genomic sequences were included in these analyses. Dr Nobubelo Ngandu (University of Cape Town) assisted with positive selection analyses.

3.5.4 Statistical tests

Fisher's exact two-tailed tests (http://www.graphpad.com/quickcalcs) were used to perform categorical statistical tests.

3.6 Results

3.6.1 Identification of early immune selection in HIV-1 subtype C near full-length genomes

Longitudinal near full-length genome sequences (n=113) generated from five acutely infected women over the first six months of infection were analyzed for mutations associated with immune selection. Eleven near full-length and four half-genomes were also included from additional time-points (**Appendix 3**). All women were HIV-1 subtype C-infected and were recruited into the study within two (CAP45, CAP63 and CAP210) to five (CAP85 and CAP239)

weeks post infection. The sequence of their t/f near full-length genome virus/es were derived (Chapter 2), and changes over time were characterized to identify the timing, type and location of the earliest immune selective pressures acting on the t/f genome.

CTL-escape mutations were identified as mutations from high frequency/nonconsensus to low frequency/consensus amino acids in or flanking known or predicted HLA Class I-restricted epitopes (see section 3.5.2) or which corresponded to known HLA-restricted polymorphisms (Allen et al., 2005, Liu et al., 2006, Li et al., 2007, Brumme et al., 2008, Matthews et al., 2008). Conversely, mutations from low frequency/nonconsensus to high frequency/consensus amino acids in epitopes not targeted by the individuals HLA were classified as reversion (Allen et al., 2005, Li et al., 2007, Brumme et al., 2008). Mutations in *env* resulting in gains or losses of PNGSs and indels in variable loops were classified as Ab-mediated selection (Wei et al., 2003; Richman et al., 2003, Rybarczik et al., 2004). Furthermore, mutational clusters in amino acid nine-mers (Kelleher et al., 2001; Jones et al., 2004) containing sites evolving under positive selection (de Oliveira et al., 2004; Liu et al., 2006) were classified as immune pressure.

Mutations conforming to the aforementioned criteria were identified in a total of 55 genome regions of viruses from the five women (**Figures 24 to 28**). Of the 55 regions, 22 were classified as evolving under CTL-mediated selection (**Table 4**), 19 as evolving under Ab-mediated selection (**Table 5**), nine as reverting (**Table 6**), and finally selection in five regions (with clustered mutations) could not be classified (**Table 7**). Positive selection (dN/dS>1) was identified in a total of 55 amino acid sites, of which 84% (46/55) fell within the 55 genome regions identified as evolving under immune selection (**Figure 24B to 28B**). A total of 30 sites under positive selection were identified in *env*, representing the majority, followed by Gag (n=9), Pol (n=7), Nef (n=5), Rev (n=3) and Tat (n=1).

Of the 22 genome regions identified as evolving under CTL-mediated selection, 19 spanned known class I HLA-restricted CTL epitopes (**Table 4**), two spanned epitopes predicted to be bound by class I HLA types identified in the participant, and finally one region contained neither a known nor a predicted epitope but was found to induce a response in IFN-γ ELISPOT assays in a companion study by Liu et al. (2013) which tested responses to autologous peptides from participants CAP45, CAP210 and CAP239. This reactive peptide was located at the beginning of Pol (FFRENLAFPEGEARELPS) and may contain a novel CTL epitope. All 11 evolving CTL epitopes (known and predicted) identified for participants CAP45, CAP210 and CAP239 were likewise confirmed to induce an immune response by IFN-γ ELISPOT (Liu et al., 2013), as well

as three of the 11 epitopes identified for participants CAP63 and CAP85 (Liu et al., unpublished data, personal communication). For CAP63 one epitope in Pol (QLTEAVHKI), which was predicted to be HLA-A*02:01-targeted, and another in Gp41 (LLDSIAITV) were reactive in IFN-γ ELISPOT, and for participant CAP85 one epitope in Nef (KAAVDLSFF) was reactive. Therefore in total 14 sites predicted by sequence data to be evolving under CTL-mediated selection were confirmed by elispot as responsive, and responses could not be confirmed to the remaining eight evolving epitopes.

Of the nine genome regions identified with reverting mutations (**Table 6**), four contained known Class I HLA-targeted epitopes in Gag, Gp41 and Nef, and may represent reversion of transmitted escape mutations from the donor. In the remaining five regions, changes from uncommon (identified in 0.19% to 19% of subtype C database sequences) to common (identified in 48% to 98% of subtype C database sequences) were observed but could not be associated with reversion of escape in known HLA-restricted epitopes.

Of the five genome regions for which immune selection could not be classified (**Table 7**), one genome region in Nef (SLHGMEDTEREVLQWKFD), in participant CAP210, overlapped with a known CD4+ T-cell peptide (PEREVLEWRFDSRLAFHH) reported to be targeted by multiple HLA-DR types (Kaufmann et al., 2004), however none of these types matched the participant's HLA-DR classification.

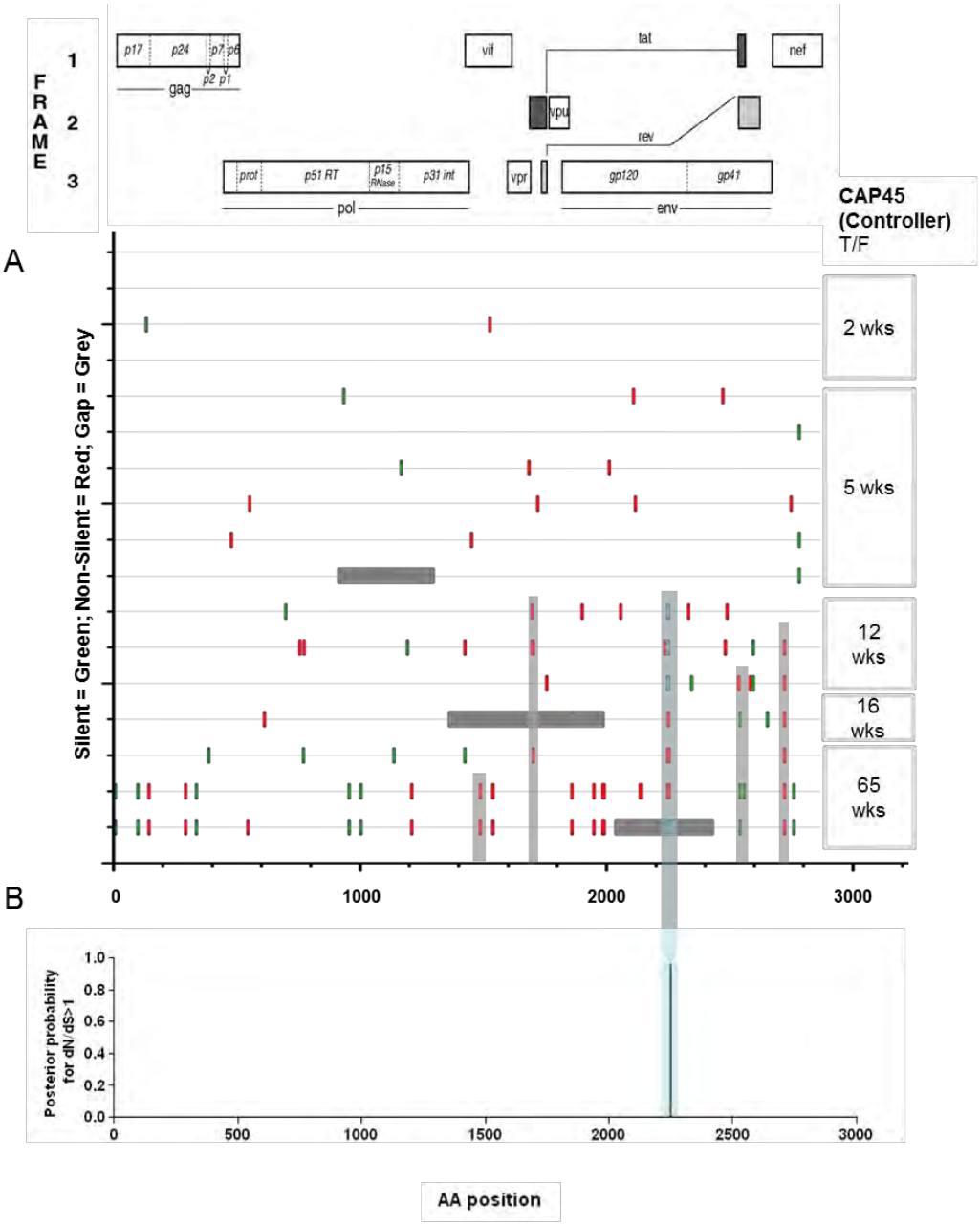


Figure 24. **Early selection in the transmitted/founder (t/f) virus of participant CAP45**. (**A**) A *Highlighter* plot of synonymous (green) and non-synonymous (red) changes from the t/f virus over time, and (**B**) amino acid sites across the genome under positive selection given as the posterior probability of a dN/dS ratio >1 for each site. Pale grey bars = putative CD8+ T-lymphocyte escape sites and blue bars = sites under putative antibody-mediated pressure. Horizontal grey bars indicate deletions/missing data within the genome.

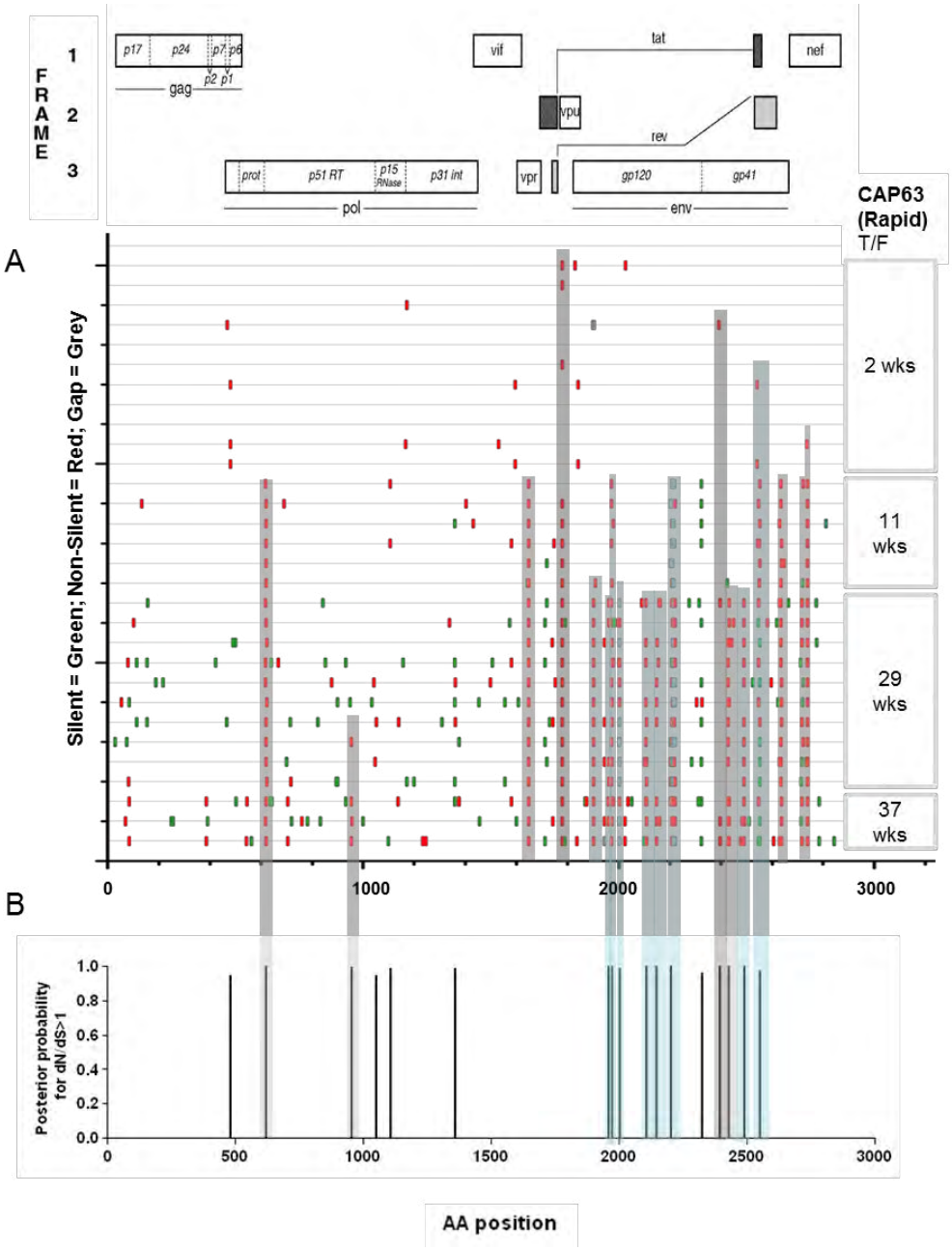


Figure 25. **Early selection in the transmitted/founder (t/f) virus of participant CAP63**. (**A**) A *Highlighter* plot of synonymous (green) and non-synonymous (red) changes from the t/f virus over time, and (**B**) amino acid sites across the genome under positive selection given as the posterior probability of a dN/dS ratio >1 for each site. Shaded bars indicate where mutating sites in the *Highlighter* plots correspond to sites under positive selection. Pale grey bars = putative CD8+ T-lymphocyte escape sites, dark grey = sites reverting to consensus, and blue bars = sites under putative antibody-mediated pressure. A [#] indicates that mutations were found to arise earlier than indicated on the *Highlighter* plot by subsequent targeted epitope sequencing.

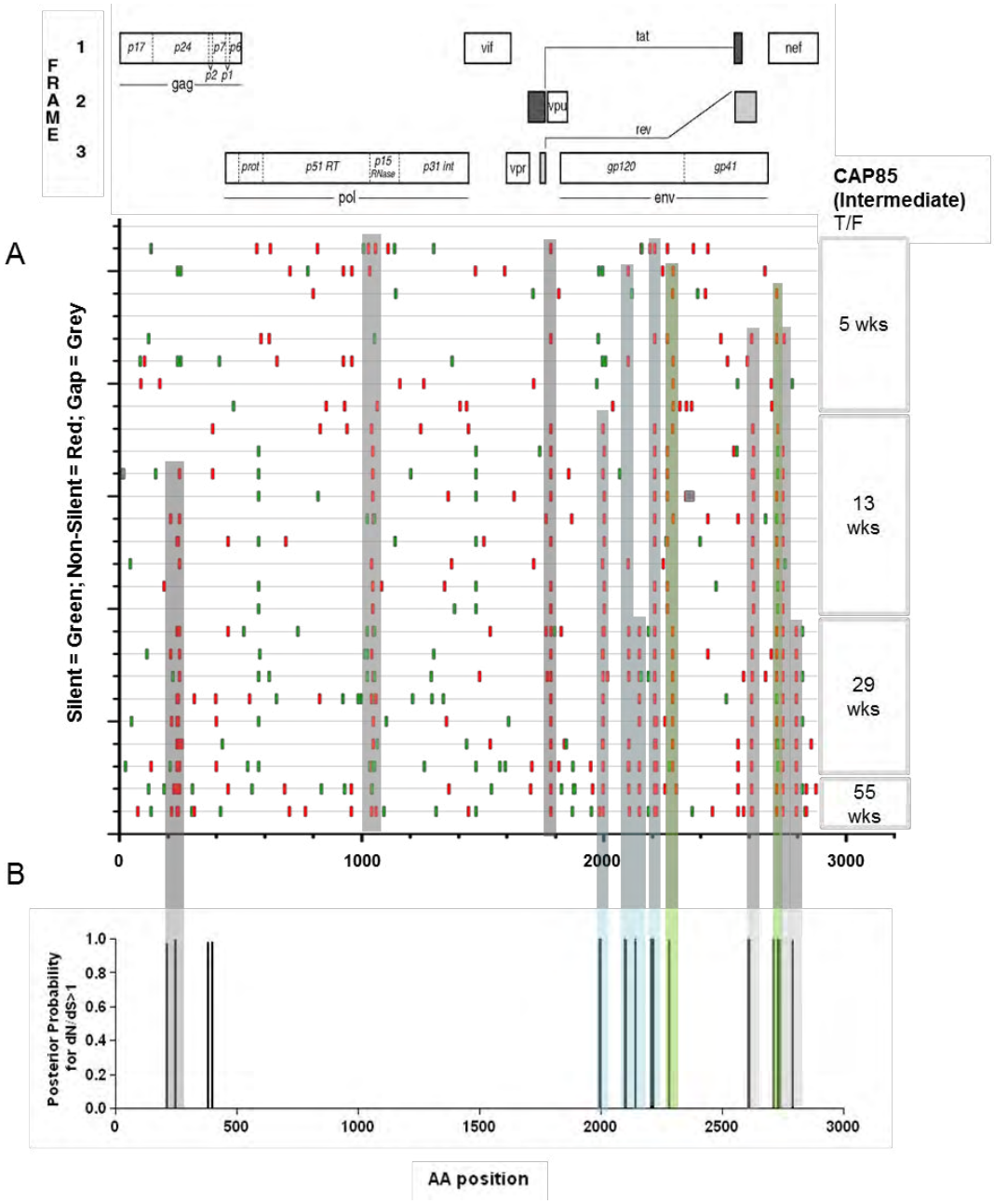


Figure 26. **Early selection in the transmitted/founder (t/f) virus of participant CAP85**. (**A**) A *Highlighter* plot of synonymous (green) and non-synonymous (red) changes from the t/f virus over time, and (**B**) amino acid sites across the genome under positive selection given as the posterior probability of a dN/dS ratio >1 for each site. Pale grey bars = putative CD8+ T-lymphocyte escape sites, dark grey = sites reverting to consensus, blue bars = sites under putative antibody-mediated pressure, and green bars = sites with changes which could not be classified. Horizontal grey bars indicate deletions/missing data within the genome.

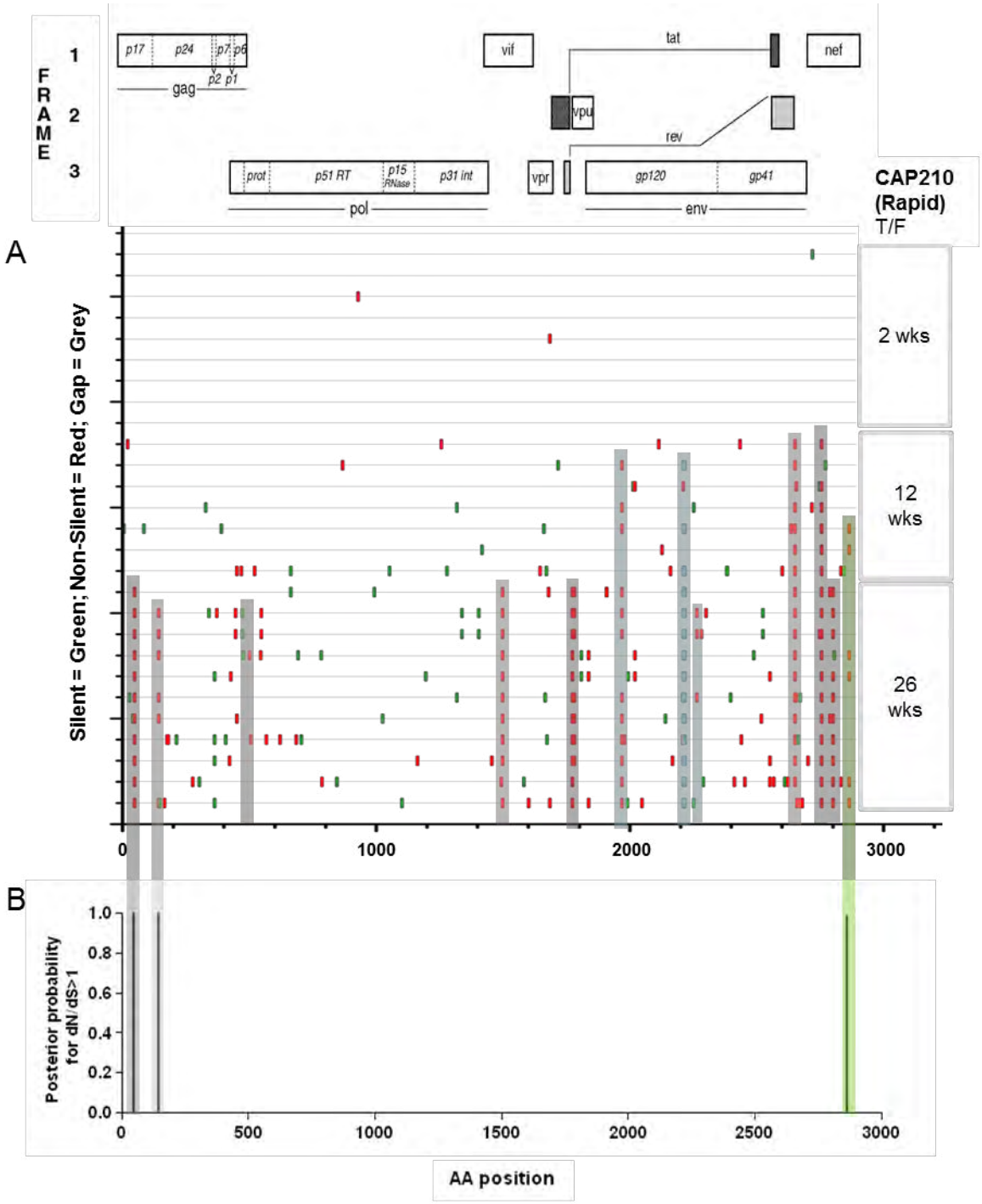


Figure 27. **Early selection in the transmitted/founder (t/f) virus of participant CAP210**. (**A**) A *Highlighter* plot of synonymous (green) and non-synonymous (red) changes from the t/f virus over time, and (**B**) amino acid sites across the genome under positive selection given as the posterior probability of a dN/dS ratio >1 for each site. Pale grey bars = putative CD8+ T-lymphocyte escape sites, dark grey = sites reverting to consensus, blue bars = sites under putative antibody-mediated pressure, and green bars = sites with changes which could not be classified.

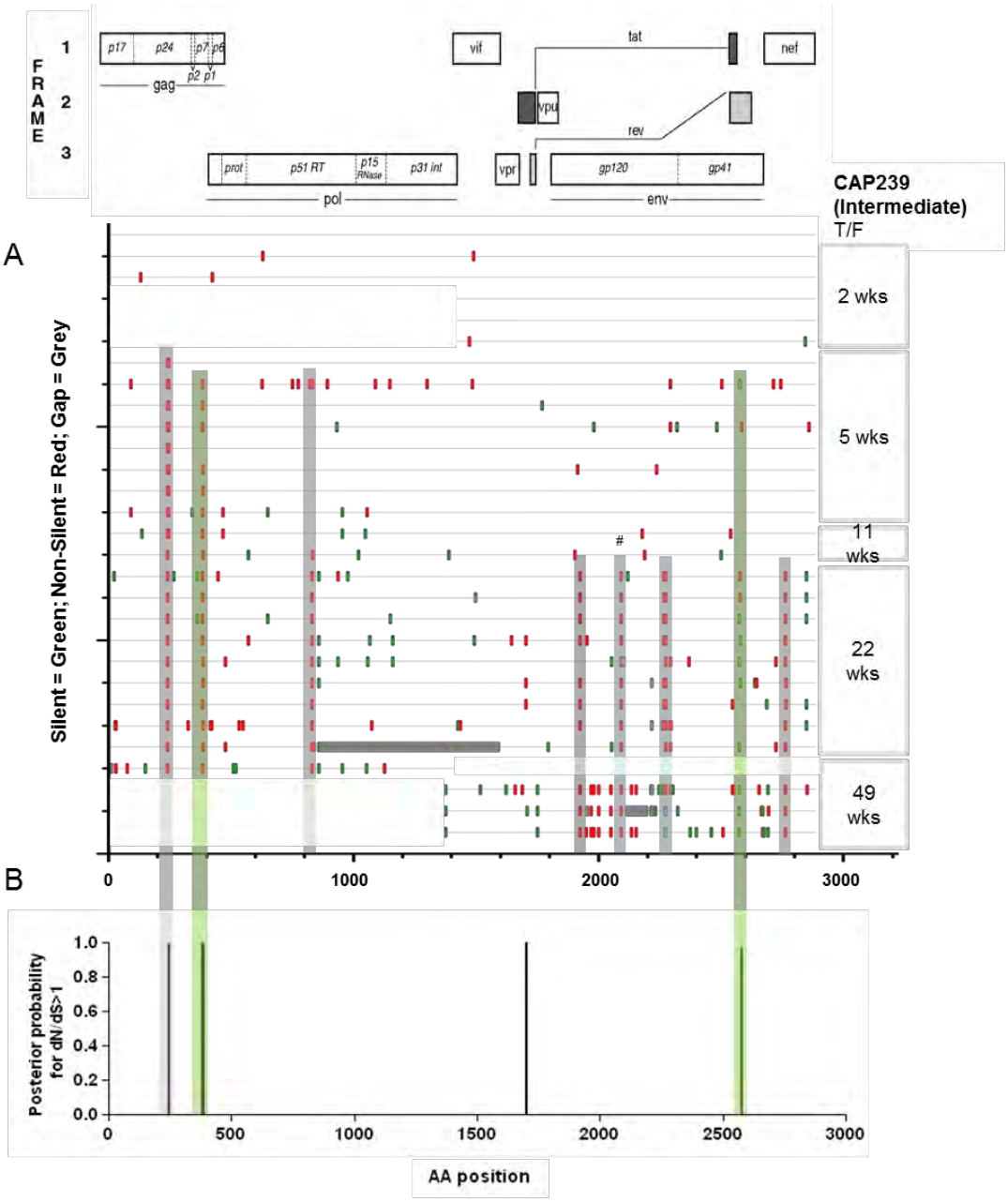


Figure 28. E**arly selection in the transmitted/founder (t/f) virus of participant CAP239**. (**A**) A *Highlighter* plot of synonymous (green) and non-synonymous (red) changes from the t/f virus over time, and (**B**) amino acid sites across the genome under positive selection given as the posterior probability of a dN/dS ratio >1 for each site. Pale grey bars = putative CD8+ T-lymphocyte escape sites, dark grey = sites reverting to consensus, blue bars = sites under putative antibody-mediated pressure, and green bars = sites with changes which could not be classified. Horizontal grey bars indicate deletions/missing data within the genome. A # indicates that mutations were found to arise earlier than indicated on the *Highlighter* plot by subsequent targeted epitope sequencing.

Table 4. Early immune selection in Class I HLA-restricted cytotoxic T-lymphocyte epitopes

Participant ID	ORF	Epitope/genome region sequence*	HXB2 psn.	Participant HLA association/s**	Reference	Time of first AA change (range)(wks)	High entropy epitope/peptide (LANL subtype C database)	Shuffling/ Toggling of AA mutations	Structural/ Functional genome region (LANL HXB2 Annotation)
CAP45	Vif	**D**WHLGHGVSI	78-87	B*15:10	LANL database	12-65	No	No	None/Unknown
	Rev	I**H**SISERI**L**	52-60	B*15:10	LANL database	5-12	Yes	Yes	None/Unknown
	Tat	**N**CYCK**H**CSY	24-32	A*29:02	LANL database	5-12	Yes	Yes	TAR RNA binding
	Nef	**EE**VGFPVRPQV	64-74	B*45:01	Matthews et al., 2008	5-9	No	Yes	Acidic domain
CAP63	Pol	ALT**E**IC**EE**M	188-196	A*02:01	LANL database	5-11	Yes	Yes	None/Unknown
	Pol	QLTE**A**V**H**KI	522-530	Predicted A*02:01		11-29	No	Yes	None/Unknown
	Vpr	AL**IRI**LQQL	59-67	A*02:01	LANL database	5-11	Yes	Yes	None/Unknown
	Gp41	SWSNKSE**E**DIW**G**NMTWMQ	102-119	A*23:01/Cw*04:01	LANL database	11-29	Yes	Yes	RRE stem I
	Gp41	L**LD**SI**A**I**TV**	303-311	A*02:01	LANL database	2-4	Yes	Yes	None/Unknown
	Nef	ALT**SS**NT**AA**	42-50	A*02:01	LANL database	5-11	Yes	Yes	None/Unknown
	Nef	**EE**VGFPVRPQV	64-74	B*45:01	Matthews et al., 2008	0-2	No	Yes	Acidic domain
CAP85	Pol	K**A**GYVTD**R**GRQKV**VS**LTE	609-626	B*08:01	Matthews et al., 2008	0-5	Yes	Yes	None/Unknown
	Gp41	**RY**L**GS**LVQ**Y**	283-291	A*30:02	LANL database	0-5	Yes	Yes	None/Unknown
	Nef	**KE**VGFP**V**RPQV	64-74	B*45:01	Matthews et al., 2008	0-5	No	Yes	Acidic domain
	Nef	**Y**FPDWQNY	120-127	A*30:02	LANL database	13-29	No	No	Dimerization
CAP210	Gag	VHQ**A**ISPRTL	143-152	B*15:10	Matthews et al., 2008	12-16	No	No	None/Unknown
	Pol	FFR**E**NLAFPEGEARELP**S**	1-18	Unknown	Novel, IFN-ɣ ELISPOT	12-16	No	Yes	None/Unknown
	Vif	**D**WHLGHG**V**SI	78-87	B*15:10	LANL database	5-12	No	Yes	None/Unknown
	Gp41	**EA**TDRI**L**E**L**	313-321	Predicted A*68:02		2-5	Yes	Yes	None/Unknown
CAP239	Gag	TS**TL**QEQV**A**W	240-249	B*58:01	Matthews et al., 2008	0-2	No	Yes	None/Unknown
	Pol	IVLP**EKE**SW	399-407	B*58:01	Matthews et al., 2008	2-5	Yes	Yes	None/Unknown
	Nef	K**AA**V**V**DLSFF	82-90	B*58:01	Matthews et al., 2008	11-22	Yes	No	None/Unknown

*Bold amino acids indicate sites undergoing mutation, ** Predicted epitopes obtained using Net MHCPan2.0 (www.cbs.dtu.dk/services/NetMHCPan)

Table 5. Genome regions evolving under antibody-mediated immune selection

Participant ID	ORF	Genome region sequence* (18 to 20-mer)	HXB2 position	Time of first mutation in hypervariable loop or PNGS# (range)(wks)
CAP45	Gp120	LTRDGG**KTDRNDT**EIFRP	454-470	0- 2
CAP63	Gp120	QE**I**VL**E**N**VI**ENFNMWKND	82-99	5-11
	Gp120	LTPLCVTLNC**A**N**AN**ITKN	122-139	11-29
	Gp120	**MIG**EI**K**NCSFN**ATTELR**D	147-167	2-5
	Gp120	L**NN**N**RSNENS**YILINC**N**S	184-198	0-2
	Gp120	IVH**F**NQSV**KIV**C**A**RPHNN	285-302	11-29
	Gp120	IRQA**H**CNIS**KTQ**W**NTT**L**E**	326-343	11-29
	Gp120	**FNSTYMPNGIHIPNGASEVIT**	396-415	2-5
	Gp41	**L**W**S**WF**NI**S**H**WL**W**YI**RI**FI	158-147	11-29
	Gp41	IE**EE**GG**EQ**D**NS**RS**I**RLVS	222-239	5-11
CAP85	Gp120	D**IVPLNNDIGNYSEY**RLI	180-194	5-13
	Gp120	IVHLN**HS**V**KIV**CTRPGNN	285-302	0-5
	Gp120	IRQAH**C**NIS**KA**EWN**NT**L**E**	326-343	13-29
	Gp120	**GSSTT**T**NGSSP**ITLPCRI	404-420	0-5
CAP210	Gp120	ICSFN**AT**TELRDKKKKEY	156-173	5-12
	Gp120	FNSTHN**ST**D**S**T**V****NSTDS**T	391-409	5-12
	Gp120	ITCIS**NI**TGLLLTRDGGE	443-460	22-26
CAP239	Gp120	DIIIRSQ**NIL**D**N**TK**TIIV**	269-286	2-5
	Gp120	GLLLT**W**DGGDS**KE**N**K**TR**H**	451-467	11-22

*Bold amino acids indicate sites undergoing mutation; underlined amino acids indicate sites evolving under positive selection; highlighted amino acids indicate sites mutating to result in a change within, or gain/loss of a potential N-linked glycosylation site gain/loss (NXS/Tx, where x is not Proline); #PNGS = potential N-linked glycosylation site

Table 6. Genome regions containing mutations associated with reversion

Participant ID	ORF	Genome region sequence*	HXB2 position	Amino acid (% frequency#) change	Timing of reversion (weeks post-infection)
CAP63	Vpu	VGALIIALILT**V**VVWIIA	9-26	V to I (0.9% to 97.85%)	0-2
	Gp41	LAVERYL**R**DQQLLGIWGC**	70-87	R to K (7.28% to 76.18%)	11-29
CAP85	Gag	TS**N**LQEQIAWMT**A**NPPVPVG**E**	240-260	N to T (10% to 85%)	5-13
				A to S (4.4% to 54%)	5-13
				E to D (31.72% to 68.04%)	5-13
	Vpu	IIAIIVWTI**T**YLEYRKVV	17-34	T to A (1.6% to 48%)	0-5
CAP210	Gag	LERFAL**D**PGLLETSGGCK	41-58	D to N (1.5% to 98%)	12-26
	Vpu	YRLGVGA**F**IVALIIAIVV	5-22	F to L (19% to 80%)	12-26
	Nef	EVGFPV**K**PQVPLRPMTYK	65-82	K to R (8.5% to 88%)	5-12
	Nef	LIYSKKRQDILDLW**I**YNT	100-117	I to V (1% to 98%)	12-26
CAP239	Gp120	NDMVDQMH**K**DIISLWDQS	98-115	K to E (0.19% to 93.2%)	11-22

*Only mutations associated with reversion are indicated (shown in **bold**). Known Call I HLA-targeted epitopes (based on Matthews et al., 2008 and the HIV Molecular Immunology Compendium (www.lanl.hiv.gov)) are underlined. Genome regions may also contain changes from high to low frequency amino acids not associated with known CTL-escape. **Participant A*02:01/A*23:01 HLA association according to Los Alamos HIV Molecular Immunology Compendium (http://www.hiv.lanl.gov). #Amino acid frequencies according to Los Alamos subtype C database sequences for each gene (http://www.lanl.gov).

Table 7. Genome regions evolving under unclassified immune selection

Participant ID	ORF	Genome region sequence* (18 to 20-mer)	HXB2 position	Time of first mutation (range) (wks)	Shuffling/ Toggling of AA mutations	Structural/ Functional genome region#
CAP85	Gp120	RPGGG**D**M**K****D**NWRSELYKY	469-486	0-5	Yes	CD4 binding region
	Nef	GVGAASQDLGK**Y**G**G**ALT**SS**	29-46	0-5	Yes	None/unknown
CAP210	Nef	SLH**G**MEDTE**R**EVLQWKFD	169-186	5-12	No	None/unknown
CAP239	Gag	SNPS**G****P**K**R****P****I**KC**F**NCGRE	382-399	2-5	Yes	Start of zinc knuckle
	Rev	GRP**A****E****P**V**P****F**QLPPIERLH	65-82	2-5	Yes	None/unknown

*Bold amino acids indicate sites undergoing mutation, and underlined amino acids indicate sites evolving under positive selection; #Structural and functional genomes regions were identified using the Los Alamos HXB2 annotation spreadsheet (http://www.hiv.lanl.gov/content/sequence/HIV/MAP/annotation.html).

3.6.2 Frequency and timing of early CTL-mediated selection and reversion

Mutations associated with CTL-mediated selection or reversions (possibly of transmitted CTL-escape) were analyzed more closely to determine how early in acute infection these mutations arose and the frequency of their location within the genome. In order to more precisely time the appearance of mutations, supplemental sequence data from clonal and SGA sub-genomic sequences (**Appendix 3**) as well as from targeted epitope sequencing (**Appendix 4**) from additional time-points within the six month period of infection were included in these analyses.

CTL-mediated selection was most frequently identified in Nef (n=6 mutating epitopes), followed by Pol and Env (n=5 mutating epitopes each), Gag and Vif (n=2 mutating epitopes each) and Rev, Tat and Vpr (n=1 mutating epitope each) (**Figure 29A**). Even when accounting for protein length, since longer proteins may contain a greater number of epitope targets or mutational clusters, Nef remained the protein with the highest frequency of mutating epitopes.

The earliest mutations associated with putative CTL-mediated selection were detected at two weeks post-infection in the Gag HLA B*58:01-restricted TW10 epitope and in the Nef HLA B*45:01-restricted EV11 epitope. Thirty-six percent of mutations were identified in the first five weeks of infection and were located in Gag, Pol, Env and Nef, and 32% were identified between five and 12 weeks and included mutations in accessory genes Tat, Rev and Vpr (**Figure 29B, Table 4**). The frequency of mutation was calculated as 1.6 putative escapes/week for the first five weeks of infection, followed by a reduction to 0.9 putative escapes/week between weeks

five and three months, and a further reduction to 0.4 putative escapes per week between three and six months (up to 29 weeks post-infection).

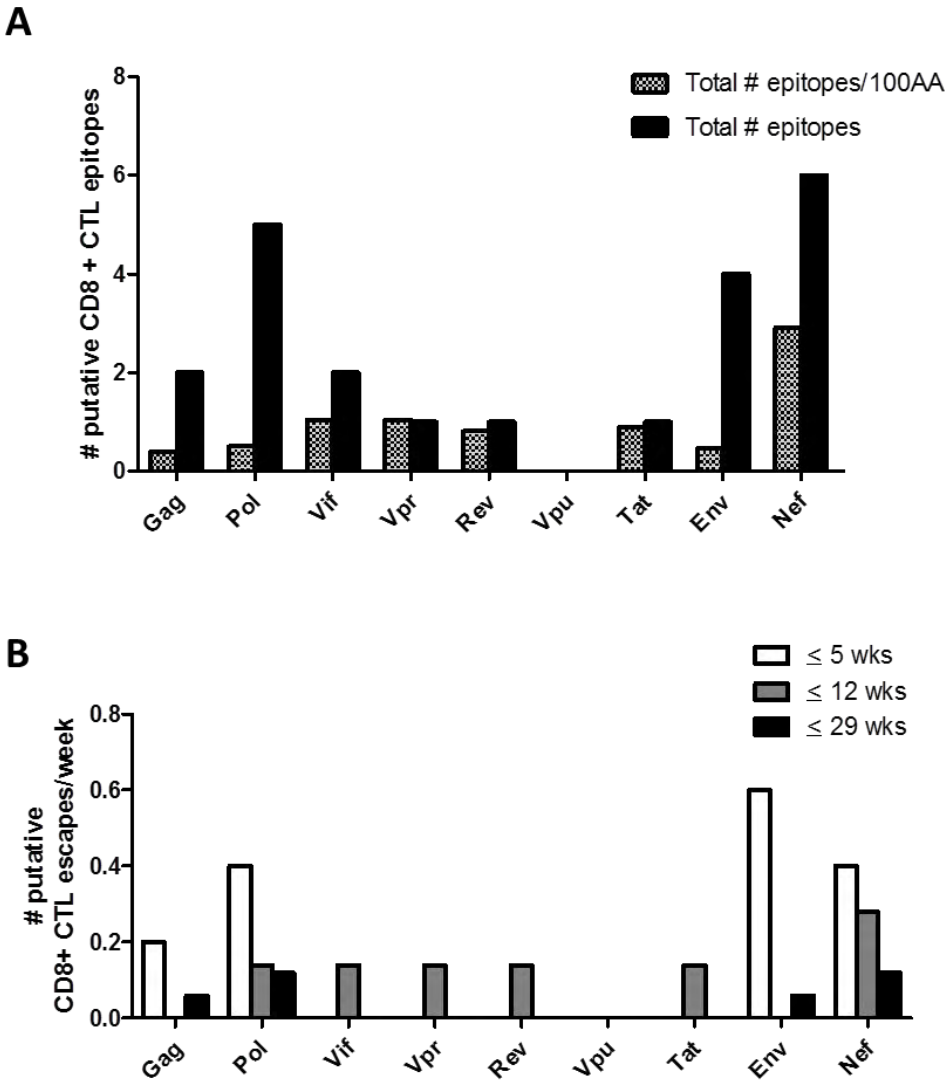


Figure 29. **Frequency and timing of CTL-mediated immune selection in five acutely subtype C-infected women**. (**A**) The number of CTL epitopes (total and total per 100 amino acids) putatively undergoing escape from CTL-mediated pressure across the HIV-1 genome, and (**B**) number of putative CTL-escapes per week over the course of acute and early infection.

Genome regions containing mutations classified as reversion were most frequent in Vpu (n=3), followed by Gag, Env and Nef (n=2 each) (**Table 6**). Four of the nine reversions were rapid and occurred within the first five weeks of infection, while the remaining reversions occurred between 13 and 29 weeks post-infection. The reverting sites in Vpu were all within the first 26 amino acids (according to HXB2 numbering), and did not overlap with another open reading frame.

3.6.3 Frequency and timing of early Ab-mediated selection

Mutations associated with Ab-mediated selection in Env were identified as mutations within PNGSs and hypervariable loops (substitutions or indels). It was of interest to determine how early in acute infection Ab-mediated selection was observed, and also whether early selection was more frequent in specific areas of this structural gene. *Env* SGA sequences were used to supplement sequence numbers at screening/transmission and additional time-points where available (see Table 3, Chapter 2), and to increase sensitivity for timing of selection.

Mutations in seven of the 19 genome regions (37%) that were classified as evolving under Ab-mediated selection were identified in the first five weeks of infection in all participants with the exception of CAP210 (**Table 5**). Earlier studies investigating nAb responses and escape in women from the CAPRISA 002 AI cohort reported, for the five women in this thesis, that the earliest nAb responses were detected at nine weeks post-infection (in participant CAP45) and the latest at 46 weeks post-infection (in participant CAP210) (Moore et al., 2008; Moore et al., 2009). It would therefore seem unlikely that these early mutations were associated with nAb selection. However, in participant CAP45, sites undergoing early mutation in the V5 loop coincided with sites in which neutralization escape was reported to later occur (mutations K460D and D462G) (**Table 5**) (Moore et al., 2009), potentially indicating the earlier presence of nAbs that were not detected. This was the only participant of the five in which very early mutations could be associated with nAb responses.

3.6.4 Mutational pathways in genome regions evolving under immune selection

To evaluate the pathways to escape in genome regions evolving under CTL-mediated selection, additional near full-length genome sequences generated from SGA and limiting dilution PCR, as well as targeted epitope sequences were analyzed (**Appendix 3 and 4**).

In 82% (18/22) of evolving CTL epitopes, mutations were seen to shift between different amino acid positions within the epitope (shuffling) and/or between different amino acids at the same position (toggling). This is illustrated for Nef epitope EV11 restricted by HLA*B45:01 in participants CAP63 and CAP85 (**Figure 30A**). In the case of participant CAP63, by 37 weeks post-infection only one variant was observed in all sequences sampled (due to low sequence numbers complete fixation of this variant cannot be assumed), whereas in the case of CAP85 more than one mutant variant was still present as late as 55 weeks post-infection.

The extent of variation within evolving CTL epitopes in acute and early infection was expressed as the number of mutant variants per epitope at each time-point (**Figure 30B**). With the exception of participant CAP239, the number of variants per epitope increased in all women over the first three months of infection. For CAP63 and CAP210, variation continued to increase over the subsequent three months, whereas for CAP85 the number of variants reached a plateau by six months and thereafter decreased. The number of variants per epitope decreased for CAP63 after six months. In the case of CAP239, an increase in variants per epitope was seen very early (between two and five weeks); however numbers reached a plateau by three months and decreased thereafter. The decrease in variants was in part owing to fixation of mutant variants. Shuffling/toggling was also seen in four of the five genome regions with unclassified selection. Fixation was observed in only 9% of mutating CTL epitopes by six months post-infection and 7% when including genome regions evolving under unclassified selection.

