Minority HIV-1 Drug Resistance Mutations in Patients Failing Highly Active Antiretroviral Therapy (HAART).

Submitted by

Phumzile Khumalo

In fulfillment of the requirements for the degree of

Master in Medical Sciences in Virology

In the

Faculty of Health Sciences

University of KwaZulu Natal

2014

PREFACE

The experimental work described in this dissertation was carried out in the Hasso Plattner Research Laboratory, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, from February 2012 to December 2013 under the supervision of Dr Michelle Gordon.

These studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma to any other University. Where use has been made of the work of others, it is duly acknowledged in the text.

Phumzile Khumalo (candidate)

9/12/2014

Date

Dr Michelle Gordon (Supervisor)

9/12/2014

Date

Declaration- Plagiarism

I, Phumzile Khumalo, declare that

(i) The research reported in this dissertation, except otherwise indicated, is my original work.

(ii) This dissertation has not been submitted for any degree or examination at any other university.

(iii) This dissertation does not contain other person's data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This dissertation does not contain other person's writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a. Their words have been re-written but the general information attributed to them has been referenced.

b. Where their exact words have been used, their writing has been placed inside quotation marks and referenced.

(v) Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was written by myself alone and have fully referenced such publications.

(vi) This dissertation does not contain text, graphics or tables copied and pasted from the internet unless specifically acknowledged, and the source being detailed in the dissertation and in the references section.

9/12/2014

Phumzile Khumalo (Candidate) Date

Chapter 1

Introduction

1.1 HIV-1

1.1.1 The Epidemiology of HIV-1

By the end of 2011, the Joint United Nation program on HIV/ AIDS (UNAIDS) estimated that 34 million people were living with HIV worldwide http://www.avert.org/aidsSouthAfrica. An estimated 2.5 million had become infected with HIV while 1.7 million people were estimated to have died of AIDS related causes. However, the epidemic of HIV continues to differ between countries and regions. Sub-Saharan Africa is one of the regions most affected by HIV-1, with an estimate of 23.5 million people living with HIV-1. South Africa still remains the country with the highest number of HIV-1 infected people with an estimate of 5.6 million by the end of 2011 (UNAIDS special report 2013). During the same year, the number of infected people who died from HIV/AIDS related causes reduced from 310 000 in year 2009 to 270 190 in 2011 http://www.avert.org/aidsSouthAfrica .

The HIV-1 prevalence in SA is 17.3 % but it varies greatly by province. KwaZulu –Natal has the highest prevalence (15.8%), mainly in the 15-49 year old age group. Western Cape and Northern Cape are the least affected provinces with percentage of 3.8% and 5.9% respectively http://www.avert.org/aidsSouthAfrica .

1.1.2 The Classification of HIV-1 and geological distribution

Human immunodeficiency virus (HIV) is subdivided into two types: the HIV-1 and HIV-2 (Cohen et al., 2008). HIV -1 has been postulated to have originated from three cross species transmissions of simian immunodeficiency virus (SIV) which infected chimpanzees while HIV-2 is thought to have originated from the sooty mangabey monkey (Lihana et al., 2012). HIV-1 genetic diversity is classified into groups: Major (M), Outlier (O) and N (non-M/N), subtypes, sub-subtypes and circulating recombinant forms (CRFs). This classification is based on phylogenic sequences from the HIV-1 genome (pol, gag, env) (Jülg and Goebel, 2005, Paraschiv et al., 2007, Wainberg, 2004). Recently, a new Putative group designated P was described in Cameroonian patients and has not shown evidence of recombination with other HIV-1 subtypes (Lihana et al., 2012, San Mauro and Agorreta, 2010).

Group M is the most common cause of HIV-1 infection worldwide and has nine subtypes A, B, C, D, F, G, H, J and K (Jülg and Goebel, 2005, Paraschiv et al., 2007). It has been reported that 48% of the infections are caused by Subtype C, followed by 12% caused by subtype A, 11% by subtype B, 5% by subtype G, 2% by subtype D and 22% caused by recombinants (Lihana et al., 2012). Subtype C is found largely in Southern and Eastern Africa, while Subtype A, which is further sub dived into A1, A2, A3, A4, and A5, is found mostly in East Africa (Lihana et al., 2012). Subtype D is also found in East Africa while Subtype B has been found to be stable in Northern Africa and is more common in the Western world (Soares et al., 2007). Subtype F which is sub dived into F1 and F2 occurs in Central Africa, South America and Eastern Europe (Jülg and Goebel, 2005, Wainberg, 2004). Subtype G and A/G recombinants also occur in Eastern Africa and in Central Europe, while H and K only occur in Central Europe. Subtype J has been found in Central America (Wainberg, 2004). The circulating recombinants forms occur due to recombination between subtypes occurring mainly in dually infected patients. Recombinants such CRFO2 _AG and CRFO6_CPX are predominant in Western Africa (Lihana et al., 2012).

Group N is a very distinctive form of virus that has only been identified in a few individuals in Cameroon and subtypes within this HIV-1 group are not yet clearly defined. Group O has very diverse viruses, but it is rarely found. Also, subtypes within O group are not yet fully defined (Lihana et al., 2012). The Geological distribution of the HIV-1 subtypes is illustrated in figure 2 below:

Figure 1.1: Geological distribution of HIV-1 subtypes. This distribution shows the most prevalent HIV-1 subtypes and their geological distribution. Source: (http://www.medscape.org/viewarticle/583361_2).

1.1.3 The Structure of HIV-1

The HIV-1 virions are spherical in shape and possess a lipid bilayer membrane that surrounds the viral Env glycoprotein and some cellular proteins (Sierra et al., 2005). The Env glycoprotein is a heterodimer of the gp120 which is attached to the gp41 (transmembrane glycoprotein). The structural proteins that form the core are: Matrix (MA/p17), capsid (CA/p24), Nucleocapsid (NC/p7) and p6. The MA forms a shell that connects directly to the inner side of the membrane. The capsid has the N- terminal domains that are arranged in hexameric ring to form a capsid. The NC is involved in the formation and stabilization of the of the genomic RNA dimers and in the nucleocapsid assembly. The p6 is the domain of p55 and is essential for the last stage of viral assembly as well as the release of the *vpr* protein into the assembled virion. The gp120 encloses the CD4- receptor binding domain as well as the coreceptor binding site (Sierra et al., 2005).

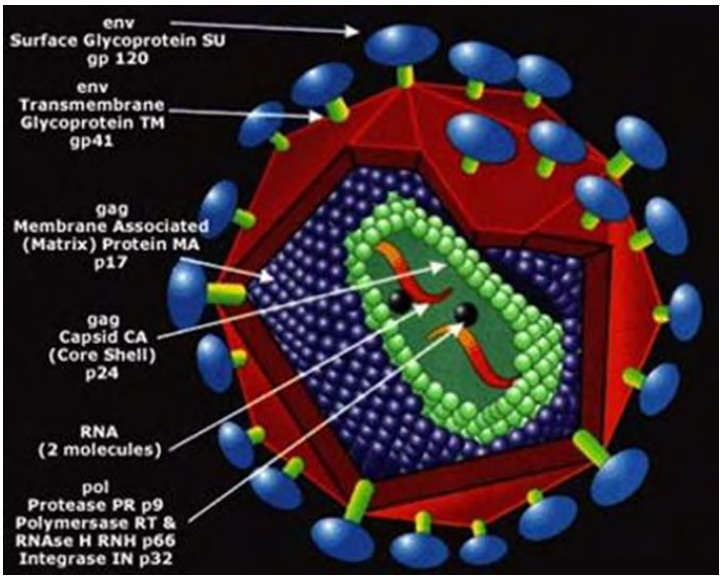


Figure 1.2: The structure of HIV-1. Source: (www.stanford.edu/.../ret_6_gpedesc.html)

The HIV genome is made up of two single stranded RNA molecules which are 9.2kb in size. It has four main regions: the LTR, gag-pol gene, env gene and the accessory genes (Figure 1.3) (Sierra et al., 2005). LTR is a regulatory region that contains the U3, R and U5 regions. The gag-pol gene consists of the Gag polyprotein that comprises the proteins of the nucleocapsid. It also consists of the gag –pol polyprotein that compromises some structural proteins and the viral enzymes: protease (PR), reverse transcriptase (RT), and integrase (IN). The *env* gene encodes the gp160 polypeptide precursor that consists of the exterior gp120 and the transmembrane gp41. The *tat, rev, rev, nef, vif, vpr* and *vpu* are accessory genes that codes for proteins that control the HIV to infect the cell and to produce new copies of the virus (Sierra et al., 2005).

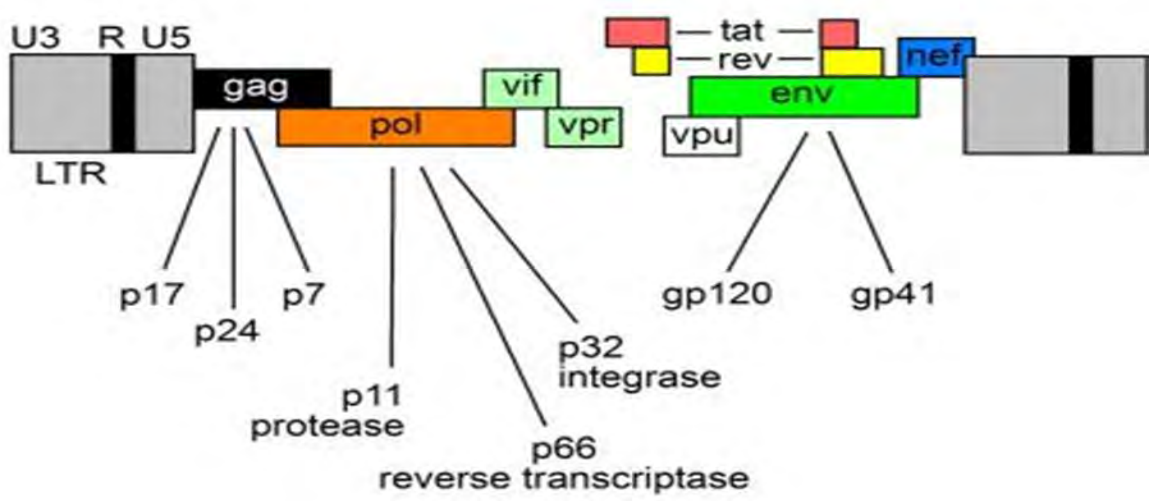


Figure 1.3: The structure illustrating the organization of the HIV-1 genome. Source http://hivbook.com/tag/viral-genome/

1.1.4 The Life Cycle of HIV-1

The first step in the life cycle of HIV-1 is the viral attachment. The HIV-1 envelope gp120 binds to the CD4 cell receptor (Arts and Hazuda, 2012). The interaction of the gp120 and the CD4 is followed by the binding to the CC chemokine receptor such as the CCR5 or CXCR4. These receptor binding causes the conformational change in the gp120 which uncovers a hydrophobic domain on the gp41 that facilitates fusion with the cell membrane thus allowing the uncoating of viral core into the host cell cytoplasm (Arts and Hazuda, 2012). The reverse transcriptase enzyme converts the single stranded RNA into double stranded DNA (Arts and Hazuda, 2012, Singh et al., 2010). The cDNA is then transported to the nucleus where the viral DNA is integrated with the host cell DNA by the enzyme integrase (Arts and Hazuda, 2012, Craigie and Bushman, 2012).

