EVOLUTION OF VIRAL POPULATIONS IN INDIVIDUALS INFECTED WITH SINGLE AND MULTIPLE HIV SUBTYPES: A STUDY OF HIV-1 DUAL INFECTION IN A HIGH RISK COHORT OF FEMALE BAR WORKERS FROM TANZANIA

ABRAHAM LUCKY MALAZA

Presented for the degree of Doctor of Philosophy in the Division of Medical Virology,
Institute of Infectious Diseases and Molecular Medicine
University of Cape Town

JUNE 2007

Supervisor : Prof. Carolyn Williamson
Co-supervisor : Dr. Darren Martin

CONTENTS

ABSTRACT

The Human Immunodeficiency virus (HIV) is characterized by high replication rate and high levels of diversity resulting in numerous quasispecies that form a 'swarm' in the infected individual. HIV diversification is driven by selection pressure exerted by the host. The role of high viral diversity in disease progression in context of infection with more than one subtype (dual infection) is not elucidated. Delineating of the evolving viral diversity in dual infection will contribute to our understanding of HIV pathogenesis and thus enable for better strategies for effective therapy and vaccine development strategies. This study forms part of the HIV Superinfection Study (HISIS, www.mmrp.org) which aimed to identify the frequency of dual and superinfection in a high risk cohort of women in Mbeya, Tanzania. The Mbeya region is characterized by a highly diverse HIV epidemic where multiple HIV-1 subtypes co-circulate including A, C, and D, as well as recombinant forms.

The aims of this thesis were: firstly, to investigate the distribution of HIV subtypes and recombinant forms in a cohort from Mbeya, and to determine the prevalence of HIV dual infection using the multi region hybridization assay (MHA); secondly to investigate the dynamics of viral evolution among dually infected individuals using a gel based screening assay (heteroduplex mobility assay, HMA) and sequencing, lastly the study aimed to investigate the role of point mutations and recombination on viral diversity and immune escape in a dually infected individual.

The MHA was used to screen for different subtypes present in the study population (n=57). Subtypes A, C, D and their recombinants forms were detected. Subtype A accounted for 5%, C 33%, D 7% and recombinants 44%. Dual infection was detected in 11% of individuals. AC recombinants accounted for 60% of all recombinant viral strains, ACD and CD strains accounted for approximately 12% each and AD recombinants accounted for 16%. This study provides additional information on the virus strains circulating in Tanzania. Overall these studies confirm that Tanzania harbours a complex diversity of viruses with a large proportion of the viruses being recombinant forms. The high prevalence of dual infections is clearly fuelling the generation of these recombinant viruses and it would be of interest to monitor the molecular epidemiology of the epidemic to determine if new circulating recombinant forms emerge.

In a selected subset of chronically infected participants monitored over 21 months, we identified dual infection in four out of the twelve individuals. The presence of dual infections was screened for using the HMA based on the *vpu* and *env* C2C3 regions of the genome. Sequencing was employed to confirm subtypes.

Four of the twelve women were dually infected with two distinct subtypes (A and C). The majority of single infections were by subtype C (n= 4), two were infected with subtype A, while subtype D was responsible for one infection. One individual was infected by a CD recombinant virus. Analysis of 20 randomly selected clones from different timepoints was used to determine diversity at selected timepoints. *Env* C2C3 region was more diverse than the *vpu* region with the mean DNA distances of *vpu* ranging from 0.9% to 8.5% compared to 0.3% and 19.5% in the *env* C2C3 region. Tz14 was found to harbour the most diverse viral strains with a maximum DNA distance of 12% (median 7%) in the *vpu* region and 14% (median 11%) in the *env* C2C3 region. This individual was selected for an indepth analysis of the role of point mutations and recombination on viral evolution in dual infection.

Analysis of clones from different timepoints demonstrated that dual infections were detectable in a minority of follow-up visits with only one individual having detectable dual infection at most timepoints. This emphasizes the importance of analyzing multiple time-points and that cross-sectional studies will likely underestimate the prevalence of dual infections. An analysis of the contribution of different viral variants in dual infection to overall viral burden illustrated large fluctuations of viral populations over time. However, generally the viral populations were relatively homogeneous at single time point suggesting that, while there is very high potential diversity, this diversity is largely constrained.

Lastly, through full-length gene (*gag* and *nef*) sequencing we investigated positive selection, recombination patterns and screened for evidence of CTL escape in a dually infected individual (Tz14) over time (21 months). No evidence of positive selection was detected in both genes during the study period. Recombination analysis revealed that *gag* sequences had similar break points at time point F0 and F7. Furthermore, it was shown that none of the sequences obtained from both genes were 'pure' subtypes. Limited evidence of predicted CTL escape was observed in the *nef* and *gag*. Based on these results we postulate that viral diversity observed in these individuals is mostly a result of recombination and to a very limited extent through point mutations.

This study confirms that multiple HIV-1 subtypes and recombinant viruses co-circulate in the population of Mbeya, and that there is a high prevalence of dual infections. While dually infected individuals harbour highly divergent viruses with high fluctuation of diversity over time, the diversity at a single timepoint is usually constrained. An investigation of full-length *gag* and *nef* suggests that the major mechanism of viral evolution is through recombination.

ABBREVIATIONS

A	adenine
AIDS	acquired immune deficiency syndrome
APS	ammonium persulphate
ATP	adenosine triphosphate
ART	antiretroviral therapy
b	bases
bp	base pair(s)
°C	degrees celsius
C	cytosine
$CaCl_2$	calcium chloride
CCR5	chemokine coreceptor
cDNA	copy DNA
CTL	cytotoxic T lymphocytes
Cpz	chimpanzee
CXCR4	CXC chemochine receptor
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanidine triphosphate
dH2O	distilled water
dTTP	deoxythymidine triphosphate
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
ds	double stranded
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
env	envelope
h	hours
HAART	highly active antiretroviral therapy
HCl	Hydrochloric acid
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
IFN	interferon
IL	interleukin
IPTG	isopropyl-β-D-thio galactosidase
kB	kilobases
kDa	kiloDaltons
LB	Luria-Bertani broth
NK	cells natural killer cells
LTNP	long-term non-progressor
LTR	long terminal repeat
NSI	non-syncytium inducing
M	Molar
µg	micrograms
µl	microlitres
MA	matrix
Mab	monoclonal antibody
MCS	multiple cloning site
MIP1-α/β	macrophage inflammatory protein-α/β

mg	milligrams
$MgCl_2$	magnesium chloride
MHC	major histocompatibility complex
min	minutes
ml	millilitres
mM	millimolar
mm	millimetres
(m)RNA	messenger ribonucleic acid
N	normal
NA	not applicable
NaOH	sodium hydroxide
NaCl	sodium chloride
NC	nucleocapsid
NF-κB	nuclear transcription factor
ng	nanograms
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pH	hydrogen potential
pol	polymerase
RNase	ribonuclease
rpm	revolutions per minute
RSA	Republic of South Africa
SI	syncytium inducing
SIV	simian immunodeficiency virus
ss	single stranded
STI	sexually-transmitted infections
T	thymidine
Taq	*Thermus aquaticus*
TBE	Tris-borate
TEMED	N,N,N^{l},N^{l} tetramethyl-
Tm	melting temperature
TNF	tumor necrosis factor
Tris	2-amino-2-(hydroxymethyl)-1,3-propandiol
USA	United States of America
UV	ultraviolet
V	volts
Vpu	viral protein u
Vpr	viral protein r
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactosidase
YT	yeast-tryptone
U	units
UCT	University of Cape Town
UNAIDS	Joint United Nations Programme on HIV/AIDS

DECLARATION

The work presented in this thesis was done at the Division of Virology, part of the Institute for Infectious Disease and Molecular Medicine at the University of Cape Town (U.C.T), under the supervision of Associate Professor Carolyn Williamson. This work is all original and my own. Where use has been made of work of others, their contribution has been acknowledged in the text.

Abraham L. Malaza

June 2007

ACKNOWLEDGEMENTS

I wish to sincerely thank my supervisor, Prof. Carolyn Williamson for her advice, patience, guidance, encouragement and invaluable teachings. I also wish to thank her for critically reading this manuscript. I wish to extend a word of gratitude to all my colleagues from the HIV-Diversity laboratory at the IIDMM at U.C.T, especially Dr. Jandre Grobler for his technical assistance without which this work would not be accomplished.

I further wish to express my sincere gratitude to my co-supervisor Dr. Darren Martin for his critical and analytical input to this body of work. I am grateful to the European Union for the scholarship awarded and to Dr. Mika Salminen of the National Health Institute in Helsinki, Finland for hosting me. I also wish to thank the Fogaty fellowship for the award and I am grateful to Dr. Francine McCutchan and Dr. Michael Hoelscher for hosting me at the Henry Jackson Foundation in Maryland.

I am grateful to Dr. Adrian Puren of the NICD for the viral load assays and I thank Mr. Carel van Heerden of the University of Stellenbosch for all the sequencing work. I would like to thank the entire Mbeya Medical Research Group especially the ground staff and the study participants involved in the HISIS study.

To all my friends and especially my family thank you ever so much for all the support. I am also grateful to Elevation-biotech for allowing me the time and space to complete my write-up.

To my kids, Nomfundo and Khutso thank you so much for allowing me the time to pursue my dreams - love you guys with all my heart and soul !

At last but not least to the Lord my provider I give YOU all the glory !!

CHAPTER 1
LITERATURE REVIEW

1.1 General Introduction

Since its first identification in the 1980s the human immunodeficiency virus (HIV)-1 has infected and killed millions of people the world over. The spread of this epidemic has left no part of the world untouched, with the African continent the worst affected. With the introduction of Highly Active Antiretroviral Therapy (HAART), Western countries have being able to decrease HIV/AIDS fatalities and the spread of the virus. In Africa however, HIV/AIDS remains a scourge due to the limitations in rolling out therapy to those who need it. Vaccine development and prevention strategies remain the only viable hope for controlling the epidemic in the developing countries.

HIV-1 is a RNA virus that belongs to the retrovirus family and Lentiviridae genus (Chiu *et al.*, 1985 and Levy *et al.*, 1995). The RNA genome is surrounded by a protein caspid surrounded by a lipid envelope with embedded surface glycoproteins. The genome contains three structural genes; *gag*, *pol* and *env* that encode for structural proteins needed to produce progeny virions. The *gag* gene codes for the nucleocapsid proteins, *pol* codes for the protease, reverse transccriptase and integrase enzymes and finally the *env* gene, which codes for the envelope proteins, the structure of HIV is depicted in Figure 1.1.1.

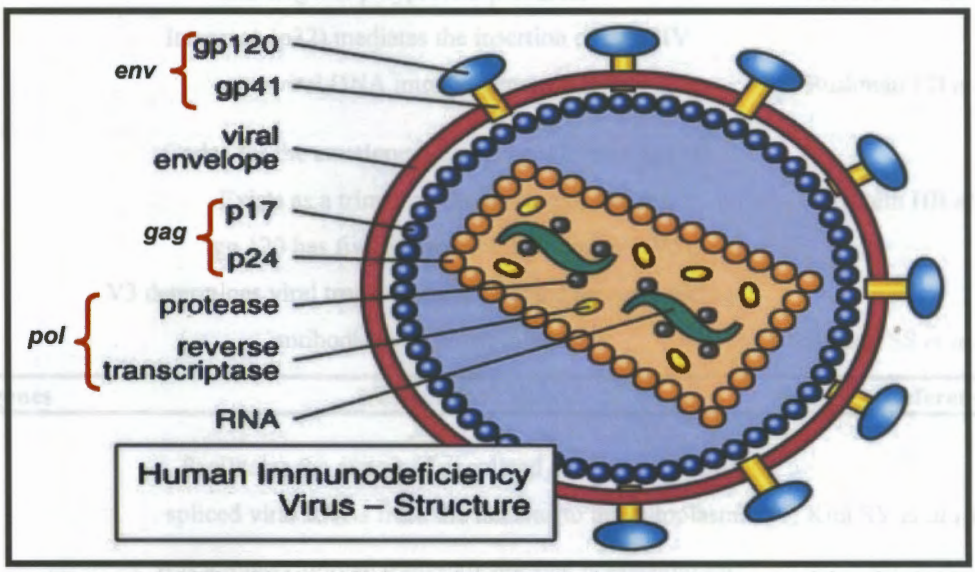


Figure 1.1.1: Physical morphology of the HIV-1 virus (www.avert.org/virus.htm).

In addition to structural genes HIV-1 has several accessory and regulatory genes namely *vif*, *vpr*, *vpu*, *nef*, *tat* and *rev*. The accessory genes *vif*, *vpr* and *vpu* code for proteins that are needed in different maturation steps of the viral particle, affecting the infectivity, replication and assembly (Vaishnav and Wong-Staal, 1991).

Proteins coded by the regulatory genes are needed to regulate different stages of viral replication (Vaishnav and Wong-Stall *et al.*, 1991). The Nef protein is important in pathogenesis, it is a multipotent protein that exerts several functions (increases viral replication, down regulates cell surface CD4 molecules and activates T-cells) (Rhee and March *et al.*, 1991). The Rev protein regulates the amount of spliced versus non-spliced mRNAs and transports viral RNA to the cytoplasm (Sodroski *et al.*, 1986). Tat increases the initiation of transcription (Sodroski *et al.*, 1985). Table 1.1.1 details in summary the functions of all HIV-1 genes.

Table 1.1.1 A: Functions of HIV-1 genes

Structural genes	Functions	References
gag	Codes for the nucleocaspid proteins Matrix (p17), Caspid (p24) and Nucleocapsid (p9 and p6)	Sodroski JG *et al.*,1989
	p17 functions to stabilizing the HIV-1 particle	Gallay P *et al.*, 1995
	p24 forms the conical core of viral particles	Franke EK *et al.*, 1994
pol	Codes for protease (p10), reverse transcriptase (p66/p51) and integrase (p32) enzymes. Reverse transcriptase (p66/p51) is responsible for cDNA synthesis. Protease (p10) plays a role in the cleavage of Gag and Gag-Pol polyprotein precursors.	Ashorn P *et al.*, 1990
	Integrase (p32) mediates the insertion of the HIV proviral DNA into the genomic DNA	Bushman FD *et al.*, 1990
env	Codes for the envelope proteins (gp120 and gp41). Exists as a trimer on the surface of virions.	Bernstein HB *et al.*, 1995
	gp 120 has five hypervariable loops (V1-V5), V3 determines viral tropism and is targeted by neutralizing antibodies that block infectivity	Hwang SS *et al.*, 1999
Regulatory genes	**Functions**	**References**
Rev	Facilitates the export of unspliced and incompletely spliced viral RNAs from the nucleus to the cytoplasm.	Kim SY *et al.*, 1989
Tat	It is a transcriptional transactivator that is essential for HIV replication.	Ruben S *et al.*, 1989

Table 1.1.1 B: Functions of HIV-1 genes

Accessory genes	Functions	Reference
nef	Down regulates CD4 molecules	Garcia JV *et al.*, 1992
	Disrupts T cell activation	Luria S *et al.*, 1991
	Stimulates HIV infectivity	Miller MD *et al.*, 1994
vpr	Plays a role in infecting non-dividing cells	Cohen EA *et al.*, 1990
vpu	Promotes CD4 degradation and enhances virion release	Klimkait T *et al.*, 1990
vif	Essential for the replication of HIV in peripheral blood lymphocytes, macrophages	Strebel K *et al.*, 1987
	Interacts with APOBEG 3 to limit viral DNA quantities	Mangeat B *et al.*, 2003

The integrated form of HIV is known as the provirus and is approximately 9.8Kb (kilobases) in length. Both ends of the provirus are flanked by a repeated sequence known as long term repeats (LTRs). All the HIV genes are located in the central region of the proviral DNA. Figure 1.1.2 depicts the overall genomic organisation of HIV-1 provirus.

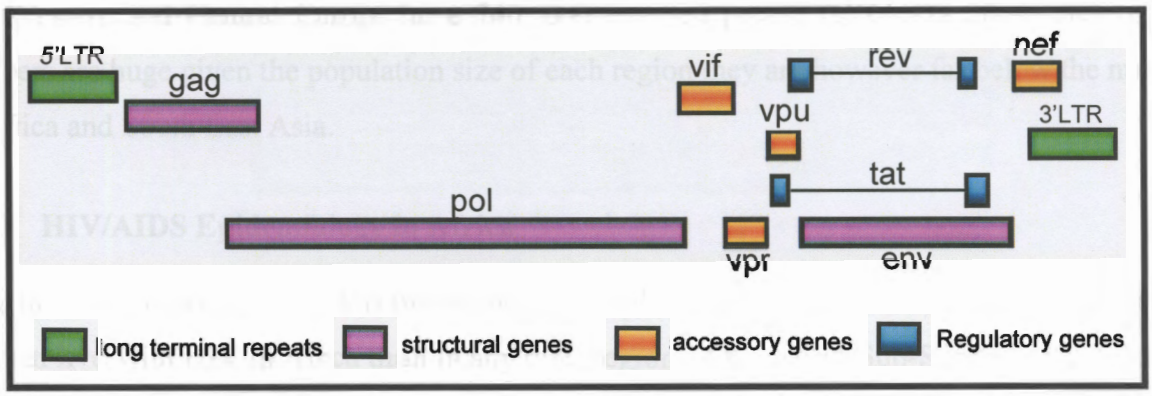


Figure 1.1.2: Genomic organization of HIV-1. The open reading frames for the structural genes (*gag*, *pol* and *env*) are shown together with regulatory and accessory genes. The identical long term repeat (LTR) regions are present at both ends of the provirus.

1.2 HIV-1 Epidemiology

1.2.1 Global Epidemiology

At the end of 2006 the World Health Organisation (WHO) and the Joint United Nations Programme on HIV/AIDS (UNAIDS) estimated that approximately 39.5 million (34.1 - 47.1 million) people were living with HIV. In 2006 alone over 4.3 million (3.6 - 6.6 million) people became infected with HIV-1 and AIDS caused 2.9 million (2.5 - 3.5 million) deaths in the world (UNAIDS 2006). The majority of people living with HIV live in developing world, with over 63% of new infections having occurred in sub Saharan Africa. Figure 1.2.1 shows the number of people living with HIV/AIDS in the world at the end of 2006.

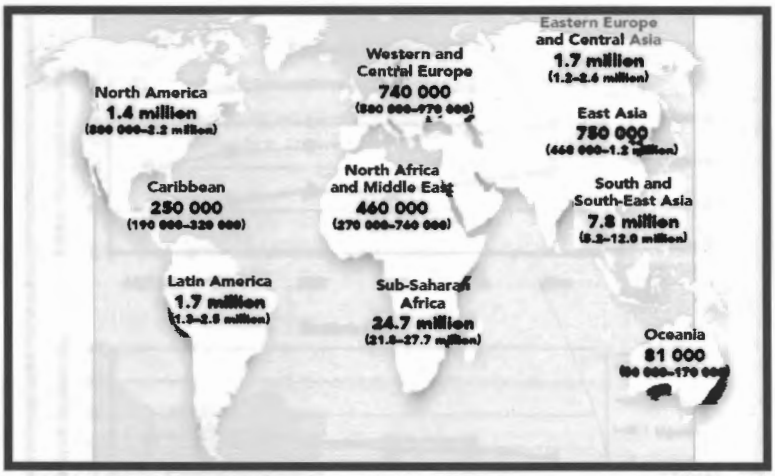


Figure 1.2.1: Global prevalence of adults and children estimated to be living with HIV/AIDS at the end of 2006 UNAIDS (http://www.who.int/hiv/facts/en/).

In South East Asia the HIV epidemic is more recent than in Africa yet the spread is as fast with 7.8 million people estimated to be living with the disease. The least affected countries include Oceania where only 81, 000 people are living with the disease, the Caribbean has 250, 000 infected people and Western and Central Europe have 740, 000 infected people (UNAIDS 2006). While these numbers are huge given the population size of each region they are however far below the numbers in Africa and South East Asia.

1.2.2 HIV/AIDS Epidemiology in Africa

In Africa the spread of HIV-1 has mostly been through heterosexual contact, and more women and children live with HIV in Africa than in any other region of the world. Sub-Saharan Africa has just over 10% of the world's population, but is home to more than 63% of all people living with HIV. An estimated 24.7 million [21.8 - 27.7 million] people are living with HIV in this region (Figure 1.2.1). In 2006, an estimated 4.3 million [3.6 - 6.6 million] people became infected and 2.9 million [2.5 - 3.5 million] adults and children died of AIDS. Among young people aged between 15 - 24 years, an estimated 4.6% [4.2 - 5.5%] of women and 1.7% [1.3 - 2.2%] of men were living with HIV at the end of 2005 (UNAIDS 2005). The severity of the epidemic varies in different parts of the continent (Figure 1.2.2).

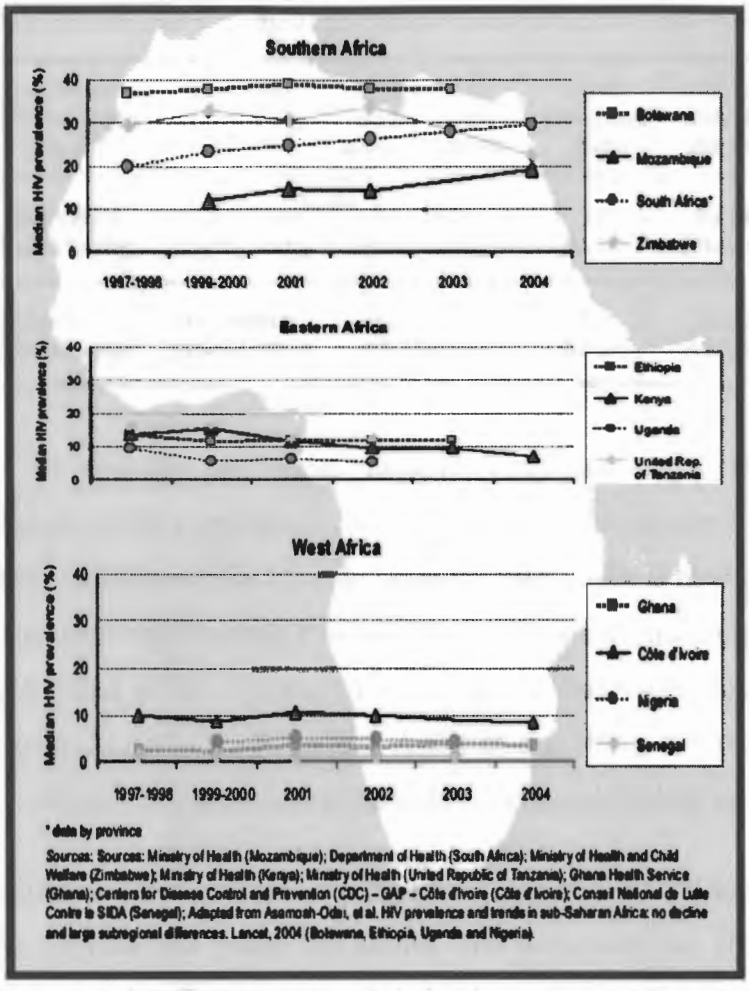


Figure 1.2.2: HIV-prevalence in Africa among pregnant women attending antenatal clinics, 197/98-2004 (UNAIDS 2005).

The West and Central Africa are less severely affected than other parts of the continent and national adult HIV prevalence is yet to exceed 10% in any West African country. Cote d'Ivoire and Nigeria are the worst affected countries with between 3.2 - 9.6 million people living with HIV at the end of 2003 respectively (UNAIDS 2005). Senegal, Togo, Burkina Faso and Mali are the least affected countries with less than 7% prevalence. The low prevalence in infections in this part of the continent is attributed to behavioural change.

Southern African countries have the high prevalence of over 30% in antenatal care clinics observed in Botswana, Mozambique, Swaziland, and South Africa. However, in some southern African countries there has been a decline in prevalence such as Zimbabwe (Gregson *et al.*, 2006) or plateau in prevalence rates, such as in Botswana.

Table1: 1.2.1: HIV/Aids statistics and features in Sub-Sahara Africa during 2003 and 2005 (UNAIDS 2005)

	Adults and children living with HIV	Number of women living with HIV	Adults and children newly infected with HIV	Adult prevalence (%)*	Adult and child deaths due to AIDS
2005	25.8 million [23.8–28.9 million]	13.5 million [12.5–15.1 million]	3.2 million [2.9–3.9 million]	7.2 [6.6–8.0]	2.4 million [2.1–2.7 million]
2003	24.9 million [23.0–27.9 million]	13.1 million [12.1–14.6 million]	3.0 million [2.7–3.7 million]	7.3 [6.7–8.1]	2.1 million [1.9–2.4 million]

East Africa continues to provide the most hopeful indications that serious Acquired immunodeficiency syndrome (AIDS) epidemics can be reversed. The country wide drop in HIV prevalence among pregnant women seen in Uganda since the mid-1990s is now being mirrored in urban parts of Kenya, Tanzania and Rwanda (UNAIDS 2006) where infection levels are dropping. Behavioural change is attributed to this reversal of the epidemic. Sadly new research indicates a possible erosion of the gains Uganda made against AIDS in the 1990s (UNAIDS 2006). In other countries in East Africa however, HIV prevalence has either decreased slightly or remained stable.

In Tanzania the HIV epidemiology is complex, with varying degrees of HIV prevalence between rural and urban areas and between the youth and adults. The socioeconomic factors play a huge role in disease transmission and influences sexual behaviour. Isolated rural areas have lower prevalence compared to overpopulated urban areas. At the end of 2005, 1.4 million (1.3 - 1.6million) people were living with HIV in Tanzania. Mbeya and Iringa are the worst affected regions in the country with HIV infection levels between 15 - 19% in urban areas (Swai *et al.*, 2006). Table 1.2.2 depicts HIV prevalence in different populations situated in different settings around the country where statistics are available.

Table1: 1.2.2: HIV-1 prevalence in Tanzania's different regions and towns

Region /Town	Area	Study Population	HIV-prevalence	Reference
Dar es Salaam	Capital Town (urban)	youth (<25yrs)	20.8% among females 15.5% among males	Chalamila G *et al.*, 2006
Kagena	Bukoba (urban)	females (15 - 24 yrs)	24% in 1987 13% in 1996	
	Muleba (rural)	females (15 - 24 yrs)	10% in 1987 4.3% in 1999	
	Karagwe (rural)	females (15 - 24 yrs)	4.5% in 1987 2.6% in 1999	Kwesigabo M *et al.*, 2005
Kisesa	(rural)	mobile	7.7 %	Kishamane C *et al.*, 2006
Moshi	Moshi (urban)	female bar worker	26.3% in 2002	Kapiga SH *et al.*, 2002
Moshi	Moshi (urban)	female bar workers	19% in 2003	Ao TT *et al.*, 2006
Mbeya	Mbeya town (urban)	female bar workers	68% in 2003	Riedner G *et al.*, 2003
Mbeya	Mbeya town (urban)	female bar workers	from 13.9/100 to 5/100 over 30month period	Riedner G *et al.*, 2006

The population of female bar workers, young people and people who travel to far away places from home constitute a high risk group than the general population. However, most studies indicate a decline in HIV-1 prevalence in the whole population.

1.3 HIV classification

AIDS is caused by two related viruses, human immunodeficiency virus (HIV) type 1 and type 2. Epidemiological analysis indicate that HIV-1 has spread all over the world, while HIV-2 is largely restricted to West Africa and is less pathogenic than HIV-1. This study will only focus on HIV-1.

1.3.1 Classification of HIV-1

HIV-1 is divided into three main groups, the major (M), outlier (O) and the non-M\non-O (N) (Myers *et al.*, 1996; Figure 1.3.1). The group M viruses are responsible for the global HIV pandemic (McCutchan *et al.*, 2000). Group M is divided into nine (n = 9) genetic pure subtypes (A-D, F-H, J, and K), which are approximately equidistant from each other. Subtypes A and F have been further subdivided into sub-subtypes (A1 and A2), and (F1 and F2) respectively this because sub-subtypes A1A2 and F1F2 are more closely related to each other than to other subtypes.

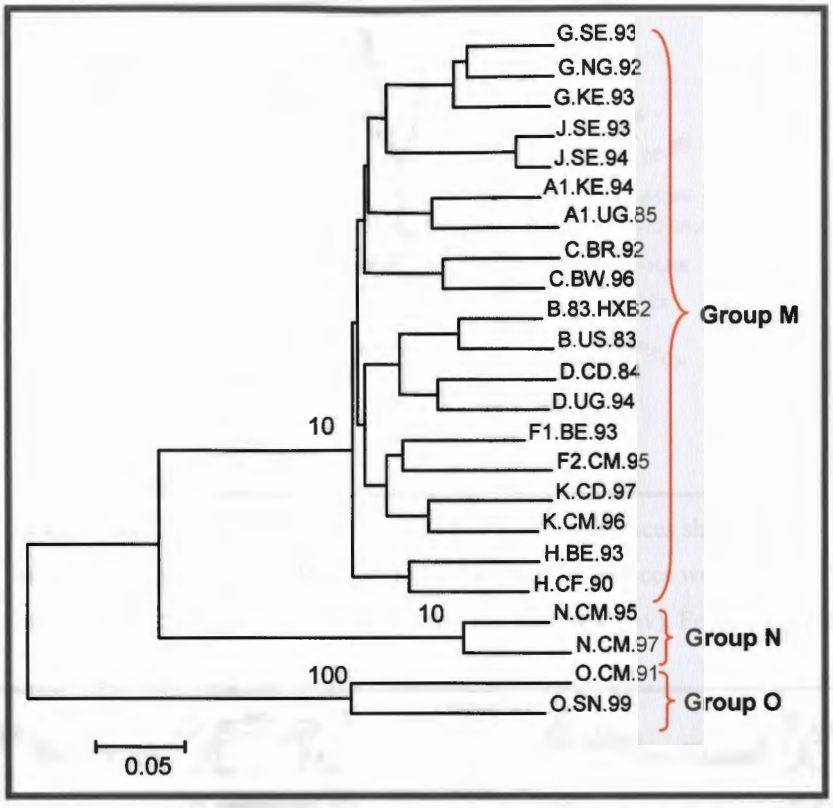


Figure 1.3.1: Neighbour joining tree (envelope region) showing the relationship among HIV-1 groups. Sequences were obtained from the HIV database at Los Alamos (http://hiv-web.lank.gov), February 2006.

In addition to the nine pure subtypes, thirty-four (n = 34) circulating recombinant forms (CRFs) have been described (Figure 1.3.3). CRFs are epidemiologically unlinked mosaic viral strains that share identical break points (Robertson *et al.*, 2000). It has become increasingly important to identify and monitor the spread of CRFs since some epidemics are solely due to CRFs, like in South East Asia where CRF02_AE dominates and in West Africa where CRF02_AG dominantes. Furthermore, hundreds of unique recombinant forms (URF) have been identified. URF refers to a mosaic viral strain that does not share its break points with any other virus. CRFs and URFs are prevalent in regions where multiple subtypes co-circulate in the population (Hoelscher *et al.*, 2002 and McCutchan *et al.*, 2006). The abundance of these mosaic forms is a reflection of the ongoing recombination of HIV and is a consequence of dual infection with more than one subtype. Figure 1.3.2 depicts the phylogenic relationship between pure subtypes and circulating recombinant forms.

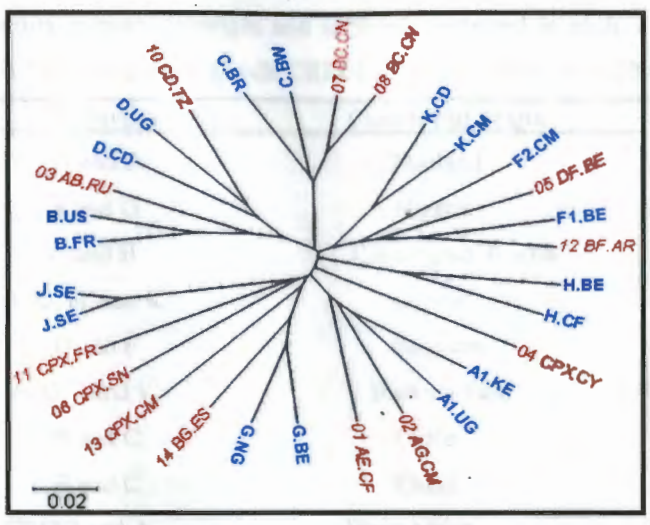


Figure 1.3.2: Phylogenetic tree based on full-length sequences showing all the HIV-1 pure subtypes (blue) and 14 CRFs (red). Sequences were obtained from the HIV database, Los Alamos (http://hiv-web.lank.gov), February 2006.

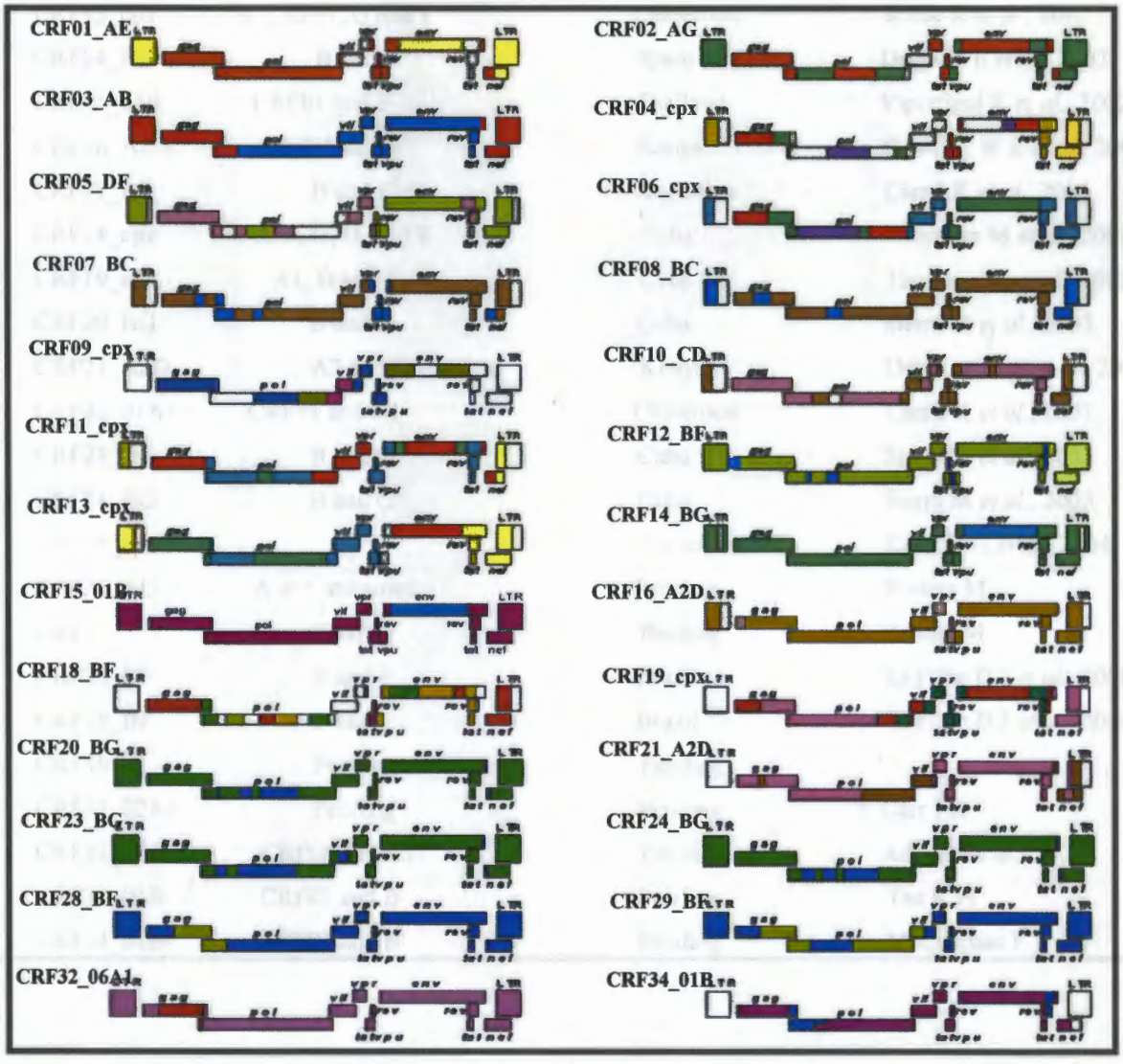


Figure 1.3.3: Mosaic structures of CRFs identified worldwide, the data was obtained from the HIV database, Los Alamos (www.hiv.lanl.gov/content/hiv-db/CRFs/CRFs.html), September 2006.

Table 1.3.1: List of all CRFs their country of origin and subtypes contained in each, data obtained from the HIV database, Los Alamos (www.hiv.lanl.gov/content/hiv-db/CRFs/CRFs.html), September 2006.

CRF	Subtypes	Country of origin	Reference
CRF01_AE	A and E	Thailand	Laukkanen T *et al.*, 1996
CRF02_AG	A and G	Nigeria	Howard T.M *et al.*, 1994
CRF03_AB	A and B	Kiliningrad, Russia	Liitsola K *et al.*, 1998
CRF04_cpx	A, G, H, and K	Cyprus	Gao F *et al.*, 1998
CRF05_DF	D and F	Belgium	Laukkanen T *et al.*, 2000
CRF06_cpx	A, G, J and K	Burkina Faso	Oelrichs R.B *et al.*, 1998
CRF07_BC	B and C	China	Shao Y *et al.*, 2004
CRF08_BC	B and C	China	Piyasirusilp S *et al.*, 2000
CRF09_cpx	CRF02 and A	West Africa	McCutchan F *et al.*, 2004
CRF10_CD	C and D	Tanzania	Koulinska I.N *et al.*, 2001
CRF11_cpx	A, CFR01, G and J	Greece	Paraskevis D *et al.*, 2000
CRF12_BF	B and F	Argentina	Carr J.K *et al.*, 2001
CRF13_cpx	A, CRF01, G and J	Cameroon	Wilbe K *et al.*, 2002
CRF14_BG	B and G	Spain	Delgado E *et al.*, 2002
CRF15_01B	CRF01 and B	Thailand	Viputtigul K *et al.*, 2002
CRF16_A2D	CRF02 and D	Kenya	Dowling W.E *et al.*, 2002
CRF17_BF	B and F	Argentina	Carr J.K *et al.*, 2001
CRF18_cpx	A1, F, G, H, and K	Cuba	Thomson M *et al.*, 2005
CRF19_cpx	A1, D and G	Cuba	Thomson M *et al.*, 2005
CRF20_BG	B and G	Cuba	Sierra M *et al.*, 2005
CRF21_A2D	A2 and D	Kenya	Dowling W.E *et al.*, 2002
CRF22_01A1	CRF01 and A1	Cameroon	Carr J.K *et al.*, 2001
CRF23_BG	B and G	Cuba	Sierra M *et al.*, 2005
CRF24_BG	B and G	Cuba	Sierra M *et al.*, 2005
CRF25_cpx		Cameroon	Kijak G.H *et al.*, 2004
CRF26_AU	A and unknown	Pending	Peeters M
CRF27	Pending	Pending	Peeters M
CRF28_BF	B and F	Brazil	Sa Filho D.J *et al.*, 2006
CRF29_BF	B and F	Brazil	Sa Filho D.J *et al.*, 2006
CRF30	Pending	Pending	
CRF31_02A1	Pending	Pending	Carr J.K
CRF32_06A1	CRF06 and A1	Estonia	Adooan *et al.*, 2005
CRF33_01B	CRF01 and B	Pending	Tee K.H
CRF34_01B	CRF01 and B	Pending	McCutchan F

1.3.2 Global subtype distribution of HIV-1 subtypes and CRFs

The global distribution of HIV-1 subtypes varies from country to country. The continents of North America, Australia and the West-Central Europe have predominantly HIV-1 subtype B (88%) in circulation (Hemelaar *et al.*, 2004). Subtypes B and F and their recombinants are largely found in South America (Carr *et al.*, 2002 and Hemelaar *et al.*, 2004). CRF03_AB is responsible for the epidemic in Russia especially in Kaliningrad (Liitsola *et al.*, 1998 and Hemelaar *et al.*, 2004). In South East Asia CRF01_AE is the dominant (84%) subtype in the population while few infections are due to subtype A (Hemelaar *et al.*, 2004). In India mainly HIV-1 subtype C (97%) is found in the population (Hemelaar *et al.*, 2004). Figure 1.3.4 depicts the global distribution of HIV-1 subtypes and recombinants forms (Hemelaar *et al.*, 2004).

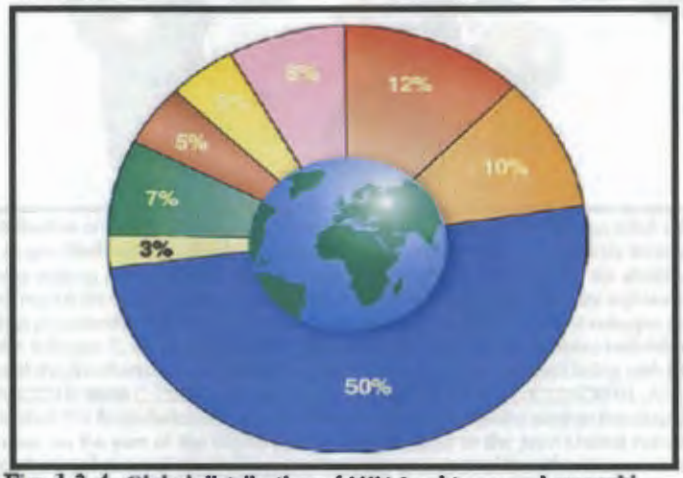


Fig. 1.3.4: Global distribution of HIV-1 subtypes and recombinants in 2004. The number of infections caused by HIV-1 subtypes and recombinants are represented as a proportion of the global total number of individuals living with HIV-1. The colours representing the different HIV-1 subtypes are indicated in the key below. Subtypes F, G, H, J and K were combined (for details see Table 2). A; B; C; D; F, G, H, J, K; CRF01_AE; CRF02_AG; CRF03_AB; other recombinants.

1.3.3 Subtype distribution in Africa

The greatest HIV-1 diversity is found in the continent of Africa, where all pure subtypes and most of the CRFs have been identified. In some parts of the continent of Africa multiple subtypes co circulate in the population simultaneously, while in other regions one subtype dominates. The central part of the continent have the most diverse epidemic, with most known HVI-1 subtypes and their recombinants having been identified in the population (Hemelaar *et al.*, 2004). In the eastern part of the continent (Kenya, Tanzania and Uganda) the epidemic is mainly due to HIV-1 subtypes A-35%, C-25%, D-11% and their recombinants (Dowling *et al.*, 2002, Hoelscher *et al.*, 2002 and Hemelaar *et al.*, 2004).

The western part of the continent is infected by subtypes A-21%, G-35% and CRF_02AG-28% and finally countries in the sub-Saharan region (Botswana, Zimbabwe, Malawi, Mozambique, Angola and South Africa) are predominantly (56%) affected by HIV-1 subtype C (Hemelaar *et al.*, 2004). Figure 1.3.5 depicts the regional distribution of HIV-1 subtypes and CRFs world-wide (Hemelaar *et al.*, 2004).

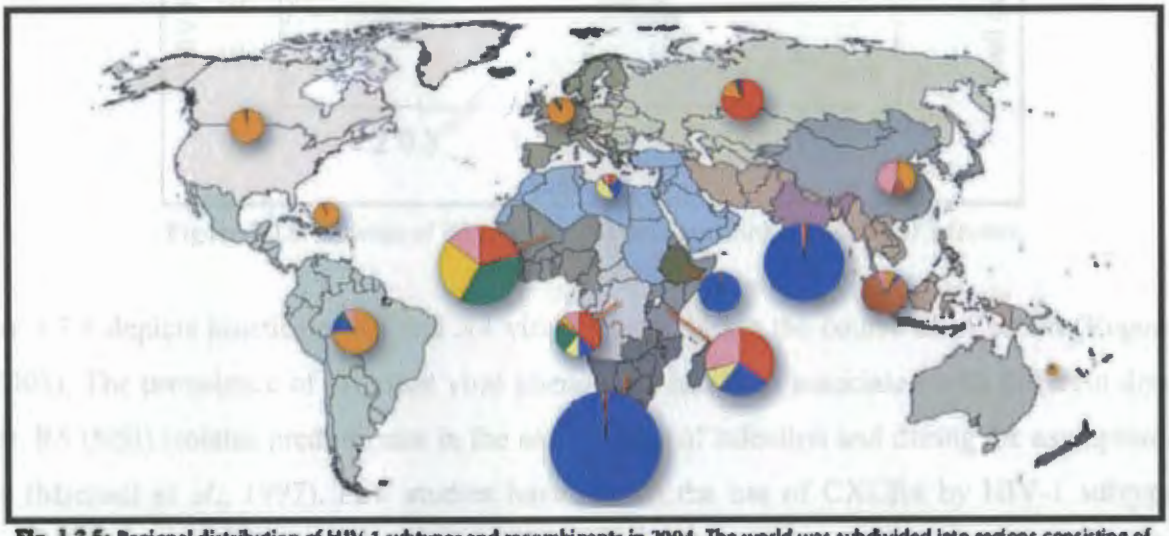


Fig. 1.3.5: Regional distribution of HIV-1 subtypes and recombinants in 2004. The world was subdivided into regions consisting of groups of countries as specified in the methods section and the legend of Table 2. Countries forming a region are shaded in the same colour. Different regions are shaded in different colours. Pie-charts representing the distribution of HIV-1 subtypes and recombinants in each region are superimposed on the regions or connected to the relevant regions by a line. The pie-charts were prepared using the data presented in Table 2. The colours representing the different HIV-1 subtypes are indicated in the key below. The proportions of the subtypes F, G, H, J and K were taken together (for details on these individual subtypes see Table 2). The relative surface areas of the pie-charts correspond to the relative numbers of individuals living with HIV in the regions (see column 2 of Table 1). A; B; C; D; F, G, H, J, K; CRF01_AE; CRF02_AG; CRF03_AB; other recombinants. Disclaimer: The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization or the Joint United Nations Programme on HIV/AIDS (UNAIDS) concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.

1.3.4 Biological characteristics of HIV-1

HIV-1 strains can be classified into syncytium inducing (SI) and non-syncytium inducing (NSI) variants, based on their capacity to induce syncytia in cell culture. Viral tropism and syncytium induction are largely determined by the sequence of the V3 loop of the viral gp120 glycoprotein. A change in the amino acid sequence of the V3 loop can lead to a change in viral tropism (Fouchier *et al.*, 1992). HIV-1 use chemokine receptors, in addition to the CD4 molecule to enter T-cells (Berger *et al.*, 1999). The two major chemokine receptors are CCR5 and CXCR4 (Dragic *et al.*, 1996). Viruses that use the β-chemokine cell receptor 5 (CCR5) and are referred to as R5 isolates, whereas viruses that use the CXCR4 co-receptor (Berger *et al.*, 1997) and are referred to as X4 isolates (Berkowitz *et al.*, 1998). Some HIV-1 isolates are capable of using both co-receptors and are referred to as R5X4 viral isolates. Over the course of infection, the co-receptor usage of HIV changes from CCR5 to CXCR4 in approximately 50% of infected individuals (Connor *et al.*, 1997 and Berger *et al.*, 1999).

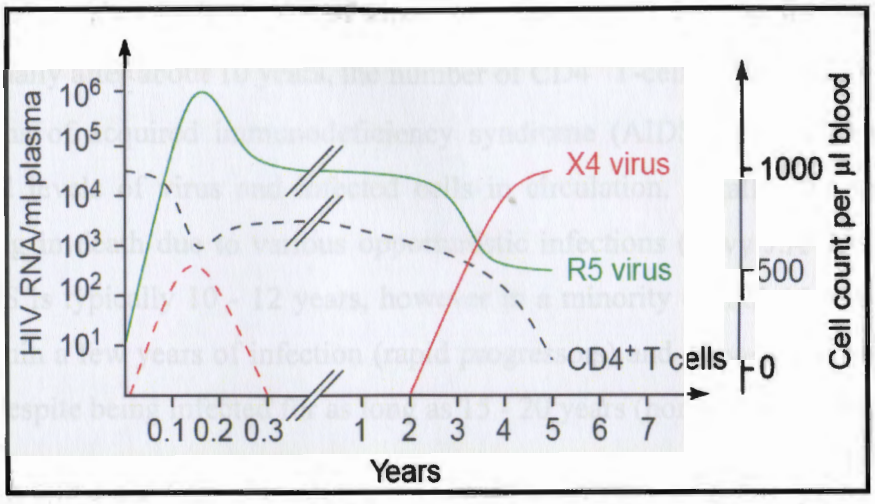


Figure 1.3.6: Kinetics of R5 and X4 viral isolates during the course of infection.

Figure 1.3.6 depicts kinetics of R5 and X4 viral isolates during the course of infection (Regoes *et al.*, 2005). The prevalence of different viral phenotypes has been associated with different disease stages. R5 (NSI) isolates predominate in the early stages of infection and during the asymptomatic phase (Michael *et al.*, 1997). Few studies have shown the use of CXCR4 by HIV-1 subtype C viruses, Cilliers *et al.*, (2003) reported on the use of both co-receptors by viruses isolated from patients at an advanced AIDS stage. The reason for R5 virus dominance is still a matter of speculation and several theories are been tested. X4 (SI) isolates on the other hand, are only prevalent during the terminal stages of the disease and are associated with decline in CD4 numbers and the onset of AIDS (Fauci *et al.*, 1996).

1.4 HIV-1 Pathogenesis

1.4.1 Natural history

The infection of humans by HIV can be categorised in three phases: acute infection, chronic infection and AIDS. The acute infection phase occurs within several weeks after infection and may be associated with clinical symptoms of acute infection which is characteristically a systemic febrile illness resembling acute mononucleosis (Schacker *et al.*, 1996). During this period there is extensive viraemia and large numbers of infected $CD4^+$ T-cells. The acute phase of HIV infection is mainly characterized by massive infection and depletion of the mucosal-associated lymph tissue (MALT, Veazey *et al.*, 2003). For 2 - 3 weeks after initial infection HIV replication in activated $CD4^+/CCR5^+$ memory T-cells proceeds relatively unchecked resulting in large-scale potentially irreversible depletion of these cells (Derdeyn and Silvertri, 2005). Following acute infection, there is the onset of antiviral immune responses and a concomitant decline in the amount of circulating viruses leading to a clinically latent phase or chronic infection which has variable duration.

Typically this is associated with control of virus replication and a gradual declining numbers of $CD4^+$ T-cells. Finally after about 10 years, the number of $CD4^+$ T-cells declines to very low values and the symptoms of acquired immunodeficiency syndrome (AIDS) appear, coincidental with sharply increased levels of virus and infected cells in circulation. Finally, the immune system collapses resulting in death due to various opportunistic infections (Levy J.A 1993). Time from infection to AIDS is typically 10 - 12 years, however in a minority of individuals progression to AIDS occurs within a few years of infection (rapid progressors) and conversely, some individuals do not progress despite being infected for as long as 15 - 20 years (non-progressors).

1.4.2 Viral load and CD4

The viral dynamics during primary HIV-1 infection is characterized by the following sequence of events: the first viraemic phase of increasing viral load occurs for one to four weeks after infection during that period a viral peak is reached. This phase is followed by a rapid decline in the viral load, until a 'steady state' is reached. During this time there may be a gradual increase in viral load which increases sharply with the onset of AIDS. There is correlation between the steady state viral load (set point) and the duration of disease progression (Lyles *et at.*, 1999 and Sterling *et al.*, 2001).

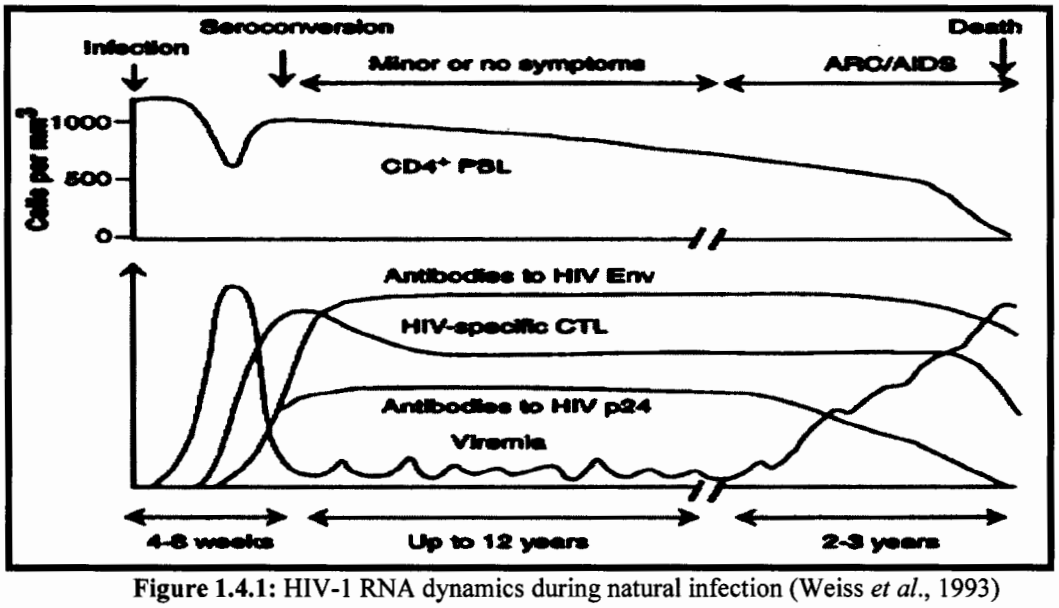


Figure 1.4.1: HIV-1 RNA dynamics during natural infection (Weiss *et al.*, 1993)

Plasma viral load and $CD4^+$ cell count are central to management of HIV-1 infection as these biological markers have been demonstrated to predict mortality among HIV-infected individuals (Mellors *et al.*, 1995 and Sterling *et al.*, 2001). While most studies on the prognostic value of viral load have been done in Europe and Africa, the relationship between viral load and disease progression has also been demonstrated in a prospective study of Kenyan women followed up

from time of infection, viral set point was strongly associated with mortality (Lavreys *et al.*, 2006). Women with a viral set point of >5 log copies/ml had a median survival duration of 7.1 years compared to women with 4 log copies/ml viral load that had a survival duration of 8.7 years. This longitudinal study of Kenyan women also revealed an association between higher initial $CD4^+$ cell count with decreased mortality and finally acute HIV-1 illness predicted mortality independent of plasma viral load.

Some studies have shown differences between subtypes and disease progression, in a study by Vasan *et al.*, (2006) in Tanzania viral heterogeneity in the rates of disease progression was observed. Patients with subtype D experienced the most rapid progression to death relative to patients infected with subtype A (HR, 2,27; 95% CI, 1.46 - 3.52) or to the World Health Organization stage 4 of illness (HR, 1.94; 95% CI, 1.20 - 3.14) and to a CD4+ cell count of <200 cells/ml (HR, 2.12; 95% CL, 1.42 - 3.17; Vasan *et al.*, 2006).

1.4.3 Host factors

Viral replication usually requires that innate intracellular lines of defence of the host be overcome, a task usually accomplished by specialized viral gene products. The host on the other hand has numerous inborn mechanisms for protection against HIV-infection and dissemination. T-cells express apolipoprotein B mRNA-editing enzyme-catalytic, polypeptide-like 3G (APOBEG3G) a member of a family of cytidine deaminases that convert cytidines to uridines in mRNA (Sheehy *et al.*, 2002). APOBEG3G works as a deaminase and exerts its antiviral effects during reverse transcription to trigger G-to-A hypermutation in the retroviral DNA (Teng *et al.*, 2002). In its attempt to counter the APOBEG3G antiviral effects, HIV uses the virion infectivity factor (Vif) protein (Sheehy *et al.*, 2002).

HIV *vif* protects the virus by preventing the incorporation of APOBEG3G into virions, by rapidly inducing its ubiquitination and proteasomal degradation (Xu *et al.*, 2004). Previous studies have demonstrated marked decrease in infectivity by HIV-1 with a defective *vif* protein (Mangeat *et al.*, 2003). In a study by Xu *et al.* (2004) a human APOBEG3G resistant to depletion induced by HIV *vif* was identified, this protein had a single D128K mutation. It was postulated that the D128K mutation reduces the HIV-1 *vif* binding affinity.

Host genetics play an essential role in HIV infection and determines the rate of disease progression. Certain polymorphisms and deletions have been reported to influence host susceptibility to HIV infection and the pathogenesis of the disease.

The most studied of these include the CCR5Δ32, 1% of the Caucasian population carries the homozygous allele for this mutation (Liu *et al.,* 1996). The deletion of the 32 amino acids in the coding region of the CCR5 genes results in the lack of expression of this receptor. CCR5 is the member of the 7-transmembrane G-protein coupled family, its functions as a co-receptor during viral entry (Raport *et al.,* 1996). Following the binding of gp120 to cellular $CD4^+$ conformational changes take place to an able the binding of CCR5 (Deng *et al.*, 1996) resulting in viral entry. Several studies have demonstrated 'resistance' to HIV infection by individuals carrying homozygous alleles of CCR5 (Dean *et al.*, 1996).

Individuals heterozygous for this mutation on the other hand have been shown to have slower rates of disease progression compared to wild type carriers. A study by Winkler *et al.* (1998) reported on the ligand polymorphism (SDF1-3'A) which delays disease progression. SDF1-3'A leads to upregulation of SDF-1 thought to be the natural ligand of CXCR4. CCR2b polymorphisms, in addition to SDF-1 variants are associated with slow progression. In addition to polymorphisms on the actual co receptors, numerous polymorphisms on the CCR5 promoter region have been reported. Some have been reported to accelerate disease progression (Mummidi *et al.*, 1998) while others have been reported to slow disease the rate of disease progression (McDermott *et al.*, 1998 and Samson *et al.*, 1997).

1.5 Immunological factors

A number of studies have shown that HIV-specific CTL responses are associated with reducing viral replication from the peak to the set point (Borrows *et al.*, 1994 and Ogg *et al.*, 1999). There is limited evidence of the role of antibodies in controlling acute infection however recent findings have shown that early neutralizing antibody responses are raised against HIV-1 *in vivo* (Richman, Wrin *et al.*, 2003; Wei, Decker *et al.*, 2003 and Li, Decker *et al.*, 2006). Immune responses place selective pressure on the infecting virus resulting in viral diversification and immune escape. Neutralizing antibody responses are thought to exert 'soft' selection pressure (Frost *et al.*, 2005) by affecting the relative fitness of different strains present in the viral quasispecies. In contrast, cellular immune responses may exert 'hard' selection pressure (Frost *et al.*, 2005), in which the absolute fitness of the population is affected suggested by the correlation of CTL responses and class I HLA alleles (Novitsky *et al.*, 2003).

Viral evolution is a consequence of multiple factors, among them immune pressure exerted by both the neutralizing antibodies (nAbs) and cytotoxic T-lymphocytes (CTLs). HIV has the potential to evolve at an extremely high rate due to the high mutation rate (Mansky *et al.*, 1995) and a large

population of productively infected cells within the host (Chun *et al.*, 1997). Evolution of escape to immune responses may also play a role in driving the rapid diversity of the viral population. Escape from cellular immune response may occur through a combination of point mutations, changes in glycosylation patterns and insertions and deletions (Frost *et al.*, 2005).

A close temporal relationship between the levels of circulating virus-specific CTLs and HIV-1 viral load was reported by Price *et al.*, (1999). CTLs exerts immune pressure on the virus leading to viral evolution as a mechanism of immune escape. Higher detection of synonymous substitutions compared to non-synonymous substitutions reflects negative selection favouring the conservation of protein structure and function (Price *et al.*, 1999). On the other hand, higher non-synonymous substitution than synonymous substitutions leads to positive selection, which in turn leads to viral evolution resulting in immune escape.

1.5.1 HLA (Human Leukocyte Antigen) restricted immune responses

The major histocompatibility complex (MHC) has been mapped to the short arm of chromosome six and contains HLA class I and class II genes (Mungall *et al.*, 2003). HLA class I molecules are coded for by three HLA genes, HLA-A, HLA-B and HLA-C. Class I molecules bind to antigenic epitopes derived from intracellular pathogens that are then presented to $CD8^+$ T-cells. Previous studies have revealed the existence of both advantageous and deleterious HLA class I alleles (Scherer *et al.*, 2004). The two most prominent HLA markers associated with delayed HIV-1 disease progression in subtype B infections are the B*27 and B*57 alleles (Goulder *et al.*, 1997). Alleles B*35 and B*53 on the other hand have been reported as markers of rapid HIV-1 disease progression, primarily in Caucasian cohort studies (Gao *et al.* 2001). Similary associations have been found in African populations where Tang *et al.*, (2002) screened 259 subtype C infected individuals from Zambia for favorable and unfavorable HLA class I alleles. The results revealed an association between low viral load and two HLA-B alleles, HLA B*39 and HLA B*57.