In order to determine whether shuffling/toggling of mutations was more frequent in variable regions of the genome, Shannon entropy scores of mutating epitopes were calculated using database subtype C sequences (www.hiv.lanl.gov). While 12 of the 18 (67%) CTL epitopes with shuffling/toggling corresponded to high entropy epitopes ($\geq$0.25) (Bansal et al., 2005), shuffling/toggling was not significantly more frequent in high entropy epitopes compared to conserved epitopes (p=0.29, Fisher's Exact categorical test).

A

Weeks post-infection	Epitope t/f **EEVGFPVRPQV**	Frequency
2	EEVGFPVRPQV	8/9
	KEVGFPVRPQV	1/9
4	E**K**VGFPVRPQV	15/19
	EEVGFPVRPQV	1/19
	E**D**VGFPVRPQV	1/19
	E**G**VGFPVRPQV	1/19
	VEVGFPVRPQV	1/19
5	E**G**VGFPVRPQV	6/13
	E**D**VGFPVRPQV	2/13
	KEVGFPVRPQV	2/13
	E**K**VGFPVRPQV	2/13
	E**A**VGFPVRPQV	1/13
7	E**G**VGFPVRPQV	9/10
	E**D**VGFPVRPQV	1/10
11	E**G**VGFPVRPQV	4/7
	E**D**VGFPVRPQV	2/7
	DEVGFPVRPQV	1/7
29	E**D**VGFPVRPQV	8/10
	E**G**VGFPVRPQV	2/10
37	E**D**VGFPVRPQV	5/5

Weeks post-infection	Epitope t/f **KEVGFPVRPQV**	Frequency
5	KEVGFPVRPQV	9/10
	KEVGFP**I**RPQV	1/10
13	**ED**VGFPVRPQV	5/9
	KEVGFPVRPQV	2/9
	EGVGFPVRPQV	1/9
	K**G**VGFPVRPQV	1/9
29	**ED**VGFPVRPQV	4/7
	EGVGFPVRPQV	2/7
	GDVGFPVRPQV	1/7
39	**EG**VGFPVRPQV	1/1
55	**ED**VGFPVRPQV	2/4
	GDVGFPVRPQV	2/4

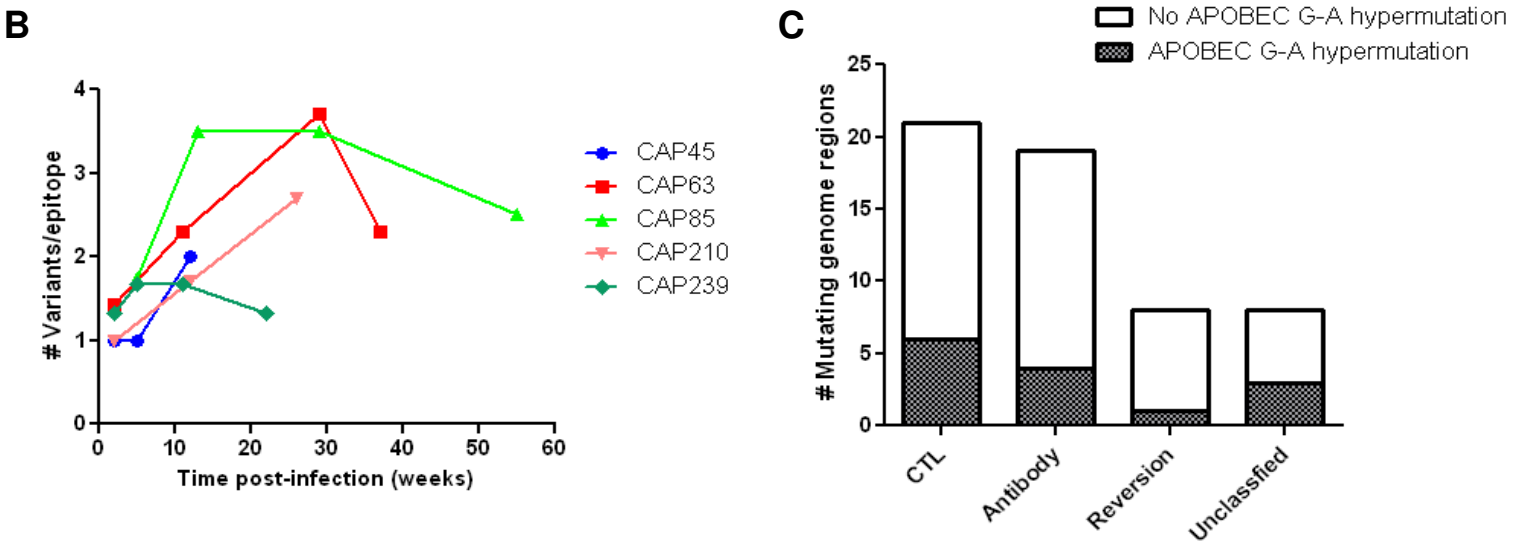


Figure 30. **Mutational variation in epitopes evolving under CTL-mediated immune selection.** (**A**) Shuffling and toggling of amino acid mutations in the B*45:01-restricted Nef EV11 epitope in participants CAP63 (left) and CAP85 (right), (**B**) the number of mutant variants within evolving CTL epitopes for each women is plotted as a function of time, and (**C**) the number of genome regions evolving under selection containing APOBEC-associated hypermutation.

Next we investigated whether shuffling/toggling may be a result of immune selection within structurally or functionally essential regions of the genome, and that the virus therefore struggles to incorporate mutations due to detrimental effects on viral fitness. The proportion of CTL epitopes with shuffling/toggling was however not greater in genome regions of structural or functional importance compared to the proportion in regions with no known structural or functional importance (p=1, Fisher's Exact categorical test). This analysis was repeated with inclusion of the five genome regions with evidence of immune selection that could not be classified, however the comparison remained statistically non-significant (p=1, Fisher's Exact categorical test). In addition, APOBEC G-to-A hypermutation was identified in 36% (8/22) of escaping CTL epitopes, and overall in 25% (13/55) of all genome regions identified as evolving under selection (**Figure 30C**).

3.6.5 Early immune selection and clinical disease progression

To determine whether the extent of mutation associated with selection was associated with disease progression, the numbers of genome regions identified as evolving under selection were compared between participants (**Figure 31**).

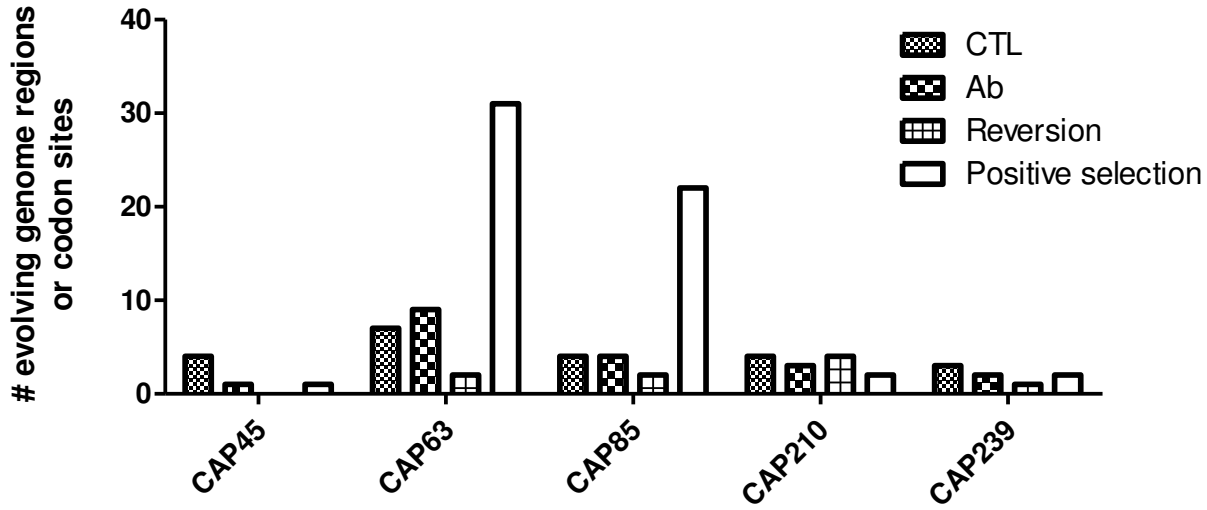


Figure 31. **Intraparticipant quantitative analysis of genome regions/codon sites evolving under selection**. The number of genome regions evolving under CTL- or Ab- mediated immune selection or reversion is indicated. Positive selection data is reported as the total number of evolving codon sites.

Sequences from rapid progressor CAP63 contained the highest number of genome regions evolving under positive (n=31), CTL-mediated (n=7) and Ab-mediated (n=9) selection. Rapid

progressor CAP210 had the highest number of genome regions with reversion (n=4). Viremic controller CAP45 had the lowest number of genome regions evolving under positive (n=1) and Ab-mediated (n=1) selection and had no reverting sites, however sequence data for this individual was lacking at six months post-infection and additional evolving sites may have arisen between three and six months.

3.7 Discussion

Elucidation of responses associated with early virological control provides information on the types of responses needed to be elicited through vaccination. Sequencing of t/f and early evolving viruses allows the identification of targets of these earliest immune responses. Further, these changes can potentially alter viral phenotype and the clinical course of disease. Four studies to date have characterized early viral evolution of near full-length viral genomes in the context of subtype B infection (Salazar-Gonzalez et al., 2009; Goonetilleke et al., 2009; Herbeck et al., 2011; Henn et al., 2012), while studies of the globally dominant subtype C strain are lacking. In this study, peptide regions evolving under selection in near full-length genomes from five subtype C-infected women in acute and early infection were identified, and the immune pressures driving the observed changes within these regions were predicted. Furthermore, the timing, frequency and pattern of early changes were described. Our sequence data illustrates the dominant role of early CTL responses in moulding viral evolution, and together with published studies evaluating early immune responses (Bernardin et al., 2005, Brumme et al., 2008, Salazar-Gonzalez et al., 2009, Goonetilleke et al., 2009), provide strong support for CD8+ T-cell responses as being the most important for virus control following infection. Some of these results were published in Abrahams et al., 2013 and Liu et al., 2013.

Selective pressure from CD8+ cytotoxic T-lymphocytes (CTLs) was predicted to account for the majority (40%) of early changes across the genome. Sixty-four percent of epitopes that were predicted to be evolving under CTL selection based on sequence data, were confirmed by IFN- γ ELISPOT (Liu et al., 2013 and Liu personal communication), providing confidence that the CTL escape prediction approach was robust. For the remaining 36% of epitopes (n=8), seven were identified in participant CAP63 whose cells had poor viability and therefore peptide CTL reactivity could not be confirmed. For the eighth putative epitope (in CAP85), CTL responses were tested at a time-point after escape had already occurred. Therefore a higher percentage of the epitopes predicted by sequence data was likely ELISPOT reactive. This finding for subtype C sequences is in keeping with subtype B studies which reported the majority of early selection to be due to CTL-mediated pressure

(Bernardin et al., 2005, Brumme et al., 2008, Salazar-Gonzalez et al., 2009, Goonetilleke et al., 2009, Herbeck et al., 2011).

Analysis of the timing and frequency of CTL-mediated selection revealed a higher frequency of mutations associated with CTL-escape in the first five weeks of infection and that the frequency of new mutations decreased four-fold over the six month period investigated. The earliest CTL-escape mutation was identified in a Gag epitope, TW10, which is targeted by the immunodominant HLA-B*58:01 response. Rate of escape is reported to be influenced by a hierarchy of responses of varying immunodominance (Liu et al., 2013), wherein more immunodominant responses exert strong selection very soon after infection resulting in early change. In addition, immunodominant responses in early infection were reported to more frequently target high entropy epitopes (Liu et al., 2013) which mutate and escape rapidly, whereas more conserved epitopes escape at slower rates (Ferrari et al., 2011). In line with these findings, for the five women in this study, in the first 12 weeks of infection, 67% of escape occurred in high entropy genome regions compared to only 29% in late infection.

CTL escape mutations were identified most frequently in Nef. This finding confers with that of a subtype B sequencing study which reported early selection to predominate in Nef and Env (Herbeck et al., 2011). While numerous studies have reported Nef to be one of the most immunogenic HIV proteins (Addo et al., 2003; Frahm et la., 2004; Brumme et al., 2007; Wang et al., 2009; Mlotshwa et al., 2010, Bartha et al., 2013, Liu et al., 2013), it should be noted that three of the five women in this study shared the B*45:01 HLA allele which is reported to target the Nef EV11 epitope (Matthews et al., 2008). Selection was identified in this epitope for all three women, possibly accounting for the observed higher frequency of escape in Nef. However, other reasons for more frequent immune targeting of Nef may include that it is expressed early in the virus life cycle, and at higher abundance compared to Gag, Pol and Env which are expressed later (Karn and Stoltzfus, 2012). Due to the major structural and enzymatic roles of Gag, Pol and Env, genome regions encoding these proteins are also likely to be less tolerant of mutations which may have a negative impact on their function.

Ab-mediated selection was predicted to account for 35% of evolving genome regions identified in the five women. In an unexpected finding, mutations typical of Ab selection were identified as early as two weeks post-infection. Since nAb pressure is generally reported to arise only some months after infection (Wei et al., 2003; Gray et al., 2007; Yeh et al., 2010) and the earliest autologous nAb response for the five women in this study was reported at nine weeks (Moore et al., 2009), these changes were initially postulated to be a result of stochastic change, reversion of transmitted Ab escape in the donor, pressure from

non-neutralizing binding Abs, or antibody-dependent cell-mediated cytotoxicity (ADCC). However, a study by Bar et al. (2012) reported evidence of low-level nAb responses not detected by routine assays. Furthermore, for participant CAP45, early changes in the V5 loop identified in this study were found to coincide with sites at which nAb escape was reported for later in infection in a separate study (Moore et al., 2009). Thus the possibility exists that these early changes were a consequence of very early, low titre nAb responses.

Reversion of transmitted mutations accounted for 16% of changes in sites evolving under selection. Transmitted CTL escape mutations from the donor possibly accounted for 44% of reverting genome regions, however this is not conclusive in the absence of donor HLA or donor viral sequence information. One third of reversions identified were in a common amino acid stretch in Vpu. These genome regions did not correspond with known CTL epitopes and contained no known functional or structural motifs, and therefore cannot be postulated to have been a result of transmitted escape or to have been fitness related. The explanation for the higher frequency of reversion in Vpu in this study is unknown. In a study by Li et al (2007), reversions were reported to arise significantly earlier than escape mutations; however this was not found to be the case for this study as well as similar studies by Goonetilleke et al. (2009) and Herbeck et al. (2011) which analyzed early selection across the genome of t/f viruses from three subtype B-infected individuals using SGA. This may be owing to differences in methodology. The derivation of t/f virus sequences together with the sensitivity of SGA of circulating RNA virus, represent a more accurate approach to identify changes to the viral genome compared to the approach used by Li et al. who amplified from proviral DNA using limiting dilution PCR and population sequencing of pooled PCR products, and who did not have the t/f virus sequence as a point of reference.

Finally, the pressure driving changes in 9% of genome regions predicted to be evolving under selection could not be classified. Changes may represent compensation of the escape mutations identified, or may have been a consequence of pressure from NK cells, CD4+ T-cells, ADCC, non-neutralizing binding Abs, or perhaps have been related to evolutionary drift or enhancement of viral fitness. These were not investigated further in this study; however they highlight that more remains to be understood regarding selection in early HIV infection and potential barriers to disease control and vaccine efficacy.

An increasing number of studies propose that epitopes which are stable and conserved, and which subsequently escape slowly, would represent good vaccine targets (Davenport et al., 2008; Matthews et al., 2008; Goonetilleke et al., 2009). A subtype B study by Goonetilleke et al. (2009) reported the presence of multiple mutant variants in 15 of 18 (83%) early escaping CTL epitopes. In the current study, 82% of mutating CTL epitopes displayed

extensive shuffling and toggling of amino acid mutations within evolving genome regions. It may be that such variation is seen in regions of the genome that are not able to tolerate change due to potential costs to viral fitness as proposed in a recent study (Herbeck et al., 2011), however no evidence was found to indicate more frequent shuffling/toggling in genome regions of known functional or structural significance. Alternatively, the opposite could be proposed, that such variation is seen since these regions *can* tolerate extensive mutation. This is supported by the fact that 61% of epitopes with shuffling/toggling corresponded to high entropy database epitopes. Henn et al. (2012) performed a detailed sequencing study of viral evolution in one participant, and also observed mutational shuffling/toggling in targeted epitopes, which they proposed could be a pattern which persist until such time that a mutation is introduced which facilitates complete viral escape that will subsequently become fixed. This would however need to be tested *in vitro*.

Furthermore, in this study it was found that APOBEC-mediated hypermutation contributed to the evolution and potentially to the escape of a significant portion (25%) of genome regions evolving under selection (36% of CTL epitopes). While the viral protein Vif functions to counteract the activity of APOBEC (Malim and Bieniasz, 2012), it would appear that the virus has evolved to use activity of this host-restriction factor to its advantage.

One advantage of this study is that the five participants display differential disease progression, which allows the description of immune selection 'profiles' in the context of viremic control, and rapid and intermediate progression. Virus from one rapid progressor, participant CAP63, who was referred to ARV treatment by nine months post-infection, had the highest number of genome regions evolving under selection. One may expect that a greater breadth of immune targeting would place more constraints on viral replication and would thus not be associated with a more rapid disease progression, unless targeting frequently resulted in immune escape. Supporting this, studies by Gray et al. (2009) and Mlotshwa et al. (2010) which examined the relationship between breadth of CTL responses and disease progression in women form the CAPRISA 002 cohort, reported a trend for increased breadth of CTL responses in rapid progressors, but a significant difference in retention of responses between rapid progressors and viremic controllers, wherein responses were lost more frequently in rapid progressors. Due to low participant numbers, the associations between breadths of overall positive, CTL- or Ab-mediated selection with set-point viral load or CD4 count at 12 months post-infection, both of which are predictors of disease progression (Mellors et al., 1997; Lyles et al., 2000), could not be tested.

In summary, this study reports early immune selection across the subtype C genome to be predominantly attributed to CTL-mediated immune pressure, that nAb-mediated pressure

may be prevalent earlier than expected, and that the mutational pathway to immune evasion or adaption is dynamic and complex. In the chapter to follow, the impact of early CTL immune selection on viral fitness will be investigated.

Chapter 4: The effect of CTL escape mutations on viral replication fitness

4.1 Abstract

Studies evaluating the impact of CTL escape mutations on viral replicative fitness typically introduce mutations into a viral backbone that does not match the naturally occurring cognate virus. This approach is limited as it does not take into consideration co-evolving mutations that may restore viral function. In this chapter we investigate the impact of early CTL escape mutations on viral replication fitness introduced into infectious molecular clones encoding the cognate transmitted/founder virus sequence.

Four transmitted/founder infectious molecular clones were available, and a fifth was constructed as part of this thesis. CTL escape mutations that evolved in HLA-B-restricted epitopes in the first six months of infection were introduced into four of the clones by site-directed mutagenesis. The replication capacities of transmitted/founder and mutant viruses were compared in parallel replication assays in peripheral blood mononuclear cells.

The transmitted/founder virus from a rapid progressor, CAP63, had the highest replication capacity ($p<0.0001$). In two individuals, HLA-B associated escape mutations did not affect replication fitness. In the third individual (CAP210), introduction of a rare mutation (A146T) in the B*15:10-restricted VL10 epitope situated in the Gag p24 capsid, together with a V85A mutation in the Vif WI9 epitope, resulted in a significant loss in viral replicative fitness relative to the cognate transmitted/founder virus ($p=0.0004$). In the fourth individual (CAP239), a Pol E343K mutation together with the well-described T242N p24 capsid mutation significantly enhanced viral replication fitness relative to the cognate transmitted/founder virus ($p=0.0072$). The CAP239 transmitted/founder virus harboured compensatory mutations H219Q and I223V, which when reversed resulted in the lowering of viral replicative capacity of the virus containing the T242N and E343K mutations to below that of the transmitted/founder virus ($p=0.0015$).

This study showed that cognate sequence information can significantly influence the interpretation of the effect of escape mutations on viral replicative capacity, as illustrated in this instance by CTL-escape mutations which increased replication fitness due to pre-existing compensatory mutations in the transmitted/founder virus.

4.2 Introduction

Viral fitness is defined as the ability of a virus to survive and propagate in a given environment (reviewed in Domingo et al., 1997). It has been proposed that, within an infected individual, viral fitness increases with increasing viral load over the course of HIV infection (Troyer et al., 2005; Arien et al., 2007). Increasing fitness may be a result of the accumulation of immune escape mutations over time (Richman et al., 2003; Allen et al., 2005) or evolution of variants to the more virulent CXCR4 receptor-utilizing phenotype (Shankarappa et al., 1999; Troyer et al., 2005).

However, the accumulation of immune escape mutations, such as those selected by CTL pressure, can also reduce viral replication fitness when arising in functionally conserved regions of the viral genome (Martinez-Picardo et al., 2006). The identification and characterization of such mutations furthers our knowledge of the vulnerabilities of HIV and of potential vaccine targets. This study examines the impact of CTL escape mutations in cognate full-length genome t/f virus backbones on replicative fitness. This approach provides an advantage over most published replicative fitness studies which have almost exclusively been done in recombinant backbones which do not include potential compensatory mutations that could restore fitness.

4.2.1 Assaying viral replicative fitness

To evaluate differences in replication capacity, viruses are either competed against each other following co-infection of cell culture, or their growth kinetics is compared independently in parallel cultures under standardized conditions (reviewed in Quinones-Mateu and Arts, 2002 and Dykes and Dementer, 2007). While competition assays are reported to be more sensitive for detecting subtle fitness differences (Collins et al., 2004; Prado et al., 2004), recombination between the competing viruses is a major confounder (Jetzt et al., 2000; Jung et al., 2002). Song et al. (2012) reported recombination as early as three to four days post culture co-inoculation. An additional drawback of competition assays is that they require expensive methods of quantification such as deep sequencing or real-time PCR to distinguish between viruses. Parallel growth kinetics assays, monitored using rapid and less costly quantitative assays such as by p24 ELISA or flow cytometry (reviewed in Quinones-Maeu and Arts, 2002 and Dykes and Dementer, 2007), have successfully been used to detect fitness costs conferred by CTL escape mutations (Brockman et al., 2007; Chopera et al., 2012; Boutwell et al., 2013), and are used in this study.

The cells used to evaluate replicative capacity can affect assay results. CD4+ T-cell-lines are frequently used as the target cells and proliferate rapidly in culture, are relatively homogeneous, and are designed to be readily infected by HIV through enhanced expression of CD4 and CCR5/CXCR4 cellular receptors. While cell-lines provide a standardized environment for comparing virus replication, and allow for more rapid assaying of viral fitness, they do not necessarily provide a realistic reflection of how viruses expand within the complex cellular environment of an infected individual. Furthermore, depending on the assay used, fitness differences may be detected in some cell-lines but not in others (Brockman et al., 2007). For these reasons, donor-derived peripheral blood mononuclear cells (PBMCs), which consist of a mixture of cells including CD4+ T-cells, CD8+ CTLs and macrophages, and are utilized for more *in vivo*-relevant fitness assays, were used in this study.

The impact of genetic mutations on viral replicative fitness is evaluated by the introduction of these mutations into a viral genetic backbone cloned into a DNA plasmid vector, known as an infectious molecular clone (IMC) (Adachi et al., 1986; Mochizuki et al., 1999; Salminen et al., 2000; Ndung'u et al., 2001). These constructs express viral genes for packaging of infectious virions in mammalian cells and can therefore be used to study viral replication. The introduction of mutations into IMCs containing full-length cognate viral genomes permits evaluation of the true impact of these mutations in the host in which they were identified. While the generation of such IMCs is challenging and costly, the resulting constructs represent valuable research tools, not only for fitness studies but also for immunological and transmission studies among others.

4.2.2 Fitness of the transmitted virus and disease progression

While genome-wide association studies (GWAS) have reported clinical outcome to be largely dictated by Class I HLA genotype, in particular HLA-B (Fellay et al., 2007, Fellay et al., 2009, Pereyra et al., 2010, Bartha et al., 2013), an increasing number of reports indicate a greater than previously appreciated contribution of the genotype of the transmitted virus (Yue et la., 2012; Lingappa et al., 2013; Fraser et al., 2014). For example, the transmission of CTL-escape mutations in epitopes targeted by advantageous HLA types, such as B*58:01, has been shown to afford a degree of disease control when transmitted to recipients with mismatched HLA types if they do not revert (Chopera et al., 2008, Goepfert et al., 2008). This may be owing to the effect of these mutations on viral replicative fitness.

Replicative capacity has also previously been correlated with disease progression. For example, viruses (both viral isolates and chimera viruses derived from IMCs) from clinical

progressors were reported to display greater replicative capacity than those of long-term survivors, and viral load was correlated with viral replication capacity (Blaak et al., 1998, Quinones-Mateu et al., 2000, Campbell et al., 2003, Trkola et al., 2003, Troyer et al., 2005, Miura et al., 2009). In two separate studies using *gag-pro* recombinant NL4-3 constructs, viruses from individuals with protective HLA-alleles B*57/58:01, B*13 and B*81 were found to have lower replication capacities (Brockman et al., 2010, Wright et al., 2011) indicating a link between disease control and lower replication fitness. Prince et al. (2012) showed that the replicative fitness of viruses expressed from chimera backbones encoding transmitted *gag* sequences from a large Zambian cohort was positively correlated with set-point viral load. However, other studies investigating a relationship between disease progression and replication fitness of *gag-pro* chimeras (Wright et al., 2011) and viral isolates (Arnott et al., 2010) derived from acute/early infection did not find a correlation with viral load.

Recently, the loss of replication capacity incurred by CTL-escape mutations was reported in a comprehensive study of mutations commonly identified in Gag epitopes, for example those in protective HLA-restricted epitopes KF11 (A136G), IW9 (A146P and I147L) and TW10 (T242N and G248A) (Boutwell et al., 2013). In a separate study, a significant association between a number of these escape mutations, including T242N, G248A and I147L with lower plasma viral loads was identified (Bartha et al., 2013). Compensation of these escape mutations has been demonstrated to effectively restore viral replicative capacity (Brockman et al., 2007; Crawford et al., 2007; Brockman et al., 2010).

In this chapter, replication capacity of viruses produced from IMCs representing the cognate t/f virus of five subtype C-infected women focused on in this study. IMCs were mutated in order to evaluate the impact of early CTL immune escape on viral fitness in parallel replication kinetics assays. Mutations introduced were selected based on their presence in known Class-I HLA-B-restricted epitopes and their being fixed or gaining dominance within the first six months of infection (immune selection data is described in Chapter 3). In addition, the replication fitness of the t/f viruses was compared in the context of disease progression profile.

4.3 Research Aim and Objectives

Aim: To determine the relationship between replication capacity (RC) of t/f viruses and viral load set-point, and the effect of early CTL-escape mutations on RC.

Objective 1: Characterize and compare RC of t/f viruses in the context of disease progression profile

Objective 2: Introduce selected CTL-escape mutations into IMCs of four women by site-directed mutagenesis and compare the RC of t/f and early CTL-escape mutant viruses in parallel replication kinetics assays

4.4 Rationale

To further our understanding of determinates of disease control, this study evaluated the replicative fitness of t/f viruses and the effect of early immune selection on replication. IMCs generated as part of this study provide a valuable biological resource for characterizing viral pathogenesis.

4.5 Methods and Materials

The IMCs containing t/f viruses of participants CAP45, CAP63, CAP210 and CAP239 were generated by F. Treurnicht (University of Cape Town) and H. Ding (University of Alabama at Birmingham) in the laboratories of Prof. Carolyn Williamson and Dr. John Kappes.

The CAP200 t/f virus IMC was constructed as part of this thesis.

4.5.1 Design and construction of the CAP200 t/f virus IMC (pCAP200.t/f)

4.5.1.1 Determination of the complete LTR sequence

A consensus of SGA near full-length viral genome sequences amplified from viral RNA from plasma drawn at approximately two weeks post-infection for participant CAP200 was previously derived by S. Goodier (University of Cape Town). These sequences contained partial LTR sequences at either end since complete LTRs may only be amplified from proviral DNA. The 5' end of the amplified sequence contained the complete LTR U5 region and the 3' end contained the complete U3 and R regions. The complete U3-R-U5 LTR sequence could thus be pieced together for both ends of the genome. Los Alamos HIV Sequence Compendium was used to identify the start and end sequences of LTR regions (http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/compendium.html).

4.5.1.2 Restriction enzyme digestion and cloning

Restriction enzyme sites within the CAP200 viral genome were identified using NEBcutter 2.0 (Vincze, Posfai and Roberts, 2003; http://tools.neb.com/NEBcutter2/index.php) and the Web Map Preferences tool at http://pga.mgh.harvard.edu/web_apps/web_map/start by S Goodier. Genome fragments were synthesized and cloned into pUC57 by GenScript, NJ, USA. Fragments digested from pUC57 were cloned into the pBR322 vector (Promega, WI, USA).

For each digestion, approximately 3-7 µg of plasmid DNA was digested using NEB restriction enzymes (New England BioSciences, UK) according to manufacturer instructions for 3 hours or overnight. Digested insert and vector bands were gel-purified using the Qiagen Gel Extraction Kit (Qiagen, Limburg, Netherlands) according to manufacturer instructions and ligated overnight at 16°C using T4 ligase (Thermo Scientific, MA, USA) in a 20 µl final reaction volume. A volume of 5 µl of ligation reaction was transformed into One Shot Stbl3 chemically competent cells (Invitrogen, CA, USA) according to manufacturer instructions, with one exception: after heat shocking, cells were incubated in 250 µl Luria broth instead of 500 µl SOC medium followed by centrifugation at 4 000 g for 5 minutes to pellet cells and all cells were resuspended in 100 µl and plated. Cells were plated onto ampicillin-containing LB agar (0.1mg/ml final concentration) and incubated at 32°C overnight or at room temperature for three days. Colonies were screened by restriction enzyme digestion following small-scale plasmid DNA isolation using the Qiagen Plasmid Mini Kit (Qiagen, Limburg, Netherlands) according to manufacturer instructions.

The In-Fusion HD Cloning Kit (Clontech, CA, USA) was used according to manufacturer instructions to clone genome fragment II into pBR322 containing fragments I and III. Primers were designed to amplify fragment II, with the forward primer designed to incorporate a 15 base pair overlap with the 3' end of fragment I and the reverse primer designed to incorporate a 15 base pair overlap with the 5' end of fragment III. Using these primers, fragment II was PCR amplified from pUC57 using Platinum Taq DNA Polymerase High Fidelity (Invitrogen, CA, USA) to minimize the introduction of PCR error.

PCR reagent mix and cycling conditions for amplification of CAP200 fragment II:

Reagent	Volume (µl)
dH_2O	15.3
10 X High Fidelity Buffer	2
$MgSO_4$ (50 mM)	0.8
dNTPs (10 mM)	0.4
CAP200T/F_F2_For (20µM)	0.2
CAP200T/F_F2_Rev (20µM)	0.2
Platinum Taq High Fidelity	0.1
pUC57_Fragment II (10 ng/ul)	1
Total	**20**

Temperature (°C)	Incubation Time	Number of Cycles
94	2 min	
94	15 sec	25
55	30 sec	
68	4 min*	
68	20 min	
4	Hold	

*An additional 15 sec of template elongation time was added on per cycle

The resulting PCR product was gel-purified as described earlier. A volume of 3 µl (66 ng) purified PCR product was combined with 4 µl of PflFI/BstBI-digested pBR322 containing fragments I and III (50 ng) and 2 µl In-Fusion HD Enzyme Premix and incubated at 50°C for 15 minutes followed by 4°C until transformation. A volume of 2.5 µl of the In-Fusion reaction was transformed into One Shot Stbl3 chemically competent cells (Invitrogen, CA, USA) as described earlier.

4.5.2 Site-directed mutagenesis

Nucleic acid mutations were introduced into IMC backbones by PCR using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies, CA, USA) according to manufacturer instructions. Unique forward and reverse mutagenesis PCR primers, which overlapped the participant-specific epitope to be mutagenized and which incorporated the desired nucleic acid mutation, were designed using the on-line QuikChange Primer Design Program (www.agilent.com/genomics/qcpd). Where necessary, primer lengths were modified so as to have a guanine or cytosine residue at the 5' and 3' termini to facilitate efficient primer binding, while ensuring that the desired mutation remained centralized within the primer sequence.

Mutagenized PCR products were transformed into XL-Gold Ultracompetent cells provided with the QuikChange II XL Site-Directed Mutagenesis Kit according to manufacturer instructions, with the exception that cells were incubated in LB Broth and not NZY^+ broth, or into One Shot Stbl3 chemically competent cells (Invitrogen, CA, USA) as described in section 4.4.1. Small and medium-scale DNA preparations from bacterial cultures were carried out using the Qiagen Plasmid Mini and Midi Kits (Qiagen, Limburg, Netherlands)

respectively according to manufacturer instructions. The full genome of all constructs were screened by sequencing at the Stellenbosch Central Analytical Facility using an ABI 3000 genetic analyser (Applied Biosystems, Foster City, CA, USA) and BigDye terminator reagents to confirm the presence of desired mutation and that no unwanted mutations were introduced by the PCR.

4.5.3 Generation of virus stocks

IMC DNA was transfected into HEK293T cells (ATCC CR-L11268) for the production of virus using PolyFect Transfection Reagent (Qiagen, Limburg, Netherlands). HEK293Ts are immortalized human embryonic kidney cells which are highly transfectable (Graham et al., 1977). Cells were maintained at 37°C with 5% CO_2 in Dulbecco's Minimal Essential Medium (D-MEM) 4.5g/L Glucose (Lonza, Basel, Switzerland) supplemented with 10% Fetal Calf Serum (Biochrom GmBH), 50 µg/ml Gentamicin antibiotic (Lonza, Basel, Switzerland) with passaging at 80-100% confluency. The PolyFect reagent is positively charged and associates with the negatively charged DNA molecules to form complexes which are taken up by the HEK293T cell membranes, followed by expression of viral proteins and packaging of virus particles within the cells.

A total of 5 x 10^5 cells were seeded per well of a 6-well cell culture plate and incubated overnight at 37°C with 5% CO_2. The following day, the medium in each well was replaced with 1.5 ml of fresh supplemented D-MEM. For each virus to be produced, a total of 2.5-5 µg of IMC DNA was combined with 100 µl of unsupplemented D-MEM and 20 µl PolyFect Transfection Reagent (Qiagen, Limburg, Netherlands) in a 15 ml conical tube, vortexed for 10 seconds and incubated for 10 minutes at room temperature to allow formation of DNA-Polyfect complexes. A volume of 600 µl of supplemented D-MEM was then added to the mixture followed by gentle mixing and slow drop-wise addition of the complexes to a single well of HEK 293T cells for each virus to be produced. Cells and DNA complexes were then incubated for 72 hours at 37°C with 5% CO_2. Viruses were harvested following aspirating of medium from each well and centrifugation at 1 500 rpm for 10 minutes to pellet cells. Supernatants containing virus particles were aliquoted into cryovials and stored at -80°C.

4.5.4 Viral titration

The titre of each virus batch harvested was determined by infection of TZM-bl cells (NIH AIDS Research and Reagent Program, Division of AIDS, NIAID, NIH, from Dr John Kappes,

Dr Xiaoyun Wu and Tranzyme Inc., catalogue number 8129). TZM-bl cells are a HeLa cell-line engineered to express luciferase and β-galactosidase under the control of a Tat-inducable HIV LTR promoter, and stably express receptors CD4+ and CCR5 for HIV infection (Platt et al., 1998; Wei et al., 2002). Infection of cells can thus be measured by either luminescence or β-galactosidase staining. Cells were maintained at 37°C with 5% CO_2 in Dulbecco's Minimal Essential Medium (D-MEM) 4.5 g/L Glucose (Lonza, Basel, Switzerland) supplemented with 10% Fetal Calf Serum (BioChrom GmBH), 50 µg/ml Gentamicin antibiotic (Lonza, Basel, Switzerland) with passaging at 80-100% confluency.

A virus aliquot from each batch was first subjected to a single freeze/thaw cycle in order for the titre to be reflective of stored vials of frozen virus from the same batch. A volume of 100 µl of supplemented D-MEM containing Dextran (40 µl per 10 ml D-MEM; the optimal concentration of Dextran used was determined based on batch serial dilution and toxicity testing) was added to each well of the first four rows of a 96-well cell culture plate. Dextran is a polysaccharide that enhances the permeability of cellular membranes thereby allowing for easier uptake of viral particles into cells. To each of the four wells in column 1, a volume of 25 µl of virus was added followed by thorough mixing. Using a multi-channel pipette, 25 µl of virus from column 1 was transferred to column 2 and this was repeated across the plate until column 11 to produce a quadruplicate 1:5 dilution series, with thorough mixing and changing of tips between transfers. The 25 µl excess from column 11 was discarded and column 12 was left as a negative column with no virus added. A volume of 100 µl of TZM-bl cells at a concentration of 1 x 10^5 cells/ml in supplemented D-MEM containing Dextran (40 µl per 10 ml D-MEM) was then added to each well of the four rows. The titration plate was then incubated at 37°C with 5% CO_2 for 48 hours.

A volume of 100 µl of medium was removed from each well and discarded. To the remaining cells and media, 100 µl of Bright-Glo luciferase substrate (Promega, CA, USA) was added per well followed by an incubation step of two minutes at room temperature to allow for cell lysis and the interaction of luciferase released from infected cells with substrate. Luciferase expression was measured on a luminometer as Relative Light Units (RLU). Viral titre was taken as the viral dilution at which 50% of wells were infected (infected wells were determined based on RLU values at least 2.5 times above background), known as TCID50 and calculated using a modification of the Reed-Meunch method.

4.5.5 Parallel replication kinetics assays

PBMCs from three HIV-negative donors were stimulated for 72 hours with the IL-2 cytokine (Gentaur, Belgium) (200 U/ml final concentration) and phytohemagglutinin-P (PHA-P) lectin (Remel, Thermo Scientific, MA, USA)(0.5 µg/ml final concentration) in RPMI 160 medium with 25 mM HEPES (Lonza, Basel, Switzerland) supplemented with 10% Fetal Calf Serum (Biochrom GmBH), 50 µg/ml Gentamicin antibiotic (Lonza, Basel, Switzerland) and 2mM L-glutamine (Lonza, Basel, Switzerland) at 37°C with 5% CO_2. The IL-2 and PHA-P activate cells and stimulate cellular proliferation.

Stimulated PBMCs from each donor were CD8-depleted using Dynabeads CD8 (Invitrogen, CA, USA) according to manufacturer instructions. For these assays, it is necessary to deplete PBMCs of $CD8^+$ CTLs prior to infection to prevent *in vitro* CTL-directed killing of HIV-infected cells and exertion of immune pressure resulting in *de novo* escape mutations evolving *in vitro*. Briefly, 8 x 10^6 to 1 x 10^7 cells from each donor were incubated in isolation buffer (PBS without calcium and magnesium, 0.1% BSA and 2 mM EDTA) with 1 x 10^7 magnetic beads coated with monoclonal anti-CD8 membrane antigen antibodies which bind human $CD8^+$ T-lymphocytes. By applying an EasySep magnet (Stemcell Technologies, Canada), bead-bound cells were sequestered while the remaining non-$CD8^+$ cells were poured off and pelleted by centrifugation at 2 000 rpm for 8 minutes. Equal amounts of CD8-depleted PBMCs from each donor were then combined and diluted to a final concentration of 750 000 cells/ml in supplemented RPMI 160 containing IL-2 (200 U/ml) referred to as IL-2 medium.

Each virus was used to infect cells in triplicate wells of a 96-well cell culture plate at a multiplicity of infection of 0.001 (75 TCID50 in 75 000 PBMCs) in a final volume of 200 µl. As a virus inoculum-only background control, 75 TCID50 of virus was added to an additional well containing only medium and no cells. Cultures were incubated at 37°C with 5% CO_2 and 50 µl of medium was removed with replacement on day five post-infection and thereafter every two days for two weeks and frozen at -20°C. No fresh cells were added over the course of the assay. The virus inoculum control wells were treated in the same manner. The Gag p24 concentration of virus supernatant was measured using a p24 sandwich ELISA described in section 4.4.6. The virus inoculum control p24 concentration was subtracted from p24 concentrations of infected cell cultures so as to determine the amount of new p24 produced as a result of viral expansion over time.

A replicative capacity (RC) score was calculated for each replicate infection growth curve using two methods: (i) the rate of viral expansion in ng p24 per day up to the point of peak p24 concentration was determined and (ii) the slope of replication was determined for day 0

to each subsequent sampling day up to the p24 concentration plateau/decrease and averaged as according to the method described by Weber et al., 2011.

4.5.6 Gag p24 antigen ELISA

An HIV-1 p24 sandwich ELISA used for measuring viral expansion in culture was modified from that originally described by Moore et al. (1990) and McKeating et al. (1991) and adapted by Dr A Trkola. In this protocol, p24 antigen from the viral lysates is bound to a polyclonal anti-HIV-1 p24 antibody immobilized on the surface of an ELISA assay plate. A monoclonal secondary anti-HIV 1 p24 antibody with an alkaline phosphatase conjugate is then bound to the p24 antigen. Finally, an alkaline phosphatase chemiluminescent substrate is added and p24 concentration is measured by light emission in relative light units (RLU). Details of the protocol used are as follows:

White, opaque high-binding 96-well ELISA plates (Porvair Sciences, United Kingdom) were coated with 100 µl/well D7320 sheep polyclonal anti-HIV-1 p24 antibody (Aalto Bio reagents, Dublin, Ireland) dissolved in 0.1M $NaHC0_3$ buffer pH 8.5 to a final concentration of 33.3 µg/ml and incubated overnight at room temperature. Plates were washed three times with 1 x Tris-buffered saline (TBS) and left to dry before freezing at -20°C. Virus supernatants were inactivated by 1:10 dilution in 1,1% Empigen detergent (Sigma-Aldrich, MO, USA) in 1 x TBS for one hour at room temperature. This incubation allowed for release of p24 antigen from virus particles. Further dilution of lysed virus particles was carried out using 10% IL-2 medium (as described in section 4.4.5), 1% Empigen in 1 x TBS. A volume of 100 µl of diluted sample was loaded onto wells of pre-coated ELISA plates. For generation of a standard curve from which to derive sample p24 concentrations, a seven-step 1:2 dilution series of recombinant HIV-1 p24 antigen standard (Aalto Bio Reagents, Dublin, Ireland) starting at 12.5 ng/ml was prepared and loaded in duplicate onto each ELISA plate (100 µl/well). Plates were incubated for two hours at room temperature to allow for binding of p24 to antibody coated on the surface of the wells and thereafter washed three times with 1 x TBS. A volume of 100 µl of BC 1017-AP anti-HIV-1 p24 secondary conjugate antibody (Aalto Bio Reagents, Dublin, Ireland) diluted 1:12 500 in 0.1% Tween-20 in 1 x TBS was then added to each well followed by incubation at room temperature for one hour and thereafter four washes with 1 x TROPIX buffer (0.2 M Tris, 10 mM $MgCl_2$, pH 9.8 with 1M HCL) containing 0.1% Tween-20. A volume of 100 µl of CPD-Star with Sapphire II enhancer substrate (Applied Biosystems, CA, USA) diluted 1:10 in 1 x TROPIX buffer was added to each well and incubated at room temperature for 30 minutes away from light. Luminescence of each well was measured in RLU using a Glomax luminometer (Promega, MA, USA).

Concentrations of p24 for all samples were extrapolated using the linear range of standard curve concentrations plotted on a log-log scale and fitted with a power regression trendline and equation. Viral replication kinetics curves were plotted as average p24 concentration (ng/ml) over time.