After integration, the provirus is transcribed by the RNA Polymerase II into mRNAs (Sierra et al., 2005). The mRNAs are translated into structural components, enzymes and genomic RNA which are transported to the cellular membrane and assemble. Immature virions are released (Sierra et al 2005). The viral PR cleaves the Gag and Gag pol polyprotein into mature Gag and Pol proteins (Sierra et al., 2005). The life cycle of HIV-1 is summarized in figure 1.4 below:

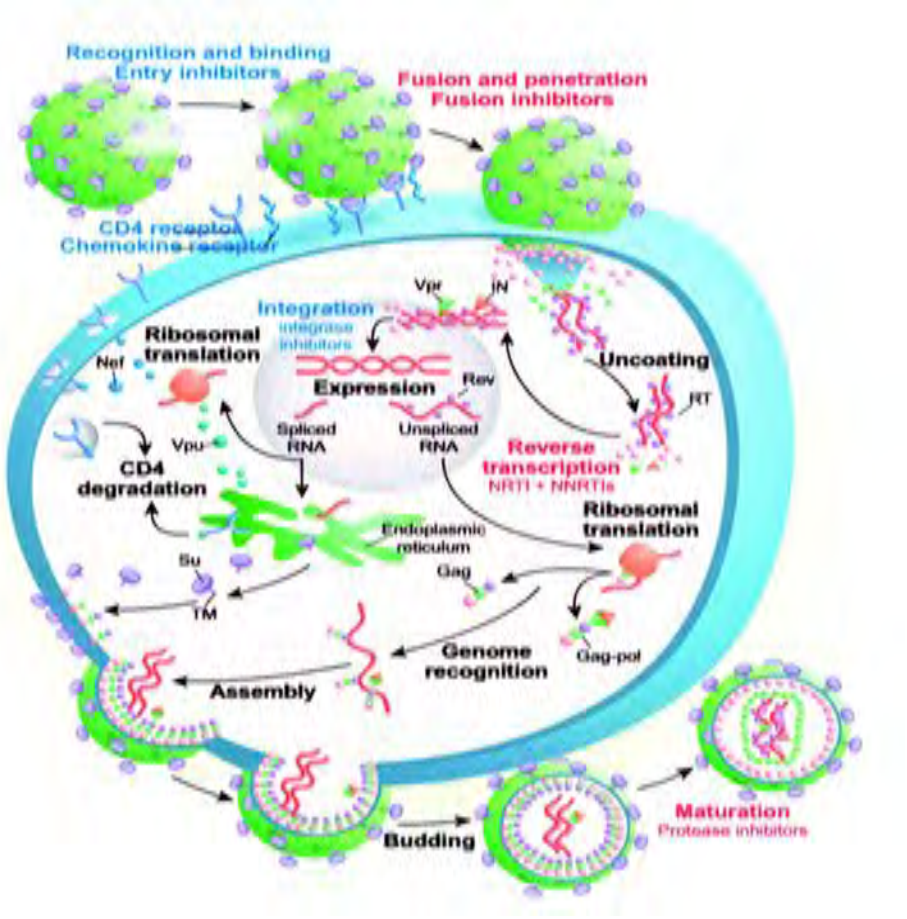


Figure 1.4: The viral life cycle of HIV. The virus binds to the CD4 receptor and to appropriate chemokine receptors resulting in the release of the viral core into the host cell cytoplasm. The viral RNA is converted into double stranded DNA and released into the nucleus where it is integrated into the host cell DNA. The DNA is transcribed to form viral proteins which are transported to the cellular membrane and assemble. The immature virus buds off the cell and mature proteins are produced (Sierra et al., 2005). Source: (Pomerantz and Horn, 2003).

1.2. Antiretroviral therapy

ARV therapy is used to suppress the replication of the HIV-1 virus by affecting the crucial steps involved in viral replication (Martinez-Cajas and Wainberg, 2008). The initial targets for ARV therapy are Protease (PR) and Reverse transcriptase (RT) due to their essential role in the HIV-1 replication cycle (Shafer, 2002). To date, there are five classes of inhibitors: Reverse transcriptase Inhibitors (RTIs), Protease Inhibitors (PIs), Fusion Inhibitors, Integrase Inhibitors, and CCR5 receptor antagonistic (Imaz et al., 2011, Tang and Shafer, 2012). Both ARV

experienced and naïve patients are treated with the Nucleoside and Nucleotide Reverse transcriptase inhibitors (NRTIs), Non-Nucleoside reverse transcriptase inhibitors (NNRTIs) and Protease inhibitors (PIs) (Martinez-Cajas and Wainberg, 2008). The Fusion, integrase and entry inhibitors are restricted to ARV experienced patients with a history of treatment failures and who are facing ARV exhaustion and who need greater access to new classes of inhibitors (Imaz et al., 2011, Martinez-Cajas and Wainberg, 2008).

The use of a combination antiretroviral therapy (cART) previously referred to as Highly Active Antiretroviral Therapy (HAART) has been considered the most successful therapy in maintaining long term suppression (Johnson et al., 2008). The combination is usually two NRTIs and one PI or NNRTI (Shafer, 2002). Ideally, these drugs should provide long term viral suppression in infected individuals, but this is unfortunately not always the case (Martinez-Cajas and Wainberg, 2008). The development of drug resistance has been an ongoing public health concern.

1.2.1 Antiretroviral therapy in South Africa

The main goal of ARV therapy is to reduce HIV related morbidity and mortality. According to the recent South African ARV treatment guidelines, the selection criteria for ARV therapy initiation in adults and adolescents include CD4 cell count ≤ 350 cells/mm^3 irrespective of their WHO clinical stage, patients classified as WHO stage 3 or 4 irrespective of CD4 count, TB patients irrespective of CD4 count, as well as patients that require fast tracking, (i.e. starting treatment within 7 days of being eligible (Department of Health, 2013). The latter include HIV-1 positive women who are pregnant or breast feeding or patients with a CD4 < 200 cells/mm^3 or classified as WHO stage 4.

The selection criteria for infants and children include: all children less than 5 years of age irrespective of their CD4 cell count as well children of 5 years to 15 years with WHO clinical stage 3 or CD4 ≤ 350 cells/mm^3 are eligible to start ARV therapy. In addition, children that require fast tracking are also eligible to start cART. These include children < 1 year of age, with a CD4 count <200 cells/mm^3, or WHO clinical stage 4 or MDR, XDR-TB (Department of Health, 2013). The current recommended ART regimens for adults and adolescent as well as for infants and children are represented in table 1.1 and 1. 2.

Table 1.1: Standardized ART regimens for adults and adolescents 2013 (Department of health, 2013)

1st line	
All new patients needing treatment, including pregnant women	TDF +FTC (or 3TC) + EFV FDC preferred
Adolescents	ABC + 3TC+ EFV
Contraindication to EFV	TDF + (FTC or 3TC) + NVP
Contraindication to TDF	AZT +3TC + EFV or (NVP)
Contraindication to TDF and AZT	d4T + 3TC +EFV (or NVP)
Contraindication to TDF, AZT, and d4T	ABC +3TC +EFV (or NVP)
Currently on d4T- based regimen	TDF + FTC (or 3TC) FDC preferred

2nd line	
Failing on TDF-based 1st line regimen (patients with anemia and renal failure switch to ABC)	AZT + 3TC +LPV/r
Failing on d4T- based 1st line regimen	TDF + 3TC (or FTC) and LPV/r
Dyslipidaemia or intractable diarrhea associated with LPV/r	Switch LPV/r to ATV/r

3rd line	
Failing any 2nd line regimen	Specialist referral
Patients failing on 2nd line therapy are managed by an expert panel and the 3rd line drugs are managed centrally	Ratelgravir/Darunavir/ Etravirine adjusted according to genotype interpretation.

Table 1.2: Standardized ART regimens for infants and children 2013 (Department of health, 2013).

1st line regimen	
All infants and children under 3 years (or < 10 kg)	ABC + 3TC +LPV/r
Children ≥ 3 years (and ≥ 10kg)	ABC+ 3TC +EFV
Currently on d4T- based regimen	Changed d4T to ABC if viral load is undetectable. If viral load > 1000 copies /ml manage as treatment failure

2nd line regimen	
Failed 1st line PI- based regimen	Recommended 2nd line regimen
ABC + 3TC + LPV/r	
d4T + 3TC + LPV/r	Consult with expert for advise
Unboosted PI-based regimen	
Failed 1st line NNRTI based regimen	
Failed 1st line NNRTI –based regimen	Recommended 2nd line regimen
ABC + 3TC +EFV (or NVP)	AZT +3TC +LPV/r
d4T + 3TC + EFV (or NVP)	AZT + ABC +LPV/r

3rd line regimens	
Failing any 2nd line regimen	Regimen based on genotype resistance testing, expert opinion and supervised care

When patients are experiencing treatment failure, they are assed for virological, immunological and clinical failure before they change regimen. Virological failure is defined as an increase of HIV-1 viral load over 1000 copies/ml after initial virological suppression (HIV-1 viral load < 400 copies/ml). Immunological failure is defined as decrease in CD4 count to 100 cell/mm^2 after

six months of therapy or 50% decrease from the initial treatment peak CD4 count value (Schoffelen et al., 2013). Clinical failure is defined as the progression of disease with the development of opportunistic infections. Patients who are experiencing virological failure to their 1st line regimen regardless of good adherence are changed to the 2nd line regimen (Department of Health, 2004). Patients failing their 2nd line regimen are switch to a 3rd line regimen based on specialist recommendation (Department of Health, 2013).

A new fixed-dose drug combination (FDC) was implemented in South Africa from the 1 April 2013. The FDC is ―one ARV pill" consisting of: Tenofovir (TDF), Emtracibine (FTC) and Efavirenz (EFV). FDC is now used as the 1st line regimen, however at the moment, first priority is given to pregnant and breast feeding women. While the advantage of a FDC is that it will limit the problem of taking multiple pills and improve adherence, side effects continue to be a problem (Department of Health, 2013).

1.3 HIV-1 drug resistance

HIV-1 drug resistance results from the accumulation of mutations in the viral genes which are targeted by the drug (Charpentier et al., 2004). These mutations occur due to the high replication error rate of RT caused by the lack of a proof reading mechanism (3‘ to 5‘ exonuclease activity). Thus mutations can occur every time the virus replicates (estimated rate of 1×10^{-9} to 1×10^{-12} mutations per cycle) resulting in a quasispecies of viral variants of related but distinctly different viruses (Abram et al., 2010, Halvas et al., 2010). Drug resistant variants emerge due to prolonged treatment and high viral fitness of the resistant variants over the wild type or by suboptimal treatment which allows continued replication of the virus. It can also be caused by transmission of resistant variants to treatment naïve patients (Martinez-Cajas and Wainberg, 2008). For most ARV drugs, resistance requires the presence of more than one mutation in the gene, and these mutations can occur continuously even if the patient is on therapy (Charpentier et al., 2004).

The mutations that cause resistance to RTIs and PIs are classified as primary or secondary mutations. Primary mutations are single mutations and appear first. They results in low sensitivity to one or more inhibitors. Secondary mutations develop later and increase resistance and viral fitness when in combination with other mutations (i.e. they cannot cause resistance on their own) (Clavel and Mammano, 2010, Nyombi et al., 2008).

1.4 HIV-1 Reverse Transcriptase

HIV-1 Reverse transcriptase (RT) enzyme is involved in RNA-dependent and DNA dependent DNA polymerization (Amiel et al., 2011, Shafer, 2002). The two main activities of the RT are polymerization and Ribonuclease activity (Singh et al., 2010). The structure of the HIV RT consists of two subunits: the p66 (p66 KDa) and p51 (p51 KDa) (figure 1.5). The p66 subunit is formed from the Gag-pol polyprotein as a result of protease excision between the PR and RT domain while the p51 is formed by protease excision of the C terminus of p66 (Sierra et al., 2005). Subunit p66 is made up of 560 amino acids and has the polymerase active site and the RNase H activity while p51 has 440 amino acids and has no enzyme activity (Shafer, 2002, Singh et al., 2010). The p66 subunit consists of three subdomains: the ―thumb", ―palm", and ―fingers" (figure1.5). The ―palm" has the polymerase active site and close to the fingers domain is the RNase subdomain (figure 1.5). The polymerase active site and the RNase activity are separated by the connection domain.

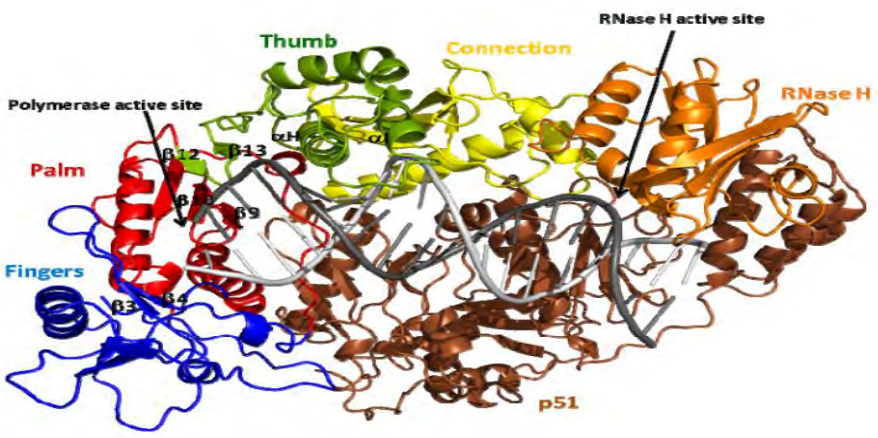


Figure 1.5: The crystal structure of HIV-1 RT illustrating the two subunits: p66 consisting of the palm (red), finger (blue) and the thumb (green) and p51 subunit (dark brown) derived from the RNaseH (orange). Source: (Singh et al., 2010).

1.4.1 HIV-1 Reverse Transcriptase Inhibitors

1.4.1.1 Nucleoside Reverse Transcriptase Inhibitors (NRTIs)

The RT inhibitors are classified into two classes: the nucleoside /nucleotide RT inhibitors (NRTIs) and the non-nucleoside RT inhibitors (NNRTIs). To date there are 7 NRTIs approved

by the US food and drug administration (FDA). This includes six nucleoside inhibitors: Zidovudine (AZT), Lamivudine (3TC), Stavudine (d4T), emtracitabin (FTC), didanosine (ddi), abacavir (ABC) and one nucleotide inhibitor tenofovir (TDF) (Tang and Shafer, 2012). The NRTIs are similar to nucleosides and they are activated by the host cellular enzymes to convert them their tri-phosphorylated form (Shafer, 2002). They therefore compete with the dNTPs and they are incorporated by the RT enzyme in the growing chain of the viral DNA causing chain termination because they lack a 3‘ hydroxyl group hence no additional nucleotide can be attached to them (figure1.6) (Clavel and Hance, 2004).