In a study from Mbeya on a cohort of HIV-seropositive bar workers, possession of HLA class I alleles B*5801, B*8101 and B*0702 was associated with low median viral load (Geldmacher *et al.*, 2007). In contrast to previous studies among subtype B infected individuals neither B*35 nor B*53 showed a clear disadvantage among subtype C infected individuals (Tang *et al.*, 2002). A recent study found supportive evidence that at a population level, the rate of disease progression is strongly associated with HLA-B expression (Kiepiela *et al.*, 2004). This study also found that

A*30-Cw*03 was associated with lower viral loads. In this study, higher viral loads in subtype C infected individuals were associated with A*02-Cw*16, A*23-B*24 and A*23-Cw*07.

1.5.2 Cytotoxic T-Lymphocyte response

Class I MHC restricted anti-HIV cytolytic T lymphocytes (CTLs) might be key in the successful control of HIV infection (Borrow *et al.*, 1994). A strong CTL response is often associated with better viral control and slower disease progression (Ogg *et al.*, 1999). $CD8^+$ T-cells have several functions, including cytolysis and production of cytokines (INFγ, TNFα and IL-2) and chemokines (MIP-1β). Infected cells present viral peptides (epitopes) in the cleft of the MHC class I molecules on their surface which are recognized by CTLs, this in turn triggers the immune system to kill all the infected cells (Berke *et al.*, 1995).

The important role played by CTLs in disease control is supported by evidence from several observations and correlative studies; the presence of $CD8^+$ T-cells during SIV infection leads to decreased viral replication and slower disease progression in rhesus macaques (Jin *et al.*, 1999). The resolution of acute viraemia was found to occur simultaneously with a major expansion of HIV specific $CD8^+$ T-cells, immunologic pressure mediated by SIV or HIV specific $CD8^+$ T-cells is often manifested by viral escape mutations (Goulder *et al.*, 1997). However, the $CD8^+$ T-cell response to HIV in most infected patients is insufficient to maintain durable control of viral load and disease progression (Betts *et al.*, 2006).

The CTL response to HIV challenge is often targeted to the Gag and Nef proteins, the significance of this choice remains a matter of speculation. Nef specific responses are considered important in HIV infection and are said to be the earliest responses seen in acute HIV infection (Alter *et al.*, 2002). The preferential targeting of Gag epitopes, may be an important marker of immune efficacy, since preferential targeting of Gag was associated with a lower viral load among subtype C infected individuals (Masemola *et al.*, 2004; Kiepela *et al.*, 2006 and Christof *et al.*, 2007).

More important than the genes targeted, is has been shown that the functional profile of HIV specific $CD8^+$ T-cells is the most important factor for viral control. In a study between progressors and non-progressors, it was found that among progressors the functional profile of HIV specific $CD8^+$ T-cells was limited compared to that of non-progressors, who consistently maintained highly functional $CD8^+$ T-cells (Betts *et al.*, 2006). Rather than quantity, the quality of the $CD8^+$ T-cell functional responses serves as an immune correlate of HIV disease progression (Betts *et al.*, 2006).

1.5.3 Neutralizing antibody defense

Neutralizing antibodies (Nabs) against the viral envelope proteins (Env) of HIV provide the first line of adaptive immune defence, by blocking infection of susceptible cells (Moore, *et al.*, 1996). The detection of HIV-specific Nabs has been reported as early as 52 days post infection (Wei *et al.*, 2003). However, their impact in protection against HIV is limited, primarily due to the variability of the *env* gene which has up to 35% sequence diversity between subtypes and 20% sequence diversity within subtypes (Shankarappa *et al.*, 1999).

Few broadly neutralizing human monoclonal antibodies (MAbs) have been isolated from HIV sero positive subtype B infected individuals (Parren *et al.*, 1999). These MAbs neutralize many primary HIV-1 isolates from varying subtypes. One group of MAbs recognize epitopes located on the gp120 surface unit of the Env spike. This group consists of two examples MAb b12, which is directed against an epitope overlapping the CD4 binding site (Roben *et al.*, 1996) and MAb 2G12, which recognizes a unique epitope in a carbohydrate-rich region on the outer domain of gp120 (Calarese *et al.*, 2003). Another group of MAbs recognize epitopes located on the membrane proximal external region of the gp41 transmembrane protein. This group consists of three examples, MAb 2F5 has been mapped to a region overlapping the conserved sequence ELDKWA (Parker *et al.*, 2001). MAbs 4E10 and Z13 recognize an epitope involving the sequence NWF(D/N)IT located adjacent to 2F5 epitope (Zwick *et al.*, 2001). These MAbs have been shown to provide protection against HIV challenge in passive transfer studies in macaques (Mascola *et al.*, 2002).

Other anti-Env MAbs of significant interest include MAbs against the gp120 V3 loop. The V3 loop is highly variable and tends to induce type-specific antibodies with limited breadth. MAb 447-52D is a V3 loop-specific MAb, which recognises the GPGR motif at the tip of the V3 loop of subtype B isolates (Gorny *et al.*, 2004). The last group of MAbs recognize discontinues epitopes that overlap the conserved co-receptor binding site of gp120 to CD4, the group is made up of MAbs X5 and 17b (Labrijn *et al.*, 2003).

In a comprehensive study by Binley *et al.*, 2004 gp120 MAbs showed greater activity against subtype B viruses than non-subtype B viruses. The anti-gp41 MAbs showed broader inter-subtype activity. In this study MAbs 2G12 and 2F5 did no neutralize subtype C viruses. To date there is no consistent evidence that HIV positive sera preferentially did not neutralize autologous viruses more effectively than they do viruses from other subtypes (Beirnaert *et al.*, 2000).

1.6 Genetic variation and HIV-1 diversity

HIV has a vast evolutionary potential in its host which could be attributed to its high mutation rate (10^{-4} - 10^{-5} mutations per nucleotide per replication cycle) associated with replication (Mansky and Temin, 1995), its rapid viral turnover rate (10^{8} to 10^{9} virions per day) (Ho *et al.*, 1995), availability of numerous target cells (10^{7} to 10^{8}) (Chun *et al.*, 1997), the high levels of recombination (Jung *et al.*, 2002) and the host's selective forces which determine a continuous process of intrahost virus evolution and diversification that gives rise to variants with different biological properties (Shankarappa *et al.*, 1999).

HIV-1 displays enormous genomic plasticity with very few nucleotide positions conserved across the full-length genome (Coffin *et al.*, 1995) and thus can exist as swarms termed quasispecies (Eigen *et al.*, 1998). The increase in HIV-1 diversity is related to increased HIV-1 replication capacity (fitness) and pathogenesis (Troyer *et al.*, 2005). This continual increase in genetic diversity enables the virus to rapidly adapt to a variety of pressures, these include escape from humoral (Richmond *et al.*, 2003) and cytotoxic T-lymphocyte (Brander et al., 2003) immune responses and the ability to be resistant against antiretroviral drugs (Johnson *et al.*, 2003).

There are two main mechanisms for HIV-1 to establish genetic variation: mutations and recombination. Replication of a retrovirus includes a reverse transcription step carried out by the reverse transcriptase enzyme (RT). During this process the single stranded RNA genome of the virus is converted to a double stranded DNA copy. Most mutations arise during this step, this is due to the lack of the 3'-5' exonuclease activity for proofreading of the RT which renders this enzyme error-prone than the cellular DNA polymerase (Preston, *et al.*, 1988). Recombination on the other hand is a consequence of the RT switching templates during reverse transcription (Goodrich and Duesberg, 1990). The switching between two templates results in a formation of a mosaic genome with characteristics from both templates.

1.6.1 Dual infection

HIV-1 dual infection can be the result of coinfection with two variants infecting at or near the same time before seroconversion. Dual infection can also result in superinfection, in which case primary infection with one variant precedes infection with a second variant some time after seroconversion (Gottlieb *et al.*, 2004). Dual infection can be defined as an infection by two strains that are no more closely related to one another than to any epidemiologically unlinked strain

(Gottlieb *et al.*, 2004). Dual infection can be with strains of the same subtype or with viruses of different subtypes (Diaz, *et al.*, 1994).

Recent studies have shown an association between dual infection and an elevated viral load set point in subtype C dually infected individuals (Grobler *et al.*, 2005) and increased disease progression (Gottlieb *et al.*, 2005). Dual infection has implications for HIV treatment, transmission and vaccine development (Goulder *et al.*, 2002), it might also an enable the viruses to efficiently escape immune surveillance.

1.6.2 Mechanism of recombination

Genetic recombination is part of the normal mechanisms of retroviral replication and as such, plays an important role in the generation of viral diversity (Najera *et at.*, 2002). Generation of recombinant retroviruses requires that two viruses infect a single cell, either simultaneously, by a single transmission event, or sequentially, in multiple transmission events. In HIV-1, recombination can occur between different strains of the same subtype (intra-subtype recombination) or different subtypes (inter-subtype recombination). HIV-1 recombinant forms are much more prevalent, geographically spread and diverse than previously thought.

HIV-1 is a diploid retrovirus, its genome consists of two RNA strands that undergo reverse transcription during its life cycle. If a cell is dually infected with two distinct HIV-1 strains, genomic RNA strands representing two different subtypes can be packaged into new progeny virions. When a virus with this kind of genome infects another cell and recombination occurs between the two strands, a mosaic genome is formed.

Recombination allows the progeny virus to acquire genetic material from its parental viruses and form a new divergent viral genome faster than through point mutations. Recombination reshuffles point mutations between quasispecies of individuals with single infection but gains new significance when individuals become infected with more than one HIV-1 strain or subtype. Recombinant viruses are becoming increasingly common and have been found in areas with high prevalence of multiple genetic subtypes, such as in Africa where a number of CRFs have been identified. The role and spread of CRFs is discussed in detail under section 1.3.1.

1.6.3 Superinfection and HIV-1 pathogenesis

HIV-1 superinfection refers to the acquisition of another strain by an individual with an established infection (McCutchan *et al.*, 2005). Superinfection studies in animals was found to be possible in studies performed in chimpanzees as 1987 which showed that HIV subtype B infected chimpanzees could be superinfected with a second strain of HIV (Fultz *et al.*, 1987). In human studies, the first proven case of superinfection was identified in an individual who was initially infected with CRF01_AE and later acquired subtype B (Jost *et al.*, 2000). The next reports described superinfection with the same subtype (subtype B; Altfeld *et al.*, 2002 and Koelsch *et al.*, 2003). All these three cases where identified in men having sex with men. Superinfection in Thailand was identified in two injecting drug users (IDU), in the first case, subtype B was acquired after an initial infection with CRF01_AE.

In the second case CRF01_AE was acquired an initial infection with subtype B (Ramos *et al.*, 2002). In the first case of superinfection identified in Africa, a recombinant AC viral strain was acquired after the initial infection by a complex ACD viral strain in an individual from Mbeya, Tanzania (McCuthan *et al.*, 2005).

These data suggest that primary HIV-1 infection does not prevent secondary infection by another viral strain. The reasons could be immunological, in that due to primary infection the immune system could be eroded by the time superinfection occurs. Other reasons could be virological, the ability of the secondary viral strain to evade the established immune response. The high prevalence of unique recombinant virus forms in regions where multiple viruses circulate suggest that superinfection is more common than originally thought, and that infection with one strain does not necessarily protect from infection with a second strain. Superinfection can occur through various routes including drug injection, heterosexual and homosexual contact. In all cases reported, superinfection was found to be associated with increased viraemia and thus increase in disease progression. Vaccine development strategies will need to take into account the phenomenon of superinfection, since protective CTL targeting against HIV-1 may vary not only by subtype but even by individual strains (Yang *et al.*, 2005).

1.7 Study scope and rationale

This study is based in Mbeya, Tanzania which is currently involved HIV clinical trial research. In this region, three different HIV subtypes A, C and D co-circulate in the population (Hoelscher *et al.*, 1998). In addition, numerous unique recombinant forms (URFs) have also been reported in the Mbeya region (Hoelscher *et al.*, 2001). In a study of 36 blood donors from this region, dual infection with more than one subtype was detected in 40% of the individuals (Arroyo *et al.*, 2004). The presence of multiple subtypes and recombinant viral strains in the population poses a challenge for vaccine development.

The overall aim of this study is to describe HIV subtype distribution in Mbeya and to investigate viral dynamics and mechanism of immune escape in individuals infected with more than one subtype.

The specific aims of this study were:

(i) To characterize subtype distribution in a cohort from Mbeya by multi region hybridization assay (MHA).

(ii) To monitor viral dynamics in individuals infected with a single or multiple HIV subtypes

(iii) To investigate the extent and mechanism of immune escape over time in a dually infected individual.

Most vaccines contain multiple gene inserts and it is important to have knowledge of subtype designation across the genome. While full-length genome sequencing remains the gold standard, this is both time consuming and expensive. In this study we evaluated the multi-region hybridization assay (MHA) capable to genotype the HIV-1 genome in five different regions, and have applied this method to characterize subtype distribution in the regions. The multi-region hybridization assay (MHA), is based on the Taqman real-time PCR format, where subtype specific exonuclease probes were used to genotype the HIV-1 genome in five different regions (Hoelscher *et al.*, 2001). This assay, accurately genotypes HIV-1 strains and can detect dual infections and HIV-1 recombinant genomes.

High risk behaviour among individuals in settings with high HIV-1 prevalence and multiple subtypes in circulation, are exposed to possible HIV-1 multiple infections. The high prevalence of recombination observed in the region is a consequence of dual infection taking place in the population. Dual infection illustrates the inability of the immune system to protect against subsequent infection. The dynamics of viral populations in dually infected individuals is not well characterized and data on the impact of dual infection on disease outcome is limited.

In a subsample of twelve individuals with either dual infection or single infection, this study aimed to investigate the dynamics of dual infections compared to single infection in a high-risk female cohort in Mbeya Town, Tanzania followed over a 21 month period. Lastly, to understand the implications on dual infection on virus evolution and the mechanism of immune escape, we performed a detailed analysis on viruses from an individual infected with A and C subtypes respectively.

This study forms part of the HIV-superinfection study (HISIS) which was funded by the European Commission Framework 5 INCO Programme. The primary aim of this project was to understand biological factors that lead to HIV superinfection which occur when an HIV infected individual becomes re-infected with a second HI-virus. This study aims to provide insights into correlates of protection which are important to the development of an effective HIV vaccine. HISIS is made up of several international institutions that include the following; Institute for Infectious Diseases and Molecular Medicine, University of Cape Town, S.A; Mbeya Medical Research Programme, Tanzania; Dept Infectious Diseases and Tropical Medicine, Ludwig Maximilian University-Munich, Germany; National Public Health Institute, Helsinki-Finland and the Henry M. Jackson Foundation, Maryland USA. All projects are funded by the European Union.

CHAPTER 2

DISTRIBUTION OF HIV-1 SUBTYPES IN MBEYA AS IDENTIFIED BY THE MULTIREGION HYBRIDIZITION ASSAY (MHA)

2.1 Introduction

The Mbeya Region is located in the Southwest of Tanzania and is bordered by three countries Zambia, Malawi and Mozambique (Fig. 2.1.1). Mbeya is the major town in the region and is located along the trans-African Highway that connects the capital Dar es Salaam with Malawi and Zambia. Mbeya is the third largest town in Tanzania and has a population of approximately 800,000 people. In a study of 3 096 volunteers recruited between 2002 and 2003 from urban areas and a nearby rural village in Mbeya, the HIV prevalence was found to be 16.6% with a range of between 15.3% and 17.9% (Arroyo *et al.*, 2005). To facilitate research in the region the Mbeya Medical Research Programme (MMRP, www.mmrp.org) was established as a partnership between the Mbeya Referral Hospital, the Mbeya regional medical office, the National institute for Medical Research, Department of Infectious Diseases and Tropical Medicine at University of Munich, and the US Military HIV Research Program. The MMRP is currently working on HIV, TB and malaria related research and is currently participating in a Phase II trial to evaluate safety and immunogenesity of a multigene (*gag*, *pol* and *env)* and multiclade (A, B and C) DNA prime Adenovirus 5 boost vaccine. The vaccine is developed by the Vaccine Research Centre, NIH-USA (VRC HIV DNA 016-00-VP prime boosted by VRC-HIV ADV014-00-VP) (ww.mmrp.org).

Fig. 2.1.1: Geographical location of Mbeya Town, Tanzania. Represented in red are the Capital, Dar es Salaam and Mbeya town joined by the trans-African Highway (www.tropinst.med.uni-muechen.de/ag/hivaids/hisis-brochure.htm).

This chapter describes a study characterizing subtypes circulating in the Mbeya region and is part of the HIV superinfection study (HISIS; see section 1.7) (www.tropinst.med.uni-muechen.de/ag/hivaids/hisis-brochure.htm HISIS network) which is part of the MMRP.

Understanding HIV genetic diversity in the region is important for informing vaccine trials and provides information on studies aimed at determining the impact of HIV-1 genetic diversity on vaccine efficacy.

Tanzania presents a genetically complex HIV epidemic with three subtypes (A, C and D) circulating in the population, together with numerous unique recombinant forms (URFs) (Hoelscher *et al.*, 2002 and Arroyo *et al.*, 2004). In particular, Mbeya which is located at the cross roads of Zambia, Malawi and Tanzania along the trans-African Highway is a meeting ground of individuals potentially infected with viruses from different regions of Africa. A study by Hoelscher *et al.*, (2001) using full-length sequencing of nine samples showed that two of the nine individuals were infected with pure subtype A viruses, two individuals were infected with pure C viruses and five individuals were infected with unique recombinant forms (URFs) comprising of subtypes A, C and D. Four of the five URFs had A and C genome sequences and one was a CD recombinant. This study provided evidence of co-circulation of pure and recombinant HIV-1 forms in the population. These results were supported by a study performed by Arroyo *et al.*, (2004) who characterized viruses among blood donors (n = 36) in Mbeya by full-length sequencing and multi-region hybridization assay (MHA). From that report 11/20 (55%) viruses were pure C, 3/20 (15%) A, 1/20 (5%) D and 5/20 (25%) were recombinant, 4/20 (20%) of the recombinant viruses contained subtype C, 3/20 (15%) of the recombinants were AC, CD and AD recombinants contributed 1/20 (5%) each. Dual infections with two subtypes was identified in 4/36 (11%) individuals in that study population (Arroyo *et al.*, 2004).

With the high frequency of recombinant viruses, full-length sequencing is the only method that can comprehensively discriminate between pure subtypes and recombinant forms, however this method is not suitable for large-scale population genotyping projects. Other methods like the heteroduplex mobility assay (HMA) (Bachman *et al.*, 1994), restriction fragment length polymorphism (RFLP) (Janini *et al.*, 1998) and sequencing of short genome segments (Gao *et al.*, 1994) are insufficient since they only screen subtype restricted regions of the genome and may not detect recombinants. Hence there is a need for a high throughput automated genotyping method capable of discriminating recombinants from pure subtype genomes. For this reason a multi-region hybridization assay (MHA) was developed to identify HIV subtypes A, C, D in five genomic regions (Hoelscher *et al.*, 2002). The principle of the MHA is to amplify five genomic regions (*gag*, *pol*, *vpu*, *env* and *gp*41) in separate first round PCR. Each amplicon is then distributed to a second round PCR, each with a subtype-specific probe.

Using a panel of 45 DNA samples from Uganda, Kenya and Tanzania of known full length genome sequence, this method was found to have 90% sensitivity and 98% specificity to discriminate between subtypes A, C and D indicating that it was a good tool for subtyping genomic regions of viruses from East Africa.

The aim of this chapter was to evaluate the reliability of MHA utilizing RNA as a template (RNA MHA), and to compare RNA-MHA to DNA-MHA (which represents integrated proviral populations). The DNA-MHA results were provided for comparative purposes by M. Hoelscher (University of Munich, Germany). Secondly, we aimed to apply this technique to define the epidemic in the Mbeya region. The study is part of the European Nations (EU) funded HIV-1 super infection study (HISIS), consisting of several research groups based in Ludwig Maximiliasn University, Germany; Mbeya Medical Research; National Public Health Institute, Helsinki-Finland and Henry M. Jackson Foundation, Maryland US.

2.2 Materials and Methods

2.2.1 Study population

A cohort of six hundred women from a high-risk population in Mbeya, Tanzania was enrolled between September and November 2000 for a prospective HIV Superinfection Study (HISIS) within the Mbeya Medical Research Programme. These women worked in bars along the trans-African Highway joining Zambia, Malawi and Mozambique. Informed consent was given by all participants and the study was approved by the University of Cape Town's ethics committee (number 198/99). For this study, blood samples were obtained from a group of fifty-seven (n = 57) HIV positive individuals.

The characteristics of this cohort have being fully described by Reigner *et al.*, (2003). Briefly, the HISIS cohort consists of about six hundred (n = 600) women of whom 408 (68%) were seropositive and 192 (32%) were seronegative on recruitment. The average age of the cohort was 28.5 years with a range of between 20 to 32 years. Sexually transmitted infections were clinically diagnosed in 54.3% of the women at recruitment. Condom usage was reported to be less than 50% on average. Participants reported engaging in sex with their steady partners as well as casual partners. The women visited the clinic every three months for evaluation, during that time they also donated blood samples. For this study samples were analyzed from 57 HIV positive individuals. Thirty two (n = 32) of the fifty seven individuals had two consecutive blood draws (T1 and T2), while twenty one (n = 21) had only the first time point (T1) and four individuals (n = 4) had the second blood draw only (T2).

2.2.2 HIV-1 diagnosis

Antibodies to HIV were detected by the dual enzyme-linked immunosorbent assays Determine HIV-1/2 (Abbot, Germany) and Enzygnost anti HIV-1/2 (Dade Behring, Germany). HIV diagnosis data was provided by M. Hoelscher, (Ludwing - Maximilians University, Germany).

2.2.3 The Principle of Multi-region Hybridization Assay (MHA)

Five short regions of the HIV-1 genome (Fig. 2.2.1) were amplified, each amplicon distributed to three individual tubes for second round PCR. Each second round tube contained a different subtype-specific probe (A, C or D), in a Taqman real-time PCR format. All probes were labeled at the 5' end with 6-carboxyfluorescein and had a quencher at the 3'end. The 5' to 3' exonuclease activity of the Taq DNA polymerase, results in cleavage of fluorescent labelled probes during PCR. PCR primers and a subtype specific TaqMan probe bind to the DNA template. In the presence of the enzyme Taq polymerase, the primer is able to extend on the template and in the process displace the TaqMan probe resulting in an increase in relative fluorescence of the reporter. Degradation of the TaqMan probe by the Taq DNA polymerase frees the reporter dye from the quenching activity of TAMRA and thus the fluorescent activity increases with an increase in cleavage of the probe. This is proportional to the amount of PCR product formed over time.

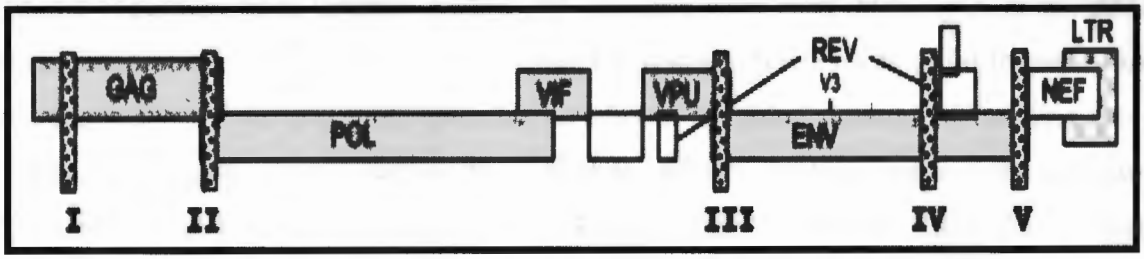


Fig. 2.2.1: The organizational structure of the HIV-1 genome, with locations of the regions targeted by MHA indicated by I, II, III, IV and V (Hoelscher, et *al.*, AIDS 2002).

Nested PCR primers for the five selected regions termed, *gag* (I), *pol* (II), *vpu* (III), *env* (IV) and *gp* (V) were made. For each of the five genomic regions selected, subtype specific probes (A, C and D) were also made (Hoelscher *et al.*, 2000). Each TaqMan probe was fluorescently labelled with a reporter dye 6-carboxyfluorescein (FAM) and a quencher 6-carboxy-tetramethyl-rhodamine (TAMRA). The probes were designed such that they do not cross react against each other. Fluorescence intensity was measured by the sequence detection system (SDS). Figure 2.2.2 depicts a typical real time PCR cycle with fluorescent accumulating over time.

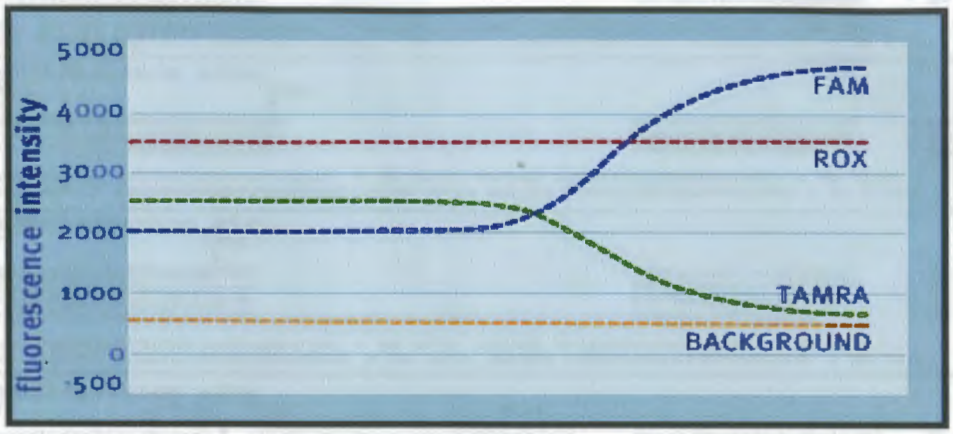


Fig. 2.2.2: The graph reflecting a single fluorescent component of the reaction. The increase in FAM is indicative of a positive determinant (www.vetscite.org).

A positive TaqMan result is reflected by an increase in the fluorescent intensity of FAM and a decrease in the fluorescent intensity of TAMRA. Samples that showed an exponential increase in fluorescence intensity over five consecutive cycles were considered positive.

2.2.4 Primer and probe Design

To design probes that could discriminate among subtypes A, C and D more than 100 virtually full-length genome sequences from Uganda, Kenya and Tanzania were used (Hoelscher, et *al.*, 2002). All primers and probes were supplied by Applied Biosystems (Foster City, California, USA). Genomic regions located in *gag*, *pol*, *vpu*, the membrane spanning domain of *env* and the C terminal part of the gp41 were selected. Nested PCR primers flanking each probe site were designed, the primers were designed such that subtypes A, C and D could be amplified to the same extent. The genome regions selected together with details of the PCR primers and probes are shown in Fig. 2.2.3.

Figure 2.2.3: Probes and primers for the MHA. The gene locations for *gag*, *pol*, *vpu*, *env* and *gp* and probes are depicted. For each location, the sequences of probes for subtypes A, C and D (underlined), to which strand of DNA they hybridize and the predicted *Tm* of the duplex formed are indicated. The sequences of outer and inner primers for nested PCR amplification of each region are shown, with direction and *Tm* indicated. The probe locations for the different subtypes are offset within *env* and *gp* regions, indicated by the break in sequence (Hoelscher *et al.*, AIDS 2002).

2.2.5 RNA Extraction

The extraction of viral RNA from plasma (200µl) was performed by the MagNA Pure total nucleic acid robotic extraction procedure (Roche Diagnostics Corp, US) following the manufacture's manual. Negative controls were included in each extraction procedure and in the analysis to test for contamination.

2.2.6 cDNA Synthesis

For the RNA-HMA, amplification of cDNA from plasma was performed using the One Step RT-PCR (Qiagen Inc, US). Extracted viral RNA was used as template for the synthesis of complementary DNA (cDNA). Amplification of the five short genomic regions (*gag*, *pol*, *vpu*, *env and gp41*) was done in separate first round RT-PCRs. First and second round fragments were kept as small as possible to provide maximum PCR efficiency. All reagents, primers and conditions for the one step RT-PCR are listed in tabular form (Table 2.2.1 and 2.2.2).

The one-step RT-PCR master mix was prepared as follows:

Table 2.2.1 Master Mix

Reagent	Amount
Primer 1 (20µM)	2µl
Primer 2 (20µM)	2µl
dNTP mix (10mM each)	2µl
5x One step RT PCR buffer	10µl
One step PCR enzyme	2µl
RNAse inhibitor	0.25µl
RNA	32µl
Total Volume	50µl

Table 2.2.2 depicts the conditions for the amplification of first round DNA fragments.

Table 2.2.2: PCR conditions of amplification

Cycling conditions	One-Step RT-PCR	Cycles
(RT)	30 minutes (50°C)	1
(activation)	15 minutes (95°C)	1
(denaturing)	30 seconds (95°C)	X 40
(annealing step)	1 minute (55°C)	
(extension step)	1 minute (72°C)	
(extension step)	10 minutes (72°C)	1
END	END (4°C)	∞

2.2.7 Second Round Polymerase Chain Reaction (PCR)

Each first round amplicon was distributed to three second round PCRs, each with a different subtype specific fluorescent probe in a Taqman real-time format. The second round PCR amplifications were performed in a 96-well spectrofluorometric ABI7700 sequence detection system (Applied Biosystem, US). All reagents, primers and conditions for second round PCRs are listed in tabular form (Table 2.2.3 and 2.2.4).

Table 2.2.3: PCR reagents and volumes for DNA amplification

PCR Reagents	Amount
Universal PCR master mix	12.5µl
Primer 3 (600 nM)	0.6µl
Primer 4 (600 nM)	0.6µ l
Subtype specific probe (220 nM)	1.1µl
deionized water	8.2µl
Template	2µl
Total	**25µl**

Table 2.2.4 depicts the conditions for the amplification of the second round DNA fragments.

Table 2.2.4: PCR conditions of amplification

Cycling conditions	*Time (temp)*	Cycles
(activation)	5 minutes (50°C)	1
(activation)	10 minutes (95°C)	1
(denaturing)	15 seconds (95°C)	X 40
(annealing step)	1 minute (55°C)	
(extension)	1 minute (60°C)	
END	END (4°C)	∞

2.2.8 MHA interpretation

For subtype detection, each sample was tested in triplicate. One part of the sample was tested against a subtype A-probe, C-probe and the D-probe. A negative control of PCR master mix without the cDNA template was used for detecting contamination. A reaction between the sample and one of the three probes indicate a positive determination. When two probes reacted to the same sample, a putative dual infection is assumed. In instances where no reaction occurred further investigation is pursued, and in most instances either PCR amplification failed or all the probes failed to bind to the sample.

A unique signal by a subtype-specific probe indicate the probe's subtype at that particular genomic region. Subtype C signals detected on all five regions would indicate a putative pure subtype C genotype. A putative recombinant is assigned to a mosaic genome that has more than one subtype-specific probe signal detected and a putative dual, is assigned when two subtype-specific probe signals are detected on the same genomic region. Figure 2.2.4 illustrates the three possible genotype outcomes namely, a pure subtype, a recombinant or a dual infection.

gag	*pol*	*vpu*	*env*	*gp41*	Outcome
A	A	A	A	A	Pure subtype A
C	C	C	C	C	Pure subtype C
D	D	D	D	D	Pure subtype D
gag	*pol*	*vpu*	*env*	*gp41*	Outcome
A	A	A	C	C	AC Recombinant
A	D	A	D	A	AD Recombinant
A	C	C	D	D	ACD Recombinant
gag	*pol*	*vpu*	*env*	*gp41*	Outcome
AC	C	C	C	C	A/C Dual
D	D	CD	D	D	C/D Dual
A	A	A	A	AD	A/D Dual

Fig. 2.2.4: Schematic interpretation of MHA results for identification of pure subtypes, recombinants and dual infections.

Other outcomes observed included samples that failed to be amplified by PCR and hence failure of the probe to bind. On the other hand some amplified samples showed no probe hybridization by any of the three subtype specific probes.

2.3 Results

2.3.1 RNA-MHA sensitivity evaluation

The multi-region hybridization assay (MHA) is a robust and high-throughput tool for HIV-1 genotyping of proviral DNA, with a sensitivity of 90% and specificity of 98% (Hoelscher *et al.*, 2002). It is an accurate and efficient technique for genotyping HIV-1 in populations of East Africa were subtypes A, C and D co-circulate (Hoelscher *et al.*, 2002). Previous studies utilized proviral DNA as template and in the current study MHA was evaluated for genotyping sensitivity on plasma RNA samples. For each of the 57 participants, all five genomic regions at each time point were genotyped. The sensitivity of the assay was calculated for 445 reactions performed. The percentage of positive determinants was stratified by the genomic regions assayed.

Table 2.3.1 depicts the break down for all regions assayed including the number of negative determinants observed. *Pol* had the highest number of positive determinants 93%, followed by *gag* (92%), *gp*41(91%), *env* (88%) and *vpu* (85%). The main factor limiting sensitivity was the failure of the probe to hybridize, rather than the lack of a PCR product. All five genomic regions showed a sensitivity of over 85% with an overall sensitivity of 90%.

Table 2.3.1: Sensitivity of the RNA - MHA assay by genomic region

		Genomic regions				
	Time point	***Gag***	***Pol***	***Vpu***	***Env***	***Gp 41***
Positive Determinations	T1	49	49	44	44	49
Negative Determinations		4	4	9	9	4
Percentage positive	**T1**	**93**	**93**	**83**	**83**	**93**
Positive Determinations	T2	33	34	32	34	32
Negative Determinations		3	2	4	2	4
Percentage positive	**T2**	**92**	**94**	**89**	**94**	**89**
Total Positive	T1/T2	82/89	83/89	76/89	78/89	81/89
Total positive (%)	**T1/T2**	**92**	**93**	**85**	**88**	**91**

2.3.2 MHA performance on DNA compared to RNA

Two independent investigators compared the performance of the assay using either DNA or RNA as the template. Hoelscher (2002) developed and tested the DNA-MHA and in this study the RNA HMA was developed and tested. Twenty-two (n = 22) individuals with two consecutive blood draws (median 3 months) for both DNA and RNA (n = 44) were compared for their positive frequency. Positive outcomes were stratified by genomic regions tested as depicted in Table 2.3.2. In all regions, except for *pol*, the number of positive determinants was great using RNA as a template compared to proviral DNA. Overall the RNA-MHA performed better than the DNA MHA with a total of 200/220 (91%) positive determinants compared to 188/220 (86%). The percentage of positive determinants at the first time point by DNA-MHA (DNA1) was 96/110 (87%) compared to 99/110 (90%) by RNA-MHA (RNA1). For the second time point, DNA-MHA (DNA2) had 92/110 (84%) positive determinants compared to 101/110 (92%) by RNA-MHA (RNA2) (described below Figure 2.3.1).

Table 2.3.2: Comparison of the performance of HMA-DNA against RNA-HMA

Template		*Gag*	*Pol*	*Vpu*	*Env*	*Gp41*	*Total*
DNA	***positive***	38	40	37	35	38	188
	negative	6	4	7	9	6	32
(%)	**positive**	**86**	**90**	**84**	**80**	**86**	**86**
RNA	***positive***	42	39	39	39	41	200
	negative	2	5	5	5	3	20
(%)	**positive**	**95**	**89**	**89**	**89**	**93**	**91**

Forty-one of the forty-four (93%) comparisons between RNA and DNA template at time point 1 (RNA1/DNA1) were concordant for subtype designation. Of the remaining three samples, one was discordant in subtype designation. This sample (SW33) was identified as dual infection in *env* and could be a consequence of higher sensitivity of RNA-HMA. The other two samples (SW62 and 86) were discordant due to a lack of subtype detection in certain genomic regions. The majority 15/41 (37%) of concordant outcomes were subtype C, 14/41 (34%) were AC recombinants, 6/41 (15%) were A, 4/41 (10%) were D and 2/41 (%) were dual. Fig. 2.3.1 depicts the 44 comparisons between DNA-HMA and RNA-HMA on samples from 22 individuals with two consecutive draw draws.

Fig. 2.3.1: Depicts the mosaic structures of all viruses which were compared between DNA1/RNA1 and DNA2/RNA2. Subtype A is represented in green, C in purple, D in blue and AC in yellow. Black out regions represent negative results where subtyping was not possible due to failure of probes to bind despite positive PCR.

2.3.3 Subtype prevalence in the study population

Due to higher sensitivity, the RNA-MHA was then applied to subtype an extended panel of 89 samples from fifty-seven (n = 57) individuals (Fig. 2.3.2). Thirty two (n = 32, demarcated by light blue) of the fifty seven individuals had two consecutive blood draws (T1 and T2), while twenty one (n = 21, demarcated by grey) had only the first time point (T1) and four individuals (n = 4, demarcated by orange) had only the second blood draw (T2). For each individual, five genomic regions (*gag*, *pol*, *vpu*, *env* and *gp*41) were assayed on either or both time points depending on sample availability at either time point.

Figure 2.3.2: RNA-MHA genotyping results of the study population. Subtype A is represented in blue, C in purple, D in green, AC in yellow, AD in orange and CD in gray. Blanks represent unavailable data while black out regions represent negative result where subtyping was not possible due to failure of probes to bind despite positive PCR.

The majority of individuals 22/57 (38.5%) were infected with recombinant viruses of which 16/22 (73%) had a subtype C genomic sequence. AC recombinants accounted for 12/22 (54.5%), 4/22 (18%) were AD while CD and ACD recombinants accounted for 3/22 (13.5%) each. Pure subtype C viruses represented 16/57 (28%) infections in the study population and were the majority among the pure subtypes. While pure subtype A represented 4/57 (7%) and subtype D 5/57 (9%). Dual infections were detected in 10/57 (17.5%) individuals with 6/10 (60%) infected with both subtypes A and C, 2/10 (20%) infected with ACD, while AD and CD duals represented 1/10 (10%) of the dual population each.

Detection of HIV genotypes in all samples from individuals with either time points (T1 or T2) revealed that HIV-1 infections with subtype A accounted for 3/57 samples, subtype C 19/57, subtype D 4/57, dual infections 6/57 and recombinants accounted for 25/57 samples genotyped (Figure 2.3.3).

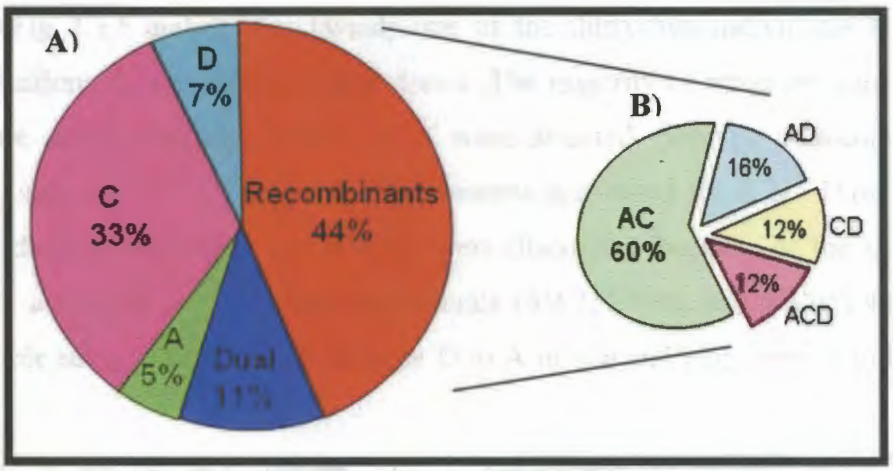


Fig. 2.3.3: Distribution of the genotypes in all samples (n = 89) provided by the study population (n = 57) individuals, assayed in time point 1 and 2) : **A)** depicts the individual contributions of pure, dual an recombinant genomes **B)** depicts the individual break down of all the unique recombinant subtypes.

Most of the recombinants forms identified did not share the same mosaic structure as identified by differences in subtype designation of genomic regions. Of the 25 recombinant viral genomes detected, 15 were AC (60%), 4/25 (16%) were AD and 3/25 (12%) were either CD or ACD (Fig. 2.3.2). Of the six dual infections detected 3/6 (50%) had both subtypes A and C, 2/6 (33%) had C and D and 1/6 (17%) subtypes C and D (Fig. 2.3.4). Some individuals showed shifts in viral subtypes over time (see section 2.3.4).

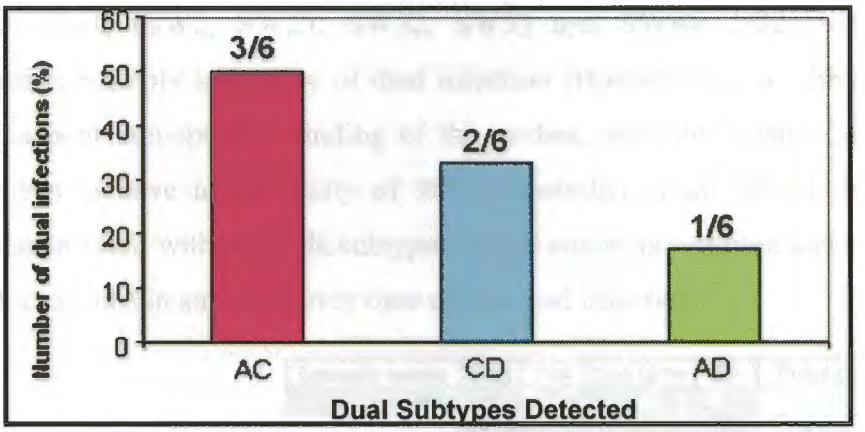


Fig. 2.3.4: Number of dual infections detected in the blood draw 1 and 2.

2.3.4 Monitoring changing viral subtypes over time

The fluctuation of viral populations was measured by determining changes in subtype determination in different genomic regions over time in 32 individuals. Samples collected over two consecutive time points, with a median of 3 months between the first and the second blood draw were analyzed (Fig 2.3.5 and 2.3.6). Twenty-one of the thirty-two individuals had concordant subtype determinations in consecutive blood draws. The majority of observed concordant results were among pure subtype viruses, where 13/21 were detected. Subtype C accounted for 7/21, subtype A 4/21, subtype D 2/21, and AC recombinants accounted for 8/21. Three (3/32) of the thirty-two individuals (Sw1, Sw62 and SW86) were discordant because of the lack of subtype determination on a specific region. Three individuals (SW2, SW81 and SW95) were discordant because of possible subtype shifts: from subtype D to A in *vpu* and *gag*; from A to C in *gag* (Fig. 2.3.5)

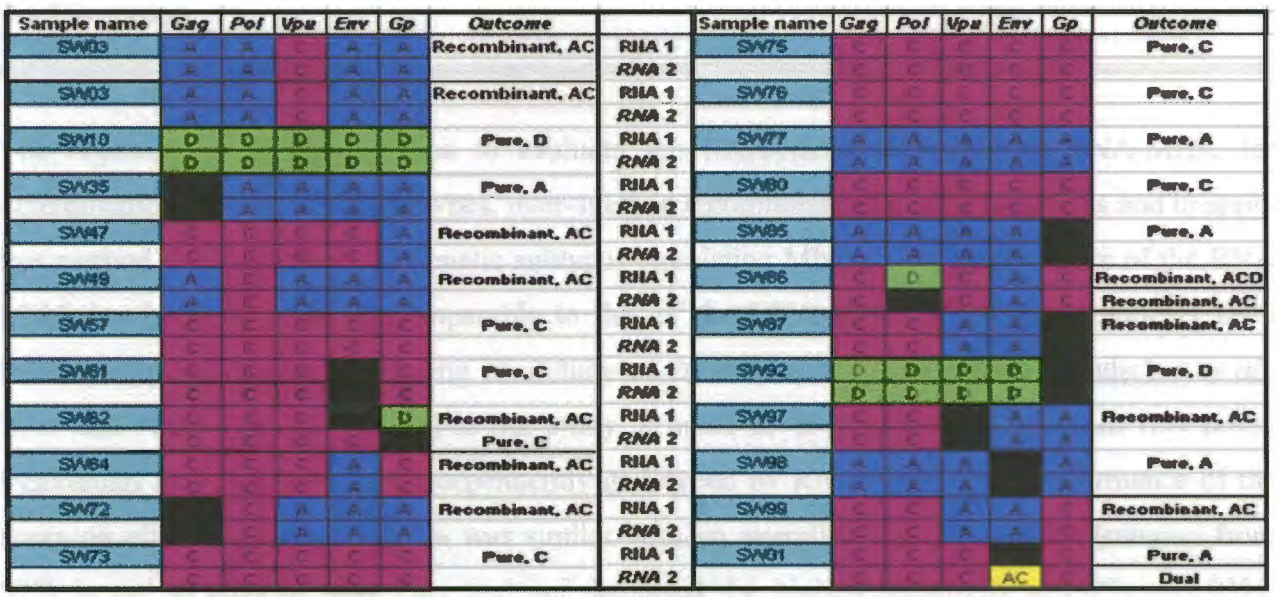


Fig. 2.3.5: Depicts the mosaic structures of all concordant and three discordant viruses detected in two consecutive blood draws. Subtype A is represented in blue, C in purple, D in green and AC in yellow. Black out regions represent negative result where subtyping was not possible due to failure of probes to bind despite positive PCR.

Five individuals (SW2, SW25, SW32, SW33 and SW84; 5/32, 16%) had multiple probe hybridization possibly indicative of dual infection (Hoelscher *et al.*, 2002; Fig. 2.3.5). While we cannot discount non-specific binding of the probes, previous studies on this cohort had shown these probes to have a specificity of 98% (Hoelscher *et al.*, 2002). In addition, detection of individuals infected with multiple subtypes in this cohort is common and studies from Mbeya have shown fluctuations in subtypes over time due to dual infection.

Sample name	Gag	Pol	Vpu	Env	Gp	Time point
SW02	C	C	D	A	C	RNA 1
	C	C	A	A	C	RNA 2
SW81	C	A	A	A	A	RNA 1
	C	C	A	A	A	RNA 2
SW95	D	A		D	A	RNA 1
	A	A			A	RNA 2
subtype shifts						
SW05	CD	A	D	C	D	RNA 1
	CD	AC	CD	AC	D	RNA 2
SW25	CD	C	C	D	C	RNA 1
	C	C	C	AD	C	RNA 2
SW32	A		C	C	A	RNA 1
	A		C	AC	A	RNA 2
SW33	C	C		AC	C	RNA 1
	C	C	C	A	C	RNA 2
SW84	C	C	C	A	C	RNA 1
	C	C	C	AC	C	RNA 2
multiple subtype detection						

Fig. 2.3.6: Depicts the mosaic structures of discordant viruses detected in two consecutive blood draws. Subtype A is represented in green, C in purple, D in blue, AC in yellow, AD in orange and CD in gray. Black out regions represent negative result where subtyping was not possible due to failure of probes to bind despite positive PCR. The first panel depicts, discordant viruses due to subtype shifts and those depicted lastly were discordant due to detection of multiple subtypes at the same region.

2.4 Discussion

The objective of this study was to evaluate the RNA-MHA compared to DNA-MHA for determination of HIV-1 pure subtypes, inter-subtype recombinants and dual infections and to apply this method to characterize the genetic subtypes circulating Mbeya. The performance of the RNA MHA has been found to be comparable to that of the DNA-MHA which has previously been validated on full length sequencing in a study of 20 individuals from Mbeya (Hoelscher *et al.*, 2001 and Arroyo *et al.*, 2004). For our study, a panel of 89 clinical samples from fifty-seven individuals from Mbeya, were independently genotyped by RNA-MHA. The performance of the assay on all five genomic regions was similar with an overall sensitivity of 89% (ranging from 93% in *pol* to 85% in *vpu*). *Gag* and *pol* displayed the highest sensitivity at (92% and 93%) respectively while the *vpu* region had a sensitivity of 85% (Table 2.3.1).

The multi-region hybridization assay had been previously optimized and assessed on DNA samples from Mbeya where the full-genome sequences were known (Hoeslscher *et al.*, 2002). We adapted the assay to subtype RNA samples and a comparison between RNA-MHA to DNA-MHA showed a concordance of 93% (41/44) (Fig. 2.3.1 results section). However subtyping using RNA was more sensitive with 91% of regions subtyped compared to 86% for DNA. Our results show the RNA-MHA to be a robust and reliable method for HIV-1 genotyping which was comparable to the DNA-MHA if not more sensitive, a view shared by Arroyo *et al.*, (2004).

Subtype prevalence in the study population was identified, with recombinant strains being most frequent comprising 44% of samples characterized, AC recombinants constituted 60% of all recombinants observed in the study and 84% of all recombinants contained a subtype C sequence. Subtype C was the dominant pure subtype at 33%, with subtype A at 5% and D constituted 7%. Dual infections were observed in 11% of samples genotyped, which is similar to the 11% of dual infections detected from a panel of 45 clinical samples from Tanzania, Kenya and Uganda used in setting up the DNA-MHA (Hoelscher *et al.*, 2002). MHA provides a unique advantage over other methods for identifying dual infections as the method can detect binding of two probes.

In conclusion, MHA is able to determine HIV-1 subtypes A, C and D in multiple genomic regions and efficiently discriminate pure from recombinant strains. This method offers a high throughput assay for large scale population genotyping studies in setting where multiple subtypes co-circulate. It has the advantage over other screening methods for detection of recombinants since it assays multiple regions. Added to that, this assay has the ability to detect dual infections. This study provides additional information on the virus strains circulating in Tanzania. Overall these studies demonstrate that Tanzania harbours a complex diversity of viruses with the majority of viruses being unique recombinant forms (URFs). The high prevalence of dual infections is clearly fuelling the generation of these URFs and it would be of interest to monitor the molecular epidemiology of the epidemic to determine if new CRFs emerge.

CHAPTER 3

HIV VIRAL DYNAMICS IN WOMEN INFECTED WITH EITHER SINGLE OR MULTIPLE HIV-1 SUBTYPES

3.1 Introduction

The high level of HIV-1 genetic diversity is caused by the ability of the virus to mutate, recombine and replicate with a short generation time. The high mutation rate of 0.3 nucleotide changes per genome per replication cycle (Jetzt *et al.*, 2000) is the result of inaccurate viral DNA synthesis by the viral enzyme reverse transcriptase, which is error-prone due to its lack of 3'→5' exonuclease proofreading activity (Boyer *at el.*, 1992). The enzyme also switches between the two RNA templates that are packaged in each virion, approximately three times per genome per replication cycle (Jetzt *et al.*, 2000). When a cell is simultaneously infected with two or more phylogenetically distinct HIV-1 strains (dual infection), recombination during replication leads to a novel viral genome, which contains genetic information from both strains.

Dual infection is responsible for the genesis of recombinant HIV-1 genomes. The number of Circulating Recombinant Forms (CRFs) and Unique Recombinant Forms (URFs) is increasing especially in areas where multiple HIV-1 subtypes circulate in the population. Dual infection can be a consequence of co-infection (infection of two strains simultaneously or within the short window period during acute infection) or superinfection (infection with a second strain after established infection). Understanding the timing, immune responses and frequency of dual infection has implications for understanding HIV transmission and correlates of protection from superinfection and can thus inform vaccine development. The frequency and pathogenic consequences of dual infection are not completely understood. Intra-subtype dual infection has been shown to be associated with higher viral load at setpoint and rapid disease progression (Grobler *et al.*, 2005 and Gottlieb *et al.*, 2004).

HIV is endemic in Tanzania where three different subtypes A, C and D have been identified (Hoelscher *et al.*, 2001, Kiwelu *et al.*, 2003 and Arroyo *et al.*, 2004). In a study of blood donors in Mbeya, more than 50% of the samples were subtype C, with a high proportion of subtype C containing inter-subtype recombinants (Arroyo *et al.*, 2004). The HIV epidemic in this region is complex with differences between urban and rural populations. The urban population harboured more recombinants viruses (OR, 2.69; 95% CI, 1.08 - 6) and duals infections (OR, 5.16; 95% CI, 1.07 - 24.9) compared to the rural population (Arroyo *et al.*, 2005). In addition, co-infection with more than one HIV-1 subtype appears to be more common in high-risk cohorts. A study showed that high-risk populations have increased percentage of HIV-1 recombinant strains (54% vs 40%, $p < 0.05$) and a higher frequency of dual infections (19% vs 9%, $p < 0.02$; Herbinger *et al.*, 2006).

A study by Gerhardt *et al.*, (2005) identified an individual from the HISIS cohort who was triply infected with AC viral strains that had two distinct subtype A sequences and depicted significant fluctuation in molecular viral forms over time. Another study (McCutchan *et al.*, 2005) from the same cohort identified a case of superinfection, wherein the study subject was first infected with a complex ACD recombinant strain and later 6 to 9 months later was infected with another recombinant (AC) viral strain. Many other potential dual infections have been identified in the HISIS study using the multi-region hybridization assay (MHA; Hoelscher *et al.*, 2002).

Information on HIV-1 genetic diversity is crucial to define requirements for an effective vaccine, especially in regions where HIV-1 vaccine trials are planned. Studies on the dynamics of viral evolution and recombination in dual infection contribute to understanding of correlates of protection and the potential contribution of high diversity to disease progression. In this study we employed a gel based screening assay heteroduplex mobility assay (HMA) for detection of dual infection. In addition, through sequence analysis we investigated the dynamics of viral evolution among dually infected individuals.

3.2 Materials and Methods

3.2.1 Study Population

A cohort of six hundred women from a high-risk population in Mbeya, Tanzania was enrolled between September and November 2000 for a prospective HIV Superinfection Study (HISIS) within the Mbeya Medical Research Programme. The women were working in bars on the Tanzanian side of the Trans-African highway, which connects Zambia and Malawi with Tanzania. Informed consent was obtained from all participants, each of which provided blood samples at enrolment and every three months thereafter for a period of to 21 months (Gerhardt *et al.*, 2005). Based on MHA results generated by M. Hoelscher (personal communication) six individuals with single infection and six with dual infection were selected to form part of our study.

3.2.2 HIV Diagnosis

Antibodies to HIV were detected by the dual enzyme-linked immunosorbent assays Determine HIV-1/2 (Abbot, Germany) and Enzygnost anti HIV-1/2 (Dade Behring, Germany). HIV diagnosis data was provided by M. Hoelscher (Ludwing - Maximilians University, Germany).

3.2.3 Viral Loads

The viral loads were measured by HIV Amplicor v5 (Roche, Germany) and were made available for this study by [Dr A. Puren, National Institute of Communicable Diseases (NICD)].

3.2.4 $CD4^+$ T-Cell Counts

Lymphocyte subsets were measured on fresh samples using FACscount (Becton Dickison, Belgium) and CD4 T-cell counts were provided by [Dr M. Hoelscher (Ludwing - Maximilians University, Germany)].

3.2.5 RNA Extraction

The isolation of viral RNA from plasma was performed manually using the QIAamp® Viral RNA Mini Spin Kit method (Qiagen, Valencia, CA, USA). Viral RNA was extracted from 280µl plasma following the manufacture's instructions. RNA was eluted in 20-150 µl of RNase free water and stored in aliquots at -80°C.

3.2.6 cDNA Synthesis

Extracted viral RNA was used as template for the synthesis of complementary DNA (cDNA) using the Invitrogen Thermoscript™ RT-PCR System (Invitrogen, GmbH, Karlsruhe, Germany). Master mix number 1 containing 20pMols of gene specific primers (MHvpu2 5'-CTTGTGGGTTGGGGTCTG -3' for the *vpu* region or ed33 5'-CCTCAGCCATTACACAGGCCTGTCCAAAAG-3' for the *env* C2C3 region; Table 3.2.1) was incubated at 65°C for 5 minutes to eliminate any RNA secondary structure before adding to Master mix 2 (Table 3.2.2) and incubating for one hour at 50°C.

The master mixes were prepared as follows:

Table 3.2.1: Master Mix 1

Reagent	Amount
MHvpu 2 or ed33 (20pmol)	1µl
RNA (10 pg -5 µg)	4µl
10 mM dNTP Mix	1µl
Total Volume	6µl

Table 3.2.2 Master Mix 2

Reagents	Amount
5X Buffer	2µl
DTT (10mM)	0.5µl
RNaseOUT™(40U)	0.5µl
Thermoscript RT enzyme (15U)	0.5µl
Water	0.5µl
Total	4µl

3.2.7 First and second round PCR for *vpu* and *env* C2C3 regions

All reagents, primers and conditions for the *vpu* and *env* C2C3 PCR are listed in tabular form (Tables 3.2.3, 3.2.4, and 3.2.5). In a first round of PCR, *vpu* was amplified as a 688bp fragment and a second round of nested PCR amplified a 501bp fragment. While the *env* C2C3 PCR amplification yielded a first round fragment of 564bp and second round of nested PCR amplified a 511bp fragment.

Table 3.2.3: Primers for both *vpu* and *env* C2C3 PCRs

Region	Primer	Position (HXB2)
First round vpu *(Forward Primer)*	**Vp-Fwd (5'-GGTGCCAACATAGCAGAATAGGCAT-3')**	5783-5807
(Reverse Primer)	**Mhvpu2 (5'-CTTGTGGGTTGGGGTCTG -3')**	6454-6471
Second round vpu *(Forward Primer)*	**Mhvpu1 (5'-CCTATGGCAGGAAGAAGCGG -3')**	5967-5986
(Reverse Primer)	**Vp-Rvs (5'-GTGGGTTGGGGTCTGTGGGTACACA -3')**	6444-6468
First round *env (Forward Primer)*	**ED31 (5'-TTACAGTAGAAAAATTCCCCTC-3')**	6817-6846
(Reverse Primer)	**ED33 (5'-CCTCAGCCATTACACAGGCCTGTCCAAAAG-3')**	7360-7381
Second round *env (Forward Primer)*	***env*BF (5'- TACACAAGCCTGTCCAAAGGT -3')**	6826-6847
(Reverse Primer)	***env*BR(5'- AATTTCTAGGTCCCCTCCTGA -3')**	7317-7337

Table 3.2.4: PCR reagents and volumes for both *vpu* and *env* C2C3 PCRs

PCR Reagents	vpu First round	vpu Second round	*env* First round	*env* Second round
5'- (Forward Primer) (20pmol/µl)	1µl	0.5µl	2µl	1µl
3'- (Reverse Primer) (20pmol/µl)	1µl	0.5µl	2µl	1µl
dNTP Mix (Roche, USA) (2.5mM each)	4µ l	2µ l	4µ l	2µ l
10x buffer	10µl	5µl	10µl	5µl
1.5mM $MgCl_2$	1.5µl	1.5µl	8µl	5µl
deionized water	72.4µl	38.4µl	63.4µl	31µl
Supertherm Polymerase (Bertec, Taiwan).	0.24µl	0.12µl	0.6µl	0.2µl
Template	10µl	5µl	10µl	5µl
Total 100	**100µl**	**50µl**	**100µl**	**50µl**

Table 3.2.5: PCR cycling conditions for *vpu* and *env* C2C3 PCRs

Cycling Conditions	*vpu* First round	*vpu* Second round	*env* First round	*env* Second round	Cycles
(denaturing)	5 minutes (95°C)	5 minutes (95°C)	5 minutes (95°C)	5 minutes (94°C)	1
(denaturing)	10 second (95°C)	10 second (95°C)	10 second (95°C)	10 second (94°C)	X 35
(annealing step)	30 seconds (55°C)	30 seconds (57°C)	45 seconds (50°C)	45 seconds (55°C)	
(extension step)	1 minutes (72°C)	1 minutes (72°C)	1 minutes (72°C)	1 minutes (72°C)	
(extension step)	5 minutes (72°C)	5 minutes (72°C)	5 minutes (72°C)	5 minutes (72°C)	1
END	END (4°C)	END (4°C)	END (4°C)	END (4°C)	∞

Standard precautions were taken to prevent contamination between samples.

3.2.8 Gel Electrophoresis

For gel electrophoresis, 5µl of PCR product including a positive and a negative control were loaded with 2µl loading dye on 2% agarose gel (agarose tablets, Bioline). Molecular weight marker VI (Roche, Germany) was also included for band size comparison. All PCR products were visualized on 2% agarose gel containing ethidium bromide (10mg/ml) (Sigma) on a transillumitor UV box (UVP, San Gabriel, CA, USA), after gel electrophoresis at 100V for 1h00. Digital pictures of all gels were captured using the Kodak digital science 1D camera.

3.2.9 Heteroduplex Mobility Assay (HMA)

Heteroduplex mobility assay was performed as described by Delwart *et al.,* (1994) with modifications. The Scie-Plas V20 vertical gel apparatus with glass plates 20cm h x 19.5cm and spacers 1.5mm thick were used (Scie-Plas Ltd, Southam Warwickshire England). Total population PCR products were mixed in a 200µl Eppendorf PCR tube, sample preparation included; 5µl PCR product, 5µl dH_2O and 1.1µl HMA buffer. For analysis of clones, equal amounts (5µl) from a reference clone and each of the screened clones. The samples were denatured at 95°C for two minutes and allowed to denature at room temperature for five minutes. Samples were mixed with 3µl 5X Ficoll/Loading Dye and loaded onto a 5% Non-denaturing polyacrylamide gel (maker) and allowed 5 hours of electrophoresis at 200V. The gels were stained by ethidium bromide (10mg/ml) followed by viewing under a UV box and digital picture captured using the Kodak digital science 1D camera.

3.2.10 PCR Clean up

PCR products were purified by the QIAquick Spin Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Briefly PB buffer (5x volume of PCR product) was mixed with the PCR product and loaded on the spin column and centrifuged for one minute at 13000rpm. Bound PCR products were washed with PE buffer and eluted by centrifugation at 13000rpm using buffer EB (50µl).