4.5.7 Statistical tests

Replicative capacity scores were compared using two-tailed unpaired Student's t-tests and one-way ANOVA tests with Bonferroni correction for multiple comparisons. All statistical tests were run using GraphPad Prism version 5 (www.graphpad.com).

4.6 Results

4.6.1 Viral replicative fitness of t/f viruses

IMCs were generated from individuals classified as infected with a single t/f *env* variant (see Chapter 2 and Appendix 1), of which four were provided (F. Treurnicht and J. Kappes) and one (pCAP200.t/f) was constructed for this thesis (see **Appendix 5** for details). IMCs were sequenced to ensure a match to the derived t/f full-length genome sequence (Chapter 2). For participant CAP63, who was infected with a second, minor variant, both t/f variant IMCs were available, however only the dominant variant was used for replication assays as the minor variant was not identified at later time-points.

T/f viruses were generated by transfection of IMCs into HEK293 cells. Replication kinetics was evaluated following infection of CD8+ T-cell-depleted, pooled PBMCs derived from three HIV-negative donors at an m.o.i. of 0.001. Replication was monitored every two days from day five to 15 post-infection by harvesting of viral supernatant and measuring Gag p24 antigen content. Replication capacity (RC) was expressed as the rate of increase of p24 concentration per day until peak p24 concentration or the mean of slope derived from day 0 to each day of sampling up to the p24 concentration plateau/decrease (Weber et a., 2011). T/f viruses from participants CAP63 and CAP239 expanded more rapidly than the other three viruses and plateaued earlier (by day 13) than viruses from participant CAP45 and CAP210 which continued to expand over the full two weeks. The t/f virus from participant CAP200 also plateaued by day 13, despite slower expansion than CAP63 and CAP239 t/f viruses.

The RC of t/f viruses from participants CAP45, CAP63, CAP200, CAP210, CAP239 were compared (**Figure 32A**). On average, CAP63.T/F had a 2.5-fold greater rate of expansion and a 3.8-fold greater mean slope than that of the other four viruses (**Figure 32B**).

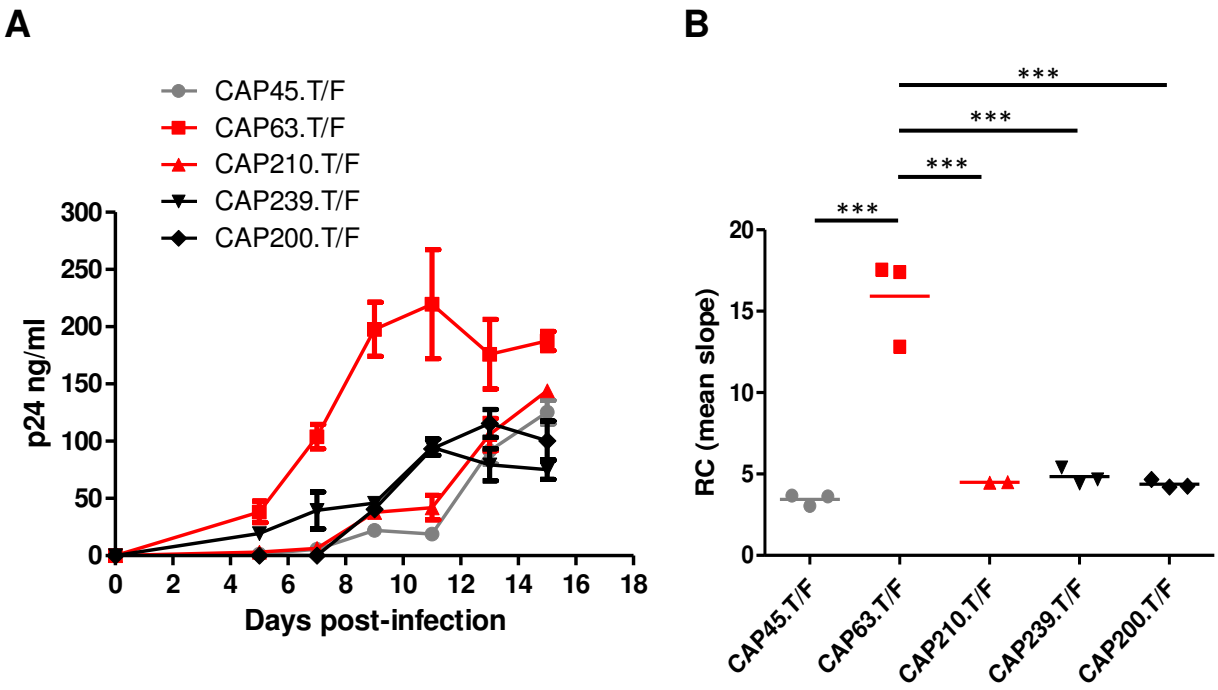


Figure 32. Replication kinetics of CAPRISA t/f viruses. (**A**) Gag p24 antigen concentrations of viral supernatants from infected PBMCs (pooled form three donors) sampled every two days from day 5 post-infection are illustrated. PBMCs were infected at an equal TCID50 m.o.i. for all viruses of 0.001. Mean p24 concentrations of triplicate infections are plotted with error bars indicating standard deviations. (**B**) Mean slopes of viral expansion from replicate infections with t/f viruses from the five women are plotted. Asterisks indicate the level of significance of differences in mean RC values obtained from unpaired parametric Student's t-tests (*** = $p<0.001$). Viruses are colour-coded according to disease progression profile: grey = viremic control, black = intermediate, red = rapid progression.

CAP45.T/F, CAP200.T/F, CAP210.T/F and CAP239.T/F displayed similar rates of increase to peak p24 concentration of 8.4, 8.68, 9.6 and 8.61 ng p24/day respectively, while CAP63.T/F had the highest rate of 21.6 ng p24/day. The difference in RC, as determined by both rate to peak p24 and mean slope of expansion, between CAP63.T/F and each of the other four viruses was highly significant [one-way analysis of variance (ANOVA) test revealed the difference in $p<0.0001$], even after correcting for multiple testing. CAP45.T/F had the lowest replication capacity but this was not significantly different to viruses other than CAP63.

When considering the RC of t/f viruses in the context of disease progression, the woman with the most replication competent virus, CAP63, progressed to CD4+ counts <200 cells/µl in the first year of infection, after which she initiated therapy according to prevailing guidelines (National Department of Health South Africa, 2004).

4.6.2 Selection of CTL-escape mutations for introduction into cognate t/f IMCs

For participants CAP45, CAP63, CAP210 and CAP239, t/f IMCs were available at the time of initiation of this study. HLA-B has been identified as the strongest determinant of HIV disease progression (Fellay et al., 2007, Fellay et al., 2009, Pereyra et al., 2010, Bartha et al., 2013) and longitudinal sequence datasets from these women were analyzed for evidence of polymorphisms presenting in Class I HLA-B targeted epitopes during the first six months of infection (Chapter 2). Nine putative escape mutations were identified and seven of the nine mutations were selected for introduction into t/f IMCs (**Table 8A-D**) based on fixation or gain of dominance of the escape variant within the population by six months post-infection.

An escape mutation, D78A, in one epitope from participant CAP45 (Vif DWHLGHGVSI) was excluded since no sequence data was available between 12 and 65 weeks in order to evaluate the frequency of the mutant population at six months. An additional mutation, A86G, in the Nef KAAVDLSFF epitope from CAP239, was excluded based on conflicting reports regarding which of the two amino acids represented the consensus form, with some studies reporting Glycine (G) to be consensus (Novitsky et al., 2002, Kiepiela et al., 2004, Brumme et al., 2007), and others reporting G to be the escape amino acid and A to be the consensus (Matthews et al., 2008, Brumme et al., 2009, Carlson et al., 2012a). This may represent an instance of frequent transmission and fixation of the escape variant as a result of HLA adaption.

CTL-escape mutations identified in participants CAP45, CAP63, CAP210 and CAP239, were introduced by site-directed mutagenesis into their cognate t/f IMCs (pCAP45.c, pCAP63.c, pCAP210.c and pCAP239.c, where ‘p’ refers to plasmid and ‘c’ to subtype C) in order to evaluate the impact of these mutations on viral fitness.

Table 8A. Selected CTL escape mutations for introduction into pCAP45.c t/f IMC

Participant ID	ORF	Weeks post-infection	Epitope	# Sequences	Epitope name	Mutation selected*	Participant HLA association
CAP45	Rev	2	IHSISERIL	19/19	IL9	L60F	B*15:10
		5	....	6/6			
		12	....	1/3			
			.R.......	2/3			
		16	........F	1/1			
		52	....	1/7			
			........F	6/7			
		65	....	1/3			
			........F	2/3			
	Nef	2	EEVGFPVRPQV	3/3	EV11	E64G	B*45:01
		5	....	7/7			
		9	G..........	6/8			
			.D.........	1/8			
			.G.........	1/8			
		12	G..........	3/3			
		16	G..........	1/1			
		65	G..........	3/3			
		93	G..........	13/13			

Table 8B. Selected CTL escape mutations for introduction into pCAP63.c t/f IMC

Participant ID	ORF	Weeks post-infection	Epitope	# Sequences	Epitope name	Mutation selected*	Participant HLA association
CAP63	Nef	2	EEVGFPVRPQV	8/9	EV11	E65D	B*45:01
			K..........	1/9			
		4	....	1/19			
			V..........	1/19			
			.K.........	15/19			
			.D.........	1/19			
			.G.........	1/19			
		5	K..........	2/13			
			.G.........	6/13			
			.D.........	2/13			
			.A.........	1/13			
			.K.........	2/13			
		7	.G.........	9/10			
			.D.........	1/10			
		11	.G.........	4/7			
			.D.........	2/7			
			D..........	1/7			
		29	.G.........	2/10			
			.D.........	8/10			
		37	.D.........	5/5			

Table 8C. Selected CTL escape mutations for introduction into pCAP210.c t/f IMC

Participant ID	ORF	Weeks post-infection	Epitope	# Sequences	Epitope name	Mutation selected*	Participant HLA association
CAP210	Gag	2	VHQAISPRTL	10/10	VL10	A146T	B*15:10
		5	....	1/1			
		12	....	8/8			
		16	....	7/8			
			...T......	1/8			
		26	....	6/12			
			...T......	6/12			
		31	....	1/13			
			...T......	11/13			
			....L.....	1/13			
		42	....	2/12			
			...T......	8/12			
			...S......	1/12			
			...P......	1/12			
		54	...T......	1/8			
			...S......	7/8			
		80	...S......	8/12			
			...P......	4/12			
	Vif	2	DWHLGHGVSI	10/10	WI9	V85A	B*15:10
		5	....	1/1			
		12	....	7/7			
		16	....	1/8			
			.......A..	7/8			
		22	.......A..	8/8			
		26	.......A..	10/11			
			E......A..	1/11			
		54	.......A..	2/2			
		80	.E.....A..	2/6			
			..L.......	4/6			

Table 8D. Selected CTL escape mutations for introduction into pCAP239.c t/f IMC

Participant ID	ORF	Weeks post-infection	Epitope	# Sequences	Epitope name	Mutation selected*	Participant HLA association
CAP239	Gag	2	TSTLQEQVAW	22/23	TW10	T242N	B*58:01
			...I....T.	1/23			
		5	....	1/36			
			..N.....T.	35/36			
		6	..N.....T.	20/20			
		11	..N.......	2/3			
			..N.....T.	1/3			
		22	..N.......	18/18			
		49	..N.......	1/1			
		79	..N.......	10/10			
		117	..N.......	3/3			
	Pol	2	IVLPEKESW	2/2	IW9	E343K	B*58:01
		5	....	9/10			
			......K..	1/10			
		11	....	1/2			
			.....R...	1/2			
		22	....K....	9/9			
		49	....K....	1/1			
		117	....K....	3/3			

*Amino acid numbering of epitopes is according to HXB2.

4.6.3 The impact of early CTL selection on viral replicative capacity

Mutant viruses were produced and parallel replication kinetics assays performed following infection of PBMCs at equal m.o.i. as described in section 4.6.2. Replication kinetics of t/f and mutant viruses were compared. The fold-difference in RC values between t/f and mutant viruses as determined by the rate of expansion to peak p24 and mean slope respectively are listed in **Table 9**.

Table 9. RC comparisons between t/f and CTL escape mutant viruses

Virus 1	Virus 2	Average expansion rate fold difference*	Student's T-test p-value	Average mean slope fold difference*	Student's T-test p-value
CAP45.T/F	CAP45.L60F/E64G	1.3	0.0549	1.23	0.0653
CAP63.T/F	CAP63.E65D	0.8	0.0269	1.06	0.6146
CAP210.T/F	CAP210.A146T/V85A	1.8	0.0004	1.81	<0.0001
CAP239.T/F	CAP239.T242N/E343K	0.5	0.0072	0.5	0.0074

*Fold differences calculated as t/f divided by mutant virus.

The CAP45.T/F and CAP45.L60F/E64G, differing at a single amino acid site in Rev and Nef respectively, expanded similarly in culture from day 5 to 11, after which CAP45.T/F expanded towards a higher-level plateau (**Figure 33A**). The average rate of expansion of CAP45.T/F was 1.3-fold greater than that of CAP45.L60F/E64G and the mean slope was 1.2-fold greater; however the difference in RCs between the two viruses did not reach statistical significance (see Table 10).

For CAP63, the t/f (CAP63.T/F) and mutant (CAP63.E65D) viruses differed only by a single amino acid position in Nef. Both viruses expanded rapidly in culture, with the CAP63.T/F expanding slightly faster in the first nine days of infection and reaching a plateau thereafter, while CAP63.E65D continued to expand up to day 11 (**Figure 33B**). At day 5, the average p24 concentration of CAP63.T/F was 3.8-fold greater than that of CAP63.E65D indicating that this virus expanded more rapidly very early in the infection and possibly explains its earlier plateau. However, despite slower expansion early on, CAP63.E65D proceeded to exceed the peak p24 concentration of the t/f virus from day 9 onwards. The rate of expansion of CAP63.E65D was significantly higher (1.3-fold) than that of CAP63.T/F (p=0.0269) (Table 10), however the mean slopes of replication between the two viruses did not differ significantly (1.06-fold, p=0.6146).

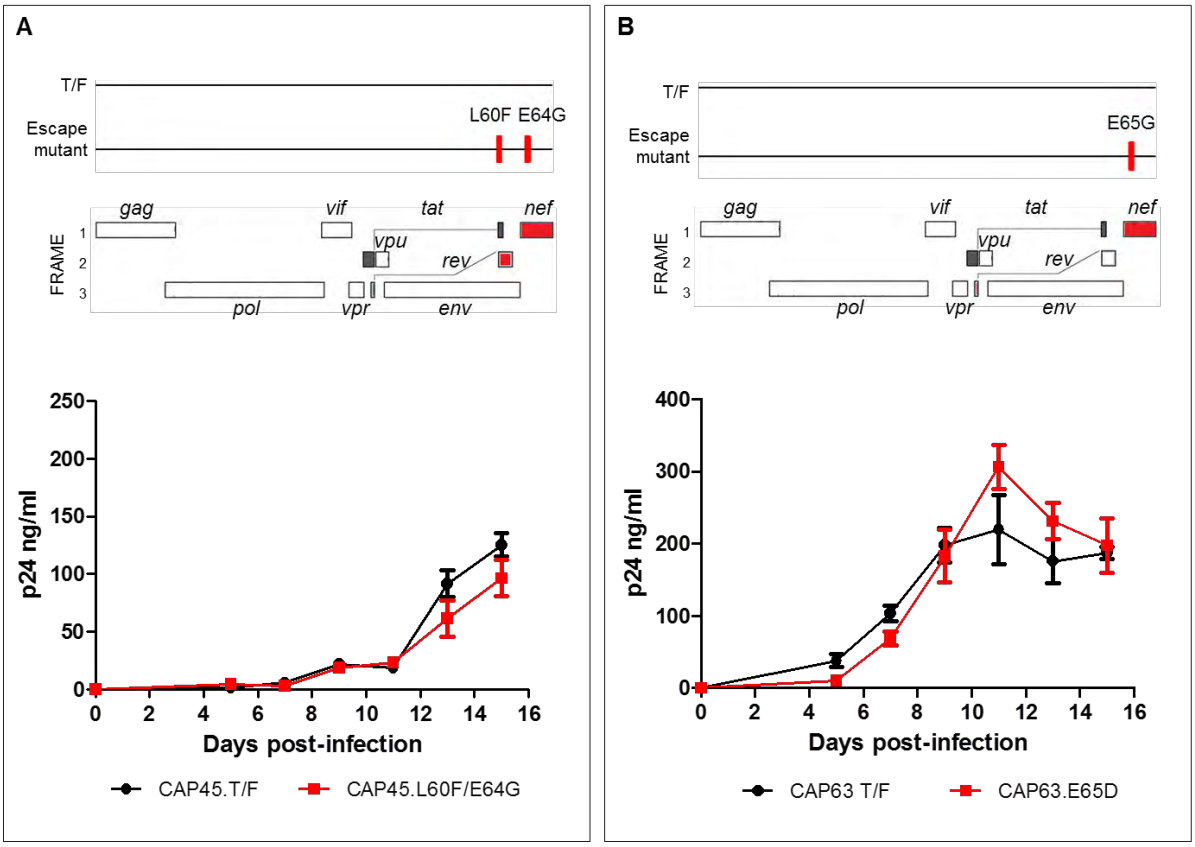


Figure 33. **Replication kinetics of t/f and CTL-escape mutants from CAP45 and CAP63.** Expansion of t/f (black) and CTL-escape mutant (red) viruses were compared in parallel PBMC infection assays. CD8-depleted, pooled PBMCs were infected at an m.o.i. of 0.001. Gag p24 antigen concentrations of viral supernatants sampled every two days from day 5 are plotted as a function of time post-infection for t/f and escape mutant pairs from CAP45 (**A**) and CAP63 (**B**). Error bars represent the standard deviation of p24 concentrations for triplicate infections.

In participant CAP210, the presence of B*15:10-associated escape mutations in each of Gag (A146T) and Vif (V85A) significantly reduced the RC of the t/f virus (1.8-fold reduction in RC as determined by both mean slope and rate of expansion, p=0.0004) (**Figure 34A**), and conversely in participant CAP239, early B*58:01-associated escape mutations T242N in Gag and E343K in Pol, significantly increased RC (**Figure 34B, Table 9**). In the case of participant CAP210, more rapid expansion of the CAP210.T/F was observed from day 7 onwards compared to CAP210.A146T/V85A.

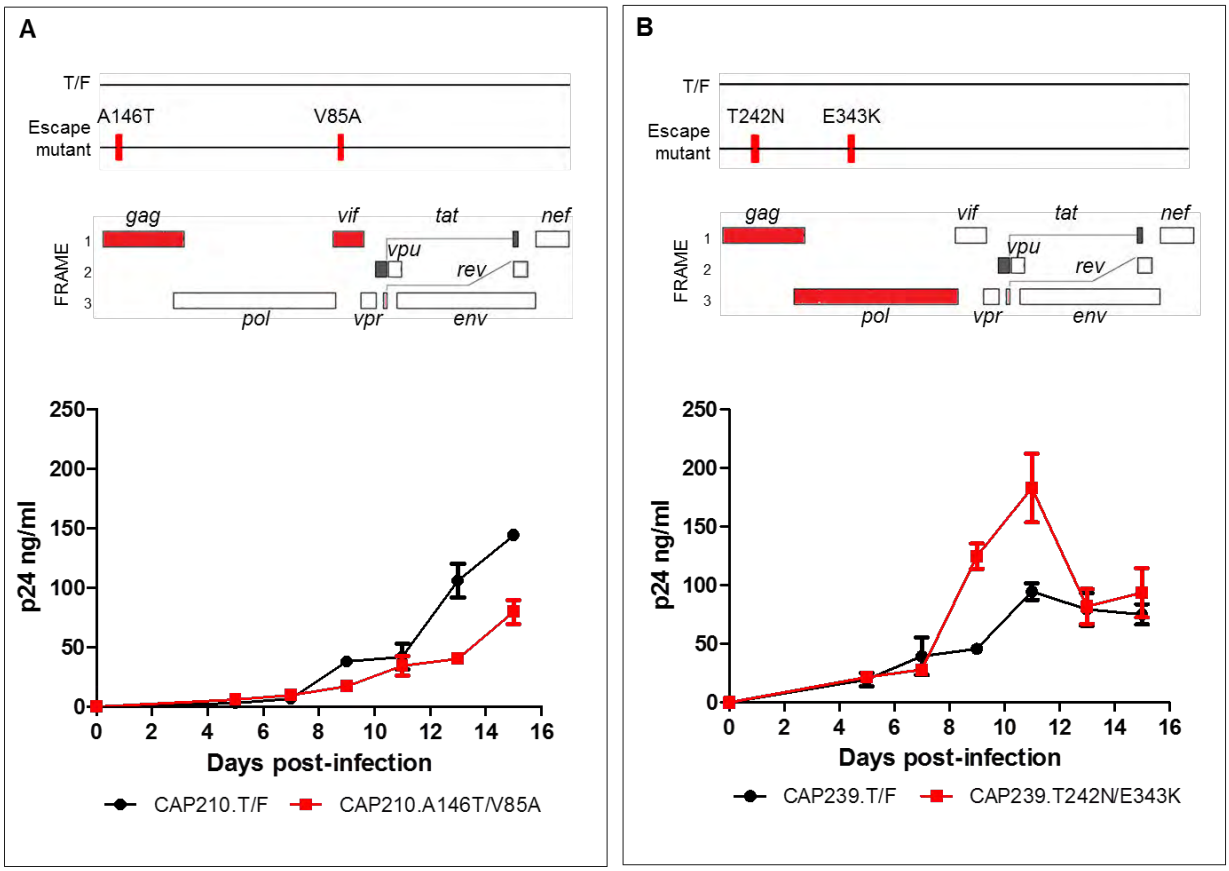


Figure 34. Replication kinetics of t/f and CTL-escape mutants from CAP210 and CAP239. Expansion of t/f (black) and CTL-escape mutant (red) viruses were compared in parallel PBMC infection assays. CD8-depleted, pooled PBMCs were infected at an m.o.i. of 0.001. Gag p24 antigen concentrations of viral supernatants sampled every two days from day 5 are plotted as a function of time post-infection for t/f and escape mutant pairs from CAP210 (**A**) and CAP239 (**B**). Error bars represent the standard deviation of p24 concentrations for triplicate infections.

Virus CAP239.T242N/E343K expanded almost twice as fast as CAP239.T/F from day 7 to day 11 after which a large decrease in p24 concentration to a similar level as that of CAP239.T/F was seen by day 13 and 15, at which point expansion of both viruses had plateaued. This was unexpected since the Gag T242N mutation is reported to significantly reduce RC (Martinez-Picardo et al., 2006; Brockman et al., 2007; Miura et al., 2009; Chopera et al., 2012; Song et al., 2012). The t/f virus sequence of CAP239 contained both the H219Q and I223V polymorphisms associated with compensation of T242N and associated with restoration of viral replicative fitness (Brockman et al., 2007). Although participant CAP239 is a B*58:01 positive individual herself, these polymorphisms in the derived t/f virus were present prior to acquisition of the T242N mutation, and are likely to have been in the virus transmitted from the donor.

These transmitted polymorphisms were back-mutated to evaluate their impact on CAP239.T242N/E343K replication. The Q219H and I223V mutations were introduced into the pCAP239.c escape mutant IMC by site-directed mutagenesis. The virus generated by transfection of this IMC, CAP239.T242N/E343K/Q219H/V223I, was used to infect the same pooled donor PBMCs as used for previous infections reported here, at the same m.o.i. of 0.001 in parallel with CAP239.T/F and CAP239.T242N/E343K. The replication kinetics of the three viruses was compared (**Figure 35A**).

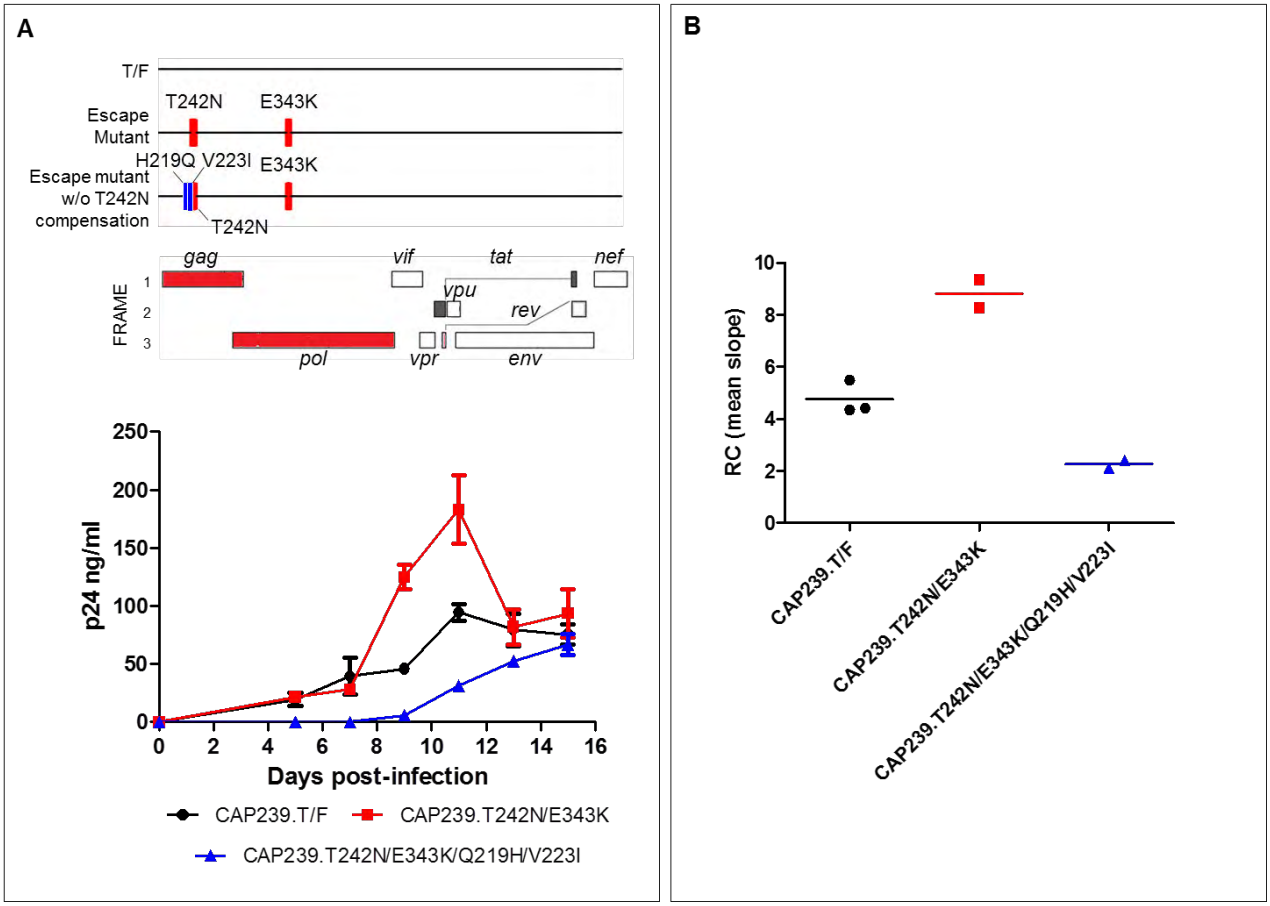


Figure 35. The impact of reversal of T242N compensatory mutations on RC of CAP239.T242N/E343K. Replication kinetics (**A**) and RC comparisons (**B**) of CAP239.T/F, CAP239.T242N/E343K and CAP239.T242N/E343K/Q219H/V223I (escape mutant with compensatory polymorphisms reversed) following parallel infections of CD8-depleted, pooled donor PBMCs at an m.o.i. of 0.001. Error bars represent the standard deviation of p24 concentrations for triplicate infections. Asterisks indicate the level of significance of differences between mean RC values obtained from unpaired parametric Student's t-tests (** = <0.01 but >0.001).

In the absence of the compensatory polymorphisms, the RC (mean slope) of the escape mutant was significantly reduced compared to both CAP239.T/F and CAP239.T242N/E343K/Q219H/V223I viruses (**Figure 35B**) [ANOVA p=0.001 with correction for multiple comparisons]. The expansion of CAP239.T242N/E343K/Q219H/V223I was slow

with the first detectable p24 antigen observed at day 9 post-infection. The mean slope RC and rate of expansion to peak p24 of this virus were 2.1- and 1.94-fold lower than that of CAP239.T/F respectively, confirming that the presence of the escape mutations did indeed confer a loss of fitness, and 3.9- and 3.7-fold lower than that of CAP239.T242N/E343K respectively indicating the extent of enhancement of replicative capacity resulting from compensation. These results show that polymorphisms associated with compensation of escape can carry a fitness cost, and that the introduction of the detrimental mutations can restore relative viral fitness.

4.7 Discussion

This study characterizes the replication capacity of full-length cognate t/f viruses in the context of differential clinical progression and reports the impact of early Class I HLA-mediated immune escape on replicative fitness of cognate subtype C viruses. We found that the participant with extremely rapid disease progression was infected with a virus which had significantly higher replication capacity compared to viruses from the other women, using a single-round, parallel PBMC infection assays with low viral multiplicity of infection (0.001). Furthermore we identify novel B*15:10-restricted escape mutations in p24 Gag and Vif which conferred a cost to replicative fitness. Counterintuitively, we found that introduction of the well-described T242N mutation as well as a Pol E343K mutation into the cognate t/f virus from one woman resulted in an increase in replicative fitness. This was due to pre-existing compensatory polymorphisms associated with T242N in the transmitted virus.

The influence of the replicative fitness of viruses overcoming the mucosal transmission bottleneck and establishing clinical infection on the rate of progression of disease is not well characterized. A recent review estimates up to 33% of variation in clinical disease to be a consequence of the transmitted virus (Fraser et al., 2014). We investigated viral fitness of t/f viruses from one viremic controller, two rapid progressors and two intermediate progressors. It was found that the replication capacity of the t/f virus from one rapid progressor, CAP63, was significantly greater than that of viruses from the other four women whose viruses had similar replication capacities. This woman represents a unique case since she progressed to CD4 count levels associated with AIDS-defining illness within approximately nine months of infection. It is possible that viral replication fitness contributed to this disease phenotype, however further work is needed to evaluate why this participant never controlled viral replication.

Many studies have evaluated the impact of immune escape mutations on viral replication in recombinant or chronic lab-adapted viral backbones (Brockman et al., 2010, Wright et al., 2010, Wright et al., 2011, Boutwell et al., 2013), however only two studies to date have investigated the impact of immune escape mutations in cognate t/f viral backbones, both of which were in subtype B (Song et al., 2012; Song et al., 2014). In two of the four women (CAP210 and CAP239), CTL-escape was found to negatively impact viral replication capacity. Since parallel replication assays are less sensitive to the detection of more subtle replicative fitness differences, the possibility exists that differences in replication capacities in viruses from the other two women (CAP45 and CAP63) were not detected as a result of the assay used.

In the case of CAP210 viral replication capacity was reduced after introduction of the B*15:10-associated Gag A146T and Vif V85A mutations. Mutations at position 146 have previously been associated with HLA-B*57/58, B*13, A*25 and B*15:10-restriction (Draenert et al., 2004; Honeyborne et al., 2007; Troyer et al., 2009; Prince et al., 2012). A146P was identified as a processing mutation flanking the B*57/58-restricted ISW9 epitope (Draenert et al., 2004), and has been shown to cause a loss of viral replication capacity (Troyer et al., 2009; Boutwell et al., 2013). Conversely, an A146S mutation has also been reported to enhance viral replication (Prince et al., 2012). In CAP210 we found that at 43 and 80 weeks, the A146T variant was replaced by A146P and A146S variants (see Table 9). Since a Threonine at position 146 is rare (present in <1% of subtype C database sequences), it is possible that variation at this site later in infection was due to the virus attempting to incorporate a more favourable amino acid while maintaining the escape phenotype. The V85A mutation in Vif is not known to be associated with a loss of fitness and is neither rare nor situated within a structural or functional domain of the genome. Based on the literature, it is likely that the Gag A146T mutation resulted in the loss of viral fitness observed, however a role for the V85A mutation needs to be tested in isolation to determine its individual contribution.

In the case of participant CAP239, replication capacity was significantly enhanced following introduction of B*58:01-associated Gag T242N and Pol E343K mutations into the t/f virus backbone. It is well established that the T242N mutation in the Gag TW10 B*57/58:01-targeted epitope reduces viral replication (Martinez-Picardo et al., 2006; Brockman et al., 2007; Miura et al., 2009; Chopera et al., 2012; Song et al., 2012). This has been attributed to interruption of viral capsid binding to host Cyclophilin A which is required for productive infection (Schaller et al., 2011); however compensatory mutations in the Cyclophilin A binding loop of the virus have been shown to restore replication fitness (Brockman et al., 2007). The presence of compensatory polymorphisms (H219Q and I223V) in the t/f virus

appeared to be detrimental to replication in the absence of T242N. When these viruses were back-mutated, the T242N-mutant replication was severely impacted. This likely indicates that the compensatory mutations are beneficial only in the context of escape and that independently they are not desirable for viral replication. These compensatory polymorphisms are accumulating in circulating subtype C viruses in southern Africa (Matthews et al., 2008) and the enhanced replication fitness associated with escape may explain why there is not a strong benefit associated with HLA-B*57. Recent studies have shown that compensatory mutations associated with T242N were found to be significantly more frequent in B*57/58:01 progressors compared to long-term non-progressors with these alleles (Gijsbers et al., 2013). In this same study, the H219Q mutation together with T242N and no additional compensatory mutations significantly enhanced replication capacity, however this was not the case for I223V (Gijsbers et al., 2013). Since the level of enhancement of replicative capacity is reported to increase as the numbers of compensatory mutations increase (Brockman et al., 2010, Gijsbers et al., 2013), the I223V mutation may have had an additive role in enhancing viral replication in the current study.

Ours is the first study to our knowledge to demonstrate that T242N escape can facilitate superior replication capacity to that of the wild-type virus counterpart. Interestingly, a study by Rolland et al. (2010) which identified co-varying amino acid sites in subtype C Gag from individuals in Durban, South Africa found that individuals with deleterious escape and compensatory mutation pairs at co-varying sites had higher viral loads than individuals with consensus mutations at these sites, thus supporting the potential enhancement of viral fitness to above 'wild-type' levels in the presence of compensation. It is possible however that in our study, the E343K mutation also contributed to enhanced viral replication capacity.

Furthermore, in a recent study, Song et al. (2012) were unable to detect a fitness difference between a subtype B cognate t/f and T242N mutant virus in a parallel infection assay using primary CD4+ T-cells. The authors proposed that more subtle fitness differences may only be detected through competition assays with multiple passaging of viruses (Song et al., 2012). In this thesis, a significant difference between the CAP239.T/F virus and CAP239.T242N/E343K mutant in the absence of compensation was detected in a parallel assay. However, the parallel assay used by Song et al. (2012) was carried out over a 5-day period compared to the 15-day assay used in this thesis which only detected p24 from day five post-infection onwards; therefore observed differences may be due to differences in assay conditions. In addition, the CAP239 escape mutant virus also contained a Pol E403K mutation in a B*58:01-targeted epitope. This mutation may have contributed to the reduced replication fitness observed.

An early escape mutation in the Nef EV11 epitope identified in viruses from both rapid progressor CAP63 and viremic controller CAP45 was introduced into the t/f virus IMC backbones for each of the two women. Both mutations were located within the Nef acidic domain which overlaps with the EV11 epitope and consists of four Glutamic Acid residues (^{62}EEEE^{65}). This domain was reported to be required for MHC Class I downregulation (Piguet et al., 2000), but was able to accommodate replacement of any two of the Glutamic acids at one time without a reduction in this viral function (Baugh et al., 2008). The E65D mutation conferred a modestly enhanced replication capacity to the t/f virus from participant CAP63, whereas in the case of participant CAP45 expansion of the t/f and mutant virus (containing the Nef E64G mutation as well as a Rev I60F mutation) was similar. It may be speculated that the type of amino acid transition influenced the impact on fitness. In the case of the E65D mutation in CAP63, the negatively charged Glutamic Acid was replaced with a similar residue, negatively charged Aspartic Acid, and in the case of the E64G mutation, the Glutamic Acid was replaced with a dissimilar residue, unique amino acid Glycine. The presence of the Aspartic Acid may be advantageous to viral functions involving the acidic domain. Interestingly, in both participants CAP45 and CAP63, the virus toggles between Glycine and Aspartic Acid at position 65, and in the case of CAP45 reverts to consensus, but in the case of CAP63, selects the Aspartic Acid (see Table 9).

While this study demonstrates the advantage of evaluating escape in cognate backbones, the generation of IMCs takes significant effort and resources, making this a difficult approach for evaluation of larger numbers of samples. The method of IMC construction (modified from that described by Salazar-Gonzalez et al. (2009) and Ochsenbauer et al. (2012)) described here (Appendix 5) holds advantages since it made only limited use of PCR and thereby reduced the risk of introduction of polymerase-induced error into DNA sequences. Furthermore, the use of cloning by means of overlapping fragments between the insert and vector, as used in the case of In-Fusion technology, allowed for more stable integration of genome inserts compared to sticky-end ligation of restriction enzyme digested insert and vector. One potential drawback of the method of IMC construction described is the high cost of synthesis of DNA fragments.

In conclusion, the significant impact of early escape resulting from Class I HLA-mediated targeting of CTL epitopes in Gag on viral replication capacity was illustrated in full-length cognate t/f viruses. The utility of cognate t/f backbones was made apparent by the demonstration of the enhancement of replicative capacity through transmitted compensatory mutations.

Chapter 5: General Discussion and Conclusion

5.1 Discussion

This thesis represents a detailed study of full-length subtype C HIV-1 genomes of the viruses establishing clinical infection in five South African women, their subsequent early evolution and the impact of early immune selection on viral replicative fitness. The study contributes to our understanding of the globally dominant subtype C strain and is one of the first to demonstrate the impact of early immune selection on subtype C viral phenotype in the context of both cognate transmitted/founder (t/f) virus backbones and of differential disease progression.

In the last six years, HIV-1 transmission studies have been accelerated with the finding that the majority of infections are established by a single viral variant (Keele et al., 2008) and that the sequence of that t/f variant can be derived and used as a basis for the mapping of early viral evolution (Salazar-Gonzalez et al., 2009). Characterization of the transmission bottleneck in subtype C infection (Abrahams et al., 2009) formed the foundation for selection of the women investigated in this thesis, and the complete viral genome analysis performed confirmed single/low-diversity transmission originally determined by *env*-only analyses.

The subsequent mapping of early selection on the genomes of these five women, revealed findings in agreement with those reported for similar near full-length genome subtype B studies (Goonetilleke et al., 2009; Herbeck et al., 2011). As with the previous studies, early selection was dominated by immune pressure mediated by CD8+ cytotoxic T-lymphocytes (CTLs) but was also contributed to by early antibody pressure and reversion, and immune selection and viral evolution was most extensive in *nef*. The high level of amino acid shuffling and toggling in HLA-targeted evolving epitopes identified in this study implies that early targets of CTL pressure are unstable and highly variable.

Long-term disease outcome may largely be dictated by the type of host immune responses mounted and the mutations selected which may alter the virus's replicative ability. The field is moving towards the use of cognate viruses as more relevant research tools for evaluating viral phenotype with two subtype B studies characterizing the effect of immune escape and compensatory mutations in cognate viral backbones published to date (Song et al., 2012; Song et al., 2014). The unique advantage of the use of cognate viruses for evaluating viral replicative capacity and the effect of CTL-escape mutations accrued very early in infection is demonstrated in this thesis. Most notably, in one individual, enhancement of viral replication to a level exceeding that of the t/f virus through a combination of immune escape and transmitted compensatory polymorphisms was demonstrated, representing a novel finding in

the field. The individual harboured the protective HLA-B*58:01 allele, but was not controlling disease.

One objective of this thesis was to compare the evolution and phenotype of viruses transmitted to individuals who go on to control disease compared to those with rapid progression. The numbers in this study were too limited to make any general statements regarding t/f viral evolution and disease progression, however one rapid progressor in this study was found to be unique as her virus displayed significantly higher viral diversification and evolution as well as a significantly higher replication capacity compared to the other women investigated. The only CTL-escape mutation identified in an HLA-B-restricted epitope in this individual did not have a significant impact on viral replication capacity.

A novel escape mutation, A146T, associated with B*15:10-restriction of the Gag VL10 epitope was identified and potentially associated with a reduction of viral replication capacity (in this instance in combination with a V85A escape mutation in Vif). This epitope and an overlapping epitope, ISW9, are targeted by a number of different HLA alleles including the protective HLA-B*57 and B*13 alleles but also by non-protective HLA-A*25 alleles. The individual in whom this mutation was identified was a rapid progressor and thus the attenuating mutation was not associated with a clinical benefit. This may be due to the fact that the restricting B*15:10 allele is non-protective, and this outweighed the benefit of attenuation of viral replication.

5.2 Conclusion

Early immune selection in t/f subtype C viruses is dominated by CD8+ T-lymphocyte responses and characterized by rapid escape and complex mutational pathways. Early escape from these responses can significantly alter viral replication fitness when introduced into cognate t/f viral backbones. The importance of evaluating immune escape mutations in the context of the cognate virus is emphasized through the demonstration of the fitness enhancing effects of compensatory mutations. Although we show that viral attenuation through immune escape is not restricted to protective HLA alleles, this attenuation was not sufficient to provide overall clinical benefit reaffirming that host genetics plays a dominant role in determining disease outcome. However, the impact of escape in Gag on replicative fitness supports a role for Gag epitopes in vaccine immunogen selection since, in the absence of vaccine protection, immune escape mutations in Gag could provide clinical benefit. Overall this thesis supports the need for eliciting CTL responses by a vaccine due to

their dominant role in early infection; however the extensive variability in early evolving epitopes would represent a challenge to viral control through vaccination.

References

Abrahams, M-R, J A Anderson, E E Giorgi, C Seoighe, K Mlisana, L-H Ping, G S Athreya, et al. 2009. “Quantitating the Multiplicity of Infection with Human Immunodeficiency Virus Type 1 Subtype C Reveals a Non-Poisson Distribution of Transmitted Variants.” *Journal of Virology* 83 (8): 3556–67.

Adachi, A, H E Gendelman, S Koenig, T Folks, R Willey, A Rabson, and M A Martin. 1986. “Production of Acquired Immunodeficiency Syndrome-Associated Retrovirus in Human and Nonhuman Cells Transfected with an Infectious Molecular Clone.” *Journal of Virology* 59 (2): 284–91.

Addo, M M, X G Yu, A Rathod, D Cohen, R L Eldridge, D Strick, M N Johnston, et al. 2003. “Comprehensive Epitope Analysis of Human Immunodeficiency Virus Type 1 (HIV-1)-Specific T-Cell Responses Directed against the Entire Expressed HIV-1 Genome Demonstrate Broadly Directed Responses , but No Correlation to Viral Load” *Journal of Virology* 77 (3): 2081–92.

Alizon, Samuel, Viktor von Wyl, Tanja Stadler, Roger D Kouyos, Sabine Yerly, Bernard Hirschel, Jürg Böni, et al. 2010. “Phylogenetic Approach Reveals That Virus Genotype Largely Determines HIV Set-Point Viral Load.” *PLoS Pathogens* 6 (9): e1001123.

Allen, Todd M, Marcus Altfeld, Shaun C Geer, Elizabeth T Kalife, Corey Moore, Kristin M O Sullivan, Ivna Desouza, et al. 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.”*Journal of Virology* 79 (21): 13239–49.