1.1.4.2 Non -Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)

Currently there are four FDA approved NNRTIs including the Nevaripine (NVP), etavirenz (EFV), etravirine (ETV) and Delavirdine (DLV). However, DLV is rarely used due to complex dosing and limited clinical data (Tang and Shafer, 2012). The NNRTIs prevent HIV-1 replication by binding to the hydrophobic pocket in the RT enzyme, close to the active site. The inhibitor binding affects the flexibility of RT thereby preventing the synthesis of DNA (figure 1.6) (Shafer, 2002).

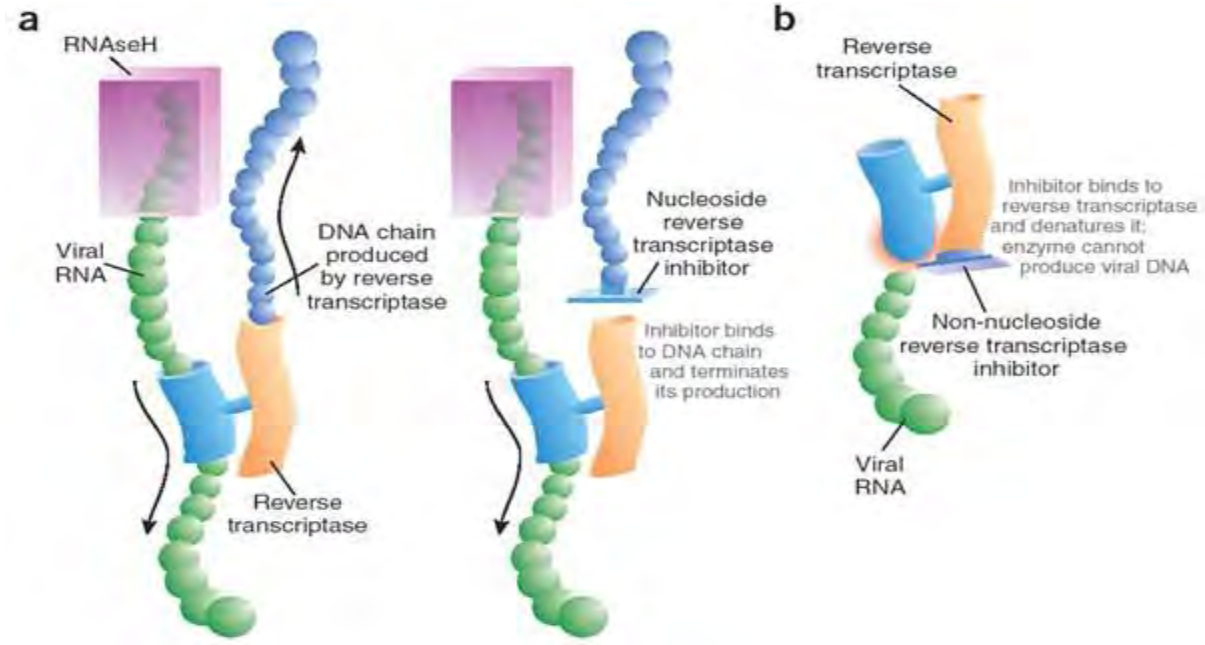


Figure 1.6: Mechanism of action of NRTIs (a) and NNRTIs (b). (a) The NRTIs are similar to the dNTPs however they lack the 3‘ hydroxyl group. When they are incorporated by the RT enzyme in the

growing chain of the viral DNA, they cause chain termination and it production. (b) The NNRTIs directly binds to the RT, close to the active site and affects the flexibility of RT thereby preventing the synthesis of DNA. Source: (Pomerantz and Horn, 2003).

1.4.1.1 Resistance to NRTIs

There are two mechanisms that cause drug resistance to NRTIs (Shafer, 2002). One of the mechanisms is caused by mutations that prevent the NRTIs from being incorporated into the growing DNA chain during syntheses (Figure1.6a). The example of such mechanism is the high level of resistance to 3TC and FTC which is caused by mutation M184V/I (Singh et al., 2010). This mutation also confers resistance to Zalcitabine (ddC), ddi, ABC and has been described to decrease viral fitness (Singh et al., 2010, Tang and Shafer, 2012). Other examples of such mutations include the K65R, L74V and the Q151M complex. The mutation K65R causes resistance to TFV, ddi, ABC and to a lesser extent to 3TC and ddC (Singh et al., 2010). The mutation L74V causes resistance to ddC, ddi and ABC. The Q151M complex cause multidrug resistance to AZT, ddi, ddC, d4T and ABC (Singh et al., 2010).

The other mechanism is caused by the nucleotide excision mutations (NEMs) that cause the removal of the NTRIs, allowing DNA synthesis to continue (figure 1.6b) (Shafer, 2002). The most common excising mutations are: M41L, D67N, K70R, L210W, T25N/F and K219Q/E (Singh et al., 2010). These mutations are also referred to as thymine analogue mutations (TAMs). TAMs are classified into two pathways referred to as type 1 and type 2 (Tang and Shafer, 2012). Type 1 includes mutations M41L, L20I and T215Y. These mutations emerge due to resistance to TDF, ABC, and ddi. Type II includes the D67N, K70R, T215F and K219Q/I/E. It has been reported that patients with many TAMs normally develop additional TAMs when receiving TDF, ABC and ddi instead of developing K65R or L74V which are the mutations associated with Tenofovir, abacavir and didanosine (Tang and Shafer, 2012). In addition, the presence of M184V together with the TAMs has been reported to be the most common pattern that causes resistance to all NRTIs (Tang and Shafer, 2012).

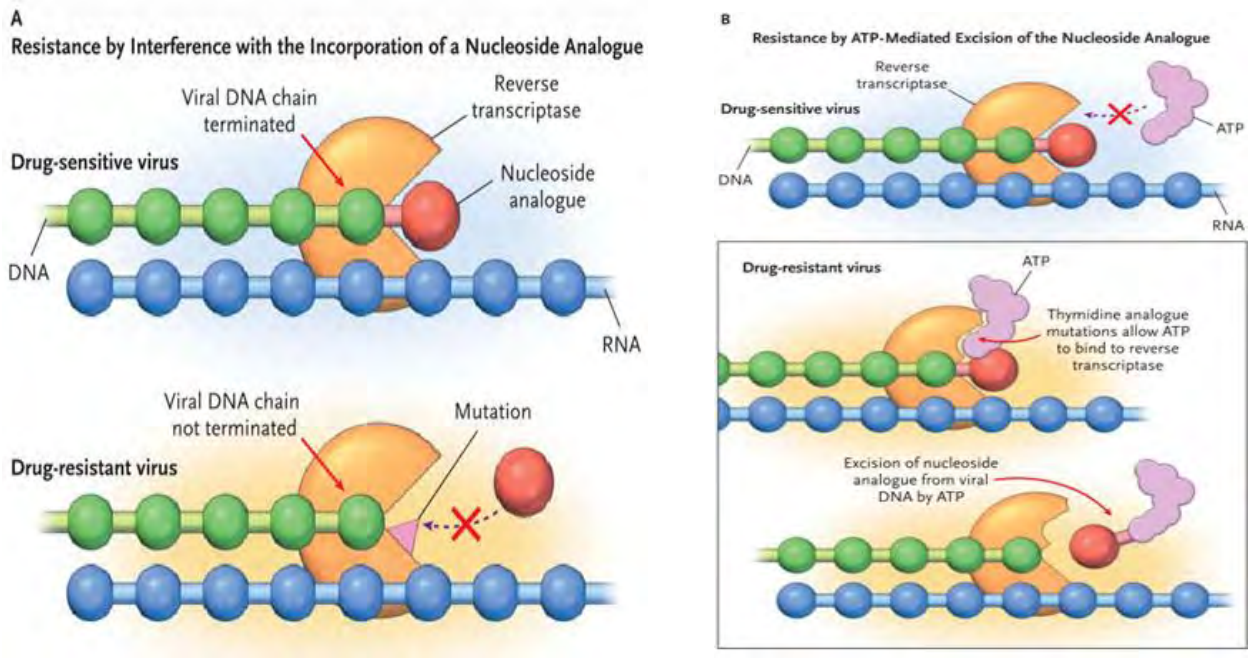


Figure 1.7: The two mechanism of resistance to NRTIs. (a) The incorporation of the nucleoside analogue into drug sensitive virus causes viral DNA termination. The mutations in the drug resistant viruses prevent the incorporation of the nucleoside analogue into the growing viral DNA. (b) ATP is prevented from binding to the RT that has formed a complex with a nucleoside analogue in the drug sensitive virus. The mutations that cause resistance to nucleoside analogues, i.e. TAMs, allow ATP to bind to the RT near the 3 ‗end of the viral DNA which has been terminated by the incorporation of a nucleoside analogue. ATP then excises the analogue, allowing reverse transcription to continue. Source: (Clavel and Hance, 2004).

1.4.1.2 Resistance to Non-Nucleoside Reverse Transcriptase Inhibitors

The mutations that result in NNRTI resistance are located in the hydrophobic pocket which blocks the NNRTIs from entering the binding pocket (Figure1.8). These mutations can result in high level resistance to one or more NNRTIs (Shafer, 2002). NNRTI resistance usually develops when the NNRTIs are used in the presence of incomplete suppression of viral replication suggesting that resistance may be due to preexisting mutants (Shafer, 2002). The commonly described NNRTI mutations include L100I, K103N, V106A, Y181C, Y188C/L and G190A (Singh et al., 2010, Tang and Shafer, 2012). Mutations K103N/S and V106M are the most common NNRTI mutations and they develop due to resistance to etravirine (ETV), and or rilpivirine (RVP) (Tang and Shafer, 2012). Mutation E138K is the recently discovered mutation

that develops in patients experiencing virological failure to RVP. However E138K also cause resistance to ETR, efavirenz (EFV) and nevaripine (NVP).

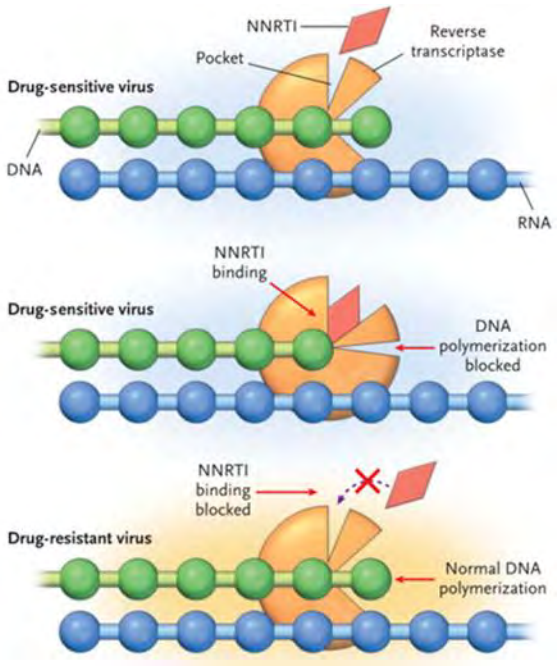


Figure 1.8: The mechanism of resistance to NNRTIs. In the drug sensitive viruses, NNRTIs binds to the pocket next to the active site of the RT and prevents polymerization of DNA by the RT. In drug resistance viruses, mutations block the binding of the NNRTIs, allowing DNA polymerization to continue. Source: (Clavel and Hance, 2004).

1.5 HIV-1 Protease

HIV-1 protease is a homodimeric aspartyl enzyme consisting of two monomers, each 99 amino acids long (Martinez-Cajas and Wainberg, 2008). The PR consists of three domains: the active site, the dimerization and the flap domain (Figure1.9). The active site confers the hydrophilic activity of the enzyme while the flap domain allows the entry of the large-gag poly proteins. The dimerization domain is responsible for dimer formation as well as stabilization of an active PR (Martinez-Cajas and Wainberg, 2008).

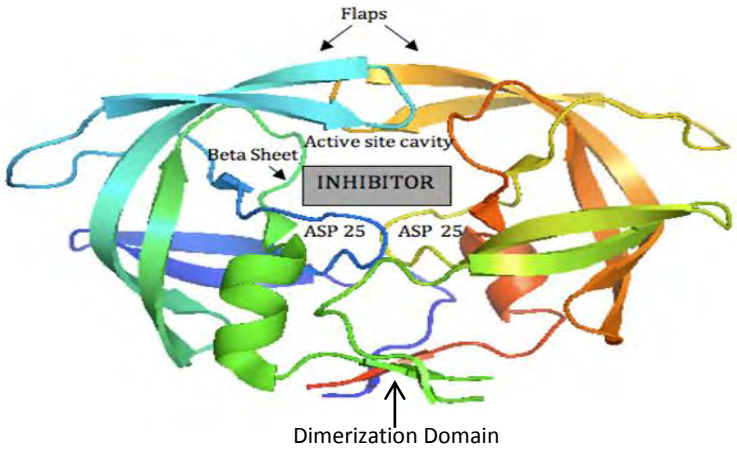


Figure1.9: The crystal structure of the HIV-1 PR. Inhibitors fit inside the active site cavity. The flaps are flexible structures that allow the entry of the large-gag poly proteins. The aspartates (ASP 25) allow hydrolytic activity of the enzyme. Source: (Martinez-Cajas and Wainberg, 2008).