3.2.11 Cloning and sequencing

Samples that showed the presence of heteroduplexes in the HMA assay and all F1 and F7 samples were cloned into the pGEM®-T Easy vector (Promega, Madison - USA) according to the manufacture's instructions. Following transformation, colonies containing amplified HIV genes were identified by colony PCR (Lee and Cooper, 1995). All bulk PCR products from each time point and selected colony PCR products (those that showed population diversity by forming heteroduplexes in HMA assays) were selected for sequencing using ABI PRISM dye terminator cycle-sequencing kits (Applied Biosystems) and a 3100 ABI 16 capillary automated sequencer at the University of Stellenbosch. All reagents, primers and conditions for both *vpu* and *env* C2C3 sequencing are listed in tabular form (Tables 3.2.6 and 3.2.7).

Table 3.2.6: Sequencing PCR reagents and volumes

PCR Reagents	Sequence reaction
(Forward Primer) or (Reverse Primer) (3.2 pmol/µl)	1µl
5x buffer	2µl
deionized water	11µl
Big Dye® enzyme (AB, USA)	4µl
Template	2µl
Total volume	**20µl**

Table 3.2.7: Sequencing PCR cycling conditions

Cycling	Conditions	Cycles
(denaturing)	10 seconds (96°C)	1
(denaturing)	30 seconds (96°C)	X 25
(annealing step)	15 seconds (50°C)	
(extension step)	4 minutes (60°C)	
END	END (4°C)	∞

3.2.12 Phylogenetic Analysis

HIV-1 *vpu* and *env* C2C3 sequences were aligned using Clustal X (default settings; Thompson *et al.*, 1994) and alignment columns with gaps were removed. Neighbour joining phylogenetic trees with 100 bootstrap iterations were constructed and DNA distances were calculated with MEGA version 2.1 using kimura 2 parameter distances and a transition:transversion ratio of 2 (Kumar *et al.*, 2001). To quality control for sequences and ensure there was no unexpected relationships, final sequences were compared phylogenetically with all sequences from the study, as well as commonly used plasmids from the laboratory.

3.3 Results

To investigate the dynamics of viral diversity in single and dual infection, samples were provided from twelve individuals: six of whom had been identified as being infected with single HIV-1 infections; and six of whom had been identified as dually infected based on the multi-region hybridization assay (MHA; samples were provided by M. Hoelscher, University of Munich, Germany). The 12 participants that form part of this study were recruited during their chronic stage of disease and the duration of infection was not known. All participants were sampled longitudinally at three months intervals over 21 months (8 time points, F0 to F7). At recruitment, only one participant had viral load less than 10 000 copies/ml, eight had viral loads between 10 000 and 100 000 copies/ml and finally three participants had viral loads greater than 100 000 copies/ml (Fig.3.3.1). CD4 counts were not available at recruitment but were measured at F9 (24 month of follow-up) and they ranged from 91 to 670 cells/µl. A detailed analysis of virological course of disease, in context of evolution of HIV molecular variants is described in section 3.3.5.

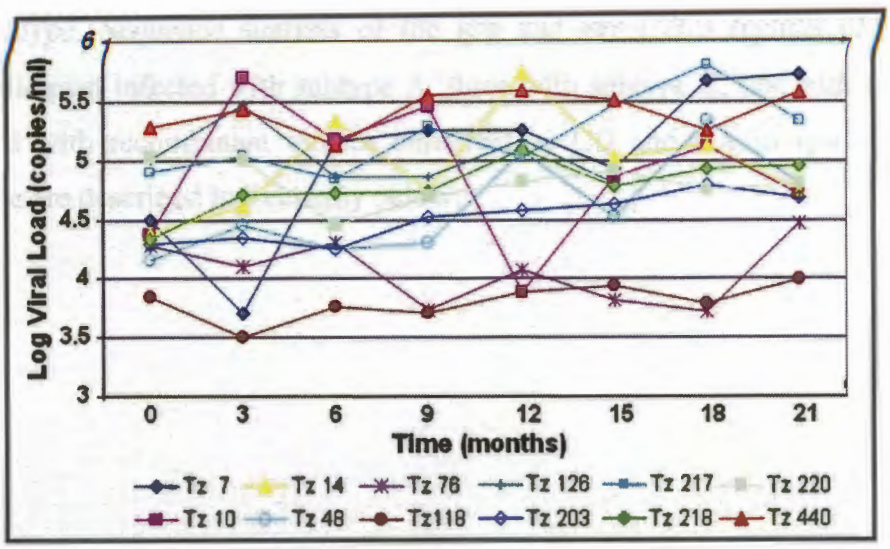


Figure 3.3.1: Viral load dynamics measured from the study cohort over 21 months of follow-up.

3.3.1 Subtype determination and detection of inter-subtype dual infection

MHA identified HIV subtypes using subtype-specific probes and therefore does not provide information on the diversity within an individual. To investigate diversity over time, the heteroduplex mobility assay (HMA) was applied. In this assay the sample is denatured and re-annealed resulting in the generation of heteroduplexes (between mismatched sequences) and homoduplexes (between homologous sequences) which are resolved by electrophoresis. Single fast migrating bands are indicative of homogeneous populations, with slow migrating bands indicative of diversity within the sample. To screen for diversity and possible shifts in viral populations, all samples were screened using HMA and amplicons were sequenced. To determine shift in viral populations samples from the first (F0) and last (F7) time points were amplified and cloned, and 20 randomly selected clones were sequenced. In addition, samples with multiple viral populations (heteroduplexes), as determined by HMA, were cloned and 20 random clones sequenced.

Two regions of the HIV genome, the *env* C2C3 and the *vpu* region, were chosen for screening diversity in this study. HMA analysis of the *env* C2C3 region has previously been shown to identify dual infection with the same subtype (Grobler *et al.*, 2004) and thus has sufficiently high diversity to enable resolution of both intra and inter-subtype dual infection populations. In addition, this region contains the variable loop three (V3) which is involved in determining coreceptor usage (Ling *et al.,* 2004 and Poignard *et al.,* 2001). The *vpu* region was selected as it corresponded to the region targeted for subtyping by the MHA, and thus enables a direct comparison between the two assays.

Single infections

Using HMA and sequencing, eight of the twelve participants were found to be infected with a single HIV-1 subtype. Sequence analysis of the *vpu* and *env* C2C3 regions of the genomes, identified one participant infected with subtype A, three with subtype C, one with subtype D and two were infected with recombinant viruses classified as CD and CA in *vpu* and *env* C2C3 respectively. These are described individually below.

Participant Tz7

Participant Tz7 was infected by a subtype A virus in both the *vpu* (red) and *env* C2C3 (green) regions of the genome (Fig.3.3.2). In both genomic regions high viral diversity was evident, as indicated by the formation of multiple heteroduplexes. However, all the viral sequences were multiple variants of the same subtype as indicated (95% bootstrap support; Fig.3.3.2).

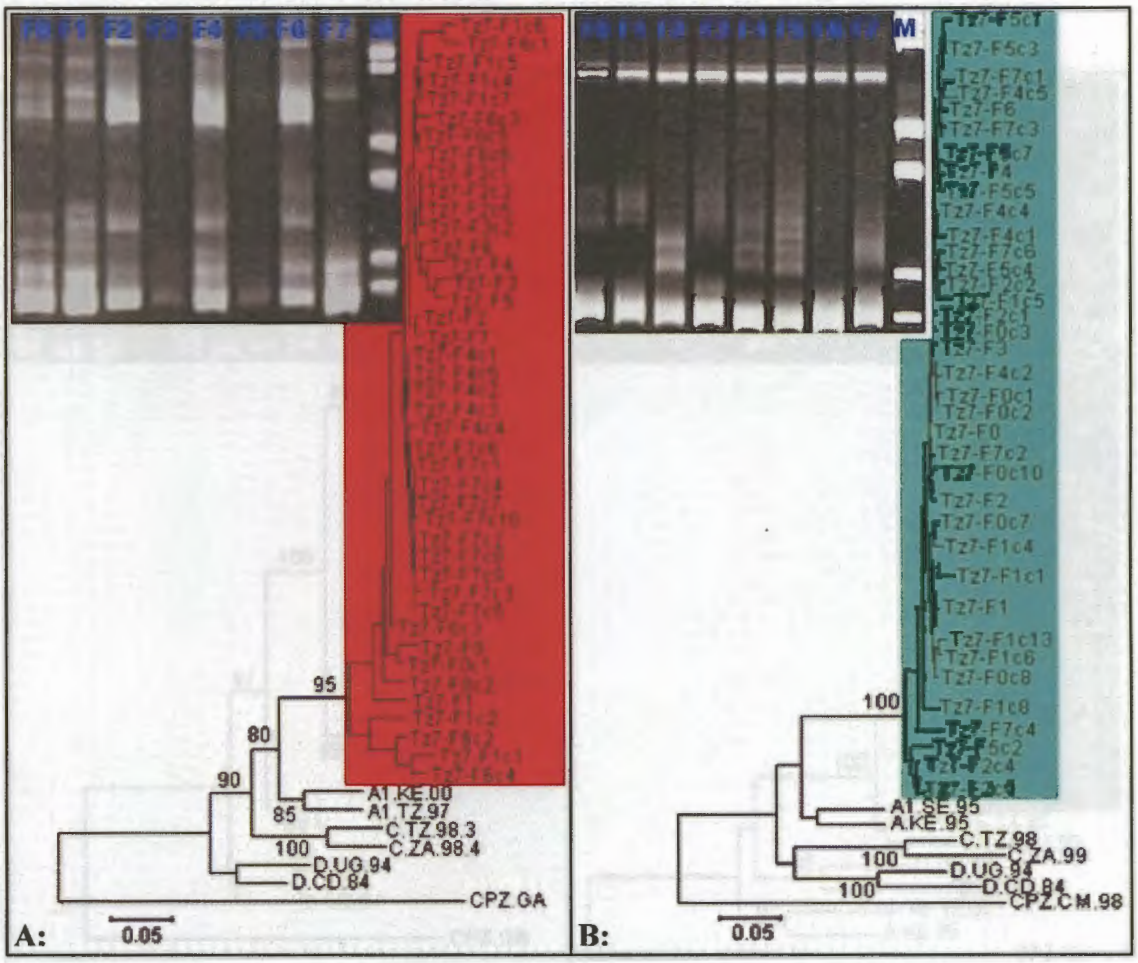


Figure 3.3.2: Tz7 *vpu* (red) and *env* C2C3 (green) neighbour joining trees from F0 (time point 0) to F7 (21 months post F0), scale bar indicates nucleotide divergence of 5% calculated from one hundred repeats. Only bootstrap values over 75% are shown. Sequences nomenclature is in the following format e.g Tz7-F1c0; where Tz7 is the participant identifier, F1 is the visit number and c0 is the clone number. Sequences with no clone numbers were generated directly from amplicons. HMA gel analysis of the *vpu* (panel A) and *env* C2C3 (panel B) regions depicting heteroduplex formations of different viral populations.

Participant Tz10

Participant Tz10 was infected by a subtype C virus in both the *vpu* (red) and *env* C2C3 (green) regions of the genome (Fig.3.3.3). Low diversity in the *vpu* HMA (Fig.3.3.3A) was detected over time compared to *env* C2C3 HMA where slow migrating heteroduplex bands were present in most samples (Fig.3.3.3B). All viral sequences were classified as subtype C (>89% bootstrap support; Fig.3.3.3).

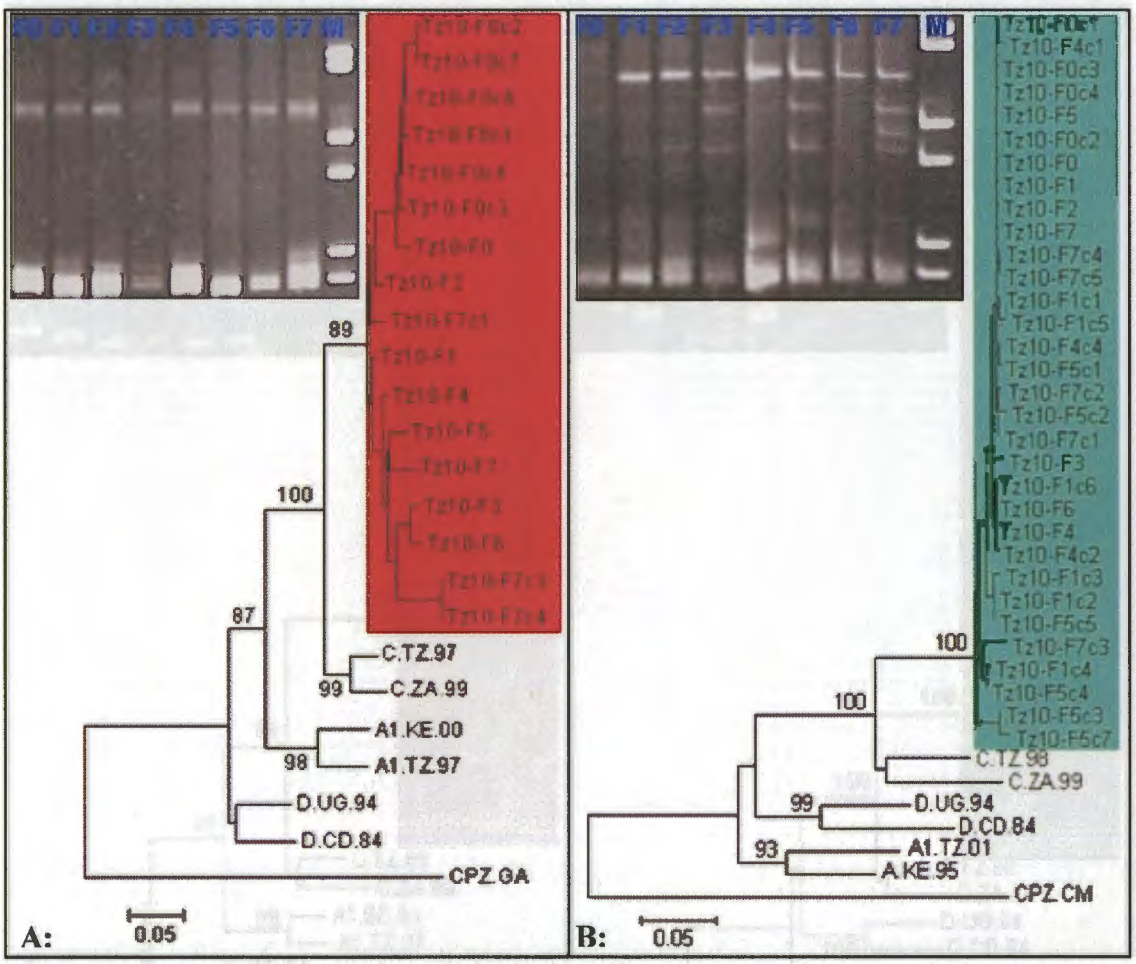


Figure 3.3.3: Tz10 *vpu* (red) and *env* C2C3 (green) neighbour joining trees from F0 (time point 0) to F7 (21 months post F0), scale bar indicates nucleotide divergence of 5% calculated from one hundred repeats. Only bootstrap values over 75% are shown. HMA gel analysis of the *vpu* (panel A) and *env* C2C3 (panel B) regions depicting homoduplex and heteroduplex formation respectively.

Participant Tz118

Participant Tz118 was infected by a subtype C virus in both the *vpu* (red) and *env* C2C3 (green) regions of the genome (Fig.3.3.4). Both the *vpu* and *env* C2C3 HMA analysis (Fig.3.3.4A and B; respectively) revealed the presence of more than one viral population at some time points, although these were relatively slow diversity as they migrated close to the homoduplex. All viral sequences were classified as subtype C (>95% bootstrap support; Fig. 3.3.4).

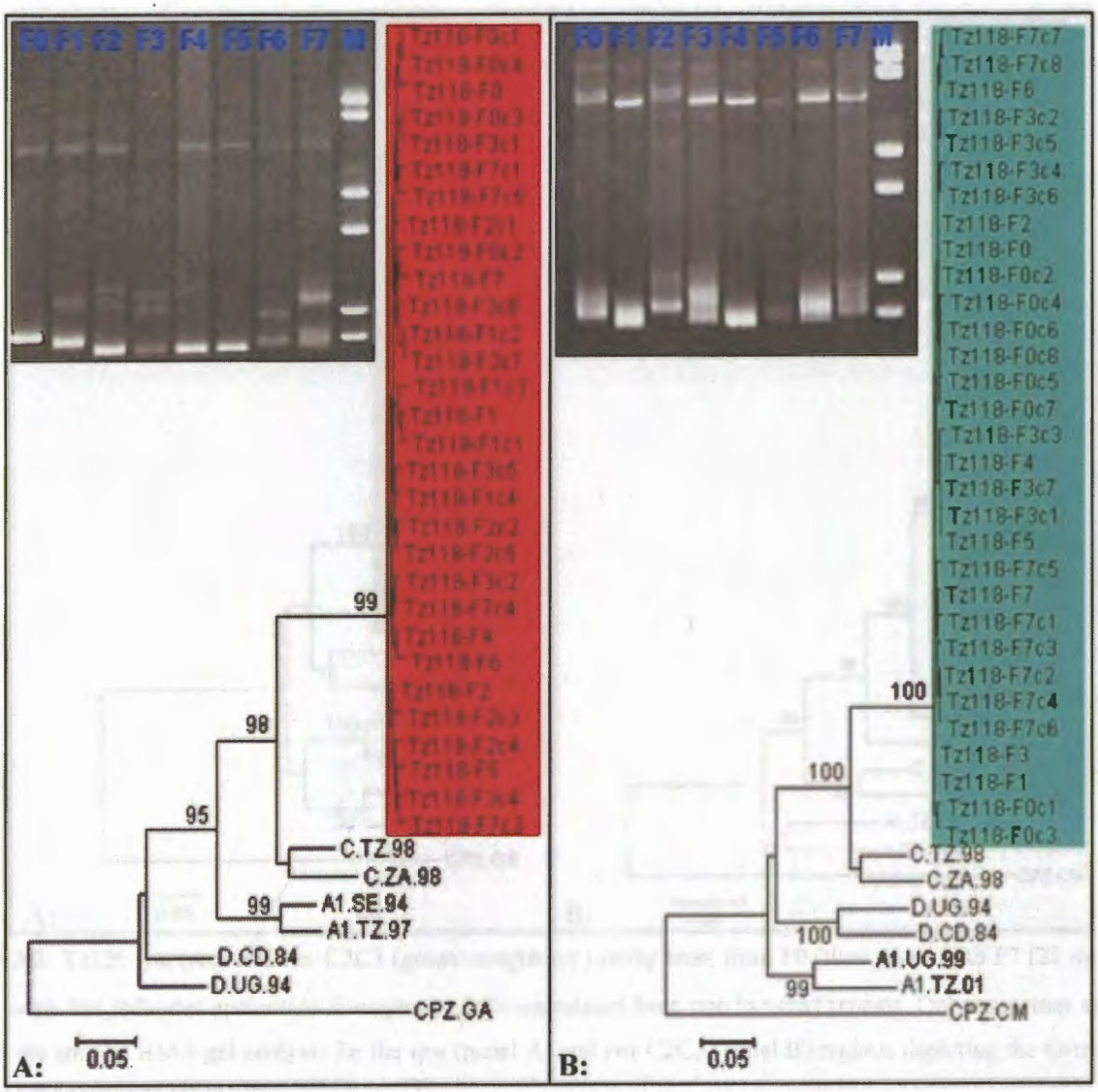


Figure 3.3.4: Tz118 vpu (red) and env C2C3 (green) neighbour joining trees from F0 (time point 0) to F7 (21 months post F0), scale bar indicates nucleotide divergence of 5% calculated from one hundred repeats. Only bootstrap values over 75% are shown. HMA gel analysis for the vpu region (panel A) and env C2C3 (panel B) depicting both the formation of homoduplexes and heteroduplexes.

Participant Tz126

Tz126 was infected with a subtype D virus in both the *vpu* (red) and *env* C2C3 (green) regions of the genome (Fig.3.3.5). The *vpu* HMA analysis (Fig.3.3.5A) revealed the presence of homogenous viral populations at all time points but in the *env* C2C3 HMA analysis (Fig.3.3.4B) some diversity was identified by slow migrating bands close to the homoduplex. All viral sequences observed were classified as subtype D (>95% bootstrap support; Fig.3.3.5).

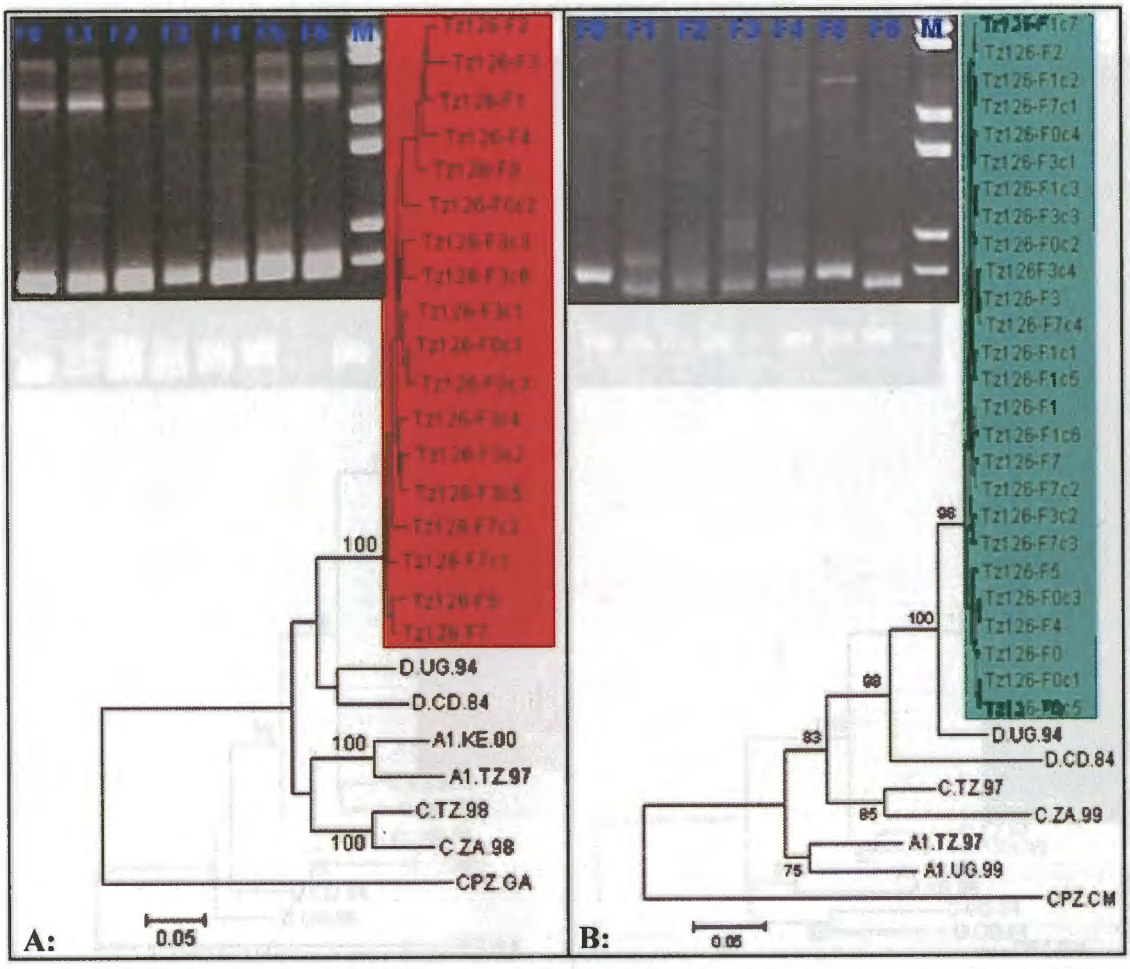


Figure 3.3.5: Tz126 *vpu* (red) and *env* C2C3 (green) neighbour joining trees from F0 (time point 0) to F7 (21 months post F0), scale bar indicates nucleotide divergence of 5% calculated from one hundred repeats. Only bootstrap values over 75% are shown. HMA gel analysis for the *vpu* (panel A) and *env* C2C3 (panel B) regions depicting the formation of homoduplexes and heteroduplexes respectively.

Participant Tz203

Participant Tz203 was infected by a subtype C virus in both the *vpu* (red) and *env* C2C3 (green) regions of the genome (Fig.3.3.6). The *vpu* HMA (Fig.3.3.5A) analysis revealed multiple viral populations in four (F0, F1, F4 and F5) time points. Multiple viral populations were also observed in all time points in the *env* C2C3 HMA analysis with slow migrating bands suggesting divergent viral populations (Fig.3.3.6B). However, all viral populations detected were classified as subtype C (>99% bootstrap support; Fig.3.3.6).

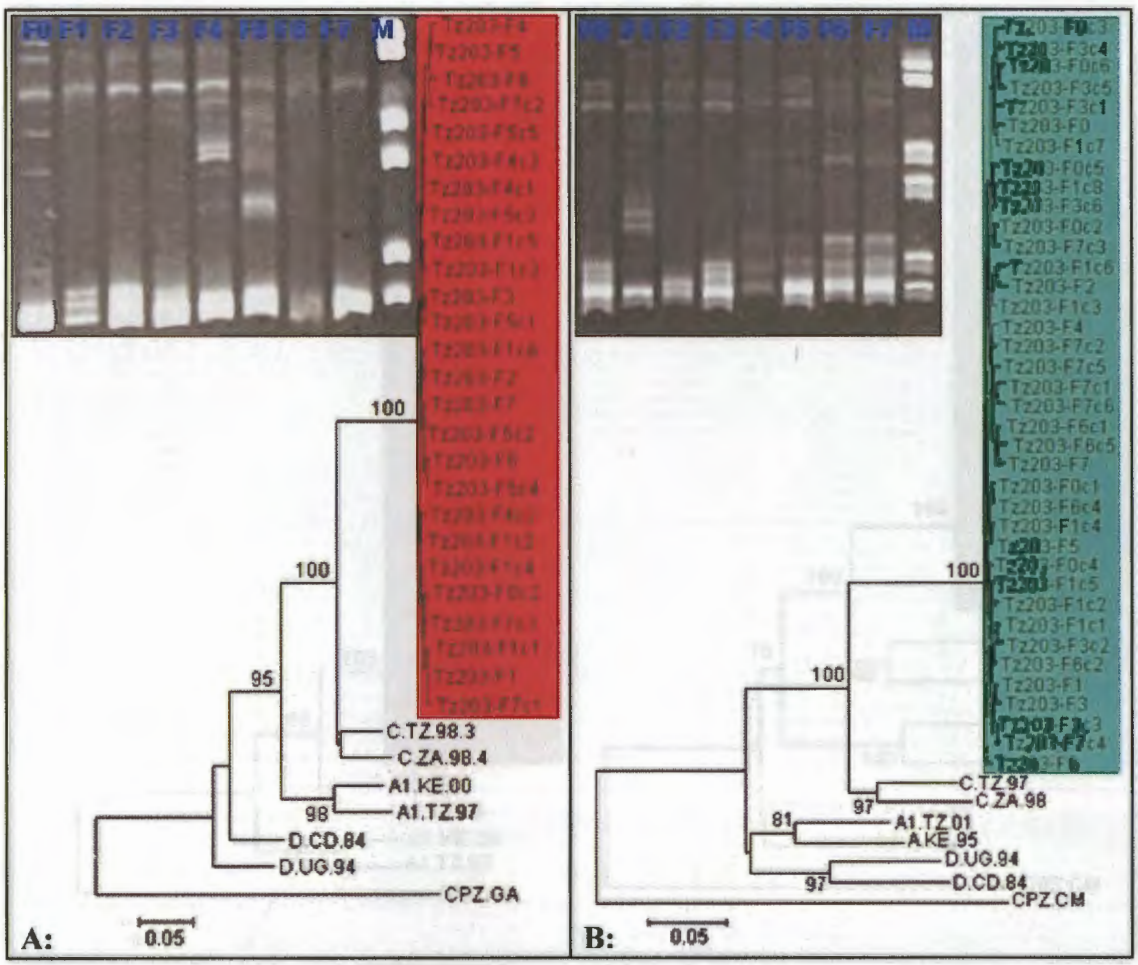


Figure 3.3.6: Tz203 *vpu* (red) and *env* C2C3 (green) neighbour joining trees from F0 (time point 0) to F7 (21 months post F0), scale bar indicates nucleotide divergence of 5% calculated from one hundred repeats. Only bootstrap values over 75% are shown. HMA gel analysis for the *vpu* (panel A) and *env* C2C3 (panel B) regions depicting the formation of homoduplexes and heteroduplexes respectively.

Participant Tz217

Tz217 was infected with a virus that displayed discordant subtype determination in the *vpu* (red) and *env* C2C3 (green) regions of the genome (Fig.3.3.7) with *vpu* been subtype C and *env* C2C3 classified as subtype D. The *vpu* HMA (Fig.3.3.7A) analysis revealed multiple viral populations but all viral sequences generated were subtype C. The *env* C2C3 HMA analysis (Fig.3.3.7B) on the other hand, revealed homogenous viral populations although faint slow migrating bands were identified in a few samples (F1, F2 and F4).

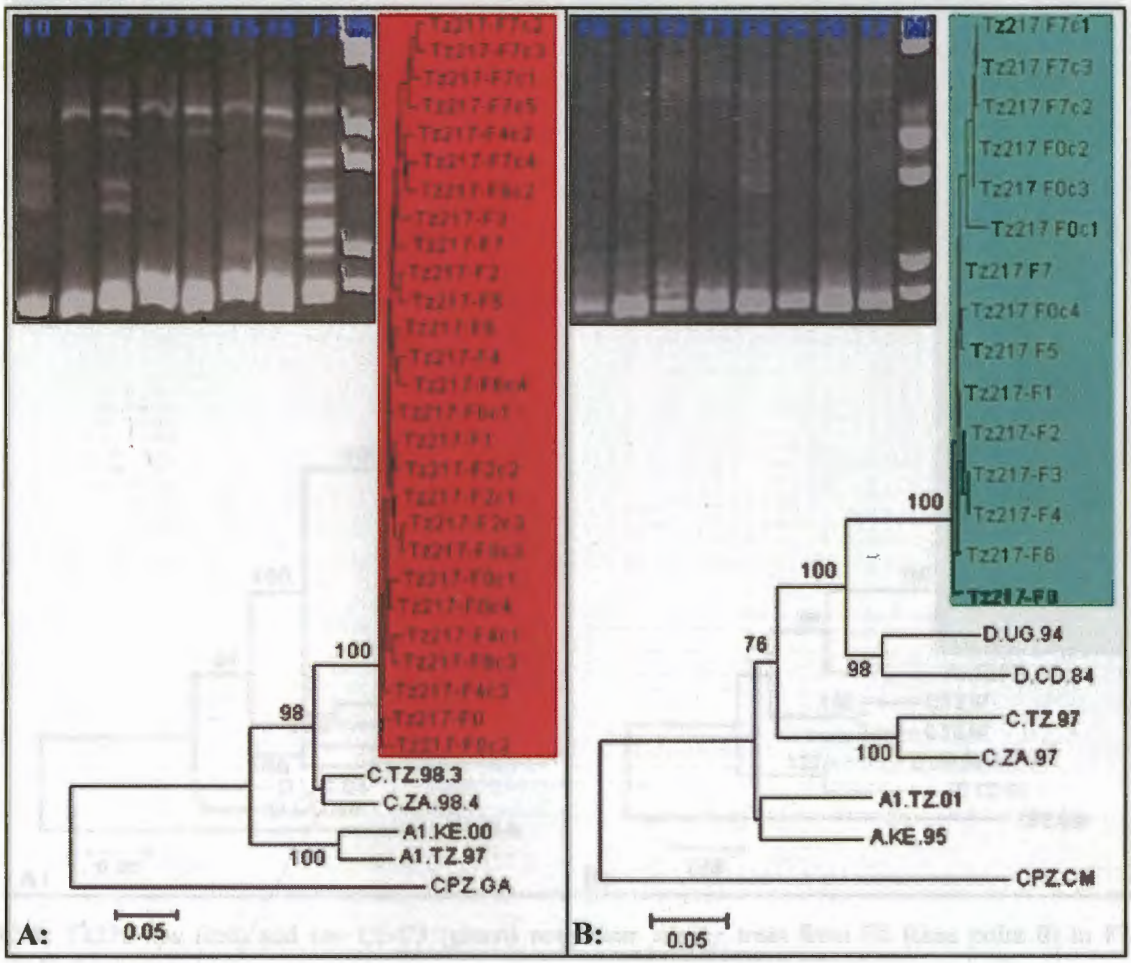


Figure 3.3.7: Tz217 *vpu* (red) and *env* C2C3 (green) neighbour joining trees from F0 (time point 0) to F7 (21 months post F0), scale bar indicates nucleotide divergence of 5% calculated from one hundred repeats. Only bootstrap values over 75% are shown. HMA gel analysis for the *vpu* region (panel A) and *env* C2C3 (panel B) depicting homoduplex and heteroduplex formations.

Participant Tz220

Participant (Tz220) was also infected with a recombinant virus that displayed discordant subtype determination in the *vpu* (red) and *env* C2C3 (green) regions of the genome (Fig.3.3.8), with *vpu* classified as subtype C and *env* C2C3 classified as subtype A. The *vpu* HMA (Fig.3.3.8A) analysis revealed the presence of multiple viral populations in some of the time points, while the *env* C2C3 HMA analysis (Fig.3.3.8B) revealed the presence of a relatively homogenous viral populations.

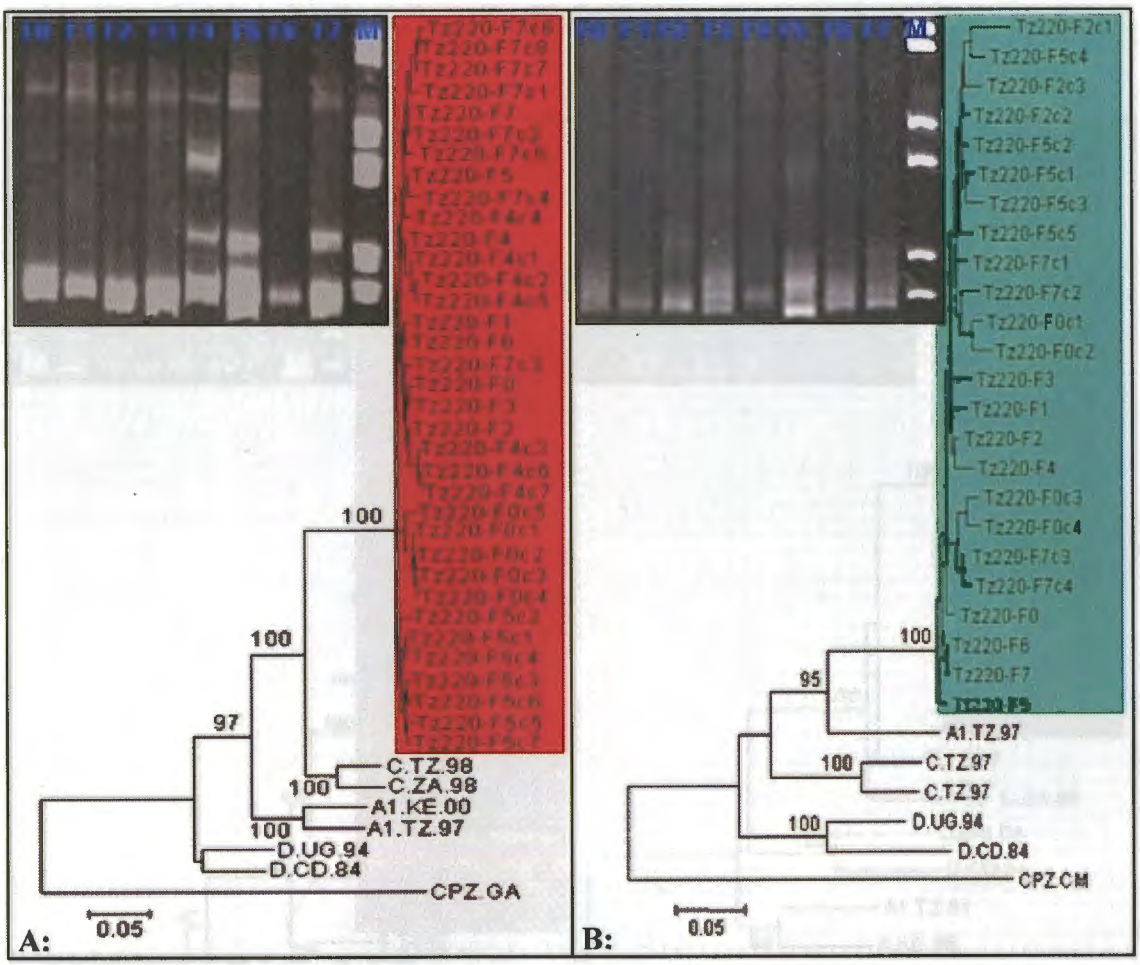


Figure 3.3.8: Tz220 *vpu* (red) and *env* C2-C3 (green) neighbour joining trees from F0 (time point 0) to F7 (21 months post F0), scale bar indicates nucleotide divergence of 5% calculated from one hundred repeats. Only bootstrap values over 75% are shown. HMA gel analysis for the *vpu* (panel A) and *env* C2C3 (panel B) regions depicting the formation of homoduplexes and heteroduplexes.

Participant Tz440

Tz440 was infected by a subtype C virus in both the *vpu* (red) and *env* C2C3 (green) regions of the genome (Fig.3.3.9). The *vpu* HMA (Fig.3.3.9A) analysis revealed the presence of multiple viral populations at most time points however, the analysis of the *env* C2C3 HMA revealed the presence of homogenous populations at most time points (Fig.3.3.9B). All the viral sequences were classified as subtype C (>93% bootstrap support; Fig.3.3.9B). Although three and two subtype C clusters were identified by phylogenetic analysis of the *vpu* and *env* C2C3 region respectively, dual infection with two subtype C strains was not confirmed.

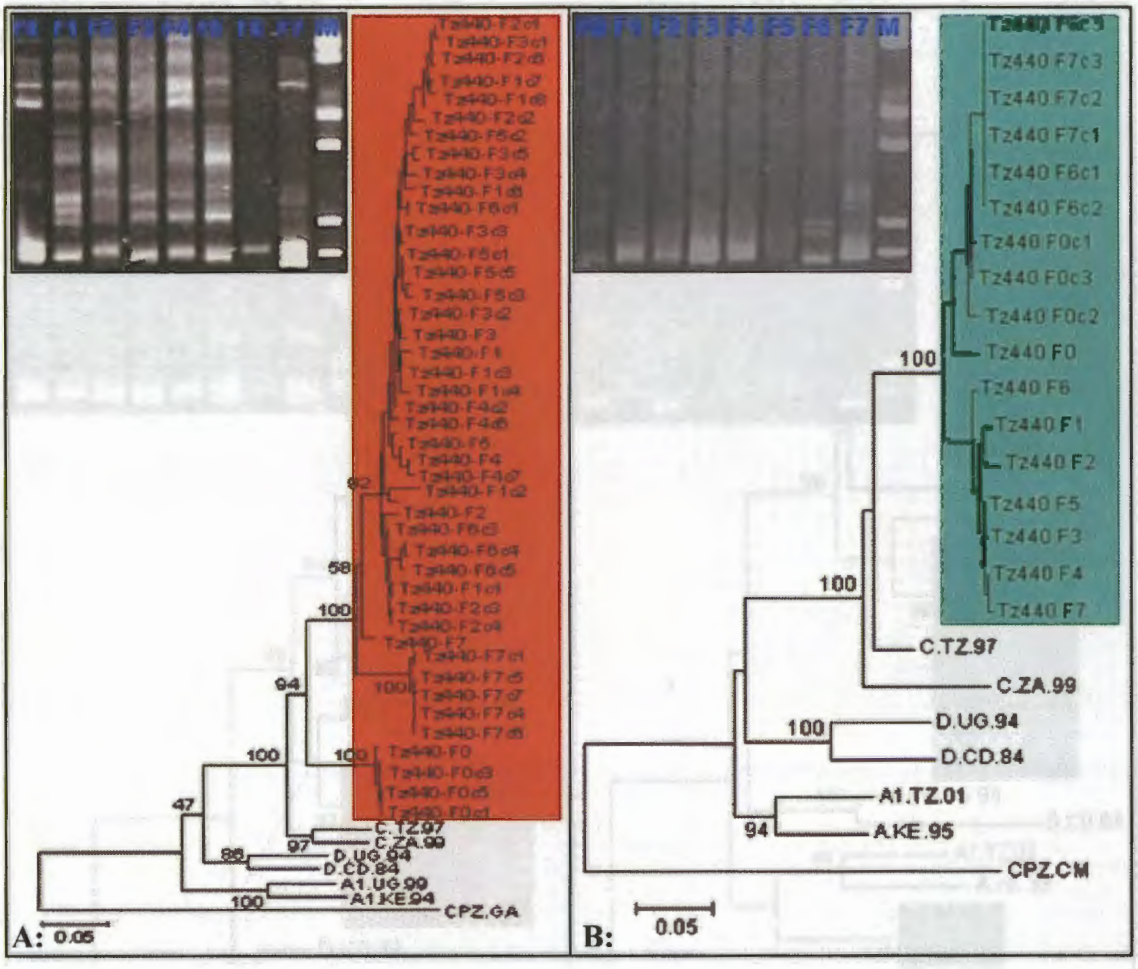


Figure 3.3.9: Tz440 *vpu* (red) and *env* C2C3 (green) neighbour joining trees from F0 (time point 0) to F7 (21 months post F0), scale bar indicates nucleotide divergence of 5% calculated from one hundred repeats. Only bootstrap values over 75% are shown. HMA gel analysis for the vpu (panel A) and *env* C2C3 (panel B) regions depicting the formation of homoduplexes and heteroduplexes.

Dual infections

Four individuals were infected with more than one subtype, with all dual infections comprising subtype A and C. These are described individually below.

Participant Tz14

Tz14 was dually infected by subtype A and C viruses in both the *vpu* (red) and *env* C2C3 (green) regions of the genome (Fig.3.3.10). In addition to the subtype A infection, two distinct subtype C clusters were identified in the *env* C2C3 region (>98% bootstrap support; Fig. 3.3.10). This individual will be investigated in more detail in chapter 4. In both genomic regions high diversity was evident, as indicated by the formation of both homoduplexes and heteroduplexes at certain time points.

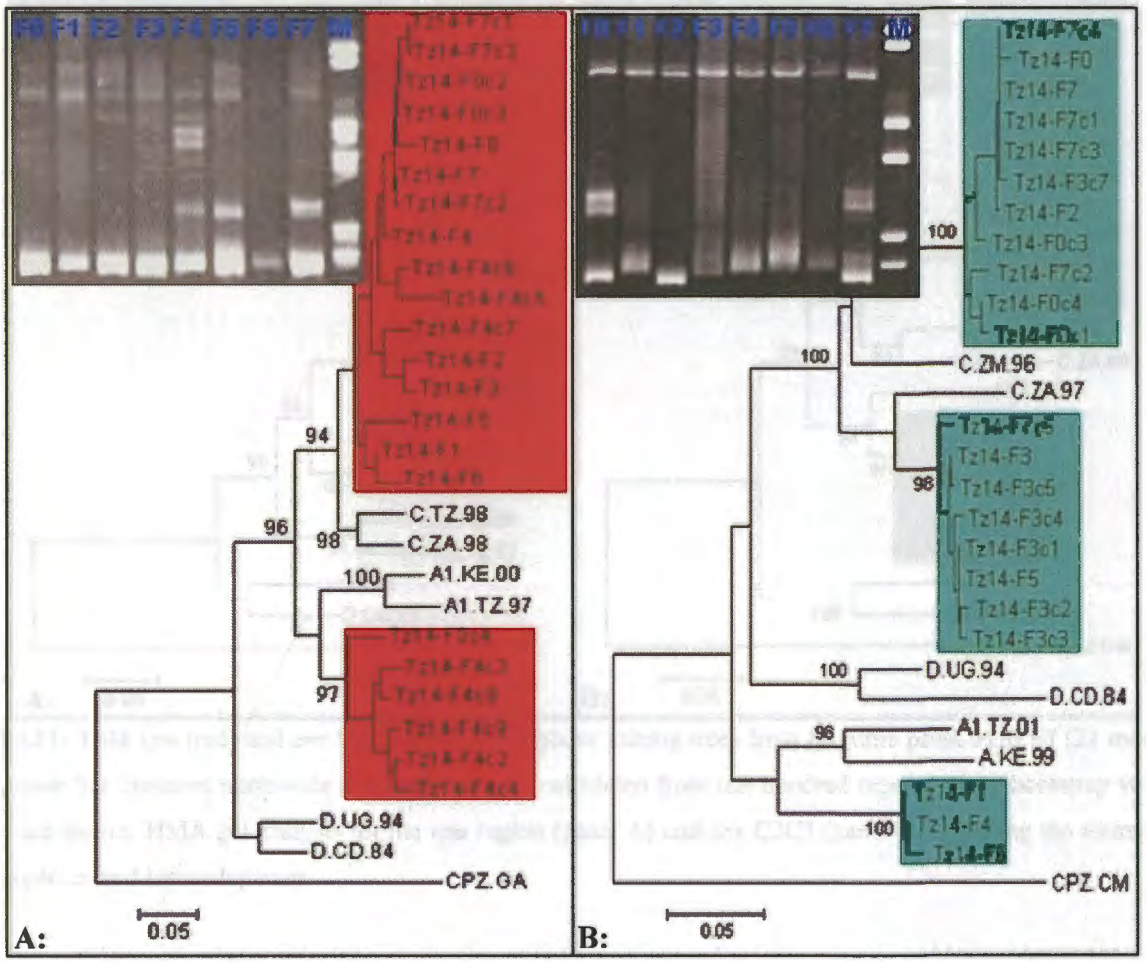


Figure 3.3.10: Tz14 *vpu* (red) and *env* C2C3 (green) neighbour joining trees from F0 (time point 0) to F7 (21 months most F0), scale bar indicates nucleotide divergence of 5% calculated from one hundred repeats. Only bootstrap values over 75% are shown. HMA gel analysis for the *vpu* region (panel A) and *env* C2C3 (panel B) depicting homoduplex and heteroduplex formations.

Participant Tz48

Participant Tz48 was infected with a virus that showed discordant subtypes in the *vpu* (red) and *env* C2C3 (green) regions of its genome (Fig.3.3.11), as well as dual infection in the *env* C2C3. All *vpu* sequences were subtype C while *env* C2C3 sequences revealed dual infection by both subtypes A and C. Vpu-HMA (Fig.3.3.11A) analysis revealed the presence of homogenous viral populations at all time points while some diversity was observed in the *env* C2C3 HMA analysis (Fig.3.3.11B).

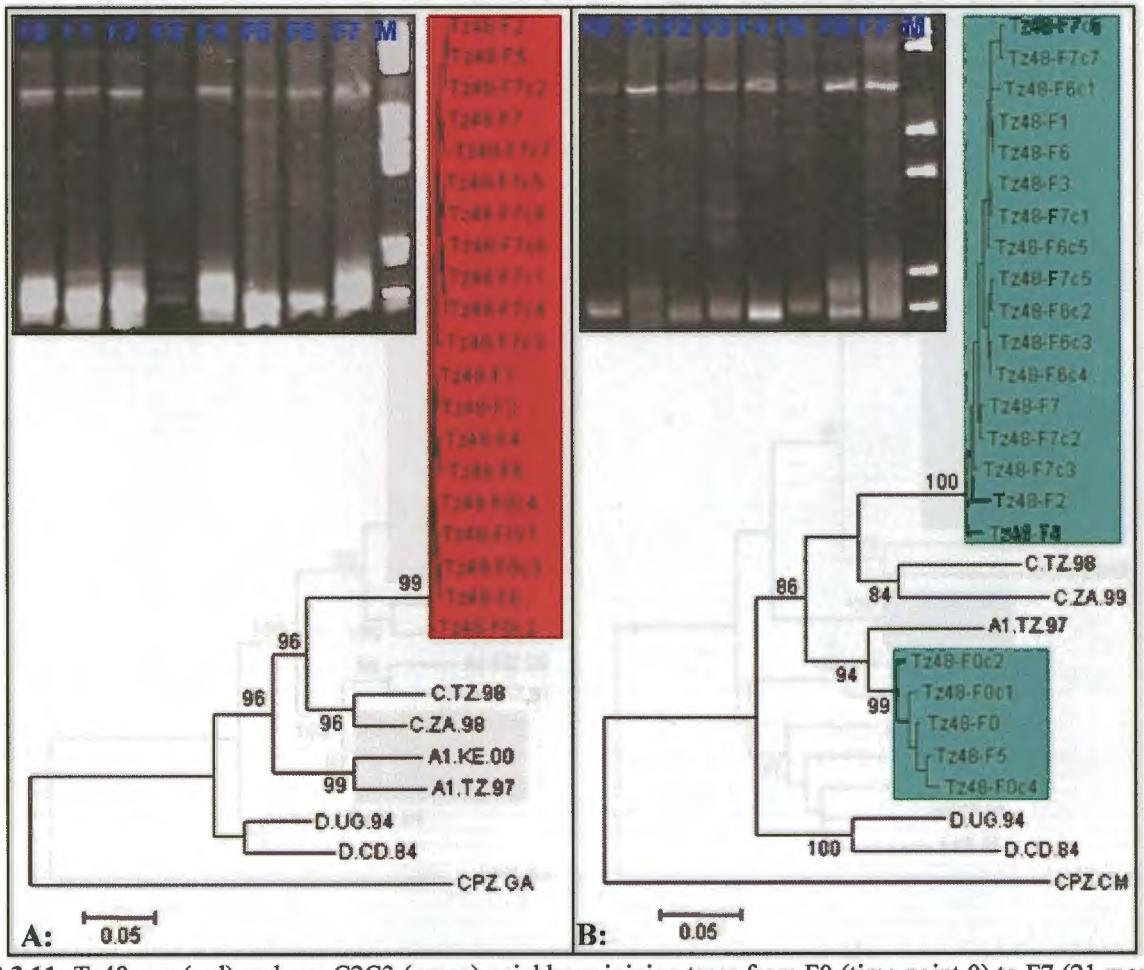


Figure 3.3.11: Tz48 *vpu* (red) and *env* C2C3 (green) neighbour joining trees from F0 (time point 0) to F7 (21 months post F0), scale bar indicates nucleotide divergence of 5% calculated from one hundred repeats. Only bootstrap values over 75% are shown. HMA gel analysis for the *vpu* region (panel A) and *env* C2C3 (panel B) depicting the formation of homoduplexes and heteroduplexes.

Participant Tz76

Participant Tz76 was dually infected by subtypes A and C viruses based on analysis of the *vpu* (red) region of the genome (Fig.3.3.12). Two *env* C2C3 sequences from the first visit did not group with any subtype however detailed analysis showed this region to contain a short stretch of subtype A-like sequences with a long stretch of subtype C-like sequences (data not shown). Both the *vpu* HMA (Fig.3.3.12A) and *env* C2C3 HMA analysis (Fig.3.3.12B) analysis revealed low diversity.

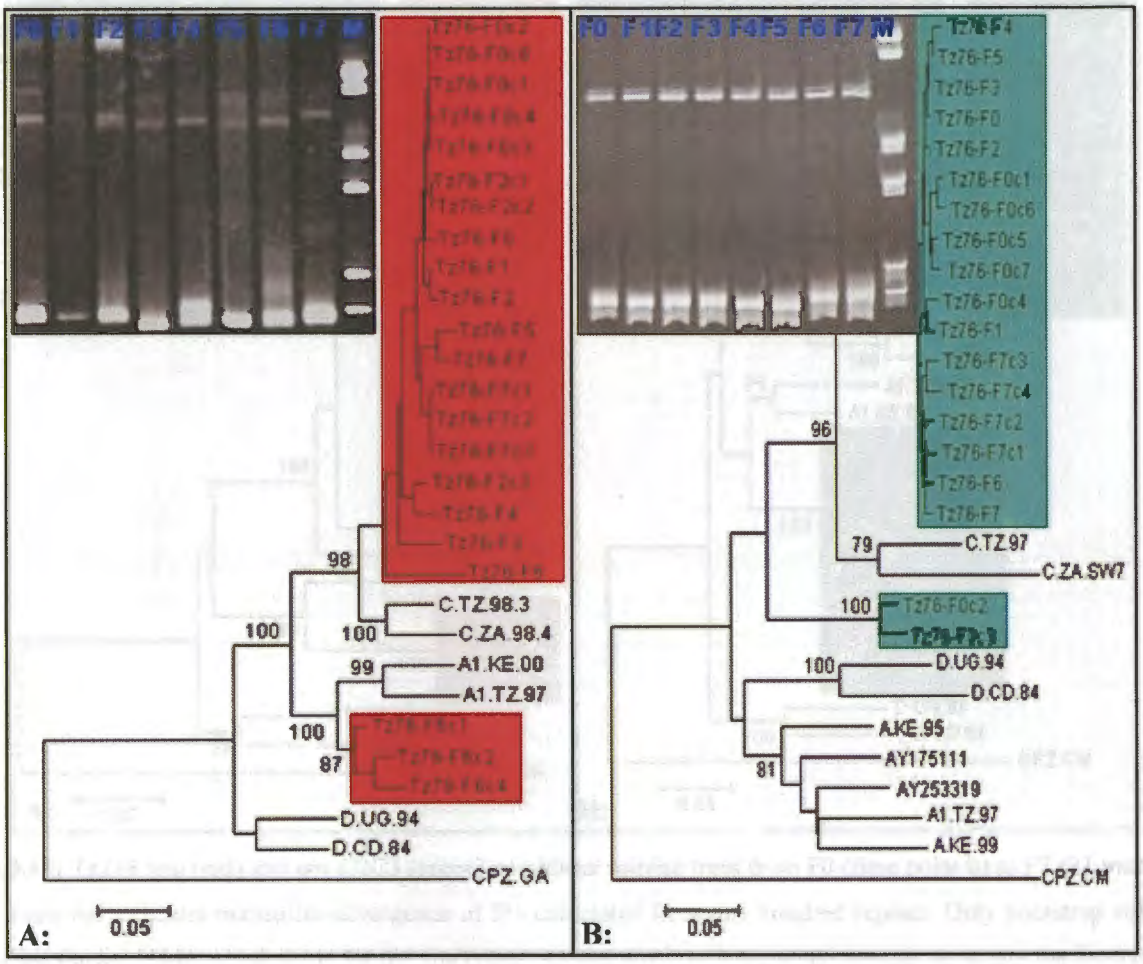


Figure 3.3.12: Tz76 *vpu* (red) and *env* C2C3 (green) neighbour joining trees from F0 (time point 0) to F7 (21 months post F0), scale bar indicates nucleotide divergence of 5% calculated from one hundred repeats. Only bootstrap values over 75% are shown. HMA gel analysis for the *vpu* region (panel A) and *env* C2C3 (panel B) depicting the formation of homoduplexes in both regions.

Participant Tz218

Participant Tz218 was dually infected with subtypes A and C viruses in both the *vpu* (red) and *env* C2C3 (green) regions of the genome (Fig.3.3.13). The *vpu* HMA and *env* C2C3 HMA (Fig.3.3.13A and B, respectively) analysis revealed diversity in most time points.

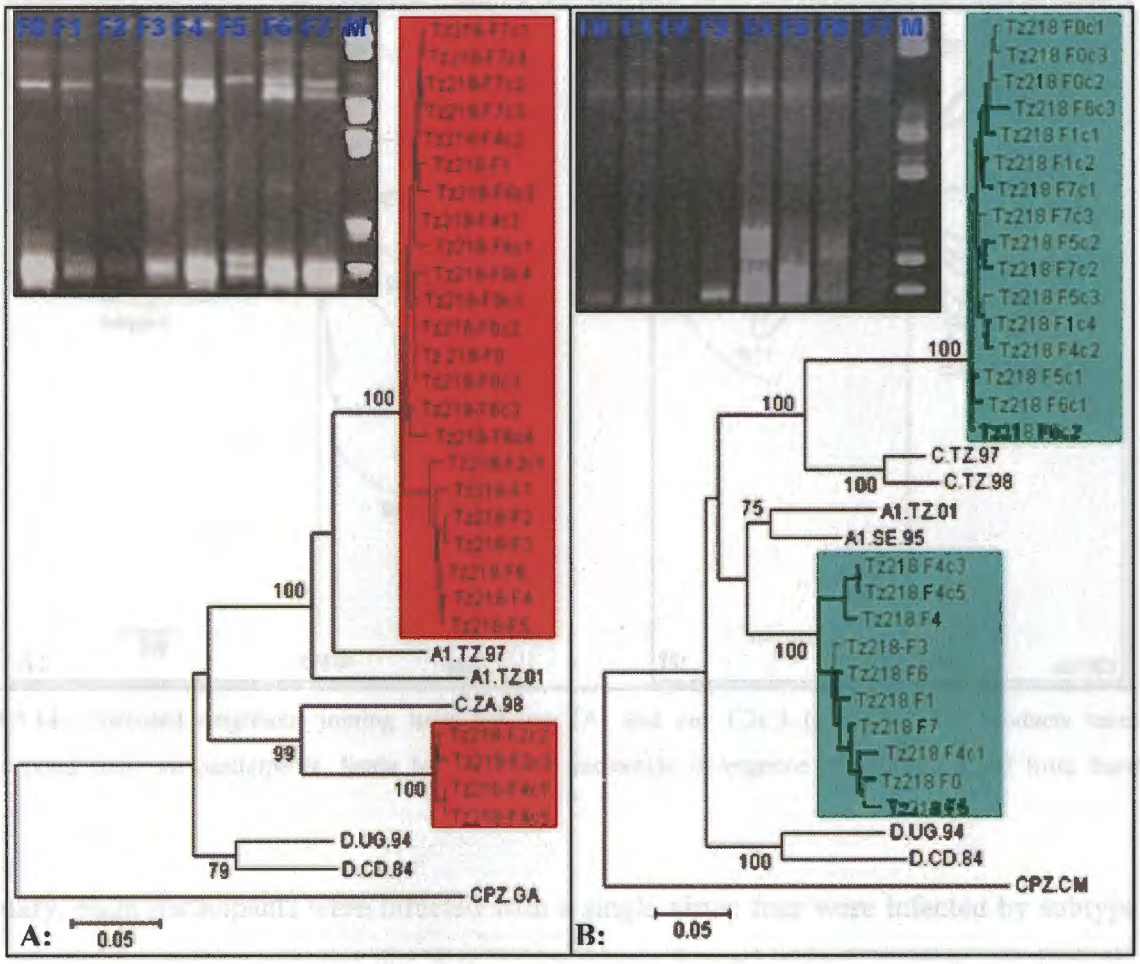


Figure 3.3.13: Tz218 *vpu* (red) and *env* C2C3 (green) neighbour joining trees from F0 (time point 0) to F7 (21 months post F0), scale bar indicates nucleotide divergence of 5% calculated from one hundred repeats. Only bootstrap values over 75% are shown. HMA gel analysis for the *vpu* (panel A) and *env* C2C3 (panel B) regions depicting the formation of homoduplexes and heteroduplexes.

The phylogenetic relationship among all bulk PCR sequences at each time point from all participants is depicted in Fig. 3.3.14. Sequences from participants grouped independently on phylogenetic trees indicating there was no evidence of contamination between sequences within the cohort.

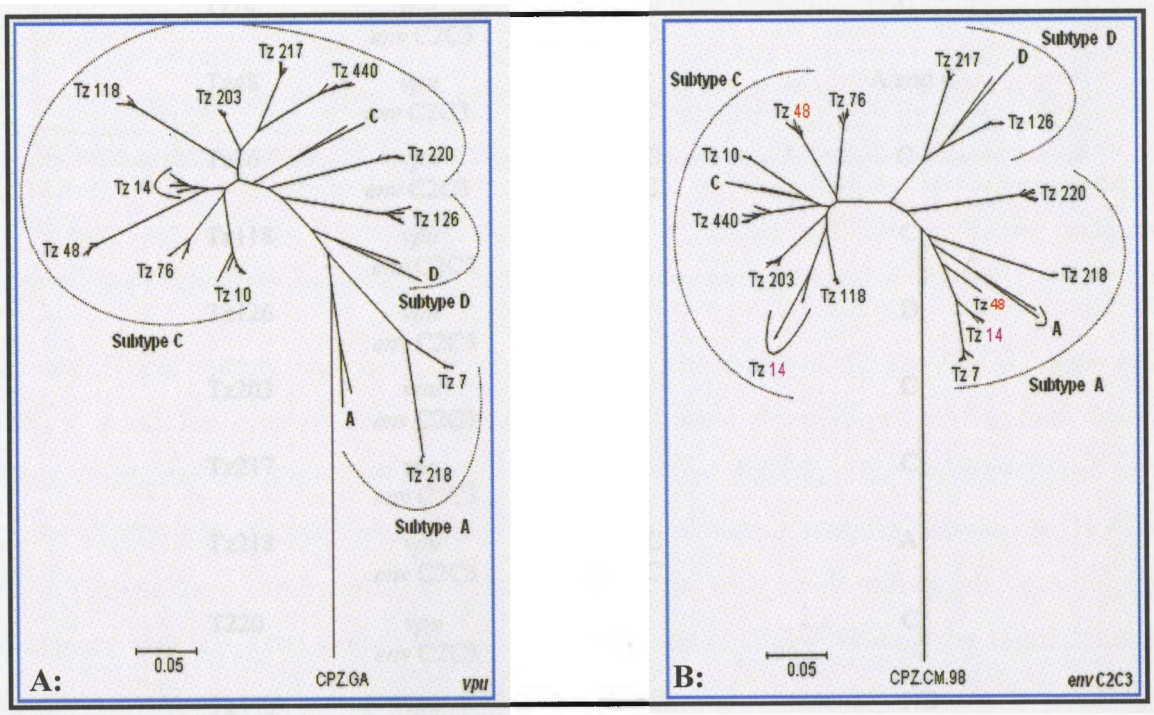


Figure 3.3.14: Unrooted neighbour joining trees for *vpu* (A) and *env* C2C3 (B) bulk PCR products taken at each time point from all participants. Scale bar indicates nucleotide divergence of 5% calculated from hundred repeats.