Allen, Todd M, Marcus Altfeld, Xu G Yu, Kristin M O Sullivan, Mathias Lichterfeld, Sylvie Le Gall, Mina John, et al. 2004. “Selection , Transmission , and Reversion of an Antigen-Processing Cytotoxic T-Lymphocyte Escape Mutation in Human Immunodeficiency Virus Type 1 Infection” *Journal of Virology* 78 (13): 7069–78.

Almeida, Sabrina EM, Rubia M de Medeiros, Dennis M Junqueira, Tiago Gräf, Caroline P B Passaes, Gonzalo Bello, Mariza G Morgado, and Monick L Guimarães. 2012. “Temporal Dynamics of HIV-1 Circulating Subtypes in Distinct Exposure Categories in Southern Brazil.” *Virology Journal* 9: 306.

Alter, Galit, David Heckerman, Arne Schneidewind, Lena Fadda, Carl M Kadie, Jonathan M Carlson, Cesar Oniangue-Ndza, et al. 2011. “HIV-1 Adaptation to NK-Cell-Mediated Immune Pressure.” *Nature* 476 (7358). Nature Publishing Group: 96–100.

Ariën, Kevin K, Guido Vanham, and Eric J Arts. 2007. “Is HIV-1 Evolving to a Less Virulent Form in Humans?” *Nature Reviews Microbiology* 5 (2): 141–51.

Arnott, Alicia, Darren Jardine, Kim Wilson, Paul R Gorry, Kate Merlin, Patricia Grey, Matthew G Law, et al. 2010. “High Viral Fitness during Acute HIV-1 Infection.” *PloS One* 5 (9): e12631.

Asmal, Mohammed, Ina Hellmann, Weimin Liu, Brandon F Keele, Alan S Perelson, Tanmoy Bhattacharya, S Gnanakaran, et al. 2011. “A Signature in HIV-1 Envelope Leader Peptide Associated with Transition from Acute to Chronic Infection Impacts Envelope Processing and Infectivity.” *PloS One* 6 (8): e23673.

Asquith, Becca. 2008. “The Evolutionary Selective Advantage of HIV-1 Escape Variants and the Contribution of Escape to the HLA-Associated Risk of AIDS Progression.” *PloS One* 3 (10): e3486.

Bansal, Anju, Ethan Gough, Steffanie Sabbaj, Doug Ritter, Karina Yusim, Greg Sfakianos, Grace Aldrovandi, et al. 2005. “CD8 T-Cell Responses in Early HIV-1 Infection Are Skewed towards High Entropy Peptides.” *AIDS* 19 (3): 241–50.

Bar, Katharine J, Hui Li, Annie Chamberland, Cecile Tremblay, Jean Pierre Routy, Truman Grayson, Chuanxi Sun, et al. 2010. “Wide Variation in the Multiplicity of HIV-1 Infection among Injection Drug Users.” *Journal of Virology* 84 (12): 6241–47.

Bar, Katharine J, Chun-yen Tsao, Shilpa S Iyer, Julie M Decker, Yongping Yang, Mattia Bonsignori, Xi Chen, et al. 2012. “Early Low-Titer Neutralizing Antibodies Impede HIV-1 Replication and Select for Virus Escape.” *PLoS Pathogens* 8 (5): e1002721.

Barouch, Dan H, Kara L O’Brien, Nathaniel L Simmons, Sharon L King, Peter Abbink, Lori F Maxfield, Ying-Hua Sun, et al. 2010. “Mosaic HIV-1 Vaccines Expand the Breadth and Depth of Cellular Immune Responses in Rhesus Monkeys.” *Nature Medicine* 16 (3): 319–23.

Barouch, Dan H, James B Whitney, Brian Moldt, Florian Klein, Thiago Y Oliveira, Jinyan Liu, Kathryn E Stephenson, et al. 2013. “Therapeutic Efficacy of Potent Neutralizing HIV-1-Specific Monoclonal Antibodies in SHIV-Infected Rhesus Monkeys.” *Nature* 503 (7475): 224–28.

Bartha, István, Jonathan M Carlson, Chanson J Brumme, Paul J McLaren, Zabrina L Brumme, Mina John, David W Haas, et al. 2013. “A Genome-to-Genome Analysis of Associations between Human Genetic Variation, HIV-1 Sequence Diversity, and Viral Control.” *eLife* 2: e01123.

Baugh, Laura L, J Victor Garcia, and John L Foster. 2008. “Functional Characterization of the Human Immunodeficiency Virus Type 1 Nef Acidic Domain.” *Journal of Virology* 82 (19): 9657–67.

Bernardin, Flavien, Denice Kong, Lorraine Peddada, Lee Ann Baxter-lowe, and Eric Delwart. 2005. “Human Immunodeficiency Virus Mutations during the First Month of Infection Are Preferentially Found in Known Cytotoxic T-Lymphocyte Epitopes” 79 (17): 11523–28.

Betts, Michael R, Martha C Nason, Sadie M West, Stephen C De Rosa, Stephen A Migueles, Jonathan Abraham, Michael M Lederman, et al. 2006. “HIV Nonprogressors Preferentially Maintain Highly Functional HIV-Specific CD8+ T Cells.” *Blood* 107 (12): 4781–89.

Blaak, H, M Brouwer, L J Ran, F de Wolf, and H Schuitemaker. 1998. “In Vitro Replication Kinetics of Human Immunodeficiency Virus Type 1 (HIV-1) Variants in Relation to Virus Load in Long-Term Survivors of HIV-1 Infection.” *The Journal of Infectious Diseases* 177 (3): 600–610.

Boeras, Debrah I, Peter T Hraber, Mackenzie Hurlston, Tammy Evans-Strickfaden, Tanmoy Bhattacharya, Elena E Giorgi, Joseph Mulenga, et al. 2011. “Role of Donor Genital Tract HIV-1 Diversity in the Transmission Bottleneck.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (46): E1156–63.

Bonsignori, Mattia, Justin Pollara, M Anthony Moody, Michael D Alpert, Xi Chen, Kwan-Ki Hwang, Peter B Gilbert, et al. 2012. “Antibody-Dependent Cellular Cytotoxicity-Mediating Antibodies from an HIV-1 Vaccine Efficacy Trial Target Multiple Epitopes and Preferentially Use the VH1 Gene Family.” *Journal of Virology* 86 (21): 11521–32.

Bor, Jacob, Abraham J Herbst, Marie-Louise Newell, and Till Bärnighausen. 2013. “Increases in Adult Life Expectancy in Rural South Africa: Valuing the Scale-up of HIV Treatment.” *Science* 339 (6122): 961–65.

Borrow, P, H Lewicki, B H Hahn, G M Shaw, and M B Oldstone. 1994. “Virus-Specific CD8+ Cytotoxic T-Lymphocyte Activity Associated with Control of Viremia in Primary Human Immunodeficiency Virus Type 1 Infection.” *Journal of Virology* 68 (9): 6103–10.

Borrow, Persephone, Hanna Lewicki, Xiping Wei, Marc S Horwitz, Nancy Peffer, Heather Meyers, Jay A Nelson, et al. 1997. “Antiviral Pressure Exerted by HIV-1-Specific Cytotoxic T Lymphocytes (CTLs) during Primary Infection Demonstrated by Rapid Selection of CTL Escape Virus.” *Nature* 3 (2): 205–11.

Borthwick, Nicola, Tina Ahmed, Beatrice Ondondo, Peter Hayes, Annie Rose, Umar Ebrahimsa, Emma-Jo Hayton, et al. 2014. “Vaccine-Elicited Human T Cells Recognizing Conserved Protein Regions Inhibit HIV-1.” *Molecular Therapy* 22 (2): 464–75.

Boutwell, Christian L, Jonathan M Carlson, Tien-Ho Lin, Aaron Seese, Karen A Power, Jian Peng, Yanhua Tang, et al. 2013. “Frequent and Variable Cytotoxic-T-Lymphocyte Escape-Associated Fitness Costs in the Human Immunodeficiency Virus Type 1 Subtype B Gag Proteins.” *Journal of Virology* 87 (7): 3952–65.

Brennan, Catherine A, F Javier Ibarrondo, Catherine A Sugar, Mary Ann Hausner, Roger Shih, Hwee L Ng, Roger Detels, et al. 2012. “Early HLA-B*57-Restricted

CD8+ T Lymphocyte Responses Predict HIV-1 Disease Progression." *Journal of Virology* 86 (19): 10505–16.

Brockman, Mark A, Zabrina L Brumme, Chanson J Brumme, Toshiyuki Miura, Jennifer Sela, Pamela C Rosato, Carl M Kadie, et al. 2010. "Early Selection in Gag by Protective HLA Alleles Contributes to Reduced HIV-1 Replication Capacity That May Be Largely Compensated for in Chronic Infection." *Journal of Virology* 84 (22): 11937–49.

Brockman, Mark A, Arne Schneidewind, Matthew Lahaie, Aaron Schmidt, Toshiyuki Miura, Ivna Desouza, Faina Ryvkin, et al. 2007. "Escape and Compensation from Early HLA-B57-Mediated Cytotoxic T-Lymphocyte Pressure on Human Immunodeficiency Virus Type 1 Gag Alter Capsid Interactions with Cyclophilin A." *Journal of Virology* 81 (22): 12608–18.

Bronke, Corine, Coral-Ann M Almeida, Elizabeth McKinnon, Steven G Roberts, Niamh M Keane, Abha Chopra, Jonathan M Carlson, David Heckerman, Simon Mallal, and Mina John. 2013. "HIV Escape Mutations Occur Preferentially at HLA-Binding Sites of CD8 T-Cell Epitopes." *AIDS* 27 (6): 899–905.

Brumme, Zabrina L, Chanson J Brumme, Jonathan Carlson, Hendrik Streeck, Mina John, Quentin Eichbaum, Brian L Block, et al. 2008a. "Marked Epitope- and Allele-Specific Differences in Rates of Mutation in Human Immunodeficiency Type 1 (HIV-1) Gag, Pol, and Nef Cytotoxic T-Lymphocyte Epitopes in Acute/early HIV-1 Infection." *Journal of Virology* 82 (18): 9216–27.

Brumme, Zabrina L, Chanson J Brumme, David Heckerman, Bette T Korber, Marcus Daniels, Jonathan Carlson, Carl Kadie, et al. 2007. "Evidence of Differential HLA Class I-Mediated Viral Evolution in Functional and Accessory/regulatory Genes of HIV-1." *PLoS Pathogens* 3 (7): e94.

Brumme, Zabrina L, Mina John, Jonathan M Carlson, Chanson J Brumme, Dennison Chan, Mark A Brockman, Luke C Swenson, et al. 2009. "HLA-Associated Immune Escape Pathways in HIV-1 Subtype B Gag, Pol and Nef Proteins." *PloS One* 4 (8): e6687.

Brumme, Zabrina L, Iris Tao, Sharon Szeto, Chanson J Brumme, Jonathan M Carlson, Dennison Chan, Carl Kadie, et al. 2008b. "Human Leukocyte Antigen-Specific Polymorphisms in HIV-1 Gag and Their Association with Viral Load in Chronic Untreated Infection." *AIDS* 22 (11): 1277–86.

Campbell, Thomas B, Kristina Schneider, Terri Wrin, Christos J Petropoulos, and Elizabeth Connick. 2003. "Relationship between In Vitro Human Immunodeficiency Virus Type 1 Replication Rate and Virus Load in Plasma." *Journal of Virology* 77 (22): 12105–12.

Carlson, J. M., M. Schaefer, D. C. Monaco, R. Batorsky, D. T. Claiborne, J. Prince, M. J. Deymier, et al. 2014. "Selection Bias at the Heterosexual HIV-1 Transmission Bottleneck." *Science* 345 (6193): 1254031–1254031.

Carlson, Jonathan M, Chanson J Brumme, Eric Martin, Jennifer Listgarten, Mark A Brockman, Anh Q Le, Celia K S Chui, et al. 2012a. “Correlates of Protective Cellular Immunity Revealed by Analysis of Population-Level Immune Escape Pathways in HIV-1.” *Journal of Virology* 86 (24): 13202–16.

Carlson, Jonathan M, and Zabrina L Brumme. 2008. “HIV Evolution in Response to HLA-Restricted CTL Selection Pressures: A Population-Based Perspective.” *Microbes Infect* 10 (5): 455–61.

Carlson, Jonathan M, Jennifer Listgarten, Nico Pfeifer, Vincent Tan, Carl Kadie, Bruce D Walker, Thumbi Ndung’u, et al. 2012b. “Widespread Impact of HLA Restriction on Immune Control and Escape Pathways of HIV-1.” *Journal of Virology* 86 (9): 5230–43.

Carrington, M. 1999. “HLA and HIV-1: Heterozygote Advantage and B*35-Cw*04 Disadvantage.” *Science* 283 (5408): 1748–52.

Chikata, Takayuki, Jonathan M Carlson, Yoshiko Tamura, Mohamed Ali Borghan, Takuya Naruto, Masao Hashimoto, Hayato Murakoshi, et al. 2014. “Host-Specific Adaptation of HIV-1 Subtype B in the Japanese Population.” *Journal of Virology* 88 (9): 4764–75.

Chohan, Bhavna, Dorothy Lang, Manish Sagar, Ludo Lavreys, Barbra Richardson, Bette Korber, and Julie Overbaugh. 2005. “Selection for Human Immunodeficiency Virus Type 1 Envelope Glycosylation Variants with Shorter V1-V2 Loop Sequences Occurs during Transmission of Certain Genetic Subtypes and May Impact Viral RNA Levels.” *Journal of Virology* 79(10):6528-31

Chopera, D R, M Mlotshwa, Z Woodman, K Mlisana, D de Assis Rosa, D P Martin, S Abdool Karim, C M Gray, and C Williamson. 2011. “Virological and Immunological Factors Associated with HIV-1 Differential Disease Progression in HLA-B 58:01-Positive Individuals.”*Journal of Virology* 85 (14): 7070–80.

Chopera, Denis R, Zenda Woodman, Koleka Mlisana, Mandla Mlotshwa, Darren P Martin, Cathal Seoighe, Florette Treurnicht, et al. 2008. “Transmission of HIV-1 CTL Escape Variants Provides HLA-Mismatched Recipients with a Survival Advantage.” *PLoS Pathogens* 4 (3): e1000033.

Chopera, Denis R, Laura A Cotton, Alexander Zawaira, Jaclyn K. Mann, Nobubelo K. Ngandu, Roman Ntale, Jonathan M. Carlson et al. 2012. "Intersubtype Differences in the Effect of a Rare p24 Gag Mutation on HIV-1 Replicative Fitness." *Journal of Virology* 86(24):13423-33

Cohen, Myron S, George M Shaw, Andrew J McMichael, and Barton F Haynes. 2011. “Acute HIV-1 Infection.” *New England Journal of Medicine* 364: 1943–54.

Collins, Jennifer A, M Gregory Thompson, Elijah Paintsil, Melisa Ricketts, Joanna Gedzior, and Louis Alexander. 2004. “Competitive Fitness of Nevirapine-

Resistant Human Immunodeficiency Virus Type 1 Mutants." *Journal of Virology* 78 (2): 603–11.

Cotton, Laura A, Xiaomei T Kuang, Anh Q Le, Jonathan M Carlson, Benjamin Chan, Denis R Chopera, Chanson J Brumme, et al. 2014. "Genotypic and Functional Impact of HIV-1 Adaptation to Its Host Population during the North American Epidemic." *PLoS Genetics* 10 (4): e1004295.

Crawford, Hayley, Wendy Lumm, Alasdair Leslie, Malinda Schaefer, Debrah Boeras, Julia G Prado, Jianming Tang, et al. 2009. "Evolution of HLA-B*5703 HIV-1 Escape Mutations in HLA-B*5703-Positive Individuals and Their Transmission Recipients." *The Journal of Experimental Medicine* 206 (4): 909–21.

Crawford, Hayley, Julia G Prado, Alasdair Leslie, Stéphane Hué, Isobella Honeyborne, Sharon Reddy, Mary van der Stok, et al. 2007. "Compensatory Mutation Partially Restores Fitness and Delays Reversion of Escape Mutation within the Immunodominant HLA-B*5703-Restricted Gag Epitope in Chronic Human Immunodeficiency Virus Type 1 Infection." *Journal of Virology* 81 (15): 8346–51.

Currier, Jeffrey R, William E Dowling, K Monique Wasunna, Uzma Alam, Carl J Mason, Merlin L Robb, Jean K Carr, Francine E McCutchan, Deborah L Birx, and Josephine H Cox. 2003. "Detection of High Frequencies of HIV-1 Cross-Subtype Reactive CD8 T Lymphocytes in the Peripheral Blood of HIV-1-Infected Kenyans." *AIDS* 17 (15): 2149–57.

Davenport, Miles P, Liyen Loh, Janka Petravic, and Stephen J Kent. 2008. "Rates of HIV Immune Escape and Reversion: Implications for Vaccination." *Trends in Microbiology* 16 (12): 561–66.

De Oliveira, Tulio, Marco Salemi, Michelle Gordon, Anne-Mieke Vandamme, Estrelita Janse van Rensburg, Susan Engelbrecht, Hoosen M Coovadia, and Sharon Cassol. 2004. "Mapping Sites of Positive Selection and Amino Acid Diversification in the HIV Genome: An Alternative Approach to Vaccine Design?" *Genetics* 167 (3): 1047–58.

De Souza, Mark S, Silvia Ratto-Kim, Weerawan Chuenarom, Alexandra Schuetz, Somsak Chantakulkij, Bessara Nuntapinit, Anais Valencia-Micolta, et al. 2012. "The Thai Phase III Trial (RV144) Vaccine Regimen Induces T Cell Responses That Preferentially Target Epitopes within the V2 Region of HIV-1 Envelope." *Journal of Immunology* 188 (10): 5166–76.

Derdeyn, Cynthia A, Julie M Decker, Frederic Bibollet-Ruche, John L Mokili, Mark Muldoon, Scott A Denham, Marintha L Heil, et al. 2004. "Envelope-Constrained Neutralization-Sensitive HIV-1 after Heterosexual Transmission." *Science* 303 (5666): 2019–22.

Domingo, E, L Menéndez-Arias, and JJ Holland. 1997. "RNA Virus Fitness." *Reviews in Medical Virology* 7 (2): 87–96.

Doria-Rose, Nicole A, Chaim A Schramm, Jason Gorman, Penny L Moore, Jinal N Bhiman, Brandon J DeKosky, Michael J Ernandes, et al. 2014. “Developmental Pathway for Potent V1V2-Directed HIV-Neutralizing Antibodies.” *Nature* 509 (7498): 55–62.

Dorrell, Lucy, Tao Dong, Graham S Ogg, Simon Lister, Steve McAdam, Tim Rostron, Chris Conlon, Andrew J McMichael, and Sarah L Rowland-Jones. 1999. “Distinct Recognition of Non-Clade B Human Immunodeficiency Virus Type 1 Epitopes by Cytotoxic T Lymphocytes Generated from Donors Infected in Africa.” *Journal of Virology* 73 (2): 1708–14.

Draenert, Rika, Sylvie Le Gall, Katja J Pfafferott, Alasdair J Leslie, Polan Chetty, Christian Brander, Edward C Holmes, et al. 2004. “Immune Selection for Altered Antigen Processing Leads to Cytotoxic T Lymphocyte Escape in Chronic HIV-1 Infection.” *The Journal of Experimental Medicine* 199 (7): 905–15.

Drummond, Alexei J, Simon Y W Ho, Matthew J Phillips, and Andrew Rambaut. 2006. “Relaxed Phylogenetics and Dating with Confidence.” *PLoS Biology* 4 (5): e88.

Drummond, Alexei J, and Andrew Rambaut. 2007. “BEAST: Bayesian Evolutionary Analysis by Sampling Trees.” *BMC Evolutionary Biology* 7: 214.

Dykes, Carrie, and Lisa M Demeter. 2007. “Clinical Significance of Human Immunodeficiency Virus Type 1 Replication Fitness.” *Clinical Microbiology Reviews* 20 (4): 550–78.

Edmonson, P F, and J I Mullins. 1992. “Efficient Amplification of HIV Half-Genomes from Tissue DNA.” *Nucleic Acids Research* 20 (18): 4933.

Esparza, José. 2013. “A Brief History of the Global Effort to Develop a Preventive HIV Vaccine.” *Vaccine* 31 (35): 3502–18.

Feeney, M E, Y Tang, K A Roosevelt, A J Leslie, K Mcintosh, N Karthas, B D Walker, and P J R Goulder. 2004. “Immune Escape Precedes Breakthrough Human Immunodeficiency Virus Type 1 Viremia and Broadening of the Cytotoxic T-Lymphocyte Response in an HLA-B27-Positive Long-Term-Nonprogressing Child.” *Journal of Virology* 78 (16): 8927–30.

Fellay, Jacques, Dongliang Ge, Kevin V Shianna, Sara Colombo, Bruno Ledergerber, Elizabeth T Cirulli, Thomas J Urban, et al. 2009. “Common Genetic Variation and the Control of HIV-1 in Humans.” *PLoS Genetics* 5 (12): e1000791.

Fellay, Jacques, Kevin V Shianna, Dongliang Ge, Sara Colombo, Bruno Ledergerber, Mike Weale, Kunlin Zhang, et al. 2007. “A Whole-Genome Association Study of Major Determinants for Host Control of HIV-1.” *Science* 317 (5840): 944–47.

Ferrari, Guido, Bette Korber, Nilu Goonetilleke, Michael K P Liu, Emma L Turnbull, Jesus F Salazar-Gonzalez, Natalie Hawkins, et al. 2011. “Relationship between Functional Profile of HIV-1 Specific CD8 T Cells and Epitope Variability with the Selection of Escape Mutants in Acute HIV-1 Infection.” *PLoS Pathogens* 7 (2): e1001273.

Fiebig, Eberhard W, David J Wright, Bhupat D Rawal, Patricia E Garrett, Richard T Schumacher, Lorraine Peddada, Charles Heldebrant, et al. 2003. “Dynamics of HIV Viremia and Antibody Seroconversion in Plasma Donors: Implications for Diagnosis and Staging of Primary HIV Infection.” *AIDS* 17 (13): 1871–79.

Frahm, Nicole, B T Korber, C M Adams, J J Szinger, R Draenert, M M Addo, M E Feeney, et al. 2004. “Consistent Cytotoxic-T-Lymphocyte Targeting of Immunodominant Regions in Human Immunodeficiency Virus across Multiple Ethnicities” *Journal of Virology* 78 (5): 2187–2200.

Fraser, Christophe, Katrina Lythgoe, Gabriel E Leventhal, George Shirreff, T Déirdre Hollingsworth, Samuel Alizon, and Sebastian Bonhoeffer. 2014. “Virulence and Pathogenesis of HIV-1 Infection: An Evolutionary Perspective.” *Science* 343 (6177): 1243727.

Frater, John, Fiona Ewings, Jacob Hurst, Helen Brown, Nicola Robinson, Sarah Fidler, Abdel Babiker, Jonathan Weber, Kholoud Porter, and Rodney E Phillips. 2014. “HIV-1-Specific CD4(+) Responses in Primary HIV-1 Infection Predict Disease Progression.” *AIDS* 28 (5): 699–708.

Frost, Simon D W, Terri Wrin, Davey M Smith, Sergei L Kosakovsky Pond, Yang Liu, Ellen Paxinos, Colombe Chappey, et al. 2005. “Neutralizing Antibody Responses Drive the Evolution of Human Immunodeficiency Virus Type 1 Envelope during Recent HIV Infection.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (51): 18514–19.

Ganeshan, S, R E Dickover, B T Korber, Y J Bryson, and S M Wolinsky. 1997. “Human Immunodeficiency Virus Type 1 Genetic Evolution in Children with Different Rates of Development of Disease.” *Journal of Virology* 71 (1): 663–77.

Gijsbers, Esther F, K Anton Feenstra, Ad C van Nuenen, Marjon Navis, Jaap Heringa, Hanneke Schuitemaker, and Neeltje A Kootstra. 2013. “HIV-1 Replication Fitness of HLA-B*57/58:01 CTL Escape Variants Is Restored by the Accumulation of Compensatory Mutations in Gag.” *PloS One* 8 (12): e81235.

Gnanakaran, S, Tanmoy Bhattacharya, Marcus Daniels, Brandon F Keele, Peter T Hraber, Alan S Lapedes, Tongye Shen, et al. 2011. “Recurrent Signature Patterns in HIV-1 B Clade Envelope Glycoproteins Associated with Either Early or Chronic Infections.” *PLoS Pathogens* 7 (9): e1002209.

Goepfert, Paul A, Wendy Lumm, Paul Farmer, Philippa Matthews, Andrew Prendergast, Jonathan M Carlson, Cynthia A Derdeyn, et al. 2008. “Transmission of HIV-1 Gag Immune Escape Mutations Is Associated with

Reduced Viral Load in Linked Recipients." *The Journal of Experimental Medicine* 205 (5): 1009–17.

Goonetilleke, Nilu, Michael K P Liu, Jesus F Salazar-Gonzalez, Guido Ferrari, Elena Giorgi, Vitaly V Ganusov, Brandon F Keele, et al. 2009. "The First T Cell Response to Transmitted/founder Virus Contributes to the Control of Acute Viremia in HIV-1 Infection." *The Journal of Experimental Medicine* 206 (6): 1253–72.

Gottardo, Raphael, Robert T Bailer, Bette T Korber, S Gnanakaran, Joshua Phillips, Xiaoying Shen, Georgia D Tomaras, et al. 2013. "Plasma IgG to Linear Epitopes in the V2 and V3 Regions of HIV-1 gp120 Correlate with a Reduced Risk of Infection in the RV144 Vaccine Efficacy Trial." *PloS One* 8 (9): e75665.

Goulder, P J, R E Phillips, RA Colbert, S McAdam, G Ogg, M A Nowak, P Giangrande, et al. 1997. "Late Escape from an Immunodominant Cytotoxic T-Lymphocyte Response Associated with Progression to AIDS." *Nature Medicine* 3 (2): 212–17.

Goulder, Philip J R, and Bruce D Walker. 2012. "HIV and HLA Class I: An Evolving Relationship." *Immunity* 37 (3): 426–40.

Gray, Clive M, Mandla Mlotshwa, Catherine Riou, Tiyani Mathebula, Debra de Assis Rosa, Tumelo Mashishi, Cathal Seoighe, et al. 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."*Journal of Virology* 83 (1): 470–78.

Gray, E S, P L Moore, I A Choge, J M Decker, F Bibollet-Ruche, H Li, N Leseka, et al. 2007. "Neutralizing Antibody Responses in Acute Human Immunodeficiency Virus Type 1 Subtype C Infection."*Journal of Virology* 81 (12): 6187–96.

Gray, Glenda E, Zoe Moodie, Barbara Metch, Peter B Gilbert, Linda-Gail Bekker, Gavin Churchyard, Maphoshane Nchabeleng, et al. 2014. "Recombinant Adenovirus Type 5 HIV Gag/pol/nef Vaccine in South Africa: Unblinded, Long-Term Follow-up of the Phase 2b HVTN 503/Phambili Study." *The Lancet. Infectious Diseases* 14 (5): 388–96.

Haaland, Richard E, Paulina A Hawkins, Jesus Salazar-Gonzalez, Amber Johnson, Amanda Tichacek, Etienne Karita, Olivier Manigart, et al. 2009. "Inflammatory Genital Infections Mitigate a Severe Genetic Bottleneck in Heterosexual Transmission of Subtype A and C HIV-1." *PLoS Pathogens* 5 (1): e1000274.

Haynes, Barton F, Peter B Gilbert, Juliana McElrath, Susan Zolla-Pazner, Georgia D Tomaras, S Munir Alam, David T Evans, et al. 2012. "Immune-Correlates Analysis of an HIV-1 Vaccine Efficacy Trial." *New England Journal of Medicine* 366 (14): 1275–86.

Hecht, Frederick M, Wendy Hartogensis, Larry Bragg, Peter Bacchetti, Robert Grant, Jason Barbour, and Steven G Deeks. 2010. “HIV RNA Level in Early Infection Is Predicted by Viral Load in the Transmission Source.” *AIDS* 24 (7): 941–45.

Hemelaar, Joris, Eleanor Gouws, Peter D Ghys, and Saladin Osmanov. 2011. “Global Trends in Molecular Epidemiology of HIV-1 during 2000-2007.” *AIDS* 25 (5): 679–89.

Hendel, H, S Caillat-Zucman, H Lebuanec, M Carrington, S O’Brien, J M Andrieu, F Schächter, et al. 1999. “New Class I and II HLA Alleles Strongly Associated with Opposite Patterns of Progression to AIDS.” *Journal of Immunology* 162 (11): 6942–46.

Henn, Matthew R., Christian L. Boutwell, Patrick Charlebois, Niall J. Lennon, Karen A. Power, Alexander R. Macalalad, Aaron M. Berlin, et al. 2012. “Whole Genome Deep Sequencing of HIV-1 Reveals the Impact of Early Minor Variants Upon Immune Recognition During Acute Infection.” *PLoS Pathogens* 8 (3): e1002529.

Herbeck, Joshua T, David C Nickle, Gerald H Learn, Geoffrey S Gottlieb, Marcel E Curlin, Laura Heath, and James I Mullins. 2006. “Human Immunodeficiency Virus Type 1 Env Evolves toward Ancestral States upon Transmission to a New Host.” *Journal of Virology* 80 (4): 1637-44.

Herbeck, Joshua T, Morgane Rolland, Yi Liu, Sherry McLaughlin, John McNevin, Hong Zhao, Kim Wong, et al. 2011. “Demographic Processes Affect HIV-1 Evolution in Primary Infection before the Onset of Selective Processes.” *Journal of Virology* 85 (15): 7523–34.

Hersperger, Adam R, Florencia Pereyra, Martha Nason, Korey Demers, Prameet Sheth, Lucy Y Shin, Colin M Kovacs, et al. 2010. “Perforin Expression Directly Ex Vivo by HIV-Specific CD8 T-Cells Is a Correlate of HIV Elite Control.” *PLoS Pathogens* 6 (5): e1000917.

Hollingsworth, T Déirdre, Oliver Laeyendecker, George Shirreff, Christl A Donnelly, David Serwadda, Maria J Wawer, Noah Kiwanuka, et al. 2010. “HIV-1 Transmitting Couples Have Similar Viral Load Set-Points in Rakai, Uganda.” *PLoS Pathogens* 6 (5): e1000876.

Honeyborne, Isobella, Andrew Prendergast, Florencia Pereyra, Alasdair Leslie, Hayley Crawford, Rebecca Payne, Shabashini Reddy, et al. 2007. “Control of Human Immunodeficiency Virus Type 1 Is Associated with HLA-B*13 and Targeting of Multiple Gag-Specific CD8+ T-Cell Epitopes.” *Journal of Virology* 81 (7): 3667–72.

Hraber, Peter, Bette T Korber, Alan S Lapedes, Robert T Bailer, Michael S Seaman, Hongmei Gao, Kelli M Greene, et al. 2014. “Impact of Clade, Geography and Age of the Epidemic on Hiv-1 Neutralization By Antibodies.” *Journal of Virology* Epub ahead of print. pii: JVI.01705-14

Huang, Jinghe, Byong H. Kang, Marie Pancera, Jeong Hyun Lee, Tommy Tong, Yu Feng, Ivelin S. Georgiev, et al. 2014. “Broad and Potent HIV-1 Neutralization by a Human Antibody That Binds the gp41–gp120 Interface.” *Nature* Epub ahead of print. doi:10.1038/nature13601.

Jetzt, Amanda E, Hong Yu, George J Klarmann, Yacov Ron, Bradley D Preston, and Joseph P Dougherty. 2000. “High Rate of Recombination throughout the Human Immunodeficiency Virus Type 1 Genome.” *Journal of Virology* 74(3):1234-40.

Jones, Nicola A, Xiping Wei, Darren R Flower, Mailee Wong, Franziska Michor, Michael S Saag, Beatrice H Hahn, Martin A Nowak, George M Shaw, and Persephone Borrow. 2004. “Determinants of Human Immunodeficiency Virus Type 1 Escape from the Primary CD8+ Cytotoxic T Lymphocyte Response.” *The Journal of Experimental Medicine* 200 (10): 1243–56.

Julien, Jean-Philippe, Albert Cupo, Devin Sok, Robyn L Stanfield, Dmitry Lyumkis, Marc C Deller, Per-Johan Klasse, et al. 2013. “Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer.” *Science* 342 (6165): 1477–83.

Jung, Andreas, Reinhard Maier, Jean-Pierre Vartanian, Gennady Bocharov, Volker Jung, Ulrike Fischer, Eckart Meese, Simon Wain-Hobson, and Andreas Meyerhans. 2002. “Multiply Infected Spleen Cells in HIV Patients.” *Nature* 418 (6894): 144.

Karasavvas, Nicos, Erik Billings, Mangala Rao, Constance Williams, Susan Zolla-Pazner, Robert T Bailer, Richard A Koup, et al. 2012. “The Thai Phase III HIV Type 1 Vaccine Trial (RV144) Regimen Induces Antibodies That Target Conserved Regions within the V2 Loop of gp120.” *AIDS Research and Human Retroviruses* 28 (11): 1444–57.

Karim, Salim S Abdool, Gavin J Churchyard, Abdool Karim, and Stephen D Lawn. 2009. “HIV Infection and Tuberculosis in South Africa: An Urgent Need to Escalate the Public Health Response.” *Lancet* 374 (9693): 921–33.

Karn, Jonathan, and C Martin Stoltzfus. 2012. “Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression.” *Cold Spring Harbor Perspectives in Medicine* 2 (2): a006916.

Kaslow, RA, M Carrington, R Apple, L Park, A Munoz, AJ Saah, JJ Goedert, et al. 1996. “Influence of Combinations of Human Major Histocompatibility Complex Genes on the Course of HIV-1 Infection.” *Nature Medicine* 2 (4): 405–11.

Kaufmann, Daniel E, Paul M Bailey, John Sidney, Philip J Norris, Mary N Johnston, Lisa A Cosimi, Marylyn M Addo, et al. 2004. “Comprehensive Analysis of Human Immunodeficiency Virus Type 1-Specific CD4 Responses Reveals Marked Immunodominance of Gag and Nef and the Presence of Broadly Recognized Peptides.” *Journal of Virology* 78 (9): 4463–77.

Kawashima, Yuka, Katja Pfafferott, John Frater, Philippa Matthews, Rebecca Payne, Marylyn Addo, Hiroyuki Gatanaga, et al. 2009. “Adaptation of HIV-1 to Human Leukocyte Antigen Class I.” *Nature* 458 (7238): 641–45.

Keele, Brandon F, Elena E Giorgi, Jesus F Salazar-Gonzalez, Julie M Decker, Kimmy T Pham, Maria G Salazar, Chuanxi Sun, et al. 2008. “Identification and Characterization of Transmitted and Early Founder Virus Envelopes in Primary HIV-1 Infection.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (21): 7552–57.

Keele, Brandon F, Hui Li, Gerald H Learn, Peter Hraber, Elena E Giorgi, Truman Grayson, Chuanxi Sun, et al. 2009. “Low-Dose Rectal Inoculation of Rhesus Macaques by SIVsmE660 or SIVmac251 Recapitulates Human Mucosal Infection by HIV-1.” *The Journal of Experimental Medicine* 206 (5): 1117–34.

Kelleher, A D, C Long, E C Holmes, R L Allen, J Wilson, C Conlon, C Workman, et al. 2001. “Clustered Mutations in HIV-1 Gag Are Consistently Required for Escape from HLA-B27-Restricted Cytotoxic T Lymphocyte Responses.” *The Journal of Experimental Medicine* 193 (3): 375–86.

Kiepiela, Photini, Alasdair J Leslie, Isobella Honeyborne, Danni Ramduth, Christina Thobakgale, Senica Chetty, Prinisha Rathnavalu, et al. 2004. “Dominant Influence of HLA-B in Mediating the Potential Co-Evolution of HIV and HLA.” *Nature* 432 (7018): 769–75.

Kiepiela, Photini, Kholiswa Ngumbela, Christina Thobakgale, Dhanwanthie Ramduth, Isobella Honeyborne, Eshia Moodley, Shabashini Reddy, et al. 2007. “CD8+ T-Cell Responses to Different HIV Proteins Have Discordant Associations with Viral Load.” *Nature Medicine* 13 (1): 46–53.

Kosakovsky Pond, Sergei L, David Posada, Michael B Gravenor, Christopher H Woelk, and Simon D W Frost. 2006. “GARD: A Genetic Algorithm for Recombination Detection.” *Bioinformatics* 22 (24): 3096–98.

Laakso, Meg M, Fang-Hua Lee, Beth Haggarty, Caroline Agrawal, Katrina M Nolan, Mark Biscone, Josephine Romano, et al. 2007. “V3 Loop Truncations in HIV-1 Envelope Impart Resistance to Coreceptor Inhibitors and Enhanced Sensitivity to Neutralizing Antibodies.” *PLoS Pathogens* 3 (8): e117.

Lazaryan, Aleksandr, Elena Lobashevsky, Joseph Mulenga, Etienne Karita, Susan Allen, Jianming Tang, and Richard A Kaslow. 2006. “Human Leukocyte Antigen B58 Supertype and Human Immunodeficiency Virus Type 1 Infection in Native Africans.” *Journal of Virology* 80 (12): 6056–60.

Leslie, A J, K J Pfafferott, P Chetty, R Draenert, M M Addo, M Feeney, Y Tang, et al. 2004. “HIV Evolution: CTL Escape Mutation and Reversion after Transmission.” *Nature Medicine* 10 (3): 282–89.

Leslie, Alasdair, Philippa C Matthews, Jennifer Listgarten, Jonathan M Carlson, Carl Kadie, Thumbi Ndung’u, Christian Brander, et al. 2010. “Additive Contribution of

HLA Class I Alleles in the Immune Control of HIV-1 Infection." *Journal of Virology* 84 (19): 9879–88.

Létourneau, Sven, Eung-Jun Im, Tumelo Mashishi, Choechoe Brereton, Anne Bridgeman, Hongbing Yang, Lucy Dorrell, et al. 2007. "Design and Pre-Clinical Evaluation of a Universal HIV-1 Vaccine." *PloS One* 2 (10): e984.

Letvin, Norman L. 2006. "Progress and Obstacles in the Development of an AIDS Vaccine." *Nature Reviews. Immunology* 6 (12): 930–39.

Li, Bin, Adrianne D Gladden, Marcus Altfeld, John M Kaldor, David A Cooper, Anthony D Kelleher, and Todd M Allen. 2007. "Rapid Reversion of Sequence Polymorphisms Dominates Early Human Immunodeficiency Virus Type 1 Evolution."*Journal of Virology* 81 (1): 193–201.

Li, Hui, Katharine J Bar, Shuyi Wang, Julie M Decker, Yalu Chen, Chuanxi Sun, Jesus F Salazar-Gonzalez, et al. 2010. "High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men." *PLoS Pathogens* 6 (5): e1000890.

Liao, Hua-Xin, Chun-Yen Tsao, S Munir Alam, Mark Muldoon, Nathan Vandergrift, Ben-Jiang Ma, Xiaozhi Lu, et al. 2013. "Antigenicity and Immunogenicity of Transmitted/founder, Consensus, and Chronic Envelope Glycoproteins of Human Immunodeficiency Virus Type 1." *Journal of Virology* 87 (8): 4185–4201.

Lingappa, Jairam R, Katherine K Thomas, James P Hughes, Jared M Baeten, Anna Wald, Carey Farquhar, Guy De Bruyn, et al. 2013. "Partner Characteristics Predicting HIV-1 Set Point in Sexually Acquired HIV-1 Among African Seroconverters." *AIDS Research and Human Retroviruses* 29 (1): 164–71.

Liu, Jinyan, Brandon F Keele, Hui Li, Sheila Keating, Philip J Norris, Angela Carville, Keith G Mansfield, et al. 2010. "Low-Dose Mucosal Simian Immunodeficiency Virus Infection Restricts Early Replication Kinetics and Transmitted Virus Variants in Rhesus Monkeys." *Journal of Virology* 84 (19): 10406–12.

Liu, Michael K P, Natalie Hawkins, Adam J Ritchie, Vitaly V Ganusov, Victoria Whale, Simon Brackenridge, Hui Li, et al. 2013. "Vertical T Cell Immunodominance and Epitope Entropy Determine HIV-1 Escape." *Journal of Clinical Investigation* 123 (1): 380–93.

Liu, Shan-lu, Timothy Schacker, Luwy Musey, Daniel Shriner, M Julie McElrath, L Corey, and JI Mullins. 1997. "Divergent Patterns of Progression to AIDS after Infection from the Same Source : Human Immunodeficiency Virus Type 1 Evolution and Antiviral Responses." *Journal of Virology* 71 (6): 4284–95.

Liu, Yi, Marcel E Curlin, Kurt Diem, Hong Zhao, Ananta K Ghosh, Haiying Zhu, Amanda S Woodward, et al. 2008. "Env Length and N-Linked Glycosylation Following Transmission of Human Immunodeficiency Virus Type 1 Subtype B Viruses." *Virology* 374 (2): 229–33.

Liu, Yi, John McNevin, Jianhong Cao, Hong Zhao, Indira Genowati, Kim Wong, Sherry McLaughlin, et al. 2006. “Selection on the Human Immunodeficiency Virus Type 1 Proteome Following Primary Infection.” *Journal of Virology* 80 (19): 9519–29.

Lukashov, Vladimir V, Carla L Kuiken, and J Goudsmit. 1995. “Intrahost Human Immunodeficiency Virus Type 1 Evolution Is Related to Length of the Immunocompetent Period.” *Journal of Virology* 69 (11): 6911–16.

Lyles, R H, A Muñoz, T E Yamashita, H Bazmi, R Detels, C R Rinaldo, J B Margolick, J P Phair, and J W Mellors. 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.” *The Journal of Infectious Diseases* 181 (3): 872–80.

Lyumkis, Dmitry, Jean-Philippe Julien, Natalia de Val, Albert Cupo, Clinton S Potter, Per-Johan Klasse, Dennis R Burton, et al. 2013. “Cryo-EM Structure of a Fully Glycosylated Soluble Cleaved HIV-1 Envelope Trimer.” *Science* 342 (6165): 1484–90.

Malim, Michael H, and Paul D Bieniasz. 2012. “HIV Restriction Factors and Mechanisms of Evasion.” *Cold Spring Harbor Perspectives in Medicine* 2 (5): a006940.

Mansky, L M, and H M Temin. 1995. “Lower in Vivo Mutation Rate of Human Immunodeficiency Virus Type 1 than That Predicted from the Fidelity of Purified Reverse Transcriptase.” *Journal of Virology* 69 (8): 5087-94.

Markham, R B, W C Wang, A E Weisstein, Z Wang, A Munoz, A Templeton, J Margolick, et al. 1998. “Patterns of HIV-1 Evolution in Individuals with Differing Rates of CD4 T Cell Decline.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (21): 12568–73.

Markowitz, Martin, Michael Louie, Arlene Hurley, Eugene Sun, Michele Di Mascio, Alan S Perelson, and David D Ho. 2003. “A Novel Antiviral Intervention Results in More Accurate Assessment of Human Immunodeficiency Virus Type 1 Replication Dynamics and T-Cell Decay In Vivo.” *Journal of Virology* 77 (8): 5037–38.

Martin, D P, C Williamson, and D Posada. 2005. “RDP2: Recombination Detection and Analysis from Sequence Alignments.” *Bioinformatics* 21 (2): 260–62.

Martinez-Picado, Javier, Julia G Prado, Elizabeth E Fry, Katja Pfafferott, Alasdair Leslie, Senica Chetty, Christina Thobakgale, et al. 2006. “Fitness Cost of Escape Mutations in p24 Gag in Association with Control of Human Immunodeficiency Virus Type 1” 80 (7): 3617–23.