The PR enzyme is responsible for the formation of enzymes and proteins of the virus through the process of post translational processing of the viral Gag and Gag-pol (Clavel and Mammano, 2010). The PR cleaves Gag into p17 matrix (MA), p24 capsid (CA), p2 (SP1), p7 nucleocapsid (NC) and p6 ^{gag}. The Gag–pol polyprotein is cleaved into p17 (MA), p24 (CA), p2 (SP1), p7 (NC), Transframe protein (TFP), p6 pol, Protease (PR), reverse transcriptase (RT), integrase (IN) and RT-RNase H (Ghosn et al., 2011, Larrouy et al., 2010) Figure 1.10.

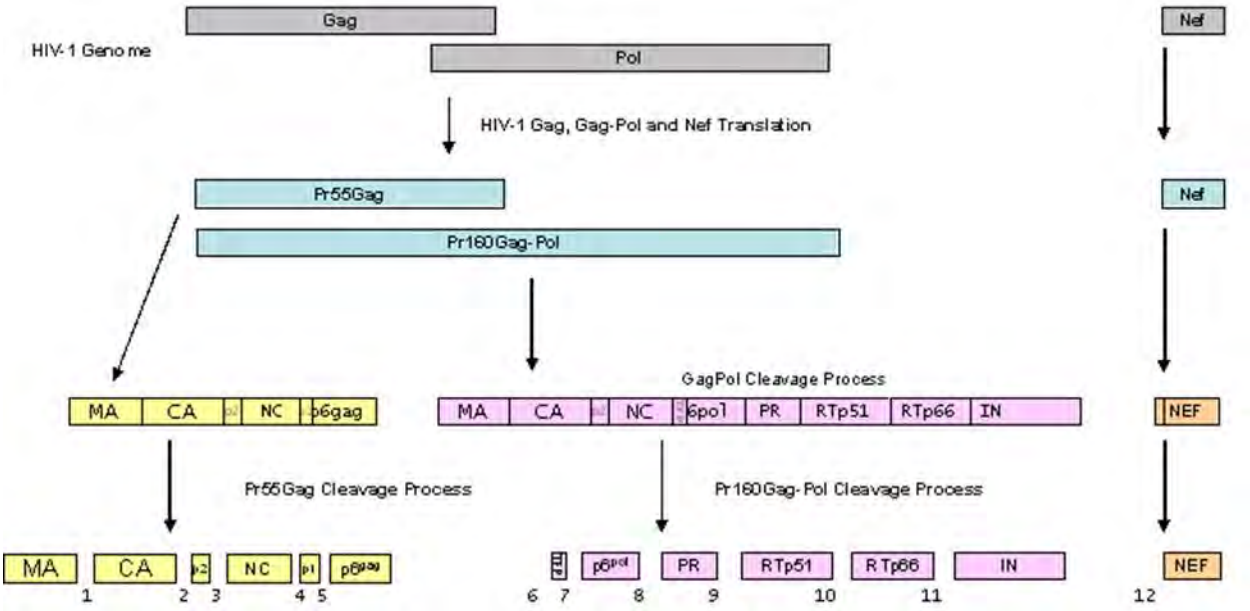


Figure 1.10: Diagram illustrating gag and gag pol processing sites. The Gag is cleaved into 5 sites (MA, p24 capsid /CA, p2 /SP1, p7/NC and p6 gag) while Gag-pol is cleaved into 6 sites (p7 /NC, TFP, $p6^{pol}$, PR, RT, IN) and single site Nef. Source: (de Oliveira et al., 2003).

1.5.1 HIV-1 Protease Inhibitors

PIs are similar in the structure to the PR substrate, thus competing with it for binding to the PR enzyme active site (Shafer, 2002). They therefore inhibit the protease enzyme function and the formation of the viral proteins (Figure1.10) (Clavel and Mammano, 2010, Shafer, 2002).

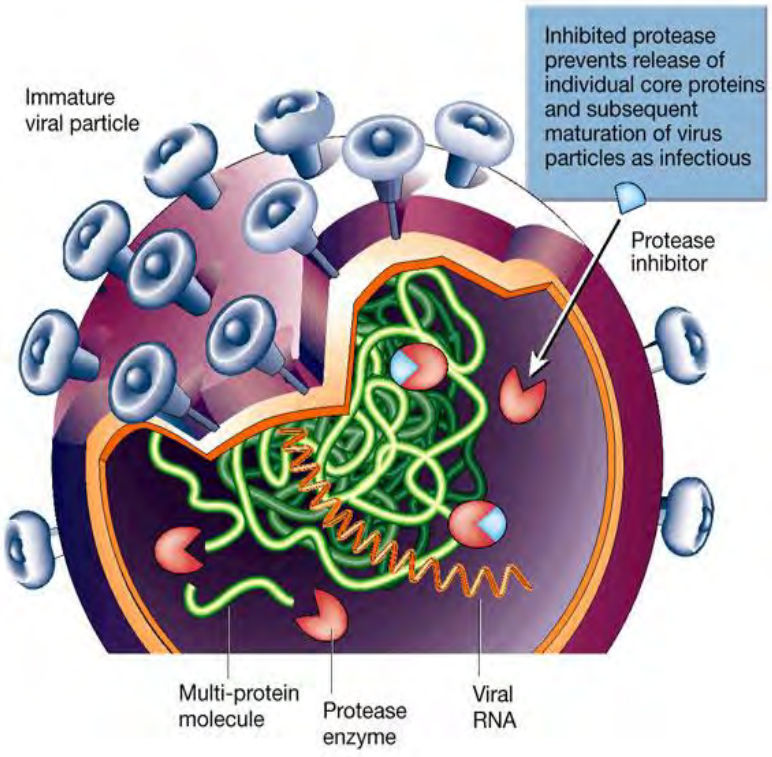


Figure 1.11: Mechanism of action Protease inhibitors (PIs). The protease inhibitor targets the protease enzyme by preventing the cleavage of the polypeptide and maturation of the virion. Source: (Richman, 2001).

To date there are nine FDA approved PIs: Saquinavir (SQV), Indinavir (IDV), Ritonavir (RTV), Nelfinar (NFV), Amprenavir (APV), Lopinavir (LPV), Atazanavir (ATV), Tipranovir (TPV) and Darunavir (DRV). These PIs are essential components of HAART due to their potency compared to the other ARVs (Ali et al., 2010). All of these PIs are boosted with RTV which is used to inhibit cytochrome P450 (cyp) 3A, except for NFV which has low antiviral efficacy (Tang and Shafer, 2012).

1.5.2 Resistance to Protease Inhibitors

Resistance to the PIs is caused by mutations in the substrate cleft which decreases the binding between mutant protease and the inhibitor. The development of PI resistance is generally a gradual process and the increase level of resistance requires the presence of many mutations in the gene itself (Charpentier et al., 2004, Shafer, 2002).

The primary mutations occur at the amino acids which are found inside the substrate cavity, thereby causing resistance by decreasing the binding affinity between the inhibitor and the mutant PR enzyme. The secondary mutations occur at amino acids which are found outside the substrate cavity, and can either compensate for the mutations found at the active site or decrease activity of the mutant protease (Amiel et al., 2011, Clavel and Hance, 2004, Shafer, 2002). The most characterized PI mutation that has been observed in patients who have developed resistance to various PIs is V82A (Clavel and Mammano, 2010). The other mutations include the D30N, V32I, M46I/L, G48V/M, I50V/L, I54V/T/J/M, L76V, I84V, N88S and L90M (Tang and Shafer, 2012). All these mutations cause resistance to one or more PIs except for D30N, and I50L which cause resistance to NFV and ATV. Most of these mutations decrease the replication capacity of the virus as the processing of the natural substrate (Gag) is decreased (Dam et al., 2009). However, drug resistance to PIs doesn‘t develop at the PR gene only but also in protease cleavage sites consisting of Gag and gag- pol polyproteins (Clavel and Mammano, 2010).

1.5.3 The Contribution of Gag Mutations to Protease Inhibitor Resistance

The process of post translational processing of the viral gag polyprotein is a crucial step in HIV viral replication (Clavel and Mammano, 2010). The gag polyprotein is the main substrate of PR and it has been shown that mutations which are located in the NC/SP2/p6 gag region play a role in the development of PI resistance (Clavel and Mammano, 2010, Dam et al., 2009). However, these gag cleavage site mutations also compensate for the viral fitness loss by facilitating the mutated PR to cleave the gag substrate (Clavel and Mammano, 2010, Dam et al., 2009, Kozisek et al., 2012).

In the study by Dam et al., 2009, the mutations that they observed to contribute to the evolution of PI resistance were, I437V and P453L at sites NC/SP and SP2/p6 respectively. In another study by Clavel and Mammano 2010, they showed that mutations L449F at P1 of the SP2/p6 cleavage site, Q430R at P3‘ and A431V at P2 (NC/SP2) also contributed to the evolution of PI resistance. They also indicated that there was a correlation between the mutations at the PR and those at the gag region. The gag mutations A431V and I437V were associated with PI mutations V82A and/or M46I and gag mutation L449F associated with I84V (Clavel and Mammano, 2010). However, some of the gag mutations occur even when the PR mutations are not observed (Ghosn et al., 2011, Nijhuis et al., 2007). In a study by Nijhuis et al 2007, they showed that NC/P1

cleavage site mutation K436E and or I437T/V were present in all the resistance viruses without PR mutations. However, when these mutations were introduced into the wild type they caused PI resistance in all the viruses that had no primary mutations. This indicates that HIV-1 can cause resistance to PIs by changing the PR substrate instead of the PR itself (Nijhuis et al., 2007).

1.6 Drug Resistance Testing

The goal of drug resistance testing is to provide information to help with the selection of effective ARV more likely to attain and maintain viral suppression (Paredes i Deiros, 2009). There are two types of resistance methods i.e. genotypic and phenotypic methods (Hirsch et al., 2008).

1.6.1 Genotypic methods

Genotypic resistance methods detect the presence of drug resistance mutations in the viral enzymes (PR, RT and IN) which are targeted by the ARV drug. The standard approach of genotypic methods involves amplifying specific regions of the HIV-1 genome, followed by direct sequencing of the amplified products by Sanger sequencing (Tang and Shafer, 2012). There are two ways in which genotypic testing can be performed: ―in-house" genotypic testing or using commercial kits (Paredes i Deiros, 2009).

The commercial kits include the TruGene HIV genotyping system assay (Siemens USA), Celera ViroSeq HIV-1 Genotyping system and the GeneSeq HIV (Monogram Biosciences) (Paredes i Deiros, 2009, Tang and Shafer, 2012). ―In-house" resistance genotyping is much cheaper than commercial kits. However, in-house testing is not standardized and inter-subject variability is likely to happen (Paredes i Deiros, 2009).

In clinical practice, the genotypic methods are more commonly used than the phenotypic methods because they are much cheaper, have less turnaround time and they have the ability to detect mutations occurring in mixtures which cannot be done by phenotypic methods (Charpentier et al., 2004, Shafer, 2002, Tang and Shafer, 2012). However, standard genotyping detects mutations in the major (dominant) viral population, but is incapable of detecting minority mutations (Gianella and Richman, 2010).

1.6.2 Phenotypic methods

Phenotypic methods involve the direct measure of mutations contained in the virus on drug susceptibility (Gianella and Richman, 2010). Therefore, the results of drug susceptibility are reported as the level of fold change in susceptibility of the investigated virus compared with a control virus (Tang and Shafer, 2012).

There are two phenotypic methods that are commercially available i.e. the Phenosense (Monogram Biosciences) and Antivirogram (Virco) (Tang and Shafer, 2012). The Phenosense assay is more sensitive and reproducible than the Antivirogram assay. However phenotypic methods are expensive and time consuming compared to genotyping methods.

Again, the phenotypic methods also only detected mutations in the major viral population while the detection of minority variants is limited (Gianella and Richman, 2010, Halvas et al., 2010, Metzner et al., 2005).

1.6.3 Testing for minority variants

As minority variants cannot be detected and quantified by standard genotyping, more sensitive methods have been developed to detect minority variants. This includes the standard cloning, Single Genome Sequencing (SGS), Allele Specific PCR (ASPCR) and the Ultra Deep Pyro-Sequencing (UDPS) (Halvas et al., 2010, Paredes i Deiros, 2009). Table 1.3 summarizes the principles of these techniques.

1.6.3.1 Standard cloning and sequencing of multiple clones

Cloning of viral fragments has been widely used to study viral populations, mainly in assessing the virus diversity and detection of HIV-1 drug resistance mutations (Eshleman et al., 2003, Ramirez et al., 2013). It allows the analysis of genetic linkage of each detected mutation. However, the sensitivity of this method depends on the number of colonies that are analyzed. Paredes i Derios, 2009 reported that in order to detect minority variants present in 10% of the viral population, at least 30 colonies per sample should be analyzed. Although this method is easy compared to the other methods, it is very time consuming and labor intensive and relatively costly (Liang et al., 2011, Ramirez et al., 2013). Additionally, polymerase induce sequencing errors can complicate the results when sequencing cloned PCR products (Liang et al., 2011).