In summary, eight participants were infected with a single virus: four were infected by subtype C, two subtype A, one subtype D and one was infected by a recombinant CD virus. All four dually infected participants were infected by subtype A and C viruses (Table 3.3.1 and Fig. 3.3.14). When sequencing and MHA results were compared (Table 3.3.1), five participants showed concordant results, while discordant results were evidenced in five participants and no MHA results were available for two participants. While MHA is reported to be sensitive in detecting dual infections (Hoelscher *et al.*, 2002 and Arroyo *et al.*, 2004), it failed to identify dual infection in three individuals. HMA and sequencing failed to detect dual infection in two individuals which were detected by MHA. While the HMA was useful to determine diversity within an individual, in this study it was not an effective tool for detecting dual infections.

Table 3.3.1: Summary table of the subtypes detected by both sequencing and MHA.

Sample ID	Genomic Region	Subtype (sequencing)	Subtype (MHA)*
Tz7	*vpu*	A	NB
	env C2C3	A	
Tz10	*vpu*	C	A and C
	env C2C3	C	
Tz14	*vpu*	A and C	C
	env C2C3	A and C	
Tz48	*vpu*	C	A and C
	env C2C3	A and C	
Tz76	*vpu*	A and C	C
	env C2C3	A and C	
Tz118	*vpu*	C	C
	env C2C3	C	
Tz126	*vpu*	D	D
	env C2C3	D	
Tz203	*vpu*	C	C
	env C2C3	C	
Tz217	*vpu*	C	C
	env C2C3	D	
Tz218	*vpu*	A and C	A
	env C2C3	A and C	
T220	*vpu*	C	C
	env C2C3	A	
Tz440	*vpu*	C	NB
	env C2C3	C	

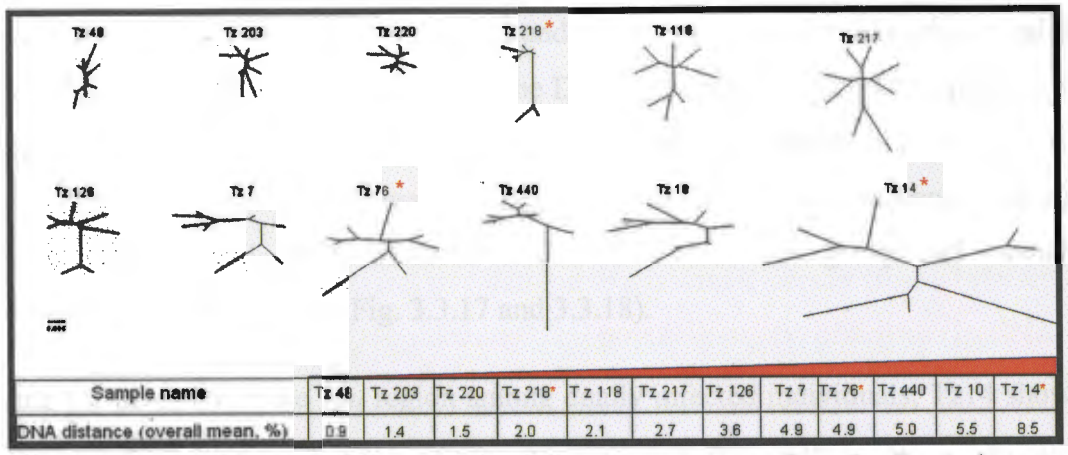
*MHA results of the *vpu* region provided by M. Hoelscher (University of Munich, Germany).

3.3.2 Measure of viral diversity

Divergence of dominant viral populations over time

To understand the divergence of the dominant virus populations over the twenty-two months of follow-up, we compared the sequences of the dominant virus represented by the bulk PCR sequence. The graphical representation of DNA distance within individuals provided a visual illustration of the space occupied by each radial tree, which represents the measure of diversity in each individual over the study period.

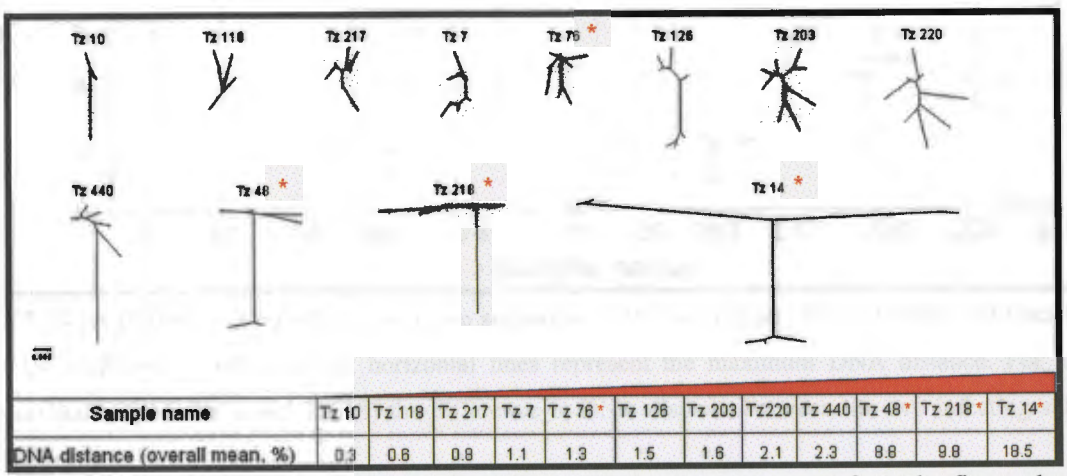
DNA distances between the dominant viruses of participants with single infections ranged from 0.9% (Tz48) to 5.5% (Tz10) while the dual infections [represented in asterisk (Figure 3.3.15)] had DNA distances ranging from 2.0% (Tz218) to 8.5% (Tz14). This lack of divergence in some dual infections suggests that infection with a second subtype is not impacting on the sequence of the dominant virus in these individuals and is not resulting in shifts in viral sequence over time.



Sample name	Tz 48	Tz 203	Tz 220	Tz 218*	T z 118	Tz 217	Tz 126	Tz 7	Tz 76*	Tz 440	Tz 10	Tz 14*
DNA distance (overall mean, %)	0.9	1.4	1.5	2.0	2.1	2.7	3.6	4.9	4.9	5.0	5.5	8.5

Figure 3.3.15: Neighbourjoining trees of the *vpu* bulk PCR product sequences from the first to last sample point (spanning 18 months of follow-up). All trees from dually infected individuals (Tz14, Tz76 and Tz218) are indicated with an asterisk. Scale bar indicates nucleotide divergence of 0.6% calculated from hundred repeats.

However a different pattern was observed when analysing diversity of the *env* C2C3 region where there was a trend for greater divergence of dominant viral populations in the dual infected individuals compared to those with single infection (Figure 3.3.16). The majority (three of the four) of the dually infected individuals had higher DNA distances ranging between 8.8% (Tz48) and 18.5% (Tz14) than those observed among individuals with single infections who displayed DNA distances ranging from 0.3% (Tz10) to 2.3 % (Tz440). Tz14 showed the highest level of diversity in this genomic region as represented by both the space occupied by its radial tree and had an overall mean DNA distance of 18.5%. Tz218 and Tz48 both dually infected had mean DNA distances of 9.8% and 8.8% respectively while Tz76 had only a mean DNA distance of 1.3% despite been dually infected. Tz10 displayed the least viral diversity in this genomic region, as represented by the space occupied by its radial tree and an overall DNA distance mean of 0.3% (Figure 3.3.16).



Sample name	Tz 10	Tz 118	Tz 217	Tz 7	T z 76*	Tz 126	Tz 203	Tz220	Tz 440	Tz 48*	Tz 218*	Tz 14*
DNA distance (overall mean, %)	0.3	0.6	0.8	1.1	1.3	1.5	1.6	2.1	2.3	8.8	9.8	18.5

Figure 3.3.16: Neighbour joining trees of the *env* C2C3 bulk PCR product sequences from the first to last sample point (spanning 18 months of follow up). All trees from dually infected individuals (Tz14, Tz76 and Tz218) are indicated with an asterisk. Scale bar indicates nucleotide divergence of 0.6% calculated from hundred repeats.

To take into consideration the viral populations and further understand the level of viral diversity within each individual over time, we measured the DNA distances between all sequences available at individual time points. For singly infected participants, the median and highest DNA distances were determined within a group. Whereas, for dually infected participants sequences were grouped according to subtypes and DNA distances calculated within each group and medians were compared between the two groups (Fig. 3.3.17 and 3.3.18).

In analyzing the diversity in the *vpu* region among dually infected individuals; Tz14 showed high viral diversity in both subtype A and C of the infecting viral populations with a maximum DNA distance of 12% and a median of 7% for subtype A and the maximum of 13% and a median of 9% for subtype C. The difference in DNA distance between the two subtypes was 20% for the Tz14 viral population (Figure 3.3.17). Tz76 was infected with subtype A and C, subtype C had a maximum DNA distance of 9% and a median of 3% while subtype A had a maximum of 4% and a median of 2%, with a difference of 21.5% in DNA distance between the two subtypes. Finally Tz218 was infected with subtype A and C and had a maximum DNA distance of 4% and a median of 2% for subtype A and the maximum of 1.9% and a median of 0.7% for subtype C with a DNA distance of 25.6% between the two subtypes (Figure 3.3.17).

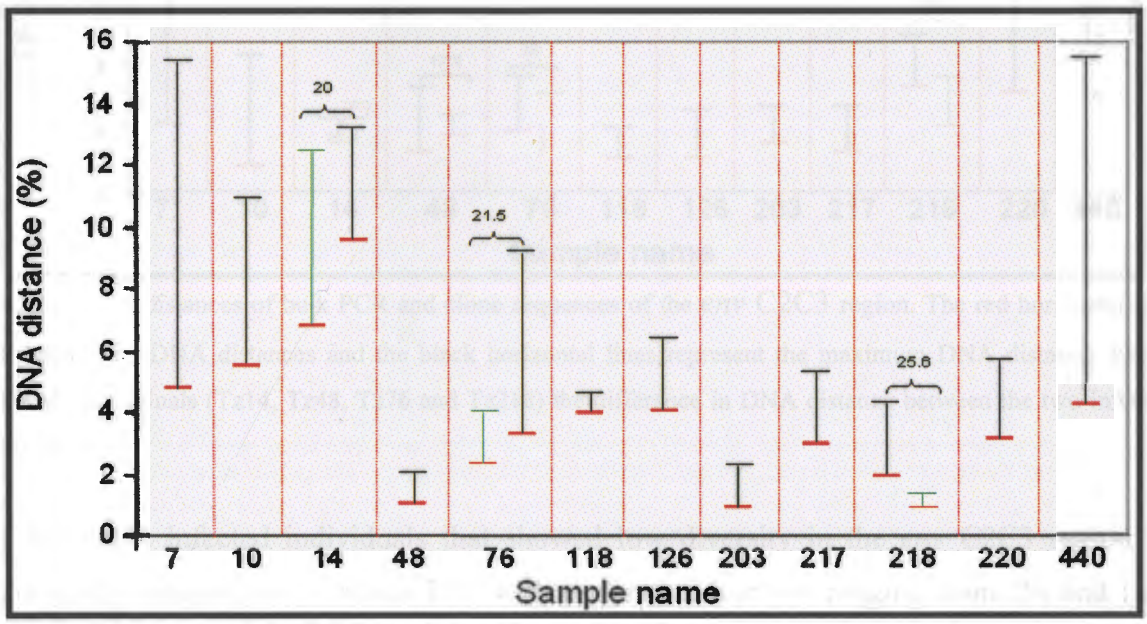


Figure 3.3.17: DNA distances of bulk PCR and clone sequences of the *vpu* region. The red horizontal lines represent the median DNA distances and the black horizontal lines represent the maximum DNA distance. For all dually infected individuals (Tz14, Tz76 and Tz218) the difference in DNA distance between the two infecting subtypes is shown.

Among the singly infected individuals that showed low diversity in the *vpu* region the median diversity ranged from 1% to 3.6% with maximum diversity ranging from 2.3% to 4.2%. Tz7, Tz10 and Tz440 had the highest diversity in this region (Figure 3.3.17).

In analyzing the diversity in the *env* C2C3 region among dually infected individuals; Tz14 was observed to harbour highly diverse viral populations. High viral diversity (max. 14% and median 12%) was observed in subtype C of the infecting viral populations, the difference in DNA distance between the two subtypes was 28.1% (Figure 3.3.18). The differences in DNA distances between the infecting subtypes in other dually infected individuals was: 17.8% for Tz48; 20% for Tz76; and 28.2% for Tz218. All single infected individuals had DNA distances of less than 2% in the *env* C2C3 region (Figure 3.3.18).

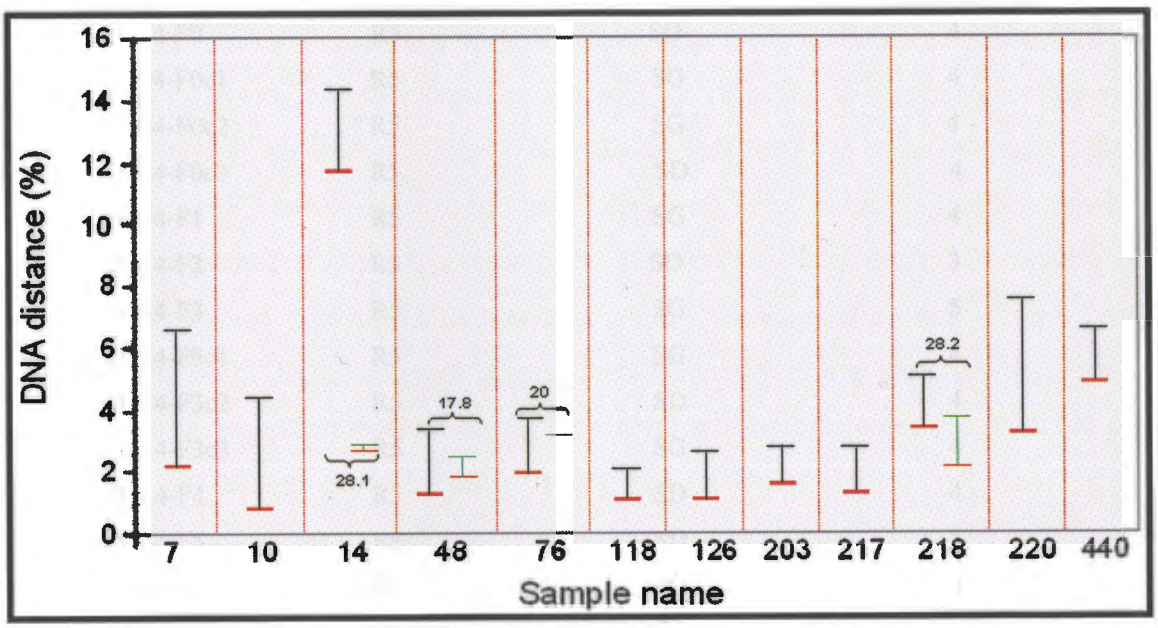


Figure 3.3.18: DNA distances of bulk PCR and clone sequences of the *env* C2C3 region. The red horizontal lines represent the median DNA distances and the black horizontal lines represent the maximum DNA distance. For all dually infected individuals (Tz14, Tz48, Tz76 and Tz218) the difference in DNA distance between the two infecting subtypes is shown.

Among the singly infected individuals that showed low diversity in the *env* C2C3 region the median diversity ranged from 1.5% to 12% with maximum diversity ranging from 2% and 14%. Tz7, Tz10 and Tz440 had the highest diversity in this region (Figure 3.3.18).

3.3.3 Viral phenotype and coreceptor usage

Co-receptor usage was predicted using the position specific scoring matrix (PSSM) method developed for HIV subtype C sequences as described by Jensen *et al.*, (2006). The C-PSSM had a reported specificity of 94% and sensitivity of 75% (Jensen *et al.*, 2006). PSSM uses background

genetic variation as a baseline comparison, or "null model," to facilitate comparison of the residues of a sequence fragment to those of a group of aligned sequences known to have the desired property. The comparison leads to a score that can be interpreted as indicating the likelihood that the sequence fragment has the property of interest. Using the PSSM a sequence can be assigned a score: the higher the score, the more closely the sequence resembles those of known X4 viruses (Jensen *et al.*, 2006). In our study all viral populations both in single and dual infections were predicted to be R5 viruses at all time points irrespective of subtype. This was true for subtypes A, C and D viruses alike. V3 viral sequences from a dually infected individual (Tz14) are illustrated in Table 3.3.2.

Table 3.3.2: V3 sequence phenotype prediction of Tz14 as detected by PSSM.

Sample ID	Subtype prediction	Residues at position 11 and 25	Net charge
Tz14-F0	R5	SG	4
Tz14-F0c1	R5	SG	4
Tz14-F0c2	R5	SG	4
Tz14-F0c3	R5	SD	4
Tz14-F1	R5	SG	4
Tz14-F2	R5	SD	3
Tz14-F3	R5	SG	5
Tz14-F3c1	R5	SG	4
Tz14-F3c2	R5	SD	4
Tz14-F3c3	R5	SG	4
Tz14-F4	R5	SD	4
Tz14-F5	R5	SD	5
Tz14-F6	R5	SG	4
Tz14-F7	R5	SG	4
Tz14-F7c1	R5	SG	4
Tz14-F7c2	R5	SG	4
Tz14-F7c3	R5	SG	3

The presence of residues lysine (K), arginine (R) and histidine (H) at positions 11 and 25 = X4 ; net charge : >5 = X4 and <5 = R5

A majority of the subtype C sequences had a conserved GPGQ V3 crown at all time points sampled, however Tz10 (Fig. 3.3.19B) was an exception having on its V3 crown the GPGR sequence. The N-linked glycosylation sites in the V3 were only conserved by Tz10 subtype C sequences (Fig. 3.3.19B), this region lays an important role in proper folding, infectivity and evasion of the immune responses (Wei *et al.*, 2003 and Li *et al.*, 2001). Figure 3.3.19 depicts examples of the V3 amino acid sequences of subtypes A (Tz7), C (Tz10) and D (Tz126) infected individuals.

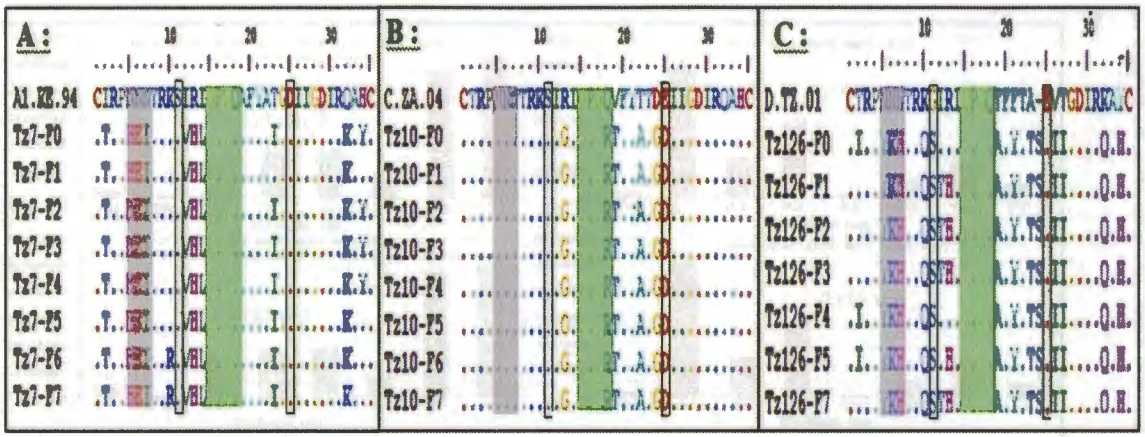


Figure 3.3.19: V3 amino acid sequences from participants Tz7, Tz10 and Tz126. Highlighted are the GPGQ crown (green) of the V3 loop, the N-linked glycosylation sites (gray) and residues at positions 11 and 25.

3.3.4 Temporal fluctuations in viral variants

To monitor the contribution of each viral variant to total viral load, we defined all sequences (bulk and clone) from both genomic regions (*vpu* and *env* C2C3) according to a defined scoring system. For each genomic region, sequences with DNA distances less than 1% were designated into one group (represented by different shades of red); sequences with DNA distances ranging between 1 - 3% were designated to another group (represented by different shades of green); DNA distances of between 3 - 5% (represented by different shades of yellow) designated a third group of variants and finally variants with DNA distances greater than 5%, representing different subtypes were designated to a fourth group (represented by different shades of blue colours). Appendix H depicts the phylogenetic trees (colour coded for each group) used for the scoring system. Thus, using the scoring system described above the proportion of each variant was determined. The contribution of each variant to the overall viral burden at each time point was plotted (Figures 3.3.20 and 3.3.21) with the height of the bar representative of viral load. This provides a visual way of assessing the amount of diversity contributing to viral burden within each individual in the different time points, as well as shifts in viral variant populations over time.

Single infections

Most of the individuals with single infection had one or two dominant variants (Fig. 3.3.20) with low viral diversity at all time points. Usually high viral load was associated with the outgrowth of a highly homogeneous viral population (red bars, Fig 3.3.20). In seven out of eight participants the viral diversity in the *vpu* and *env* C2C3 was similar, however in participant Tz7, although viral populations represented by *vpu* were highly homogenous, heterogenous populations were identified based on the *env* C2C3. This suggesting different selective pressure on the *env* C2C3 compared to the *vpu* resulting in differential diversification.

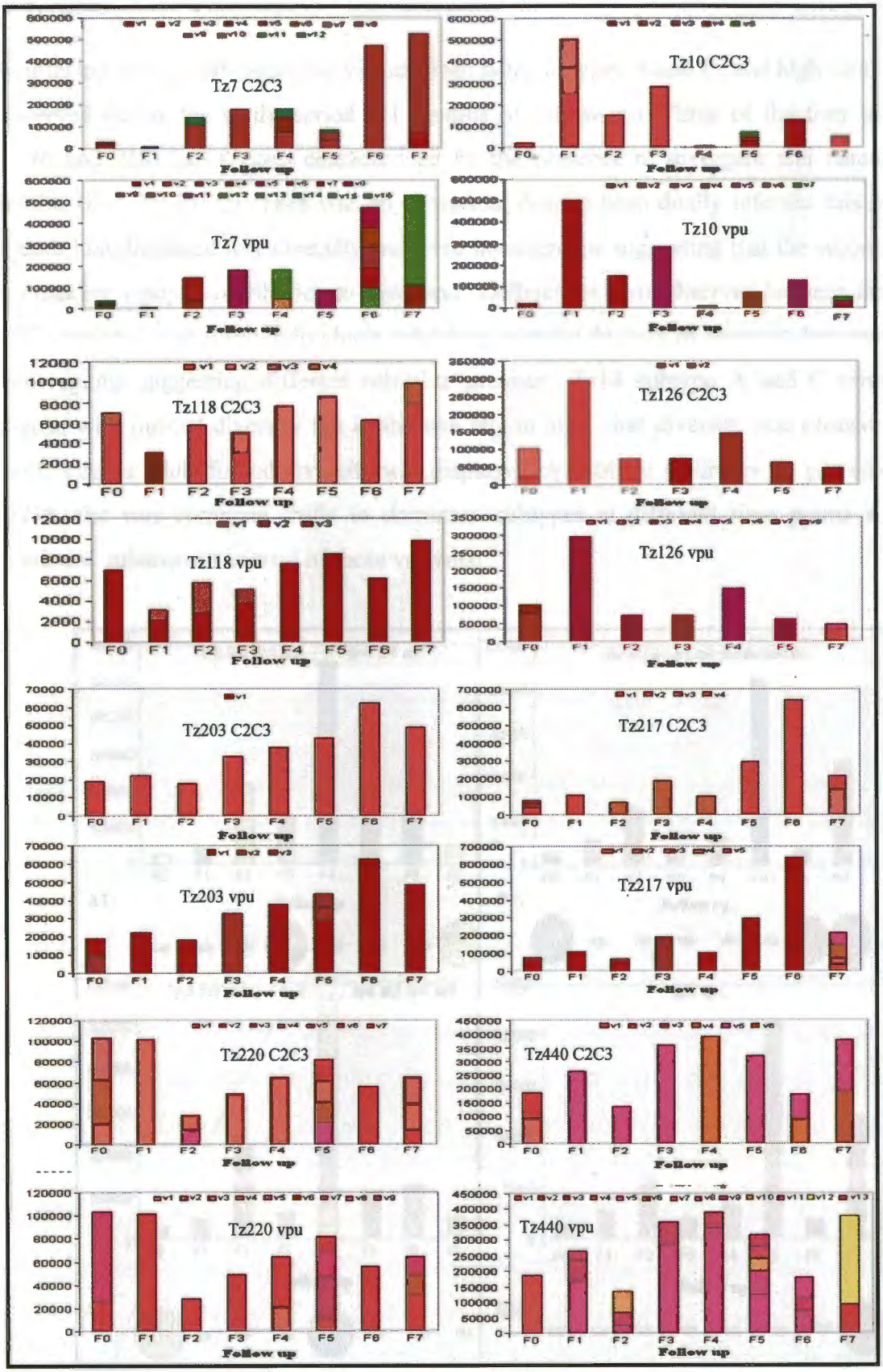


Figure 3.3.20: Compositions of the viral variants from F0 to F7 are shown for all individuals with single infection. The height of the bars is representative of the viral load (y-axis). The legend indicates the colours used to represent the various viral variants.

Dual infections

Dually infected individuals displayed viruses from both subtypes A and C, and high viral diversity was observed during the study period (21 months of follow-up). Three of the four individuals (Tz14, 76 and 218) had viruses characterized by the presence of divergent and heterogeneous populations (Fig. 3.3.21/22). Tz48 was an exception, despite been dually infected this individual had viruses that displayed low diversity and were homogenous suggesting that the second subtype was not making a major contribution to viral load. Differences were observed between the *vpu* and *env* C2C3 regions, with some individuals exhibiting varying degrees of diversity between the two genomic regions suggesting different selective pressure. Tz14 subtype A and C viruses were homogeous with limited diversity but in the *vpu* region high viral diversity was observed among subtype C viruses while limted diversity was displayed by subtype A viruses. In participant Tz76 and Tz218, the was complete shifts in dominant subtypes at different time points suggesting outgrowth and subsequent control of these variants.

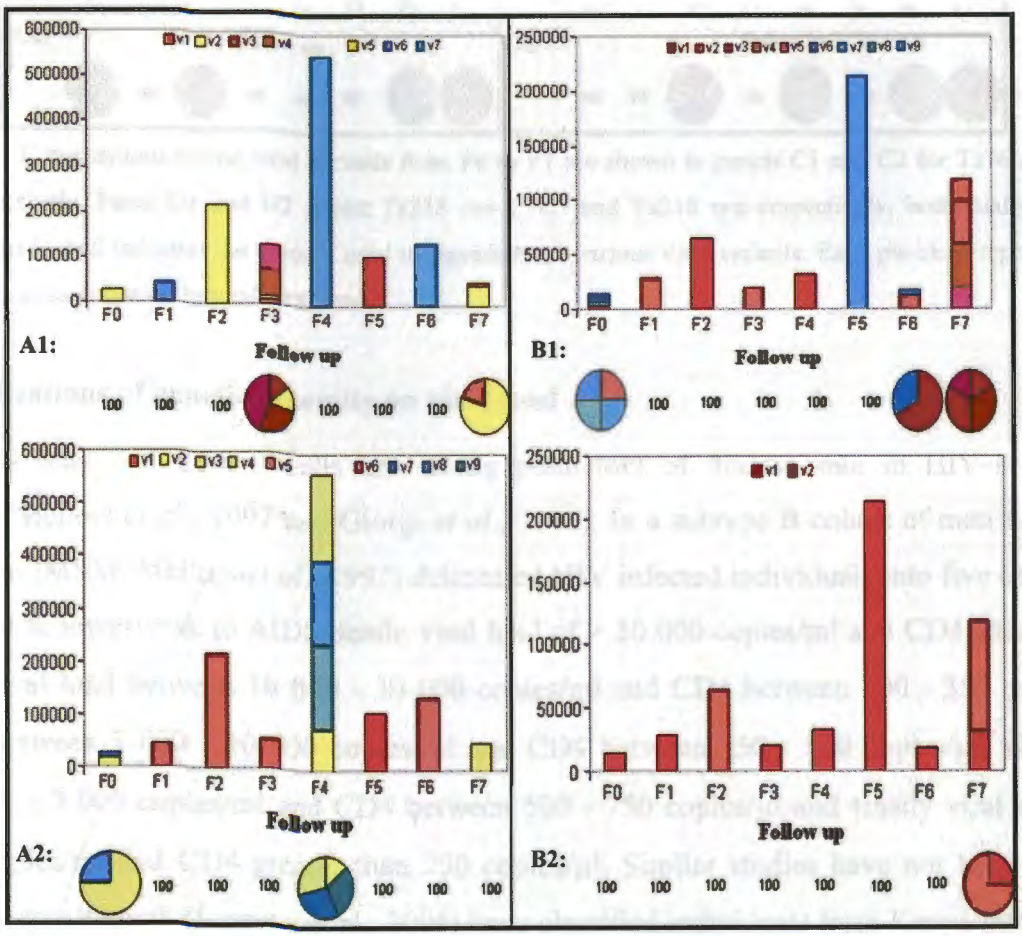


Figure 3.3.21: Compositions of the viral variants from F0 to F7 are shown in panels A1 and A2 for Tz14 *env* C2C3 and *vpu* respectively. Panels B1 and B2 depict Tz48 *env* C2C3 and Tz48 *vpu* respectively, both dually infected individuals. The legend indicates the colours used to represent the various viral variants. Each pie chart represents the actual percentage contribution to total viral load.

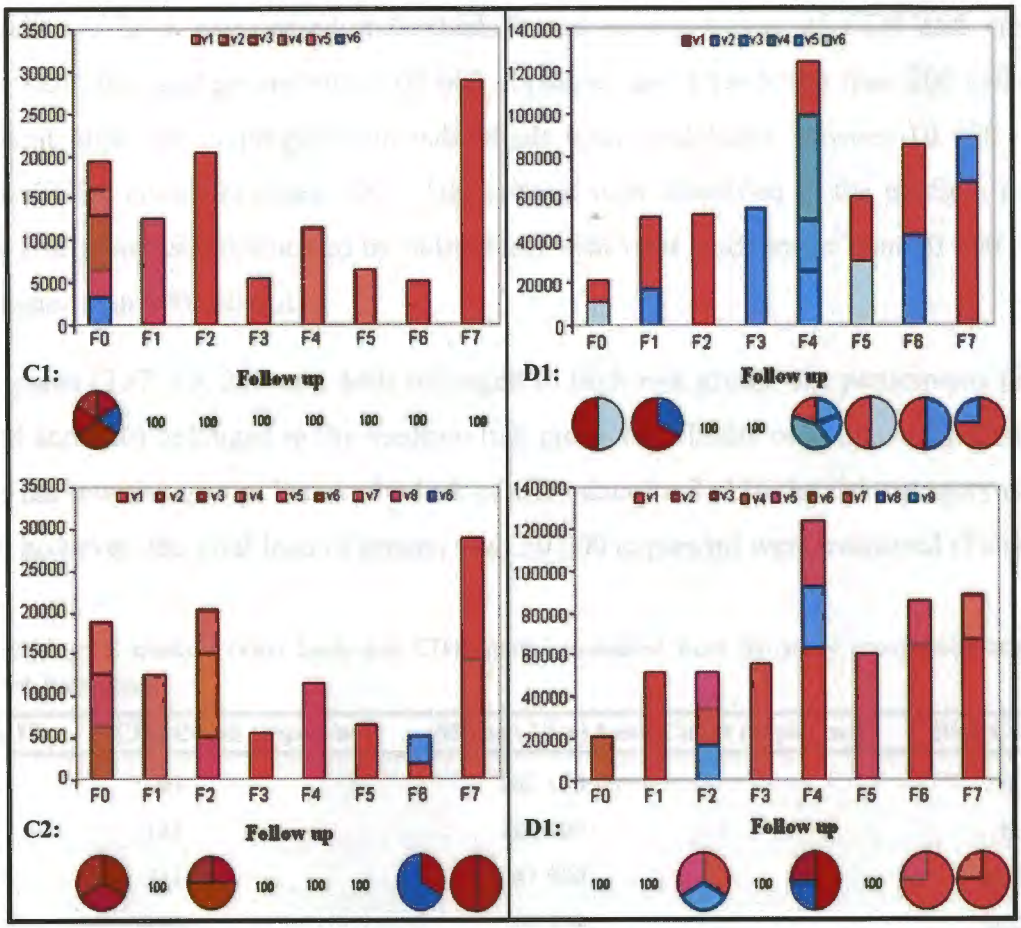


Figure 3.3.22: Compositions of the viral variants from F0 to F7 are shown in panels C1 and C2 for Tz76 *env* C2C3 and *vpu* respectively. Panel D1 and D2 depict Tz218 *env* C2C3 and Tz218 *vpu* respectively, both dually infected individuals. The legend indicates the colours used to represent the various viral variants. Each pie chart represents the actual percentage contribution to total viral load.

3.3.5 Implications of genetic diversity on viral load

Plasma viral load and CD4 T-cells are strong predictors of disease rate in HIV-1 infected individuals (Mellors *et al.*, 1997 and Giorgi *et al.*, 2002). In a subtype B cohort of men who have sex with men (MSM, Mellors *et al.*, 1997) delineated HIV infected individuals into five categories from highest to lowest risk to AIDS/death: viral load of > 30 000 copies/ml and CD4 count < 200 copies/µl, viral load between 10 000 - 30 000 copies/ml and CD4 between 200 - 350 copies/µl, viral load between 3 000 - 10 000 copies/ml and CD4 between 350 - 500 copies/µl, viral load between 500 - 3 000 copies/ml and CD4 between 500 - 750 copies/µl and finally viral load less than 500 copies/ml and CD4 greater than 750 copies/µl. Similar studies have not been done in African cohorts although (Lavreys et al., 2006) have classified individuals from Kenya infected for less than 2 years with viral loads lower than 10 000 copies/ml as in low risk of progression, individuals with viral loads greater than 100 000 copies/ml were classified as high risk and those with viral loads in between were designated to medium risk group.

For our study we have categorised individuals based on composite of CD4 and viral loads: individuals with viral load greater than 100 000 copies/ml and CD4 lesser than 200 cells/μl were classified as at high risk of progression; individuals with viral loads between 10 000 – 100 00 copies/ml and CD4 counts between 200 – 500 cells/μl were classified in the medium risk group and the low risk group was delineated by individuals with viral loads lesser than 10 000 copies/ml and CD4 greater than 500 cells/μl.

Four participants (Tz7, 10, 217 and 440) belonged to high risk group; six participants (Tz14, 48, 76, 203, 218 and 220) belonged to the medium risk group and finally only one individual (Tz118) belonged to the low risk group. Due to the lack of CD4 data for Tz126 the risk category can not be determined, however, the viral load of greater than 30 000 copies/ml were measured (Table 3.3.3).

Table 3.3.3: Biological markers (viral loads and CD4 counts) measured from the study group delineating the risk category of each individual.

Sample ID	CD4 Count (copies/ml)	Median Viral Load Count (copies/ml)	Risk Category[1]
Tz7	91	160 500	high
Tz10	143	101 900	high
Tz14*	211	81 900	medium
Tz48*	350	30 750	medium
Tz76*	400	12 100	medium
Tz118	670	6 605	low
Tz126	ND	73 100	ND
Tz203	152	35 200	high
Tz217	336	150 500	medium
Tz218*	497	58 050	medium
Tz220	468	64 650	medium
Tz440	115	293 500	high

* Denotes dually infected individuals ; ND = data unavailable ; CD4 count after 24 months ; riskk category as set up by Mellors *et al.*, 1997.

Previous studies (Grobler *et al.*, 2005 and Gottlieb *et al.*, 2004) have reported an associated between dual infection and rapid disease progression, however, that associated was not seen in our study cohort. Figure 3.3.23 depicts viral dynamics over the study period among both single and dually infected individuals and no obvious difference in the viral load trajectory is evidenced between the two groups.

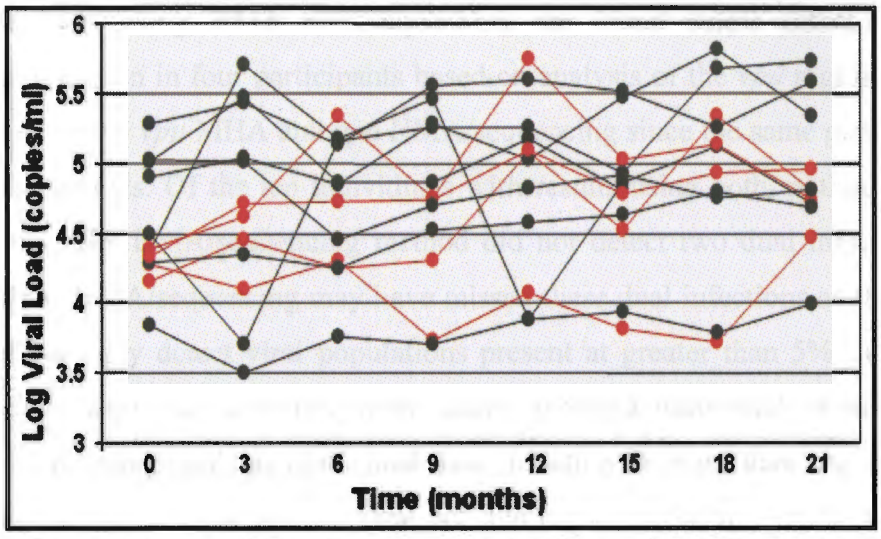


Figure 3.3.23: Viral load dynamics measured from the study cohort over 21 months of follow-up, depicting single (black) and dually (red) infected individuals.

3.4 Discussion

HIV-1 disease outcome is dependent on the interplay between viral and host factors. Viral subtype, viral phenotype, high viral diversity at infection and dual infection have all been shown to affect disease progression (Kaleebu *et al.*, 2002, Kupfer *et al.*, 1998, Wang *et al.*, 2000, Sager *et al.*, 2003, Gottlieb *et al.*, 2004 and Grobler *et al.*, 204). Longitudinal studies of cohorts that are at high risk of HIV infection in settings where multiple subtypes co-circulate enables identification of individuals infected with multiple HIV subtypes. Investigation of viral dynamics in samples from such individuals can provide insights into selective forces driving viral evolution, as well as the role of high diversity on disease progression. In this chapter we screened a selected subset of samples from a high-risk cohort in Tanzania provided through the HIV Superinfection Study (HISIS, www.mmrp.org). We identified dual infection in four out of the twelve individuals. Dual infections were detectable at a minority of follow-up visits with only one individual having detectable dual infection at most time points. This emphasises the importance of analyzing multiple timepoints and that cross-sectional studies will likely under-estimate the prevalence of dual infections. Viral dynamics studies illustrated large fluctuations of viral populations over time however, generally there were relatively homogeneous populations at each timepoint suggesting that, while there is very high potential for diversity, diversity was largely constrained.

To investigate the dynamics of viral diversity in single and dual infection, samples were provided from twelve individuals: six of whom had been identified as being singly infected and six of whom had been identified as dually infected based on the multi-region hybridization assay (MHA, samples provided by M. Hoelscher, University of Munich, Germany).

However in our analysis using HMA and sequencing we found single infection in eight participants and dual infection in four participants based on analysis of the *vpu* region. The direct comparison was based on the *vpu* MHA and *vpu* HMA/sequencing since the same genomic region was assayed by both methods. Of the ten individuals with results using both methods, only five had concordant results. The HMA/sequencing method did not detect two dual infections which were detected by MHA. HMA/sequencing may have missed these dual infections as this approach is less sensitive and can only detect viral populations present at greater than 5% in the sample material. However, this approach identified three dually infected individuals which were not detected by MHA. Sequencing analysis confirmed dual infection in more than one timepoint in these individual suggesting that these results were not due to contamination. This highlights the limitation of techniques such as the MHA, where detecting viral populations is based on utilization of subtype specific-probes. Since both methods are PCR based, their sensitivity will be impacted by primer design which will influence the amplification of amplicons.

In this study we showed that most single infections harboured homogeneous viral populations that exhibited limited viral diversity; these individuals also displayed mostly homoduplex HMA bands (indicative of limited diversity) compared to dually infected individuals. DNA distances among single infections ranged from 0.9% to 5.5%. Among dually infected individuals, while being infected with highly diverse viruses, in some instances the contribution to overall viral burden was characterized by low diversity at certain timepoints similar to single infection. The low viral diversity exhibited by some of these individuals show containment of viral diversity.

No evidence of superinfection was detected in this study as most of the dually infected individuals showed evidence for the presence of the two infecting viruses at first follow-up. Unlike other previous studies (Gottlieb *et al.,* 2004 and Sagar *et al.,* 2003) no difference in viral load trajectory was observed between dual and single infections. However, as this was a chronic cohort it is difficult to determine the impact of dual infection on disease progression. Viral dynamics over time was characterized by huge fluctuations, replacement of one variant by another or by a different subtype. The dominant viral variants contributed significantly to total viral load. This observed fluctuation in viral populations could be the result of selective pressure. To adequately elucidate the role of immune pressure as a driving force behind viral evolution whole gene and indeed entire genome analysis is the ideal way to monitor. This study has provided some insight into the dynamics of viral population diversity and elucidated the contribution of different viral variants toward total viral burden.

CHAPTER 4

ROLE OF RECOMBINATION AND POINT MUTATION IN VIRAL DIVERSITY AND CTL EPITOPE ESCAPE IN TWO FULL GENES OF A DUALLY INFECTED INDIVIDUAL

4.1 Introduction

HIV evolves mainly through two mechanisms: point mutations that are a result of the error prone nature of reverse transcription (Preston *et al.*, 1988); and recombination that occurs primarily through template switching during reverse transcription (Liu *et al.*, 2002). HIV is constantly under pressure to diversify in order to escape immune pressures exerted by the host. The cellular immune system's $CD8^+$ cytotoxic T lymphocyte (CTL) directly influences the evolution of circulating viruses within a population by selecting for immune escape mutants (Moore *et al.*, 2002) although the influence of founder effects on identification of the frequency of immune escape mutants suggests that the impact of escape on circulating viruses is less common than originally reported (Bhattacharya *et al.*, 2007). In addition, antibodies have also been shown to exert selective pressure on the virus (Wei *et al.*, 2003).

Previous studies have indicated the important role played by CTLs in both the control of HIV-1 viremia during the acute phase of infection, and in determining an HIV infected person's clinical outcome. The role of CTL responses in controlling viral replication (Koup *et al.*, 1994 and Novitsky *el al.*, 2007) have been shown to drive the evolution of escape variants capable of evading CTL recognition (Goulder *et al.*, 1996 and Kelleher *et al.*, 2001). In SIV models, CTL responses correlated with the control of viremia in both naïve and previously immunized animals (Barouch *et al.*, 1999). The importance of CTL responses is illustrated in a SIV-infected rhesus macaque model where depletion of CTLs was correlated with loss of viral replication control (Schmitz *et al.*, 1999). The pattern of epitope recognition by CTL in response to viral infection is strictly dependant on the class I HLA alleles expressed by the individual (Ferrari *et al.*, 2004). As a result HLA class I alleles strongly impact the progression of disease in HIV infected individuals (Carrington *et al.*, 1999). The two most prominent alleles (B*27 and B*57) have been found to induce immunodominant CTL responses to conserved HIV-1 epitopes (Tang *et al.*, 2002). A recent study by Geldmacher *et al.*, (2007) indicated that the possession of HLA class I alleles B*5801, B*8101 and B*0702 was associated with a low median viral load in a cohort of HIV seropositive bar workers.

One of the most significant factors obstructing the development of an effective HIV vaccine is the extreme genetic diversity of the virus (McCutchan *et al.*, 2000) that, in turn, is a consequence of its high mutation and recombination rates. Recombinant HIVs containing mixtures of genetic material from different HIV-1 subtypes are being discovered with increasing frequency, particularly in parts of the developing world where the pandemic is at its worst (Najera *et al.*, 2002).

Currently over thirty circulating recombinant forms (CRFs) and numerous unique recombinant forms (URFs) have been identified the world over. Unlike the slow accumulation of mutations that occur through replication errors (Malim *et al.*, 2001), recombination enables very rapid increases in viral sequence diversity. The emergence of recombinant HIV genotypes with novel properties is therefore particularly worrisome in the context of both drug and vaccine development.

Inter-subtype recombination must occur within individuals simultaneously infected with two different subtypes. The diverse viral populations harbored by these, so-called dually infected individuals, potentially have a much greater opportunity to explore novel sequence variation than those found in individuals infected with single viral variants. Previous studies on a limited number of dually infected individuals have suggested that recombinant viruses may have a selective fitness advantage in dual infections (Liu *et al.*, 2002), presumably through recombination drawing from a richer pool of variation in dually infected individuals.

In this chapter we focused our attention on the *gag* and *nef* genes from viruses longitudinally sampled from a dually infected individual from Tanzania. These proteins are the most frequently targeted with over 90% of individuals targeting one or both proteins (Ferrai *et al.*, 2001 and Addo *et al.*, 2002) the role of CTL responses to these genes is debated in the literature. *Gag* is particularly an interesting gene because CTL responses against its expressed protein is strongly correlated with control of viral replication (Edwards *et al.*, 2002 and Geldemacher *et al.*, 2007), other studies however, have shown contradictory results (Masemola *et al.*, 2003 and Walker *et al.*, 2002). Nef CTL responses have also been reported to control and lower viral loads (Mashishi *et al.*, 2004) but other studies have found an inverse relation between Nef CTL responses and viral load (Novisky et al., 2003). Anti-Gag and Nef responses are also quite broad in that they are the most consistently reactive against viruses from subtypes other than that of the virus eliciting the response (cross clade response: Ferrari *et al.*, 2004).

The role of recombination and point mutations on viral diversity was studied by monitoring known CTL epitopes over time in the selected Tanzanian individual. Besides being responsible for the vast array of obviously recombinant HIV-1 genotypes detectable in Tanzania (Hoelscher *et al.*, 2001), dual infections and subtle, less obvious recombination events may also be responsible, at least in part, for generating the enormous degrees of HIV diversity observed in that country. The aim of this chapter was to decipher the role of point mutations and recombination on viral diversity and immune escape in a dually infected individual.

4.2.1 cDNA Synthesis

Extracted viral RNA was used as template for the synthesis of complementary DNA (cDNA) using the Invitrogen Thermoscript™ RT-PCR System (Invitrogen, GmbH, Karlsruhe, Germany). Master mix number 1 containing 20pMols of gene specific primers (gagDR 5'-AATTCCTCCTATCATTTTTGG-3' or nef-OR 5'-AGGCAAGCTTTATTGAGG- 3', Table 4.2.1) was incubated at 65°C for 5 minutes to eliminate any RNA secondary structure before adding it to Master mix 2 (table 4.2.2) and incubating for one hour at 50°C.

The different master mixtures were prepared as follows:

Table 4.2.1 Master Mix 1:

Reagent	Amount
Primer (20pmol)	1µl
10 mM dNTP Mix	1µl
RNA (10 pg -5 µg)	4µl
Total Volume	**6µl**

Table 4.2.2 Master Mix 2:

Reagents	Amount
5X cDNA Buffer	2µl
DTT (10mM)	0.5µl
RNaseOUT™ (40U)	0.5µl
Thermoscript RT enzyme (15U)	0.5µl
Water	5µl
Total	**4µl**

4.2.2 Amplification of full-length genes by PCR

All reagents, primers and conditions for the amplification of the whole *gag* and *nef* genes are listed in tables 4.2.3, 4.2.4 and 3.2.4. In a first round of PCR, *gag* was amplified as a 1.78Kb fragment. A second round of nested PCR amplified a 1.68Kb fragment. For amplification of *nef* the first round PCR yielded a fragment of 700bp whereas a second round yielded a 650bp fragment.

Table 4.2.3: Primers for both *gag* and *nef* PCRs

Region	Primer	Position (HXB2)
First round gag *(Forward Primer)*	**gag-DF (5'- CTCTAGCAGTGGCGCCCG - 3')**	621 - 644
(Reverse Primer)	**gag-DR (5'-AATTCCTCCTATCATTTTTGG- 3')**	2383 - 2403
Second round gag *(Forward Primer)*	**gag-AF (5'-CTCTCGACGCAGGACTCGGCTT- 3')**	678 - 704
(Reverse Primer)	**gag-CR (5'-TCTTCTAATACTGTATCATCTGC-3')**	2334 - 2361
First round *nef (Forward Primer)*	**SQ15FC (5'- GAGAGCGGTGGAACTTCTGG- 3')**	8561 - 8580
(Reverse Primer)	**Nef OR (5'-AGGCAAGCTTTATTGAGG - 3')**	9243 - 9261
Second round *nef (Forward Primer)*	**nef Fwd (5'-CCTAGA1AGAATAAGACAGGGCTT - 3')**	8754 - 8776
(Reverse Primer)	**nef Rvs (5'-CCTGGAACGCCCCAGTGG -3')**	9386 - 9401

Table 4.2.4: PCR reagents and volumes for both *gag* and *nef* PCRs

PCR Reagents	*gag* First round	*gag* Second round	*nef* First round	*nef* Second round
5'- (Forward Primer) (10pmol/µl)	1µl	1µl	1µl	1µl
3'- (Reverse Primer) (10pmol/µl)	1µl	1µl	1µl	1µl
dNTP Mix (Roche, USA) (2.5mM each)	1µ l	1µ l	1µ l	1µ l
10x buffer	5µl	5µl	5µl	5µl
1.5mM $MgCl_2$	6µl	6µl	6µl	6µl
deionized water	30.2µl	33.2µl	30.2µl	33.2µl
Expand polymerase (Bertec, Taiwan)	0.8µl	0.8µl	0.8µl	0.8µl
Template	5µl	2µl	5µl	2µl
Total	**50µl**	**50µl**	**50µl**	**50µl**

Table 4.2.5: PCR cycling conditions for *gag* and *nef* PCRs

Cycling Conditions	*Gag/nef* First round	Number of cycles	*Gag/nef* Second round	Number of cycles
(denaturing)	2 minutes (94°C)	1	2 minutes (95°C)	1
(denaturing)	15 second (94°C)		15 second (94°C)	
(annealing step)	30 seconds (48°C)	x 10	30 seconds (55°C)	x 30
(extension step)	1 minutes (72°C)		1 minutes (72°C)	
(denaturing)	15 second (94°C)		5 minutes (72°C)	1
(annealing step)	30 seconds (50°C)	x 15	END (4°C)	∞
(extension step)	1.5 minutes (72°C)			
extension step)	5 minutes (72°C)	1		
END	END (4°C)	∞		

4.2.3 PCR Clean up

PCR products were purified by the QIAquick Spin Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Briefly PB buffer (5x volume of PCR product) was mixed with the PCR product and loaded on the spin column and centrifuged for one minute at 13000rpm. Bound PCR products were washed with PE buffer and eluted by centrifugation at 13000rpm using buffer EB (50µl).

4.2.4 Cloning and sequencing

Samples were cloned into the pGEM®-T Easy vector (Promega, Madison, USA) according to the manufacturer's instructions. Following transformation, colonies containing amplified HIV genes were identified by colony PCR (Lee and Cooper, 1995). Twenty colonies containing *nef* and twenty colonies containing *gag* genes were selected along with bulk, untransformed *gag* and *nef* PCR products from each time point for sequencing using ABI PRISM dye terminator cycle-sequencing kits (Applied Biosystems; Tables 4.2.6 and 4.2.7) and a 3100 ABI 16 capillary automated sequencer at the University of Stellenbosch. The full set of *gag* sequencing primers are presented in Appendix G, *nef* sequencing was carried out using the PCR primers presented in Table 4.2.3.

Table 4.2.6: Sequencing PCR reagents and volumes

PCR Reagents	Sequence reaction
(Forward Primer) or (Reverse Primer) (3.2 pmol/µl)	1µl
5x buffer	2µl
deionized water	11µl
Big Dye® enzyme (AB, USA)	4µl
Template	2µl
Total volume	**20µl**

Table 4.2.7: Sequencing PCR cycling conditions

Cycling	Conditions	Cycles
(denaturing)	10 seconds (96°C)	1
(denaturing)	30 seconds (96°C)	X 25
(annealing step)	15 seconds (50°C)	
(extension step)	4 minutes (60°C)	
END	END (4°C)	∞

4.2.5 Phylogenetic analysis

Full-length gene (*gag* and *nef*) sequences were aligned using Clustal X (default settings) (Thompson *et al.,* 1994) and alignment columns with gaps were removed. Neighbour joining phylogenetic trees with 100 bootstrap iterations were constructed with MEGA version 2.1 using kimura 2 parameter distances and a transition:transversion ration of 2 (Kumar *et al.,* 2001).

4.2.6 Recombination analysis

Recombination was detected and analysed using the RIP method implemented at http://hivweb.lanl.gov/RIP/RIPsubmit.html and the RDP (Martin and Rybicki, 2000) GENECONV (Padidam *et al.*, 2001), BOOTSCAN (Martin *et al.*, 2005a) MAXIMUM CHI SQUARE (Maynard and Smith, 1992), CHIMAERA (Posada and Cradall, 2001) and SISCAN (Gibbs *et al.*, 2001) methods implemented in RDP3 (Martin *et al.*, 2005b) with default settings.

4.2.7 Positive Selection

Dr Konrad Scheffler of the South African National Bioinformatics (SANBI) node at U.C.T used a new method for the detection of positive selection in recombinant viral sequences (Scheffer *et al.*, 2006). Other phylogenetic methods commonly applied to detect positive selection have been shown to give misleading results when applied to recombining sequences. Briefly, this method involves partitioning of alignments using detectable recombination breakpoints as a guide to split alignments into sub-alignments each of which is then assumed to include no further evidence of obvious recombination. Also included in this method is a codon substitution model that incorporates synonymous rate variation.

Firstly the positions of recombination breakpoints were estimated using the non parametric RDP (Martin and Rybicki, 2000), GENECONV (Padidam *et al.*, 1999) and MAXIMUM CHI SQUARED (Maynard and Smith, 1992) methods implemented in RDP3RDP3 (Martin et al., 2005). Once the recombination breakpoints were detected, they were used to partition the alignment into separate segments. Finally, topologies and branch lengths were estimated as in the baseline method, except that a separate topology and set of branch lengths were used for each segment during detection of selection using the *HyPhy* method (http://www.hyphy.org).

4.2.8 HLA Genotyping

HLA genotyping was carried out in the laboratory of Dr Michael Hoelscher at Munich University, Germany. DNA was isolated from peripheral blood mononucleocytes (PBMC) and patient genotypes determined at low resolution using sequence specific primer polymerase chain reaction (SSP-PCR) (Pel Freeze, USA).

4.2.9 Potential CTL epitopes

HLA Binding Motif Scanner (http://www.hiv.lanl.gov.contant/immunology/motif_scan/motif_scan) was used to identify putative CTL epitopes in sequenced and *in silico* translated *gag* and *nef* genes. This program identifies HLA anchor residue motifs within input amino acid sequences for specified HLA types and yields potential epitope sequences as output.

4.3 Results

To determine the role of point mutations and recombination on viral diversity in a dually infected individual, we longitudinally sampled and sequenced full *gag* and *nef* genes and performed phylogenetic, recombination, selection and CTL epitope prediction analyses. Tz14 was identified for in depth analysis because two different subtypes (A and C) were detected in the *vpu* region of viruses infecting this individual, more interestingly, a possible triple infection was detected but not proven in the *env* C2C3 wherein two separate clusters of subtype C sequences were found in addition to a subtype A cluster (Chapter 3). In addition, HLA data was available on this individual enabling us to investigate putative CTL escape.

4.3.1 Full gene subtype determination

Full-length *nef* sequences were amplified from three time points [F0 (0 months), F4 (12 months) and F7 (21 months)] and *gag* sequences were amplified from two time points [F0 (0 months) and F7 (21 months)]. Twenty randomly selected clones were sequenced from each time point for each gene. Phylogenetic based HIV-1 subtype determination was carried out on the 20 full-length gene sequences from each time point of both genes. Appendices F and G depict all the *nef* and *gag* full length sequences respectively. Whereas at F0 all *nef* sequences were pure subtype C, at F4 both subtypes A and C sequences were detected and at F7 only subtype A sequences were detected (Figure 4.3.1).

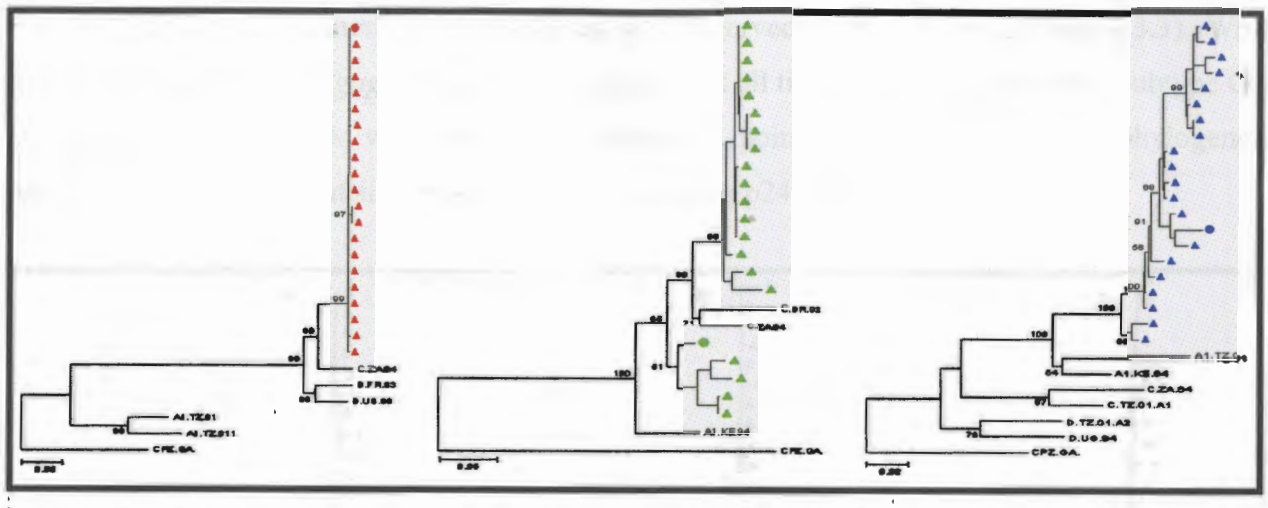


Figure 4.3.1: Tz14 *nef* eighbour joining trees from: F0 (time point 0, red); F4 (four months post F0, green) and F7 (21 months post F0, blue). Scale bar indicates nucleotide divergence of 5% calculated from hundred repeats and only bootstrap values over 75% are shown. Circles depict the bulk PCR sequences and the triangles represent the clone sequences.

This data clearly depicts a progressive displacement of subtype C *nef* sequences with subtype A *nef* sequences over the 21 month sampling period. It also confirms the previous assessment based on *vpu* and *env* C2C3 sequences (Chapter 3). Phylogenetic analysis of the *gag* sequences (Appendix G) from F0 and F7 also indicated that the *gag* genes of viruses infecting Tz14 were mostly subtype A-C recombinants. At F0, full-length *gag* and *gag* p17 sequences cluster with subtype C reference sequences but *gag* p24 sequences cluster with subtype A reference sequences (Fig 4.3.2).