Mascola, J R, G Stiegler, T C VanCott, H Katinger, C B Carpenter, C E Hanson, H Beary, et al. 2000. “Protection of Macaques against Vaginal Transmission of a

Pathogenic HIV-1/SIV Chimeric Virus by Passive Infusion of Neutralizing Antibodies." *Nature Medicine* 6 (2): 207–10.

Mascola, John R, and Barton F Haynes. 2013. "HIV-1 Neutralizing Antibodies: Understanding Nature's Pathways." *Immunological Reviews* 254 (1): 225–44.

Mascola, John R, Mark G Lewis, Gabriela Stiegler, Thomas C Vancott, Deborah Hayes, Mark K Louder, Charles R Brown, et al. 1999. "Protection of Macaques against Pathogenic Simian / Human Immunodeficiency Virus 89.6PD by Passive Transfer of Neutralizing Antibodies." *Journal of Virology* 73 (5): 4009–18.

Masharsky, Alexey E, Elena N Dukhovlinova, Sergei V Verevochkin, Olga V Toussova, Roman V Skochilov, Jeffrey A Anderson, Irving Hoffman, Myron S Cohen, Ronald Swanstrom, and Andrei P Kozlov. 2010. "A Substantial Transmission Bottleneck among Newly and Recently HIV-1-Infected Injection Drug Users in St Petersburg, Russia." *The Journal of Infectious Diseases* 201 (11): 1697–1702.

Matthews, P. C., A. Prendergast, A. Leslie, H. Crawford, R. Payne, C. Rousseau, M. Rolland, et al. 2008. "Central Role of Reverting Mutations in HLA Associations with Human Immunodeficiency Virus Set Point."*Journal of Virology* 82 (17): 8548–59.

Mccaffrey, Ruth A, Cheryl Saunders, Mike Hensel, and Leonidas Stamatatos. 2004. "N-Linked Glycosylation of the V3 Loop and the Immunologically Silent Face of gp120 Protects Human Immunodeficiency Virus Type 1 SF162 from Neutralization by Anti-gp120 and Anti-gp41 Antibodies" *Journal of Virology* 78 (7): 3279–95.

McDonald, Richard A, Douglas L Mayers, Raymond C Chung, Kenneth F Wagner, Silvia Ratto-Kim, Deborah L Birx, and Nelson L Michael. 1997. "Evolution of Human Immunodeficiency Virus Type 1 Env Sequence Variation in Patients with Diverse Rates of Disease Progression and T-Cell Function." *Journal of Virology* 71 (3): 1871–79.

McKeating, J A, A McKnight, and J P Moore. 1991. "Differential Loss of Envelope Glycoprotein gp120 from Virions of Human Immunodeficiency Virus Type 1 Isolates: Effects on Infectivity and Neutralization." *Journal of Virology* 65 (2): 852–60.

McMichael, Andrew J, Persephone Borrow, Georgia D Tomaras, Nilu Goonetilleke, and Barton F Haynes. 2010. "The Immune Response during Acute HIV-1 Infection: Clues for Vaccine Development." *Nature Reviews Immunology* 10 (1): 11–23.

Mellors, J W, A Muñoz, J V Giorgi, J B Margolick, C J Tassoni, P Gupta, L A Kingsley, et al. 1997. "Plasma Viral Load and CD4+ Lymphocytes as Prognostic Markers of HIV-1 Infection." *Annals of Internal Medicine* 126 (12): 946–54.

Migueles, S A, M S Sabbaghian, W L Shupert, M P Bettinotti, F M Marincola, L Martino, C W Hallahan, et al. 2000. “HLA B*5701 Is Highly Associated with Restriction of Virus Replication in a Subgroup of HIV-Infected Long Term Nonprogressors.” *Proceedings of the National Academy of Sciences of the United States of America* 97 (6): 2709–14.

Migueles, Stephen A, Alisha C Laborico, W Lesley Shupert, M Shirin Sabbaghian, Ronald Rabin, Claire W Hallahan, Debbie Van Baarle, et al. 2002. “HIV-Specific CD8+ T Cell Proliferation Is Coupled to Perforin Expression and Is Maintained in Nonprogressors.” *Nature Immunology* 3 (11): 1061–68.

Migueles, Stephen A, Christine M Osborne, Cassandra Royce, Alex A Compton, Rohan P Joshi, Kristin A Weeks, Julia E Rood, et al. 2008. “Lytic Granule Loading of CD8+ T Cells Is Required for HIV-Infected Cell Elimination Associated with Immune Control.” *Immunity* 29 (6):1009–21.

Miura, Toshiyuki, Mark A Brockman, Arne Schneidewind, Michael Lobritz, Florencia Pereyra, Almas Rathod, Brian L Block, et al. 2009. “HLA-B57/B*5801 Human Immunodeficiency Virus Type 1 Elite Controllers Select for Rare Gag Variants Associated with Reduced Viral Replication Capacity and Strong Cytotoxic T-Lymphocyte Recognition.” *Journal of Virology* 83 (6): 2743–55.

Miura, Toshiyuki, Zabrina L Brumme, Mark A Brockman, Pamela Rosato, Jennifer Sela, Chanson J Brumme, Florencia Pereyra, et al. 2010. “Impaired Replication Capacity of Acute/early Viruses in Persons Who Become HIV Controllers.”*Journal of Virology* 84 (15): 7581–91.

Mlisana, Koleka, Nivashnee Naicker, Lise Werner, Linidi Roberts, Francois van Loggerenberg, Cheryl Baxter, Jo-Ann S. Passmore, et al. 2012. "Symptomatic Vaginal Discharge Is a Poor Predictor of Sexually Transmitted Infections and Genital Tract Inflammation in High-Risk Women in South Africa." *Journal of Infectious Diseases* 206: 6-14.

Mlotshwa, Mandla, Catherine Riou, Denis Chopera, Debra de Assis Rosa, Roman Ntale, Florette Treunicht, Zenda Woodman, et al. 2010. “Fluidity of HIV-1-Specific T-Cell Responses during Acute and Early Subtype C HIV-1 Infection and Associations with Early Disease Progression.”*Journal of Virology* 84 (22): 12018–29.

Mochizuki, N, N Otsuka, K Matsuo, T Shiino, A Kojima, T Kurata, K Sakai, et al. 1999. “An Infectious DNA Clone of HIV Type 1 Subtype C.” *AIDS Research and Human Retroviruses* 15 (14): 1321–24.

Moore, Corey B, Mina John, Ian R James, Frank T Christiansen, Campbell S Witt, and Simon A Mallal. 2002. “Evidence of HIV-1 Adaptation to HLA-Restricted Immune Responses at a Population Level.” *Science* 296 (5572): 1439–43.

Moore, J P, J A McKeating, R A Weiss, and Q J Sattentau. 1990. “Dissociation of gp120 from HIV-1 Virions Induced by Soluble CD4.” *Science* 250 (4984): 1139–42.

Moore, P. L., E. S. Gray, I. A. Choge, N. Ranchobe, K. Mlisana, S. S. Abdool Karim, C. Williamson, and L. Morris. 2008. “The C3-V4 Region Is a Major Target of Autologous Neutralizing Antibodies in Human Immunodeficiency Virus Type 1 Subtype C Infection.”*Journal of Virology* 82 (4): 1860–69.

Moore, Penny L, Elin S Gray, C Kurt Wibmer, Jinal N Bhiman, Molati Nonyane, Daniel J Sheward, Tandile Hermanus, et al. 2012. “Evolution of an HIV Glycan – Dependent Broadly Neutralizing Antibody Epitope through Immune Escape.” *Nature Medicine* 18 (11): 1688–92.

Moore, Penny L, Nthabeleng Ranchobe, Bronwen E Lambson, Elin S Gray, Eleanor Cave, Melissa-Rose Abrahams, Gama Bandawe, et al. 2009. “Limited Neutralizing Antibody Specificities Drive Neutralization Escape in Early HIV-1 Subtype C Infection.” *PLoS Pathogens* 5 (9): e1000598.

Münch, Jan, Elke Rücker, Ludger Ständker, Knut Adermann, Christine Goffinet, Michael Schindler, Steffen Wildum, et al. 2007. “Semen-Derived Amyloid Fibrils Drastically Enhance HIV Infection.” *Cell* 131 (6): 1059–71.

National Department of Health South Africa. 2004. *National Antiretroviral Treatment Guidelines*.

Nawaz, Fatima, Claudia Cicala, Donald Van Ryk, Katharine E. Block, Katja Jelicic, Jonathan P. McNally, Olajumoke Ogundare, et al. 2011. "The Genotype of Early-Transmitting HIV gp120s Promotes α4β7 -Reactivity, Revealing $\alpha4\beta7^{+}/CD4^{+}$ T cells As Key Targets in Mucosal Transmission." *PLoS Pathogens* 7 (2): e1001301.

Ndung’u, Thumbi, and Robin A Weiss. 2012. “On HIV Diversity.” *AIDS* 26 (10): 1255–60.

Ndung'u, Thumbi, Yichen Lu, Boris Renjifo, Neal Touzjian, Nicholas Kushner, Victor Pena-cruz, Vladimir A Novitsky, Tun-hou Lee, and Max Essex. 2001. “Infectious Simian/Human Immunodeficiency Virus with Human Immunodeficiency Virus Type 1 Subtype C from an African Isolate : Rhesus Macaque Model.” *Journal of Virology* 75 (23): 11417–25.

Ngumbela, K C, C L Day, Z Mncube, K Nair, D Ramduth, C Thobakgale, E Moodley, et al. 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 Research and Human Retroviruses* 24 (1): 72–82.

Nickle, David C, Morgane Rolland, Mark A Jensen, Sergei L Kosakovsky Pond, Wenjie Deng, Mark Seligman, David Heckerman, James I Mullins, and Nebojsa Jojic. 2007. “Coping with Viral Diversity in HIV Vaccine Design.” *PLoS Computational Biology* 3 (4): e75.

Nofemela, Andile, Gama Bandawe, Ruwayhida Thebus, Jinny Marais, Natasha Wood, Oliver Hoffmann, Leonard Maboko, Michael Hoelscher, Zenda Woodman, and Carolyn Williamson. 2011. “Defining the Human

Immunodeficiency Virus Type 1 Transmission Genetic Bottleneck in a Region with Multiple Circulating Subtypes and Recombinant Forms." *Virology* 415 (2):107–13.

Nowak, Martin A. 1992. "What Is a Quasispecies?" *Trends in Ecology and Evolution* 7 (4): 118–21.

Ochsenbauer, Christina, Tara G Edmonds, Haitao Ding, Brandon F Keele, Julie Decker, Maria G Salazar, Jesus F Salazar-Gonzalez, et al. 2012. "Generation of Transmitted/founder HIV-1 Infectious Molecular Clones and Characterization of Their Replication Capacity in CD4 T Lymphocytes and Monocyte-Derived Macrophages." *Journal of Virology* 86 (5): 2715–28.

Ogg, G S, X Jin, S Bonhoeffer, P R Dunbar, M A Nowak, S Monard, J P Segal, et al. 1998. "Quantitation of HIV-1-Specific Cytotoxic T Lymphocytes and Plasma Load of Viral RNA." *Science* 279 (5359): 2103–6.

Palmer, Sarah, Mary Kearney, Frank Maldarelli, K Elias, Christian J Bixby, Holly Bazmi, Diane Rock, et al. 2005. "Multiple, Linked Human Immunodeficiency Virus Type 1 Drug Resistance Mutations in Treatment-Experienced Patients Are Missed by Standard Genotype Analysis." *Journal of Clinical Microbiology* 43 (1): 406–13.

Parker, Zahra F, Shilpa S Iyer, Craig B Wilen, Nicholas F Parrish, Kelechi C Chikere, Fang-Hua Lee, Chuka A Didigu, et al. 2013. "Transmitted/founder and Chronic HIV-1 Envelope Proteins Are Distinguished by Differential Utilization of CCR5." *Journal of Virology* 87 (5): 2401–11.

Parrish, Nicholas F, Feng Gao, Hui Li, Elena E Giorgi, Hannah J Barbian, Erica H Parrish, Lara Zajic, et al. 2013. "Phenotypic Properties of Transmitted Founder HIV-1." *Proceedings of the National Academy of Sciences of the United States of America* 110 (17): 6626–33.

Parrish, Nicholas F, Craig B Wilen, Lauren B Banks, Shilpa S Iyer, Jennifer M Pfaff, Jesus F Salazar-Gonzalez, Maria G Salazar, et al. 2012. "Transmitted/founder and Chronic Subtype C HIV-1 Use CD4 and CCR5 Receptors with Equal Efficiency and Are Not Inhibited by Blocking the Integrin α4β7." *PLoS Pathogens* 8 (5): e1002686.

Pereyra, Florencia, Marylyn M Addo, Daniel E Kaufmann, Yang Liu, Toshiyuki Miura, Almas Rathod, Brett Baker, et al. 2008. "Genetic and Immunologic Heterogeneity among Persons Who Control HIV Infection in the Absence of Therapy." *The Journal of Infectious Diseases* 197 (4): 563–71.

Pereyra, Florencia, Xiaoming Jia, Paul J McLaren, Amalio Telenti, Paul I W de Bakker, Bruce D Walker, Stephan Ripke, et al. 2010. "The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation." *Science* 330 (6010): 1551–57.

Perez, Lautaro G, Haiyan Chen, Hua-Xin Liao and David C. Montifiori. 2014. "Envelope Glycoprotein Binding to the Integrin α4β7 Is Not a General Property of Most HIV-1 Strains." *Journal of Virology* 88 (18): 10767-10777.

Piguet, V, L Wan, C Borel, A Mangasarian, N Demaurex, G Thomas, and D Trono. 2000. "HIV-1 Nef Protein Binds to the Cellular Protein PACS-1 to Downregulate Class I Major Histocompatibility Complexes." *Nature Cell Biology* 2 (3): 163–67.

Pillay-van Wyk, Victoria, William Msemburi, Ria Laubscher, Rob E Dorrington, Pam Groenewald, Richard Matzopoulos, Megan Prinsloo, et al. 2013. "Second National Burden of Disease Study South Africa: National and Subnational Mortality Trends, 1997–2009." *The Lancet* 381: S113.

Ping, Li-Hua, Sarah B Joseph, Jeffrey A Anderson, Melissa-Rose Abrahams, Jesus F Salazar-Gonzalez, Laura P Kincer, Florette K Treurnicht, et al. 2013. "Comparison of Viral Env Proteins from Acute and Chronic Infections with Subtype C Human Immunodeficiency Virus Type 1 Identifies Differences in Glycosylation and CCR5 Utilization and Suggests a New Strategy for Immunogen Design." *Journal of Virology* 87 (13): 7218–33.

Pond, Sergei Kosakovsky, and Spencer V Muse. 2005. "Site-to-Site Variation of Synonymous Substitution Rates." *Mol Biol Evol* 22 (12): 2375–85.

Pond, Sergei L Kosakovsky, Simon D W Frost, and Spencer V Muse. 2005. "HyPhy: Hypothesis Testing Using Phylogenies." *Bioinformatics* 21 (5): 676–79.

Prado, Julia G, Sandra Franco, Tania Matamoros, Lidia Ruiz, Bonaventura Clotet, Luis Menéndez-Arias, Miguel Angel Martínez, and Javier Martinez-Picado. 2004. "Relative Replication Fitness of Multi-Nucleoside Analogue-Resistant HIV-1 Strains Bearing a Dipeptide Insertion in the Fingers Subdomain of the Reverse Transcriptase and Mutations at Codons 67 and 215." *Virology* 326 (1): 103–12.

Price, D A, P J Goulder, P Klenerman, A K Sewell, P J Easterbrook, M Troop, C R Bangham, and R E Phillips. 1997. "Positive Selection of HIV-1 Cytotoxic T Lymphocyte Escape Variants during Primary Infection." *Proceedings of the National Academy of Sciences of the United States of America* 94 (5): 1890–95.

Prince, Jessica L, Daniel T Claiborne, Jonathan M Carlson, Malinda Schaefer, Tianwei Yu, Shabir Lahki, Heather A Prentice, et al. 2012. "Role of Transmitted Gag CTL Polymorphisms in Defining Replicative Capacity and Early HIV-1 Pathogenesis." *PLoS Pathogens* 8 (11): e1003041.

Quiñones-Mateu, M E, S C Ball, A J Marozsan, V S Torre, J L Albright, G Vanham, G van Der Groen, R L Colebunders, and E J Arts. 2000. "A Dual Infection/competition Assay Shows a Correlation between Ex Vivo Human Immunodeficiency Virus Type 1 Fitness and Disease Progression." *Journal of Virology* 74 (19): 9222–33.

Quiñones-Mateu, Miguel E, and Eric J Arts. 2002. "Fitness of Drug Resistant HIV-1: Methodology and Clinical Implications." *Drug Resistance Updates* 5 (6): 224–33.

Raboni, Sonia Mara, Sérgio Monteiro De Almeida, Indianara Rotta, Cléa Elisa Lopes Ribeiro, Debra Rosario, Luine Rosele Vidal, Meri Bordignon Nogueira, et al. 2010. “Molecular Epidemiology of HIV-1 Clades in Southern Brazil.” *Memórias Do Instituto Oswaldo Cruz* 105 (8): 1044–49.

Redd, Andrew D, Aleisha N Collinson-Streng, Nikolaos Chatziandreou, Caroline E Mullis, Oliver Laeyendecker, Craig Martens, Stacy Ricklefs, et al. 2012. “Previously Transmitted HIV-1 Strains Are Preferentially Selected during Subsequent Sexual Transmissions.” *The Journal of Infectious Diseases* 206 (9): 1433–42.

Reitter, Julie N, Robert E Means, and Ronald C Desrosiers. 1998. “A Role for Carbohydrates in Immune Evasion in AIDS.” *Nature Medicine* 4 (6): 679–84.

Rerks-Ngarm, Supachai, Punnee Pitisuttithum, Sorachai Nitayaphan, Jaranit Kaewkungwal, Joseph Chiu, Robert Paris, Nakorn Premsri, et al. 2009. “Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand.” *New England Journal of Medicine* 361 (23): 2209–20.

Richman, Douglas D, Terri Wrin, Susan J Little, and Christos J Petropoulos. 2003. “Rapid Evolution of the Neutralizing Antibody Response to HIV Type 1 Infection.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (7): 4144–49..

Rolland, Morgane, Jonathan M Carlson, Siriphan Manocheewa, J Victor Swain, Erinn Lanxon-Cookson, Wenjie Deng, Christine M Rousseau, et al. 2010. “Amino-Acid Co-Variation in HIV-1 Gag Subtype C: HLA-Mediated Selection Pressure and Compensatory Dynamics.” *PloS One* 5 (9): 1–10.

Rolland, Morgane, Paul T Edlefsen, Brendan B Larsen, Sodsai Tovanabutra, Eric Sanders-Buell, Tomer Hertz, Allan C deCamp, et al. 2012. “Increased HIV-1 Vaccine Efficacy against Viruses with Genetic Signatures in Env V2.” *Nature* 490 (7420): 417–20.

Rolland, Morgane, David C Nickle, and James I Mullins. 2007. “HIV-1 Group M Conserved Elements Vaccine.” *PLoS Pathogens* 3 (11): e157.

Rousseau, Christine M, Brian A Birditt, Angela R McKay, Julia N Stoddard, Tsan Chun Lee, Sherry McLaughlin, Sarah W Moore, et al. 2006. “Large-Scale Amplification, Cloning and Sequencing of near Full-Length HIV-1 Subtype C Genomes.” *Journal of Virological Methods* 136 (1-2): 118–25.

Rousseau, Christine M, Marcus G Daniels, Jonathan M Carlson, Carl Kadie, Hayley Crawford, Andrew Prendergast, Philippa Matthews, et al. 2008. “HLA Class I-Driven Evolution of Human Immunodeficiency Virus Type 1 Subtype c Proteome: Immune Escape and Viral Load.” *Journal of Virology* 82 (13): 6434–46.

Rusert, Peter, Anders Krarup, Carsten Magnus, Oliver F Brandenberg, Jacqueline Weber, Anna-Katharina Ehlert, Roland R Regoes, Huldrych F Günthard, and

Alexandra Trkola. 2011. “Interaction of the gp120 V1V2 Loop with a Neighboring gp120 Unit Shields the HIV Envelope Trimer against Cross-Neutralizing Antibodies.” *The Journal of Experimental Medicine* 208 (7): 1419–33.

Rybarczyk, Brian J, David Montefiori, Philip R Johnson, Ande West, Robert E Johnston, and Ronald Swanstrom. 2004. “Correlation between Env V1/V2 Region Diversification and Neutralizing Antibodies during Primary Infection by Simian Immunodeficiency Virus Sm in Rhesus Macaques.”*Journal of Virology* 78 (7): 3561–71.

Sáez-Cirión, Asier, Christine Lacabaratz, Olivier Lambotte, Pierre Versmisse, Alejandra Urrutia, Faroudy Boufassa, Françoise Barré-Sinoussi, et al. 2007. “HIV Controllers Exhibit Potent CD8 T Cell Capacity to Suppress HIV Infection Ex Vivo and Peculiar Cytotoxic T Lymphocyte Activation Phenotype.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (16): 6776–81.

Sagar, Manish, Oliver Laeyendecker, Sandra Lee, Jordyn Gamiel, Maria J Wawer, Ronald H Gray, David Serwadda, et al. 2009. “Selection of HIV Variants with Signature Genotypic Characteristics during Heterosexual Transmission.” *The Journal of Infectious Diseases* 199 (4): 580–89.

Sagar, Manish, Xueling Wu, Sandra Lee, and Julie Overbaugh. 2006. “Human Immunodeficiency Virus Type 1 V1-V2 Envelope Loop Sequences Expand and Add Glycosylation Sites over the Course of Infection, and These Modifications Affect Antibody Neutralization Sensitivity.” *Journal of Virology* 80 (19): 9586–98.

Salazar-Gonzalez, Jesus F, Elizabeth Bailes, Kimmy T Pham, Maria G Salazar, M Brad Guffey, Brandon F Keele, Cynthia A Derdeyn, et al. 2008. “Deciphering Human Immunodeficiency Virus Type 1 Transmission and Early Envelope Diversification by Single-Genome Amplification and Sequencing.”*Journal of Virology* 82 (8): 3952–70.

Salazar-Gonzalez, Jesus F, Maria G Salazar, Brandon F Keele, Gerald H Learn, Elena E Giorgi, Hui Li, Julie M Decker, et al. 2009. “Genetic Identity, Biological Phenotype, and Evolutionary Pathways of Transmitted/founder Viruses in Acute and Early HIV-1 Infection.” *Journal of Experimental Medicine* 206 (6): 1273–89.

Salgado, Maria, Robert W Buckheit, Gail V Berkenblit, and Joel N Blankson. 2014. “Evolution of an Attenuated HIV-1 Isolate in an Elite Suppressor.” *AIDS Research and Human Retroviruses* 30 (3): 284–88.

Salminen, M O, P K Ehrenberg, J R Mascola, D E Dayhoff, R Merling, B Blake, M Louder, et al. 2000. “Construction and Biological Characterization of Infectious Molecular Clones of HIV-1 Subtypes B and E (CRF01_AE) Generated by the Polymerase Chain Reaction.” *Virology* 278 (1): 103–10.

Santra, Sampa, Hua-Xin Liao, Ruijin Zhang, Mark Muldoon, Sydeaka Watson, Will Fischer, James Theiler, et al. 2010. “Mosaic Vaccines Elicit CD8+ T Lymphocyte

Responses That Confer Enhanced Immune Coverage of Diverse HIV Strains in Monkeys." *Nature Medicine* 16 (3): 324–28.

Schaller, Torsten, Karen E Ocwieja, Jane Rasaiyaah, Amanda J Price, Troy L Brady, Shoshannah L Roth, Stéphane Hué, et al. 2011. "HIV-1 Capsid-Cyclophilin Interactions Determine Nuclear Import Pathway, Integration Targeting and Replication Efficiency." *PLoS Pathogens* 7 (12): e1002439.

Scheffler, Konrad, Darren P Martin, and Cathal Seoighe. 2006. "Robust Inference of Positive Selection from Recombining Coding Sequences." *Bioinformatics* 22 (20): 2493–99.

Schneidewind, Arne, Mark A Brockman, John Sidney, Yaoyu E Wang, Huabiao Chen, Todd J Suscovich, Bin Li, et al. 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." *Journal of Virology* 82 (11): 5594–5605.

Shankarappa, R, J B Margolick, S J Gange, A G Rodrigo, D Upchurch, H Farzadegan, P Gupta, et al. 1999. "Consistent Viral Evolutionary Changes Associated with the Progression of Human Immunodeficiency Virus Type 1 Infection." *Journal of Virology* 73 (12): 10489–502.

Shibata, R, T Igarashi, N Haigwood, A Buckler-White, R Ogert, W Ross, R Willey, M W Cho, and M A Martin. 1999. "Neutralizing Antibody Directed against the HIV-1 Envelope Glycoprotein Can Completely Block HIV-1/SIV Chimeric Virus Infections of Macaque Monkeys." *Nature Medicine* 5 (2): 204–10.

Shingai, Masashi, Yoshiaki Nishimura, Florian Klein, Hugo Mouquet, Olivia K Donau, Ronald Plishka, Alicia Buckler-White, et al. 2013. "Antibody-Mediated Immunotherapy of Macaques Chronically Infected with SHIV Suppresses Viraemia." *Nature* 503 (7475): 277–80.

Shioda, T, S Oka, X Xin, H Liu, R Harukuni, A Kurotani, M Fukushima, et al. 1997. "In Vivo Sequence Variability of Human Immunodeficiency Virus Type 1 Envelope gp120 : Association of V2 Extension with Slow Disease Progression." *Journal of Virology* 71 (7): 4871–81.

Simmonds, P, P Balfe, C A Ludlam, J O Bishop, and A J Brown. 1990. "Analysis of Sequence Diversity in Hypervariable Regions of the External Glycoprotein of Human Immunodeficiency Virus Type 1." *Journal of Virology* 64 (12): 5840–50.

Slatkin, Montgomery, and Richard R Hudson. 1991. "Pairwise Comparisons of Mitochondria DNA Sequences in Stable and Exponentially Growing Populations." *Genetics* 562: 555–62.

Song, Hongshuo, Bhavna Hora, Tanmoy Bhattacharya, Nilu Goonetilleke, Michael K P Liu, Kevin Wiehe, Hui Li, et al. 2014. "Reversion and T Cell Escape Mutations Compensate the Fitness Loss of a CD8+ T Cell Escape Mutant in Their Cognate Transmitted/founder Virus." *PloS One* 9 (7): e102734.

Song, Hongshuo, Jeffrey W Pavlicek, Fangping Cai, Tanmoy Bhattacharya, Hui Li, Shilpa S Iyer, Katharine J Bar, et al. 2012. “Impact of Immune Escape Mutations on HIV-1 Fitness in the Context of the Cognate Transmitted/founder Genome.” *Retrovirology* 9 (1): 89.

Statistics South Africa. 2013. “Statistical Release: Mid-Year Population Estimates.”

T.A.Hall. 1999. “BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT.” *Nucleic Acids Symposium Series* 41: 95–98.

Tamura, Koichiro, Joel Dudley, Masatoshi Nei, and Sudhir Kumar. 2007. “MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0.” *Molecular Biology and Evolution* 24 (8): 1596–99.

Tang, Jianming, Shenghui Tang, Elena Lobashevsky, Angela D Myracle, Ulgen Fideli, Grace Aldrovandi, Susan Allen, Rosemary Musonda, Richard A Kaslow, and Zambia-UAb HIV Research Project. 2002a. “Favorable and Unfavorable HLA Class I Alleles and Haplotypes in Zambians Predominantly Infected with Clade C Human Immunodeficiency Virus Type 1.” *Journal of Virology* 76 (16): 8276–84.

Tang, Jianming, Shenghui Tang, Elena Lobashevsky, Isaac Zulu, Grace Aldrovandi, Susan Allen, and Richard A Kaslow. 2004. “HLA Allele Sharing and HIV Type 1 Viremia in Seroconverting Zambians with Known Transmitting Partners.” *AIDS Research and Human Retroviruses* 20 (1): 19–25.

Tang, Jianming, Craig M Wilson, Shreelatha Meleth, Angela Myracle, Elena Lobashevsky, Mark J Mulligan, Steven D Douglas, Bette Korber, Sten H Vermund, and Richard A Kaslow. 2002b. “Host Genetic Profiles Predict Virological and Immunological Control of HIV-1 Infection in Adolescents.” *AIDS* 16 (17): 2275–84.

Tanser, Frank, Till Bärnighausen, Erofili Grapsa, Jaffer Zaidi, and Marie-Louise Newell. 2013. “High Coverage of ART Associated with Decline in Risk of HIV Acquisition in Rural KwaZulu-Natal, South Africa.” *Science* 339 (6122): 966–71.

Tenzer, Stefan, Edmund Wee, Anne Burgevin, Guillaume Stewart-Jones, Lone Friis, Kasper Lamberth, Chih-hao Chang, et al. 2009. “Antigen Processing Influences HIV-Specific Cytotoxic T Lymphocyte Immunodominance.” *Nature Immunology* 10 (6): 636–46.

Tomaras, Georgia D, Guido Ferrari, Xiaoying Shen, S Munir Alam, Hua-Xin Liao, Justin Pollara, Mattia Bonsignori, et al. 2013. “Vaccine-Induced Plasma IgA Specific for the C1 Region of the HIV-1 Envelope Blocks Binding and Effector Function of IgG.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (22): 9019–24.

Tomaras, Georgia D, Nicole L Yates, Pinghuang Liu, Li Qin, Genevieve G Fouda, Leslie L Chavez, Allan C Decamp, et al. 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." *Journal of Virology* 82 (24): 12449–63.

Treurnicht, F K, C Seoighe, D P Martin, N Wood, M-R Abrahams, D De Assis Rosa, H Bredell, et al. 2010. "Adaptive Changes in HIV-1 Subtype C Proteins during Early Infection Are Driven by Changes in HLA-Associated Immune Pressure." *Virology* 396 (2): 213–25.

Trkola, Alexandra, Herbert Kuster, Christine Leemann, Claudia Ruprecht, Beda Joos, Amalio Telenti, Bernhard Hirschel, Rainer Weber, Sebastian Bonhoeffer, and Huldrych F Günthard. 2003. "Human Immunodeficiency Virus Type 1 Fitness Is a Determining Factor in Viral Rebound and Set Point in Chronic Infection." *Journal of Virology* 77 (24): 13146–55.

Troyer, Ryan M, Kalonji R Collins, Awet Abraha, Erika Fraundorf, Dawn M Moore, Randall W Krizan, Zahra Toossi, et al. 2005. "Changes in Human Immunodeficiency Virus Type 1 Fitness and Genetic Diversity during Disease Progression" *Journal of Virology* 79 (14): 9006–18.

Troyer, Ryan M, John McNevin, Yi Liu, Shao Chong Zhang, Randall W Krizan, Awet Abraha, Denis M Tebit, et al. 2009. "Variable Fitness Impact of HIV-1 Escape Mutations to Cytotoxic T Lymphocyte (CTL) Response." *PLoS Pathogens* 5 (4): e1000365.

"UNAIDS AIDSinfo Epidemiological Status." http://www.unaids.org/en/dataanalysis/datatools/aidsinfo/.

Van Loggerenberg, Francois, Koleka Mlisana, Carolyn Williamson, Sara C Auld, Lynn Morris, Clive M Gray, Quarraisha Abdool Karim, et al. 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.

Vincze, T. 2003. "NEBcutter: A Program to Cleave DNA with Restriction Enzymes." *Nucleic Acids Research* 31 (13): 3688–91.

Walker, Bruce D, and Xu G Yu. 2013. "Unravelling the Mechanisms of Durable Control of HIV-1." *Nature Reviews Immunology* 13 (7): 487–98.

Walker, Laura M, Michael Huber, Katie J Doores, Emilia Falkowska, Robert Pejchal, Jean-Philippe Julien, Sheng-Kai Wang, et al. 2011. "Broad Neutralization Coverage of HIV by Multiple Highly Potent Antibodies." *Nature* 477 (7365): 466–70.

Wang, Yaoyu E, Bin Li, Jonathan M Carlson, Hendrik Streeck, Adrianne D Gladden, Robert Goodman, Arne Schneidewind, et al. 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." *Journal of Virology* 83 (4): 1845–55.

Weber, Jan, Ana C Vazquez, Dane Winner, Justine D Rose, Doug Wylie, Ariel M Rhea, Kenneth Henry, et al. 2011. "Novel Method for Simultaneous Quantification of Phenotypic Resistance to Maturation, Protease, Reverse Transcriptase, and Integrase HIV Inhibitors Based on 3'Gag(p2/p7/p1/p6)/PR/RT/INT-Recombinant Viruses: A Useful Tool in the Multitarget Era of Antiretroviral Therapy." *Antimicrobial Agents and Chemotherapy* 55 (8): 3729–42.

Wei, Xiping, Julie M Decker, Shuyi Wang, Huxiong Hui, John C Kappes, Xiaoyun Wu, Jesus F Salazar-Gonzalez, et al. 2003. "Antibody Neutralization and Escape by HIV-1." *Nature* 422 (6929): 307–12.

"WHO Global Health Observatory." http://www.who.int/gho/hiv/en/.

Wilen, Craig B, Nicholas F Parrish, Jennifer M Pfaff, Julie M Decker, Elizabeth a Henning, Hillel Haim, Josiah E Petersen, et al. 2011. "Phenotypic and Immunologic Comparison of Clade B Transmitted/founder and Chronic HIV-1 Envelope Glycoproteins." *Journal of Virology* 85 (17): 8514–27.

Wolfs, T F, J J de Jong, H Van den Berg, J M Tijnagel, W J Krone, and J Goudsmit. 1990. "Evolution of Sequences Encoding the Principal Neutralization Epitope of Human Immunodeficiency Virus 1 Is Host Dependent, Rapid, and Continuous." *Proceedings of the National Academy of Sciences of the United States of America* 87 (24): 9938–42.

Wolinsky, S M, B T Korber, A U Neumann, M Daniels, K J Kunstman, A J Whetsell, M R Furtado, Y Cao, D D Ho, and J T Safrit. 1996. "Adaptive Evolution of Human Immunodeficiency Virus-Type 1 during the Natural Course of Infection." *Science* 272 (5261): 537–42.

Wolinsky, Steven M, Carla M Wike, Bette T M Korber, Cecelia Hutto, P Wade, Lisa L Rosenblum, Kevin J Kunstman, Manohar R Furtado, and José L Muñoz. 1992. "Infants Selective Transmission of Human Immunodeficiency Virus Type-I Variants from Mothers to Infants." *Science* 255 (5048): 1134–37.

Wood, Natasha, Tanmoy Bhattacharya, Brandon F Keele, Elena Giorgi, Michael Liu, Brian Gaschen, Marcus Daniels, et al. 2009. "HIV Evolution in Early Infection: Selection Pressures, Patterns of Insertion and Deletion, and the Impact of APOBEC." *PLoS Pathogens* 5 (5): e1000414.

World Health Organization. 2011. "Global HIV/AIDS Response: Epidemic Update and Health Sector Progress towards Universal Access."

Wright, Jaclyn K, Zabrina L Brumme, Jonathan M Carlson, David Heckerman, Carl M Kadie, Chanson J Brumme, Bingxia Wang, et al. 2010. "Gag-Protease-Mediated Replication Capacity in HIV-1 Subtype C Chronic Infection : Associations with HLA Type and Clinical Parameters."*Journal of Virology* 84 (20): 10820–31.

Wright, Jaclyn K, Vladimir Novitsky, Mark A Brockman, Zabrina L Brumme, Chanson J Brumme, Jonathan M Carlson, David Heckerman, et al. 2011. “Influence of Gag-Protease-Mediated Replication Capacity on Disease Progression in Individuals Recently Infected with HIV-1 Subtype C.” *Journal of Virology* 85 (8): 3996–4006.

Yeh, Wendy W, Ishita Rahman, Peter Hraber, Rory T Coffey, Daiva Nevidomskyte, Ayush Giri, Mohammed Asmal, et al. 2010. “Autologous Neutralizing Antibodies to the Transmitted/founder Viruses Emerge Late after Simian Immunodeficiency Virus SIVmac251 Infection of Rhesus Monkeys.” *Journal of Virology* 84 (12): 6018–32.

York, Ian A, and Kenneth L Rock. 1996. “Antigen Processing Anf Presentation by the Class I Major Histocompatibility Complex.” *Annual Review of Immunology* 14: 369–96.

Yue, Ling, Heather A Prentice, Paul Farmer, Wei Song, Dongning He, Shabir Lakhi, Paul Goepfert, et al. 2013. “Cumulative Impact of Host and Viral Factors on HIV-1 Viral-Load Control during Early Infection.” *Journal of Virology* 87 (2): 708–15.

Yusim, Karina, Can Kesmir, Brian Gaschen, Marylyn M Addo, Marcus Altfeld, Søren Brunak, Alexandre Chigaev, Vincent Detours, and Bette T Korber. 2002. “Clustering Patterns of Cytotoxic T-Lymphocyte Epitopes in Human Immunodeficiency Virus Type 1 (HIV-1) Proteins Reveal Imprints of Immune Evasion on HIV-1 Global Variation” *Journal of Virology* 76 (17): 8757–68.

Zhu, Tuofu, Hongmei Mo, Ning Wang, Daniel S Nam, Yunzhen Cao, Richard A Koup, and David D Ho. 1993. “Genotypic and Phenotypic Characterization of HIV-1 in Patients with Primary Infection.” *Science* 261 (5125): 1179–81.

Zolla-Pazner, Susan, Allan DeCamp, Peter B Gilbert, Constance Williams, Nicole L Yates, William T Williams, Robert Howington, et al. 2014. “Vaccine-Induced IgG Antibodies to V1V2 Regions of Multiple HIV-1 Subtypes Correlate with Decreased Risk of HIV-1 Infection.” *PloS One* 9 (2): e87572.

Zuniga, Rosario, Aldo Lucchetti, Patricia Galvan, Shyla Sanchez, Carmen Sanchez, Ana Hernandez, Hugo Sanchez, et al. 2006. “Relative Dominance of Gag p24-Specific Cytotoxic T Lymphocytes Is Associated with Human Immunodeficiency Virus Control.” *Journal of Virology* 80 (6): 3122–25.

Appendix 1: Published manuscript Abrahams et al., 2009 *J Virol*

JOURNAL OF VIROLOGY, Apr. 2009, p. 3556–3567 Vol. 83, No. 8
0022-538X/09/$08.00+0 doi:10.1128/JVI.02132-08

Quantitating the Multiplicity of Infection with Human Immunodeficiency Virus Type 1 Subtype C Reveals a Non-Poisson Distribution of Transmitted Variants▽†

M.-R. Abrahams,[1] J. A. Anderson,[2] E. E. Giorgi,[3,4] C. Seoighe,[1] K. Mlisana,[5] L.-H. Ping,[2] G. S. Athreya,[3] F. K. Treurnicht,[1] B. F. Keele,[6] N. Wood,[1] J. F. Salazar-Gonzalez,[6] T. Bhattacharya,[3,7] H. Chu,[2] I. Hoffman,[2] S. Galvin,[2] C. Mapanje,[8] P. Kazembe,[8] R. Thebus,[1] S. Fiscus,[2] W. Hide,[9] M. S. Cohen,[2] S. Abdool Karim,[5] B. F. Haynes,[10] G. M. Shaw,[6] B. H. Hahn,[6] B. T. Korber,[3,7] R. Swanstrom,[2] and C. Williamson[1]* for the CAPRISA Acute Infection Study Team and the Center for HIV-AIDS Vaccine Immunology Consortium

Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa[1]; University of North Carolina at Chapel Hill, Chapel Hill, North Carolina[2]; Los Alamos National Laboratory, Los Alamos, New Mexico[3]; University of Massachusetts, Amherst, Massachusetts[4]; Centre for the AIDS Programme of Research in South Africa, University of Kwa-Zulu Natal, Durban, South Africa[5]; University of Alabama at Birmingham, Birmingham, Alabama[6]; Santa Fe Institute, Santa Fe, New Mexico[7]; Kamuzu Central Hospital, Lilongwe, Malawi[8]; South African Bioinformatics Institute, University of Western Cape, Cape Town, South Africa[9]; and Duke University Medical Center, Durham, North Carolina[10]

Received 9 October 2008/Accepted 24 December 2008

Identifying the specific genetic characteristics of successfully transmitted variants may prove central to the development of effective vaccine and microbicide interventions. Although human immunodeficiency virus transmission is associated with a population bottleneck, the extent to which different factors influence the diversity of transmitted viruses is unclear. We estimate here the number of transmitted variants in 69 heterosexual men and women with primary subtype C infections. From 1,505 env sequences obtained using a single genome amplification approach we show that 78% of infections involved single variant transmission and 22% involved multiple variant transmissions (median of 3). We found evidence for mutations selected for cytotoxic-T-lymphocyte or antibody escape and a high prevalence of recombination in individuals infected with multiple variants representing another potential escape pathway in these individuals. In a combined analysis of 171 subtype B and C transmission events, we found that infection with more than one variant does not follow a Poisson distribution, indicating that transmission of individual virions cannot be seen as independent events, each occurring with low probability. While most transmissions resulted from a single infectious unit, multiple variant transmissions represent a significant fraction of transmission events, suggesting that there may be important mechanistic differences between these groups that are not yet understood.

The development of a human immunodeficiency virus (HIV) vaccine remains a global priority. Since vaccines need to target viruses at, or close to, the time of infection, there is great interest in understanding both the genetic characteristics of viruses that are successfully transmitted and the genetic diversification that ensues during the earliest stages of infection. There is evidence from cross-sectional studies of individuals with acute and early HIV infections that transmission is associated with a population bottleneck (4, 9, 16, 35, 36, 39, 46–48). The best evidence of transmission bottlenecks is obtained from studies of HIV discordant couples, where infection is associated with transmission of one or a few genetic variants despite high degrees of viral diversity in the transmitting donors (13). A recent study was able to quantify this bottleneck by using methodologies that could identify precisely the transmitted, founder virus population (16). That study found that as many as 76% of transmission events involved productive clinical infection by only a single genetic variant in HIV-1 subtype B transmission by either heterosexual or homosexual routes. These findings are encouraging since in the majority of infections a vaccine-mediated response may have to protect against only a low-multiplicity infection event that is associated with low viral diversity in the initial days and weeks of infection.

However, it is also apparent that the transmission bottleneck can be overcome, as evidenced by the transmission of multiple viral variants (9, 10, 12, 16, 35, 36). The clinical implications of multiple variant transmission are potentially serious because coinfection with genetically divergent viral lineages has been associated with more severe disease progression (10, 12, 37). Current estimates of the frequency of multiple variant transmissions vary widely with between 0 and 50% of successful sexual transmissions estimated to involve the transfer of multiple genetic variants (9, 16, 35).

It is still an open question as to whether there are distinctive features of viruses that enhance their transmissibility and whether different risk behaviors are associated with different

* Corresponding author. Mailing address: Institute of Infectious Diseases and Molecular Medicine, Division of Medical Virology, Faculty of Health Sciences, University of Cape Town, Observatory, Cape Town 7925, South Africa. Phone: 27-21-406 6683. Fax: 27-21-406 6682. E-mail: carolyn.williamson@uct.ac.za.

† Supplemental material for this article may be found at http://jvi.asm.org/.

▽ Published ahead of print on 4 February 2009.

numbers of transmitted viruses. For example, intravenous infection has been associated with more heterogeneous infections than intravaginal infection in a rhesus macaque model (11). Besides the potential identification of previously unrecognized transmission risk factors, attempts to discover the variables influencing multivariant transmission event frequencies should also provide insights into natural barriers to HIV infection. Unfortunately, because of methodological variations between studies it has not been possible to directly compare results and to assess the impacts of biological factors such as gender, routes of transmission, the presence of sexually transmitted diseases and viral subtype on multiple variant transmission frequencies (39). Complex cofactors in a given risk category may impact rates of transmission of multiple variants just as they impact overall transmission rates (32).