1.6.3.2 Single genome sequencing

Single genome sequencing (SGS) was developed by Palmer and colleagues (Palmer et al., 2005), based on previous limiting dilution assays (Brown and Cleland, 1996, Holmes et al., 1992, Simmonds et al., 1990, Zhang et al., 1991). This method allows a more refined analysis of the whole gene with the sequencing of the single virus particles obtained from the original HIV-1 particle, allowing the analysis of genetic linkage of each detected mutation. In this method, the number of DNA sequences that can be analyzed per samples is generally 20 to 40 single virus particles per sample, although the number can be increased (Palmer et al., 2005). The sensitivity of SGS in detecting minority variants can be as low to 2% of the viral population (Table1.3), however this method is expensive, time and labor consuming (Paredes i Deiros, 2009).

1.6.3.3 Allele Specific PCR

Allele specific PCR (ASPCR) is the most sensitive method in detecting minority variants (Paredes i Deiros, 2009). It can detect minority variants to as low as 0.003-0.4 % of the viral population (Table1.2). It is less labour intensive, cheap and less time consuming than SGS or clonal sequencing (Paredes i Deiros, 2009). However, ASPCR is very specific and sensitive in detecting minority variants, as it does not give information on other drug resistance mutations present in the virus (Gianella and Richman, 2010). In addition, with ASPCR it is difficult to establish a clinically relevant threshold for minority variant, due to a loss of linearity in the measuring of normal proportions below 0.10-1% (Paredes i Deiros, 2009). However, ASPCR continues to be an important and affordable method to detect minority variants (Paredes i Deiros, 2009). **1.6.3.4 Ultra deep Pyro- Sequencing**

Ultra deep Pyro- Sequencing (UDPS) was developed by researchers at life 454 sciences Roche Co (Brandford, CT, USA) (Margulies et al., 2005). This method uses emulsion PCR together with massive parallel pyrosequencing technique where many sequences of individual molecules generated from RT PCR products are sequenced in a single run (Simen et al., 2009). This type of technology results in huge number of sequence, allowing the detection of multiple drug resistance mutation. The sensitivity to detect minority variants depends on the coverage and depth attained i.e. the average number of times a gene has been sequenced during a run. In

addition, its sensitivity to detect minority variants can be as low as 0.5- 1% in viral population (Paredes i Deiros, 2009).

UDPS also allows for sequencing of the individual template and can determine the genetic linkage within the same viral genome in the same way as cloning and SGS (Gianella and Richman, 2010). In additions, UDPS is simpler and less time consuming. Although UDPS remains one of the most sensitive methods, it is more expensive and generates a lot of data that requires extensive bioinformatics expertise in order to fully analyze the data (Paredes i Deiros, 2009, Ramirez et al., 2013).

Table 1.3: Summary of commonly used techniques to detect HIV-1 minority variants (Paredes i Deiros, 2009).

	Standard Cloning	**Single Genome sequencing (SGS)**	**Allele Specific PCR (ASPCR)**	**Ultra deep Pyrosequencing (UDPS**
Principle	Analysis of single CFUs with individual clones	Massive sequencing of single genome molecules	Differential amplification of mutants vs WT in real time PCR	Massively parallel microfluid Solid surface sequencing of single molecules
Sensitivity	> 10 %	2%	0.003- 0.4%	0.5 %
# mutations	Multiple	Multiple	1	300-400bp
Linked mutations	Yes	Yes	No	Yes
Labor intensity	↑↑↑	↑↑↑↑↑	↑↑	↑
Cost	↑	↑↑↑↑	↑↑	↑↑↑↑↑↑
Best	Experience, PPV	Enables linkage of mutations	S, PPV, NPV, Affordable	Linkage, accuracy, S, Rapidity of results
Worst	S, NPV	Only 1 mutant, SP, effect of polymorphism	Cost, labor intensity,	Short sequence , background noise, SP

S: sensitivity; Sp: Specificity; PPV: Positive predictive value; NPV: negative positive value. (Paredes i Deiros, 2009).

1.7. Minority drug resistance mutations

HIV-1 has a high replication rate and errors are likely to happen. The reverse transcriptase enzyme involved in the replication of RNA viruses has a high error rate due to the lack of a proof reading mechanism (3‘ to 5‘ exonuclease activity). Thus mutations can occur every time the virus replicates (estimate rate of 1×10^{-9} to 1×10^{-12} mutations per cycle) resulting in a quasispecies of viral variants of related but distinctly different viruses (Abram et al., 2010, Halvas et al., 2010). A small proportion of these variants (minority variants) could harbor mutations related to drug resistance.

Minority variants are important as they impact on ARV therapy and some studies have found that these minority mutations can later emerge as the major viral population (Metzner et al., 2005). This can be either because they have developed more resistance mutations, or can be due to partially suppressive ARV therapy, allowing the minority population to have a higher growth advantage over the majority population (Charpentier et al., 2004). Many studies have shown that minority drug resistance mutations can exits both in treatment naïve and treatment experience patients (Gianella and Richman, 2010). But as expected, these minority mutations are more frequently detected in treatment experience patients and may be the cause of virological failure (Gianella and Richman, 2010).

In a study by Metzner et al., 2005, quantitative real time PCR was used to detect minority variants using allele discriminator oligonucleotides which were specific for the L90M resistance mutation in PR and K103N and M184V mutations in RT. In five of the ten patients with detectable drug resistance, the detected mutations represented a minor viral population and were not detected by standard genotyping. This suggested that minor HIV-1 drug resistance mutations are underrated when detected only by standard genotyping methods (Metzner et al., 2005).

Roquebert et al., 2006 also did a study on minority variants where they investigated the role of PR and gag cleavage sites in patients failing a PI-inclusive regimen. No PI or gag mutations were detected using standard genotyping before virological failure, however, the D30N and N88D mutations were detected in some of the patients investigated after virological failure. The protease and gag regions were cloned and analyzed. After virological failure, PI minority variants (M46I and L90M) were present among clones which had mutations D30N and N88D.

Gag mutations L449P, S451N and P453L were also present in clones with PI mutations D30N and N88D (Roquebert et al., 2006a). This suggested that HIV can use other mechanisms to cause resistance to PIs by affecting the gag and thus these mutations could compensate for PI resistance (Nijhuis et al., 2007, Roquebert et al., 2006b).

1.8 HIV analysis using online tools

1.8.1 Stanford University HIV Drug Resistance Database (HIVdb)

The Stanford University HIVdb which is also called the HIV RT and PR sequence database is an online database that is used for drug resistance interpretation. It consists of PR and RT sequences from published data on: 1) genotype-treatment correlation, 2) genotype- phenotype correlation and 3) Genotype-outcome correlations (Shafer, 2006). HIVdb aligns the submitted RT and PR sequence and compares it to the subtype B reference strain (HXB2) (Tang and Shafer, 2012). The programme generates a list of drug associated mutations where mutations are assigned drug penalty scores based genotypic-phenotypic correlative data. The programme also gives the level of susceptibility to ARV drugs as well as comments about the ARV resistance mutations found (Shafer, 2002, Tang and Shafer, 2012).

1.8.2 RegaDB sequence analysis tools

RegaDB is a database with software tools, including a sequence analysis tool which is used for drug resistance interpretation (http://jose.med.kuleuven.be/sequencetools). The sequence analysis tool performs a codon-corrected alignment, corrects detected frameshifts and translates the sequences into its 3 open reading frame (ORF). It also performs a Needleman-Wunsch amino acid alignment which calculates the alignment scores. The output of the RegaDB shows the number of short sequences, scores, corrected frameshifts as well as the failed sequences. Finally, it gives a list of tables in XML format showing the amino acids, nucleotides, amino acids mutations and the diagnostic mutations list.

1.8.4 HyPhy package in Datamonkey

Hypothesis testing using phylogenies (HyPhy) is a software package in Datamonkey that performs likelihood based analysis to study the pattern of sequence evolution (Delport et al., 2010, Kosakovsky Pond and Frost, 2005). One of the standard analyses methods found in the

HyPhy package is the Positive selection analysis. HyPhy uses 4 approaches to determine sites under Positive selections. This includes the single likelihood ancestor counting (SLAC), fixed effect likelihood (FEL), internal branch FEL (IFEL) and a random effect likelihood (REL) (Delport et al., 2010, Kosakovsky Pond and Frost, 2005). These methods estimates the rate of nonsynonymous (*dn*) and synonymous (*ds*) changes occurring at each site in the sequence to identify sites under positive or negative selection (Kosakovsky Pond and Frost, 2005).

SLAC is the most conventional counting method that involves reconstruction of the ancestral sequences using a single most likely ancestral reconstruction that considers all possible ancestral reconstructions or sampling from ancestral reconstructions (Kosakovsky Pond and Frost, 2005). The FEL method estimates the substitution rate of nonsynonymous and synonymous at each site of the branch while the IFEL determine the selection pressure that occurs on the internal branches of a tree (Kosakovsky Pond and Frost, 2005). Generally, the FEL and IFEL are the best methods in terms of statistical performance and computational expense. The REL is an extensive codon- based selection analysis that allows nonsynonymous and synonymous rate of variation with the selection pressure at an individual site (Kosakovsky Pond and Frost, 2005).

1.9. An overview of Phylogenetic analysis

Phylogenetic is the scientific study which estimates the evolutionary relationships between groups (e.g. sequences) based on the comparison of amino acids or nucleotides. To represent this relationship, a phylogenetic tree is constructed. A phylogenetic tree is composed of branches (edges) and nodes (figure 1.12). Branches connect the nodes which are the points at which two or more branches diverge. Both the branches and the nodes can be either internal or external. An internal node (hypothetical ancestor) represents everything arising from it while the external nodes (leaf) represent the taxa also called the Operational taxa units (OTUs) (Baldauf, 2003). In addition, phylogenetic trees are drawn with proportional branch lengths whereby the branch length corresponds to the amount of evolution. Thus the longer the branch length the more diversity in the sequence compared to the others (Baldauf, 2003).

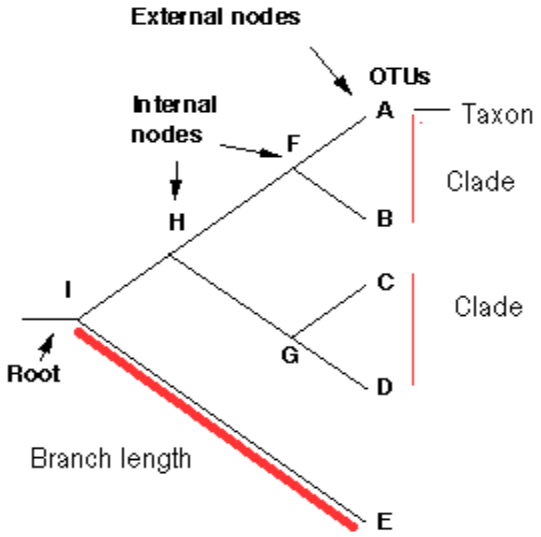


Figure 1.12: Example of a phylogenetic tree. A-E represents external nodes while F-I represent internal nodes (ancestral unit). Clade represents a group of taxa that share a common ancestor (Freed, 2001).

Phylogenetic trees can be rooted or unrooted (Figure 1.13). Rooted trees have a root that indicates common ancestor of all the OTUs thus illustrating the direction of evolutionary process while the unrooted only positions the taxa relative to each other without showing the direction of evolutionary process (Baldauf, 2003, Rizzo and Rouchka, 2007).

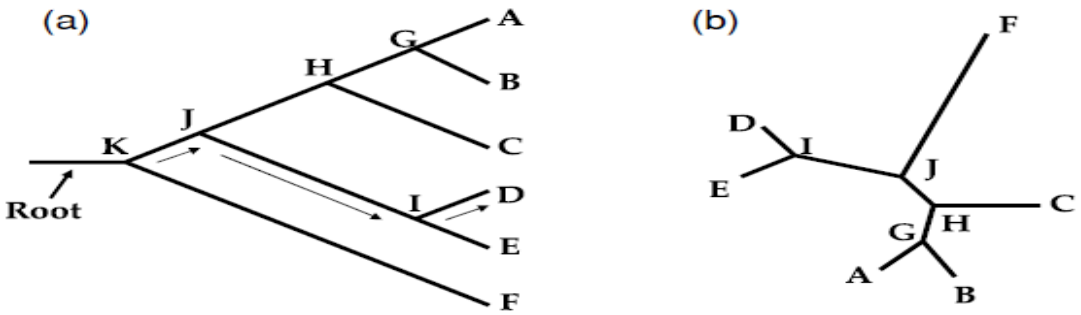


Figure 1.13: (a) Example of a rooted tree and (b) of unrooted tree. Source: (Salemi et al., 2009)

1.9.1 Methods for estimating phylogenetic trees

1.9.1.1 Neighbor joining method

The neighbor joining (NJ) method is the distance method that uses the genetic distance between sequences to construct a phylogenetic tree (Bos and Posada, 2005). This is based on the

hypothesis that the difference between two sequences is directly related to their phylogenetic relationship (San Mauro and Agorreta, 2010). The difference between the sequences is due to a number of changes that have occurred along the branches, that is evolutionary distance (San Mauro and Agorreta, 2010). The genetic distance can also be the differences between the pair of sequences based on the transition (purines to purines or pyrimidines to pyrimidines) and transversion (pyrimidine to purines) substitution rate (figure 1.14) (Bos and Posada, 2005). An example of a distance measure is the Kimura‘s 2-parameter model, which is also a substitution model. The distance methods are relatively quicker compared to the other methods. However, they don‘t give information about the sequences as they are distance based (San Mauro and Agorreta, 2010).