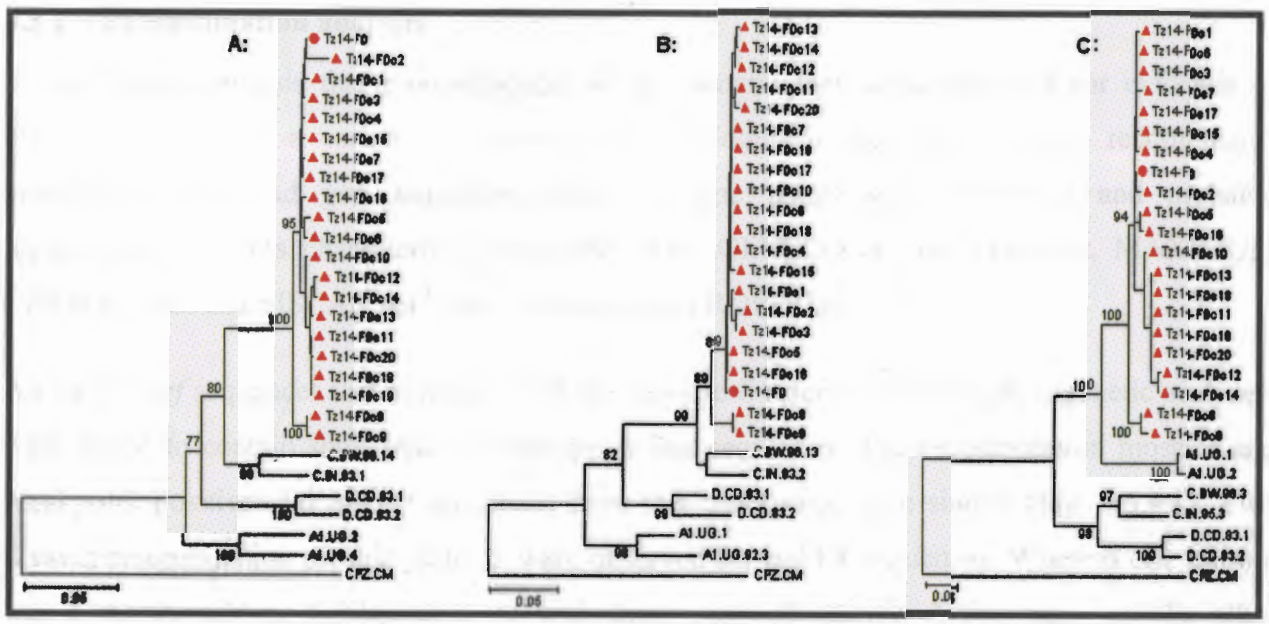


Figure 4.3.2: Tz14 *gag* neighbour joining trees at time point F0, full gene sequences (A), *gag* p17 sequences (B) and *gag* p24 sequences (C). Scale bar indicates nucleotide divergence of 5% calculated from hundred repeats and only bootstrap values over 75% are shown. Circles depict the bulk PCR sequences and the triangles represent the clone sequences. The circles depict the bulk PCR sequence and the triangles represent the clone sequences.

Similar evidence of chimeric *gag* sequences was observed in F7 sequences (Fig 4.3.3). While evidence of recombinant *gag* sequences was obtained at all time points no completely subtype C or subtype A *gag* sequences were observed at any time point. This is based on the phylogenetic analysis showing different subtypes for *gag* p17 and *gag* p24 regions.

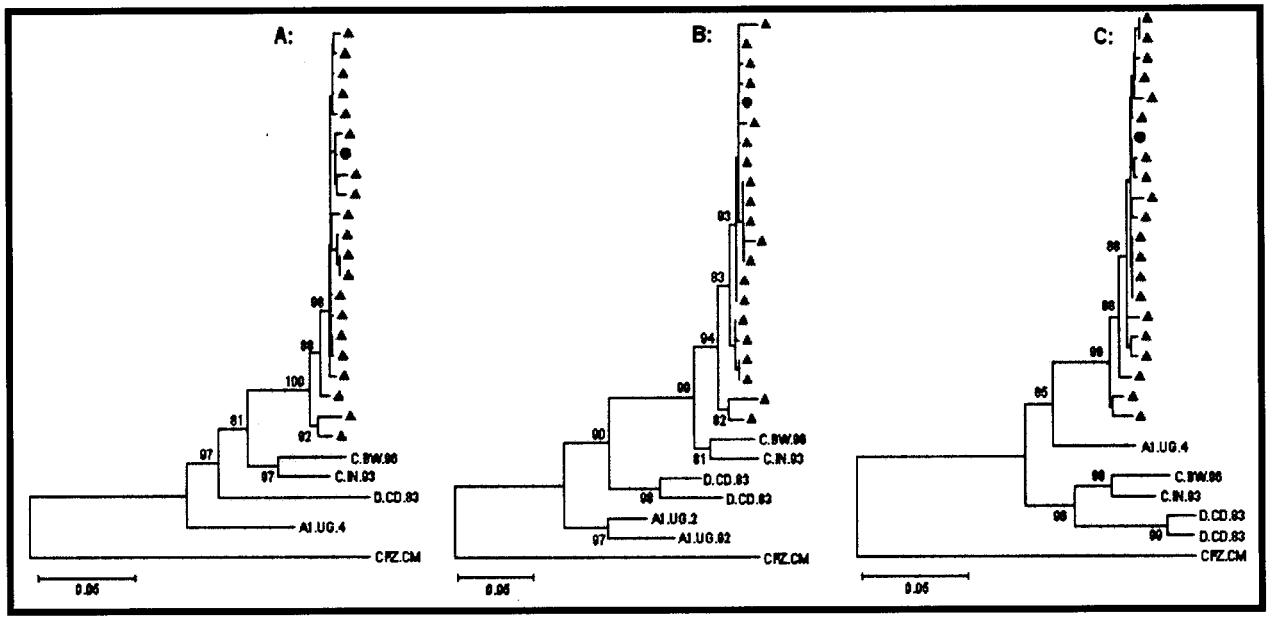


Figure 4.3.3: Tz14 *gag* neighbour joining trees at time point F7, full gene sequences (A), *gag* p17 sequences (B) and *gag* p24 sequences (C). Scale bar indicates nucleotide divergence of 5% calculated from hundred repeats and only bootstrap values over 75% are shown. Circles depict the bulk PCR sequences and the triangles represent the clone sequences. The circles depict the bulk PCR sequence and the triangles represent the clone sequences.

4.3.2 Recombination analysis

To test the hypothesis that *gag* sequences were chimeric they were analyzed for evidence of recombination using a variety of statistical and phylogenetic methods. Potential recombinants amongst all *nef* and *gag* sequences from all time points were identified and probable recombination events characterized using the RDP, GENECONV, BOOTSCAN, MAXIMUM CHI SQUARE and SISCAN methods implemented in RDP3RD3.

All the F0 *nef* sequences that clustered with the subtype C references in the phylogenetic analyses, were found to contain short tracts of subtype A-like sequences. The recombination mosaics and breakpoint positions for all *nef* sequences from this time point were similar (Fig. 4.3.4A). Two distinct recombination mosaic patterns were observed for *nef* F4 sequences. Whereas one mosaic was a mostly subtype C-like sequence with short tracts of subtype A-like sequence, the other mosaic was mostly subtype A-like with short tracts of subtype C-like sequence. Interestingly, breakpoint positions in both mosaics were similar and they therefore had a similar, albeit inverted, appearance (Fig 4.3.4B). F7 *nef* sequences were predominantly subtype A-like with small tracts of

subtype C-like sequences. The recombination mosaics and breakpoints of all of the F7 *nef* sequences were similar to one another (Fig. 4.3.4C).

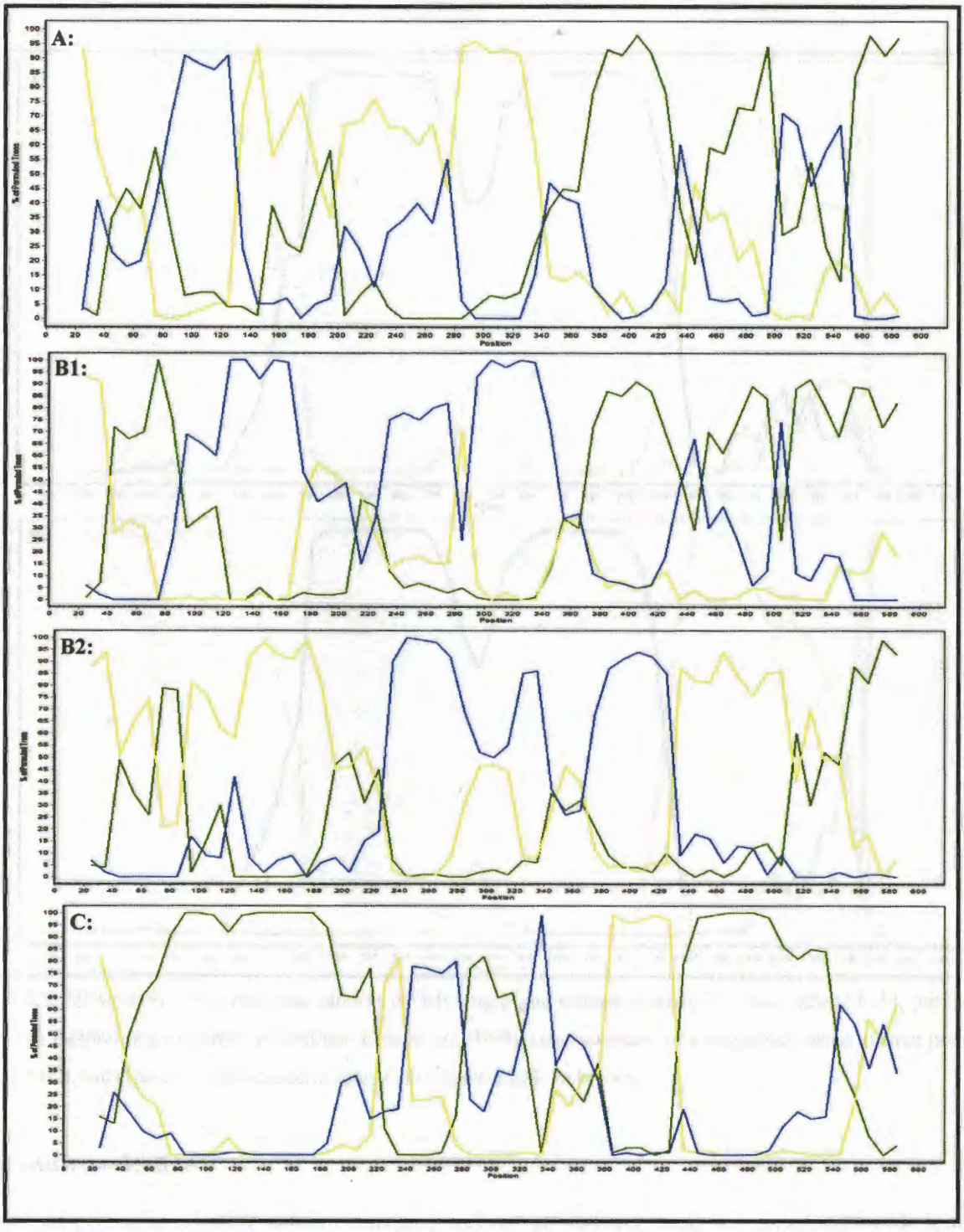


Figure 4.3.4: Illustration of recombinant patterns of full-length *nef* sequences sampled from patient Tz14, performed by bootscan method implemented in SimPlot (Lole *et al.*, 1999), representation of a sequences taken at time point F0 (A) C-like sequences, F4 (B1) A-like and (B2) C-like sequences finally F7 (C) A-like sequences. Subtype A is represented in green, C in yellow and D in blue.

Tz14 *gag* sequences indicated that the first 400 nucleotides of all *gag* sequences from all time points had subtype C origins whereas the remainder of the gene had subtype-A origin (Figure 4.3.5).

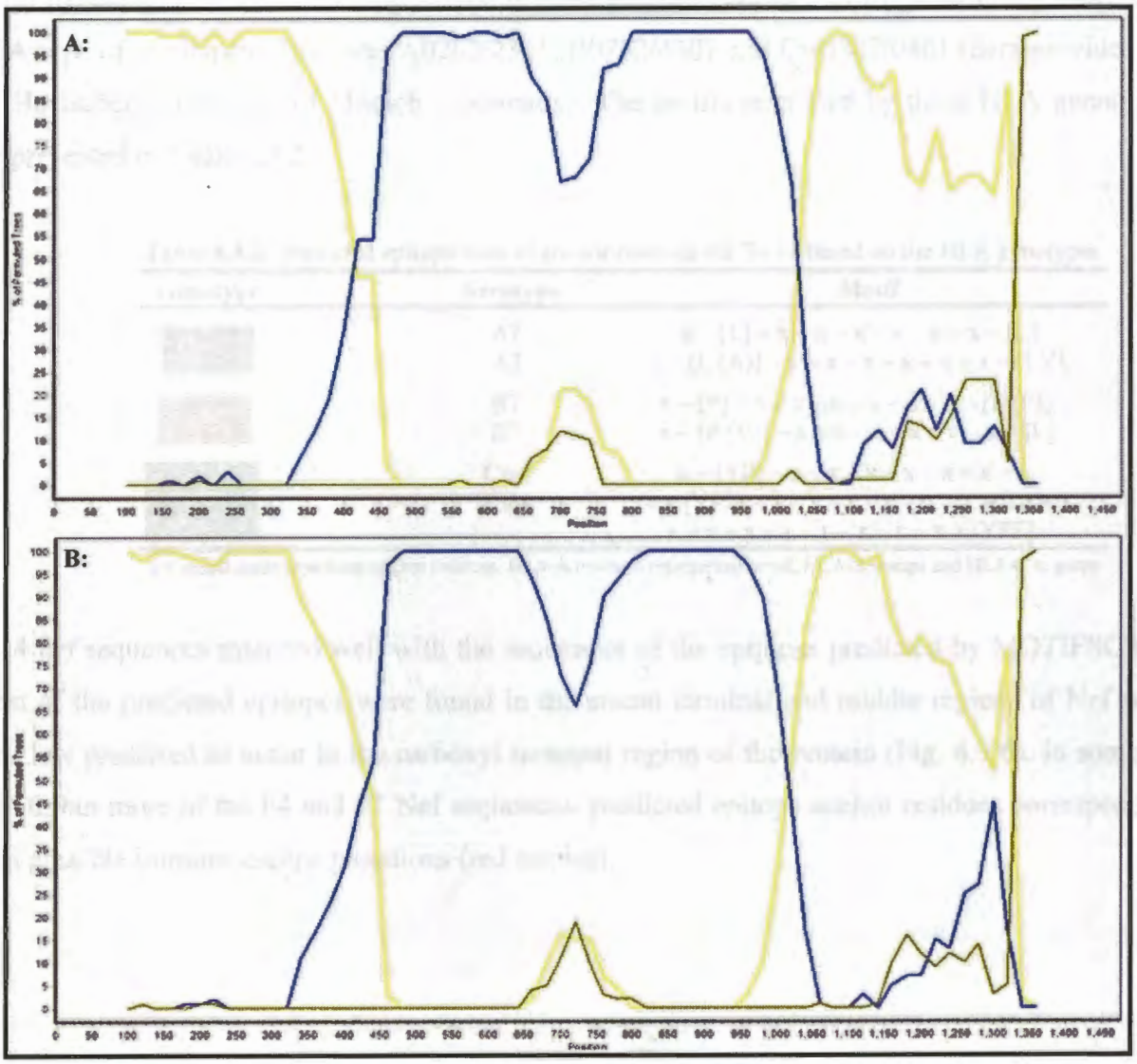


Figure 4.3.5: Illustration of recombinant patterns of full-length *gag* sequences sampled from patient Tz14, performed by bootscan method implemented in SimPlot (Lole *et al.*, 1999), representation of a sequences taken at time point F0 (A) and F7 (C). Subtype A is represented in blue, C in yellow and D in brown.

4.3.3 Positive selection

Evidence of potential positive selection in *gag* and *nef* genes over time was assessed by Dr Konrad Scheffler of the South African National Bioinformatics Network node at U.C.T using custom developed software for the analysis of selection in recombinant sequences. There was no signal of positive selection in either the *nef* or *gag* gene sequences during the 21 months of follow-up.

4.3.4 Potential CTL epitopes

In the absence of any evidence for positive selection acting on the *gag* and *nef* sequences sampled from patient Tz14, potential *gag* and *nef* epitopes restricted by this individuals's HLA were identified using MOTIFSCAN (http://www.hiv.lanl.gov/immunology/motif_scan/motif_scan). The HLA-type of participant Tz14 was A0202/2301; B0702/4501 and Cw0702/0401 (data provided by M. Hoelscher, University of Munich - Germany). The motifs restricted by these HLA genotypes are presented in Table 4.3.2.

Table 4.3.2: Predicted epitope scan of anchor residues for Tz 14 based on the HLA genotypes

Genotype	Serotype	Motif
	A2	x – [L] – x – x – x – x – x – x – [L]
	A2	x – [L (A)] – x – x – x – x – x – x – [LV]
	B7	x – [P] – x – x – x – x – x – x – [L (F)]
	B7	x – [P (V)] – x – x – x – x – x – x – [L]
	Cw4	x – [YP] – x – x – x – x – x – x – x
	Cw4	x – [YPF] – x – x – x – x – x – x – [LFM]
		x – x – x – x – x – x – x – x – [YPF]

x = amino acids separating anchor residues, HLA-A epitopes represented in red, HLA-B orange and HLA-C in green

Tz14 *nef* sequences matched well with the sequences of the epitopes predicted by MOTIFSCAN. Most of the predicted epitopes were found in the amino terminal and middle regions of Nef with very few predicted to occur in the carboxyl terminal region of the protein (Fig. 4.3.6). In some of the F0, but more of the F4 and F7 Nef sequences, predicted epitope anchor residues corresponded with possible immune escape mutations (red arrows).

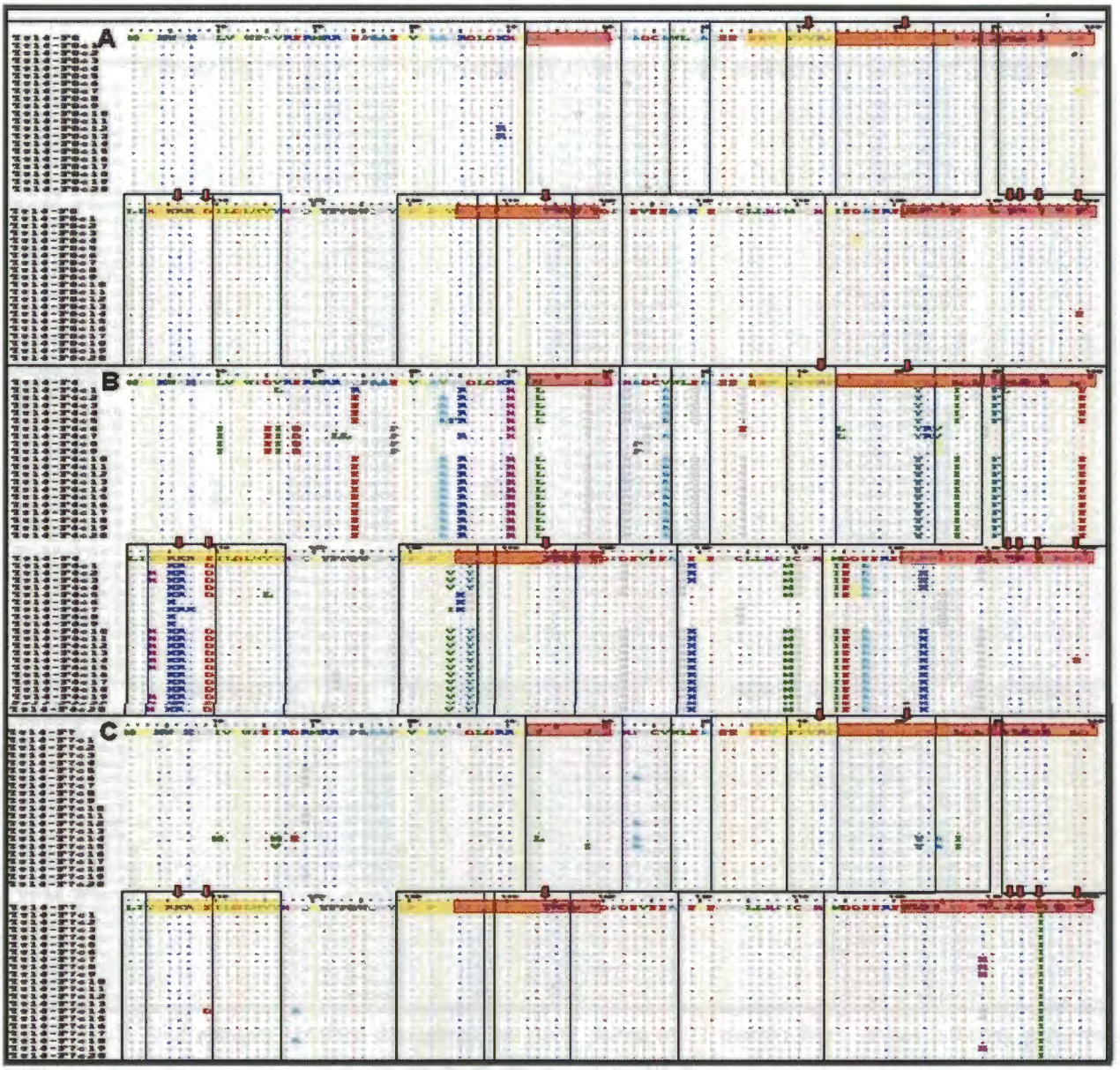


Figure 4.3.6: Predicted epitopes based on anchor residues are highlighted according to HLA genotype for Tz14 Nef protein at three time-points. Boxed regions highlight previously defined epitopes that are restricted by HLA-A and B alleles (htt://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan). Red arrows indicate possible immune escape anchor residues in the *nef* amino acid sequence (A; F0, B; F4 and C; F7) for HLA A*0202 (red) , B*0702 (orange) and Cw*0702 (green).

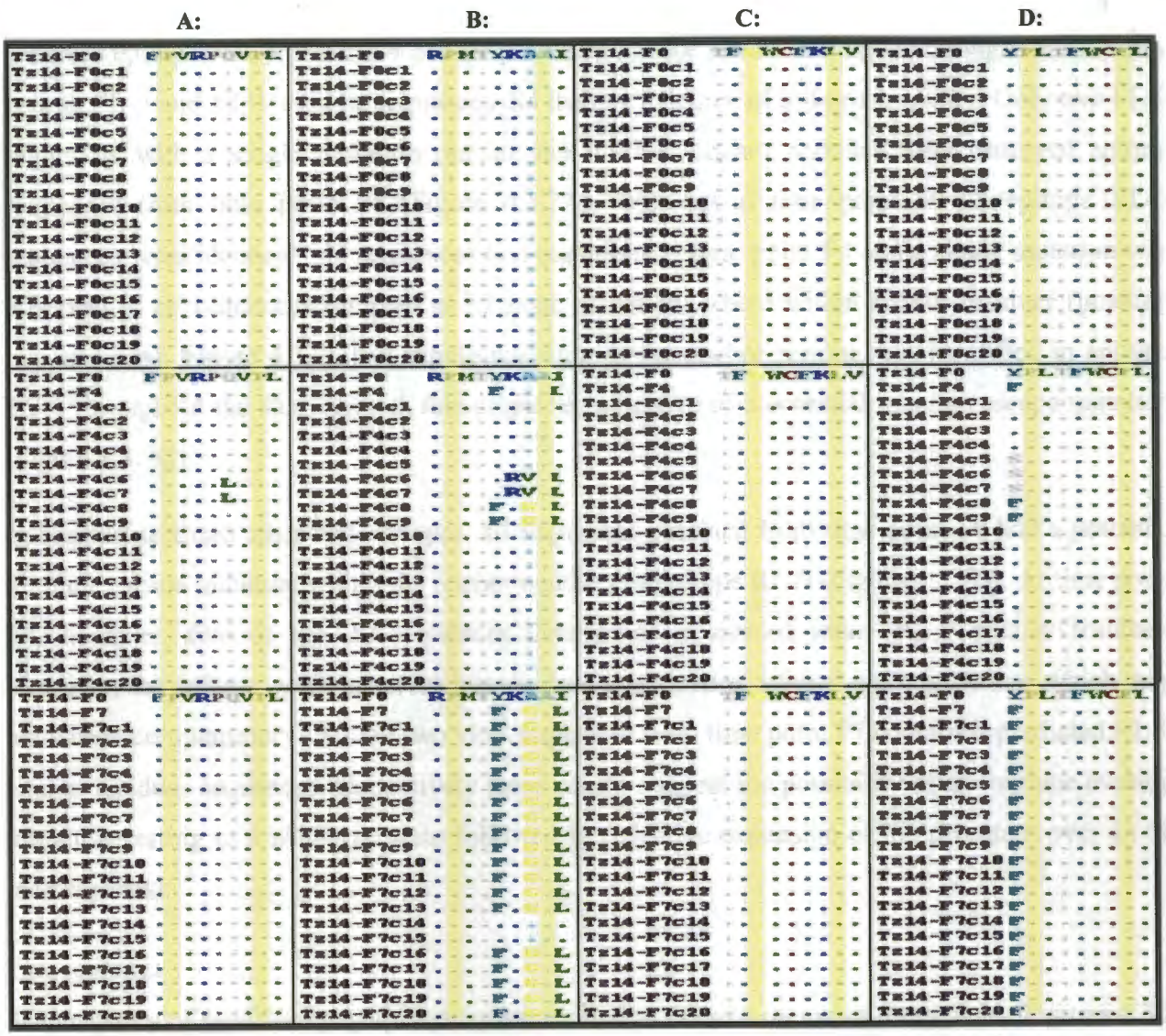


Figure 4.3.7: Nef epitopes obtained throughout the study period of 21 months from viruses infecting participant Tz14, panel A; shows epitope FL-9, B; RL-9, C; TV-9 and D; YL-9. Anchor residues at position two and nine highlighted in yellow.

Figure 4.3.7 depicts four common *nef* epitopes detected at three time points in viruses from patient Tz14. Variable and conserved predicted Gag and Nef epitope binding motifs were examined in greater detail to both identify portions of Gag and Nef that might be important for inducing broadly cross-reactive immune responses and provide some insight into the role of anchor residue mutation in the evolution of HIV within an individual patient.

Nef epitopes

Whereas epitopes FL-9 and YL-9 had no predicted anchor residue mutations throughout the study period, FL-9 and YL-9 epitope sequences did display a degree of minor variability. Only two FL-9 sequences with a single mutation outside the inferred anchor residues were observed among sequences from time point F4 (Figure 4.3.7A). Whereas at time point F0 all epitope YL-9 sequences were identical, six sequences had mutations at time point F4 and a single mutation was observed in all sequences sampled at F7 none of which where within anchor residues (position two and nine; Figure 4.3.7D). A single mutation in the anchor residue was identified in epitope TV-9 throughout the study period, this might be indicative of a potential immune escape mutation (Figure 4.3.7C).

In contrast to these other Nef epitopes, all sequences obtained from time point F0 had a potential immune escape mutation within the anchor residues of epitope RL-9 (Figure 4.3.7B). At time point F4, however, five of the 20 sequences determined possessed intact RL-9 anchor residues. However, the epitopes with intact anchor residues contained additional point mutations which may have been compensatory. All but two viral sequences from time point F7 contained predicted RL-9 anchor residues sequences. Collectively these results suggest the possibility of an immune evasion mutant reverting to a wild-type state followed by selective expansion of the reventant over an 18 month period.

Gag epitopes

The majority of predicted *gag* epitopes were located in the p24 region of the protein (Figures 4.3.8 and 4.3.9). All Gag epitopes predicted from sequences obtained from both time point F0 and F7 showed good overlap between predicted and experimentally defined subtype C epitopes (Masemola *et al.*, 2004). Few point mutations were observed within *gag* sequences throughout the study and there were therefore very few predicted epitope sequences that contained potential anchor residue immune escape mutations (red arrows).

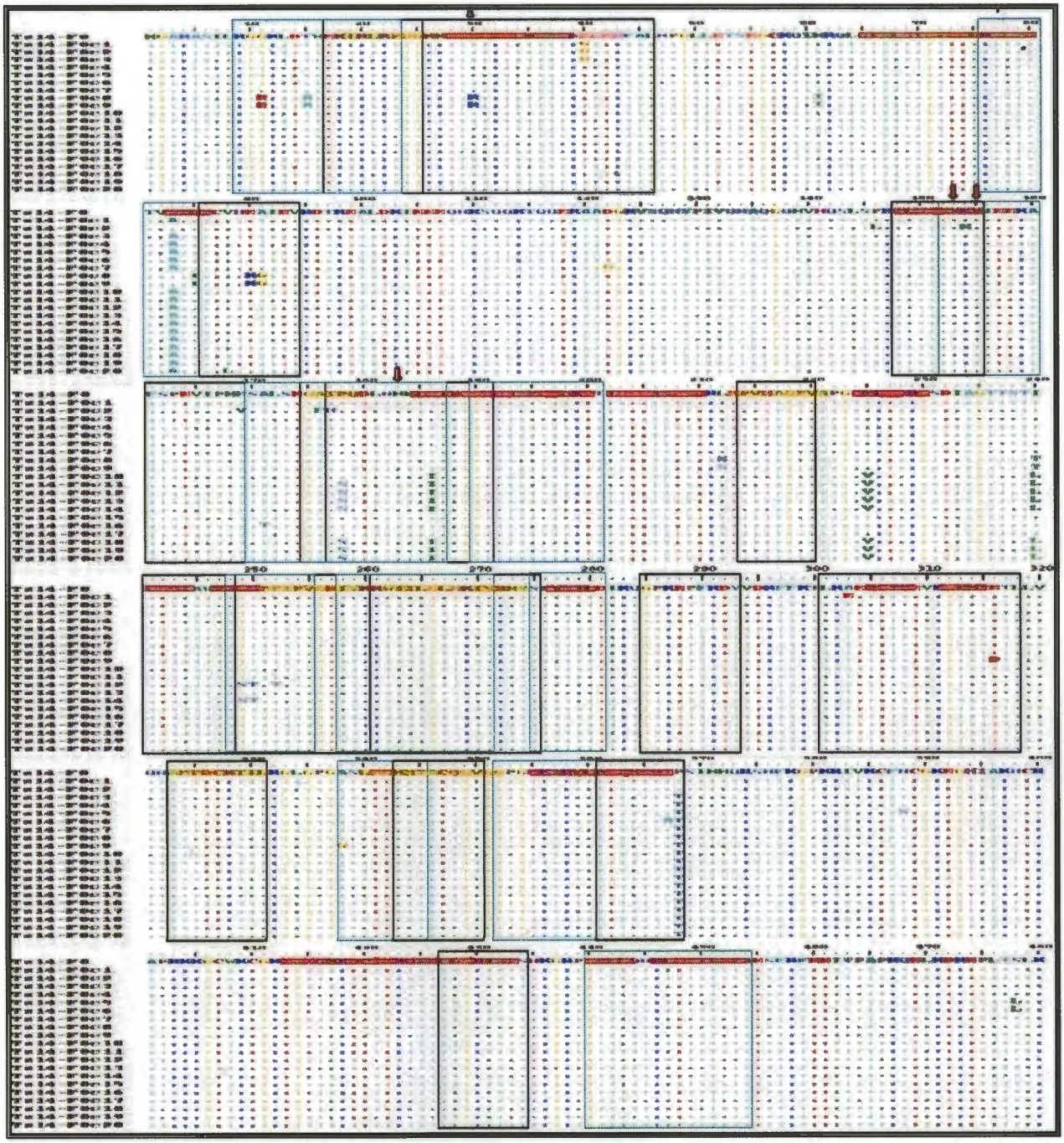


Figure 4.3.8 Predicted epitopes based on anchor residues are highlighted according to HLA genotype for Tz14 Gag protein from first time F0. Boxed regions highlight previously defined epitopes that are restricted by HLA-A and B alleles (htt://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan). Red arrows indicate the regions of possible escape in the *gag* amino acid sequence for HLA-A*0202 (red), B*0702 (orange) and Cw*0702 (green).

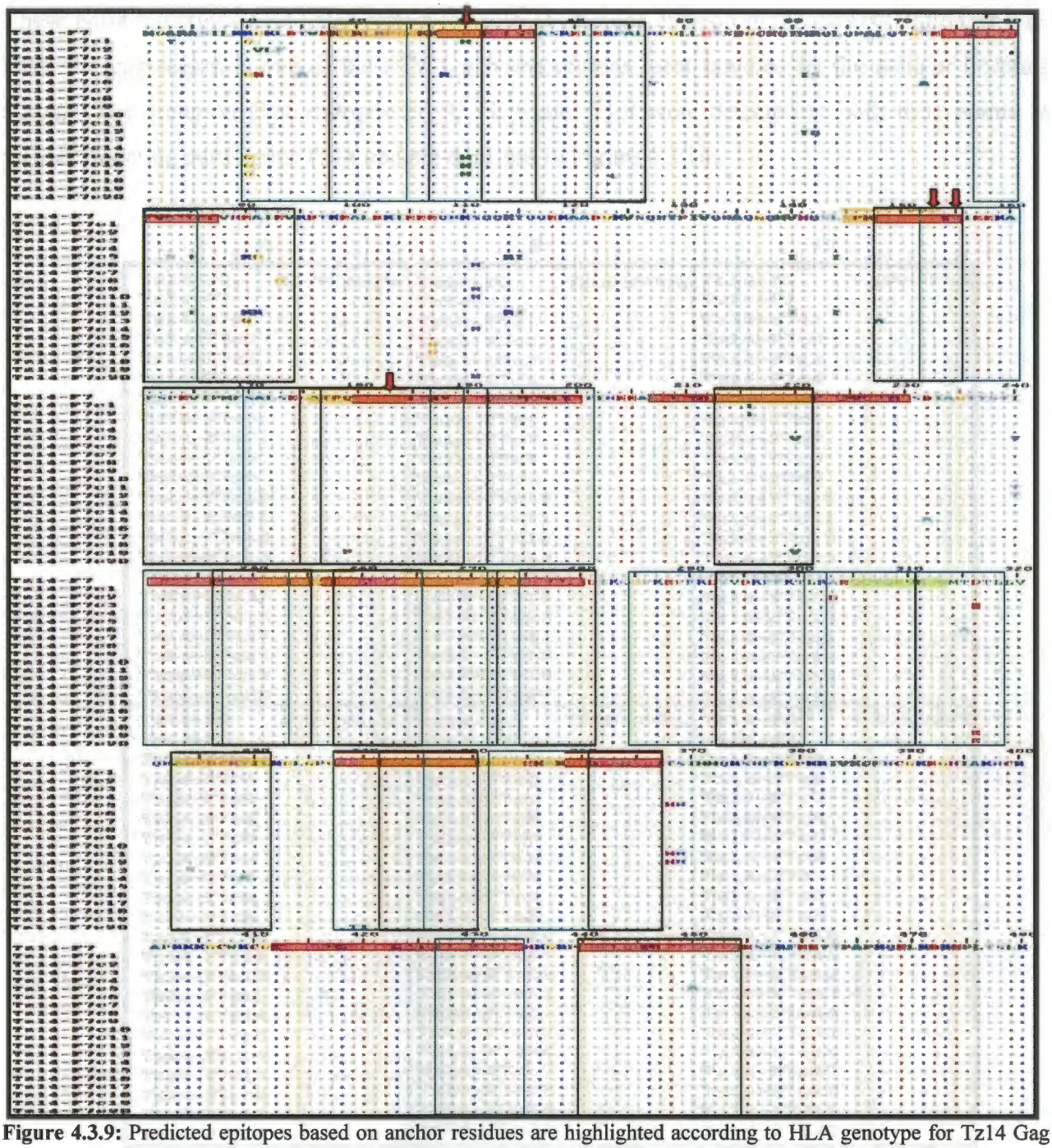


Figure 4.3.9: Predicted epitopes based on anchor residues are highlighted according to HLA genotype for Tz14 Gag protein from time point F7. Boxed regions highlight previously defined epitopes that are restricted by HLA-A and B alleles (htt://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan). Red arrows indicate the regions of possible escape in the *gag* amino acid sequence for HLA-A*0202 (red), B*0702 (orange) and Cw*0702 (green).

Anchor residues of potential Gag epitopes restricted by HLA-B0702 known to be associated with low viral loads in HIV-1 infected individuals (Geldmacker *et al.*, 2007) were screened for possible immune escape mutations at the beginning (F0) and end (F7) of the study period 18 months apart.

These point mutations could also reflect the role of immune evasion in the diversification of *gag*, since the *gag* region is one of the HIV-1 genome regions most targeted by the cellular immune system. The common Gag epitopes SL-9, TL 9 and TI-9 were detected at both time points in viruses infecting participant Tz14 and are depicted in Figure 4.3.10.

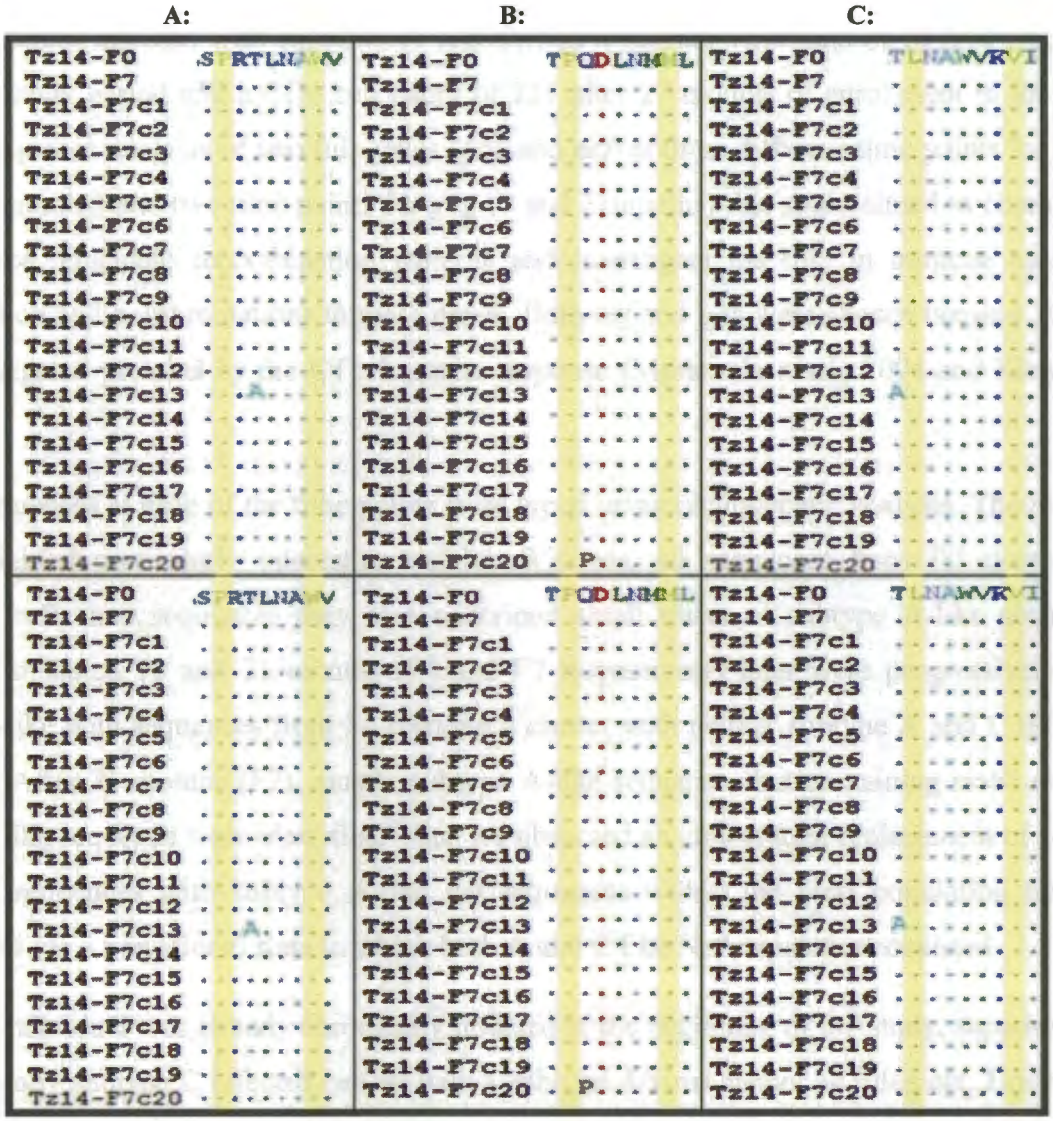


Figure 4.3.109: *Gag* epitopes obtained throughout the study period of 21 months from viruses infecting participant Tz14, panel A; shows epitope SL-9, B; TL-9 and C; TI-9. Anchor residues at position two and nine highlighted in yellow.

Gag epitope SL-9 (Figure 4.3.10A) contained potential anchor residue escape mutations in all *gag* sequences sampled at both time point F0 and F7, with respectively 20 of 19 of Gag epitope sequences containing one potential anchor residue escape mutation. One of the F0 SL-9 epitope sequences had mutations in both anchor residues. One potential anchor residue escape mutation was detected in seven of the twenty epitope TL-9 sequences at F0 however, no mutations were

detected in this epitope at F7 (Figure 4.3.10B). Finally, Gag epitope TI-9 sequences had one potential anchor residue escape mutation in all sequences from both F0 and F7 (Figure 4.3.10C).

4.4 Discussion

Tz14 was dually infected with subtypes A and C, had a median viral load of 81 900 (log 4.91) during the study period and a CD4 cell count of 211 after 27 months of enrollment in the study. Through sequence analysis of two full genes (*gag* and *nef*) at three different time points for *nef* (0, 12 and 21 months) and two time points for *gag* (0 and 21months) this study aimed to characterize viral subtype, elucidate recombination patterns and investigate the role in immune escape of recombination and point mutations in these genes. Both *nef* and *gag* were chosen because they are the most targeted proteins by the CTL immune response (Masemola *et al.*, 2004 and Kiepiela *et al.*, 2007).

Twenty sequences at each of the time points were typed using phylogenetic analysis. The *nef* gene analysis yielded particularly interesting results. Whereas, all sequences from F0 cluster with subtype C reference sequences they also contained small tracts of subtype A-like sequences. Sequences obtained 12 and 21 months (F4 and F7 respectively) later were progressively more subtype A-like with sequences from F4 forming a cluster with neither subtype A and C reference sequences. After 21 months (F7), mostly subtype A-like sequences but containing small tracts of subtype C-like sequence were identified. Thus we observed an almost total replacement of subtype C-like *nef* sequences with subtype A-like *nef* sequences within the viral population infecting patient Tz14 via a transitional state in which both A and C like Nef sequences coexisted.

Since this individual was already chronically infected at the beginning of the study, superinfection of an originally subtype C infected patient with a subtype A virus cannot be ruled out. However, it is unlikely to be superinfected as it is apparent from the recombination analysis that both subtype A and C-like virus sequences were already circulating within Tz14 at F0, the first sampling time point. In fact, none of the *nef* or *gag* sequences sampled from this paticipant were 'pure' non-recombinants with all of them displaying some degree of subtype A/C mosaic. In contrast to the shift from subtype-C to subtype-A like sequences in *nef* over the study period, however, the relative proportions of subtype A and C-like sequences comprising the recombinant *gag* sequences remained constant. Over the entire 18-month sampling period the *gag* p17 region remained subtype C-like whereas the p24 region remained subtype A-like.

As expected from previous studies (Novisky *et al.*, 2002) low degrees of sequence diversity were observed in *gag* and *nef* relative to that observed in other HIV genome regions such as the *env*. Possibly because of this low degree of diversity, no evidence of positive selection was detected in either *gag* or *nef*.

Both Gag and Nef CTL epitopes are known to be conserved across subtypes (Inwoley *et al.*, 2005 and Thakar *et al.*, 2005) and, as we have predicted here with the detection of numerous potential CTL epitopes, these proteins are generally highly targeted by the cellular immune system (Masemola *et al.*, 2003). Importantly, most of the predicted epitopes identified in the Nef sequences from patient Tz14 were concentrated in the n-terminal and middle region of the protein and overlapped substantially with experimentally confirmed epitopes (Mashishi *et al.*, 2001). In Gag the predicted epitopes, which were mostly concentrated in the p24 region, also overlapped substantially with those that have been experimentally verified (Masemola *et al.*, 2004).

Importantly, the predicted CTL epitopes detected in *gag* were highly conserved throughout the study period. It is perhaps not surprising therefore, that no strong evidence of immune escape mutation was detected. Three of the four Nef epitopes screened showed limited diversity and were thus highly conserved throughout the study period. However, epitope RL-9 was present as an escape mutant at F0 with the anchor residue L (at position 9) been replaced by I in all sequences detected at that time point. At time point F4, five of the twenty sequences had the correct wild type sequence and finally only two of the twenty sequences Nef sequences at F7 harbored the escape mutant epitope, with all others (18/20) having the correct RL-9 epitope sequence. Despite this, however, some of the mutational variation detected here, particularly at the later time points, involved changes in one or both predicted epitope binding anchor residues that may have made these epitopes less recognizable to patient Tz14's CTLs (Leggatt *et al.*, 1998).

On the whole, epitopes restricted by HLA-B were better (75%) conserved than those restricted by HLA-A which may point to the more important role played by HLA-B targeted epitopes in CTL recognition and disease control (Kiepiela *et al.*, 2005). It is evident from this dual infection case study that, in the *gag* and *nef* genomic regions of viruses infecting patient Tz14 at least, viral diversification appears to have been driven largely by recombination between the dually infecting subtypes. Unlike with the step-wise accumulation of diversity achievable via point mutation, recombination provides a mechanism by which instantaneous leaps in diversity are achievable (Mulim *et al.*, 2001). While it is unclear from this study whether recombination has facilitated the concentration of potential immune escape mutations detectable in *gag* and *nef* over the study period, it is evident that, in *nef* at least, it has resulted in the displacement of an almost entirely

subtype C-like sequences by subtype A-like sequences. Although physical linkage of *nef* and *gag* is not demonstrated here, it is possible that recombination has, conversely, enabled preservation of the same *gag* sequence lineage as the predominant population member in Tz14. This study therefore provides good evidence of how, in a single patient, recombination facilitates the rapid diversification of some genome regions at the same time enabling the conservation of others.

CHAPTER 5
SUMMARY AND FINAL CONCLUSIONS

Since its first identification in the 1980s the human immunodeficiency virus (HIV-1) has infected multitudes and killed more than 10 million people the world over, with 4.3 million people infected in 2006 alone and 2.9 million lives lost in the same year (UNAIDS 2006). Since the advent of antiretroviral therapy the tide against HIV-1 has turned for the better especially in the developed world. In Africa, however, antiretroviral therapy has had limited impact; the continent is still the most affected with sub-Saharan Africa carrying the worst burden of the disease. An effective HIV vaccine remains our greatest hope of controlling the epidemic. Despite great efforts and improvements in medical science, some scientific questions on developing an effective vaccine remain unanswered. High HIV-1 viral diversity has been reported to pose a major challenge facing vaccine development (Novitsky *et al.*, 2002; Rowland-Jones *et al.*, 1998 and Williamson *et al.*, 2003). Africa harbours the highest HIV-1 diversity globally with various subtypes and their recombinants found in her populations. To further understand the implications of infection with highly divergent HIV strains, this study aimed to characterise viral dynamics in singly or dually infected individuals. This project formed part of a larger study investigating correlates of protection in infected individuals who become superinfected with a second HIV-1 subtype (www.mmrp.org).

The multi-region hybridisation assay (MHA) is based on a principle wherein five genomic regions are genotyped simultaneously, in that way both dual infections and recombinant forms can be detected without sequencing the whole genome (Hoelscher *et al.*, 2002). Using this assay we found that recombinant strains were responsible for most infections comprising 44% of samples characterized, with AC recombinant constituting 6o% of all recombinants. Subtype C containing recombinants constituted 84% of all recombinant forms detected, while subtype C was the dominant subtype at 33%, with subtype A at 5% and D constituted 7%. Dual infections were observed in 11% of samples genotyped. Our study adds to available data which elucidates the genetically complex HIV-1 epidemic where multiple subtypes and their recombinants co-circulate (Arroyo *et*

al., 2004 and Hoelscher *et al.*, 2001). Together these studies show that viruses circulating in Mbeya, Tanzania are predominately subtype C containing recombinants, subtypes A, C and D plus numerous unique recombinant forms. The high prevalence of dual infections is clearly fuelling the generation of recombinant viruses and it would be of interest to monitor the molecular epidemiology of the epidemic to determine if new circulating recombinant forms emerge.

We performed a detailed characterization of evolving viral populations over a 21 month period in twelve women. Four women were found to be dually infected, two were infected by recombinant viral strains, four harboured subtype C viruses, and subtypes A and D infections accounted for one individual each. All viral strains obtained were predicted to be R5 viruses and no co-receptor switch to X4 was observed irrespective of disease status or subtype. Viral strains were characterized by fluctuations over time, with varying degrees of viral diversity observed among individuals. In addition, diversification differed between the two genomic regions suggesting different selective pressures. An analysis of the of the contribution of different viral variants in dual infection to overall viral burden illustrated large fluctuations of viral populations over time. However, generally the viral populations were relatively homogenous at a single time point suggesting that, while there is very high potential for diversity, this diversity is largely constrained.

We assessed the mechanism of *gag* and *nef* evolution in one dually infected individual (Tz14), diversity through point mutations was compared to diversity due to recombination. None of the viruses from this individual were 'pure' subtypes in these regions, and both subtypes A and C short sequences were detected at different proportions. However, *nef* sequences showed a displacement of subtype C-like viruses by subtype A-like viruses through an intermediate where both subtypes A and C were detected. While there was no subtype displacement of *gag* sequences, sequences revealed the presence of recombinant viral strains through out the study period. Recombination was evident in both *gag* and *nef* regions of viruses from this dually infected individual. Few point mutations were detected in conserved CTL epitopes of both regions but no

positive election was detected towards these epitopes. We propose that the increase viral diversity in dual infection is enabled mostly by recombination and to a lesser extent by point mutations. Some evidence of CTL escape, facilitated through point mutations in some viral sequences from this individual, were observed in the *gag* and *nef* regions. It has been hypothesized that the ability to undergo major genetic shift in viral populations may be the reason behind rapid disease progression in individuals with dual infection (Grobler *et al.*, 2004).

In conclusion, this study provided insights into a genetically complex HIV-1 epidemic where multiple subtypes co-circulate among a high risk group population. The presence of multiple subtypes is the contributing factor to the high prevalence of both dual infections and recombinant viral strains in this region. Viral dynamics over time where shown to be characterized by fluctuation of different viral populations. Although higher viral diversity was exhibited by the dually infected individuals compared to individuals with single infections over time, at certain time points diversity was often contained. Recombination seems to be the mechanism involved in driving viral diversity among dually infected individuals. The presence of multiple subtypes in a population and their influence in the formation of recombinant strains poses another challenge for both HV-1 vaccine development strategies and therapies.

APPENDIX A : STANDARD DNA AND RNA TECHNIQUES

A1. Extraction of viral RNA

RNA was manually extracted using the QIAamp® Viral RNA Mini Kit for the purification of viral RNA from plasma (Qiagen, Valencia, CA). A 140-280µl sample aliquot was added to 560µl prepared Buffer AVL containing carrier RNA in a 1.5ml micro-centrifuge tube, mixed by pulse-vortexing for 15 seconds, incubated at room temperature (15 -25°C) for 10- minutes and briefly centrifuged to remove drops from the inside of the lid. Ethanol (560µl;96-100%) was added and mixed by pulse vortexing for 15 seconds, and briefly centrifuged to remove drops form the inside of the lid and the solution was carefully applied to the QIAmp spin columns and centrifuged at 6000x g (8 000rpm) for 1 minute. The column was placed in a clean 2ml collection tube and the tube containing the filtrate was discarded. Next 500µl Buffer AW1 was added and centrifuged at 6000x g (8 000rpm) for 1 minute, the column placed in a clean collection tube and the tube containing the filtrate was discarded. Next 500µl Buffer AW2 was added and centrifuged at 20 000x g (14 000rpm) for 3 minutes and the tube containing the filtrate was discarded. The column was then placed in a clean 2ml collection tube and centrifuged for an additional 1 minute at full speed. The sample was then eluted after a one minute incubation step with 60µl AVE buffer, aliquoted to 10µl-20µl and either used directly in the cDNA synthesis or stored at –80°C.

A2. PCR Purification

PCR products were purified by the QIAquick Spin Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Briefly PB buffer (5x volume of PCR product) was mixed with the PCR product and loaded on the spin column and centrifuged for a minute at 13000rpm. PE buffer (750µl) was used to wash by centrifugation at 13000rpm for one minute. Clean PCR product was eluted by centrifugation at 13000rpm using buffer EB (50µl).

A3. Transformation

Preparation of competent cells

Both commercial available competent cells (cells supplied with the kits) and *E. coli* DH5α cells, prepared according to the dimethyl sulphoxide (DMSO) method (Chung & Miller, 1988) were used for transformation. A 5ml volume of 2YT broth (Appendix B) was inoculated and incubated overnight at 37°C on a shaker. A dilution of 1/100 was made from the culture into 100ml 2YT broth in a litre culture flask and grown on the shaker to early log phase (OD_{600} 0.2 to 0.4). The cells were then harvested by centrifugation at 5000 rpm for 5 min at 4°C in a Beckman J2-21 centrifuge. The pellet was resuspended in ice cold TSB

buffer (Appendix B) and placed on ice for 10 min. Sterile glycerol was added to a final concentration of 10% v/v and 100 µl aliquots were stored at -80°C.

Transformation

Frozen cells were defrosted on ice and the 3µl ligation mix was then mixed with the cells and left on ice between 10 to 20 min. The cells were heat shocked at 42 °C for 2 min. One milliliter of 2YT medium was added and the cells were incubated for 30 to 45 min at 37°C. A volume of 50µl to 100µl was plated on selective plates containing either ampicillin or kanamycin (Appendix B). Blue/White screening was done with pGEM®-T Easy vector by the addition of Xgal (5-bromo-4-chloro-3-indolyl-β-D-galactosidase) and IPTG (isopropyl-β-D-thio-galactopyranoside) to plates (Appendix B). The plates were then incubate at 37°C and single colonies were either inoculated into 2 ml 2YT broth for minipreparations or for colony PCR screening.

A4. Colony Screening

Following transformation colonies were screened by colony PCR as described by (Lee and Cooper, 1995). The colonies were picked using a plastic pipette tip, replicated on another plate for future reference and inserted into a 25µl to 50µl PCR master mix as described in chapter 2 (2.3.6). The PCR conditions and master mixes were prepared the same way as in second round PCRs and 5µl of the PCR product was run on a gel to confirm the presence of the insert. Posive PCR fragments were cleaned up for sequencing.

A5. Agarose gel electrophoresis

Fragments sizes of 510bp - 700bp were visualized in 2% agarose gels. Agarose gel electrophoresis was performed using horizontal gel apparatus (Stratagene, La Jolla, USA). The agarose gel was prepared by melting the appropriate weight per volume agarose (Agarose Di LE, Hispanagar, Burgos, Spain) in 1X TBE (Appendix B1). Once melted 10mg/ml ethidium bromide stock solution was added to a final concentration of 1µg/100ml agarose. The agarose was poured into gel setting trays and cooled down to room temperature to allow setting and then placed in a gel apparatus submerged in 1X TBE. Before loading, 1-5µl PCR product was mixed with 2µl of 6X agarose gel electrophoresis loading dye (Appendix B2). In order to determine the size of the amplicons a DNA molecular weigh marker VI (Roche, GmbH, Mannheim, Germany, Appendix C2) was included in the first lane of all gels. The gel was electrophoresed at 100 to 120V according to gel size for 60 minutes or until sufficient separation of bands. The DNA fragments were then visualized on a UVP transilluminator (UVP, San Gabriel, California, USA) at 256nm wavelength and photographed with a Kodak ds 1D, Electrophoresis Documentation and Analysis System 120, V 2.0.3. computerized gel imager (Kodak ds 1D digital science, version 2.03).

A6. HMA

HMA was used to assess the genetic diversity of each PCR-amplified sample. In a 200 µl Eppendorf PCR tube approximately equal amounts of amplified fragments from the unknown and each reference were mixed and heated to 94 - 96°C for 2 minutes in a thermocycler (or in a boiling water bath). This was allowed to cool at room temperature for 5 min. Heteroduplexes was kept at room temperature until loading loading the gel. The reaction was mixed with 3 µl 5X Ficoll loading dye and loaded onto a 5% non-denaturing polyacrylamide gel. The gel appatarus was the Scie-plas/apelex CDC series of Vertical Gel Electrophoresis Maxi V20-CDC 20 x 20cm – V20-CDC (Scie-Plas Ltd, Southam Warwickshire, England). This is the Scie-plas/apelex standard Maxi Vertical Gel Unit comparable to the Biorad Protean II units (Biorad-Laboroaries, Hercules, CA). The active gel dimensions are 16.5 x 17.5cm.

Gel electrophoresis of HMA reactions

The glass plates were assembled as a sandwich first with two 3mm spacers. The glass plates were then slipped into the gel running unit with the notched plates facing inwards to create the inner buffer tank. The gel was made up according to appendix B and poured between the plates. It was left to polymerize for at least 90 minutes. After polymerization the gel was placed in the buffer tank and 1X TBE buffer into the tank so that at least 2cm of the glass plate was covered with buffer. The upper buffer tank was filled until the wells were covered (top of the notched plate). The wells were flushed with a needle and syringe. A P20 PIPETMAN Gilson® micropipette (Gilson Medical Electronics, France) was used to load 15µl of the HMA reaction onto each well with a long 1-2mm tip. The lid was placed on the electrophoretic box. The gel was run at 200V for 4 hours. The gel was removed by inserting a spatula between the glass plates and the plates gently lifted. Both the gel and plate was placed in an agitating ETBR 100µl (10mg/ml) solution until the gel separates and was agitated for a further 30min at room temperature. The DNA fragments were then visualized on a UVP transilluminator (UVP, San Gabriel, California, USA) at 256nm wavelength and photographed with a Kodak ds 1D, Electrophoresis Documentation and Analysis System 120, V 2.0.3 computerized gel imager (Kodak ds 1D digital science, version 2.03).

APPENDIX B: STANDARD BUFFERS AND SOLUTIONS

B1. 10X TBE (Tris-Boric acid EDTA) buffer:

108g Tris-HCl, 55g Boric acid 20ml 1,5M EDTA, made up to one litre with dH_2O.

B2. 6X Agarose Gel Electrophoresis Loading Dye:

0,25% bromophenol blue, 0,25% xylene cyanol FF 30% glycerol (In deionised water)

B3. X-Gal, IPTG, Ampicillin LB agar plates:

Per litre: 10g tryptone, 5g yeast extract, 10g NaCl, 15g agar after agar has been sterilised by autoclavation and allowed to cool down to ~ 50°C the following selective regents were added: 15µg/ml tetracycline, 100µg/ml ampicillin, 80µg/ml X-gal, 0,5mM IPTG.

B4. 6X Polyacrylamide gel electrophoresis loading dye:

0,25% Bromophenolblue, 0,25% Xylene Xylanol, 15% Ficoll in deionised water.

B5. HMA Annealing Buffer:

100mM Tris/HCl, 1M NaCl, 20mM EDTA

B6. LB Broth:

10g NaCl, 5g Yeast extract, 10g Tryptone, up to 1lµl with deionised water

B7. TSB:

1.6g peptone, 1.0g yeast extract, 0.5g NaCl, 10g polyethylene glycol (PEG) 3350-4000, 5ml DMSO, 1ml $1MgCl_2$, 1ml 1M Mg SO_4 plus water to 100ml. Store in 10ml aliquots.

B8. 2YT (yeast-tryptone) liquid medium (pH7.0):

16g tryptone, 10g yeast extract, 5g NaCl

B9. 2YT plates:

As for liquid 2YT, with an additional 15g agar per litre

B10. Antibiotic and Xgal/IPTG selection:

Ampicillin (Sigma, MO, USA) 100µg/ml was added to liquid media and plates, kanamycin (Km; Nova Nordisk, Johannesburg, R.S.A) at a concentration of 15 µg/ml. Xgal (Biosolve, Netharlands) 20 mg/ml stock of 0.4g dissolved in 5ml DMSO and made up to 10ml and IPTG (Roche, Germany) 200mg/ml stock were both added to plates at a final concentration of 0.1 mg/ml.

B11. 5% Non-denaturing Polyacrylamide Gel

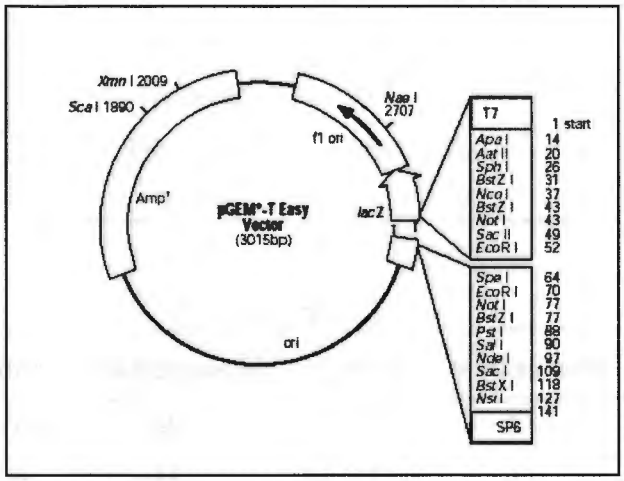
8.3 ml acrylamide stock, (using a stock of 30% Acrylamide, 0.8% Bis-acrylamide) 5 ml 10X TBE and 36.7 ml H2O. Polymerization is initiated by mixing in 50 mg of Ammonium persulfate and 33 µl of TEMED (50 ml gel solution)

B12. 1X Electrophoresis Buffer

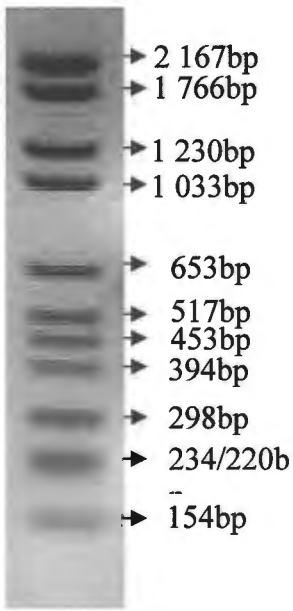
(Dilute from 10X TBE Gel Electrophoresis Buffer above)

Dilute 1/10 to a final concentration of 89 mM Tris-borate, pH 8.0, 89 mM Boric Acid 2 mM EDTA in deionised water. 108g Tris-HCl, 55g Boric acid 20ml 1,5M EDTA, made up to one litre with dH_2O.