In the present study, we investigated the genetic characteristics of subtype C variants transmitted via a heterosexual route in order to understand mechanisms involved in HIV transmission at genital sites. We used here the approach described by Keele et al. (16), which allowed us to directly compare multiple transmission frequencies between subtype B and C viruses. Our results indicate that similar proportions of subtype B and subtype C infections involve the transmission of multiple variants. Analysis of a total of 171 transmission events in both the subtype B and C studies showed that infection with multiple variants does not follow a Poisson distribution, suggesting that transmissions of multiple variants are not independent events in a setting of a low probability of infection. The question of the biological basis of the mucosal viral transmission bottleneck and factors associated with its breach, however, remains unanswered.

MATERIALS AND METHODS

Cohort description. Plasma was obtained from 69 individuals experiencing early HIV-1 subtype C infection. Twenty-six were identified through prospective monthly monitoring of HIV-negative women using rapid HIV tests and RNA PCR (44) (CAPRISA 002 cohort, Durban, South Africa); 22 were identified (i) with negative rapid HIV or enzyme immunoassay assay (EIA) antibody test results but positive HIV RNA and/or p24 antigen test results or (ii) who were antibody positive but antibody negative or Western blot indeterminate within the previous 45 days (CHAVI 001 cohorts, South Africa and Kamazu Central Hospital Sexually Transmitted Disease Clinic, Malawi). An additional 21 individuals enrolled from the Sexually Transmitted Disease Clinic in Malawi were identified as HIV PCR positive with a rapid HIV or EIA antibody-negative test or who had indeterminate HIV Western blot results (30).

This study was approved by institutional review boards, and all participants provided written informed consent.

Staging of HIV infection. The durations of HIV-1 infection were categorized into six stages based on evolving HIV-1 RNA or antibody profiles developed by Fiebig et al. (8). Plasma was tested for HIV RNA by using Roche Amplicor vRNA assays (Rotkreuz, Switzerland) and for antibodies by EIA (BEP 2000 [Dade Behring, Marburg, Germany] or Determine AntiHIV-1/2 3rd Generation EIA [Abbott, Illinois] and Uni-Gold Recombigen [Trinity Biotech, Ireland]) and a GS HIV-1 Western blot analysis kit (Bio-Rad, Washington). Individuals classified as being in stage I were viral RNA positive and p24 antigen and EIA antibody negative, those in stage II were RNA and p24 antigen positive but EIA antibody negative, those in stage III were EIA antibody positive but Western blot negative, those in stage IV had an indeterminate Western blot result, those in stage V were Western blot positive but without reactivity to the p31 integrase band, and those in stage VI were Western blot positive with a p31 band present.

PCR amplification and sequencing. HIV-1 RNA was extracted from 140 to 200 μl of ACD or EDTA plasma and eluted in a 50-μl final volume. The full volume of RNA was reverse transcribed to cDNA (100-μl reaction volume) by using a Superscript III reverse transcriptase (RT) system (Invitrogen, California) with an OFM-19 primer as described previously (16) or with oligo(dT). The cDNA was then serially diluted to obtain no more than 30% positive amplification reactions so that each amplicon would theoretically be amplified from a single template more than 80% of the time (39). This limiting dilution approach of *env* gene amplification was initially described by Simmonds et al. (41) and Edmonson and Mullins (7) and was thereafter modified by Palmer et al. (29) by sequencing of the amplicon directly and finally modified by Salazar-Gonzalez et al. (39), who showed that single genome amplification of the *env* gene with direct sequencing precludes recombination and Taq-induced error, and provides proportional representation of each viral sequence. PCR products were directly sequenced by using an ABI 3000 genetic analyzer (Applied Biosystems, Foster City, CA) and BigDye terminator reagents. To ensure that sequences reflected single templates from the viral populations in vivo, amplicons with sequence chromatograms with "double peaks," indicative of coamplification of more than one template, were excluded. Sequences with deletions larger than 100 nucleotides compared to the intraparticipant consensus were excluded.

Sequence analysis. Differences in sequences were visualized by using neighbor-joining trees (MEGA 3.1) (43) and Highlighter nucleotide transition and transversion plots (www.hiv.lanl.gov). Pairwise DNA distances were computed by using MEGA 3.1.

Conformance of intraparticipant sequence diversity to a mathematical model of random evolution was evaluated as described by Keele et al. (16) and Lee et al. (H. Y. Lee, E. E. Giorgi, B. F. Keele, B. Gaschen, G. S. Athreya, J. F. Salazar-Gonzalez, K. T. Pham, P. A. Geopfert, J. M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, B. H. Hahn, G. M. Shaw, B. T. Korber, T. Bhattacharya, and A. S. Perelson, submitted for publication) whereby exponential viral replication from a single lineage is assumed to fix mutations at a constant rate, in the absence of positive selection, using the following parameters: an HIV-1 generation time of 2 days (24), a reproductive ratio of 6 (i.e., assuming that the virus replicates exponentially, for each currently infected cell six new cells will be infected in the next generation) (42), and a replication error rate of 2.16×10^{-5} substitutions per site per replication cycle (23).

Time of divergence from the most recent common ancestor (MRCA) was estimated by using BEAST (i.e., Bayesian Evolutionary Analysis Sampling Trees, v1.4.7) (5, 6) with a relaxed (uncorrelated exponential) molecular clock and general time-reversible substitution model, with relative substitution rate parameters estimated by using HyPhy (31), as described previously (16). We used a gamma distribution with four categories and a proportion of invariant sites to model rate heterogeneity across sites. Substitution rates were unlinked across codon positions with a mean fixed at 2.16×10^{-5} substitutions per site per generation (23).

Sequences were analyzed for evidence of APOBEC3G-induced hypermutation by using the Hypermut 2.0 tool (www.hiv.lanl.gov). Sequences with a P value of ≤ 0.1 were considered enriched for mutations consistent with APOBEC3G signatures. In sequence sets showing evidence of enrichment for APOBEC3G-driven G-to-A transitions but with no single significantly hypermutated sequence, hypermutation was tested for after superimposition of all mutations within that sequence set onto a single representative sequence.

We used randomization to test whether there was evidence of clustering of mutations within 10-amino-acid stretches putatively associated with cytotoxic-T-lymphocyte (CTL) immune responses (14). Sites were classified as mutated in a given patient if there was at least one sequence from the patient differing from the patient consensus at that site. For each mutated site, we calculated its nearest-neighbor distance as the distance between the site and the closest mutated site to the left or right of it on the intrapatient sequence alignment. We then compared the number of mutations with a nearest neighbor within 10 amino acids to the number expected by chance by randomizing the locations of the mutated sites 1,000 times. A P value was calculated as the fraction of randomized datasets for which the proportion of mutated sites within 10 amino acids of another mutated site was equal to or greater than the observed proportion. The null hypothesis here is random distribution of mutated sites, which is consistent with the model of neutral drift of the infecting virus in acute infection, proposed by Keele et al. (16). Rejection of the null hypothesis suggests clustering of the mutations on a scale consistent with escape from CTL responses.

Multivariant transmission was considered if (i) within-patient *env* diversity was heterogeneous, with multimodal distribution of pairwise Hamming distances (HDs; that is, the number of differing sites between sequences) and structure within the phylogenetic trees, and (ii) if these deviations from the model of random evolution from a single founder virus could not be accounted for by APOBEC3G mutations, immune pressure, or stochastic events. For individuals infected with more than one variant, our expectation was that when multiple variants were transmitted, the sequences in the recipient should coalesce at a time predating the estimated time of infection. The number of infecting variants was enumerated, after accounting for recombination, by identification of distinct

TABLE 1. Demographic, clinical, and virological characteristics of individuals infected with a single variant compared to infection with multiple HIV-1 variants

Parameter	Individuals infected with:	
	A single variant	More than one variant
Cohort		
Total no. of subjects	54	15
No. of female subjects	32	7
No. of male subjects	22	8
No. of subjects at laboratory stage of infection:		
I/II	17	5
III/IV	14	6
V/VI	23	4
Mean age in yr (female) ± SD	29 ± 10	25 ± 7
Infecting virus		
Maximum % DNA distance range	0.08–0.64	0.32–8.49
Median no. of infecting variants (range)	1	3 (2–5)
Viral load at 12 mo postinfection[a]	39,300 ± 106,072 (400–376,000)	70,100 ± 532,413 (400–1,230,000)
CD4 counts at 12 mo postinfection[b]	468 ± 193 (303–1,030)	274 ± 251 (188–695)

[a] Expressed as median copies/ml ± the standard deviation ($n = 24$).
[b] Expressed as median cells/μl ± the standard deviation ($n = 24$).

lineages on phylogenetic trees, together with examination of Highlighter transition and transversion plots (www.hiv.lanl.gov).

N-linked glycosylation sites were identified by using the N-glycosite program (www.hiv.lanl.gov). HXB2 *env* protein amino acid locations for putative epitopes were obtained by using the HIV sequence locator tool (www.hiv.lanl.gov). Motif Scan (www.hiv.lanl.gov) was used to detect HLA anchor residue motifs within putative epitopes and to search for matching potential epitopes.

Recombination analysis. The GARD (www.datamonkey.org/GARD/) (17, 18), RAP Beta version (www.hiv.lanl.gov), and RDP version 3.27 (http://darwin.uvigo.es/rdp/rdp.html) (25) tools were used to detect recombination in intraparticipant sequence sets.

Modeling the probability of multivariant infection. We have assumed that established infection with a single genotypic variant signifies that a single virus particle was involved in the transmission event and that, in the setting of low probability of transmission, this represents the minimal infectious dose. The Poisson distribution was used to model the frequency of transmission of one, two, or more variants under the assumption that the transmission of each variant occurs with the same probability, i.e., as independent events. A left truncated Poisson model was used (since the zero events, i.e., no transmission, are not observed), and the model fitted to all of the data with a maximum-likelihood method; this then allowed the frequency of zero events to be estimated given the distribution of one, two, etc., variants being detected. A corresponding transmission probability was estimated as the sum of all probabilities for observing one or more variants.

Statistical testing. Categorical variables were compared by using Fisher exact tests (two-tailed). P values of <0.05 were considered significant.

GenBank accession numbers. The GenBank accession numbers for the 1,505 *env* sequences are FJ443128 to FJ444362.

RESULTS

Cohort description and timing of infection. Blood plasma samples were obtained from 69 study participants (30 men and 39 women) from South Africa and Malawi who had recently acquired HIV-1 infection through heterosexual contact (Table 1). A total of 42 of the participants were in the very early stages of infection: 22 were in stage I/II (viral RNA positive, p24 antigen and EIA antibody negative, or RNA and p24 antigen positive but EIA antibody negative), 6 were in stage III (EIA antibody positive but negative by Western blotting), and 14 were in stage IV (EIA antibody positive with an indeterminate Western blot result). Individuals in these stages had high viral loads (mean of >6 million copies/ml) typical of acute infection (Fig. 1). Twenty-seven participants were in the later stages of primary infection, eighteen in stage V (Western blot positive but without reactivity to the p31 integrase band) and eight in stage VI (Western blot positive with a p31 band present), with lower viral loads (medians of 84,572 and 21,176 copies/ml, respectively); one individual was classified as being in stage V or VI, with reactivity to p31 integrase not yet determined.

Characterization of *env* gene diversity. A total of 1,505 *env* gene sequences encoding gp160 (with an average of 22 se-

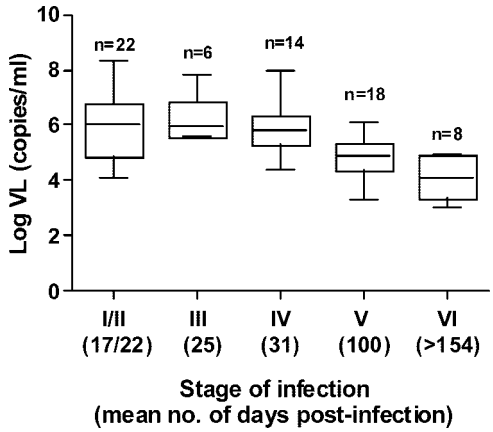
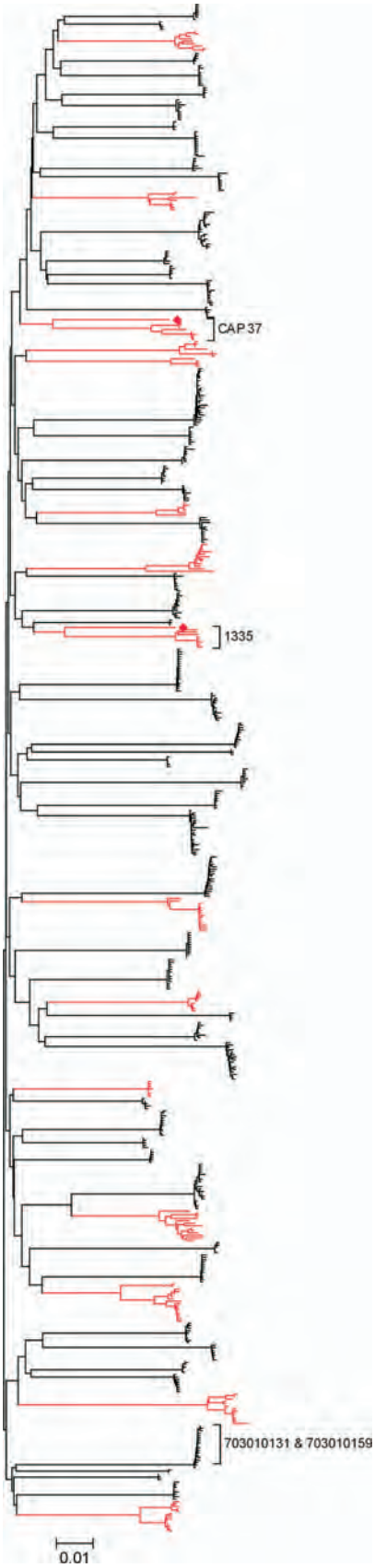


FIG. 1. Log viral loads of 68 participants categorized into stage I/II (viral RNA positive, p24 antigen and EIA antibody negative, or RNA and p24 antigen positive but EIA antibody negative), stage III (EIA antibody positive but negative by Western blot), stage IV (EIA antibody positive with an indeterminate Western blot result), stage V (Western blot positive but without reactivity to the p31 integrase band), and stage VI (Western blot positive with a p31 band present) (8). Boxes represent the 70th percentile; horizontal bars represent the median, and whisker bars correspond to minimum and maximum values. p24 antigen tests, which differentiate stages I and II of infection, were not carried out for all participants; therefore, HIV RNA-positive, enzyme-linked immunosorbent assay-negative individuals were classified as stage I/II. One participant was classified as being in stage V or VI of infection and is thus not included in this figure. The mean number of days postinfection was determined by Fiebig et al. (8) and modified by Keele et al. (16).



quences per participant; range, 15 to 42) were generated from the 69 participants using single genome amplification and direct sequencing of amplicons. All sequences were classified as HIV-1 subtype C (Fig. 2). Fourteen participants clearly harbored highly heterogeneous viral populations (maximum intrapatient DNA distances of 0.73 to 8.49%) forming more than one phylogenetic lineage. These individuals were postulated to be infected with multiple variants and were further characterized to quantify the number of infecting variants. The remaining 55 participants harbored virus populations with lower diversity (maximum intrapatient DNA distances ranging from 0.08 to 0.64%) with sequences clustering as single lineages with no, or very little, structure within phylogenetic trees (see Tables S1A to S1C in the supplemental material). On further analysis, one of the 55 individuals with a low diversity *env* population, participant 1176, was subsequently found to be infected with three closely related viruses.

All sequences were initially analyzed for evidence of APOBEC3G-induced G-to-A hypermutation and, after removal of these mutations from the relevant sequence sets, all sequences were compared to a model of random evolution in which infection with a single founder virus is followed by neutral expansion characterized by exponential population growth (16). This is expected to result in a star-like phylogeny and a Poisson distribution for the intrapatient HD. Sequence sets that did not conform to the model after the removal of the G-to-A hypermutation were analyzed for evidence of immune pressure, selection, or evidence of scattered APOBEC3G-related hypermutation, although not detected as significant ($P > 0.1$), to determine possible reasons for the deviation from the model. Finally, the time of infection estimated from laboratory staging was compared to the estimated time to the MRCA of the sequences.

Two study participants were identified as being a donor-recipient pair based on their sexual history and the close phylogenetic linkage of their viruses (Fig. 2). Participant 703010131 was assumed to be the donor since he was classified as being in stage III infection, while the assumed recipient, participant 703010159, was found to be in stage I/II.

Low *env* diversity after transmission. Of the 55 individuals harboring viruses with low diversity, sequences from 30 individuals displayed Poisson distributed HDs and a star-like phylogeny consistent with infection with a single founder virus and with the subsequent incorporation of randomly distributed mutations (Fig. 3A and see Table S1A in the supplemental material). Sequences from a further five individuals conformed to the model following the removal of G-to-A substitutions embedded in APOBEC3G signature patterns after showing sig-

FIG. 2. Neighbor-joining tree of *env* sequences from each of the 69 study participants from South Africa and Malawi. To limit the size of the tree, only 626 sequences of the 1,505 sequences generated were analyzed: where individuals harbored viruses with multiple identical *env* sequences, only the unique sequences were included. Red branches represent sequences from participants infected with more than one variant. Diamonds represent sequences with >6% divergence from the participant sequence population. The sequences for participants 703010131 and 703010159, a donor-recipient pair, cluster together, as indicated.

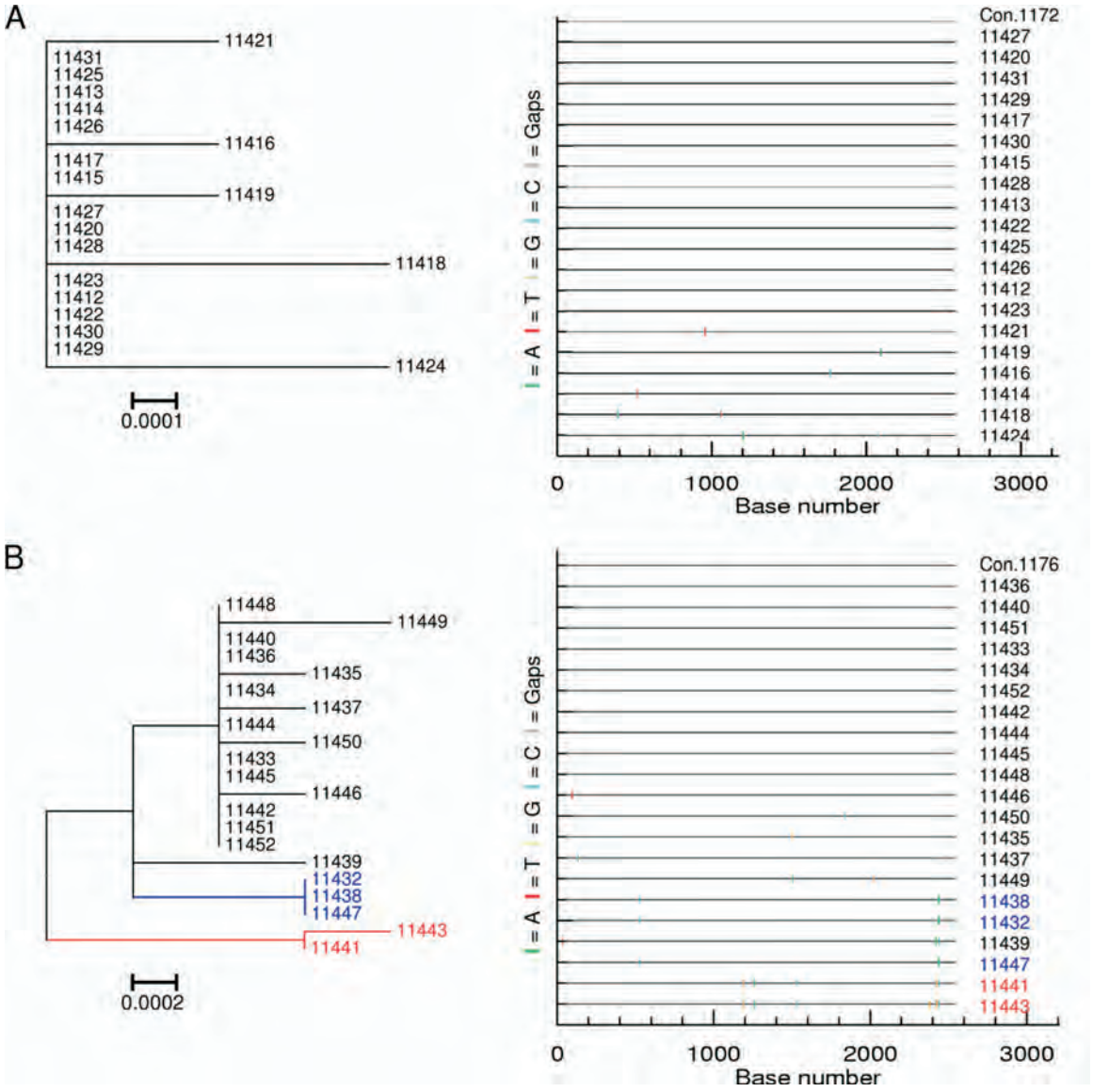


FIG. 3. *env* sequence diversity was visually determined by the structure of the phylogenetic tree (left), and the pattern of nucleotide base mutations within sequences was observed on a Highlighter plot (right). The Highlighter plots compare sequences from each participant's sequence set to an intraparticipant consensus (uppermost sequence) and illustrate the positions of nucleotide base transitions and transversions using short, color-coded bars. (A) Participant 1172 with a highly homogeneous *env* sequence population displaying limited structure on a tree and a few nucleotide changes from the intraparticipant consensus. (B) Participant 1176 infected with three closely related *env* populations (indicated in black, blue, and red, respectively) based on the clustering of sequences into individual clades on a tree and the shared patterns of mutations observed on a Highlighter plot. Both participants were viral RNA positive but ELISA negative (stage I/II of infection).

nificant enrichment of these substitutions within a single sequence and/or within the overall participant sequence set with no single overtly hypermutated sequence ($P < 0.1$) (see Table S1B in the supplemental material). In a sixth individual with significant APOBEC3G-driven hypermutation ($P = 0.0067$), participant CAP85, sequences did not conform to a star-like phylogeny even after removal of APOBEC-driven substitutions (data not shown). This was not investigated further. Sequence sets for an additional two individuals (CAP225 and 704810053) conformed to the model following removal of scattered G-to-A mutations, despite the fact that this hypermutation was not significant ($P > 0.1$) (see Table S1B in the supplemental material).

Thus, in seven of eight individuals, the model violation prior to the removal of G-to-A substitutions would likely be due to an increased relative mutation rate in these sites. This result is similar to that reported for a small set of individuals identified in the acute subtype B-infected cohort of Keele et al. (16), where an overall enrichment of G-to-A substitutions in APOBEC motifs was scattered throughout the available sequences in 7 of 81 homogeneous infections, with an additional six subjects carrying one or more overtly hypermutated sequence. The

balance between HIV-1 Vif and APOBEC3G (22) may be altered in such cases. In addition, a recent study suggests that this type of pattern may be associated with improved clinical outcome (19).

The remaining 17 of the 55 individuals with low-diversity *env* populations had no evidence of enrichment for G-to-A hypermutation but harbored virus populations that did not conform to the model due to their having either non-Poisson-distributed HDs and/or non-star-like phylogenies. Six of these individuals were classified as infected with a single founder virus based on divergence within a single lineage ($n = 2$) and/or on patterns of shared mutations between sequences ($n = 4$) (i.e., internal branching in the intrapatient tree inconsistent with the expected star-like phylogeny). These shared mutations were possibly due to stochastic accumulation of neutral mutant alleles, either as neutral mutations that occurred very early in the infection and could thus be established and retained in a high enough frequency to be sampled or due to very early CTL escape mutations.

A further 7 of these 17 individuals had evidence of early antibody pressure with sequences showing loss or gain of N-linked glycosylation sites or changes in envelope loop length; all 7 were recruited in stages IV to VI of infection (Fig. 4). In four of the seven participants, these changes occurred in the V1/V2 loop region, which may indicate that this area is an early target for immune pressure during primary infection.

Three of the seventeen individuals had evidence of cellular immune pressure. In one individual there was a mutation associated with a known epitope (CAP8, HLA-B*0801). In the other two individuals there was clustering of mutations within 10-mers (participants 70310054 and 705010015) (Fig. 4), a phenomenon previously found to be associated with CTL escape (14). The clustered mutations within these two participants' 10-mers fell within sequences found to contain HLA anchor residue motifs for a range of HLA genotypes and which were also matched to potential epitopes by Motif Scan (www.hiv.lanl.gov). However, since the HLA types for these two individuals are not known, the inference of putative CTL escape remains speculative, and the relevance of these mutations unknown.

Finally, the last individual harboring low diversity virus, subject 1176, had sequences that exhibited shared mutations between sequence subsets and an estimated time to MRCA of 88 days, which exceeded the estimated time of infection since this individual was in stage I/II infection; thus, we infer that this subject was infected with three closely related viruses (average DNA distance, 0.1%) (Fig. 3B).

In summary, of the 55 individuals with low diversity after transmission, 54 were classified as likely to be infected with a single variant, and 1 was classified as likely to be infected with three closely related variants. For participants identified as infected with a single infectious unit, the estimated time to the MRCA was consistent with the expected time of infection with one exception, participant CAP217, whose infecting virus displayed lower genetic diversity (mean, 0.02%) than expected, given this participant was in Fiebig stage IV of infection (see Table S1A in the supplemental material).

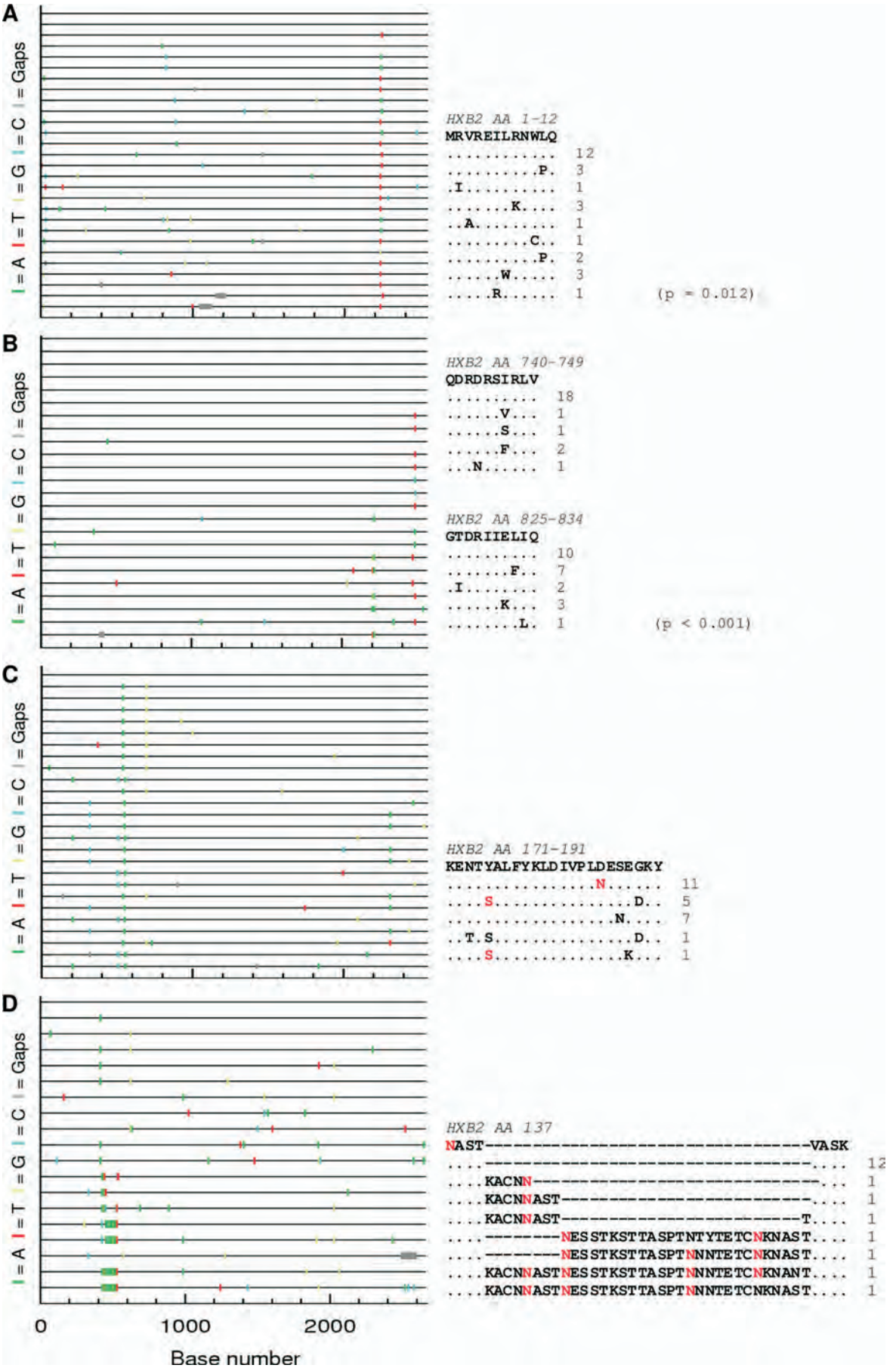
Infection with multiple unique infectious units. After correcting for recombination and hypermutation, phylogenetic analyses of the high diversity sequence sets indicated that all 14 were infected with more than one viral variant (see Table S1D in the supplemental material). Including participant 1176 infected with three closely related viruses brings the total number of individuals with multivariant infections to 15. Interlineage recombination between transmitted variants was observed in 10 of the 14 individuals, using the Recombination Analysis Program (www.hiv.lanl.gov) (Fig. 5). In 9 of these 10 cases, recombination was also detected by GARD or RDP3.27 (18, 19, 26) or both. However, since donor samples were not available for these individuals, the transmission of recombinants cannot conclusively be ruled out. In all 14 sequence sets, the estimated number of days since the MRCA significantly exceeded the period for which the associated individual could realistically have been infected (MRCA range, 605 to 5,998 days).

Therefore, including the individual infected with three closely related variants, a total of 15 individuals were inferred to be infected with multiple viruses. The median number of readily distinguishable infectious units per individual was three (range, two to five), with each viral lineage having a unique pattern of nucleotide variation (Fig. 5). This represents the minimum number of transmitted viruses in these individuals, given the limitations in sampling. While we cannot exclude the possibility that multiple variants were transmitted and a single variant grew out at the time of detection, there is no bias for the detection of multiple variants at earlier Fiebig stages (Table 1).

Two of the fourteen participants, CAP37 and 1335, were both infected with viruses with *env* sequences differing from one another by more than 6% nucleotide sequence identity (Fig. 2). Although the complete *env* sequences from these individuals clustered as an outlier to the participant sequences, the region encoding gp41 separated into distinct phylogenetic branches separated by epidemiologically unlinked sequences. This suggests that these two subjects had dual infections, although it is not possible to determine whether these individuals were infected by two independent transmission events from different donors or if variants were cotransmitted from a donor who had a dual infection.

Recombinant genomes result from the dual infection of cells, followed by recombination in a subsequent round of infection and thereafter outgrowth, allowing detection. We sought to determine whether the detection of recombination was related to the Fiebig stage. To have a large enough data set, we pooled data from Fiebig stage I to III to compare them to Fiebig stage IV to VI and also included the data reported by Keele et al. (16). We detected the presence of recombinant viruses at a significantly higher rate in the later Fiebig samples ($P = 0.0015$). In most individuals, we detected each recombinant genome once. While we can conclude from these results that the detection of recombinants becomes more likely with later stages of infection, without corresponding donor samples the possibility that detected recombinants were transmitted cannot be ruled out. In particular, this could be the case with participant CAP69 (stage I/II of infection) in whom we saw the outgrowth of some recombinants (Fig. 5B).

Clustering of mutations. We screened all sequences for evidence of the clustering of mutations. We identified tight clusters of mutations by visual inspection, as well as by means of a test based on randomizing the locations of mutated sites within each patient in order to determine whether these mutations


A
B
C
D
I = A I = T = G I = C I = Gaps
HXB2 AA 1-12
MRVREILRNWLQ
. 12
. . . . . P. 3
.I. 1
. . . . . K. . . . 3
..A. 1
. . . . . C. . 1
. . . . . P. 2
. . . . . W. 3
. . . . . R. . . . . 1 (p = 0.012)
HXB2 AA 740-749
QDRDRSIRLV
. 18
. . . . . V. . . 1
. . . . . S. . . 1
. . . . . F. . . 2
. . . N. 1
HXB2 AA 825-834
GTDRIIELIQ
. 10
. . . . . F. . 7
.I. 2
. . . . . K. . . 3
. . . . . L. 1 (p < 0.001)
HXB2 AA 171-191
KENTYALFYKLDIVPLDESEGKY
. . . . . N. 11
. . . . S. . . . . D. . 5
. . . . . N. . . . 7
..T.S. . . . . D. . 1
. . . . S. . . . . K. . . 1
HXB2 AA 137
NAST----------VASK
. . . . ---------- 12
. . . . KACNN---------- . . . 1
. . . . KACNNAST---------- 1
. . . . KACNNAST----------T. . . . 1
. . . . --------NESSTKSTTASPTNTYTETCNKNAST. . . . 1
. . . . --------NESSTKSTTASPTNNNTETCNKNAST. . . . 1
. . . . KACNNASTNESSTKSTTASPTNNNTETCNKNANT. . . . 1
. . . . KACNNASTNESSTKSTTASPTNNNTETCNKNAST. . . . 1
0
1000
2000
Base number

were significantly more clustered than would be expected by chance. We focused specifically on clustering on a length scale of 10 amino acids (roughly the size of a CTL epitope), since the presence of clusters of mutations within a region of approximately this size is consistent with selective pressure resulting in evasion of early cellular immune responses such as CTL responses (14). We determined the number of mutations with a nearest neighbor within 10 amino acids and compared this to the number expected by chance, estimated by randomizing the locations of the mutated sites 1,000 times. Under a model of neutral divergence (16), we expect mutated sites to be distributed randomly throughout the sequence. In addition, tight clustering of mutations is unlikely even if purifying selection were to affect a proportion of mutations. This would thus suggest that mutations within these clusters would be favored by selection.

Significant clustering was identified in 16 individuals (Fig. 4). Nine of these individuals with clustered mutations had single variant infections; six of these subjects (CAP129, CAP217, 0626, 703010193, 703010217, and 706010164) harbored sequences consistent with the model of neutral evolution from a single founder virus (see Table S1A in the supplemental material), suggesting that the virus populations from these individuals were under early selection despite the failure to reject the model of neutral evolution. Clustered mutations were detected in seven individuals with multivariant transmission.

As expected, most participants harboring sequences with evidence of clustered mutations were in later stages of primary infection (stages IV to VI) although, interestingly, three individuals with very early infection (stages I to III) harbored sequences with clustered mutations, providing evidence for very early selective pressure. In the absence of sequence data from donor individuals, however, it is unknown whether any of these mutations were transmitted.

Multiple variants are not transmitted as independent events with low probability. Transmission of HIV-1 is infrequent (32), and the homogeneity of sequences in acute infection suggests that most transmission events represent the minimal infectious dose. For both the subtype B (16) and the subtype C datasets we found that ca. 80% of transmissions result in infection by a single genotypic variant or a single virus particle. To determine whether transmission of multiple variants represents independent but concurrent infectious events where transmission of the first variant had no effect on the probability of transmission of the second (or third) variant, we modeled the number of transmitted variants using the Poisson distribution and estimated a putative transmission probability as the likelihood that one or more variants are transmitted (equal to 1 − the probability that no variants are transmitted). In Fig. 6 we show the expected number of infections with one, two, or more than two variants for when the probability of transmission in a single exposure is 0.1, 0.25, and 0.4. Transmission probabilities in this range are needed to observe the transmission of multiple variants at significant levels. In contrast, a more realistic transmission rate of 0.01, for example, would result in the transmission of two variants once in 10,000 transmission events.

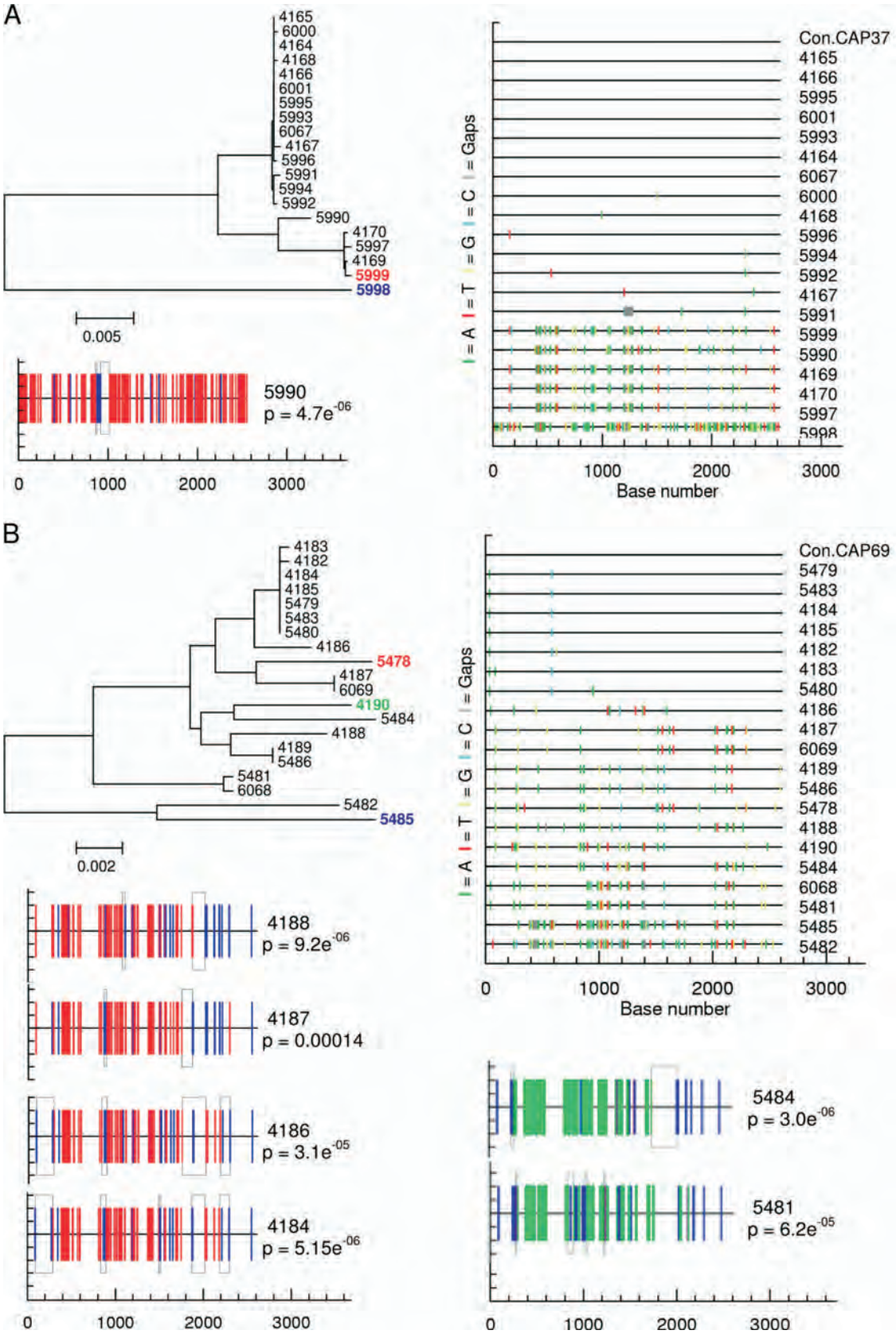
In Fig. 6 we also show the observed number of individuals infected with one, two, or more than two variants in the subtype B and subtype C primary infection cohorts. When these data are fitted to the Poisson model we estimate a transmission rate of 0.48 for the subtype C data set (Poisson mean of 0.65 with a standard error of 0.12) and a transmission rate of 0.55 for the subtype B data set (Poisson mean of 0.80 with a standard error of 0.05). These rates of transmission are unreasonably high compared to reported sexual transmission rates, which have a lower bound of ~0.001 (reviewed in Powers et al. [32]). We therefore conclude that the transmission of multiple variants does not represent low probability independent events but rather results from either transiently high transmission rates or linked transmission of multiple virions.

Multiple variant infection and disease progression. Of the 24 individuals monitored for 1 year postinfection, we found no significant difference between viral RNA load set point or $CD4^+$ T-cell counts in participants who were infected with a single variant (19 of 24) compared to those infected with multiple variants (5 of 24) ($P = 0.3198$ and $P = 0.2232$, respectively) (Table 1). However, 4 of 6 (67%) individuals with multiple variant infections had $CD4^+$ T-cell counts consistently below 350 cells/μl and were classified as rapid progressors, whereas only 4 of 20 (20%) individuals that were infected with single variants fell into this category. This association between rapid disease progression and multiple variant transmission ($P = 0.051$) supports previous studies that have shown that high diversity after infection is associated with increased rates of disease progression (10, 12, 37).

DISCUSSION

This study is the first to provide a comprehensive assessment of the multiplicity of HIV-1 subtype C infection in the context of heterosexual transmission in men and women. Our findings mirror observations of subtype B transmissions via heterosexual or homosexual routes using the same methodology (16). Together with the subtype B study, these data demonstrate that in 171 individuals a single virus is responsible for infection of 77% of individuals, with 23% of individuals infected with

FIG. 4. (A and B) Clustering of mutations within stretches of 10 amino acid residues associated with putative CTL pressure illustrated in participant 703010054 (stage V of infection) (A) and participant 705010015 (stage V of infection) (B). (C and D) Changes in the V1/V2 loop in *env* associated with putative antibody pressure illustrated in participant 704010017 (stage VI of infection) displaying clustered mutations within the V2 loop resulting in the gain of three N-linked glycosylation sites (C) and participant CAP269 displaying multiple amino acid insertions within the V1 loop resulting in the gain of one to four N-linked glycosylation sites (D). Highlighter plots (left) compare sequences from each participant's sequence set to an intraparticipant consensus (uppermost sequence) and illustrate the positions of nucleotide base transitions and transversions using short, color-coded bars. Amino acid identity alignments (right) illustrate regions of clustering of mutations in sequences aligned with an intraparticipant subtype C consensus sequence with corresponding HXB2 *env* protein locations indicated above. The number of sequences harboring a particular mutation is displayed alongside each sequence. N-linked glycosylation sites are indicated as red amino acid residues. Where significant mutational clustering was detected by a randomization test, *P* values are provided in parentheses.


A
4165
6000
4164
4168
4166
6001
5995
5993
6067
4167
5996
5991
5994
5992
5990
4170
5997
4169
5999
5998
0.005
5990
p = 4.7e-06
0
1000
2000
3000
Con.CAP37
4165
4166
5995
6001
5993
4164
6067
6000
4168
5996
5994
5992
4167
5991
5999
5990
4169
4170
5997
5998
I = A I = T I = G I = C I = Gaps
0
1000
2000
3000
Base number
B
4183
4182
4184
4185
5479
5483
5480
4186
5478
4187
6069
4190
5484
4188
4189
5486
5481
6068
5482
5485
0.002
4188
p = 9.2e-06
4187
p = 0.00014
4186
p = 3.1e-05
4184
p = 5.15e-06
0
1000
2000
3000
Con.CAP69
5479
5483
4184
4185
4182
4183
5480
4186
4187
6069
4189
5486
5478
4188
4190
5484
6068
5481
5485
5482
I = A I = T I = G I = C I = Gaps
0
1000
2000
3000
Base number
5484
p = 3.0e-06
5481
p = 6.2e-05
0
1000
2000
3000

multiple variants. Based on the frequency of the transmission of multiple variants, we find that infection with more than one variant does not occur as independent events at low probability. This implies that transmission of the second variant is linked to transmission of the first variant. Understanding the frequency and cause of multivariant transmission is relevant since individuals infected with multiple variants would require a vaccine that protects against greater initial viral diversity instead of a single homogeneous virus population. In addition, it is clinically important since high diversity following transmission has been associated with faster disease progression (10, 12, 37). In keeping with this observation, the present study found an association between multivariant transmission and disease progression.