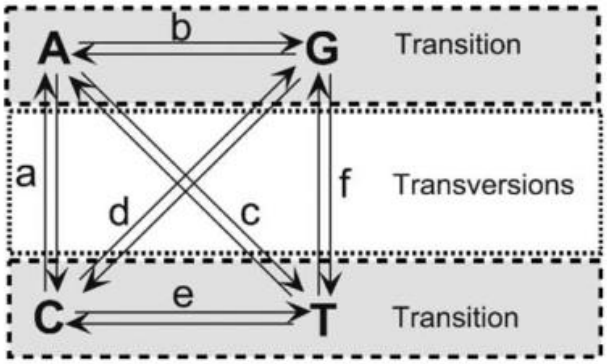


Figure 1.14: A substitution matrix illustrating the two possible different rates of transitions (A-G, C-T) and four possible transversion (A-T, A-C, C-G, and G-T). In this substitution matrix, substitution parameters are reversible, so that the rate of change from nucleotide *I* to nucleotide *j* is the same as rate of change from *j* to *I* (Bos and Posada, 2005).

1.9.1.2. Maximum likelihood methods

Maximum likelihood (ML) uses the sequence data directly to select the tree that best explains the data based on the specific substitution model (Bos and Posada, 2005). The likelihood of a tree is the probability of observing the data provided with the tree and the model of evolution (San Mauro and Agorreta, 2010). The likelihood can thus be estimated for different substitution models and the ML found. The ML has more statistical power in comparison to genetic distance methods, therefore allowing a robust way of estimating phylogenies and understanding sequence evolution (Bos and Posada, 2005, San Mauro and Agorreta, 2010). However, ML methods are

more computational intensive especially when there are large numbers of sequences to be analyzed (Bos and Posada, 2005).

1.10. Nucleotide substitution model

A nucleotide substitution model provides an outline whereby the phylogenetic construction method estimates the parameters used to find the best fit tree. The models differ according to the number of parameters used to represent evolutionary change. Model parameters can represent differences in nucleotide frequencies, substitution rate (figure 1.14) and among site variation. Some of these parameters have been combined, resulting in other models which share some of the parameters (Bos and Posada, 2005). The most commonly used nucleotide substitution models include: the Jukes and Cantor model (JC69), Felsenstein (F81), Kimura 2-parameter (K2P), K80, K81, Hasagawa, Kishino and Yano (HKY85), symmetrical model (SYM), transition model (TIM), transversion model (TVM) and general time reversible (GTR) (Bos and Posada, 2005, Posada and Crandall, 1998).

The JC69 is the simplest model and it considers that all nucleotides substitution have an equal probability. The F81 method determines the probability of nucleotide change using the equilibrium nucleotide frequencies. The K2P uses a substitution matrix that allows for transition and transversion rates (figure 1.14). The K80 model is similar to the K2P, while K81 is similar K3P. The HKY85 allows for different rates in nucleotide pairs i.e. transversion/transition. The GTR has other models nested within it. It allows six substitution rates to have different rates and also allows for different nucleotide substitutions (Gatto et al., 2006, Lio and Goldman, 1998).

1.11 Model selection

The most commonly used methods for selecting the appropriate model includes the Likelihood ratio test (LTR) and the Akaike Information Criterion (AIC) (Bos and Posada, 2005, Posada and Crandall, 1998). These methods provide rigorous statistical support to select and justify the appropriate model (Bos and Posada, 2005). The LRT is calculated by comparing the likelihood scores of the null model (L_0) i.e. the simple model and the alternative model (L_1) that is the more complex and parameter rich model (Posada and Crandall, 1998). The significance of the LTR is also determined.

The AIC calculates the likelihood of the models and gives a score based on the number of model parameters (Bos and Posada, 2005, Posada and Crandall, 1998). The appropriate model is given to the one with the smaller AIC value. The AIC has an important advantage over LRT by comparing all the models, instead of performing sequential pairwise comparison (Bos and Posada, 2005).

1.12 MODELTEST

To test the model of DNA substitution, a MODELTEST is performed. This programme compares the different DNA substitution models using a hierarchical hypothesis testing framework (Posada and Crandall, 1998). The programme calculates the LTR statistics and its associated *P*-values as well as the AIC values (Posada and Crandall, 1998). MODELTEST is performed in a phylogenetic package called PAUP* (Phylogenetic Analysis Using Parsimony). Initially, a file called modelblockPAUPb10.txt is added to the end of the sequence alignment file. This file contains a matrix of log likelihood scores resulting from the execution of a block of PAUP* (Posada and Crandall, 1998). PAUP* then tests the sequence alignment against the 56 models of evolution and gives an output called model. Scores. MODELTEST is then executed to find the best model. The MODELTEST output includes the *P*-values related to the test performed and the *P*-values are interpreted and the best model selected (Posada and Crandall, 1998).

1.13 Aims and objectives of the study

To determine the minority drug resistant variants that were not detected by population sequencing in patients failing HAART

Objectives:

- To clone PCR products that were previously genotyped from patients failing HAART.
- To sequence the clones to detect the minority mutations.
- To perform Ultra Deep Pyrosequencing (UDPS) on selected patients.
- To compare the minority mutations detected in cloning vs UDS.
- To amplify the gag region in these patients.
- To correlate mutations in gag and protease.

Chapter 2

Minority HIV drug resistance variants by cloning

2.1. Introduction

The goal of genotyping resistance testing is to assist physicians /healthcare provider to monitor the patients‘ response to ARV therapy and make choices of successful ARV treatment for patients with a resistant virus (Martinez-Cajas and Wainberg, 2008, Shafer, 2009). The genotyping method that has been routinely used is the standard genotyping (population sequencing) which involves RNA extraction, reverse transcription, PCR amplification and the direct sequencing of the PCR products using the Sanger sequencing approach (Shafer, 2009).

As HIV-1 has a high replication error rate due to the lack of a proof reading mechanism of the reverse transcriptase enzyme, mutations can occur every time the virus replicates resulting in a quasispecies of viral variants of related but distinctly different viruses (Abram et al., 2010, Halvas et al., 2010). A small proportion of these variants (minority variants) could harbor mutations related to drug resistance. Therefore this complicates the analysis of standard genotyping results as these method only detects mutations >20% of the HIV-1 population but not the minority variants which occur at low levels in the HIV-1 population (Johnson and Geretti, 2010). Thus more sensitive methods such as the cloning and sequencing of multiple clones, Allele Specific PCR (AS-PCR), Single Genome Sequencing (SGS) and Ultra-Deep Pyrosequencing (UDPS) have been used to detect the minority variants (Gianella and Richman, 2010, Halvas et al., 2010). In this study, PCR products that were previously genotyped were cloned and multiple clones were sequenced using the Sanger sequencing method. In addition, UDPS was performed to compare minority variants detected by cloning and UDPS. This is an important study as the impact of minority variants related to drug resistance on clinical outcomes remains unclear although there is evidence that these variants can influence treatment efficacy.

2.2. Materials and methods

2.2.1 Study design

This is a retrospective study and this study was approved by the Biomedical Research Ethics Committee of the University of KwaZulu Natal (REF: BE191/12).

Stored plasma samples were obtained from patients failing HAART. These patients were enrolled in the study title: Molecular characterized of HIV-1 protease cleavage steps in South

African patients failing HAART-viral fitness and drug resistance in HIV-1 subtype C. Ethical approval for the study was obtained by the Biomedical Research Ethics Committee of the University of KwaZulu Natal (BREC reference: BF 068/08) as well as McCords Hospital Research ethics committee numbers 2305086/6.1 mg and 280907/4.2 mt.

2.2.2. Study Participants

Six of the stored samples in the study had genotypic data from earlier time points. The samples were genotyped at two time points. They were separated into those who were genotyped at 1st line failure which include (2 NRTIs and 1 NRTI), (1st time point) as well as at 2nd line failure (2 NRTIs and PI) (2nd time point) and those that were genotyped at 2nd line failure only (both time points). The 1st group was studied to investigate the resistance profile before and after starting a PI inclusive regimen while the 2nd group was studied to investigate the resistance profile with prolonged PI therapy. The overview of the study participants is represented in the figure below:

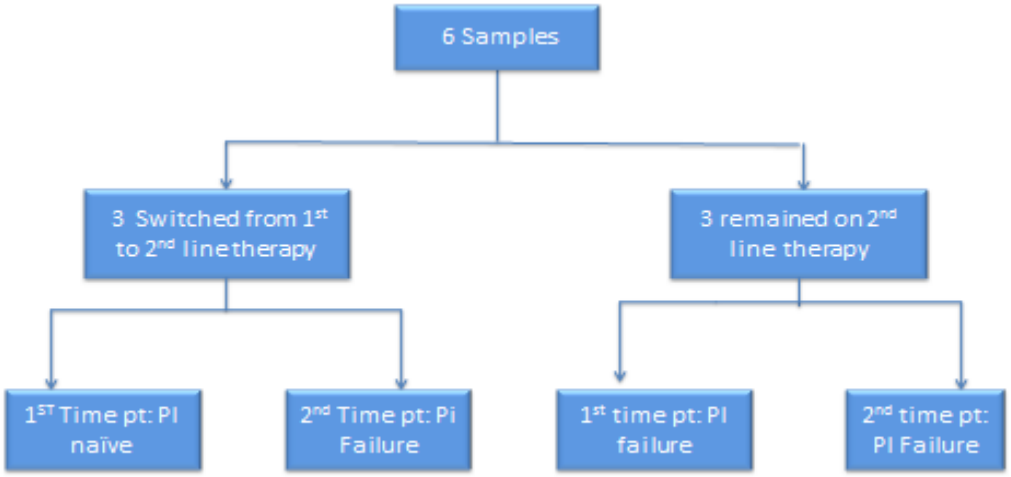


Figure 2.1: Diagram showing the overview of the study participants.

2.2.3. Patients characteristics

2.2.3.1 Characteristic of patients that were switched from 1st to 2nd line therapy.

The 1st line treatment for PCSM004 included D4T, 3TC and EFV from April 2005 before replacing EFV with NVP in October 2007. In July 2008 a 1st genotype was performed and the viral load was 620 cpm while the CD4 cell count was 270 cells/ml. In September 2008, the

patient was then changed to 2nd line treatment (AZT, DDI and Kaletra). The viral load had increased to 7200 cpm while the CD4 cell count decreased to112 cells/ml. In April 2009, a 2nd genotype was performed.

The 1st line treatment for PCSM009 also included D4T, 3TC and EFV from September 2005 to June 2008. The patient was then change to 2nd line treatment (AZT, DDI and Kaletra). The 1st genotype was performed in July 2008. The viral load was 4100 cpm while the CD4 cell count was 204 cells/ml. In July 2009, a 2nd genotype was performed and during this time, the viral loads had increased to 8300 cpm while the CD4 cell count decreased to167 cells/ml.

The 1st line treatment for PCSM012 also included D4T, 3TC and EFV from February 2007 to September 2008. The patient was then changed to 2nd line treatment (AZT, 3TC and Kaletra). In November 2008, a 1st genotype was performed and the viral load was 3700 cpm while the CD4 cell count was 721 cells/ml. In July 2009, a 2nd genotype was performed. During this time, the viral load had increase to 13000 cpm while the CD4 cell count decreased to 82 cells/ml. The summary of these patients‘ characteristics is represented in the table below:

Table 2.1: Characteristics of patients that were switched from 1st to 2nd line therapy.

	1st time point				2nd time point			
Patient	**CD4 cell count (cells/ml)**	**Viral load (cpm)**	**Treatment**	**Time on treatment**	**CD4 cell count (cells/ml)**	**Viral loads (cpm)**	**Treatment**	**Time on treatment**
PCSM004	n/a	n/a	D4T, 3TC, EFV	04/05-10/07	112	72000	AZT, DDI, Kaletra	09/08-04/09
	270	620	D4T, 3TC, NVP	10/07-09/08				
PCSM009	204	4100	D4T,3TC, EFV	09/05- 6/08	167	83000	AZT, DDI, Kaletra	6/08-06/09
PCSM012	721	3700	D4T,3TC, EFV	02/07-09/08	82	130000	AZT, 3TC, Kaletra	09/08-7/09

*CD4 counts and VLs are at the time of genotyping.