C1. pGEM® -T Easy (Promega)



C2. Roche Molecular Weight Marker VI (Roche, GmbH, Mannheim, Germany)



C3. AMINO ACID CODE

Alanine	Ala	A	Leucine	Leu	L	Arginine	Arg	R
Lysine	Lys	K	Asparagine	Asn	N	Methionine	Met	M
Aspartic acid	Asp	D	Phenyalanine	Phe	F	Cysteine	Cys	C
Proline	Pro	P	Glutamine	Gln	Q	Serine	Ser	S
Glutamic acid	Glu	E	Threonine	Thr	T	Glycine	Gly	G
Tryptophan	Trp	W	Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V			

C4. CODONS

Codon	Amino acid	Codon	Amino acid	Codon	Amino acid	Codon	Amino acid
AAA	K	CAA	Q	GAA	E	TAA	*
AAC	N	CAC	H	GAC	D	TAC	Y
AAG	K	CAG	Q	GAG	E	TAG	*
AAT	N	CAT	H	GAT	D	TAT	Y
ACA	T	CCA	P	GCA	A	TCA	S
ACC	T	CCC	P	GCC	A	TCC	S
ACG	T	CCG	P	GCG	A	TCG	S
ACT	T	CCT	P	GCT	A	TCT	S
AGA	R	CGA	R	GGA	G	TGA	*
AGC	S	CGC	R	GGC	G	TGC	C
AGG	R	CGG	R	GGG	G	TGG	W
AGT	S	CGT	R	GGT	G	TGT	C
ATA	I	CTA	L	GTA	V	TTA	L
ATC	I	CTC	L	GTC	V	TTC	F
ATG	M	CTG	L	GTG	V	TTG	L
ATT	I	CTT	L	GTT	V	TTT	F

APPENDIX D: VIRAL LOADS

Table D1: Viral loads supplied by Dr. Adrian Puren from the NICD, Johannesburg.

Sample number	Follow up	Viral load (copies/ml)	Log viral load
Tz 7	F0 (1st time point)	307 000	4.48
	F1 (3months)	5 040	3.70
	F2 (6months)	142 000	5.15
	F3 (9months)	179 000	5.25
	F4 (12months)	181 000	5.25
	F5 (15months)	84 900	4.92
	F6 (18months)	472 000	5.67
	F7 (21months)	527 000	5.72
	Median	**160 500**	**5.20**
Tz 10	F0 (1st time point)	23 400	4.36
	F1 (3months)	503 000	5.70
	F2 (6months)	149 000	5.17
	F3 (9months)	284 000	5.45
	F4 (12months)	7 590	3.88
	F5 (15months)	72 800	4.86
	F6 (18months)	131 000	5.11
	F7 (21months)	50 900	4.70
	Median	**101 900**	**5.00**
Tz 14	F0 (1st time point)	24 100	4.38
	F1 (3months)	41 200	4.61
	F2 (6months)	217 000	5.33
	F3 (9months)	56 800	4.75
	F4 (12months)	554 000	5.74
	F5 (15months)	107 000	5.02
	F6 (18months)	137 000	5.13
	F7 (21months)	53 900	4.73
	Median	**81 900**	**4.91**
Tz 48	F0 (1st time point)	14 400	4.15
	F1 (3months)	28 300	4.45
	F2 (6months)	18 000	4.25
	F3 (9months)	20 100	4.30
	F4 (12months)	119 000	5.07
	F5 (15months)	33 200	4.52
	F6 (18months)	214 000	5.33
	F7 (21months)	64 000	4.80
	Median	**30 750**	**4.48**
Tz 76	F0 (1st time point)	19 100	4.28
	F1 (3months)	12 500	4.09
	F2 (6months)	20 300	4.30
	F3 (9months)	5 480	3.73
	F4 (12months)	11 700	4.06
	F5 (15months)	6 460	3.81
	F6 (18months)	5 320	3.72
	F7 (21months)	28 900	4.46
	Median	**12 100**	**4.08**

Tz 118	F0 (1st time point)	6 990		3.84	
	F1 (3months)		3 180		3.50
	F2 (6months)		5 770		3.76
	F3 (9months)		5 080		3.70
	F4 (12months)		7 630		3.88
	F5 (15months)		8 580		3.93
	F6 (18months)		6 220		3.79
	F7 (21months)		9 840		3.99
Tz 126	F0 (1st time point)		102 000		5.00
	F1 (3months)		295 000		5.46
	F2 (6months)		73 100		4.86
	F3 (9months)		72 700		4.86
	F4 (12months)		1 49 000		5.17
	F5 (15months)		63 900		4.80
	F7 (21months)		47 800		4.67
	Median		**73 100**		**4.86**
Tz 203	F0 (1st time point)		19 200		4.28
	F1 (3months)		22 200		4.34
	F2 (6months)		17 900		4.25
	F3 (9months)		32 700		4.51
	F4 (12months)		37 700		4.57
	F5 (15months)		42 800		4.63
	F6 (18months)		62 500		4.79
	F7 (21months)		48 500		4.68
	Median		**35 200**		**4.54**
Tz 217	F0 (1st time point)		78 600		4.89
	F1 (3months)		110 000		5.04
	F2 (6months)		70 500		4.84
	F3 (9months)		191 000		5.28
	F4 (12months)		107 000		5.02
	F5 (15months)		295 000		5.46
	F6 (18months)		640 000		5.80
	F7 (21months)		218 000		5.33
	Median		**150 500**		**5.17**
Tz 218	F0 (1st time point)		21 300		4.32
	F1 (3months)		51 500		4.71
	F2 (6months)		52 300		4.71
	F3 (9months)		55 500		4.74
	F4 (12months)		124 000		5.09
	F5 (15months)		60 600		4.78
	F6 (18months)		85 500		4.93
	F7 (21months)		89 100		4.94
	Median		**58 050**		**4.76**
Tz 220	F0 (1st time point)		104 000		5.01
	F1 (3months)		102 000		5.00
	F2 (6months)		28 200		4.45
	F3 (9months)		49 400		4.69
	F4 (12months)		64 900		4.81
	F5 (15months)		81 700		4.91
	F6 (18months)		55 800		4.74
	F7 (21months)		64 400		4.80
	Median		**58 050**		**4.81**

Tz 440	F0 (1st time point)	188 000	5.27
	F1 (3months)	267 000	5.42
	F2 (6months)	137 000	5.13
	F3 (9months)	358 000	5.55
	F4 (12months)	392 000	5.59
	F5 (15months)	320 000	5.50
	F6 (18months)	182 000	5.26
	F7 (21months)	379 000	5.57
	Median	**293 500**	**5.47**

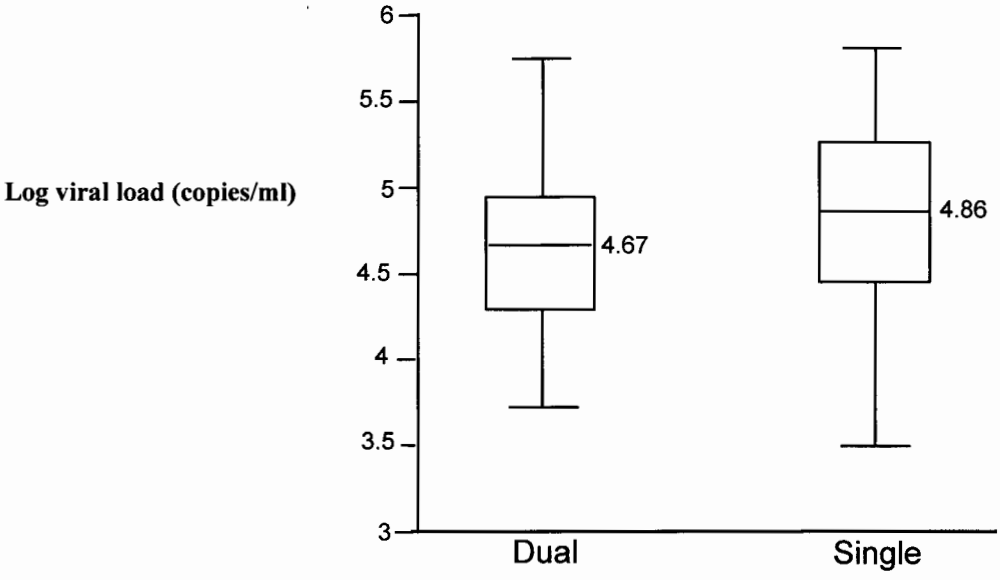


Figure D1: Viral load differences between single and dual infections. The highest and lowest values are presented by the upper and lower bars, the median values are indicated by the parallel bars within boxes and the 25th and 75th percentiles are indicated by the top and bottom of the boxes. Individuals with single infections had higher viral loads than dually infected individuals ($p > 0.05$).

APPENDIX E: DNA DISTANCES

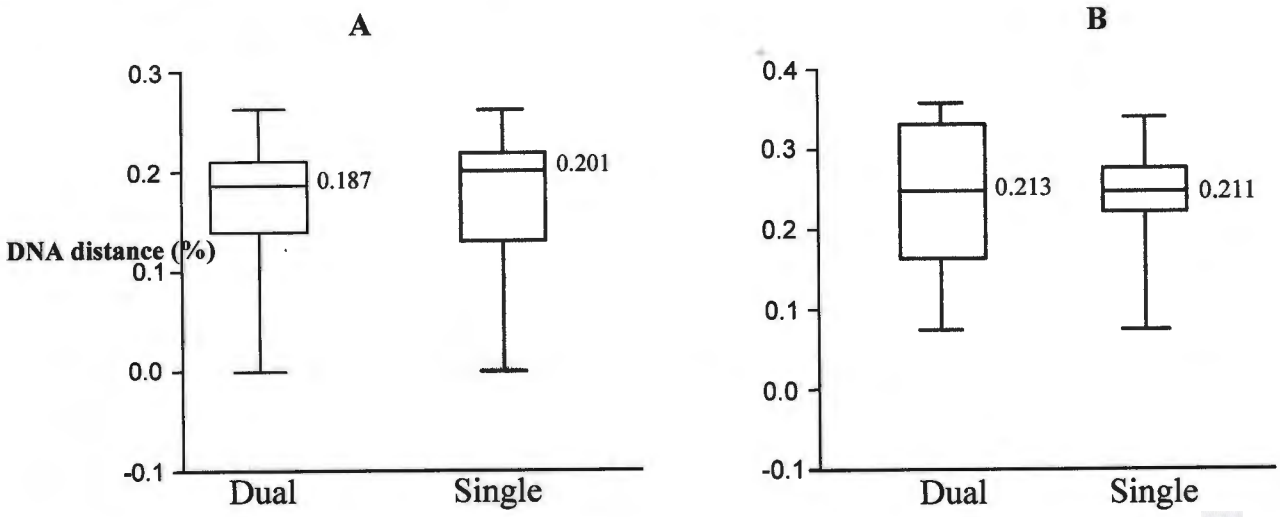
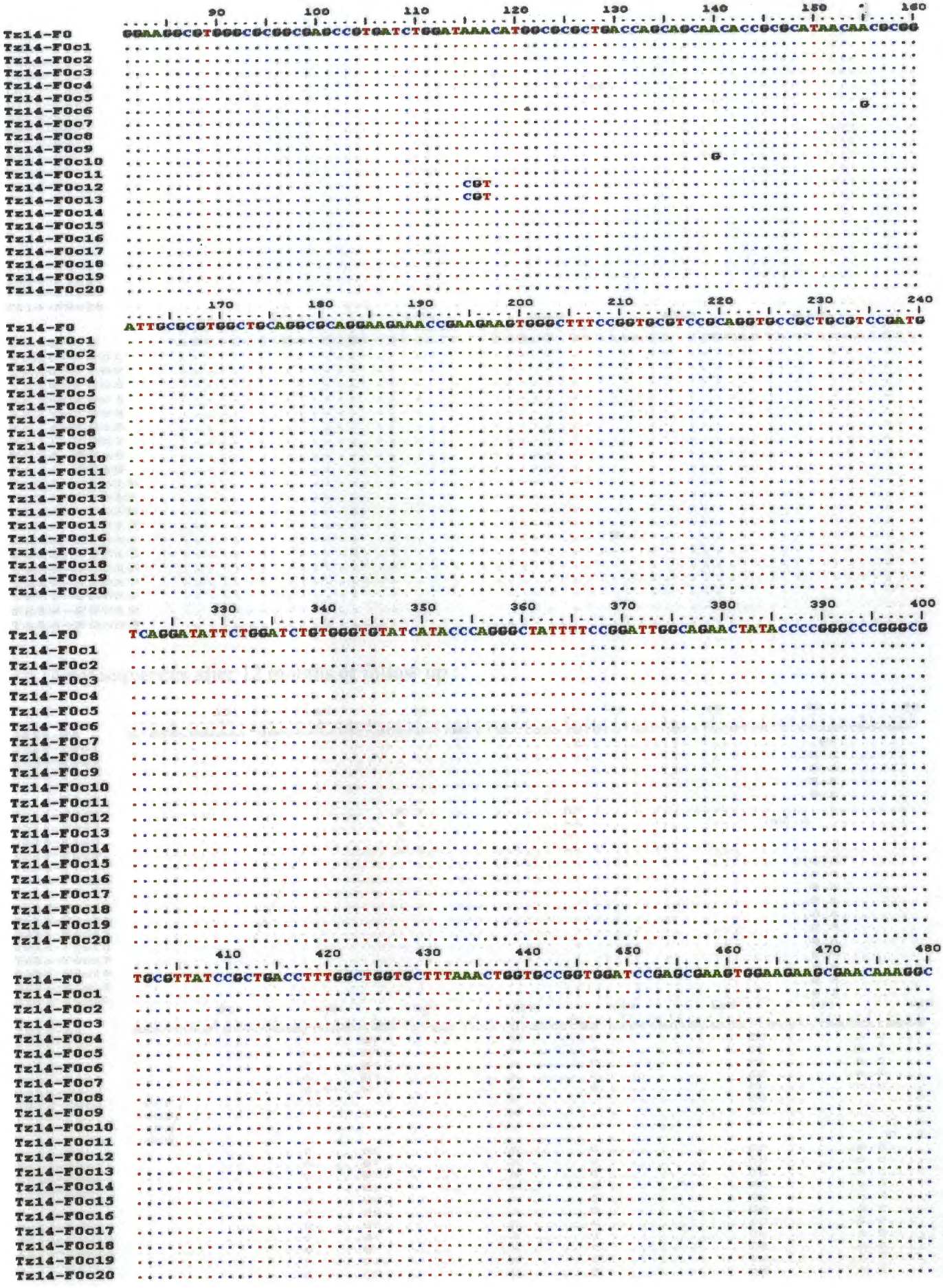


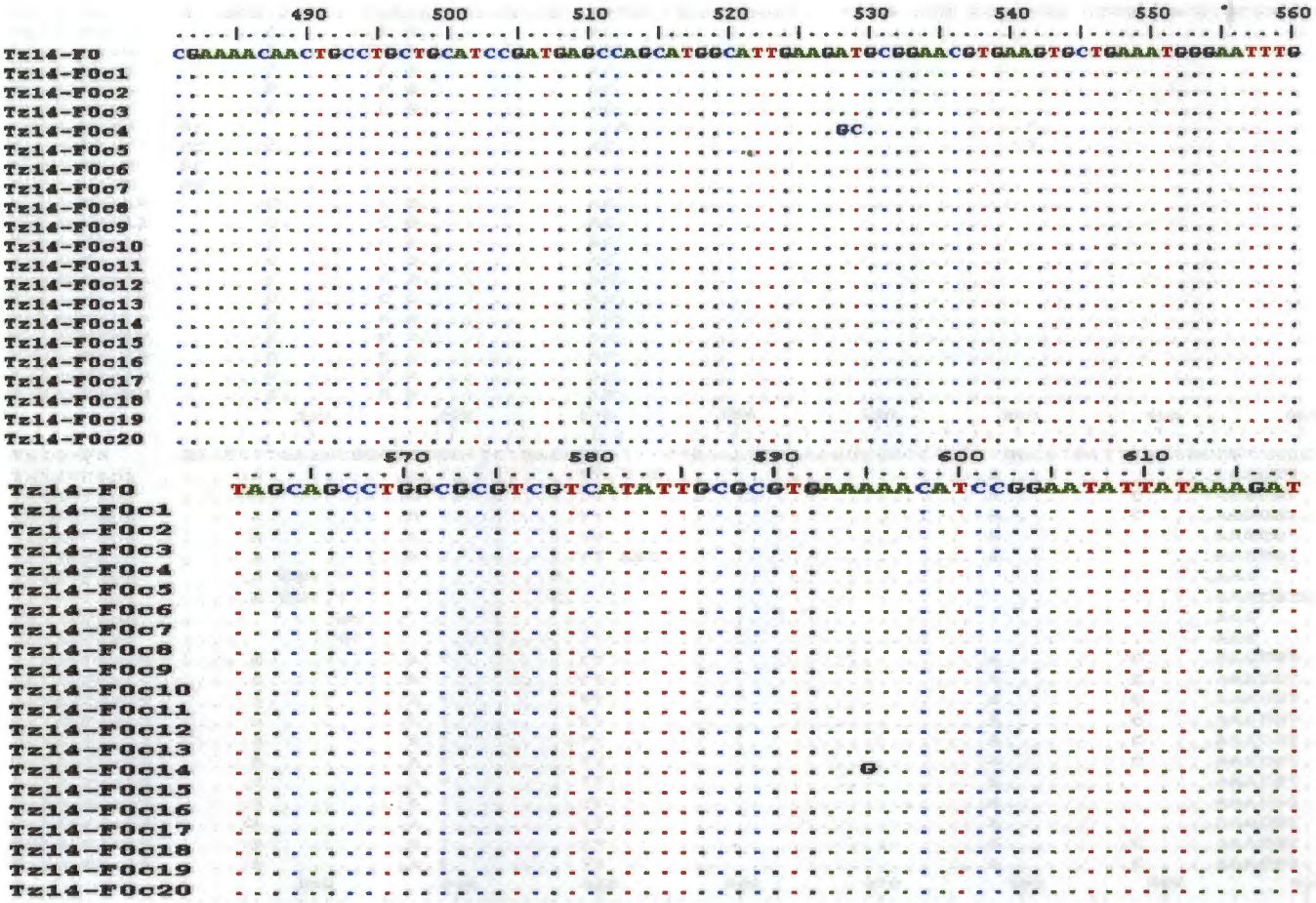
Figure E1: Median DNA distances comparison of single and dually infected individuals in the *env* C2-C3 (A) and *vpu* (B) genomic regions. Dually infected individuals showed lower DNA distance than individuals with single infection in the *env* C2-C3 while they showed higher DNA distance in the *vpu* region compared to individuals with single infections. DNA distances were calculated from the dominant viral population at each time point. The median values are indicated by the parallel bars within boxes and the 25th and 75th percentiles are indicated by the top and bottom of the boxes.

APENDIX F: Tz 14 FULL LENGTH NEF CLONE SEQUENCES FROM TIME POINT ZERO, TWELVE AND 21 MONTHS OF FOLLOW UP

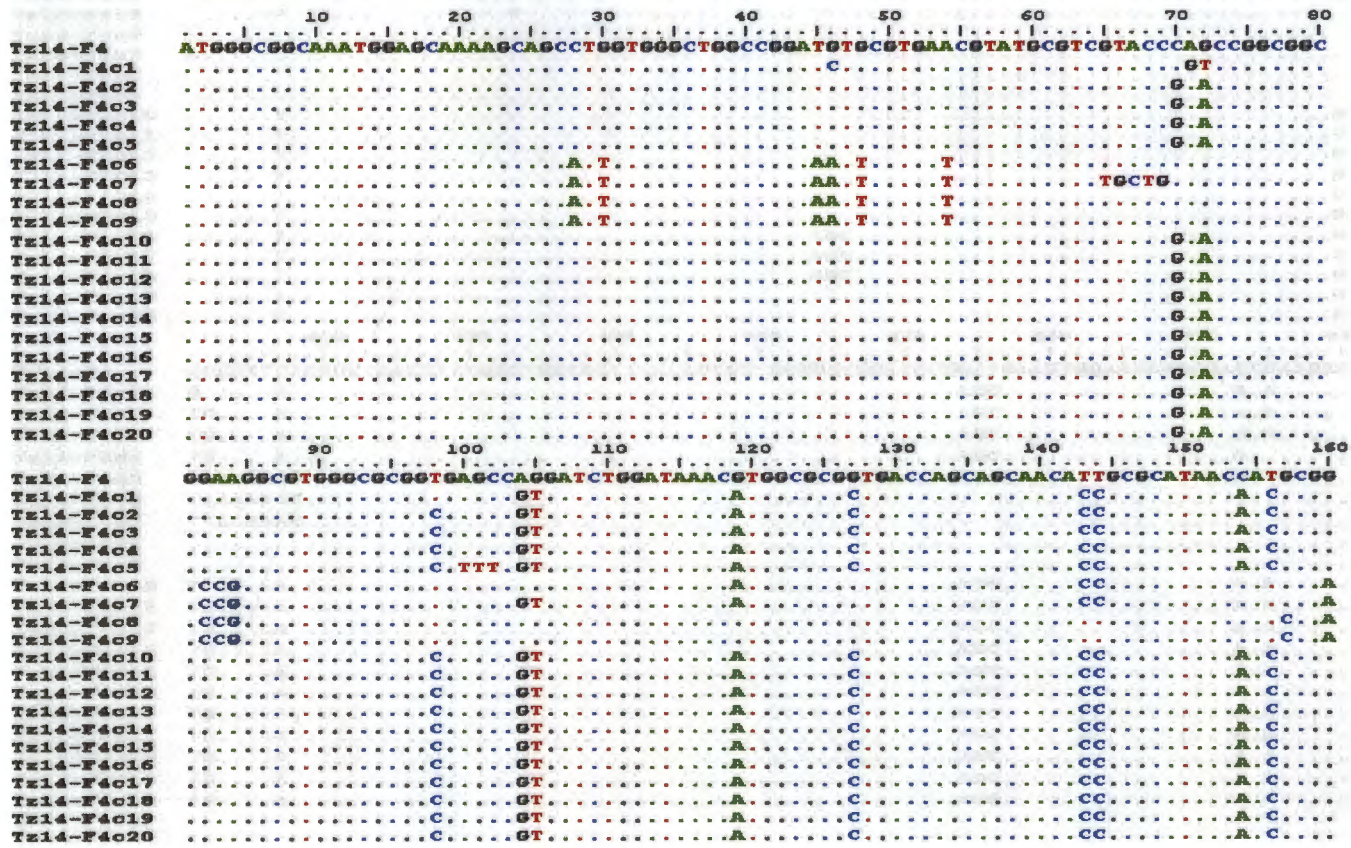
Tz 14 *nef* sequences from the first time point :

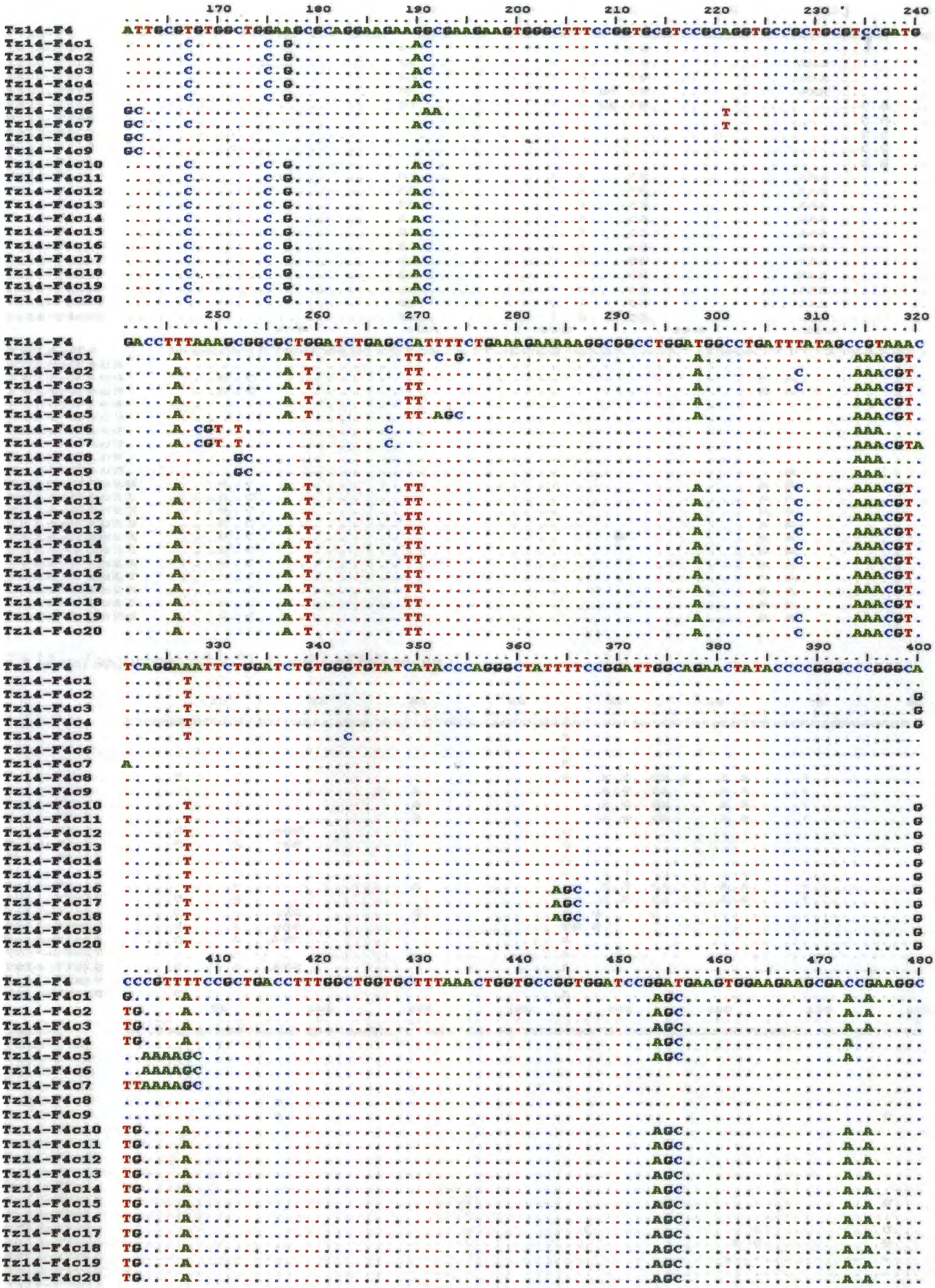
```
                    10        20        30        40        50        60        70        80
                    |....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Tz14-F0             ATGGGCGGCAAATGGAGCAAAAGCAGCCTGGTGGGCTGGCCGAACGTGCGTGAACGTATGCGTCGTACCGAACCGGCGGC
Tz14-F0c1           ................................................................................
Tz14-F0c2           ................................................................................
Tz14-F0c3           ................................................................................
Tz14-F0c4           ................................................................................
Tz14-F0c5           ................................................................................
Tz14-F0c6           ................................................................................
Tz14-F0c7           ................................................................................
Tz14-F0c8           ................................................................................
Tz14-F0c9           ................................................................................
Tz14-F0c10          ................................................................................
Tz14-F0c11          ................................................................................
Tz14-F0c12          ................................................................................
Tz14-F0c13          ................................................................................
Tz14-F0c14          ................................................................................
Tz14-F0c15          ................................................................................
Tz14-F0c16          ................................................................................
Tz14-F0c17          ................................................................................
Tz14-F0c18          ................................................................................
Tz14-F0c19          ................................................................................
Tz14-F0c20          ................................................................................
```

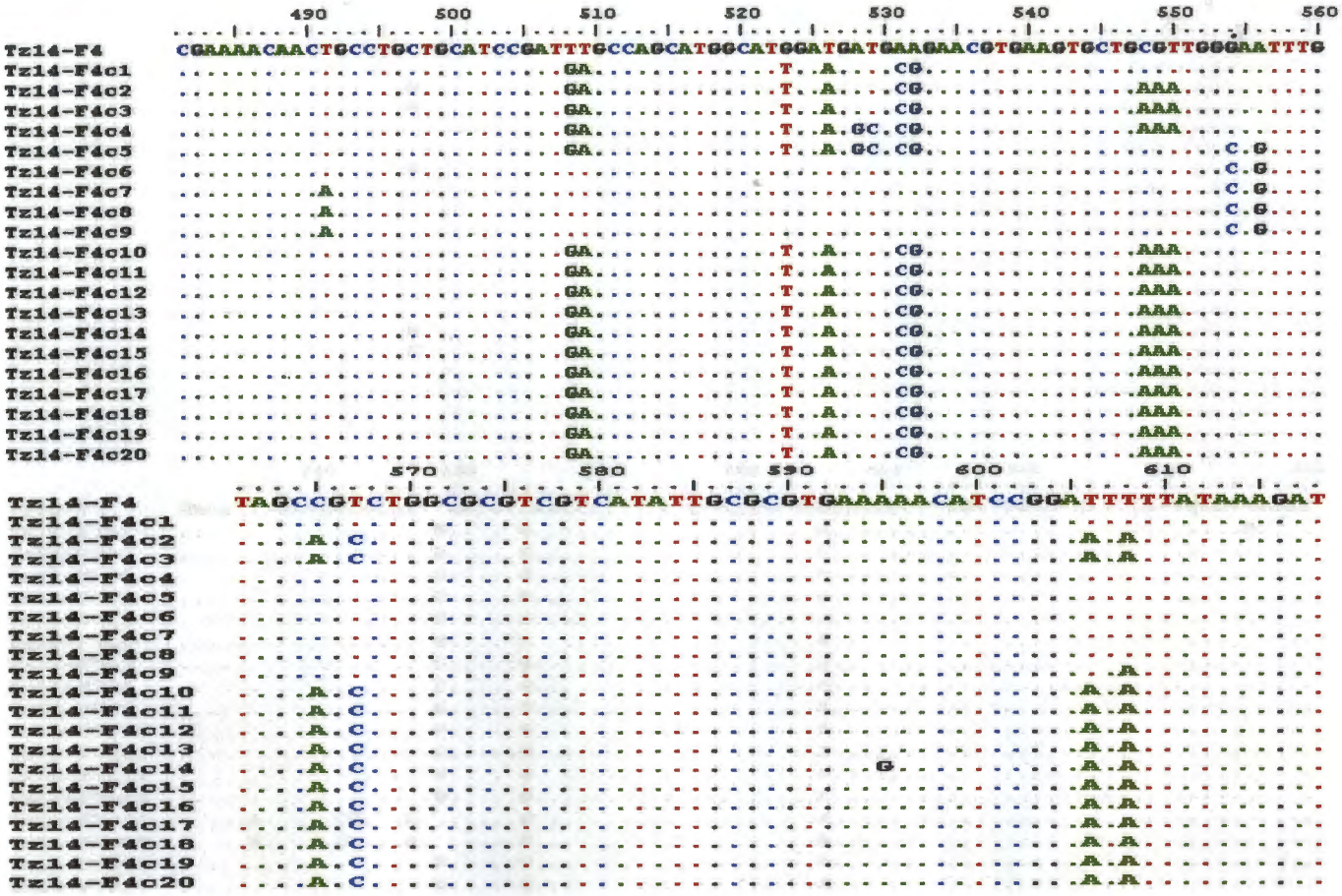

90 100 110 120 130 140 150 160
Tz14-F0
Tz14-F0c1
Tz14-F0c2
Tz14-F0c3
Tz14-F0c4
Tz14-F0c5
Tz14-F0c6
Tz14-F0c7
Tz14-F0c8
Tz14-F0c9
Tz14-F0c10
Tz14-F0c11
Tz14-F0c12
Tz14-F0c13
Tz14-F0c14
Tz14-F0c15
Tz14-F0c16
Tz14-F0c17
Tz14-F0c18
Tz14-F0c19
Tz14-F0c20
170 180 190 200 210 220 230 240
Tz14-F0
Tz14-F0c1
Tz14-F0c2
Tz14-F0c3
Tz14-F0c4
Tz14-F0c5
Tz14-F0c6
Tz14-F0c7
Tz14-F0c8
Tz14-F0c9
Tz14-F0c10
Tz14-F0c11
Tz14-F0c12
Tz14-F0c13
Tz14-F0c14
Tz14-F0c15
Tz14-F0c16
Tz14-F0c17
Tz14-F0c18
Tz14-F0c19
Tz14-F0c20
330 340 350 360 370 380 390 400
Tz14-F0
Tz14-F0c1
Tz14-F0c2
Tz14-F0c3
Tz14-F0c4
Tz14-F0c5
Tz14-F0c6
Tz14-F0c7
Tz14-F0c8
Tz14-F0c9
Tz14-F0c10
Tz14-F0c11
Tz14-F0c12
Tz14-F0c13
Tz14-F0c14
Tz14-F0c15
Tz14-F0c16
Tz14-F0c17
Tz14-F0c18
Tz14-F0c19
Tz14-F0c20
410 420 430 440 450 460 470 480
Tz14-F0
Tz14-F0c1
Tz14-F0c2
Tz14-F0c3
Tz14-F0c4
Tz14-F0c5
Tz14-F0c6
Tz14-F0c7
Tz14-F0c8
Tz14-F0c9
Tz14-F0c10
Tz14-F0c11
Tz14-F0c12
Tz14-F0c13
Tz14-F0c14
Tz14-F0c15
Tz14-F0c16
Tz14-F0c17
Tz14-F0c18
Tz14-F0c19
Tz14-F0c20



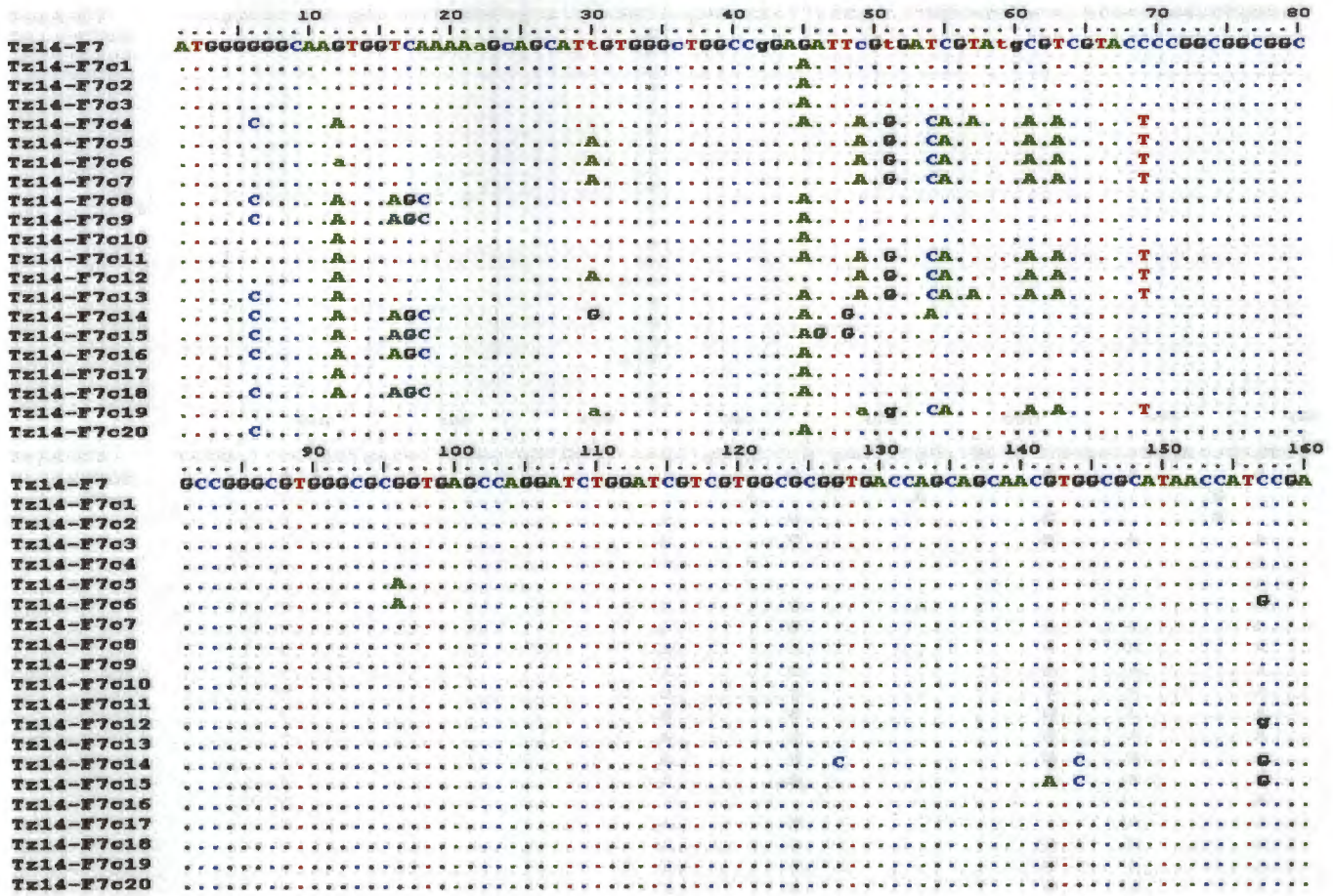
Tz 14 *nef* sequences after 12 months of follow up :

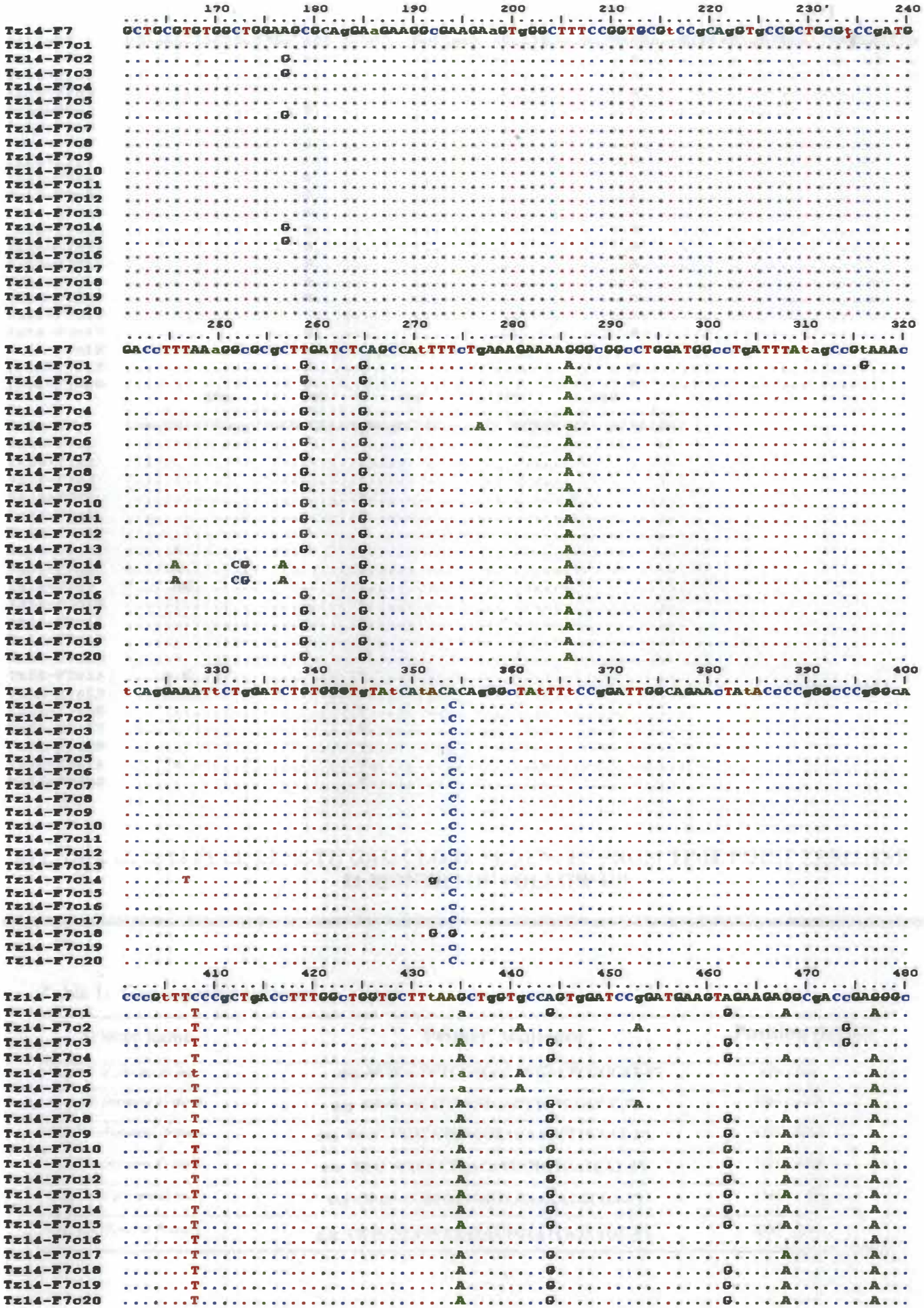


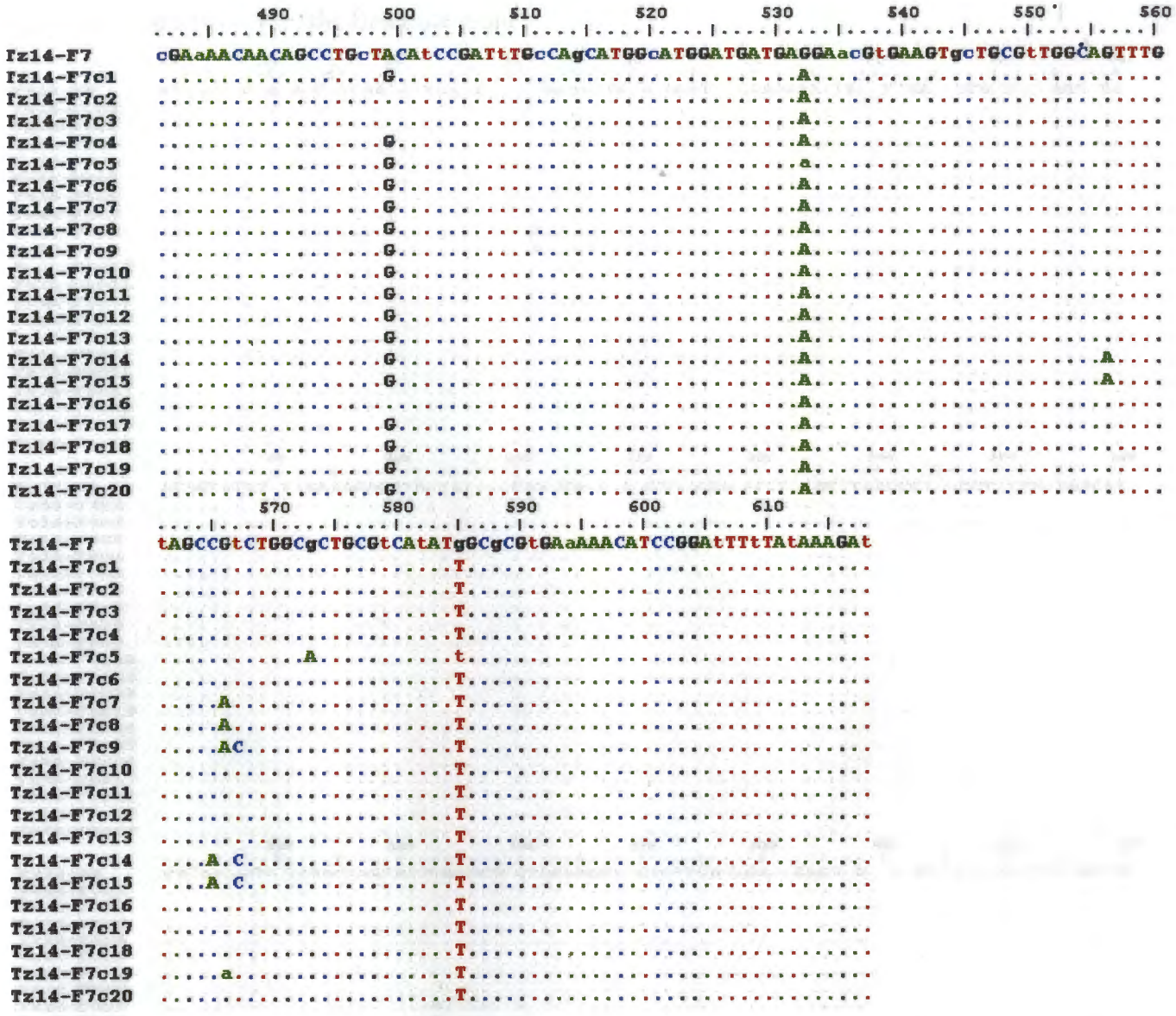

170 180 190 200 210 220 230 240
Tz14-F4
Tz14-F4c1
Tz14-F4c2
Tz14-F4c3
Tz14-F4c4
Tz14-F4c5
Tz14-F4c6
Tz14-F4c7
Tz14-F4c8
Tz14-F4c9
Tz14-F4c10
Tz14-F4c11
Tz14-F4c12
Tz14-F4c13
Tz14-F4c14
Tz14-F4c15
Tz14-F4c16
Tz14-F4c17
Tz14-F4c18
Tz14-F4c19
Tz14-F4c20
250 260 270 280 290 300 310 320
Tz14-F4
Tz14-F4c1
Tz14-F4c2
Tz14-F4c3
Tz14-F4c4
Tz14-F4c5
Tz14-F4c6
Tz14-F4c7
Tz14-F4c8
Tz14-F4c9
Tz14-F4c10
Tz14-F4c11
Tz14-F4c12
Tz14-F4c13
Tz14-F4c14
Tz14-F4c15
Tz14-F4c16
Tz14-F4c17
Tz14-F4c18
Tz14-F4c19
Tz14-F4c20
330 340 350 360 370 380 390 400
Tz14-F4
Tz14-F4c1
Tz14-F4c2
Tz14-F4c3
Tz14-F4c4
Tz14-F4c5
Tz14-F4c6
Tz14-F4c7
Tz14-F4c8
Tz14-F4c9
Tz14-F4c10
Tz14-F4c11
Tz14-F4c12
Tz14-F4c13
Tz14-F4c14
Tz14-F4c15
Tz14-F4c16
Tz14-F4c17
Tz14-F4c18
Tz14-F4c19
Tz14-F4c20
410 420 430 440 450 460 470 480
Tz14-F4
Tz14-F4c1
Tz14-F4c2
Tz14-F4c3
Tz14-F4c4
Tz14-F4c5
Tz14-F4c6
Tz14-F4c7
Tz14-F4c8
Tz14-F4c9
Tz14-F4c10
Tz14-F4c11
Tz14-F4c12
Tz14-F4c13
Tz14-F4c14
Tz14-F4c15
Tz14-F4c16
Tz14-F4c17
Tz14-F4c18
Tz14-F4c19
Tz14-F4c20



Tz 14 *nef* sequences after 21 months of follow up :




170 180 190 200 210 220 230 240
Tz14-F7
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20
250 260 270 280 290 300 310 320
330 340 350 360 370 380 390 400
410 420 430 440 450 460 470 480

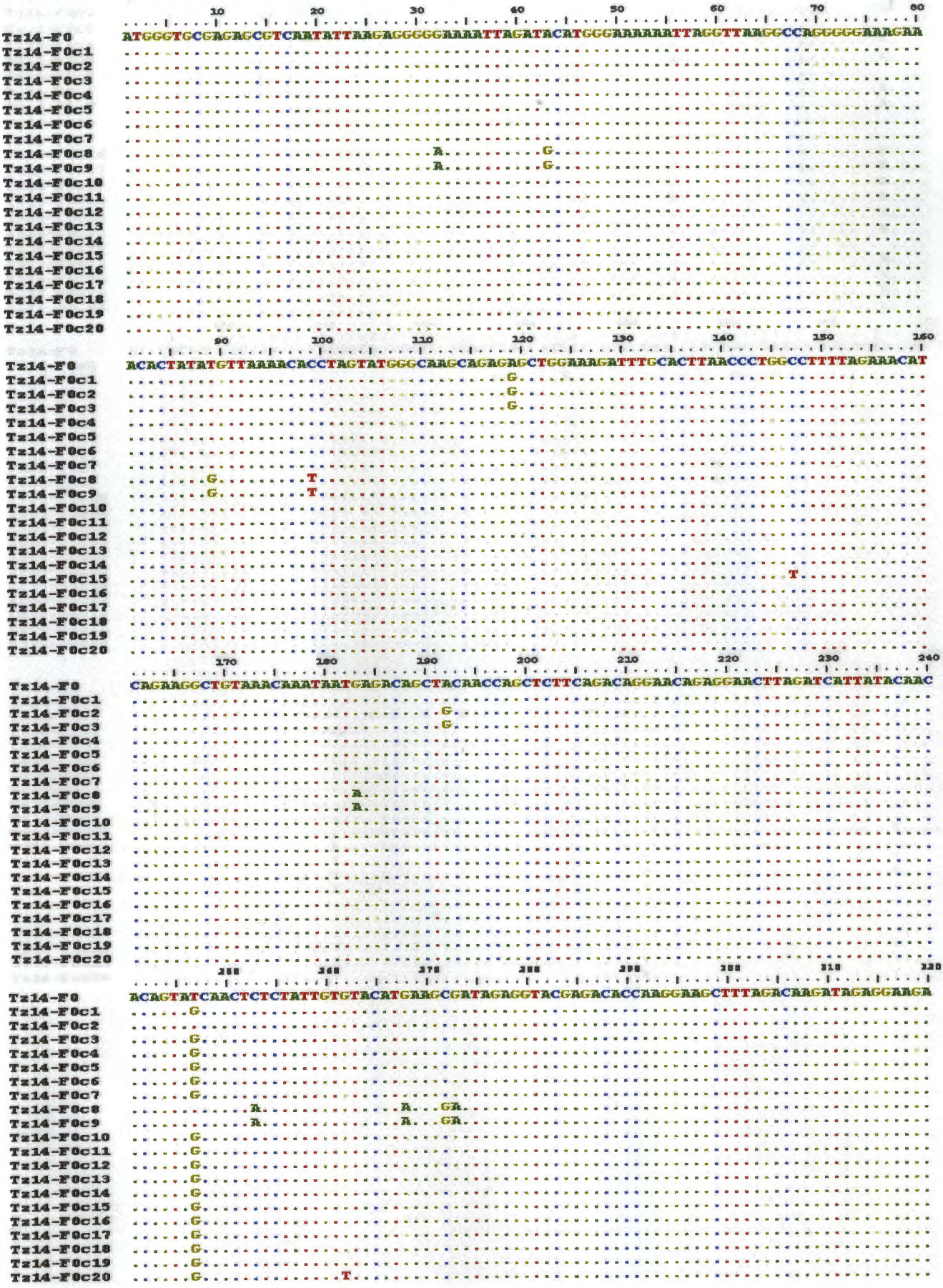


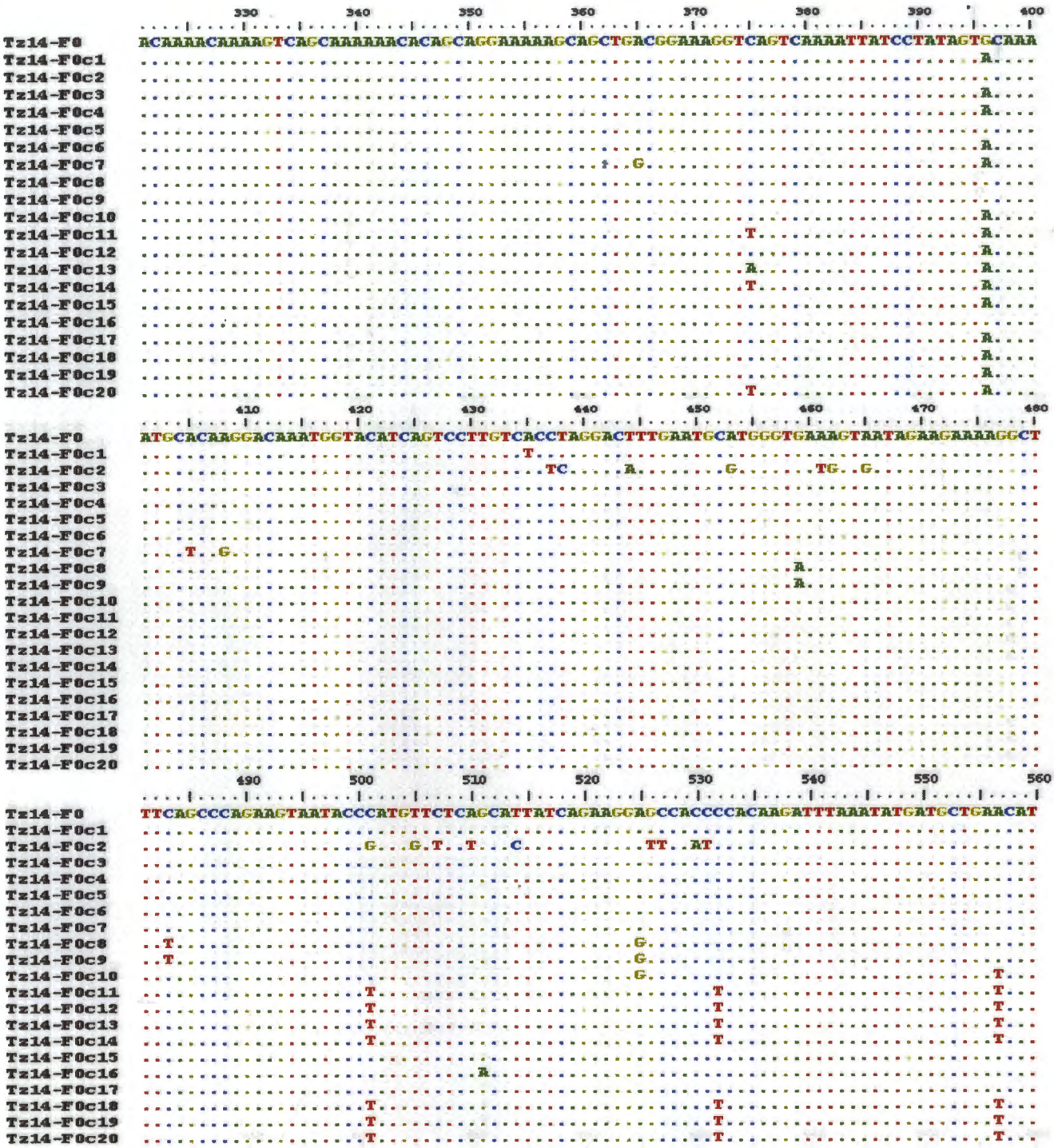
APENDIX G: Tz 14 FULL LENGTH GAG CLONE SEQUNCES FROM TIME POINT ZERO AND 21 MONTHS OF FOLLOW UP

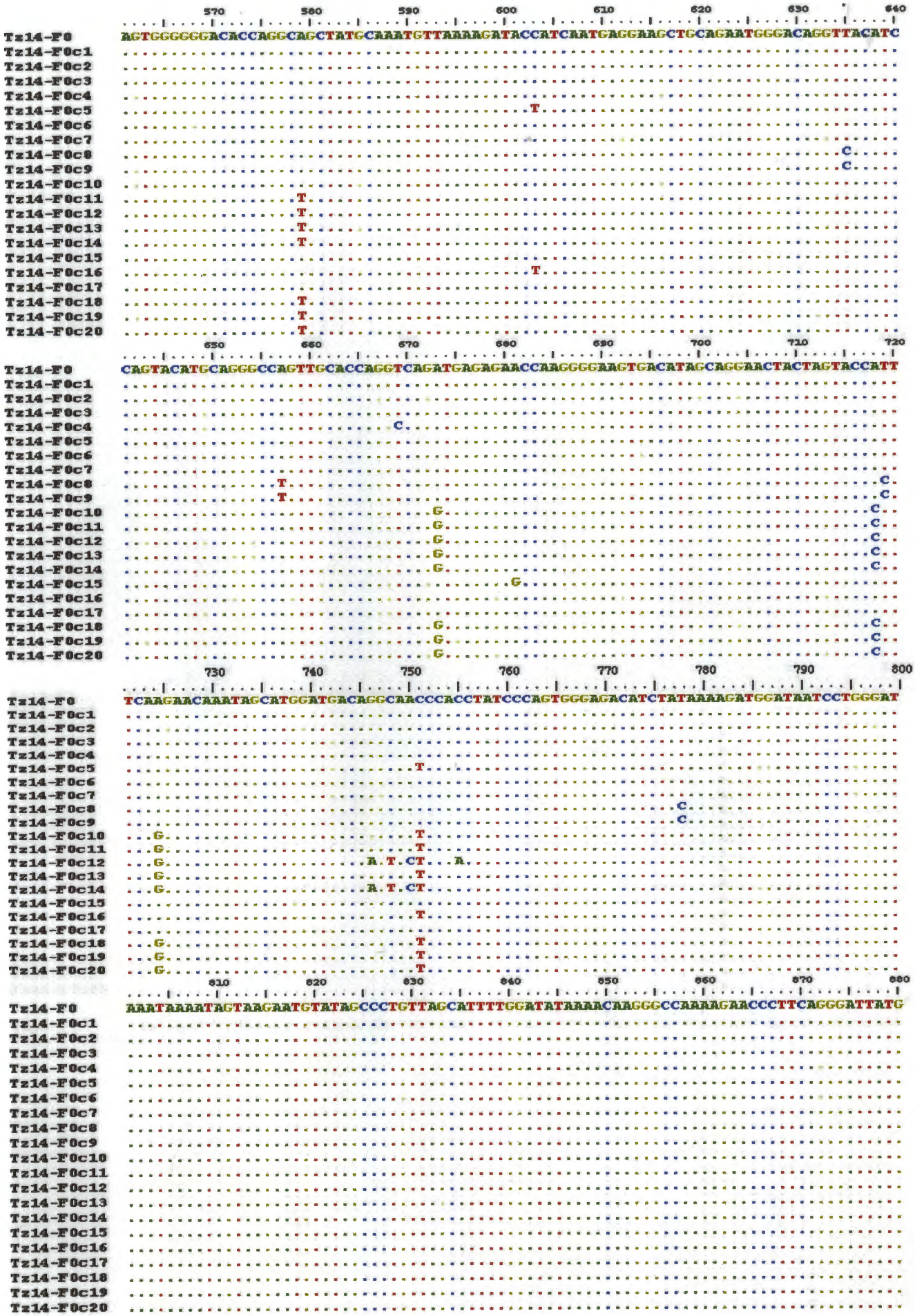
Table 1: *Gag* sequencing primers.