Discrepant results due to differences in methodological approaches have hindered a clear understanding of multivariant transmission. A key advantage of our study is that we used the same methodological and analytical approaches to define the founder virus population that was used recently to study subtype B acute infection (16), thus enabling us to clearly enumerate the infecting viruses and also directly compare results. Despite different infecting subtypes and routes of transmission, the frequencies of multivariant transmission were strikingly similar: we report 22% in subtype C heterosexually infected men and women compared to the 24% of participants infected with subtype B via homosexual and heterosexual transmission reported by Keele et al. (16). Phylogenetic analysis indicated that the multiple variants came from a single donor in 87% of the cases (13 of 15 subjects), and the time to the MRCA demonstrates that the variants diverged at times significantly before the transmission event.

This estimated frequency of multivariant transmissions should, however, be considered a minimum. Many infections in highly epidemic regions have been attributed to transmissions during the acute stage of infection (30). Since this stage is generally associated with a highly homogeneous viral population, multiple variant transmissions in these instances could be missed. In addition, we may miss variants present at a low frequency (with a sample size of 20 sequenced amplicons, there is 95% confidence of detecting sequences present at frequencies greater than 15%) (16).

Although we used a model which assumes neutral evolution (16), deleterious mutations will be lost through purifying selection, and early innate and adaptive host responses are likely to impact the apparent mutation rate, especially in participants sampled after peak viremia. We did in fact identify putative immune pressure in acute infection, with a third of the sequence sets containing evidence of putative CTL pressure (based on clustered mutations) or antibody pressure (based on changes in N-glycosylation sites or variable loop length). The rates of mutation were also influenced by APOBEC3G-mediated hypermutation observed in eight individuals with single variant infections. In addition, sequences were under purifying selection with a higher rate of synonymous (dS) compared to nonsynonymous (dN) substitutions (mean dN/dS ratio of 0.79; variance, 0.44). A mean dN/dS ratio of <1 suggests that the rate of diversification of the sequences could be slightly less than the rate estimated under a strict assumption of neutrality. However, the impact of a relatively small departure from neutrality on the estimated times to the last common ancestors of intrapatient sequence sets is likely to be minor.

The rate of HIV transmission is in the range of one transmission event per 1,000 exposures (34; reviewed in Powers et al. [32]), although two studies reported rates of 31 per 1,000 exposures (26), (40) and 97 per 1,000 exposures(3). However, even rates as high as 0.03 to 0.1 cannot account for a frequency of multivariant transmission of 22 to 24%, if multiple variants are transmitted independently. This suggests that transmission of each variant is not an independent event in the context of a low transmission probability. One explanation of the frequency of multiple variant transmissions is that different cofactors transiently change the rate of multivariant transmission. The distribution of frequency of one, two, and more than two variants can be explained if two rates are incorporated: one rate would account for 70 to 75% of transmission events and have a low probability of transmission with only rare occurrences of the transmission of multiple variants; the second rate would account for 25 to 30% of transmission events. However, a probability of transmission of $\sim$0.8 would be required to result in equal numbers of transmission events of two variants and more than two variants, which would approximate the observed data. It is likely that increased transmission occurs as a result of sexually transmitted infections (38) or traumatic breaks in the epithelium. However, it is also possible that the transmission of multiple variants represents a linked event, i.e., infection by one particle (or genome) is in some way linked to an increased probability of infection with a second particle (or genome). If the infectious unit were an infected $CD4^+$ T cell, which can be infected with multiple viruses (15), this could account for at least some of the multivariant transmission events. A recent report has shown the potential for infected cells to penetrate a disturbed epithelium (45), and the apparent need for infection via infected cells in the case of HTLV-1 provides further support for a cell-mediated mechanism of transmission (33). Alternatively, virus particles could be aggregated by biological molecules such as SEVI (for semen-derived enhancer of virus infection) (27) or tetherin during budding (28), potentiating infection with two particles in a single, rare transmission event.

A previous study from Kenya showed women were generally

FIG. 5. Highly heterogeneous, multiple variant *env* sequence populations are visually represented by phylogenetic trees with extensive branch structure and discernible clades (left) and Highlighter plots with diverse patterns of nucleotide base mutations compared to the intraparticipant consensus (right). *env* variants resulting from recombination between clades are displayed below with RAP (for recombination analysis program) plots. Parental strains for each recombinant are color coded on the trees, with regions within each recombinant likewise color coded to correspond to respective parental strains. Recombination breakpoints are illustrated by empty boxes on RAP plots. (A) Participant CAP37 was infected with three distinct variants. Sequence 5998 differs by up to 6% from the remaining sequences from this participant and is a suspected dual infection. (B) Participant CAP69 was infected with at least five distinct viruses with extensive recombination between clades. Six recombinant strains are illustrated here.

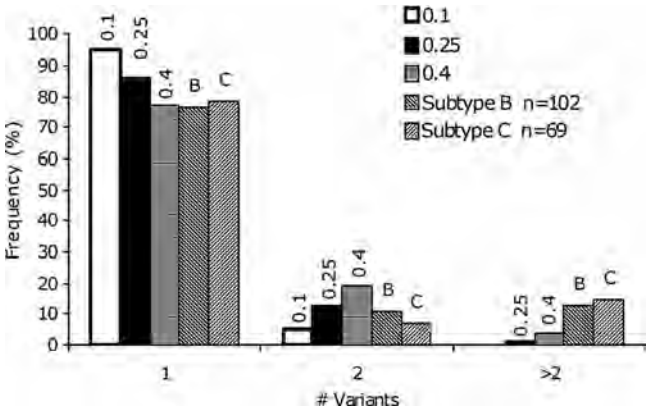


FIG. 6. Model of the rates of transmission of multiple variants using a Poisson distribution. Transmission rates, shown above each bar, are modeled to account for differences in the frequency of transmissions of one, two, or more than two variants using transmission probabilities of 0.1 (Poisson mean, 0.1), 0.25 (Poisson mean, 0.5), and 0.4 (Poisson mean, 0.5). Transmission probability in this setting is defined as the sum of the probabilities of all nonzero events in the Poisson distribution. The frequency of transmission of one, two, or more than two variants is shown for the subtype C cohort described in the present study and from the subtype B cohort described in Keele et al. (16) and indicated by a "C" or "B" above each bar. In modeling the Poisson distribution to fit the cohort data all data were used as counts, not percentages, and values greater than two variants were not pooled but rather modeled in total.

infected with more heterogeneous virus populations compared to men (21). Although in our study, we did not find that females were infected with higher-diversity viral populations compared to men (data not shown), there were differences in frequency of sexually transmitted infections, with genital ulcerative disease being much more common in men from our study than in women (82% versus 13%). Thus, since genital ulcerative disease has an impact on transmission, this could confound our analysis. In addition, uncircumcised men are more susceptible to infection (1, 2); thus, this difference in results between studies may be due to the fact that most of the men in our study were from Malawi where there is a very low frequency of circumcision, whereas the Kenyan study recruited from a cohort of individuals where 87% were circumcised (34).

In conclusion, infection with a single virus in the majority of individuals demonstrates the severity of the genetic bottleneck at transmission. These data in conjunction with the subtype B analysis suggests a universal observation that mucosal HIV-1 infection most frequently originates from a single infectious unit. Less frequently, multiple viral variants are transmitted, which not only increases the genetic diversity, but this increased diversity also provides the virus with greater opportunity to escape early selective pressure through recombination. Although the biological basis for the transmission of multiple variants remains unknown, possible explanations include transiently high rates of transmission due to cofactors, transmission via a multiply infected cell, or transmission of viral aggregates. Since one in five individuals will become infected with multiple infectious variants, it is important to translate how this information impacts on the breadth and targeting needed for protective vaccination.

ACKNOWLEDGMENTS

This study was funded by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, and the U.S. Department of Health and Human Services (AI51794, CAPRISA; DK49381 [M.S.C.], CHAVI), as well as by the National Research Foundation (no. 67385) (South Africa), the South African AIDS Vaccine Initiative, and amFAR grant 106997-43.

We thank the clinical staff and participants from the CAPRISA, CHAVI, and Malawi STI cohorts; Darren Marten for critical comments; and Leslie Arney for assistance with the graphics. We also thank the clinical staff from the CHAVI Lilongwe cohorts, including Francis Martinson, Gift Kamanga, Happiness Kanyamula, and Deborah Kamwendo, for their support.

REFERENCES

1. **Auvert, B., D. Taljaard, E. Lagarde, J. Sobngwi-Tambekou, R. Sitta, and A. Puren.** 2005. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial. PLoS Med. **2:**e298.
2. **Bailey, R. C., S. Moses, C. B. Parker, K. Agot, I. Maclean, J. N. Krieger, C. F. Williams, R. T. Campbell, and J. O. Ndinya-Achola.** 2007. Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. Lancet **369:**643–656.
3. **Cameron, D. W., J. N. Simonsen, L. J. D'Costa, A. R. Ronald, G. M. Maitha, M. N. Gakinya, M. Cheang, J. O. Ndinya-Achola, P. Piot, and R. C. Brunham.** 1989. Female to male transmission of human immunodeficiency virus type 1: risk factors for seroconversion in men. Lancet **ii:**403–407.
4. **Derdeyn, C. A., J. M. Decker, F. Bibollet-Ruche, J. L. Mokili, M. Muldoon, S. A. Denham, M. L. Heil, F. Kasolo, R. Musonda, B. H. Hahn, G. M. Shaw, B. T. Korber, S. Allen, and E. Hunter.** 2004. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. Science **303:**2019–2022.
5. **Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut.** 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. **4:**e88.
6. **Drummond, A. J., and A. Rambaut.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. **7:**214.
7. **Edmonson, P. F., and J. I. Mullins.** 1992. Efficient amplification of HIV half-genomes from tissue DNA. Nucleic Acids Res. **20:**4933.
8. **Fiebig, E. W., D. J. Wright, B. D. Rawal, P. E. Garrett, R. T. Schumacher, L. Peddada, C. Heldebrant, R. Smith, A. Conrad, S. H. Kleinman, and M. P. Busch.** 2003. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. AIDS **17:**1871–1879.
9. **Gottlieb, G. S., L. Heath, D. C. Nickle, K. G. Wong, S. E. Leach, B. Jacobs, S. Gezahegne, A. B. van't Wout, L. P. Jacobson, J. B. Margolick, and J. I. Mullins.** 2008. HIV-1 variation before seroconversion in men who have sex with men: analysis of acute/early HIV infection in the multicenter AIDS cohort study. J. Infect. Dis. **197:**1011–1015.
10. **Gottlieb, G. S., D. C. Nickle, M. A. Jensen, K. G. Wong, J. Grobler, F. Li, S. L. Liu, C. Rademeyer, G. H. Learn, S. S. Karim, C. Williamson, L. Corey, J. B. Margolick, and J. I. Mullins.** 2004. Dual HIV-1 infection associated with rapid disease progression. Lancet **363:**619–622.
11. **Greenier, J. L., C. J. Miller, D. Lu, P. J. Dailey, F. X. Lu, K. J. Kunstman, S. M. Wolinsky, and M. L. Marthas.** 2001. Route of simian immunodeficiency virus inoculation determines the complexity but not the identity of viral variant populations that infect rhesus macaques. J. Virol. **75:**3753–3765.
12. **Grobler, J., C. M. Gray, C. Rademeyer, C. Seoighe, G. Ramjee, S. A. Karim, L. Morris, and C. Williamson.** 2004. Incidence of HIV-1 dual infection and its association with increased viral load set point in a cohort of HIV-1 subtype C-infected female sex workers. J. Infect. Dis. **190:**1355–1359.
13. **Haaland, R. E., P. A. Harkins, J. F. Salazar-Gonzalez, A. Johnson, A. Tichacek, E. Karita, O. Manigart, J. Mulenga, B. F. Keele, G. M. Shaw, B. H. Hahn, S. Allen, C. A. Derdeyn, and E. Hunter.** 2009. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. PloS Pathog. **5:**e1000274.
14. **Jones, N. A., X. Wei, DR. Flower, M. Wong, F. Michor, M. S. Saag, B. H. Hahn, M. A. Nowak, G. M. Shaw, and P. Borrow.** 2004. Determinations of human immunodeficiency virus type 1 escape from the primary $CD8^+$ cytotoxic T lymphocyte response. J. Exp. Med. **200:**1243–1256.
15. **Jung, A., R. Maier, J. P. Vartanian, G. Bocharov, V. Jung, U. Fischer, E. Meese, S. Wain-Hobson, and A. Meyerhans.** 2002. Multiply infected spleen cells in HIV patients. Nature **418:**144.
16. **Keele, B. F., E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham, M. G. Salazar, C. Sun, T. Grayson, S. Wang, H. Li, X. Wei, C. Jiang, J. L. Kirchherr, F. Gao, J. A. Anderson, L. H. Ping, R. Swanstrom, G. D. Tomaras, W. A. Blattner, P. A. Goepfert, J. M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, M. S. Cohen, D. C. Montefiori, B. F. Haynes, B. Gaschen, G. S. Athreya, H. Y. Lee, N. Wood, C. Seoighe, A. S. Perelson, T. Bhattacharya, B. T. Korber, B. H. Hahn, and G. M. Shaw.** 2008. Identification and char-

acterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc. Natl. Acad. Sci. USA **105:**7552–7557.
17. **Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. Frost.** 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. **23:**1891–1901.
18. **Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. Frost.** 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics **22:**3096–3098.
19. **Land, A. M., T. B. Ball, M. Luo, R. Pilon, P. Sandstrom, J. E. Embree, C. Wachihi, J. Kimani, and F. A. Plummer.** 2008. Human immunodeficiency virus (HIV) type 1 proviral hypermutation correlates with CD4 count in HIV-infected women from Kenya. J. Virol. **82:**8172–8182.
20. Reference deleted.
21. **Long, E. M., H. L. Martin, Jr., J. K. Kreiss, S. M. Rainwater, L. Lavreys, D. J. Jackson, J. Rakwar, K. Mandaliya, and J. Overbaugh.** 2000. Gender differences in HIV-1 diversity at time of infection. Nat. Med. **6:**71–75.
22. **Malim, M. H.** 2006. Natural resistance to HIV infection: the Vif-APOBEC interaction. C. R. Biol. **329:**871–875.
23. **Mansky, L. M., and H. M. Temin.** 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:**5087–5094.
24. **Markowitz, M., M. Louie, A. Hurley, E. Sun, M. M. Di, A. S. Perelson, and D. D. Ho.** 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:**5037–5038.
25. **Martin, D. P., C. Williamson, and D. Posada.** 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics **21:**260–262.
26. **Mastro, T. D., G. A. Satten, T. Nopkesorn, S. Sangkharomya, and I. M. Longini, Jr.** 1994. Probability of female-to-male transmission of HIV-1 in Thailand. Lancet **343:**204–207.
27. **Munch, J., E. Rucker, L. Standker, K. Adermann, C. Goffinet, M. Schindler, S. Wildum, R. Chinnadurai, D. Rajan, A. Specht, G. Gimenez-Gallego, P. C. Sanchez, D. M. Fowler, A. Koulov, J. W. Kelly, W. Mothes, J. C. Grivel, L. Margolis, O. T. Keppler, W. G. Forssmann, and F. Kirchhoff.** 2007. Semen-derived amyloid fibrils drastically enhance HIV infection. Cell **131:**1059–1071.
28. **Neil, S. J., T. Zang, and P. D. Bieniasz.** 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature **451:**425–430.
29. **Palmer, S., M. Kearney, F. Maldarelli, E. K. Halvas, C. J. Bixby, H. Bazmi, D. Rock, J. Falloon, R. T. Davey, Jr., R. L. Dewar, J. A. Metcalf, S. Hammer, J. W. Mellors, and J. M. Coffin.** 2005. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. J. Clin. Microbiol. **43:**406–413.
30. **Pilcher, C. D., H. C. Tien, J. J. Eron, Jr., P. L. Vernazza, S. Y. Leu, P. W. Stewart, L. E. Goh, and M. S. Cohen.** 2004. Brief but efficient: acute HIV infection and the sexual transmission of HIV. J. Infect. Dis. **189:**1785–1792.
31. **Pond, S. L., S. D. Frost, and S. V. Muse.** 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics **21:**676–679.
32. **Powers, K. A., C. Poole, A. E. Pettifor, and M. S. Cohen.** 2008. Rethinking the heterosexual infectivity of HIV-1: a systematic review and meta-analysis. Lancet Infect. Dis. **8:**553–563.
33. **Proietti, F. A., A. B. Carneiro-Proietti, B. C. Catalan-Soares, and E. L. Murphy.** 2005. Global epidemiology of HTLV-1 infection and associated diseases. Oncogene **24:**6058–6068.
34. **Rakwar, J., L. Lavreys, M. L. Thompson, D. Jackson, J. Bwayo, S. Hassanali, K. Mandaliya, J. Ndinya-Achola, and J. Kreiss.** 1999. Cofactors for the acquisition of HIV-1 among heterosexual men: prospective cohort study of trucking company workers in Kenya. AIDS **13:**607–614.
35. **Ritola, K., C. D. Pilcher, S. A. Fiscus, N. G. Hoffman, J. A. Nelson, K. M. Kitrinos, C. B. Hicks, J. J. Eron, Jr., and R. Swanstrom.** 2004. Multiple V1/V2 env variants are frequently present during primary infection with human immunodeficiency virus type 1. J. Virol. **78:**11208–11218.
36. **Sagar, M., E. Kirkegaard, E. M. Long, C. Celum, S. Buchbinder, E. S. Daar, and J. Overbaugh.** 2004. Human immunodeficiency virus type 1 (HIV-1) diversity at time of infection is not restricted to certain risk groups or specific HIV-1 subtypes. J. Virol. **78:**7279–7283.
37. **Sagar, M., L. Lavreys, J. M. Baeten, B. A. Richardson, K. Mandaliya, B. H. Chohan, J. K. Kreiss, and J. Overbaugh.** 2003. Infection with multiple human immunodeficiency virus type 1 variants is associated with faster disease progression. J. Virol. **77:**12921–12926.
38. **Sagar, M., L. Lavreys, J. M. Baeten, B. A. Richardson, K. Mandaliya, J. O. Ndinya-Achola, J. K. Kreiss, and J. Overbaugh.** 2004. Identification of modifiable factors that affect the genetic diversity of the transmitted HIV-1 population. AIDS **18:**615–619.
39. **Salazar-Gonzalez, J. F., E. Bailes, K. T. Pham, M. G. Salazar, M. B. Guffey, B. F. Keele, C. A. Derdeyn, P. Farmer, E. Hunter, S. Allen, O. Manigart, J. Mulenga, J. A. Anderson, R. Swanstrom, B. F. Haynes, G. S. Athreya, B. T. Korber, P. M. Sharp, G. M. Shaw, and B. H. Hahn.** 2008. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J. Virol. **82:**3952–3970.
40. **Satten, G. A., T. D. Mastro, and I. M. Longini, Jr.** 1994. Modelling the female-to-male per-act HIV transmission probability in an emerging epidemic in Asia. Stat. Med. **13:**2097–2106.
41. **Simmonds, P., P. Balfe, C. A. Ludlam, J. O. Bishop, and A. J. Brown.** 1990. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. J. Virol. **64:**5840–5850.
42. **Stafford, M. A., L. Corey, Y. Cao, E. S. Daar, D. D. Ho, and A. S. Perelson.** 2000. Modeling plasma virus concentration during primary HIV infection. J. Theor. Biol. **203:**285–301.
43. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596–1599.
44. **van Loggernberg, F., K. Mlisana, C. Williamson, S. C. Auld, L. Morris, C. M. Gray, K. Q. Abdool, A. Grobler, N. Barnabas, I. Iriogbe, and S. S. Abdool Karim.** 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:**e1954.
45. **Weiler, A. M., Q. Li, L. Duan, M. Kaizu, K. L. Weisgrau, T. C. Friedrich, M. R. Reynolds, A. T. Haase, and E. G. Rakasz.** 2008. Genital ulcers facilitate rapid viral entry and dissemination following intravaginal inoculation with cell-associated simian immunodeficiency virus SIVmac239. J. Virol. **82:**4154–4158.
46. **Wolfs, T. F., G. Zwart, M. Bakker, and J. Goudsmit.** 1992. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. Virology **189:**103–110.
47. **Wolinsky, S. M., C. M. Wike, B. T. Korber, C. Hutto, W. P. Parks, L. L. Rosenblum, K. J. Kunstman, M. R. Furtado, and J. L. Munoz.** 1992. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. Science **255:**1134–1137.
48. **Zhu, T., H. Mo, N. Wang, D. S. Nam, Y. Cao, R. A. Koup, and D. D. Ho.** 1993. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. Science **261:**1179–1181.

Appendix 2: Published manuscript Abrahams et al., 2013 *AIDS*

Rapid, complex adaptation of transmitted HIV-1 full-length genomes in subtype C-infected individuals with differing disease progression

Melissa-Rose Abrahams[a], Florette K. Treurnicht[a], Nobubelo K. Ngandu[a], Sarah A. Goodier[a], Jinny C. Marais[a], Helba Bredell[a], Ruwayhida Thebus[a], Debra de Assis Rosa[b], Koleka Mlisana[c,d], Cathal Seoighe[e], Salim Abdool Karim[c], Clive M. Gray[b] and Carolyn Williamson[a]

Objective(s): There is limited information on full-length genome sequences and the early evolution of transmitted HIV-1 subtype C viruses, which constitute the majority of viruses spread in Africa. The purpose of this study was to characterize the earliest changes across the genome of subtype C viruses following transmission, to better understand early control of viremia.

Design: We derived the near full-length genome sequence responsible for clinical infection from five HIV subtype C-infected individuals with different disease progression profiles and tracked adaptation to immune responses in the first 6 months of infection.

Methods: Near full-length genomes were generated by single genome amplification and direct sequencing. Sequences were analyzed for amino acid mutations associated with cytotoxic T lymphocyte (CTL) or antibody-mediated immune pressure, and for reversion.

Results: Fifty-five sequence changes associated with adaptation to the new host were identified, with 38% attributed to CTL pressure, 35% to antibody pressure, 16% to reversions and the remainder were unclassified. Mutations in CTL epitopes were most frequent in the first 5 weeks of infection, with the frequency declining over time with the decline in viral load. CTL escape predominantly occurred in *nef*, followed by *pol* and *env*. Shuffling/toggling of mutations was identified in 81% of CTL epitopes, with only 7% reaching fixation within the 6-month period.

Conclusion: There was rapid virus adaptation following transmission, predominantly driven by CTL pressure, with most changes occurring during high viremia. Rapid escape and complex escape pathways provide further challenges for vaccine protection.

AIDS 2013, **27**:507–518

Keywords: acute infection, Africa, cytotoxic T-lymphocytes, genome, HIV-1, progression

[a]Division of Medical Virology, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, [b]AIDS Research Unit: Immunology, National Institute of Communicable Diseases, Johannesburg, [c]Centre for the AIDS Program of Research in South Africa, [d]Department of Medical Microbiology, University of KwaZulu Natal, Durban, South Africa, and [e]School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway, Ireland.

Correspondence to Carolyn Williamson, Division of Medical Virology, Institute of Infectious Diseases and Molecular Medicine, Faculty of Health Sciences, University of Cape Town and National Health Laboratory Services, Observatory 7925, South Africa.
Tel: +27 21 406 6683; fax: +27 21 406 6682; e-mail: Carolyn.Williamson@uct.ac.za
Received: 17 July 2012; revised: 5 November 2012; accepted: 15 November 2012.

DOI:10.1097/QAD.0b013e32835cab64

Introduction

Early host selective pressures drive genetic diversification of the transmitted HIV and potentially influence the course of disease. In heterosexual infection, it is estimated that approximately 80% of individuals are infected with a single virus or virus-infected cell [1,2]. Vaccines that do not block the establishment of this initial infecting virus will need to target the early diversifying virus and thus an understanding of early viral evolution is important. The frequency and speed at which the transmitted virus changes in response to host immune pressures are of key interest to vaccine immunogen design as these provide insights into its strengths and vulnerabilities.

Recent methodologies using single genome amplification (SGA) applied to individuals with acute HIV-1 infection have enabled identification of the sequence of transmitted/founder (t/f) full-length viral genomes [1,3,4]. This approach uses mathematical modeling to derive the sequence of the virus(es) responsible for productive clinical infection, and was proven in the simian immunodeficiency virus (SIV) model wherein the derived t/f was identical, or differed by a few nucleotides, to the virus in the inocula [5].

Using this approach, early evolutionary changes following transmission in three subtype B infections from the USA have been mapped [3]. No longitudinal studies elucidating early evolution in full-length subtype C genomes have been reported despite subtype C being the dominant subtype both globally and in southern Africa where large vaccine and microbicide trials take place [6,7]. Studies in different population settings are essential given that differences in host genetics influence viral evolution [8].

The transmitted virus encounters immune selective pressures almost immediately following infection. Cytotoxic T lymphocyte (CTL) and neutralizing antibody pressure are the proven driving forces for viral diversification [9–16]. Evidence of CTL pressure on the viral genome has been identified in the very first weeks of subtype B and C infection [1–3,17–23], and CTL activity has been associated with control of viremia early in infection [24–26]. Most recently, a subtype B full-length genome study using mathematical modeling to determine the killing rate of CTL during acute viremia suggested that these cells have a role in controlling peak viremia [21].

In this study, we investigated changes observed across the genomes of t/f subtype C viruses from five heterosexually infected women with differing disease progression profiles. We extrapolated the near full-length genome sequence of the t/f viruses and quantified genetic mutations associated with positive selection from humoral and cellular immune pressures over the first 6 months following infection.

Methods

Study participants

Samples were obtained from the Centre for the AIDS Program of Research in South Africa (CAPRISA) 002 Acute Infection cohort (Durban, South Africa) [27]. Date of infection was estimated as the midpoint between last negative and first positive HIV antibody test and as 14 days prior for individuals who were RNA positive/antibody negative. Human leukocyte antigen (HLA) A, B and C types were determined using four-digit high-resolution HLA typing as described [28]. The study was approved by the Universities of Cape Town, Witwatersrand and KwaZulu Natal.

PCR amplification and sequencing

RNA was extracted from 200–400 μl plasma using the Qiagen Viral RNA Mini Kit (Qiagen, Valencia, California, USA) and reverse transcribed to cDNA using superscript III reverse transcriptase and Oligo$(dT)_{20}$ (Invitrogen, GmbH, Karlsruhe, Germany). SGA [4] and sequencing of near full-length genome amplicons was done using Expand Long Template Taq (Roche Diagnostics, Rotkreuz, Switzerland) as reported [4] with primers described by Rousseau *et al.* [29] optimized for subtype C. Sequences less than 6000 bp were excluded. Salazar-Gonzalez *et al.* [3] attributed one to five ambiguities within a genome obtained at less than 20% PCR positivity predominantly to PCR Taq error. We accepted amplicons obtained at less than 66% positivity with up to six ambiguities. Targeted epitope sequencing was done following gene-specific limiting dilution PCR. *Gag* and *nef* clones were generated using limiting dilution PCR and the pGEM-T Easy system as described [30]. *Env* SGA was described previously [4]. All products were directly sequenced using the ABI 3000 genetic analyzer (Applied Biosystems, Foster City, California, USA) and BigDye terminator reagents.

Sequence analysis

Analyses performed were sequence alignments, amino acid identity and frequency plots and consensus sequence derivation (BioEdit version 7.0.8.0 [31]); subtyping (REGA HIV Subtyping Tool; http://dbpartners.stanford.edu/RegaSubtyping/); phylogenetic and pairwise DNA distance analyses (Mega 4 [32]); Highlighter plots (http://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter); CTL epitope prediction (Epitope Location Finder (ELF) (http://www.hiv.lanl.gov/content/sequence/ELF/epitope_analyzer) and NetMHCpan 2.2 (http://www.cbs.dtu.dk/services/NetMHCpan)[33]); Hypermut detection of APOBEC hypermutation (http://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermnut.html); Shannon entropies (Entropy One; http://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy_one) and mapping of functionally/structurally relevant genome regions (http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark). A

high average Shannon entropy score was taken as more than 0.25 as described by Bansal *et al.* [34] for variable HIV proteins. Where an optimal epitope (9–11-mer) has not yet been described, the entropy of an 18-mer peptide encompassing the mutating region was used.

Time to most recent common ancestor (MRCA) was determined using Bayesian Evolutionary Analysis Sampling Trees (BEAST) v1.4.7 [35,36] as described previously [1,2]. A relaxed (uncorrelated exponential) molecular clock with general time-reversible substitution model, mean of 2.16×10^{-5} substitutions per site per generation with rates unlinked across codon sites [37] and gamma distribution with four categories and a proportion of invariant sites was used.

Known HLA class I restricted epitopes or class I HLA-associated polymorphisms were identified using the Los Alamos HIV Molecular Immunology 2008 Compendium (http://www.hiv.lanl.gov/content/immunology/compendium.html) and Matthews *et al.* [38]. Subtype C database alignments were obtained from the Los Alamos HIV database (http://www.hiv.lanl.gov/components/sequence/HIV/).

Positive selection and statistical analysis

Nonsynonymous to synonymous substitution (dN/dS) rate ratios per codon site were estimated using the MG94xHKY85 codon model [39]. We allowed dS to vary across codon sites and employed the Dual model which takes into account that dS may vary independently of dN [40]. Models were implemented within HyPhy [41] and ensured that correct phylogenetic relationships were used for regions separated by recombination breakpoints [42].

Categorical statistical tests were carried out using the Fisher's exact two-tailed test (http://www.graphpad.com/quickcalcs).

Results

This study characterized virus evolution in five subtype C-infected women recruited 2–5 weeks following transmission. These women were selected based on infection with a single transmitted/founder virus [2] and clinical disease progression profile (Table 1). One woman was classified as a viremic controller (CAP45; viral loads consistently <2000 copies/ml and CD4 cell counts >350 cells/μl in the absence of antiretroviral therapy) [43], two as rapid progressors (CAP63 and CAP210; viral loads >100 000 copies/ml and CD4 cell counts <350 cells/μl on consecutive visits within the first year of infection) [23,44], and two as intermediate progressors (CAP85 and CAP239; not fitting either controller or rapid status) (Fig. 1). A total of 112 near full-length genomes were generated, with an average of nine at screening per enrolment (2–5 weeks after infection), six at 3 months (11–13 weeks after infection) and nine at 6 months (22–29 weeks after infection). No sequences could be generated at 6 months for viremic controller CAP45 due to low viral loads. Additional sequences (half-genome, SGA and clonal) were generated from various time points ranging from 2 to 117 weeks after infection (Table 1).

Derivation of transmitted/founder virus sequences

All sequences were classified as subtype C along the entire length of the genome. Mean intraparticipant DNA distances ranged from 0.008 to 0.25% (median 0.03%) at the first time point and mean number of days since MRCA ranged from 18 to 53 days (Table 1), indicating limited sequence diversification since transmission (see Fig. S1, Supplemental Digital Content 1, http://links.lww.com/QAD/A282, illustrating intraparticipant sequence diversity in a Neighbour-Joining tree). The t/f sequences were defined as the consensus of sequences from the earliest time point where all genes had an intact open reading frame and no ambiguities [3]. Although CAP63 and CAP85 were classified as infected with a single t/f variant based on *env* [2], we identified a minor early variant in each (not detected at later time points) which, in both cases, differed from the derived t/f at five nucleotide positions, suggesting that these individuals may each have been infected with two very closely related variants.

The majority of early genetic changes are due to cytotoxic T lymphocyte pressure or reversion

Using longitudinal near full-length genome and *env* screening/enrolment SGA sequences, viral evolution from the t/f was analyzed for evidence of immune escape or reversion. Substitutions from high-frequency/consensus to lower frequency/nonconsensus amino acids within or adjacent to known class I HLA-restricted epitopes, or corresponding to known HLA-associated polymorphisms, were classified as CTL pressure [16,19,45,46]. Mutations within the hypervariable loops and potential N-linked glycosylation sites (PNGSs) in *env* were classified as antibody pressure [9]. Mutations from low/nonconsensus to higher frequency/consensus subtype C amino acids within CTL epitopes not restricted by the participants HLA were classified as reversion of transmitted CTL escape mutations [16,45,46]. In addition, clustered mutations within amino acid 9-mers, previously reported to be associated with immune selection [18], or single amino acid mutations persisting to fixation and corresponding to sites under positive selection were identified as putative immune escape.

In viruses from the five women, immune pressure was identified in 55 genome regions (see Figs S2–S6, Supplemental Digital Content 2, http://links.lww.com/QAD/A282, which illustrate genome regions

Table 1. Demographic and sequencing information for five CAPRISA participants.

Participant ID	Age	Disease progression	Fiebig stage[a] at first sequenced time point	BEAST mean no. of days since MRCA (95% CI)[b]	HLA type	Sample date	Weeks postinfection	Whole (half) genome sequences	Subgenomic clone/SGA sequences
CAP45	41	Slow	I/II	18 (4–35)	A*23:01, 29:02	20 April 2005	2	3	16
					B*15:10, 45:01	11 May 2005	5	6	1
					Cw*06:02, 16:01	7 June 2005	9	3	8
						28 June 2005	12	1	7
						27 July 2005	16	3	
						4 April 2006	52		
						4 July 2006	65		
CAP63	32	Rapid	III	30 (10–53)	A*02:01, 23:01	6 January 2005	2	11	19
					B*45:01	26 January 2005	5	7	
					Cw*04:01, 16:01	9 March 2005	11	10	26
						13 July 2005	29	5	
						7 September 2005	37		
CAP85	24	Intermediate	V	53 (20–101)	A*30:02	22 June 2005	5	8	21
					B*08:01, 45:01	18 August 2013	13	9	
					Cw*07:01, 16:01	7 December 2005	29	7	11
						16 February 2006	39		1
						10 May 2006	51	2	9
						7 June 2006	55		9
						7 December 2006	81		1
						6 June 2007	107		7
CAP210	43	Rapid	I/II	11 (3–25)	A*68:02	3 May 2005	2	9	21
					B*15:10	13 June 2005	12	7	
					Cw*03:04	21 September 2005	22		8
						19 October 2005	26	11	3
						23 November 2005	31		3
CAP239	44	Intermediate	V[c]	34 (11–58)	A*01:23, 29:02	19 July 2005	2	2 (3)	21
					B*42:01, 58:01	10 August 2005	5	8	75
					Cw*06:02, 17:01	17 August 2005	6		19
						21 September 2005	11	2	
						7 December 2005	22	9	9
						14 June 2006	49	0 (4)	
						11 January 2007	79		9
						4 October 2007	117	3	

BEAST, Bayesian Evolutionary Analysis Sampling Trees; HLA, human leukocyte antigen; MRCA, most recent common ancestor; SGA, single genome amplification.
[a]Fiebig stage I/II [63] are HIV RNA positive but antibody negative, III is ELISA positive but nonreactive western blot and V is reactive western blot without p31 band.
[b]Bayesian Evolutionary Analysis Sampling Trees estimated period of infection to observed *env* diversity from a single infecting strain.
[c]Fiebig stage determined for 5 weeks postinfection

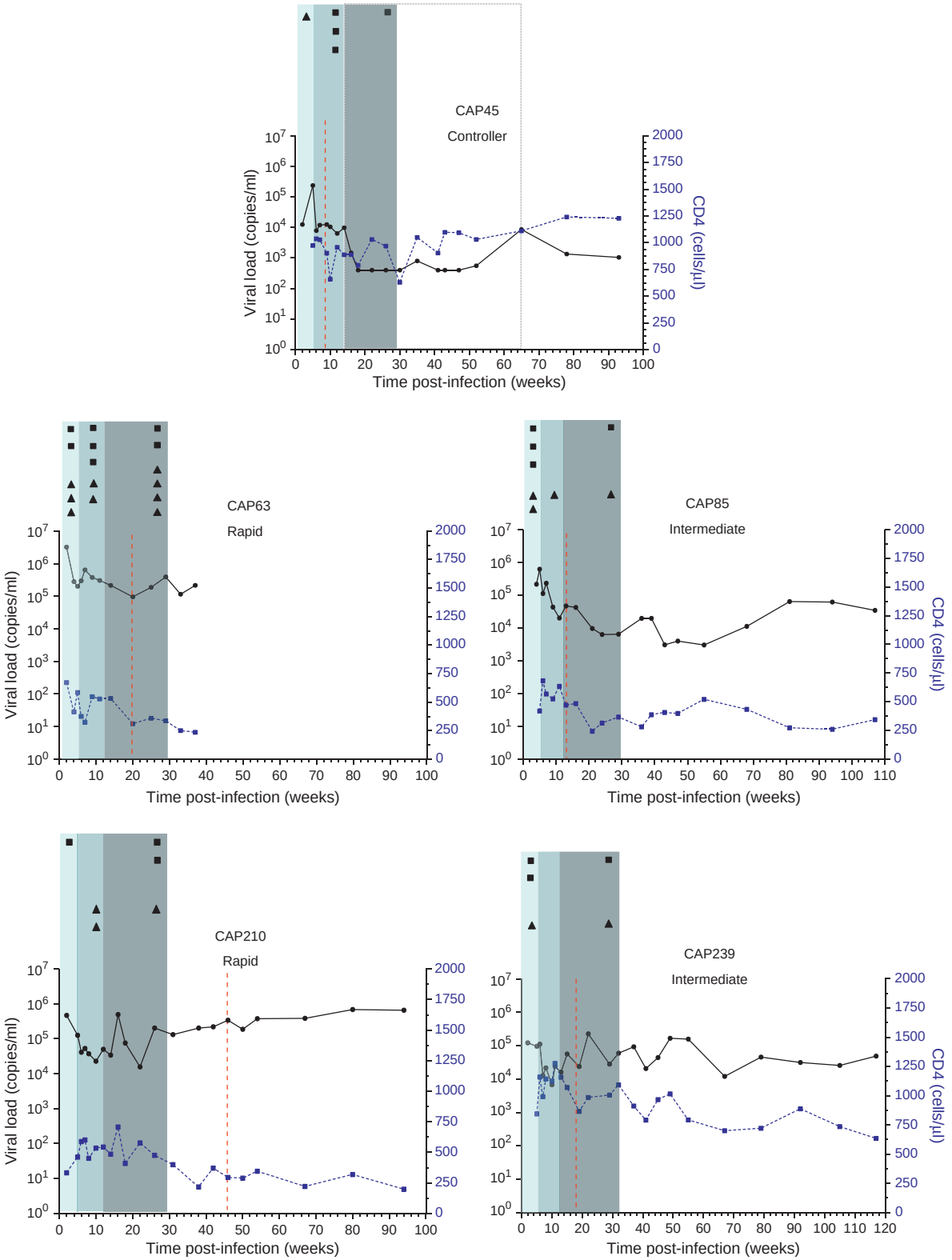


Fig. 1. Clinical and immune selection profiles for each of the five women illustrated with viral load and CD4 cell count graphs. The number of epitopes/genome regions identified as under putative cytotoxic T-lymphocyte pressure (black squares) and antibody pressure (black triangles) are illustrated for approximate periods after infection of 0–5 weeks (pale blue shaded region), 5–12 weeks (darker blue shaded region) and 12–29 weeks (gray shaded region). The dashed gray box is used for participant CAP45 as the black square represents an epitope which may have mutated anywhere between 12 and 65 weeks after infection. Time of first detection of autologous neutralizing antibodies is indicated by dashed red lines.

under immune pressure): 21 were classified as under CTL pressure (Table 2), 19 under antibody pressure (Table 3), nine as reversion (see Table S1, Supplemental Digital Content 3, http://links.lww.com/QAD/A282, which tabulates genome regions undergoing reversion) and further six regions contained mutations (clustered within amino acid 9-mers or single persisting to fixation in sites under positive selection) which did not conform to criteria described for CTL or antibody pressure or reversion (Table 3). Identification of regions under immune pressure was supported by selection analysis (for which *gag*, *env* and *nef* supplemental sequences were included), which found a total of 55 sites under positive selection (dN/dS >1) of which 84% (46/55) were situated within genome regions identified as under immune pressure (see Tables 2 and 3; Table S1, Supplemental Digital Content 3 and Figs S2–S6, Supplemental Digital Content 2, http://links.lww.com/QAD/A282, which illustrate positive selection sites corresponding to genome regions under immune pressure). The majority of positively selected sites were identified in *env* ($n = 30$), followed by *gag* ($n = 9$), *pol* ($n = 7$), *nef* ($n = 5$), *rev* ($n = 3$) and *tat* ($n = 1$). APOBEC-mediated G to A hypermutation was identified in 24% (13/55) of regions under immune selection.

Of the mutating genome regions associated with CTL pressure, 19 spanned known CTL epitopes, and two spanned predicted epitopes (Table 2). The highest frequency of CTL-driven mutations was located in *nef* (even when normalizing for amino acid length) ($n = 6$), followed by *pol* and *env* ($n = 4$), *gag* and *vif* ($n = 2$) and finally *rev*, *tat* and *vpr* ($n = 1$) (Fig. 2a). Reversion was identified most frequently in *vpu* ($n = 3$) followed by *gag*, *env* and *nef* ($n = 2$) (see Table S1, Supplemental Digital Content 3, http://links.lww.com/QAD/A282).

Cytotoxic T lymphocyte escape is most frequent in acute infection

We supplemented near full-length genomes with subgenomic (*gag*, *env*, *nef* and targeted epitope) sequence data to better elucidate timing of mutations associated with CTL pressure and escape. The majority of mutating epitopes was identified within the first 5 weeks of infection in structural genes *gag*, *pol*, *env* and *nef* (Figs 1 and 2b). The earliest of these was identified at 2 weeks postinfection in the *gag* HLA B*58 : 01 restricted TW10 epitope; and the *nef* HLA B*45 : 01 restricted EV11 epitope. The frequency of mutation associated with escape slowed over time with an initial 1.6 total escapes per week for the first 5 weeks of infection, to 0.9 escapes per week between 5 and 12 weeks postinfection, and 0.4 escapes per week between 12 and 29 weeks postinfection (Fig. 2b).

Of the 19 regions of *env* with changes in hypervariable loops and PNGSs, mutations in seven (37%) arose in the first 5 weeks of infection (Fig. 1 and Table 3). Since first detection of autologous neutralizing antibodies (nAb) for the women in this study ranged from 9 to 46 weeks [44,47] (Fig. 1), these early changes are unlikely to be the result of nAb pressure. However, in one instance, early mutating sites in V5 of participant CAP45 corresponded to sites where neutralization escape mutations (K460D and D462G) were identified later in infection (see Table 3) [44].

Finally, of the nine reversions identified, four occurred in the first 5 weeks of infection and the remaining between 13 and 29 weeks postinfection (data not shown).

Shuffling and toggling of amino acid mutations

Shifting of mutations between different positions within an epitope (shuffling) or between different amino acids at the same position (toggling) was observed in 81% (17/21) of mutating CTL epitopes (see Table 2 and Table S2, Supplemental Digital Content 4, http://links.lww.com/QAD/A282, illustrating shuffling and toggling in a Nef epitope). For four of the five participants, the number of mutant variants per CTL epitope increased over the 6 months of infection after which a plateau or decrease was seen owing in part to eventual fixation of escape mutants (Fig. 2c).