2.2.3.2 Characteristic of patients that remained on 2nd line therapy in both time points

The 2nd line treatment for PCSM001 included DDI, AZT and Kaletra from December 2006 before replacing AZT with 3TC in January 2008. In March 2009, the 1st genotyped was performed and during this time point the viral load was 3200 cpm and the CD4 cell count 280 cells/ml. In April 2009, a 2nd genotype was performed and at this time point the viral load had increase to 21000 cpm while the CD4 cell count also increased to 388 cells/ml.

The 2nd line treatment for PCSM002 included D4T, 3TC and Kaletra from March 2008 to April 2009 before replacing D4T and 3TC with AZT and DDI to June 2009. In February 2009, a 1st genotype was performed and the viral load was 84000 cpm and the CD4 cell count was 1562 cells/ml. In April 2009, a 2nd genotype was performed. During this time point, the viral load had decrease to 8000 cpm while the CD4 cell count increased to 1967cells/ml.

Patient PCSM007 has been on 1st line treatment D4T, 3TC and EFV from March 2007 before replacing D4T with AZT in March 2008. The patient was then changed to 2nd line treatment EFV and Kaletra. In March 2009, a 1st genotyped was performed and the viral load was 7900 cpm while the CD4 cell was 480 cells/ml. The 2nd genotype was performed in May 2009. During this time point the viral load increased to 61000 cpm while the CD4 cell count decrease to 179 cells/ml. The summary of these patient‘s characteristics is represented in the table below:

Table 2.2: Characteristic prior the 1st genotype and the 2nd genotype of patients that remained on 2nd line therapy.

	1st time point				2nd time point			
Patient	CD4 cell count (cells/ml)	Viral load (cpm)	Treatment	Time on treatment (prior 1st time point)	CD4 cell count (cells/ml)	Viral load (cpm)	Treatment	Time on treatment
PCSM001	280	3200	AZT, DDI, Kaletra DDI, **3TC**, Kaletra	12/06- 03/09 01/08-03/09	388	21000	DDI, 3TC, Kaletra	03/09-04/09
PCSM002	1562	84000	D4T,3TC, Kaletra	02/08- 03/09	1967	8000	AZT,DDI, Kaletra	03/09-05/09
PCSM007	480	7900	D4T,3TC, EFV **AZT**, 3TC, EFV	03/07-03/08 03/08-10/08	179	610000	EFV and Kaletra	10/08-04/09

2.2.4 Generation of PCR products by In-house resistance genotyping assay

PCR products were generated using an in-house resistance genotyping assay. Briefly, Viral RNA was extracted from plasma using the QIAamp Viral RNA Mini Kit (QIAGEN, Germany) according to the manufacturer‘s protocol. RNA was reverse transcribed to cDNA with Supercript III and amplification was done with primers as previously described in Jennifer Giandhari thesis (2010).

2.2.5 Cloning of the PCR products

Cloning was done in the cloning room. To avoid resampling and PCR bias, 3 PCR products per isolate and time point that were previously genotyped were used for cloning. The PCR products were quantified to optimal concentration of 20 ng/µl before cloning using a Nanodrop 2000 spectrophotometer (Thermo Scientific, South Africa).

The PCR products were cloned into PCR2.1 TOPO vector, following the manufacturer‘s instructions (Invitrogen TOPO TA cloning kit, California). Briefly, in the ligation step the TOPO

cloning vector containing the deoxythymidine (T) overhangs was ligated to the PCR product which has deoxyadenosine (A) sticky ends. This is illustrated in the figure below:

Figure 2.2: Illustration of the ligation step source: www.b2b.invitrogen.com

The ligation reaction was prepared in 0.5ml RNase/DNase free microcentrifuge tubes as follows:

Table 2.3: TOPO cloning reaction.

Reagent	Volume (µl)
PCR product	2
Salt solution	1
Water	2
TOPO cloning vector	1
Final volume	6

The reaction was incubated at room temperature for 30 minutes. For the transformation step, 2 µl of the ligation reaction was added to 1 vial of competent TOPO 10 *E.coli* cells. Care was taken not to mix by pippeting up and down as this would kill the cells. Transformations was carried out on ice for 30 minutes, heat shocked for 30 seconds in a water bath at 42ºC and cooled on ice for 2 minutes. A volume of 250 µl of prewarmed SOC medium was added to the transformation and incubated in a Shaking incubator set at 230 rpm and at 37ºC for 1 hour. While waiting for the 1 hour, 40 µl of X-gal was plated onto prewarmed LB/amp plates and incubated until use. For each

transformation, 100 µl and 50 µl was plated onto two separate LB/amp plates to ensure that one plate would have well-spread colonies. The plates were incubated at 37ºC for 16-18 hours. After incubation the positive transformants were indicated by white colonies.

2.2.6 PCR screening of positive transformants.

A master mix for the PCR screening was prepared using the reagents for the 2nd round of PCR amplification of HIV-1 pol region as described in Jennifer Giandhari thesis (2010). The primers used for the amplification is represented in the table below:

Table 2.4: The HXB2 position and primers used to amplify the 1.7kb of HIV-1 pol.

Primer name	HXB2 position	Primer sequences 5'-3'
AV150 (10pmol/µl)	2036-2062 F	GTGGAAAGGAAGGACACCAAATGAAAG
Pol M4 (10pmol/µl)	3892-3870 R	CTATTAGCTGCCCCATCTACATA

PCR master mix was prepared in a pre-amplification room. The PCR reagents were thawed and vortex for 5 seconds to mix and centrifuged briefly to collect content to the bottom of the tube. The PCR master mix was prepared in a 1.5 ml RNase/DNase free microcentrifuge tube as follows:

Table 2.5: Reagents for the PCR master mix for screening colonies.

Reagents	Volume for 1 reaction (µl)	Final Concentration
10× buffer	2.5	1X
dNTPs (10mM)	0.5	0.2mM
AV150 (10pmol/µl)	0.5	0.2pmol/µl
Pol M4 (10pmol/µl)	0.5	0.2pmol/µl
Amplitaq Gold 5U/µl	0.125	0.025U/µl
$MgCl_2$ 25mM	1.5	1.5mM
DEPC treated water	9.375	-
Final volume	15	-

The PCR Master Mix was kept at -80ºC until added to the reaction tubes. The LB/amp plates that were used as master plates for the colonies were prewarmed in an incubator at 37ºC for 30 minutes. The plates were labeled by drawing 40 grids lines and the grids were labeled 1 to 40.

After incubation, 40 white colonies were picked up using a 10 µl sterile pipette tip and touched on a master plate before proceeding with the PCR screening. The master plates were incubated at 37ºC for 16-18 hours. Thirty colonies per sample were screened. A volume of 15 µl of distilled water was added to each reaction tubes. The colonies were transferred to each reaction tubes and heated in thermocycler for 5 minutes at 99.9ºC. The reaction tubes were centrifuged for 5 minutes at 12000 ×g to separate the supernant from the pellet. A volume of 15 µl of the PCR master mix was added to each PCR reaction tubes. A volume of 10 µl of the supernant of the colonies was added to each reaction tube making it to final volume of 25 µl. Care was taken to open one tube at a time. The PCR screening was run on the thermocycler using the following conditions:

Table 2.6 Cycling conditions for the 1st round PCR.

Temperature	Time	Cycles
94°C	2 minutes	1
94°C	10 seconds	
50°C	30 seconds	10
68°C	2 minutes	
94°C	15 seconds	
52°C	30 seconds	52
68°C	*2 minutes	*plus 20 sec/cycle
68°C	7 minutes	1
4°C	>10 minutes	

After the run, the PCR screening products were run on a 1% agarose gel to verify the presence positive insert with correct size band.

A 1% agarose gel was prepared by adding 0.5 g of agarose tablet to 50 ml of 1×TBE buffer. The agarose was allowed to dissolve and heated in a microwave until all the agarose was completely dissolved and the solution was clear. The solution was cooled and poured into prepared casting trays with combs and the gel was left to set for 30 minutes at room temperature. After 30 minutes, the combs were removed and the gel was placed in the electrophoresis tank. The 1×TBE buffer was poured into the electrophoresis tank up to the level that covered the gel. The

gel loading dye was prepared in a 1.5ml tube by adding 1 µl of the gel red to 50 µl of gel loading. Five µl of the PCR product was mixed with 1µl of the gel loading dye and loaded into the wells. Two µl of 1kb DNA ladder was also mixed with 1 µl of gel loading dye and loaded to the first well of the gel. The gel was run for 45 minutes at 100V on Enduro electrophoresis power supply (Denville Scientific Inc). After the run, the gels were viewed with GelVue UV transilluminator (GENEsnap) and the images were captured. The correct size of the PCR product was estimated by comparing it to the size of the different products of the 1kb DNA ladder as illustrated below:

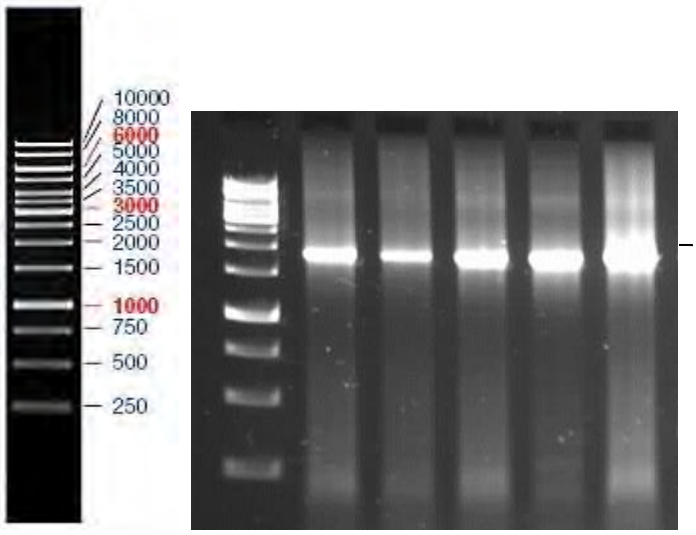


Figure 2:3. An Agarose gel electrophoresis of the PCR screen with the correct size of the insert 1.7 kb (right). O‘ Generuler 1kb DNA ladder (left) showing the size of the different products of the gel/ladder (Invitrogen).

2.2.7 Growing of positive clones and Plasmid DNA (pDNA) extractions

Thirty positive colonies per isolate were grown. This has been statistically proven in other studies that in order to detect minority variants at 10% of the viral population, at least 30 clones have to be analysed.

LB broth was prewarmed in incubator at 37ºC for 30 minutes. A volume of 5ml of LB broth and 5 µl of ampicillin was added to 15ml centrifuge tube. The positive clones were picked from the master plates using quadloop and added to the centrifuge tube with LB and ampicillin. The tubes were place in a rack positioned horizontally in shaking incubator at 37ºC and 230 rpm for 16-18

hours. After incubation, the glycerol stocks for each clone were prepared as follows: The glycerol was heated in a microwave until it was warm. A volume of 150 µl of glycerol was added to 1.5 ml cryovial tubes. A volume of 850 µl of cell suspension was added to the cryovial tubes with glycerol making the total volume to 1000 µl. Care was taken to open one tube at a time. The glycerol stocks were stored at -80ºC for long term use. The remaining cell suspension was centrifuge at 12000 ×g for 2 minutes as follows:

A volume of 1000 µl of cell suspension was added to 1.5 ml tube using a Plastic Pasteur pipette and centrifuge for 2 minutes at 12000 × g. The supernatant was discarded and the pellet was kept. Additional volume of 1000 µl was added to the same tube with the pellet and process was repeated until all the cell suspension was finished. The pellet (pDNA) was purified using GeneJet plasmid miniprep kit following manufactures instructions (fermantas miniprep kit). Distilled water was used to elute the pDNA instead of the provided elution buffer. The Plasmid DNA was stored at -20^{o}C until used.

To confirm the present of both the vector and the insert, a restriction digest was performed. The pDNAs were quantified using a Nanodrop. The restriction digest mix was prepared as follows:

Table 2.7: Reagents for the restriction digest.

Reagent	Volume for 1 reaction (1µl)	Final concentration
10× Ecori Buffer	1	1X
Ecori Enzyme 2000U/µl	0.5	20U/µl
Distilled water	7.5	-
Final volume	9	-

A volume of 9 µl of the mix was added to each reaction tubes. One µl of the pDNA was added to each reaction tube making it to final volume of 10 µl. The reaction was run on the thermocycler using the following conditions:

Table 2.8 Cycling conditions for the restriction digest.