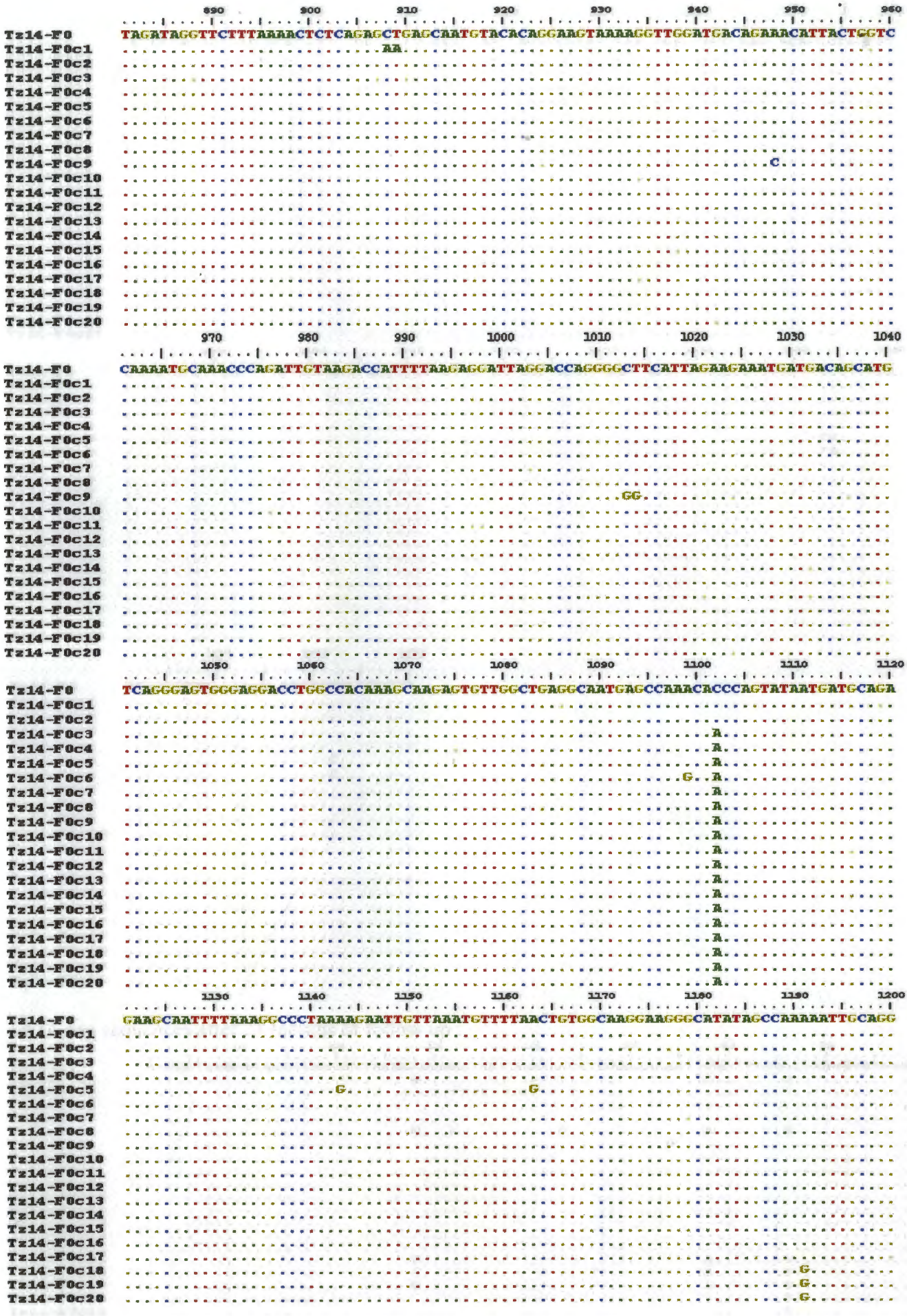
Primer name	Primer sequence	Position (HXB2)
Gag AF *(Forward Primer)*	gag AF (5'-CTCTCGACGCAGGACTCGGCTT-3')	678 - 704
Gag AR *(Reverse Primer)*	gag AR (5'-ACATGGGTCACTTCTGGGCT-3')	1283 - 1309
Gag BF *(Forward Primer)*	gag BF (5'-CCATATCACCTAGAACTTTGAAT-3')	1222 - 1249
Gag BR *(Reverse Primer)*	gag BR (5'-CTCCCTGACATGCTGTCATCAT-3')	1828 - 1849
Gag CF *(Forward Primer)*	gag CF (5'-CCTTGTTGGTCCAAAATGCGA-3')	1746 - 1771
Gag CR (Reverse Primer)	gag CR (5'-TCTTCTAATACTGTATCATCTGC-3')	2337 - 2364

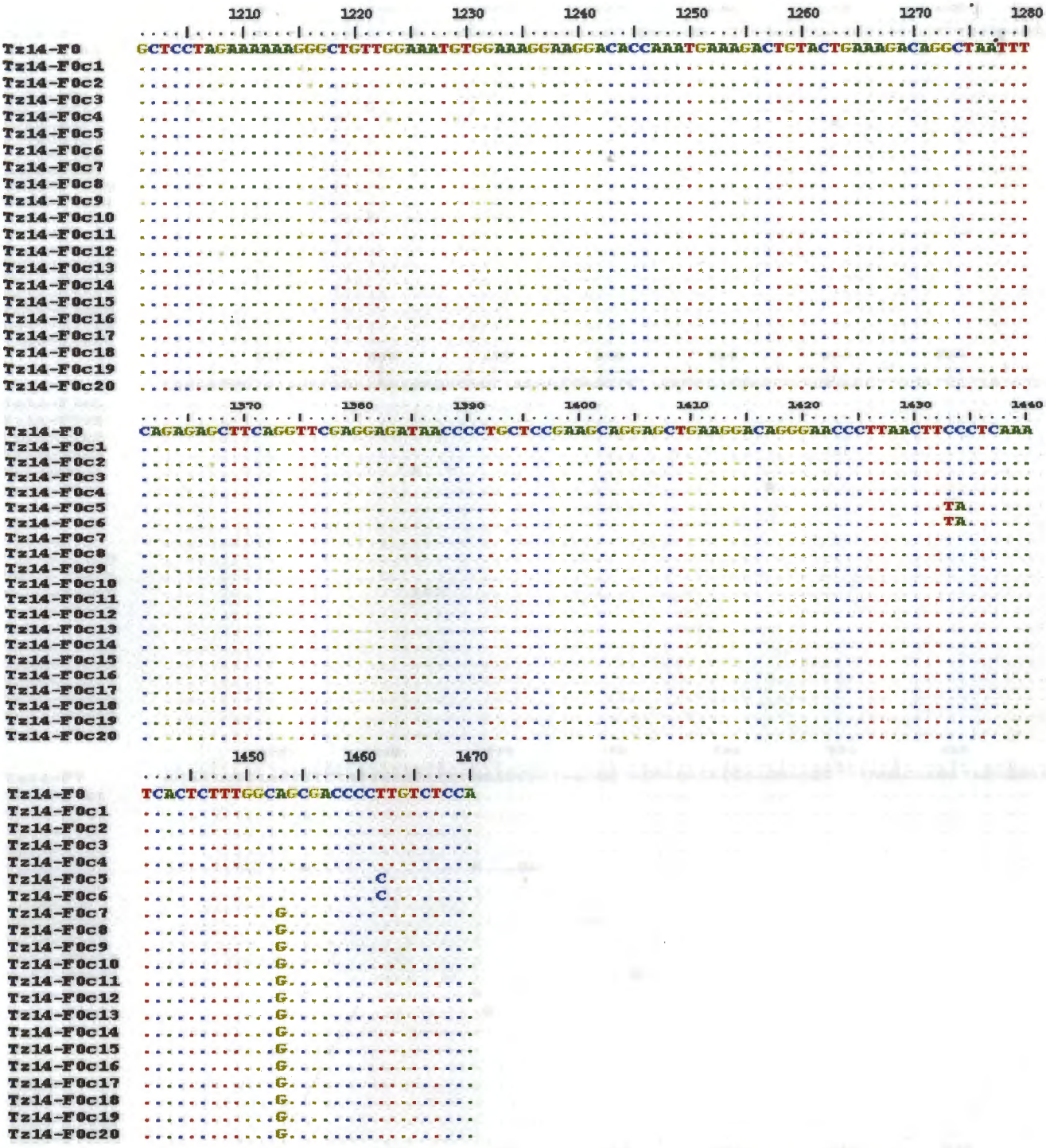
Tz 14 *gag* sequences from the first time point :



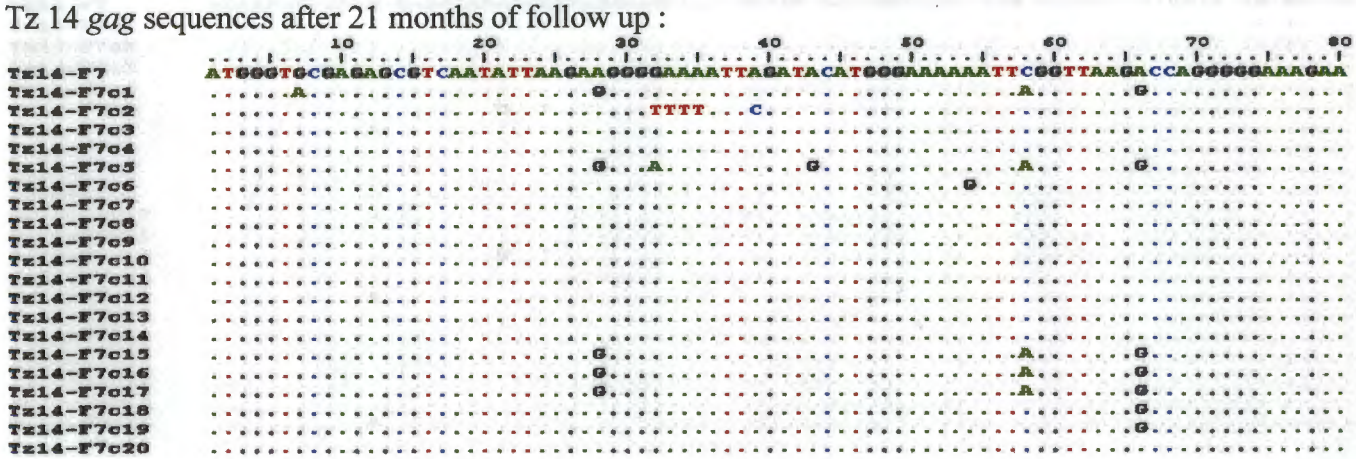

330 340 350 360 370 380 390 400
Tz14-F0
Tz14-F0c1
Tz14-F0c2
Tz14-F0c3
Tz14-F0c4
Tz14-F0c5
Tz14-F0c6
Tz14-F0c7
Tz14-F0c8
Tz14-F0c9
Tz14-F0c10
Tz14-F0c11
Tz14-F0c12
Tz14-F0c13
Tz14-F0c14
Tz14-F0c15
Tz14-F0c16
Tz14-F0c17
Tz14-F0c18
Tz14-F0c19
Tz14-F0c20
410 420 430 440 450 460 470 480
Tz14-F0
Tz14-F0c1
Tz14-F0c2
Tz14-F0c3
Tz14-F0c4
Tz14-F0c5
Tz14-F0c6
Tz14-F0c7
Tz14-F0c8
Tz14-F0c9
Tz14-F0c10
Tz14-F0c11
Tz14-F0c12
Tz14-F0c13
Tz14-F0c14
Tz14-F0c15
Tz14-F0c16
Tz14-F0c17
Tz14-F0c18
Tz14-F0c19
Tz14-F0c20
490 500 510 520 530 540 550 560
Tz14-F0
Tz14-F0c1
Tz14-F0c2
Tz14-F0c3
Tz14-F0c4
Tz14-F0c5
Tz14-F0c6
Tz14-F0c7
Tz14-F0c8
Tz14-F0c9
Tz14-F0c10
Tz14-F0c11
Tz14-F0c12
Tz14-F0c13
Tz14-F0c14
Tz14-F0c15
Tz14-F0c16
Tz14-F0c17
Tz14-F0c18
Tz14-F0c19
Tz14-F0c20

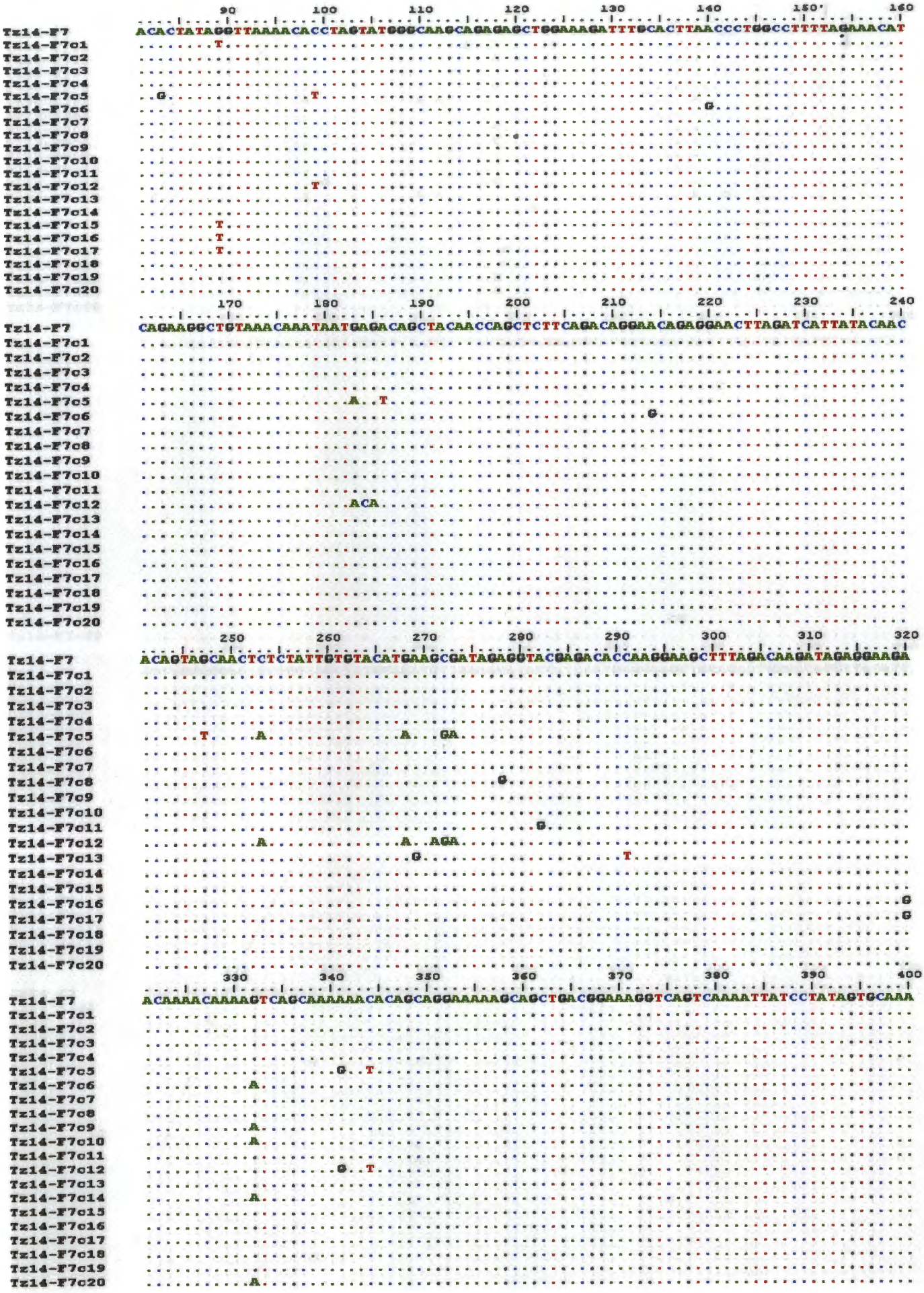

570 580 590 600 610 620 630 640
Tz14-F0 AGTGGGGGGGACACCAGGCAGCTATGCAAATGTTAAAAGATACCATCAATGAGGAAGCTGCAGAATGGGACAGGTTACATC
Tz14-F0c1
Tz14-F0c2
Tz14-F0c3
Tz14-F0c4
Tz14-F0c5
Tz14-F0c6
Tz14-F0c7
Tz14-F0c8
Tz14-F0c9
Tz14-F0c10
Tz14-F0c11
Tz14-F0c12
Tz14-F0c13
Tz14-F0c14
Tz14-F0c15
Tz14-F0c16
Tz14-F0c17
Tz14-F0c18
Tz14-F0c19
Tz14-F0c20
650 660 670 680 690 700 710 720
Tz14-F0 CAGTACATGCAGGGCCAGTTGCACCAGGTCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGAACTACTAGTACCATT
Tz14-F0c1
Tz14-F0c2
Tz14-F0c3
Tz14-F0c4
Tz14-F0c5
Tz14-F0c6
Tz14-F0c7
Tz14-F0c8
Tz14-F0c9
Tz14-F0c10
Tz14-F0c11
Tz14-F0c12
Tz14-F0c13
Tz14-F0c14
Tz14-F0c15
Tz14-F0c16
Tz14-F0c17
Tz14-F0c18
Tz14-F0c19
Tz14-F0c20
730 740 750 760 770 780 790 800
Tz14-F0 TCAAGAACAAATAGCATGGATGACAGGCAACCCACCTATCCCAGTGGGAGACATCTATAAAAGATGGATAATCCTGGGAT
Tz14-F0c1
Tz14-F0c2
Tz14-F0c3
Tz14-F0c4
Tz14-F0c5
Tz14-F0c6
Tz14-F0c7
Tz14-F0c8
Tz14-F0c9
Tz14-F0c10
Tz14-F0c11
Tz14-F0c12
Tz14-F0c13
Tz14-F0c14
Tz14-F0c15
Tz14-F0c16
Tz14-F0c17
Tz14-F0c18
Tz14-F0c19
Tz14-F0c20
810 820 830 840 850 860 870 880
Tz14-F0 AAATAAAATAGTAAGAATGTATAGCCCTGTTAGCATTTTGGATATAAAACAAGGGCCAAAAGAACCCTTCAGGGATTATG
Tz14-F0c1
Tz14-F0c2
Tz14-F0c3
Tz14-F0c4
Tz14-F0c5
Tz14-F0c6
Tz14-F0c7
Tz14-F0c8
Tz14-F0c9
Tz14-F0c10
Tz14-F0c11
Tz14-F0c12
Tz14-F0c13
Tz14-F0c14
Tz14-F0c15
Tz14-F0c16
Tz14-F0c17
Tz14-F0c18
Tz14-F0c19
Tz14-F0c20

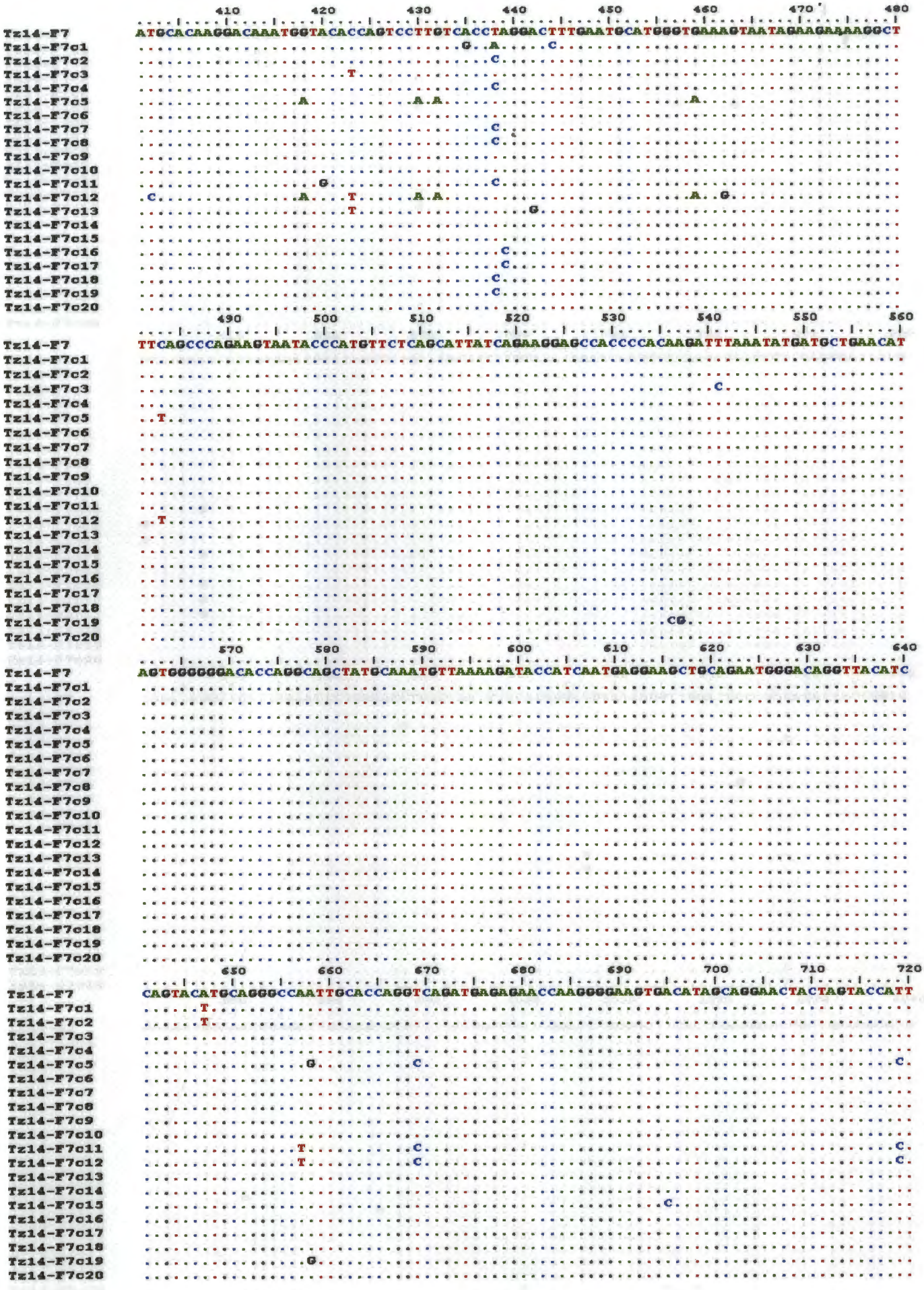

890 900 910 920 930 940 950 960
Tz14-F0 TAGATAGGTTCTTTAAAACTCTCAGAGCTGAGCAATGTACACAGGAAGTAAAAGGTTGGATGACAGAAACATTACTGGTC
Tz14-F0c1 AA
Tz14-F0c9 C
970 980 990 1000 1010 1020 1030 1040
Tz14-F0 CAAAATGCAAACCCAGATTGTAAGACCATTTTAAGAGGATTAGGACCAGGGGCTTCATTAGAAGAAATGATGACAGCATG
Tz14-F0c9 GG
1050 1060 1070 1080 1090 1100 1110 1120
Tz14-F0 TCAGGGAGTGGGAGGACCTGGCCACAAAGCAAGAGTGTTGGCTGAGGCAATGAGCCAAACACCCAGTATAATGATGCAGA
Tz14-F0c3 A
Tz14-F0c4 A
Tz14-F0c5 A
Tz14-F0c6 G A
Tz14-F0c7 A
Tz14-F0c8 A
Tz14-F0c9 A
Tz14-F0c10 A
Tz14-F0c11 A
Tz14-F0c12 A
Tz14-F0c13 A
Tz14-F0c14 A
Tz14-F0c15 A
Tz14-F0c16 A
Tz14-F0c17 A
Tz14-F0c18 A
Tz14-F0c19 A
Tz14-F0c20 A
1130 1140 1150 1160 1170 1180 1190 1200
Tz14-F0 GAAGCAATTTTAAAGGCCCTAAAAGAATTGTTAAATGTTTTAACTGTGGCAAGGAAGGGCATATAGCCAAAAATTGCAGG
Tz14-F0c5 G G
Tz14-F0c18 G
Tz14-F0c19 G
Tz14-F0c20 G

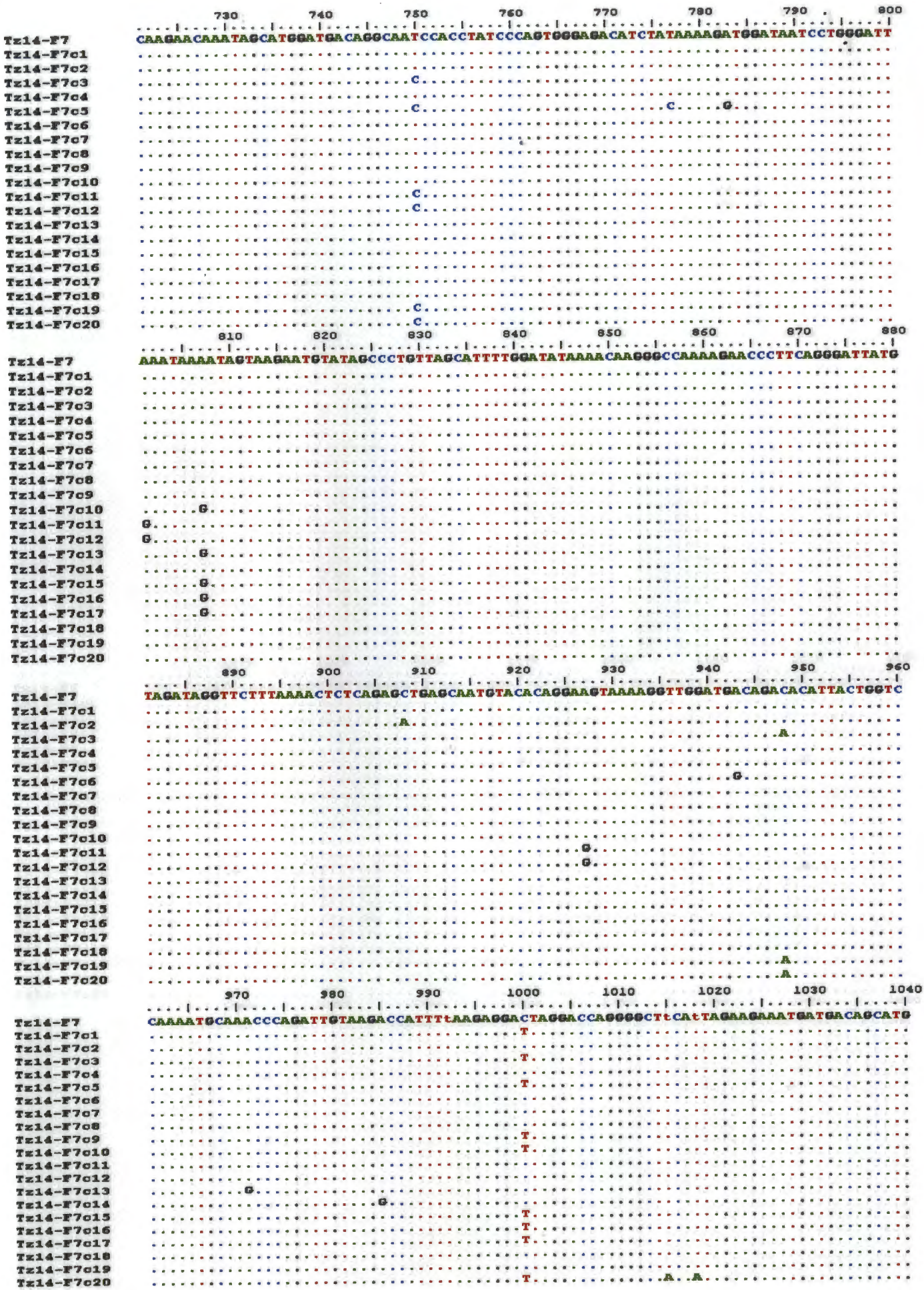


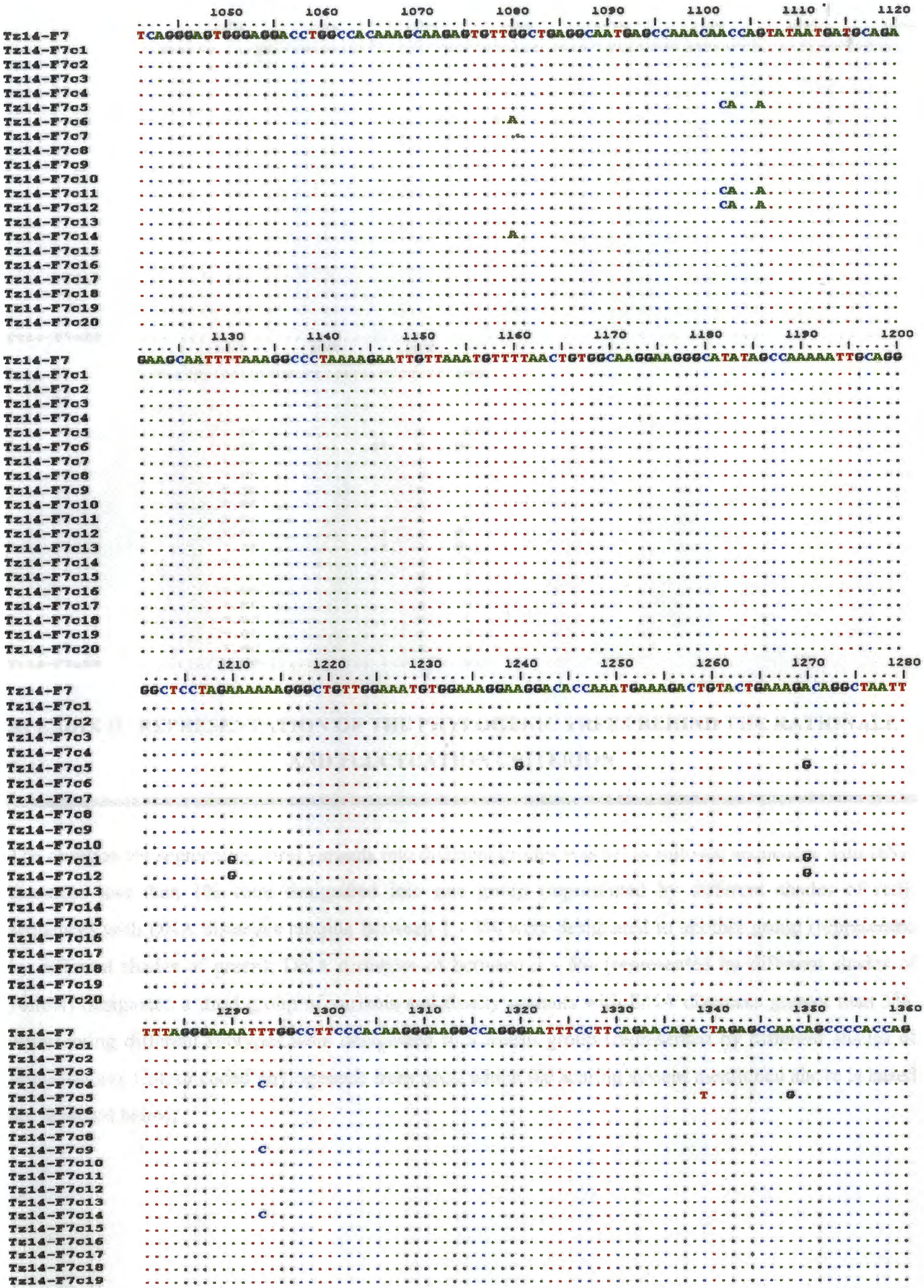
Tz 14 *gag* sequences after 21 months of follow up :




Tz14-F7
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20

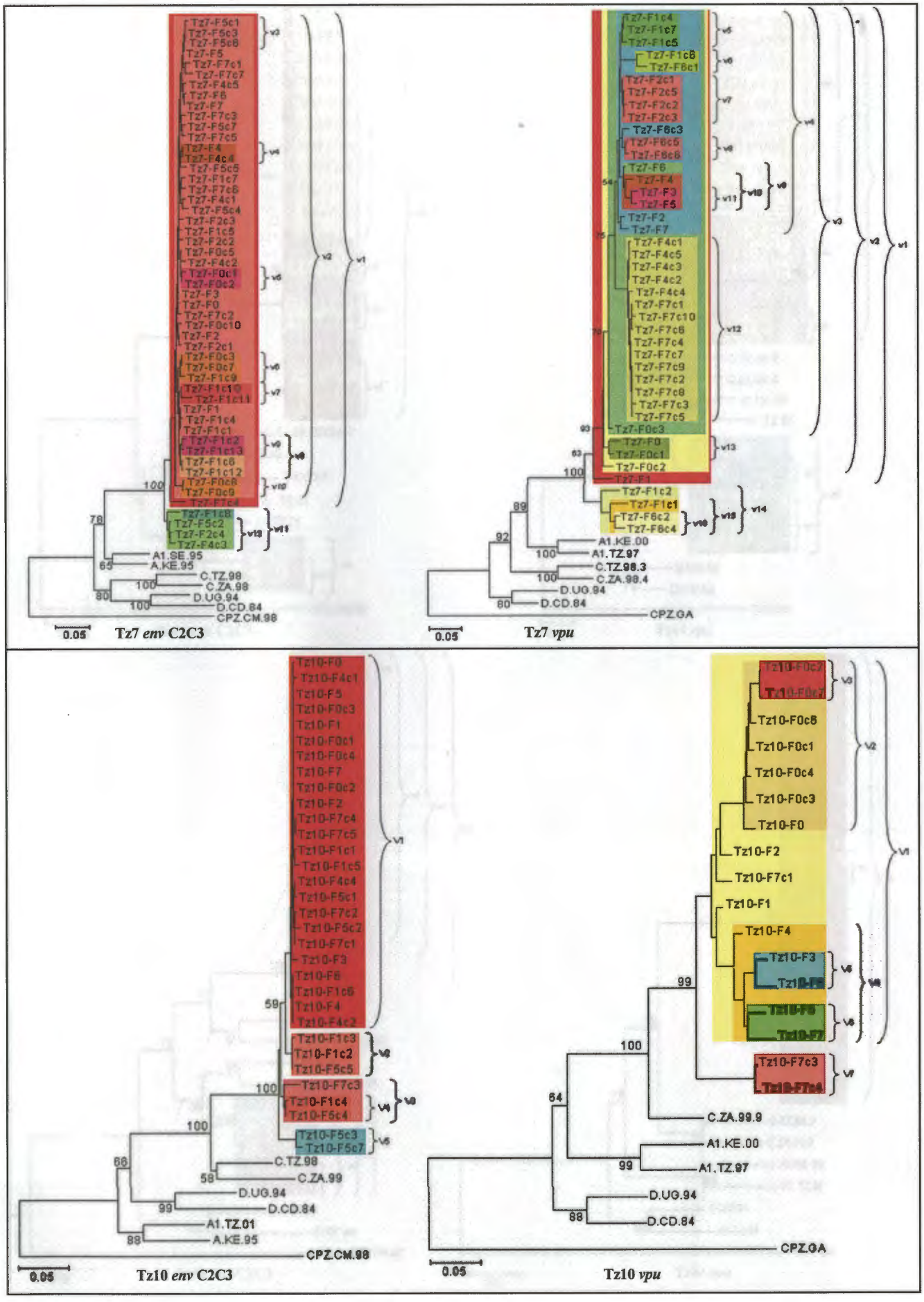

410 420 430 440 450 460 470 480
Tz14-F7 ATGCACAAGGACAAATGGTACACCAGTCCTTGTCACCTAGGACTTTGAATGCATGGGTGAAAGTAATAGAAGAAAAGGCT
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20
490 500 510 520 530 540 550 560
Tz14-F7 TTCAGCCCAGAAGTAATACCCATGTTCTCAGCATTATCAGAAGGAGCCACCCCACAAGATTTAAATATGATGCTGAACAT
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20
570 580 590 600 610 620 630 640
Tz14-F7 AGTGGGGGGACACCAGGCAGCTATGCAAATGTTAAAAGATACCATCAATGAGGAAGCTGCAGAATGGGACAGGTTACATC
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20
650 660 670 680 690 700 710 720
Tz14-F7 CAGTACATGCAGGGCCAATTGCACCAGGTCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGAACTACTAGTACCATT
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20

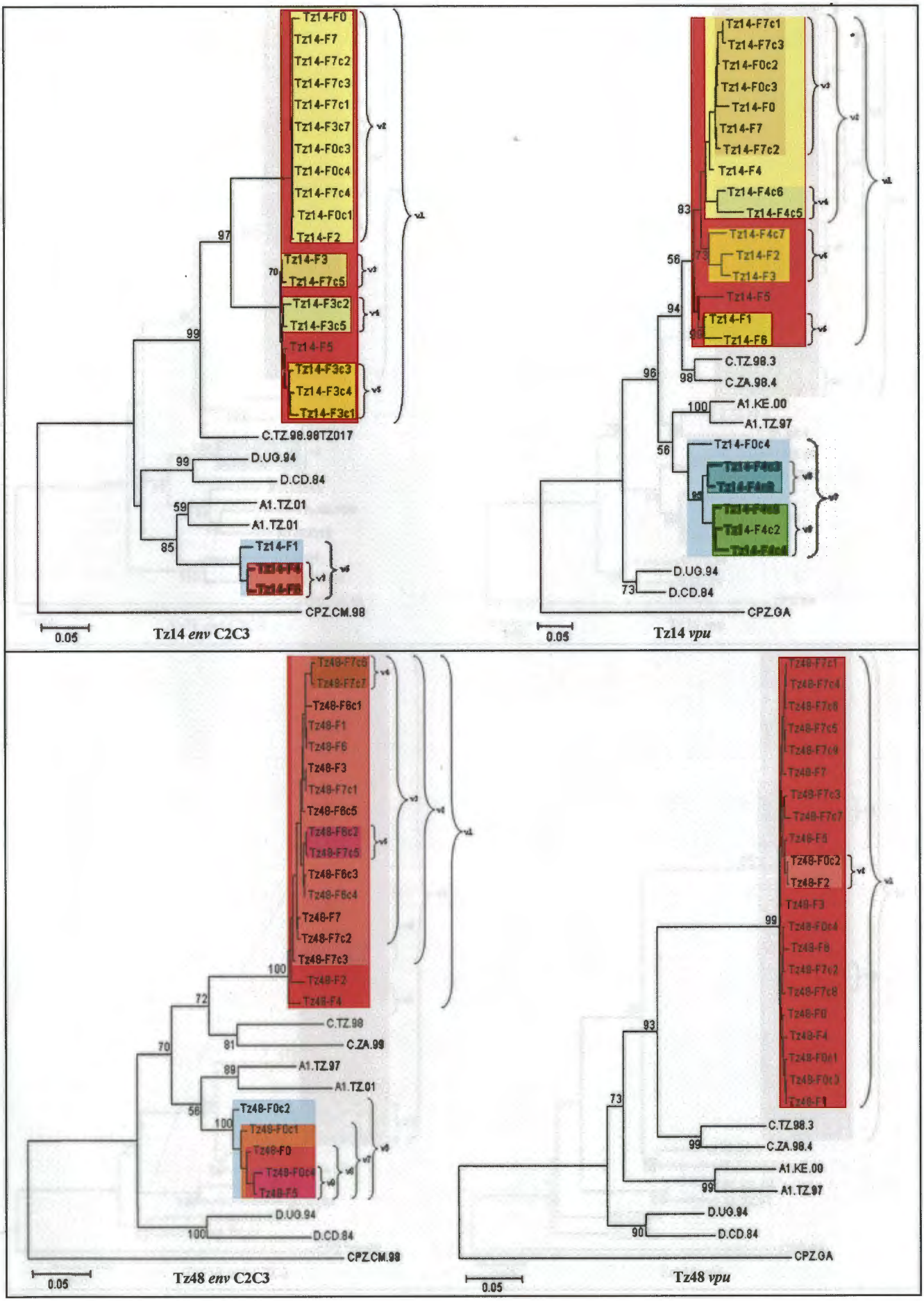

730 740 750 760 770 780 790 800
Tz14-F7 CAAGAACAAATAGCATGGATGACAGGCAATCCACCTATCCCAGTGGGAGACATCTATAAAAGATGGATAATCCTGGGATT
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20
810 820 830 840 850 860 870 880
Tz14-F7 AAATAAAATAGTAAGAATGTATAGCCCTGTTAGCATTTTGGATATAAAACAAGGGCCAAAAGAACCCTTCAGGGATTATG
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20
890 900 910 920 930 940 950 960
Tz14-F7 TAGATAGGTTCTTTAAAACTCTCAGAGCTGAGCAATGTACACAGGAAGTAAAAGGTTGGATGACAGACACATTACTGGTC
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20
970 980 990 1000 1010 1020 1030 1040
Tz14-F7 CAAAATGCAAACCCAGATTGTAAGACCATTTtAAGAGGACTAGGACCAGGGGCTtCAtTAGAAGAAATGATGACAGCATG
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20

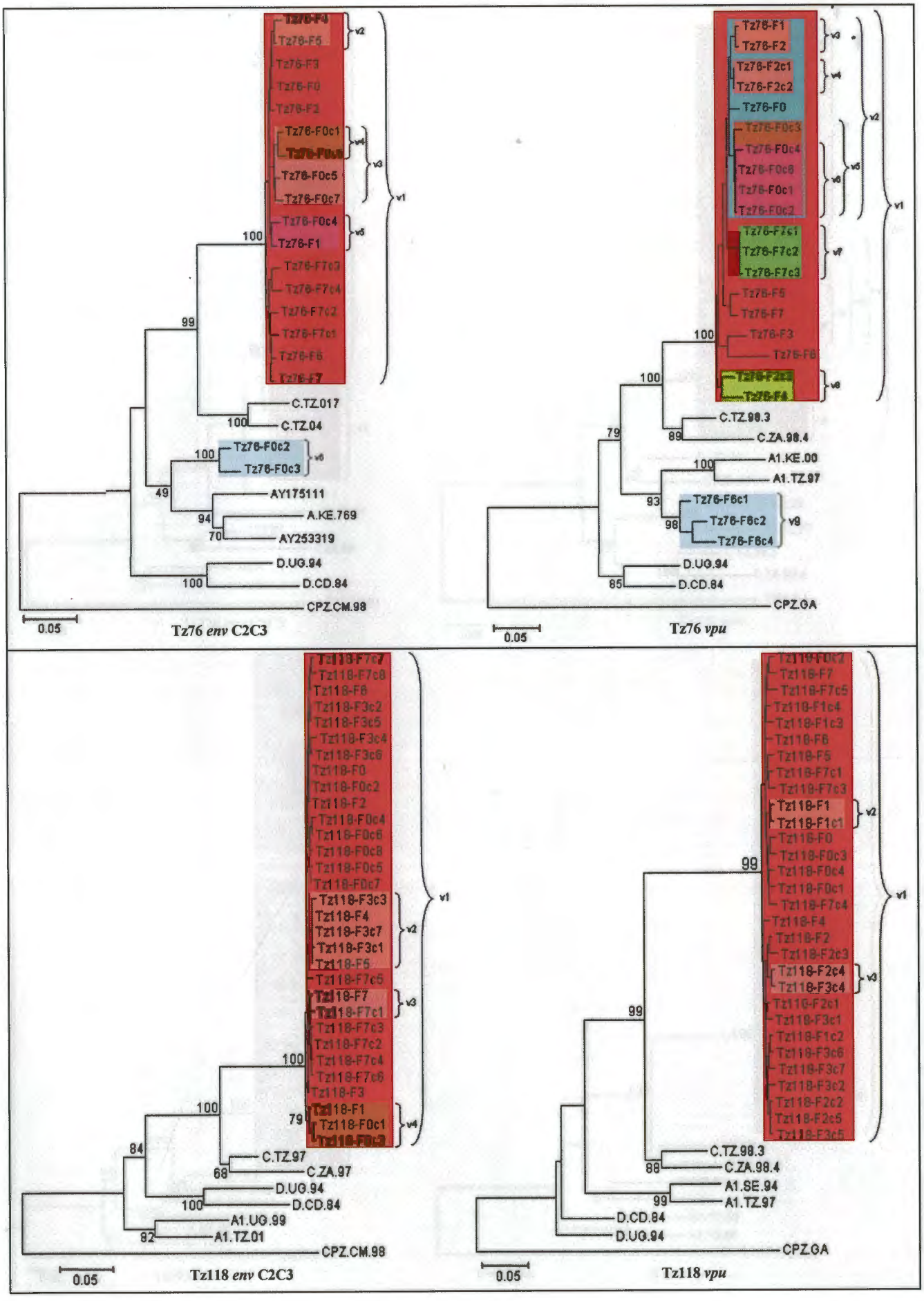

1050 1060 1070 1080 1090 1100 1110 1120
Tz14-F7 TCAGGGAGTGGGAGGACCTGGCCACAAAGCAAGAGTGTTGGCTGAGGCAATGAGCCAAACAACCAGTATAATGATGCAGA
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20
1130 1140 1150 1160 1170 1180 1190 1200
Tz14-F7 GAAGCAATTTTAAAGGCCCTAAAAGAATTGTTAAATGTTTTAACTGTGGCAAGGAAGGCATATAGCCAAAAATTGCAGG
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20
1210 1220 1230 1240 1250 1260 1270 1280
Tz14-F7 GGCTCCTAGAAAAAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGACTGTACTGAAAGACAGGCTAATT
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20
1290 1300 1310 1320 1330 1340 1350 1360
Tz14-F7 TTTAGGGAAAATTTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTCCTTCAGAACAGACTAGAGCCAACAGCCCCACCAG
Tz14-F7c1
Tz14-F7c2
Tz14-F7c3
Tz14-F7c4
Tz14-F7c5
Tz14-F7c6
Tz14-F7c7
Tz14-F7c8
Tz14-F7c9
Tz14-F7c10
Tz14-F7c11
Tz14-F7c12
Tz14-F7c13
Tz14-F7c14
Tz14-F7c15
Tz14-F7c16
Tz14-F7c17
Tz14-F7c18
Tz14-F7c19
Tz14-F7c20

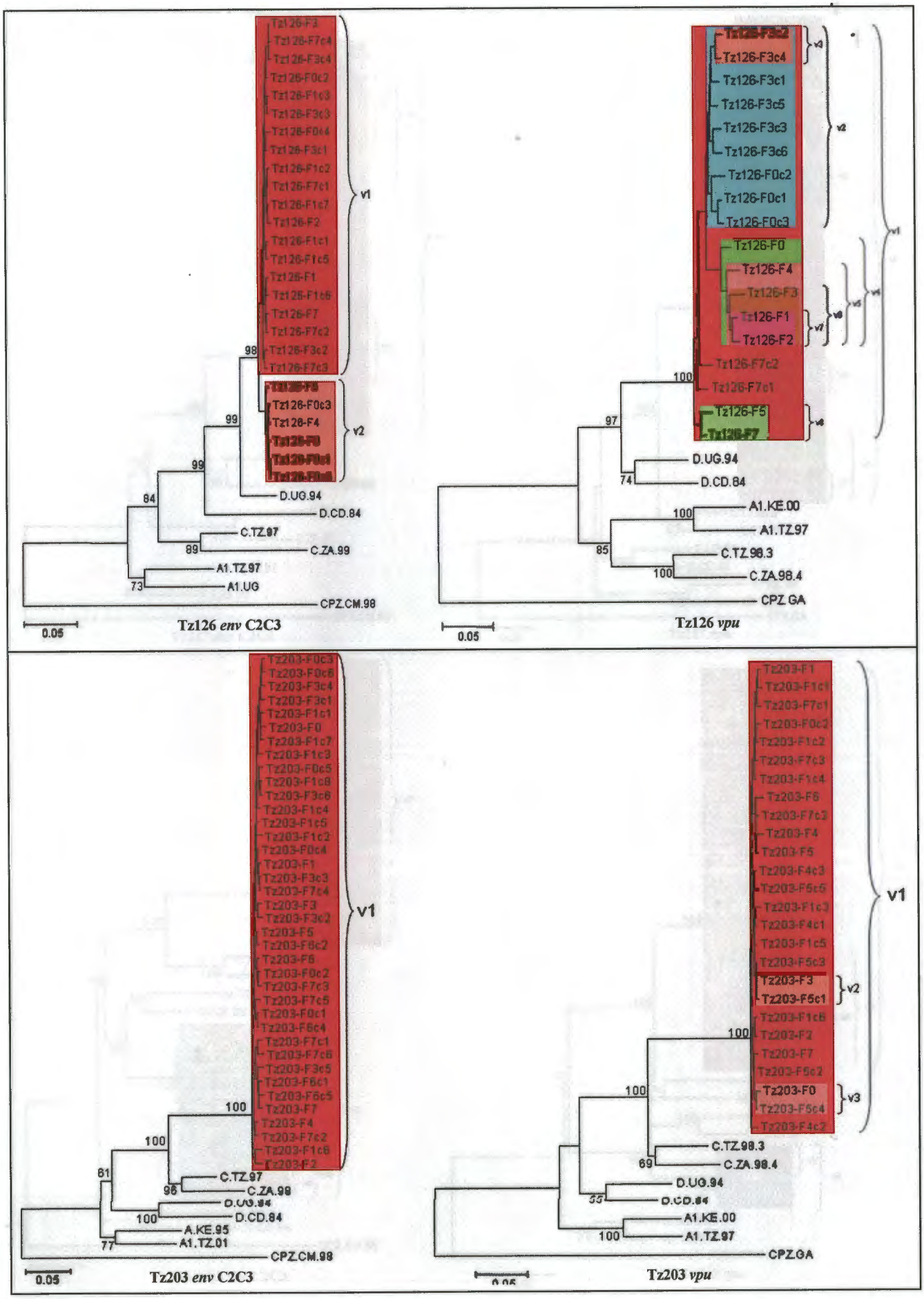
APENDIX H: REPRESENTATION OF THE PHYLOGENIC TREES BEHIND THE RATIONALE AND FLUCTUATION CRITERION

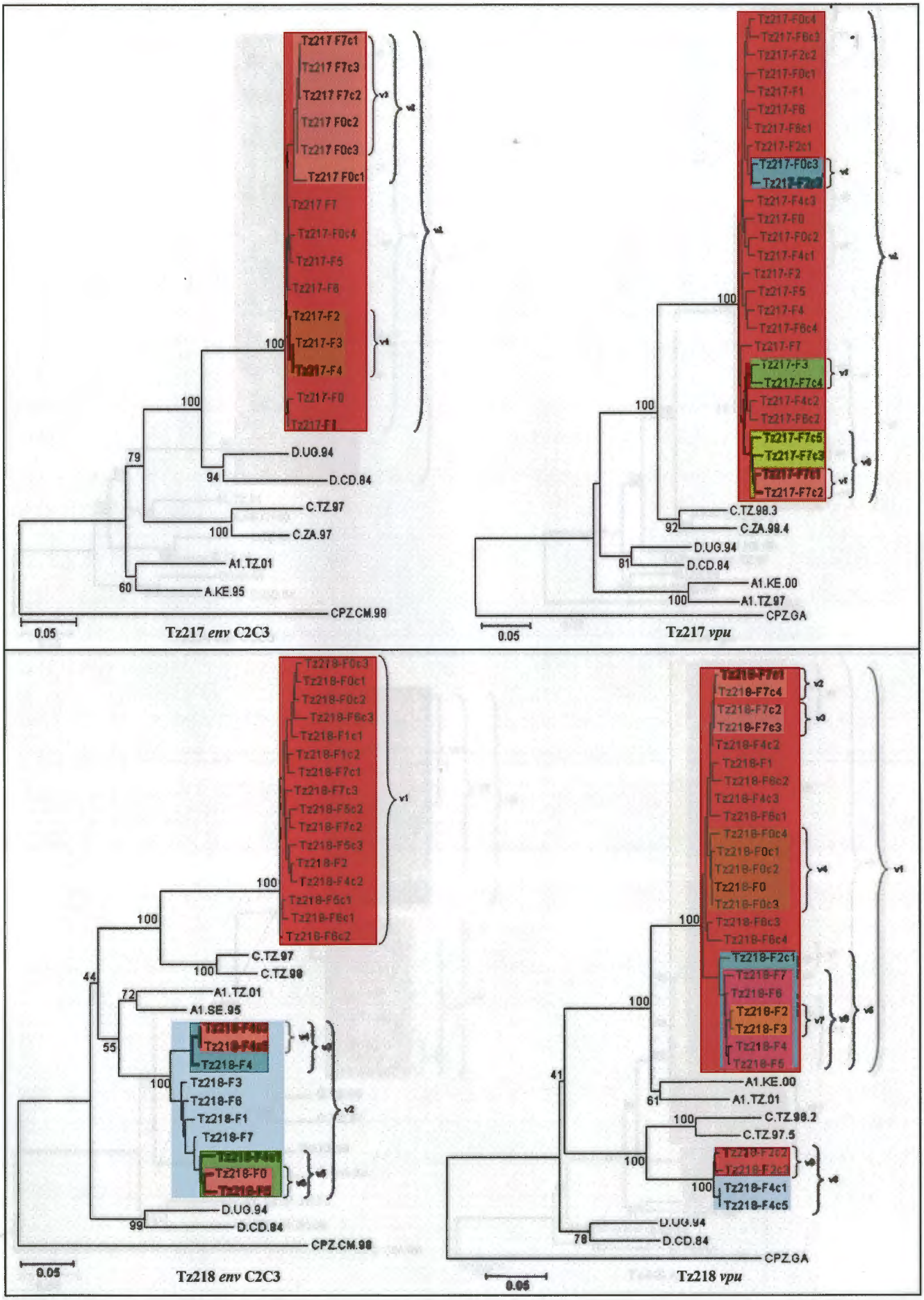
The criterion for segmenting viral variants into different groups was set as follows; sequences with DNA distances less than 1% were designated into one group (represented by different shades of red); sequences with DNA distances ranging between 1 - 3% were designated to another group (represented by different shades of green); DNA distances of between 3 - 5% (represented by different shades of yellow) designated a third group of variants and finally variants with DNA distances greater than 5%, representing different subtypes were designated to a fourth group (represented by different shades of blue colours). Colour coded phylogenetic trees upon which the scoring system mentioned above is based are depicted below.