Twelve of the 17 CTL epitopes with shuffling/toggling mutations corresponded to subtype C database epitopes with high Shannon entropy (scores >0.25 [34]). Shuffling/toggling was however not more frequent in epitopes corresponding to high-entropy regions than in more conserved regions ($P = 0.25$). We investigated whether mutational shuffling/toggling was due to mutations arising in structurally or functionally essential sites, which when altered may abrogate efficient RNA folding or alter protein structure and function. We therefore compared the proportion of epitopes with shuffling/toggling in functionally or structurally relevant sites with the proportion in sites with no currently known functional or structural significance. We found no significant difference between the two categories ($P = 0.53$), even when including the six genome regions under unclassified immune pressure ($P = 0.28$).

Immune pressure, sequence diversification and disease progression

To determine whether rate of genetic diversification from t/f sequences differed with different disease progression profiles, we compared gene-specific tree lengths using sequences from the first 3 months of infection. No significant differences between rates of diversification were found (data not shown). Furthermore, no significant association between number of genome regions under immune pressure and viral load over time was identified (data not shown); possibly owing to low participant numbers.

Table 2. Putative cytotoxic T lymphocyte escape epitopes and polymorphisms.

Participant ID	ORF	Epitope/genome region sequence[a]	HXB2 position	Participant HLA association(s)[b]	Reference	Time of first AA change (range) (weeks)	High entropy epitope/peptide (LANL[c] subtype C database)	Shuffling/toggling of AA mutations
CAP45	Vif	**D**WHLGHGVSI	78–87	B*15:10	LANL database	12–65	No	No
	Rev	I**H**SISERI**L**	52–60	B*15:10	LANL database	5–12	Yes	Yes
	Tat	**N**CYCK**H**CSY	24–32	A*29:02	LANL database	5–12	Yes	Yes
	Nef	**EE**VGFPVRPQV	64–74	B*45:01	Matthews *et al.* [38]	5–9	No	Yes
CAP63	Pol	ALT**E**IC**EE**M	188–196	A*02:01	LANL database	5–11	Yes	Yes
	Pol	QLTE**AV**H**K**I	522–530	Predicted A*02:01		11–29	No	Yes
	Vpr	AL**IRI**LQQL	59–67	A*02:01	LANL database	5–11	Yes	Yes
	Gp41	SWSNKSE**E**DIW**G**NMTWMQ	102–119	A*23:01/Cw*04:01	LANL database	11–29	Yes	Yes
	Gp41	L**LD**SI**A**I**TV**	303–311	A*02:01	LANL database	2–4	Yes	Yes
	Nef	ALT**SS**NT**AA**	42–50	A*02:01	LANL database	5–11	Yes	Yes
	Nef	**EE**VGFPVRPQV	64–74	B*45:01	Matthews *et al.* [38]	0–2	No	Yes
CAP85	Pol	K**A**GYVTD**R**GRQKV**VS**LTE	609–626	B*08:01	Matthews *et al.* [38]	0–5	Yes	Yes
	Gp41	**RY**L**GS**L**V**Q**Y**	283–291	A*30:02	LANL database	0–5	Yes	Yes
	Nef	**KE**VGFP**V**RPQV	64–74	B*45:01	Matthews *et al.* [38]	0–5	No	Yes
	Nef	**Y**FPDWQNY	120–127	A*30:02	LANL database	13–29	No	No
CAP210	Gag	VHQ**A**ISPRTL	143–152	B*15:10	Matthews *et al.* [38]	12–16	No	No
	Vif	**D**WHLGHG**V**SI	78–87	B*15:10	LANL database	5–12	No	Yes
	Gp41	**EA**TDR**I**L**EL**	313–321	Predicted A*68:02		2–5	Yes	Yes
CAP239	Gag	TS**TL**QEQV**A**W	240–249	B*58:01	Matthews *et al.* [38]	0–2	No	Yes
	Pol	IVLP**EKE**SW	399–407	B*58:01	Matthews *et al.* [38]	2–5	Yes	Yes
	Nef	K**AA**V**V**DLSFF	82–90	B*58:01	Matthews *et al.* [38]	11–22	Yes	No

[a]Bold amino acids indicate sites undergoing mutation.
[b]Predicted epitopes obtained using Net MHCPan2.0 (www.cbs.dtu.dk/services/NetMHCPan).
[c]Los Alamos National Laboratory (LANL) database (www.hiv.lanl.gov) HIV Molecular Immunology 2008 compendium used.

Table 3. Genome regions under putative antibody-mediated or unclassified immune pressure.

Participant ID	ORF	Genome region sequence[a] (18–20 mer)	HXB2 position	Putative immune selective pressure	Time of first mutation in hypervariable loop or PNGS (range) (weeks)
CAP45	Gp120	LTRDGG**KTDR**N**DT**EIFRP	454–470	Antibody	0–2
CAP63	Gp120	QE**I**VLE**N**V**I**ENFNMWKND	82–99	Antibody	5–11
	Gp120	LTPLCVTLNCANA**N**I**T**KN	122–139	Antibody	11–29
	Gp120	**MI**GEI**K**NCSFN**ATTELR**D	147–167	Antibody	2–5
	Gp120	L**NNNRSNENS**YILINC**N**S	184–198	Antibody	0–2
	Gp120	IVH**F**NQSV**K**IVC**A**RPHNN	285–302	Antibody	11–29
	Gp120	IRQA**H**CNIS**KTQ**W**NTT**LE	326–343	Antibody	11–29
	Gp120	**FNSTYMPNGIHIPNGASEVIT**	396–415	Antibody	2–5
	Gp41	**L**WS**S**WFN**I**S**H**WL**W**YI**RI**FI	158–147	Antibody	11–29
	Gp41	IE**EE**GG**EQ**DN**S**RS**I**RLVS	222–239	Antibody	5–11
CAP85	Gp120	D**IVPLNNDIGNYSEY**RLI	180–194	Antibody	5–13
	Gp120	IVHLN**HS**V**K**IVCTRPGNN	285–302	Antibody	0–5
	Gp120	IRQAH**C**NIS**KA**EWN**NT**LE	326–343	Antibody	13–29
	Gp120	**GSSTT**T**NGSSP**ITLPCRI	404–420	Antibody	0–5
	Gp120	RPGGG**DM**K**D**NWRSELYKY	469–486	Unclassified	0–5
	Nef	GVGAASQDLGK**YG**ALT**SS**	29–46	Unclassified	0–5
CAP210	Pol	FFR**E**NLAFP**EGE**AR**ELPS**	1–18	Unclassified	12–16
	Gp120	ICSFN**AT**TELRDKKKKEY	156–173	Antibody	5–12
	Gp120	FNSTHN**ST**D**S**T**V**N**STDS**T	391–409	Antibody	5–12
	Gp120	ITCIS**NIT**GLLLTRDGGE	443–460	Antibody	22–26
	Nef	SLH**G**MEDTE**R**EVLQWKFD	169–186	Unclassified	5–12
CAP239	Gag	SNPS**GP**K**RPI**KC**F**NCGRE	382–399	Unclassified	2–5
	Rev	GRPA**EPV**P**F**QLPPIERLH	65–82	Unclassified	2–5
	Gp120	DIIIRSQ**NIL**DN**T**K**TIIV**	269–286	Antibody	2–5
	Gp120	GLLLT**W**DGGDS**KE**N**KT**R**H**	451–467	Antibody	11–22

PNGS, potential N-linked glycosylation site.
[a]Bold amino acids indicate sites undergoing mutation; underlined amino acids indicate sites evolving under positive selection; highlighted amino acids indicate sites mutating to result in a change within, or gain/loss of a potential N-linked glycosylation site gain/loss (NXS/Tx, where x is not Proline).

Discussion

Design of a globally relevant HIV-1 vaccine requires an understanding of transmitted viruses and their early immune adaptation in different populations. Here, we report the first study to comprehensively identify and classify the earliest changes to subtype C transmitted/founder full-length genome viruses. We provide a detailed analysis of the timing, frequency and patterns of these changes in five women with differing clinical disease progression. The predominant and earliest host selective pressure was from cytotoxic T lymphocytes. Despite increasing breadth of CTL responses over time [23], the frequency of mutations associated with CTL escape declined as viral load declined. We show that complex mutational pathways are used to escape in the majority of epitopes.

We predicted that CTL pressure, or reversion of transmitted escape, accounted for the majority (54%) of changes across the genome, reaffirming the importance of this response in early infection. Supporting our predictions, autologous peptide screening using IFN-γ ELISPOT in three of the five participants confirmed all nine epitopes predicted for these individuals (Liu *et al.*, in press). A further four of the 21 epitopes were likewise confirmed to be responsive (Liu *et al.*, unpublished data). The remaining eight were not confirmed either due to poor cell quality or screening after escape had already occurred. Mutations associated with CTL escape occurred earliest in *gag* and most frequently in *nef*, although this was likely influenced by the fact that three of the five individuals were HLA B*45 : 01 positive and targeted the same *nef* epitope. A further 35% of changes were associated with antibody pressure. We could not classify 11% of changes which may be associated with novel CTL epitopes (possibly in alternate reading frames [48,49]), compensation of escape, changes in antigen processing, $CD4^+$ T-cell pressure, natural killer (NK) cell pressure [50], nonneutralizing binding antibody activity, antibody-dependent cellular cytotoxicity (ADCC), viral fitness or evolutionary drift/hitchhiking. One unclassified mutating region in *pol* of participant CAP210 was subsequently found to be responsive by IFN-γ ELISPOT and may represent a novel CTL epitope (Liu *et al.*, in press). Another mutating region in *nef* of CAP210 corresponded to a known HLA-DR $CD4^+$ T-cell epitope; however, it was not restricted by the participants' HLA-DR (data not shown).

Three findings from this study illustrate the immense pressure the virus is under following infection. First, in these five women, escape was rapid and occurred at high frequency in acute infection (<5 weeks postinfection), with a four-fold reduction in the number of escapes per week over the 6-month period. Seventy percent of early

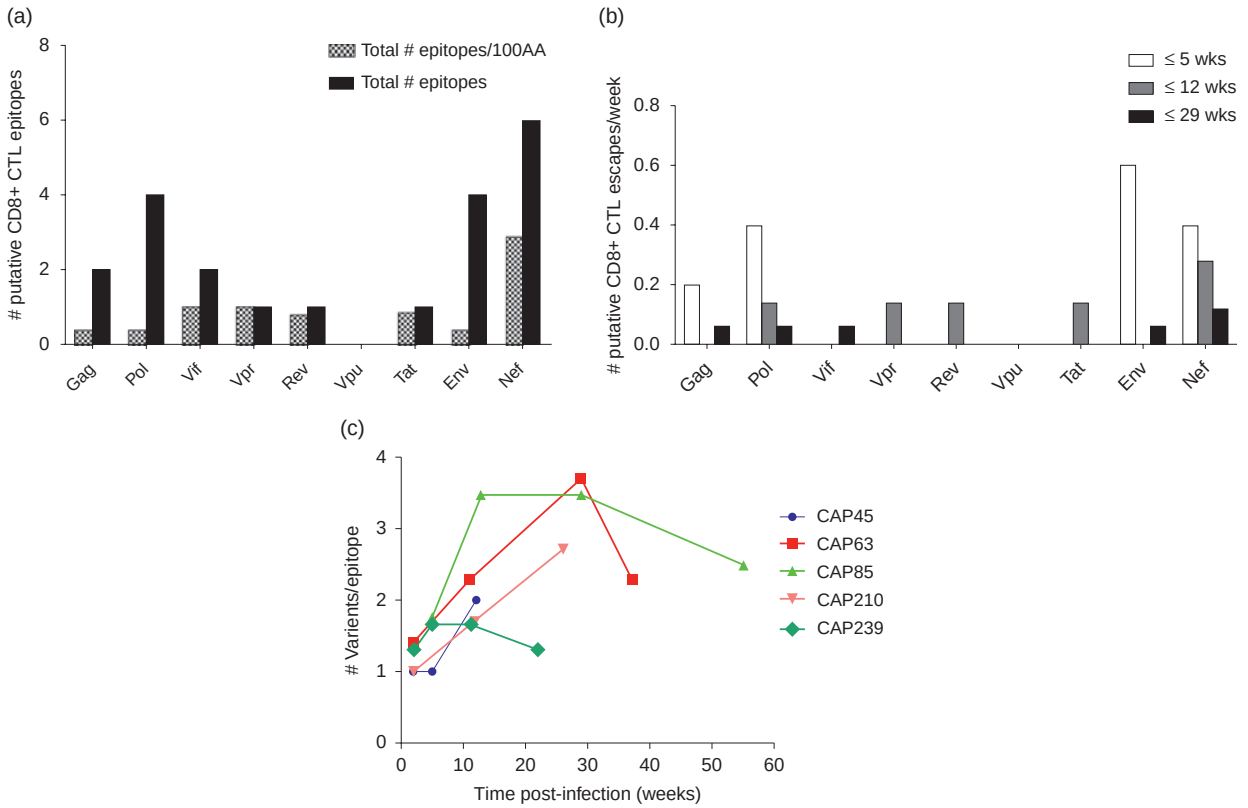


Fig. 2. Frequency, timing and complexity of putative cytotoxic T lymphocyte escape in the first 6 months of infection. (a) Total number of putative escapes per gene and per 100 amino acids (AAs). (b) Total number of putative escapes per week. (c) Total number of variants per CTL epitope over time for each of the five women. CTL, cytotoxic T lymphocyte.

escape (<12 weeks postinfection) occurred in high-entropy locations compared with only 33% after 12 weeks. These observations, although not statistically significant due to low sample size, consolidate reports that strong CTL pressure and rapid escape prevail early in infection [21,51] and that epitopes with higher entropy escape faster than more conserved epitopes [52]. Thus, the observed reduction in mutations associated with CTL escape later in infection seen in these women may be associated with targeting of epitopes that escape slowly.

Second, the pathway to CTL escape is complex, with this study showing an unprecedented level (81%) of genome regions under CTL pressure with mutations shifting between different amino acid sites or between different amino acids at the same site resulting in late or no fixation over the period investigated. A recent study proposed that initial mutations are replaced by secondary mutations which are less costly to the virus with respect to its survival [53]; however, we found no statistical support for more frequent shuffling/toggling in virus regions of structural or functional importance. It may be that this pattern of variation is seen simply because these regions of the genome are able to tolerate multiple changes (supported by the fact that 71% of the epitopes with shuffling/toggling corresponded to high-entropy database epitopes). Alternatively, fixation may only occur once mutations are introduced into sites facilitating complete escape as postulated in a detailed study of one participant by Henn *et al.* [54].

Third, and perhaps surprisingly since neutralizing antibody pressure typically emerges some weeks or months after infection [9,55,56], we identified escape patterns typically associated with antibody pressure as early as 2 weeks postinfection. Autologous nAb response data for these five participants indicate the earliest response to have arisen at 9 weeks postinfection [44]. These early changes could possibly be attributed to reversion of antibody escape mutations in the donor, early stochastic changes, nonneutralizing binding antibody activity, ADCC, *env* fitness or as very recently reported in a subtype B study by Bar *et al.* [57], may in fact be a result of early low level neutralizing antibody responses from which the virus escapes. We saw evidence for this in V5 of CAP45 wherein early changes corresponded to sites where nAb escape later in infection was confirmed in this participant [44].

When examining participant sequence data in the context of disease progression, we found no significant differences in early virus sequence diversification between individuals in our study. However, larger studies would be needed to evaluate the role of CTL escape in disease

progression. Recent studies revealed that clinical disease profiles are heritable in donor–recipient pairs [58,59] and that viral genotype can be a determinant of viral load set point [60]. Furthermore, transmission of a virus harboring a mutation with a fitness cost can also contribute to viral control [28,61]. We observed that viremic controller CAP45 was infected with a virus harboring mutations flanking the B*57/58 Gag TW10 and ISW9 epitopes previously reported to be associated with disease control [28]. It is worth mentioning that the two rapid progressors in this study were HLA-B (CAP63) and HLA-A, HLA-B and HLA-C (CAP210) homozygous. HLA homozygosity has previously been associated with poor clinical outcome [62].

Numerous CTL studies propose stable CTL epitopes that escape slowly or late in infection due to high fitness costs to be good vaccine targets [21,38,51]. This study suggests that the process of immune escape holds greater complexities which vaccine design strategies will need to take into account. Frequent and persistent shuffling and toggling of mutations within targeted epitopes may indicate high levels of epitope instability early in infection. Early APOBEC-mediated hypermutation, identified in more than a quarter of genome regions under putative immune pressure in this study, represents an efficient mechanism of CTL escape [22]. Not to be discounted is the unclassified 11% of genome changes which demonstrate that much more is at play very early in infection, possibly such as pressure from NK cells which has recently been emphasized [50]. These additional forces likely play a significant role in shaping the early virus.

Our findings provide novel data on the dynamic interplay between virus and host very early in infection and the complex pathways of escape in response to the earliest immune pressures acting on transmitted/founder subtype C viruses. These processes of immune adaptation are likely to pose further challenges that vaccines will need to overcome. Furthermore, this study demonstrates the highly sensitive nature of viral sequencing as a tool for the identification and characterization of early immune selective pressures that mould early HIV-1 evolution. Examining early host–virus interactions in the context of disease progression will enable us to identify those changes to the virus which are associated with better disease outcome and may therefore be incorporated into vaccine design.

Acknowledgements

This work is funded by the National Institute of Allergy and Infectious Disease (NIAID), National Institutes of Health (NIH) and the US Department of Health and Human Services (DHHS) (#AI51794, CAPRISA; #DK 49381 (MSC), CHAVI), as well as by the National Research Foundation (#67385), SA; the South African AIDS Vaccine Initiative; and amfAR grant 106997-43. We thank the clinical staff and participants from the CAPRISA 002 Acute Infection Study. We would like to thank Denis Chopera, Gama Bandawe and Roman Ntale for subgenomic sequence data and Ziyaad Valley-Omar for assistance with data analysis.

Author contributions: M.-R.A.: first author, manuscript writing, full-length genome and focused epitope DNA amplification and sequencing, data collation, sequence analyses, statistical analyses; F.K.T.: study design and coordination, full-length genome DNA amplification and sequencing, sequence analysis; N.K.N.: positive selection analyses and selection methods writing; S.A.G.: full-length genome and focused epitope DNA amplification and sequencing, statistical and sequence analyses; J.C.M.: full-length genome sequencing, focused epitope amplification and sequencing, sequence analysis; H.B.: half-genome DNA amplification and sequencing, sequence analysis; R.T.: full-length genome amplification and sequencing; D.d.A.R.: HLA typing of CAPRISA participants; K.M.: clinical site project director; C.S.: advising and guidance of clustered mutation and selection analyses; S.A.K.: CAPRISA study design; C.M.G.: immunology consultant and assistance with manuscript writing; and C.W.: principal investigator, corresponding author, manuscript writing.

Conflicts of interest

There are no conflicts of interest.

References

1. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, *et al*. **Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection.** *Proc Natl Acad Sci U S A* 2008; **105**:7552–7557.
2. Abrahams M-R, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping L-H, *et al*. **Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a nonpoisson distribution of transmitted variants.** *J Virol* 2009; **83**:3556–3567.
3. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, *et al*. **Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection.** *J Exp Med* 2009; **206**:1273–1289.
4. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, *et al*. **Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing.** *J Virol* 2008; **82**:3952–3970.
5. Keele BF, Li H, Learn GH, Hraber P, Giorgi EE, Grayson T, *et al*. **Low-dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1.** *J Exp Med* 2009; **206**:1117–1134.
6. Pantaleo G. **HIV-1 T-cell vaccines: evaluating the next step.** *Lancet Infect Dis* 2008; **8**:82–83.
7. Abdool Karim Q, Abdool Karim SS, Frohlich JA, Grobler AC, Baxter C, Mansoor LE, *et al*. **Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women.** *Science* 2010; **329**:1168–1174.

8. Yu XG, Addo MM, Perkins BA, Wej F, Rathod A, Geer SC, *et al*. **Differences in the expressed HLA class I alleles effect the differential clustering of HIV type 1-specific T cell responses in infected Chinese and caucasians.** *AIDS Res Human Retroviruses* 2004; **20**:557–564.
9. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, *et al*. **Antibody neutralization and escape by HIV-1.** *Nature* 2003; **422**:307–312.
10. Treurnicht FK, Seoighe C, Martin DP, Wood N, Abrahams M-R, Rosa DDA, *et al*. **Adaptive changes in HIV-1 subtype C proteins during early infection are driven by changes in HLA-associated immune pressure.** *Virology* 2010; **396**:213–225.
11. Rybarczyk BJ, Montefiori D, Johnson PR, West A, Johnston RE, Swanstrom R. **Correlation between env V1/V2 region diversification and neutralizing antibodies during primary infection by simian immunodeficiency virus sm in rhesus macaques.** *J Virol* 2004; **78**:3561–3571.
12. Richman DD, Wrin T, Little SJ, Petropoulos CJ. **Rapid evolution of the neutralizing antibody response to HIV type 1 infection.** *Proc Natl Acad Sci U S A* 2003; **100**:4144–4149.
13. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. **Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level.** *Science* 2002; **296**:1439–1443.
14. Frost SDW, Wrin T, Smith DM, Kosakovsky Pond SL, Liu Y, Paxinos E, *et al*. **Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection.** *Proc Natl Acad Sci U S A* 2005; **102**:18514–18519.
15. Carlson JM, Brumme ZL. **HIV evolution in response to HLA-restricted CTL selection pressures: a population-based perspective.** *Microbes Infect* 2008; **10**:455–461.
16. Allen TM, Altfeld M, Geer SC, Kalife ET, Moore C, Sullivan KMO, *et al*. **Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution.** *J Virol* 2005; **79**:13239–13249.
17. Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, *et al*. **Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus.** *Nature* 1997; **3**:205–211.
18. Jones NA. **Determinants of human immunodeficiency virus type 1 escape from the primary CD8+ cytotoxic T lymphocyte response.** *J Exp Med* 2004; **200**:1243–1256.
19. Liu Y, McNevin J, Cao J, Zhao H, Genowati I, Wong K, *et al*. **Selection on the human immunodeficiency virus type 1 proteome following primary infection.** *J Virol* 2006; **80**:9519–9529.
20. Gray CM, Mlotshwa M, Riou C, Mathebula T, de Assis Rosa D, Mashishi T, *et al*. **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* 2009; **83**:470–478.
21. Goonetilleke N, Liu MKP, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Ganusov VV, *et al*. **The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection.** *J Exp Med* 2009; **206**:1253–1272.
22. Wood N, Bhattacharya T, Keele BF, Giorgi E, Liu M, Gaschen B, *et al*. **HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC.** *PLoS Pathog* 2009; **5**:e1000414.
23. Mlotshwa M, Riou C, Chopera D, de Assis Rosa D, Ntale R, Treunicht F, *et al*. **Fluidity of HIV-1-specific T-cell responses during acute and early subtype C HIV-1 infection and associations with early disease progression.** *J Virol* 2010; **84**:12018–12029.
24. Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, *et al*. **Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome.** *J Virol* 1994; **68**:4650–4655.
25. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. **Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection.** *J Virol* 1994; **68**:6103–6110.
26. Ogg GS, Jin X, Bonhoeffer S, Dunbar PR, Nowak MA, Monard S, *et al*. **Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA.** *Science* 1998; **279**:2103–2106.
27. van Loggerenberg F, Mlisana K, Williamson C, Auld SC, Morris L, Gray CM, *et al*. **Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study.** *PLoS One* 2008; **3**:e1954.
28. Chopera DR, Woodman Z, Mlisana K, Mlotshwa M, Martin DP, Seoighe C, *et al*. **Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage.** *PLoS Pathog* 2008; **4**:e1000033.
29. Rousseau CM, Birditt BA, McKay AR, Stoddard JN, Lee TC, McLaughlin S, *et al*. **Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes.** *J Virol Methods* 2006; **136**:118–125.
30. Chopera DR, Mlotshwa M, Woodman Z, Mlisana K, de Assis Rosa D, Martin DP, *et al*. **Virological and immunological factors associated with HIV-1 differential disease progression in HLA-B 58:01-positive individuals.** *J Virol* 2011; **85**:7070–7080.
31. Hall TA. **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Ser* 1999; **41**:95–98.
32. Tamura K, Dudley J, Nei M, Kumar S. **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007; **24**:1596–1599.
33. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, *et al*. **NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence.** *PLoS One* 2007; **2**:e796.
34. Bansal A, Gough E, Sabbaj S, Ritter D, Yusim K, Sfakianos G, *et al*. **CD8 T-cell responses in early HIV-1 infection are skewed towards high entropy peptides.** *AIDS* 2005; **19**:241–250.
35. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. **Relaxed phylogenetics and dating with confidence.** *PLoS Biol* 2006; **4**:e88.
36. Drummond AJ, Rambaut A. **BEAST: Bayesian evolutionary analysis by sampling trees.** *BMC Evol Biol* 2007; **7**:214.
37. Mansky LM, Temin HM. **Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase.** *J Virol* 1995; **69**:5087–5094.
38. Matthews PC, Prendergast A, Leslie A, Crawford H, Payne R, Rousseau C, *et al*. **Central role of reverting mutations in HLA associations with human immunodeficiency virus set point.** *J Virol* 2008; **82**:8548–8559.
39. Kosakovsky Pond SL, Frost SDW. **Not so different after all: a comparison of methods for detecting amino acid sites under selection.** *Mol Biol Evol* 2005; **22**:1208–1222.
40. Pond SK, Muse SV. **Site-to-site variation of synonymous substitution rates.** *Mol Biol Evol* 2005; **22**:2375–2385.
41. Pond SLK, Frost SDW, Muse SV. **HyPhy: hypothesis testing using phylogenies.** *Bioinformatics* 2005; **21**:676–679.
42. Scheffler K, Martin DP, Seoighe C. **Robust inference of positive selection from recombining coding sequences.** *Bioinformatics* 2006; **22**:2493–2499.
43. Pereyra F, Addo MM, Kaufmann DE, Liu Y, Miura T, Rathod A, *et al*. **Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy.** *Int J Infect Dis* 2008; **197**:563–571.
44. Moore PL, Ranchobe N, Lambson BE, Gray ES, Cave E, Abrahams M-R, *et al*. **Limited neutralizing antibody specificities drive neutralization escape in early HIV-1 subtype C infection.** *PLoS Pathog* 2009; **5**:e1000598.
45. Li B, Gladden AD, Altfeld M, Kaldor JM, Cooper DA, Kelleher AD, *et al*. **Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution.** *J Virol* 2007; **81**:193–201.
46. Brumme ZL, Brumme CJ, Carlson J, Streeck H, John M, Eichbaum Q, *et al*. **Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection.** *J Virol* 2008; **82**:9216–9227.
47. Moore PL, Gray ES, Choge IA, Ranchobe N, Mlisana K, Abdool Karim SS, *et al*. **The C3-V4 region is a major target of autologous neutralizing antibodies in human immunodeficiency virus type 1 subtype C infection.** *J Virol* 2008; **82**:1860–1869.

48. Bansal A, Carlson J, Yan J, Akinsiku OT, Schaefer M, Sabbaj S, *et al*. **CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription.** *J Exp Med* 2010; **207**:51–59.
49. Berger CT, Carlson JM, Brumme CJ, Hartman KL, Brumme ZL, Henry LM, *et al*. **Viral adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting epitopes in HIV frameshift sequences.** *J Exp Med* 2010; **207**:61–75.
50. Alter G, Heckerman D, Schneidewind A, Fadda L, Kadie CM, Carlson JM, *et al*. **HIV-1 adaptation to NK-cell-mediated immune pressure.** *Nature* 2011; **476**:96–100.
51. Davenport MP, Loh L, Petravic J, Kent SJ. **Rates of HIV immune escape and reversion: implications for vaccination.** *Trends Microbiol* 2008; **16**:561–566.
52. Ferrari G, Korber B, Goonetilleke N, Liu MKP, Turnbull EL, Salazar-Gonzalez JF, *et al*. **Relationship between functional profile of HIV-1 specific CD8 T cells and epitope variability with the selection of escape mutants in acute HIV-1 infection.** *PLoS Pathog* 2011; **7**:e1001273.
53. Herbeck JT, Rolland M, Liu Y, McLaughlin S, McNevin J, Zhao H, *et al*. **Demographic processes affect HIV-1 evolution in primary infection before the onset of selective processes.** *J Virol* 2011; **85**:7523–7534.
54. Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, *et al*. **Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection.** *PLoS Pathog* 2012; **8**: e1002529.
55. Gray ES, Moore PL, Choge IA, Decker JM, Bibollet-Ruche F, Li H, *et al*. **Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection.** *J Virol* 2007; **81**:6187–6196.
56. Yeh WW, Rahman I, Hraber P, Coffey RT, Nevidomskyte D, Giri A, *et al*. **Autologous neutralizing antibodies to the transmitted/ founder viruses emerge late after simian immunodeficiency virus SIVmac251 infection of rhesus monkeys.** *J Virol* 2010; **84**:6018–6032.
57. Bar KJ, Tsao C-Y, Iyer SS, Decker JM, Yang Y, Bonsignori M, *et al*. **Early low-titer neutralizing antibodies impede HIV-1 replication and select for virus escape.** *PLoS Pathog* 2012; **8**:e1002721.
58. Hecht FM, Hartogensis W, Bragg L, Bacchetti P, Grant R, Barbour J, *et al*. **HIV RNA level in early infection is predicted by viral load in the transmission source.** *AIDS* 2010; **24**:941–945.
59. Hollingsworth TD, Laeyendecker O, Shirreff G, Donnelly C a, Serwadda D, Wawer MJ, *et al*. **HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda.** *PLoS Pathog* 2010; **6**:e1000876.
60. Alizon S, von Wyl V, Stadler T, Kouyos RD, Yerly S, Hirschel B, *et al*. **Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load.** *PLoS Pathog* 2010; **6**:e1001123.
61. Miura T, Brumme ZL, Brockman M a, Rosato P, Sela J, Brumme CJ, *et al*. **Impaired replication capacity of acute/early viruses in persons who become HIV controllers.** *J Virol* 2010; **84**:7581–7591.
62. Tang J, Costello C, Keet IP, Rivers C, Leblanc S, Karita E, *et al*. **HLA class I homozygosity accelerates disease progression in human immunodeficiency virus type 1 infection.** *AIDS Res Human Retroviruses* 1999; **15**:317–324.
63. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, *et al*. **Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection.** *AIDS* 2003; **17**:1871–1879.

Appendix 3: Supplemental near full-length and sub-genomic sequences

Participant ID	Sample date	Weeks post-infection[1]	Whole (half) genome sequences[2]	Env SGA sequences[3]	Gag clone sequences[4]	Nef clone sequences[4]	Partial genome sequences[5]
CAP45	20-Apr-05	2		16			
	11-May-05	5					1
	07-Jun-05	9				8	
	04-Apr-06	52		7			
	04-Jul-06	65	3				
CAP63	06-Jan-05	2		19			
	13-Jul-05	29		26			
	07-Sep-05	37	3				
CAP85	22-Jun-05	5		21			
	07-Dec-05	29		10	1		
	16-Feb-06	39				1	
	10-May-06	51			9		
	07-Jun-06	55	2	8	1		
	07-Dec-06	81			1		
	06-Jun-07	107			7		
CAP210	03-May-05	2		21			
	21-Sep-05	22		8			
	19-Oct-05	26			1	2	
	23-Nov-05	31				3	
CAP239	19-Jul-05	2			21		
	10-Aug-05	5		24	27	23	1
	17-Aug-05	6			19		
	07-Dec-05	22			9		
	14-Jun-06	49	0 (4)				
	11-Jan-07	79			9		
	04-Oct-07	117	3				

[1]Time from mid-point date between first positive and last negative HIV EIA test to date of sampling, or time since first HIV RNA positive date plus 14 days in antibody negative individuals

[2]Whole genome sequence numbers include smaller genomes no less than 6 000 bp in size

[3]Sequences from 2-5 weeks are published in Abrahams et al., 2009; the remaining sequences were obtained from G. Bandawe

[4]Obtained from D. Chopera and R.Ntale

[5]Near full-length genome amplification products that were <6 000 bp in size

Appendix 4: Limiting dilution amplification of *gag*, *vif* and *nef*

Viral RNA extraction and cDNA synthesis are described in Chapter 2 section 2.5.2. For *gag* and *nef*, cDNA was serially diluted and nested PCR amplification was carried out using Expand High Fidelity Taq (Roche Diagnostics, Rotkreuz, Switzerland). For *vif*, a hemi-nested PCR amplification was carried using methods described in Chapter 2 section 2.5.3 with the exception that primer VIF-1 was used in both the first and second round of amplification. Amplicons obtained at ≤60% positivity were directly sequenced using the ABI 3000 genetic analyzer (Applied Biosystems, Foster City, CA, USA) and BigDye terminator reagents. In the case of *gag* and *vif*, only the region containing the epitope/s of interest was sequenced, and in the case of *nef* the whole gene was sequenced. Primer details, reaction mixes and PCR cycling conditions for *gag* and *nef* amplification are provided below.

Amplification of *gag*:

1st round forward primer:	Gag D forw	5'-TCT CTA GCA GTG GCG CCC G-3'
1st round reverse primer:	Gag D rev	5'-AAT TCC TCC TAT CAT TTT TGG-3'
2nd round forward primer:	Gag A forw	5'-CTC TCG ACG CAG GAC TCG GCT T-3'
2nd round reverse primer:	Gag A rev	5'-TCT TCT AAT ACT GTA TCA TCT GC-3'

1st round PCR reagent mix and cycling conditions:

Reagent	Volume (µl)
dH_2O	15.5
10 X Buffer 1 + $MgCl_2$ (15mM)	2
dNTPs (10 mM)	0.4
Gag D forw (10µM)	0.4
Gag D rev (10µM)	0.4
Expand High Fidelity (5U/µl)	0.3
cDNA	1
Total	**20**

Temperature (°C)	Incubation Time	Number of cycles
95	2 min	
95	15 sec	10
52	30 sec	
72	1 min	
95	15 sec	15
55	30 sec	
72	1.5 min	
72	7 min	
4	Hold	

2nd round PCR reagent mix and cycling conditions:

Reagent	Volume (µl)
dH_2O	15.5
10 X Buffer 1 + $MgCl_2$ (15mM)	2
dNTPs (10 mM)	0.4
Gag A forw (10µM)	0.4
Gag C rev (10µM)	0.4
Expand High Fidelity (5U/µl)	0.3
1st round product	1
Total	**20**

Temperature (°C)	Incubation Time	Number of cycles
95	2 min	
95	15 sec	35
55	30 sec	
72	1.5 min	
72	7 min	
4	Hold	

Amplification of *nef*:

1st round forward primer:	SQ15FC	5′-GAG AGC GGT GGA ACT TCT-3′
1st round reverse primer:	Nef OR	5′-AGG CAA GCT TTA TTG AGG -3′
2nd round forward primer:	Nef F	5′-CCT AGA AGA ATA AGA CAG GGC TT-3′
2nd round reverse primer:	Nef R	5′-CCT GGA ACG CCC CAG TGG-3′

1st round PCR reagent mix and cycling conditions:

Reagent	Volume (µl)
dH_2O	15.5
10 X Buffer 1 + $MgCl_2$ (15mM)	2
dNTPs (10 mM)	0.4
SQ15FC (10µM)	0.4
Nef OR (10µM)	0.4
Expand High Fidelity (5U/µl)	0.3
cDNA	1
Total	**20**

Temperature (°C)	Incubation Time	Number of cycles
95	2 min	
95	15 sec	10
52	45 sec	
72	1 min	
95	15 sec	15
55	30 sec	
72	1 min	
72	7 min	
4	Hold	

2nd round PCR reagent mix and cycling conditions:

Reagent	Volume (µl)
dH_2O	15.5
10 X Buffer 1 + $MgCl_2$ (15mM)	2
dNTPs (10 mM)	0.4
Nef F (10µM)	0.4
Nef R (10µM)	0.4
Expand High Fidelity (5U/µl)	0.3
cDNA	1
Total	**20**

Temperature (°C)	Incubation Time	Number of cycles
95	2 min	
95	15 sec	35
55	30 sec	
72	1 min	
72	7 min	
4	Hold	

Appendix 5: Construction of IMC pCAP200.t/f

The t/f virus full-length genome (including cognate LTRs) of CAPRISA AI 002 cohort participant CAP200 was selected for cloning into the low-copy number plasmid pBR322 for use in characterization of transmitted/founder virus phenotype. The method described is a modification of that described by Salazar-Gonzalez et al. (2009) and Ochsenbauer et al. (2012). The sequence of the t/f virus was derived as the consensus of 11 SGA near full-length genome sequences from the earliest time-point post-infection by S Goodier (approximately two weeks). Derivation of the complete 5' and 3'LTRs is outlined in Methods and Materials section 4.5.1.

The IMC, named pCAP200.t/f was designed so as to sequentially clone the complete genome in three fragments into a low copy number vector pBR322 using restriction enzyme digestion and sticky-end ligation. Fragment synthesis eliminates error induced by polymerases during PCR amplification of genome fragments and thus eliminates the need for back-mutation of unwanted polymorphisms following cloning. Furthermore, the use of a low copy number vector enhances the stability of the large IMC construct (approximately 14 kb) which is often unstable following transformation into bacterial cells.

The design and construction of the IMC, named pCAP200.t/f, was as follows:

(i) Genome partitioning

The CAP200 full-length viral sequence was analysed for the presence of restriction enzyme sites which would partition the genome into three fragments. Two single-cutter restriction enzyme sites (PflFI and BstBI), which cut the genome into three fragments of approximately 2 100 bp (fragment I, 5'LTR start to *gag*), 4 700 bp (fragment II, *gag* to *env*) and 2 700 bp (fragment III, *env* to 3'LTR end). Since *env* may prove toxic to bacterial cells resulting in complete or partial loss or shuffling of the gene within bacterial plasmid vectors, the second restriction enzyme site selected interrupted the *env* open reading frame to ensure that the complete gene would only be present once the final of the three fragments was inserted into pBR322. The selected restriction enzyme sites and their positioning within the CAP200 viral genome are indicated in **Figure 36A**.

(i) Optimization of fragment design for 3-step cloning

In order to insert the genome in three steps using restriction enzyme digestion and sticky-end ligation, a SalI site was incorporated at the 3' end of each fragment to allow

for cloning into pBR322 (SalI has a single recognition site within pB322). Fragment I was therefore designed to incorporate a PflFI site flanked by a SalI site at the 3'end; fragment II was designed to incorporate a PflF1 site at the 5' end and a BstBI site flanked by a SalI site at the 3' end; and fragment III was designed to incorporate a BstBI site at the 5' end and a SalI site at the 3' end. The digestion and fragment synthesis design are illustrated in **Figure 36B**.

(ii) Restriction enzyme digestion and cloning

The step-wise cloning procedure is illustrated in **Figure 36C**. Fragment III was inserted into pBR322 first following SapI/PflFI digestion. Likewise, fragment I was inserted into pBR322 containing fragment III using ClaI/PflFI digestion and ligation. Fragment II could not be stably inserted into pBR322 containing fragments I and II by restriction enzyme digestion and sticky-end ligation, and was therefore inserted into pBR322 containing fragments I and III using In-Fusion cloning technology (Clontech, CA, USA). Briefly, fragment II was PCR amplified from pUC57 using a high-fidelity Taq polymerase and primers which incorporated 15 base pair overhangs which matched its 5' and 3' flanking regions within the CAP200 genome. During amplification, the forward primer incorporated a 15 base pair overlap with the 3' end of fragment I and the reverse primer incorporated a 15 base pair overlap with the 5' end of fragment III. The pBR322 vector containing fragments I and III was then PflFI/BStBI digested and the fragment II PCR product was fused with the linearized vector by the In-Fusion HD enzyme. Sequencing of the final pCAP200.T/F IMC revealed the cloned viral sequence to be identical to the derived t/f sequence, and no unwanted mutations were introduced by PCR.

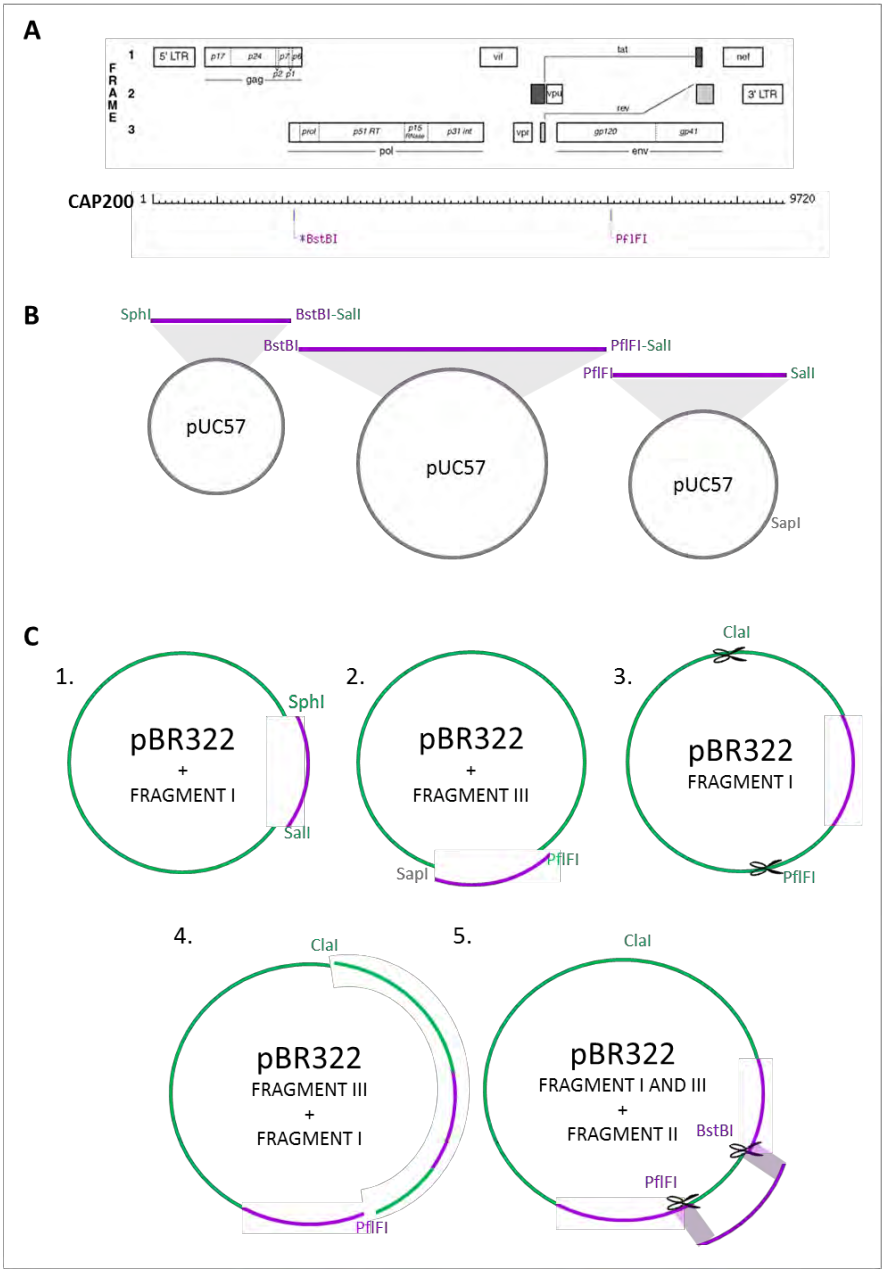


Figure 36. **Construction of pCAP200c transmitted/founder IMC**. (**A**) The location of selected single-cutter restriction enzyme sites partitioning the viral genome into three fragments are indicated, (**B**) synthesized fragments with restriction enzyme sites incorporated on 5' and 3' ends to facilitate step-wise cloning were provided cloned into pUC57 and, (**C**) step-wise cloning of each fragment into pBR322 is illustrated (steps 1 to 4, which incorporated fragments I and III into pBR322 respectively, were carried out by restriction enzyme digestion and ligation, and step 5 was carried out by In-Fusion cloning using 15 bp overlaps illustrated by shaded purple bars).