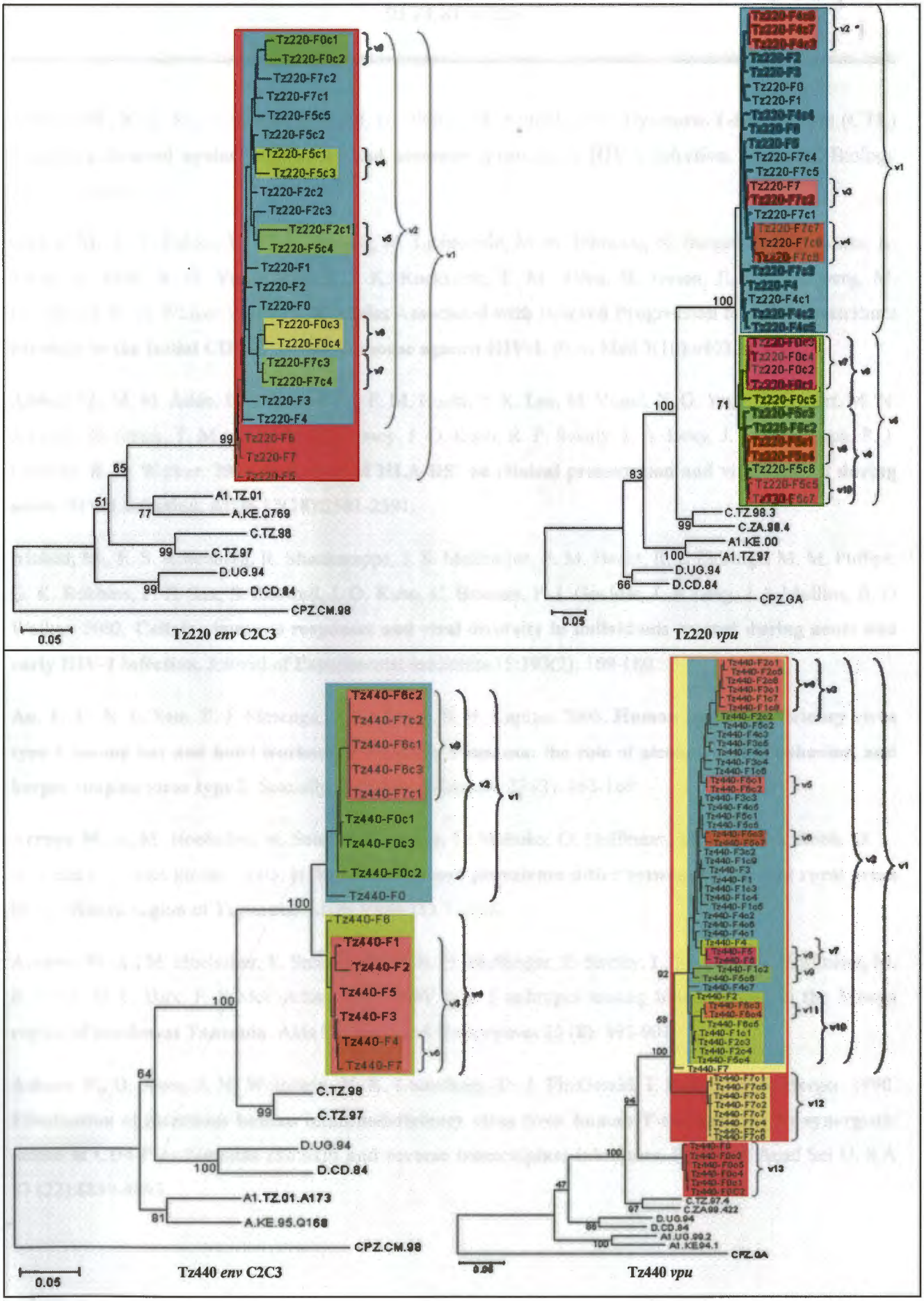

Tz7 env C2C3
Tz7 vpu
Tz10 env C2C3
Tz10 vpu
0.05
A1.SE.95
A.KE.95
C.TZ.98
C.ZA.98
D.UG.94
D.CD.84
CPZ.CM.98
A1.KE.00
A1.TZ.97
C.TZ.98.3
C.ZA.98.4
CPZ.GA
A1.TZ.01
C.ZA.99
C.ZA.99.9


Tz14-F0
Tz14-F7
Tz14-F7c2
Tz14-F7c3
Tz14-F7c1
Tz14-F3c7
Tz14-F0c3
Tz14-F0c4
Tz14-F7c4
Tz14-F0c1
Tz14-F2
Tz14-F3
Tz14-F7c5
Tz14-F3c2
Tz14-F3c5
Tz14-F5
Tz14-F3c3
Tz14-F3c4
Tz14-F3c1
C.TZ.98.98TZ017
D.UG.94
D.CD.84
A1.TZ.01
A1.TZ.01
Tz14-F1
CPZ.CM.98
0.05
Tz14 env C2C3
Tz14-F7c1
Tz14-F7c3
Tz14-F0c2
Tz14-F0c3
Tz14-F0
Tz14-F7
Tz14-F7c2
Tz14-F4
Tz14-F4c6
Tz14-F4c5
Tz14-F4c7
Tz14-F2
Tz14-F3
Tz14-F5
Tz14-F1
Tz14-F6
C.TZ.98.3
C.ZA.98.4
A1.KE.00
A1.TZ.97
Tz14-F0c4
Tz14-F4c2
D.UG.94
D.CD.84
CPZ.GA
0.05
Tz14 vpu
Tz48-F6c1
Tz48-F1
Tz48-F6
Tz48-F3
Tz48-F7c1
Tz48-F6c5
Tz48-F6c3
Tz48-F6c4
Tz48-F7
Tz48-F7c2
Tz48-F7c3
Tz48-F2
Tz48-F4
C.TZ.98
C.ZA.99
A1.TZ.97
A1.TZ.01
Tz48-F0c2
Tz48-F0c1
Tz48-F0
Tz48-F5
D.UG.94
D.CD.84
CPZ.CM.98
0.05
Tz48 env C2C3
Tz48-F0c2
Tz48-F2
Tz48-F3
Tz48-F4
C.TZ.98.3
C.ZA.98.4
A1.KE.00
A1.TZ.97
D.UG.94
D.CD.84
CPZ.GA
0.05
Tz48 vpu


Tz76-F4
Tz76-F5
Tz76-F3
Tz76-F0
Tz76-F2
Tz76-F0c1
Tz76-F0c6
Tz76-F0c5
Tz76-F0c7
Tz76-F0c4
Tz76-F1
Tz76-F7c3
Tz76-F7c4
Tz76-F7c2
Tz76-F7c1
Tz76-F6
Tz76-F7
v1
v2
v3
v4
v5
v6
100
99
C.TZ.017
100
C.TZ.04
100
Tz76-F0c2
Tz76-F0c3
v6
49
AY175111
94
A.KE.769
70
AY253319
D.UG.94
100
D.CD.84
CPZ.CM.98
0.05
Tz76 env C2C3
Tz76-F1
Tz76-F2
Tz76-F2c1
Tz76-F2c2
Tz76-F0
Tz76-F0c3
Tz76-F0c4
Tz76-F0c6
Tz76-F0c1
Tz76-F0c2
Tz76-F7c1
Tz76-F7c2
Tz76-F7c3
Tz76-F5
Tz76-F7
Tz76-F3
Tz76-F6
v3
v4
v2
v5
v6
v1
v7
100
100
C.TZ.98.3
79
89
C.ZA.98.4
A1.KE.00
100
A1.TZ.97
93
Tz76-F6c1
98
Tz76-F6c2
Tz76-F6c4
v8
D.UG.94
85
D.CD.84
CPZ.GA
0.05
Tz76 vpu
Tz118-F7c7
Tz118-F7c8
Tz118-F6
Tz118-F3c2
Tz118-F3c5
Tz118-F3c4
Tz118-F3c6
Tz118-F0
Tz118-F0c2
Tz118-F2
Tz118-F0c4
Tz118-F0c6
Tz118-F0c8
Tz118-F0c5
Tz118-F0c7
Tz118-F3c3
Tz118-F4
Tz118-F3c7
Tz118-F3c1
Tz118-F5
Tz118-F7c5
Tz118-F7
Tz118-F7c1
Tz118-F7c3
Tz118-F7c2
Tz118-F7c4
Tz118-F7c6
Tz118-F3
Tz118-F1
Tz118-F0c1
v1
v2
v3
v4
100
100
79
84
C.TZ.97
68
C.ZA.97
D.UG.94
100
D.CD.84
A1.UG.99
82
A1.TZ.01
CPZ.CM.98
0.05
Tz118 env C2C3
Tz118-F0c2
Tz118-F7
Tz118-F7c5
Tz118-F1c4
Tz118-F1c3
Tz118-F6
Tz118-F5
Tz118-F7c1
Tz118-F7c3
Tz118-F1
Tz118-F1c1
Tz118-F0
Tz118-F0c3
Tz118-F0c4
Tz118-F0c1
Tz118-F7c4
Tz118-F4
Tz118-F2
Tz118-F2c3
Tz118-F2c4
Tz118-F3c4
Tz118-F2c1
Tz118-F3c1
Tz118-F1c2
Tz118-F3c6
Tz118-F3c7
Tz118-F3c2
Tz118-F2c2
Tz118-F2c5
Tz118-F3c5
v1
v2
v3
99
99
C.TZ.98.3
88
C.ZA.98.4
A1.SE.94
99
A1.TZ.97
D.CD.84
D.UG.94
CPZ.GA
0.05
Tz118 vpu


Tz126-F3
Tz126-F7c4
Tz126-F3c4
Tz126-F0c2
Tz126-F1c3
Tz126-F3c3
Tz126-F0c4
Tz126-F3c1
Tz126-F1c2
Tz126-F7c1
Tz126-F1c7
Tz126-F2
Tz126-F1c1
Tz126-F1c5
Tz126-F1
Tz126-F1c6
Tz126-F7
Tz126-F7c2
Tz126-F3c2
Tz126-F7c3
v1
Tz126-F0c3
Tz126-F4
v2
98
99
99
84
89
73
D.UG.94
D.CD.84
C.TZ.97
C.ZA.99
A1.TZ.97
A1.UG
CPZ.CM.98
0.05
Tz126 env C2C3
Tz126-F3c4
Tz126-F3c1
Tz126-F3c5
Tz126-F3c3
Tz126-F3c6
Tz126-F0c2
Tz126-F0c1
Tz126-F0c3
Tz126-F0
Tz126-F4
Tz126-F3
Tz126-F1
Tz126-F2
Tz126-F7c2
Tz126-F7c1
Tz126-F5
v3
v2
v4
v5
v6
v7
v8
100
97
74
100
85
100
D.UG.94
D.CD.84
A1.KE.00
A1.TZ.97
C.TZ.98.3
C.ZA.98.4
CPZ.GA
0.05
Tz126 vpu
Tz203-F0c3
Tz203-F0c6
Tz203-F3c4
Tz203-F3c1
Tz203-F1c1
Tz203-F0
Tz203-F1c7
Tz203-F1c3
Tz203-F0c5
Tz203-F1c8
Tz203-F3c6
Tz203-F1c4
Tz203-F1c5
Tz203-F1c2
Tz203-F0c4
Tz203-F1
Tz203-F3c3
Tz203-F7c4
Tz203-F3
Tz203-F3c2
Tz203-F5
Tz203-F6c2
Tz203-F6
Tz203-F0c2
Tz203-F7c3
Tz203-F7c5
Tz203-F0c1
Tz203-F6c4
Tz203-F7c1
Tz203-F7c6
Tz203-F3c5
Tz203-F6c1
Tz203-F6c5
Tz203-F7
Tz203-F4
Tz203-F7c2
Tz203-F1c6
Tz203-F2
V1
100
100
61
95
100
77
C.TZ.97
C.ZA.99
D.UG.94
D.CD.84
A.KE.95
A1.TZ.01
CPZ.CM.98
0.05
Tz203 env C2C3
Tz203-F1
Tz203-F1c1
Tz203-F7c1
Tz203-F0c2
Tz203-F1c2
Tz203-F7c3
Tz203-F1c4
Tz203-F6
Tz203-F7c2
Tz203-F4
Tz203-F5
Tz203-F4c3
Tz203-F6c5
Tz203-F1c3
Tz203-F4c1
Tz203-F1c5
Tz203-F5c3
Tz203-F3
Tz203-F5c1
Tz203-F1c6
Tz203-F2
Tz203-F7
Tz203-F5c2
Tz203-F0
Tz203-F5c4
Tz203-F4c2
V1
v2
v3
100
100
69
55
100
C.TZ.98.3
C.ZA.98.4
D.UG.94
D.CD.84
A1.KE.00
A1.TZ.97
CPZ.GA
Tz203 vpu


Tz217 F7c1
Tz217 F7c3
Tz217 F7c2
Tz217 F0c2
Tz217 F0c3
Tz217 F0c1
Tz217 F7
Tz217-F0c4
Tz217-F5
Tz217-F6
Tz217-F2
Tz217-F3
Tz217-F4
Tz217-F0
Tz217-F1
D.UG.94
D.CD.84
C.TZ.97
C.ZA.97
A1.TZ.01
A.KE.95
CPZ.CM.98
0.05
Tz217 env C2C3
Tz217-F0c4
Tz217-F6c3
Tz217-F2c2
Tz217-F0c1
Tz217-F1
Tz217-F6
Tz217-F6c1
Tz217-F2c1
Tz217-F0c3
Tz217-F2c3
Tz217-F4c3
Tz217-F0
Tz217-F0c2
Tz217-F4c1
Tz217-F2
Tz217-F5
Tz217-F4
Tz217-F6c4
Tz217-F7
Tz217-F3
Tz217-F7c4
Tz217-F4c2
Tz217-F6c2
Tz217-F7c5
Tz217-F7c3
Tz217-F7c1
Tz217-F7c2
C.TZ.98.3
C.ZA.98.4
D.UG.94
D.CD.84
A1.KE.00
A1.TZ.97
CPZ.GA
0.05
Tz217 vpu
Tz218-F0c3
Tz218-F0c1
Tz218-F0c2
Tz218-F6c3
Tz218-F1c1
Tz218-F1c2
Tz218-F7c1
Tz218-F7c3
Tz218-F5c2
Tz218-F7c2
Tz218-F5c3
Tz218-F2
Tz218-F4c2
Tz218-F5c1
Tz218-F6c1
Tz218-F6c2
C.TZ.97
C.TZ.98
A1.TZ.01
A1.SE.95
Tz218-F4
Tz218-F3
Tz218-F6
Tz218-F1
Tz218-F7
Tz218-F0
D.UG.94
D.CD.84
CPZ.CM.98
0.05
Tz218 env C2C3
Tz218-F7c4
Tz218-F7c2
Tz218-F7c3
Tz218-F4c2
Tz218-F1
Tz218-F6c2
Tz218-F4c3
Tz218-F6c1
Tz218-F0c4
Tz218-F0c1
Tz218-F0c2
Tz218-F0
Tz218-F0c3
Tz218-F6c3
Tz218-F6c4
Tz218-F2c1
Tz218-F7
Tz218-F6
Tz218-F2
Tz218-F3
Tz218-F4
Tz218-F5
A1.KE.00
A1.TZ.01
C.TZ.98.2
C.TZ.97.5
Tz218-F4c1
Tz218-F4c5
D.UG.94
D.CD.84
CPZ.GA
0.05
Tz218 vpu


Tz220-F0c1
Tz220-F0c2
Tz220-F7c2
Tz220-F7c1
Tz220-F5c5
Tz220-F5c2
Tz220-F5c1
Tz220-F5c3
Tz220-F2c2
Tz220-F2c3
Tz220-F2c1
Tz220-F5c4
Tz220-F1
Tz220-F2
Tz220-F0
Tz220-F0c3
Tz220-F0c4
Tz220-F7c3
Tz220-F7c4
Tz220-F3
Tz220-F4
Tz220-F6
Tz220-F7
A1.TZ.01
A.KE.Q769
C.TZ.98
C.TZ.97
D.UG.94
D.CD.84
CPZ.CM.98
0.05
Tz220 env C2C3
Tz220-F4c3
Tz220-F6c7
Tz220-F4c3
Tz220-F2
Tz220-F3
Tz220-F0
Tz220-F1
Tz220-F4c4
Tz220-F8
Tz220-F6
Tz220-F7c4
Tz220-F7c5
Tz220-F7
Tz220-F7c2
Tz220-F7c1
Tz220-F7c7
Tz220-F7c0
Tz220-F7c8
Tz220-F7c9
Tz220-F4
Tz220-F4c1
Tz220-F4c2
Tz220-F4c5
Tz220-F0c3
Tz220-F0c4
Tz220-F0c2
Tz220-F0c1
Tz220-F0c5
Tz220-F5c3
Tz220-F5c2
Tz220-F5c1
Tz220-F5c4
Tz220-F5c6
Tz220-F5c5
Tz220-F5c7
C.TZ.98.3
C.ZA.98.4
A1.KE.00
A1.TZ.97
D.UG.94
D.CD.84
CPZ.GA
0.05
Tz220 vpu
Tz440-F6c2
Tz440-F7c2
Tz440-F6c1
Tz440-F6c3
Tz440-F7c1
Tz440-F0c1
Tz440-F0c3
Tz440-F0c2
Tz440-F0
Tz440-F6
Tz440-F1
Tz440-F2
Tz440-F5
Tz440-F3
Tz440-F4
Tz440-F7
C.TZ.98
C.TZ.97
D.UG.94
D.CD.84
A1.TZ.01.A173
A.KE.95.Q168
CPZ.CM.98
0.05
Tz440 env C2C3
Tz440-F2c1
Tz440-F2c6
Tz440-F2c8
Tz440-F3c1
Tz440-F1c7
Tz440-F1c8
Tz440-F2c2
Tz440-F5c2
Tz440-F4c3
Tz440-F3c5
Tz440-F4c4
Tz440-F3c4
Tz440-F1c5
Tz440-F6c1
Tz440-F6c2
Tz440-F3c3
Tz440-F4c5
Tz440-F5c1
Tz440-F5c6
Tz440-F6c3
Tz440-F6c7
Tz440-F3c2
Tz440-F1c9
Tz440-F3
Tz440-F1
Tz440-F1e3
Tz440-F1c4
Tz440-F1c6
Tz440-F4c2
Tz440-F4c6
Tz440-F4c1
Tz440-F4
Tz440-F5
Tz440-F8
Tz440-F1c2
Tz440-F5c8
Tz440-F4c7
Tz440-F2
Tz440-F8c3
Tz440-F8c4
Tz440-F8c5
Tz440-F1c1
Tz440-F2c3
Tz440-F2c4
Tz440-F5c4
Tz440-F7
Tz440-F7c1
Tz440-F7c5
Tz440-F7c3
Tz440-F7c2
Tz440-F7c7
Tz440-F7c4
Tz440-F7c6
Tz440-F0
Tz440-F0c2
Tz440-F0c5
Tz440-F0c4
Tz440-F0c1
Tz440-F0c3
C.TZ.97.4
C.ZA99.422
D.UG.94
D.CD.84
A1.UG.99.2
A1.KE94.1
CPZ.GA
Tz440 vpu

Addo, NM., X. G. Yu, E. S. Rosenberg, B. D. Walker, M. Altfeld. 2002. **Cytotoxic T-lymphocyte (CTL) responses directed against regulatory and accessory proteins in HIV-1 infection.** DNA cell Biology 21(9): 671-678.

Altfeld M., E. T. Kalike, Y. Qi, H. Streeck, M. Lichterfeld, M. N. Johnston, N. Burgett, M. E. Swartz, A. Yang, G. Alter, X. G. Yu, A. meier, J. K. Rockstroh, T. M. Allen, H. Jessen, E. S. Rosenberg, M. Carrington, B. D. Walker. 2006. **HLA Alleles Associated with Delayed Progression to AIDS Contribute Strongly to the Initial CD8(+) T Cell Response against HIV-1.** PLos Med 3(10):e403.

Altfeld M., M. M. Addo, E. S. Rosenberg, F. M. Hecht, P. K. Lee, M. Vogel, X. G. Yu, R. Draenert, M. N. Johnson, D. Strick, T. M. Allen, M. E. Feeney, J. O. Kahn, R. P. Sekaly, J. A. Levy, J. K. Rockstroh, P. J. Goulder, B. D. Walker. 2003. **Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection.** AIDS 17(18):2581-2591.

Altfeld, M., E. S. Rosenberg, R. Shankarappa, J. S. Mukherjee, F. M. Hecht, R. L Eldridge, M. M. Philips, G. K. Robbins, P. E. Sax, S. Boswell, J. O. Kahn, C. Brander, P. J. Goulder, J. A Levy, J. I. Mullins, B. D Walker. 2002. **Cellular immune responses and viral diversity in individuals treated during acute and early HIV-1 infection.** Journal of Experimental mediceine15:193(2): 169-180.

Ao, T. T., N. E. Sam, E. J. Masenga, G. R. Seage, S. H. Kapiga. 2006. **Human immunodeficiency virus type 1 among bar and hotel workers in northern Tanzania: the role of alcohol, sexual behavior, and herpes simplex virus type 2**. Sexually Transmitted Disease 33 (3): 163-169.

Arroyo M. A, M. Hoelscher, w. Sateren, E Samky, L. Maboko, O. Hoffmann, G. Kijak, M. Robb, D. L. Birx and F. E. McCutchan. 2005. **HIV-1 diversity and prevalence differ between urban and rural areas in the Mbeya region of Tanzania.** AIDS 19(4):1517-1524.

Arroyo, M. A., M. Hoelscher, E. Sanders-Buell, K. H. Herbinger, E. Samky, L. Maboko, O. Hoffmann, M. R. Robb, D. L. Birx, F. E McCutchan. 2004. **HIV type 1 subtypes among blood donors in the Mbeya region of southwest Tanzania.** Aids Research and Retroviruses 20 (8): 895-901.

Ashorn P., B. Moss, J. N. Weinstein, V. K. Chaudhary, D. J. FitzGerald, I. Pastan, E. A. Berger. 1990. **Elimination of infectious human immunodeficiency virus from human T-cell cultures by synergistic action of CD4-Pseudomonas exotoxin and reverse transcriptase inhibitors.** Proc Natl Acad Sci U. S.A 87 (22):8889-8893.

Beirnaert E., P. Nyabi, B. Willems, L. Heyndrickx, R. Colebunders, W. Janssens, G. van der Groen. 2000. **Identification and characterization of sera from HIV-infected individuals with broad cross-neutralizing activity against group M (env clade A-H) and group O primary HIV-1 isolates.** Journal of Medical Virology 62 (1): 14-24.

Berger E. A., P. M. Murphy, J. M. Farber. 1999. **Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease.** Annual Review of Immunology 17:657-700.

Berkowitz R. D., S. Alexander, C. Bare, V. Linquist-Stepps, M. Bogan, M. E. Moreno, L Gibson, E. D. Wieder, J. Kosek, C. A Stoddart and J. M. McCune. 1998. **CCR5- and CXCR4-utilizing strains of human immunodeficiency virus type 1 exhibit differential tropism and pathogenesis in vivo.** Journal of virology 72(12):10108-10117.

Bernstein H. B., S. P. Tucker, S. R. Kar, S. A. McPherson, D. T. McPherson, J. W. Lebowitz, R. W. and E. Hunter. 1995. **Oligomerization of the hydrophobic heptad repeat of gp41.** Journal of virology 69 (5):2745-2750.

Betts M. R., M. C. Nason, S. M. West, S. C. De Rosa, S. A. Migueles, J. Abraham, M. M. Lederman, J. M. Benito, M. Connors, M. Roederer and R. A. Koup. 2006. **HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells.** Blood 107(12):4781-4789.

Bhattacharya T., M. Daniels, D. Heckerman, B. Foley, N. Frahm, C. Kadie, J. Carlson, K. Yuslim, B. McMahon, B. Gaschen, S. Mallal, J. I. Mullins, D. C. Nickle, J. Herbeck, C. Rousseau, G. H. Learn, T. Miura, C. Brader, B. Walker and B. Korber. 2007. **Founder effects in the assessment of HIV polymorphisms and HLA allele associations.** Science 315(5818):1583-1586.

Binley J. M., T. Wrin, B. Korber, M. B. Zwick, M. Wang C. Chappey, G. Stiegler, R. Kunert, S. Zolla-Pazner, H. Katinger, C. J. Petropoulos and D. R. Burton. 2004. **Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies.** Journal of virology 78(23):13232-13252.

Boyer J. C., K. Bebenek, T. A. Kunkel. 1992. **Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates.** Proc Natl Acad Sci USA 89(15):6919-6923.

Brander C. and Walker B. D. 2003. **Gradual adaptation of HIV to human host populations: good or bad news?** Nature 9(11):1359-1362.

Bushman F. D., T. Fujiwara, R. Craigie. 1990. **Retroviral DNA integration directed by HIV integration protein in vitro.** Science 249(4976): 1555-1558.

Carr J. K., M. Avila, M. Gomez Carrillo, H. Salomon, J. Hierholzer, V. Watnaveeradej, M. A. Pando, M. Negrete, K. L. Russell, J. Sanchez, D. L. Birx, R. Andrade, J. Vinoles and F. E. McCutchan. 2001. **Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America.** AIDS 15(15):F41-47.

Carrington M., G. W. Nelson, M. P. Martin, T. Kissner, D. Vlahov, J. J. Goedert, R. Kaslw, S. Buchbinder, K. Hoots and S. J. O'Brien. 1999. **HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage**. Science 283: 253-258.

Chiu I. M., A. Yaniv, J. E. Dahlberg, A. Gazit, S. F. Skuntz, S. R. Tronickand S. A Aaronson. 1985. **Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses.** Nature 317(6035):366-368.

Chun T. W., L. Stuyver, S. B. Mizell, L. A. Ehler, J. A. Mican, M. Baseler, A. L. Llyod, M. A. Nowak and A. S. Fauci. 1997. **Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy.** Proc Natlt Acad Sci USA 94(24):13193-13197.

Cilliers T., J. Nhlapho, M. Coetzer, D. Orlovic, T. Ketas, W. C. Olson, J. P. Moore, A. Trkola and L. Morris. 2003. **The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C.** Journal of Virology 77(7):4449-4456.

Cilliers, T., J. Nhlapo, M. Coetzer, D. Orlovic, T. Ketas, W. C. Olson, J. P. Moore, A. Trkola, and L. Morris. 2003. **The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C**. Journal of Virology 77:4449-4456.

Cilliers, T., S. Willey, W. M. Sullivan, T. Patience, P. Pugach, M. Coetzer, M. Papathanasopoulos, J. P. Moore, A. Trkola, P. Clapham, and L. Morris. 2005. **Use of alternate coreceptors on primary cells by two HIV-1 isolates.** Virology **339**:136-144.

Coffin, J. M. 1995. **HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy.** Science 271(5249):670-671.

Cohen E. A.., G. Dhni, J. G. Sodroski and W. A. Haseltine. 1990. **Human immunodeficiency virus vpr product is a virion-associated regulatory protein.** Journal of Virology 64(6):3097-3099.

Connor R. I., K. E Sheridan, D. Ceradini, S. Choes and N. R. Landaau. 1997. **Change in coreceptor use coreceptor use correlates with disease progression in HIV-1 infected individuals.** Journal of Experimental Medicine 185(4):621-628.

Dean M., M. Carrington, G. A. Huttley, M. W. Smith, R. Allikmets, J. J. Goedert, S. P. Buchbinder, E. Vittinghoff, E. Gomperts, S. Donfield, D. Vlahov, R. Kaslow, A. Saah, C. Rinaldo, R. Detels and S. J. O'Brien. 1996. **Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study.** Science 273(5283):1856-1862.

Decker J. M., F. Bibollet-Ruche, X. Wei, S. Wang, D. N. Levy, W. Wang, E. Delaporte, M. Peeters, C. A. Derdeyn, S. Allen, E. Hunter, M. S. Saag, J. A. Hoxie, B. H. Hahn, P. D. Kwong, J. E. Robinson and G.M Shaw. 2005. **Antigenic conservation and immunogenicity of the HIV coreceptor binding site.** Journal of Experimental Medicine 201(9):1407-1419.

Delgado E., M. M. Thomson, M.L Villahermosa, M. Sierra, A. Ocampo, C. Miralles, R. Rodriguez-Perez, J. Diz-Aren, R. Ojea-de Castro, E. Losada, E. Vazquez-de Parga, R. Carmona, L. Perez-Alvarez, L. Medrano, L. Cuevas, J. A. Taboada and R. Najera. 2002. **Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure.** Journal of Immune Deficiency Syndrome 29(5):536-543.

Delwart E. L., H. W. Sheppard, B. D. Walker, J. Goudsmith and J. I. Mullins. 1994. **Human immunodeficiency virus type 1 evolution in vivo tracked by DNA heteroduplex mobility assays.** Journal of virology 68(10):6672-6683.

Deng H., R. Liu, W. Ellmeier, S. Choe, D. Unutmaz and M. Burkhart. 1996. **Identification of a major co-receptor for primary isolates of HIV-1.** Nature 381: 661-666.

Derdeyn C. A. and G. Silvertri. 2005. Viral and host factors in the pathogenesis of HIV infection. Current Opinions in Immunology 17(4):366-373.

Diaz T., S. Y. Chu, L. Conti, F. Sorvillo, P. J. Checko, P. Hermann, S. A. Fann, M. Frederick, D. Boyd and E. Mokotoff. 1994. **Risk behaviors of persons with heterosexually acquired HIV infection in the United States: results of a multistate surveillance project.** Journal of Acquired Immune Deficiency Syndrome 7(9):958-963.

Dowling W. E., B. Kim, C. J. Mason, K. M. Wasunna, U. Alam, L. Elson, D. L. Birx, M. L. Robb, F. E. McCutchan and J. K. Carr. 2002. **Forty-one near full-length HIV-1 sequences from Kenya reveal an epidemic of subtype A and A-containing recombinants.** Virology 247: 22-31.

Dragic T., V. Litwin, and G. P. Allaway. 1996. **HIV-1 entry into CD4 cells is mediated by the chemokine receptor CC-CKR5.** Nature 381:667-673.

Edwards, C. T. T., K. J. Pfafferott, P. J. R. Goulder, R. E. Phillips, and E. C. Holmes. 2005. **Intrapatient escape in the A*0201-restricted epitope SLYNTVATL drives evolution of human immunodeticiency virus type 1 at the population level.** Journal of Virology 79:9363-9366.

Fauci A. S. 1996. **Host factors in the pathogenesis of HIV disease.** Antibiot Chemother 48:4-12. **Fauci, A. S.** 2003. **HIV and AIDS: 20 years of science.** Nature Medicine 9:839-843.

Ferrai G., W. Neal, A. Jones, N. Olender, J. Ottinger, R. Ha, M. J. McElrath, P. Goepfert and K. J. Weinhold. 2001. **CD8 CTL responses in vaccines: emerging patterns of HLA restriction and epitope recognition.** Immunology leters79: 37-45.

Ferrari, G., W. Neal, J. Ottinger, A. M. Jones, B. H. Edwards, P. Goepfert, M. R. Betts, R. A. Koup, S. Buchbinder, M. J. McElrath, J. Tartaglia, and K. J. Weinhold. 2004. **Absence of immunodominant anti-gag p17 (SL9) responses among gag CTL-Positive, HIV-uninfected vaccine recipients expressing the HLA-A*0201 allele.** Journal of Immunology 173:2126-2133.

Fouchier R.A, M. Groenink, N. A. Kootstra, M. Termette, H. G. Huisman, F. Miedema and H. Schuitemaker. 1992. **Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule.** Journal of virology 66(5):3183-3187.

Franke E. K., H. E. Yuan and J. Luban. 1994. **Specific incorporation of cyclophilin A into HIV-1 virions.** Nature 372(6504):359-362.

Frost S. D. W., T. Wrin, D. M. Smith, S. L. Kosakovsky Pond, Y. Liu, E. Paxinos, C. Chappey, J. Galovich, J. Beauchaine, C. J. Petropoulos, S. J. Little and D. D. Richman. 2005. **Neutralizing antibody responses drive the evolution of HIV-1 envelope during recent HIV-1 infection.** PNAS 102(5): 18514-18519.

Fultz P. N., A. Srinivasan, C. R. Greene, D. Butler, R. B. Swenson and H. M. McClure. 1987. **Superinfection of a chimpanzee with a second strain of human immunodeficiency virus.** Journal of virology.61: 402-409.

Gallay P., S. Swingler and J. Song *et al.*, 1995. **HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase.** Cell 83(4):569-576.

Gao F., D. L. Robertson and C. D. Carruthers. 1998. **A comprehensive panel of near-full-length clones and reference immunodeficiency virus type 1.** Journal of virology 72:5680-5698.

Gao F., L. Yue, S. Craig, C. L. Thornton, D. L. Robertson, F. E. McCutchan, J. A. Brada, P. M. Sharp and B. H. Hahn. 1994. **Genetic variation of HIV type 1 in four World Health Organization-sponsored vaccine evaluation sites: generation of functional envelope (glycoprotein 160) clones representative of sequence subtypes A, B, C, and E. WHO Network for HIV Isolation and Characterization.** Aids Research and Human Retroviruses 10(11):1359-1368.

Gao F., S. A. Trask,H. Hui, O. Mamaeve, Y. Chen, T. S. Theodore, B. T Korber, G. M. Shaw and B. H. Hahn. 2001. **Molecular characterization of a highly divergent HIV type 1 isolate obtained early in the AIDS epidemic from the Democratic Republic of Congo.** Aids Research and Human Retroviruses 17(12):1212-1222.

Garcia J. V. and A. D. Miller. 1992. **Downregulation of cell surface CD4 by nef.** Res Virology 143(1):52-55.

Geldemacher C., J. R. Currier, M. Gerhardt, A. Haule, L. Maboko, D. Birx, C. Gray, A. Meyerhans, J. Cox and M. Hoelscher. 2007. **In a mixed subtype epidemic, the HIV-1 Gag-specific T-cell response is biased towards the infecting subtype.** AIDS 21(2):135-143.

Gerhardt, M., D.Mloka, S. Tovanabutra, E. Sanders-Buell, O. Hoffmann, L. Moboko, D. Mmbando, D. L. Birx, F. E. McCutchan and M. Hoelscher. 2005. **In-depth, longitudinal analysis of viral quasispecies from an individual triply infected with late-stage human immunodeficiency virus type 1, using a multiple PCR primer approach.** Journal of Virology 79(13)8249-8261.

Giorgi J. V., R. H. Lyles, J. L. Matud, T. E. Yamashita, J. W. mellors, L. E. Hultin, B. D. Jamieson, J. B. Margolick, C. R. Rinaldo Jr, J. P. Phair and R. Detels 2002. **Predictive value of immunologic and virologic markers after long or short duration of HIV-1 infection.** J. Acquir Immune deficiency Syndrome 29(4):346-355.

Goodrich D.W. and P. H. Duesberg. 1990. **Retroviral recombination during reverse transcription.** Proc Natl Acad Sci USA 87(6):2052-2056.

Gorny M. K., K. Revesz, C. Williams, B Volsky, M. K. Louder, C. A. Anyangwe, C. Krachmarov, S. C. Kayman, A. Pinter, A. Nadas, P. N. Nyambi, J. R. Mascola and S. Zolla-pazner. 2004. **The V3 loop is accessible on the surface of most HIV-1 primary isolates and servesas a neutralization epitope.** Journal of virology 76: 9035-9045.

Gottlieb, G. S., D. C. Nickle, M. A. Jensen, K. G. Wong, J. Grobler, F. S. Li, S.L. Liu, C. Rademeyer, G. H. Learn, S. S. A. 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.

Goulder P. J. and B. D. Walker. 1996. **HIV-1 superinfection-a word of caution.** New England Journal of Medicine 347(10):756-758.

Goulder P. J. and B. D. Walker. 2002. **HIV-1 superinfection – a word of caution.** New England Journal of Medicine 347: 756-758.

Goulder P., D. Price, M. Nowak, S. Rowland-Jones, R. Phillips and A. McMichael. 1997. **Co-evolution of human immunodeficiency virus and cytotoxic T-lymphocyte responses.** Immunology Reviews 159:17-29.

Goulder, P. J. R. 2000. **Rapid characterization of HIV clade C-specific cytotoxic T lymphocyte responses in infected African children and adults.** Prevention and Treatment of HIV Infection in Infants and Children 918:330-345.

Gregson S., G. P. Garnett, C. A. Nyamukapa, T. B. Hallett, J. J. Lewis, P. R. Mason, S. K. Chandiwana and R. M. Anderson. 2006. **HIV decline associated with behaviour change in eastern Zimbabwe.** Science 311(576):664-666.

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.** Journal of Infectious Diseases 190:1355-1359.

Hemelaar J., E. Gouws, P. D. Ghys and S. Osmanov. 2004. **Global and regional distribution of CRF_ 02 HIV-1 genetic subtypes and recombinants in 2004.** AIDS 20(16):W13-23.

Herbinger K. H., M. Gerhardt, S. Piyasirisilp, D. Mloko, M. A. Arroyo, O. Hoffmann, L. Maboko, D. L. Birx, D. Mmbando, F. E. McCutchan and M. Hoelscher. 2006. **Frequency of HIV type 1 dual infection and HIV diversity: analysis of low- and high-risk populations in Mbeya Region, Tanzania.** AIDS Res Hum Retroviruses 22(7):599-606.

Ho D. D., A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard and M. Markowitz. 1996. Nature 373: 123-126.

Hoelscher M., S. Hanker and F. Barin. 1998. **HIV type 1 V3 serotyping of Tanzanian samples: probable reasons for mismatching with genetic subtypes.** AIDS Research and Human Retroviruses 14:139-149.

Hoelscher, M., B. Kim, L. Maboko, F. Mhalu, F. von Sonnenburg, D. L. Birx, and F. E. McCutchan. 2001. **High proportion of unrelated HIV-1 intersubtype recombinants in the Mbeya region of southwest Tanzania.** Aids 15:1461-1470.

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

Howard T. M., D. O. Olaylele and S. Rasheed. 1994. **Sequence analysis of the glycoprotein 120 coding region of a new HIV type 1 subtype A strain (HIV-1IbNg) from Nigeria.** Aids Research and Retroviruses 10(12):1755-1757.

Inwoley A., P. Recordo-Pinson, M. Dupuis, J. Gaston, M. Genete, A. Minga, F. Letourneur, F. Rouet, J. Choppin, H. Fleury, J. G. Guillet and M. Andrieu. ANRS 1120 PRIMO-CI Study Grroup. 2005. **Cross-clade conservation of HIV type 1 Nef immunodominant regions recognized by CD8+ T cells of HIV type 1 CRF02_AG-infected Ivorian (West Africa).** Aids Research and Retroviruses 21(7):620-628.

Inwoley, A., P. Recordon-Pinson, M. Dupuis, J. Gaston, M. Genete, A. Minga, F. Letourneur, F. Rouet, J. Choppin, H. Fleury, J. G. Guillet, and M. Andrieu. 2005. **Cross-clade conservation of HIV type 1 nef immunodominant regions recognized by CD8(+) T cells of HIV type 1 CRF02_AG-infected Ivorian (West Africa).** Aids Research and Human Retroviruses 21:620-628.

Janini L. M., A. Tanuri, M. Schechter, J.M. Peralta, A. C. Vicente, N. Dela Torre, N. J. Pieniazek, C. C. Luo, A. Ramos, V. Soriano, G. Schochetman, M. A. Rayfield and D. Pieniazek. 1998. **Horizontal and vertical transmission of human immunodeficiency virus type 1 dual infections caused by viruses of subtypes B and C.** Journal of Infectious Diseases 177(1):227-231.

Jensen, M. A., M. Coetzer, A. B. van 't Wout, L. Morris, and J. I. Mullins. 2006. **A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences.** Journal of Virology 80:4698-4704.

Jetzt A. E., H. Yu, G. J. Klarmann, Y. Ron, B. D. Preston and J. P. Dougherty. 2000. **High rate of recombination throughout the human immunodeficiency virus type 1 genome.** Journal of virology 74:(3): 1234-1240.

Jin X., D. E. Bauer, S. E. Tuttleton, S. Lewin, A. Gettie, J. Blanchard, C. E. Irwin, J. T. Safrit, J. Mittler, L. Weinberger, L. G. Kostrikis, L. Zhang, A. S. Perelson and D. Ho. 1999. **Dramatic rise in plasma viremia after CD8 T-cell depletion in simian immunodeficiency virus-infected macaques.** Journal of Experimental Medicine 189; 991-998.

Johnson M., A. Phillips and D. Wilkinson. 2003. **HIV infection.** Clin Evid (8): 693-701.

Jost S., M. C. Bernard, L. Kaiser, S. Yerly, B. Hirschel, A. Samri, B. Autran and L. E Perrin. 2002. **A patient with HIV-1 superinfection.** New England Journal of Medicine 347(10):731-736.

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(6894):144.

Kaleebu P., N. Frech, C. Mahe, D. Yirrell, C. Watera, F. Lyagoba, J. Naiyingi, A. Rutebemerwa, D. Morgan, L. Wbster, C. Gilks and J. Whitworth. 2002. **Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda.** Journal of Infectious Disease 185(9):1244-1250.

Kapiga S. H., and J. L. Lugalla. 2002. **Sexual behaviour patterns and condom use in Tanzania: results from the 1996 Demographic and Health Survey.** AIDS Care 14(4):455-469.

Kelleher A. D., C. Long, E. C. Holmes, r. l. allen, J. Wilson, C. Conlon, c. workman, S. Shaunak, K. Olson, P. Goulder, C. Brander, G. Ogg, J. S. Sullivan, W. Dyer, I. Jones, A. J. McMichael and R. E. Philips. 2001. **Clustered mutations in HIV-1gag are consistently required for escape from HLA-B27 restricted cytotoxic T-lymphocytes responses.** Journal of Experimental Medicine 193; 75-386.

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

Kiepiela P., A. J. Leslie, I. Honeyborne, D. Ramduth, C. Thobakgale, S. Chetty, P. Rathnavalu, C. Moore, K. J. Pfafferott, L. Hilton, P. Zimbwa, S. Moore, T. Allen, C. Brander, M. M. Addo, M. Altfeld, I. James, S. Mallal, M. Bunce, L. D. Barber, J. Szinger, C. Day, P. Klenerman, J. I. Mullins, B. Korber, H. M. Coovadia, B. D. Walker, P. Goulder. 2004. **Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA.** Nature 432(7018):769-775.

Kiepiela P., A. N. Smith and E. Rosenberg. 2005. **Laboratory markers associated with progression of HIV infection.** Best Pract Res Clin Obstet Gynaecol 19(2):243-254.

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

Kijak G.H., E. Sanders-Buell and N. D. Wolfe. 2004. **Development and application of a high-throughput HIV type 1 genotyping assay to identify CRF02_AG in West/West Central Africa.** Aids Research and Human Retroviruses. 20: 521-530.

Kim S. Y., R. Byrn, J. Groopma and D. Baltimore. 1989. **Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression.** Journal of Virology 63(9):3708-3713.

Kishamawe C., D. C. Visser, M. Urassa, R. Isingo, G. Mwaluko, G. J. Borsboom, H. A Voeten, B. Zaba, J. D. Habbema and S. J. de Vlas. 2006. **Mobility and HIV in Tanzanian couples: both mobile persons and their partners show increased risk.** AIDS 20(4): 601-608.

Kiwelu I. E., B. Renjiro, B. Chaplan, N. Sam, W. N. Nkya, J. Shao, S. Kapiga and M. Essex. 2003. **HIV Types among Bar and Hotel Workers in Moshi, Tanzania**. Aids Research and Human Retroviruses 19(1):57-64.

Klimkait T., K. Strebel, M. D. Hoggan, M. A. Martin and J. M Orenstein. 1990. **The human immunodeficiency virus type 1-specific protein vpu is required for efficient virus maturation and release.** Journal of Virology 64(2):621-629.

Koelsch K. K., D. M. Smith, S. J. Little, C. C. Ignacio, T. R. Macaranas, A. J. Brown, C. J. Petropolos , D. D. Richman and J. K. Wong. 2003. **Clade B HIV-1 superinfection with wild-type virus after primary infection with drug-resistsnt clade B virus.** AIDS 17: F11-F16.

Koulinska I. N., T. Ndung'u, D. Mwakagile, G. Msamanga, C. Kagoma, W. Fawzi, M. Essex and B. Renjifo. 2001. **A New Human Immunodeficiency Virus Type 1 Circulating Form from Tanzania.** AIDS Research and retrovivuses17(5):423-431.

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

Kumar S. K., K. Tamura, I. B. Jakobsen and M. Nei. 2001. **MEGA2: molecular evolutionary genetics analysis software.** Bioinformatics 17: 1244-1245.

Kupfer B., R. Kaiser, J. k. Rockstroh, B. Matz and K. E. Schneweis. 1998. **Role of HIV-1 phenotype in viral pathogenesis and its relation to viral load and CD4+ T-cell count.** Journal of Medical Virology 56(3):259-263.

Kwesigabo G., J. Killewo, W. Urassa, J. Lugalla, M. Emmelin, A. Mutembei, F. Mhalu, G. Biberfeld, S. Walls and A Sandstrom. 2005. **HIV-1 infection prevalence and incidence trends in areas of contrasting levels of infection in the Kagera region, Tanzania, 1987-2000.** J Acquir Immune Defic Syndr 40(5):585:591.

Labrijn A. F., P. Poignard, A. Raja, M. B. Zwick, K. Delgado, M. Franti, J. Binley, V. Vivona, C. Grundner, C. C. Huang, M. Venturi, C. J. Petropoulos, T. Wrin, D. S. Dimitrov, J. Robinson, P. D. Kwong, R. T. Wyatt, J. Sodroski and D. R. Burton. 2003. **Access of antibody molecules to the conserved coreceptor binding site on glycoprotein gp120 is sterically restricted on primary human immunodeficiency virus type 1.** Journal of Virology 77(19):10557-10565.

Laukkanen T., J. K. Carr, W. Janssesns, K. Liitsola, D. Gotte, F. E McCutchan, E. Op de Coul, M. Cornelissen, L. Heyndrickx, G. van der Groen and M. O. Salminen. 2000. **Virtually full-length subtype F and F/D recombinant HIV-1 from Africa and South America.** Virology 269(1):95-104.

Laukkanen T., K. Liitsola, M. Salminen and P. Leinikki. 1996. **HIV-1 D subtype viruses in Finland.** Clin Diagn Virol 5(2-3):205-210.

Lavreys L., J. M. Baeten, D. D. Panteleeff, B. A. Richardson, R. S. McClelland, K. Mandaliya, J. O. Ndinya-Achola and J. Overbaugh. 2006. **High levels of cervical HIV-1 RNA during early HIV-1 infection.** AIDS 20 (18): 2389-2390.

Leggatt G. R., A. Hosmalin, C. D. Pendleton, A. Kumar, S. Hoffman and J. A. Berzofsky. 1998. **The importance of pairwise interactions between peptide residues in the delineation of TCR specificity.** Journal of Immunology 161(9):4728-4735.

Levy J.A., 1993. **Pathogenesis of human immunodeficiency virus infection.** Microbiological Reviews 57: 183-189.

Levy L.A. 1995. **History and epidemiology of acquired immune deficiency syndrome.** J Am Podiatr Assoc 85(7)346-351.

Li Pira, G., L. Bottone, D. Fenoglio, P. Terranova, E. Pontali, F. Ivaldi, F. Del Galdo, L. Mortara, A. Loregian, G. Palu, A. Kunkl, R. Accolla, R. De Palma, and F. Manca. 2001. **Analysis of the antigen specific T cell repertoires in HIV infection**. Immunology Letters 79:85-91.

Liitsola K., I. Tashkinova and T. Laukkanen. 1998. **HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad**. AIDS 12: 1907-1919.

Liu M. X., F.S. Wang, W. G. Hong B. Wang, L. Jin, Z. Y. Lei and J. Hou. 2002. **Single nucleotide polymorphism loci of HIV-1 coreceptor CCR5 gene in Chinese Han people.** Za Zhi 20 (1):72-74.

Liu R., W. A. Paxton and S. Choe. 1996. **Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection.** Cell 86: 376-377.

Lole K. S., R. C. Bollinger, R. S. Paranjape, D. geldkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard and S. C. Ray. 1999. **Full-length HIV-1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtpye recombination.** Journal of Virology 73(1):152.

Luria S., I. Chambers and P. Berg. 1991. **Expression of the type 1 human immunodeficiency virus Nef protein in T cells prevents antigen receptor-mediated induction of interleukin 2 mRNA.** Proc Natl Acad Sci USA 88(12):5326-5330.

Lyles C. M., M. Dorrucci, D. Vlahov, P. Pezzotti, G. Angarano. A. Sinicco, F. Alberici, T. M. Alcorn, S. Vella and G. Rezza. 1999. **Longitudinal human immunodeficiency virus type 1 load in the italian seroconversion study: correlates and temporal trends of virus load.** Journal of Infectious Diseases 180(4):1018-1024.

Lyles R. H., A. Munoz and T. E. Yamasita. 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**. Journal of Infectious Diseases. 181: 872-880.

Malim M. H and M. Emerman. 2001. **HIV-1 sequence variation: drift, shift and attenuation**. Cell 104:469-472.

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

Mamadou M., C. Montavon, A. Ben, A. Djibo, S. Rabiou, S. Mboup, E. Delaporte and M. Peeters. 2000. **Predominance of CRF02_AG and CRF06_cpx in Niger, West Africa.** AIDS Research and Retrovivuses 18(10): 723-726.

Mangeat B., P. Turelli, G. Caron, M/ Friedli, L. Perrin and D. Trono. 2003. **Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts**. Nature 424: 99-103.

Mansky L.M and H. M. Temin. 1995. **Lower in vivo mutation rate of HIV-1 than that predicted from the fidelity of purified reverse transciptase.** Journal of virology 69(8): 5087-5094.

Martin D. P. and **E. Rybicki**. 2000. **RDP: detection of recombination amongst aligned sequences.** Bioinformatics 16:562-563.

Martin D. P., Posada D. Cradall K. A. and C. Williamson. 2005b **A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints.** Aids Research and Retroviruses 21(1):98-102.

Martin D.P., 2005. **RDP2: detection of recombination amongst aligned sequences.** Bioinformatics 21:5260-262.

Mascola J. R., M. G. Lewis, T. C VanCott, G. Stiegler, H. Katinger, M. Seaman, K. Beaudry, D. H. Barouch, B. Korioth-Schmitz, G. Krivulka, A. Sambor, B. Welcher, D. C. Montefiori, J. W. Shiver, P. Poignard, D. R. Burton, N. L. Letvin. 2003. **Cellular immunity elicited by human immunodeficiency virus type 1/ simian immunodeficiency virus DNA vaccination does not augment the sterile protection afforded by passive infusion of neutralizing antibodies.** Journal of Virology 77 (19):10348-10356.

Masemola, A. M., T. N. Mashishi, G. Khoury, H. Bredell, M. Paximadis, T. Mathebula, D. Barkhan, A. Puren, E. Vardas, M. Colvin, L. Zijenah, D. Katzenstein, R. Musonda, S. Allen, N. Kumwenda, T. Taha, G. Gray, J. Mcintyre, S. A. Karim, H. W. Sheppard, and C. M. Gray. 2004. **Novel and promiscuous CTL epitopes in conserved regions of Gag targeted by individuals with early subtype CHIV type infection from southern Africa**. Journal of Immunology **173**:4607-4617.

Masemola, A., T. Mashishi, G. Khoury, P. Mohube, P. Mokgotho, E. Vardas, M. Colvin, L. Zijenah, D. Katzenstein, R. Musonda, S. Allen, N. Kumwenda, T. Taha, G. Gray, J. Mcintyre, S. A. Karim, H. W. Sheppard, and C. M. Gray. 2004. **Hierarchical targeting of subtype C human immunodeficiency virus type 1 proteins by CD8(+) T cells: Correlation with viral load**. Journal of Virology 78:3233-3243.

Mashishi, T., S. Loubser, W. Hide, G. Hunt, L. Morris, G. Ramjee, S. bdool-Karim, C. Williamson, and C. M. Gray. 2001. **Conserved domains of subtype C Nef from South African HIV type 1-infected individuals include cytotoxic T lymphocyte epitope-rich regions**. Aids Research and Human Retroviruses 17:1681-1687.

McCutchan F. E., 2000. **Understanding the genetic diversity of HIV-1.** AIDS 14 Supplementary 3:S32-44.

McCutchan F. E., J. L. Sankale, S. M'Boup, B. Kim, S. Tovanabutra, D. J. Hamel, S. K. Brodine, P. J. Kanki and D. L. Birx. 2004. **HIV type 1 circulating recombinant form CRF09_cpx from west Africa combines subtypes A, F, G, and may share ancestors with CRF02_AG and Z321.** AIDS Research and Human Retroviruses 20(8):819-826.

McCutchan, F. E., M. Hoelscher, S. Tovanabutra, S. Piyasirisilp, E. Sanders-Buell, G. Ramos, L. Jagodzinski, V. Polonis, L. Maboko, D. Mmbando, O. Hoffmann, G. Riedner, F. von Sonnenburg, M. Robb, and D. L. Birx. 2005. **In-depth analysis of a heterosexually acquired human immunodeficiency virus type 1 superinfection: Evolution, temporal fluctuation, and intercompartment dynamics from the seronegative window period through 30 months postinfection**. Journal of Virology **79**:11693-11704.

McDermott D. H., P. A. Zimmerman and F. Guignard. 1998. **Nucleotide, OMIM CCR5 promoter poymorphism and HIV-1 disease progession. Multicenter AIDS Cohort Study (MACS).** Lancet 352:866-870.

Mellors J. W., A. Munoz, J. V. Giorgi, J. B. Margolick, C. J. Tassoni, P. Gupta, L. A. Kingley, J. A. Todd, A. J. Saah, R. Detels, J. P. Phair and C. R. Rinaldo Jr. 1997. **Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection.** Annals of Internal Medicine 126(12):946-954.

Mellors J. W., L. A. Kingsley and C. R. RinaldoJr. 1995. **Quantification of HIV-1 RNA in plasma predicts outcome after seroconversion.** Ann Intern Med 122: 573-579.

Michael N. L., G. Chang and L. G. Louie 1997. **The role of viral phenotype and CCR-5 gene defects in HIV-1 transmission and disease progression.** Nature Medicine 3: 338-340.

Miller M. D., M. B. Feinberg and W. C. Greene. 1994. **The HIV-1 nef gene acts as a positive viral infectivity factor.** Trends Microbiology 2(8):294-298.

Montavon C., L. Vergne, A. Bourgeois, E. Mpouni-Ngole, G. Malonga-Mouellet, C. Butel, C. Toure-Kane, E. Delaporte and M. Peeters. **Identification of a New Circulating Recombinant Form of HIV Type 1, CRF11_cpx, Involving Subtypes A, G, J and CRF01_AE, in Central Africa.** AIDS Research and Human Retroviruses 18(3):231-236.

Moore C. B., M. John, I. R. James, F. T. Christiansen, C. S. Witt and S. A. Mallal. 2002. **Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level.** Science 296(5572):1439-1443.

Mummidi S., S. S. Ahuja, E. Gonzalez, S. A Anderson, E. N. Santiago, K. T. Stephan, F. E. Craig, P. O'Connell, V. Tryon, R. A. Clark, M. L. Doaln and S. K. Ahuja. 1998. **Genealogy of the CCR5 locus and chemokine system gene variants associated with altered rates of HIV-1 disease progression.** Nature Medicine 4(7):786-793.

Najera, R., E. Delgado, L. Perez-Alvarez, and M. M. Thomson. 2002. **Genetic recombination and its role in the development of the HIV-1 pandemic**. Aids 16:S3-S16.

Novitsky, V., H. Cao, N. Rybak, P. Gilbert, M. F. Mclane, S. Gaolekwe, T. Peter, I. Thior, T. Ndung'u, R. Marlink, T. H. Lee, and M. Essex. 2002. **Magnitude and frequency of cytotoxic T-lymphocyte responses: Identification of immunodominant regions of human immunodeficiency virus type 1 subtype C**. Journal of Virology **76**:10155-10168.

Novitysky V., P. Gilbert, T. Peter, M. F. McLane, S. Gaolekwe, N. Rybak, I. Thior, T. Ndung'u, R. Marlink, T. H. Lee and M. Essex. 2003. **Association between virus-specific T-cell responses and plasma viral load in human immunodeficiency virus type 1 subtype C infection.** Journal of Virology 77(2):882-890.

Oelrichs R B., C. Workman, T. Laukkanen, F. E. McCutchan and N. J. Deacon. 1998. **A novel subtype A/G/J recombinant full-length HIV type 1 genome from Burkina Faso.** Aids Research and Human Retroviruses 14(16):1495-500.

Ogg G.S., S. Kostense, M. R. Klein, S. Jurriaans, D. Hamann, A. J. McMichael and F. Miedema. 1999. **Longitudinal phenotypic analysis of HIV-1 specific cytotoxic T-lymphocytes: correlation with disease progression.** Journal of virology 73: 9153-9160.

Paraskevis D., M. Magiorkinis, V. Paparizos, G. N. Pavlakis and A. Hatzakis. 2000. **Molecular characterization of a recombinant HIV type 1 isolate (A/G/E/?): unidentified regions may be derived from parental subtype E sequences.** Aids Research and Human Retroviruses 16(9):845-855.

Parker C. E., L. J. Deterding, C. Hager-Braun, J.M. Binley, N. Schulke, H. Katinger, J.P. Moore and K. B. Tomer. 2001. **Fine definition of the epitope on the gp41 glycoprotein of human immunodeficiency virus type 1 for the neutralizing monoclonal antibody 2F5.** Journal of Virology 75(22):10906-10911.

Peeters M. 2000. **Recombination HIV sequences: their role in the global epidemic.** Theoretical biology and biophysics group, Los Alamos National Laboratory. 39-54.

Piyasirusilp S., F. E. McCutchan, J. K. Carr, E. Sanders-Buell, W. Liu, J. Chen, R. Wagner, H. Wolf, Y. Shao, S. Lai, C. Beyrer and X. F. Yu. 2000. **A recent outbreak of HIV-1 infection in southern China was initiated by two highly homogenous, geographically separated strains, circulating recombinant form AE and a novel BC recombinant.** Journal of virology 74: 11286-11295.

Posada D. and K. A. Cradall. 2001. **Evaluation of methods for detecting recombination from DNA sequences: computer simulations.** Proc Natl Acad Sci USA 98(24):13757-13762.

Preston B. D., B. J Poiesz and L. A Loeb. 1988. **Fidelity of HIV-1 reverse transcriptase.** Science 242(4882):1168-1171.

Ramos A., D. J. Hu, L. Nguyeni, K. Phan, S. Vanichseni, N. Promadej, K. Choopanya, M. Callahan, L. Young and S. Subbarao. 2002. **Intersubtye Human Immunodeficiency Virus Type 1 Superinfection following Seroconversion to Primary Infection in Two Injecting Drug Users.** Journal of Virology 76: (15) 7444-7452.

Raport C. J., J. Gosling, V. L. Schweickart, P. W. Gray and I. F. Charo. 1996. **Molecular cloning and functional characterization of a novel human CC chemokine receptor (CCR5) for RANTES, MIP-1beta, and MIP-1alpha.** Journal of Biological Chemistry 271(29):17161-17166.

Regoes R. R., and S. Bonhoeffer. 2005. **The HIV coreceptor switch: a population dynamical perspective.** Trends in Microbiology 13(6): 269-277.

Riedner G., M. Busizoka, O. Hoffmann, F. Nichombe, E. Lyamuya, D. Mmbando, L Maboko, P, Hay, J. Todd, R. Hayes, M. Hoelscher and H. Grosskurth. 2003. **Baseline survey of sexually transmitted infections in a cohort of female bar workers in Mbeya Region, Tanzania.** Sexually Transmitted Infections 79(5):382-387.

Riedner G., O. Hoffmann, M. Rusizola, D. Mmbando, L Maboko, H. Grosskurth, J. Todd, R. Hayes and M. Hoelscher 2006. **Decline in sexually transmitted infection prevalence and HIV incidence in female barworkers attending prevention and care services in Mbeya Region, Tanzania.** AIDS 20(4):609-615.

Robertson D. L., J. P. Anderson, J. A. Bradac, J. K. Carr, B. Foley and R. K. Funkhouser. 2000. **HIV-1 nomenclature proposal.** Science 288: 55-56.

Rowland-Jones, S.L., T. Dong, K. R. Fowke, J, Kimani, P. Krausa, H. Nwell, T. Blanchard, K Ariyoshok, J. Oyugi, J. Ngugi, J. Bwayo, K. S. MacDonald, A. J. McMichael and F. A Plummer. 1998. **Cytotoxic T cell responses to multiple conserved HIV epitopes in HIV-resistant prostitutes in Nairobi.** Jouranal of Clinical Investigations 102(9):1643-1644.

Ruben S., A. Perkins, R. Purcell, K. Joung, R. Sia, R. Burghoff, W. A Haseltine and C. A Rosen. 1989. **Structural and functional characterization of human immunodeficiency virus tat protein.** Journal of virology 63(1):1-8.

Sa Filho D. J., M. C. Sucupira, M. M. Caseiro, E. C. Sabino and L. M. Janini. 2006. **Identification of two HIV type 1 circulating recombinant forms in Brazil.** AIDS Research and Human Retroviruses 22(1):1-3.

Sagar M., L. Lavreys, J. M. Baeten, B. A. Richrdson, 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.** Journal of virology 77923):12921-12926.

Samson M., G. LaRosa, F. Libert, P. Paindavoine, M. Detheux, G. Vassart and M. Parmentier. 1997. **The second extracellular loop of CCR5 is the major determinant of ligand specificity.** J Biol Chem 272(40):24934-24941.

Schacker T., A. C. Collier, J. Hughes, T. Shea and L. Corey. 1996. **Clinical and epidemiologic features of primary HIV infection.** Ann Intern Med 125(4):257-260.

Scheffler K., D. P. Martin and C. Seoighe. 2006. **Robust inference of positive selection from recombining coding sequences.** Phylogenetics 0(0):1-7.

Schmitz J. E., M. J. Kuroda, S. Santra, V. G. Sasseville, M. A. Simon, M. A. Lifton, P. Racz, K. Tenner-Racz, m. Dalesadro, B. J. Scallon, J. Ghrayeb, M. A Forman, D. C. Montefiori, E. P. Rieber, N. L. Letvin and K. A. Reimann. 1999. **Control of viremia in simian immunodeficiency virus infection by CD8 lymphocytes.** Science 283: 857-860.

Shankarappa, R., J. B. Margolick, S. J. Gange, A. G. Rodrigo, D. Upchurch, H. Farzadegan, P. Gupta, C. R. Rinaldo, G. H. Learn, X. He, X. L. Huang, and J. I. Mullins. 1999. **Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection.** Journal of Virology **73**:10489-10502.

Sheehy A. M., N. C. Gaddis, J. D. Choi and M. H. Malim. 2002. **Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein.** Nature 418(6898):646-650.

Sierra M., M. M. Thomson, M. Rios, G. Casado, R. O. Castro, E. Delgado, G. Echevarria, M. Munoz, J. Colomina, R. Carmona, Y. Vega, L. Megrona, L. Perez-Alvarez, G. Contreras and R. Najera. 2005. **The analysis of near full-length genome sequences of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Chile, Venezuela and Spain reveals their relationship to diverse lineages of recombinant viruses related to CRF12_BF.** Infect Genet Evol 5(3):209-217.

Smith M. W., M. Dean and M. Carrington. 1997. **Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression.** Science 277: 959-965.

Sodroski J. G., W. C. Goh, C. A. Rosen, S. Z. Salahuddin, A. Aldovini, G. Franchini, F. Wong-Staal, R. C. Gallo, K. Sugamura and Y. Hinuma. 1985. **trans-Activation of the human T-cell leukemia virus long terminal repeat correlates with expression of the x-lor protein.** Journal of virology 55(3):831-835.

Sterling T., D. Vlahov, J. Astemborski, D. Hoover, J. Margolick and T. Quinn. 2001. **Initial plasma HIV-1 RNA levels and progression to AIDS in women and men.** New England Journal of Medicine 344: 720-725.

Strebel K., D. Daugherty, K. Clouse, D. Cohen, T. Folks and M. A. Martin. 1987. **The HIV 'A' (sor) gene product is essential for virus infectivity.** Nature 328(61320:728-730.

Swai R. O., G. R. Somi, M. I. Matee, J. Killewo, E. F. Lyamuya, G. Kwesigabo, T. Tulli, T. K. Kabalimu, L. Ng'ang'a, R. Isingo and J. Ndayongeje. 2006. **Surveillance of HIV and syphilis infections among antenatal clinic attendees in Tanzania-2003/2004.** BMC Public Health 6:91.

Tang J., C. M. Wilson, S. Meleth, A. Myracle, E. Lobashevsky, M. J. Mulligan, S. D. Douglas, B. Korber, S. H. Vermund, R. A. Kaslow. 2002. **Host genetic profiles predict virological and immunological control of HIV-1 infection in adolescents.** AIDS 16(17):2275-2284.

ten Haaft P., E. J. Verschoor, B. Verstrepen, H. Niphuis, R. Dubbes, W. Koornstra, W. Bogers, B. Rosenwirth and J. L. Heeney. 2004. **Readily acquired secondary infections of human and simian immunodeficiency viruses following single intravenous exposure in non-human primates.** Journal of general virology 85(12):3735-3745.

Thakar M. R., L. S. Bhonge, S. K. Lakhashe, U. Shankarkumar, S. S. Kulkarni, B. A. Mahajan and R. S Paranjape. 2005. **Cytolytic T lymphocytes (CTLs) from HIV-1 subtype C-infected Indian patients recognize CTL epitopes from a conserved immunodominant region of HIV-1 Gag and Nef.** Journal of Infectious Diseases 192(5):749-759.

Thompson C. 1994. **Fiery entrance for the HIV co-receptor.** Lancet 343(8888)48.

Thomson M and R. Najera. 2005. **Molecular epidemiology of HIV-1 variants in the global AIDS pandemic: an update.** AIDS reviews 7(4):210-224.

Troyer R. M., K. R. Collins, A. Abraha, E. Fraundorf, D. M. Moore, R. W. Krizan, Z. Toossi, R. L. Colebunders, M. A. Jensen, G. Vanham and E. J. Arts. 2005. **Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression.** Journal of virology 79(14):9006-9018.

Tsui R., B. L. Herring and J. D. Barbour. 2004. **Human immunodeficiency virus type 1 superinfection was not detected following 215 years of injection drug user exposure.** Journal of virology 78: 94-103.

Vaishnav Y. N., M. Vaishnav and F. Wong-Staal, 1991. **Identification and characterization of a nuclear factor that specifically binds to the Rev response element (RRE) of human immunodeficiency virus type 1 (HIV-1).** New Biol 3(2):142-150.

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

Veazey R. and A. Lackner. 2003. **The mucosal immune system and HIV-1 infection.** AIDS Reviews 5(4):245-252.

Wang F., L. Jin, Z. Lei, H. Shi, W. Hong, D. Xu, M. Shi, J. Jiang, Y. Wang, B. Zhang, M. Liu, Y. Li. 2000. **Distribution of HIV resistance CCR5-delta 32, CCR2-64 I and SDF1-3'A alleles and their polymorphisms in the Han population in China.** 21(4):256-260.

Wei B.L., V. K. Arora, J. L Foster, D. L. Sodora, J. V. Garcia. 2003. **In vivo analysis of Nef function.** Current HIV Research 19(1):41-50.

Wilbe K., C. Casper, J. Albert, T. Leitner. 2002. **Identification of two CRF11-cpx genomes and two preliminary representatives of a new circulating recombinant form (CRF13-cpx) of HIV type 1 in Cameroon.** AIDS Research Human Retroviruses 18(12):849-856.

Williamson, C., L. Morris, M. F. Maughan, L. H. Ping, S. A. Dryga, R. Thomas, E. A. Reap, T. Cilliers, J. van Harmelen, A. Pascual, G. Ramjee, G. Gray, R. Johnston, S. A. Karim, and R. Swanstrom. 2003. **Characterization and selection of HIV-1 subtype C isolates for usé in vaccine development**. AIDS Research and Human Retroviruses 19:133-144.

Winkler C., W. Modi, M. W. Smith, G. W. Nelson, X. Wu, M. Carrington, M. Dean, T. Honjo, K. Tashiro, D. Yabe, S. Buchbinders, E. Vittinghoff, J. J. Goedert, T. R. O'brien, L. P. Jacobson, R. Detels, S. Donfield, A. Willoughby, E. Gomperts, D. Vlahov J. Phair, S. J. O'brien. 1998. **Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. ALIVE Study, Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC).** Science 16: 279(5349):327.

Xu H., E. S. Svarovskaia, R. Barr, Y. Zhang, M. A. Khan, K, Strebel and V. K. Pathak. 2004. **A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion.** Proc Natl Acad Sci USA 101(15): 5652-5657.

Yang O. O., E. S. Daar, B. D. Jamieson, A. Balamurugan, D. M. Smith, J. A. Pitt, C. J. Petropoulos, D. D. Richman, S. J. Little and A. J. Brown. 2005. **Human immunodeficiency virus type 1 clade B superinfection: evidence for differential immune containment of distinct clade B strains**. Journal of virology 79: 860-868.

Zwick M. B., A. F. Labrijn, M. Wang, C. Spenlehauer, E. O. Saphire, J. M. Binley, J. P. Moore, G. Stiegler, H. Katinger, D. R. Burton, P. W. Parren. 2001. **Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41.** Journal of virology 75(22):10892-10905.