Temperature (^{0}C)	Time (Minutes)
37^{0}C	60
37^{0}C	60
65^{0}C	20

After the run, the restriction digest was analyzed by running in a 1% gel electrophoresis as described in 2.2.5. This is illustrated in the figure below:

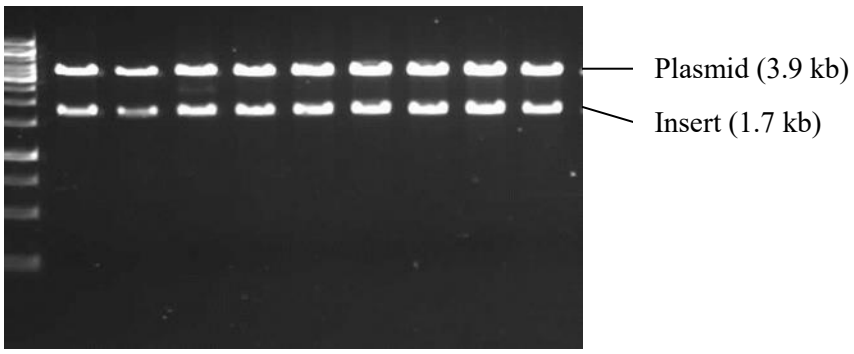


Figure 2.4: An Agarose gel electrophoresis of the restriction digest.

2.2.8 Sequencing of Plasmid DNA (pDNA)

The pDNAs were diluted to 50 ng/µl which is an optimal concentration for sequencing of the pDNA. The HXB2 position and sequences of the 4 primers used for the sequencing of HIV-1 pol are represented in the table below:

Table 2.9: The HXB2 position and primers used to sequence the 1.7 kb of HIV-1 pol.

Primer name	HXB2 Position	Primer sequence (5'-3')
AV150 (3.2 pmol/µl)	2036→2062 F	GTGGAAAGGAAGGACACCAAATGAAAG
Pol M4 (3.2 pmol/µl)	3892←3870 R	CTATTAGCTGCCCCATCTACATA
Pol MG (3.2 pmol/µl)	2823←2798 R	ATTGAACTTCCCAGAAGTCTTGAGTT
Pol M1 (3.2 pmol/µl)	2251→2272 F	GTTAAACAATGGCCATTGACAGA

The figure below illustrates the outline of the 4 primers, showing how the forward and the reverse primers overlap.

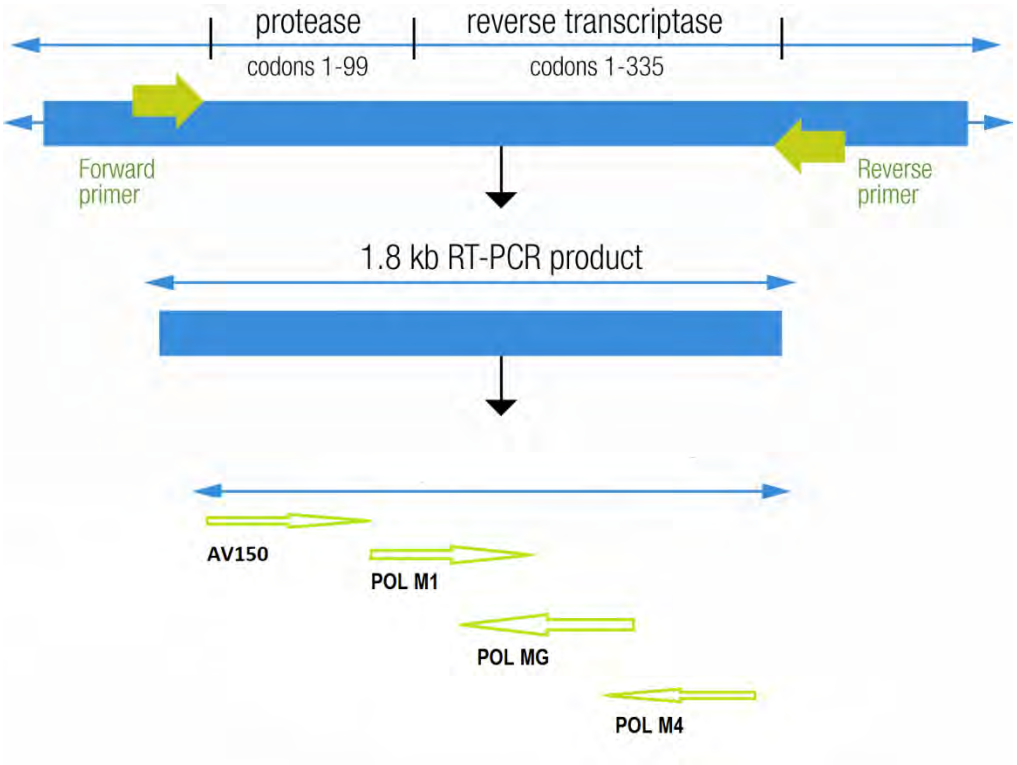


Figure 2.5: Diagram illustrating the layout of primers used for sequencing of HIV-1 Pol gene.

The reagents for the sequencing reaction were thawed, vortex (except for the Ready reaction mix) and centrifuge briefly. A master mix for each primer was prepared as follows:

Table 2.10: Reagents for the sequencing reaction.

Reagents	Volume (µl)	Final concentration
Big Dye ready reaction mix	0.4	-
5×sequencing buffer	2	1×
Primer (3.2 pmol/µl)	1	0.32 pmol/µl
Water	3.6	-
Final Volume	7	-

The sequencing reaction was set up in a MicroAmp optical 96-well reaction plate. A volume of 7 µl of each reaction primer mix was added to the selected wells and 3 µl of diluted PCR product (50 ng/µl) was also added and mixed with pipette tips. The plates were sealed with a covering mat and vortexed briefly. The sequencing reaction was run in the thermocycler under the following conditions:

Table 2.11 Cycling conditions for the sequencing reaction.

Temperature	Time	Cycles
96 ^{0}C	1 min	1
96 ^{0}C	10 sec	
50 ^{0}C	5 sec	35
60 ^{0}C	4 min	
4 ^{0}C	Hold	-

After the run, the sequencing products were purified on the same day. Firstly, a mixture of 3M Sodium acetate (NaOAc) pH 5.2 and 100% Ethanol was prepared as follows:

Reagent	1 well (µl)
3M NaOAc pH 5.2	1
100% Ethanol	25
Final Volume	26

A volume of 1 µl of 125 mM EDTA pH8.0 was added to each well and mixed with pipette tips. A Volume of 26 µl of the mixture prepared above was added to each well using multi- channel pipette tips. The plate was covered with an adhesive foil and was vortex briefly before centrifuged at 3000 × g for 20 minutes using an Eppendorf plate centrifuge. The plate was carefully inverted onto paper towel and centrifuged at 150 × g for 5 minutes. A 70% cold ethanol was prepared and 35 µl of the solution was immediately added to the each well and the plate was covered once again with an adhesive foil and centrifuge at 3000 × g for 5 minutes. The plate was inverted once more onto paper towel and centrifuge for 1 minute at 150 × g and dry in the thermocycler at 50^{o}C for 5 minutes in dark place. The plate was sealed with an adhesive foil and

covered with a foil and stored at -20^{o}C until sequenced. Before sequencing, 10 µl of Formamide was added to the dried pellet in each well and heated in a thermocycler for 3 minutes at 95^{o}C. The plates were loaded onto the ABI 3130 Genetic analyzer (Applied Biosystems, California).

2.2.9 DNA Sequence analysis

Sequence analysis was done on a Linux computer using the Phred and Phraph and consed sequence editing programme. The sequences were aligned using CLUSTALW and manually edited by codon alignment using the Genetic Data Environment (GDE) program or BioEdit. Final consensus sequences were submitted to Stanford drug resistance database & REGADB sequence analysis tools which gave a list of protease and reverse transcriptase mutations associated with RT and PR inhibitors (Chapter 1.8.1 and 1.8.2).

2.2.9.1 Phylogenetic analysis

Preliminary phylogenetic trees were performed for quality control of the sequences i.e. to exclude contamination between samples. Trees were generated using PAUP* 4b.10 by means of the neighbor joining method (Chapter 1.9.1.1). A F84 model of substitution was used. Trees were viewed using figtree v1.4.

Maximum likelihood trees for each patients were constructed in PAUP*4.0b10 using the appropriated evolutionary model selected by the AIC in MODELTEST 3.7. With this method, a neighbor joining tree was calculated using JC69 model of evolution. The likelihood ratios test statistic was also calculated and the programme selected the appropriate model based on the *P*-Values of the likelihood ratio test. Trees were rooted using the ―best fit root" as estimated using the Path-O-Gen v.1.4. Trees were viewed using Figtree v.1.4.

2.2.9.2 Genetic diversity

The genetic diversity (evolutionary distance) within and between the sequences from the two time points was determined using the Maximum Composite Likelihood (MCL) method implemented in MEGA software version 4.0 (Tamura et al., 2007).

2.2.9.3 Analysis of selective pressure.

Positive selection was assessed using HyPhy software package in Datamonkey (chapter 1.7.4). The IFEL model was used to estimate the ratio of nonsynonymous (dN) to synonymous (dS) amino acid substitution as a determinations of natural selection. Positive selection was characterized as dN/dS ratio greater than 1. The HKY85 model of substitution was used with 0.05 significant levels.

2.2.9.4 Identification of the signature amino acid difference between the sequences at two time points using VESPA

The Viral Epidemiology Signature Pattern Analysis (VESPA) analysis tool is used to detect signature amino acid difference between two groups i.e. the query sequences and the background sequences VESPA determine the frequency of amino acids at each position in the aligned sequence in each group. The VESPA report shows the signature sequence that indicates the common positions in the query sequence that differ from the reference sequence. It also shows the signature amino acids frequency and the information for each position in the sequences. http://www.hiv.lanl.gov/content/sequence/VESPA/vespa.html

2.3 Results

2.3. Minority variants of the HIV-1 Pol region.

2.3.1.1 Longitudinal analysis of patients that were switched from 1st to 2nd line therapy

None of these patients developed PI resistance mutations when switched on to PI inclusive regimen. Additional RT mutations that were not detected by the standard genotyping were detected before and after switching to PI inclusive regimen. Specifically, this included NNRTI mutations from the 1st line failure and NRTI from 2nd line failure (PCSM012) and the NNRTI mutations from 2nd line failure (PCSM004 and PCSM009).

2.3.1.1.1 Longitudinal analysis of PCSM004

For PCSM004, a total of 58 HIV-1 Pol gene sequence were analyzed: 30 sequences at the 1st and 28 sequences at the 2nd time point. At the 1st time point, the RT drug resistance mutations detected by standard genotyping included the major NRTI mutations (M41L, M184V) and NNRTI mutations (V106M, V179D, K103R and F227L). No additional mutations were detected by cloning. The frequency of most of these mutations were also high (>97%) at this time point. After nine months, a 2nd genotype was performed. No mutations were detected by standard genotyping at the 2nd time point (Figure 2.6). However, the NNRTI mutations (V106M and V179D) were detected by cloning (Figure 2.6). The frequency of these mutations decreased from 97% to 4% (V106M) and from 97% to 14% (V179D) at the 2nd time point (Table 2.12). No mutations were detected in PR at either time point for both standard genotyping as well as cloning.

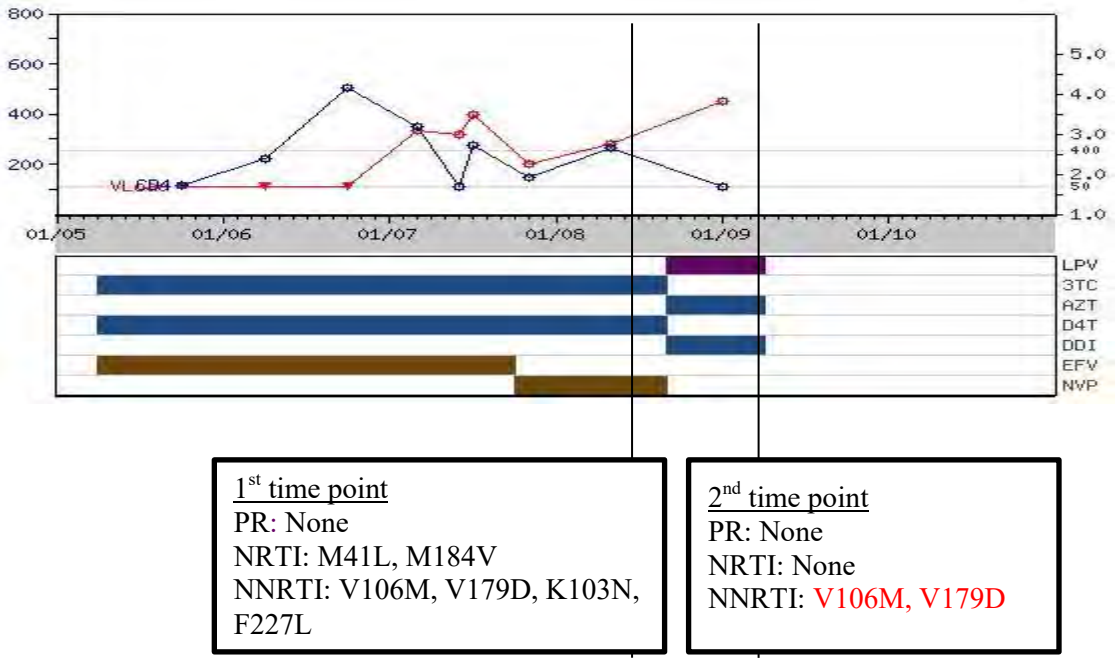


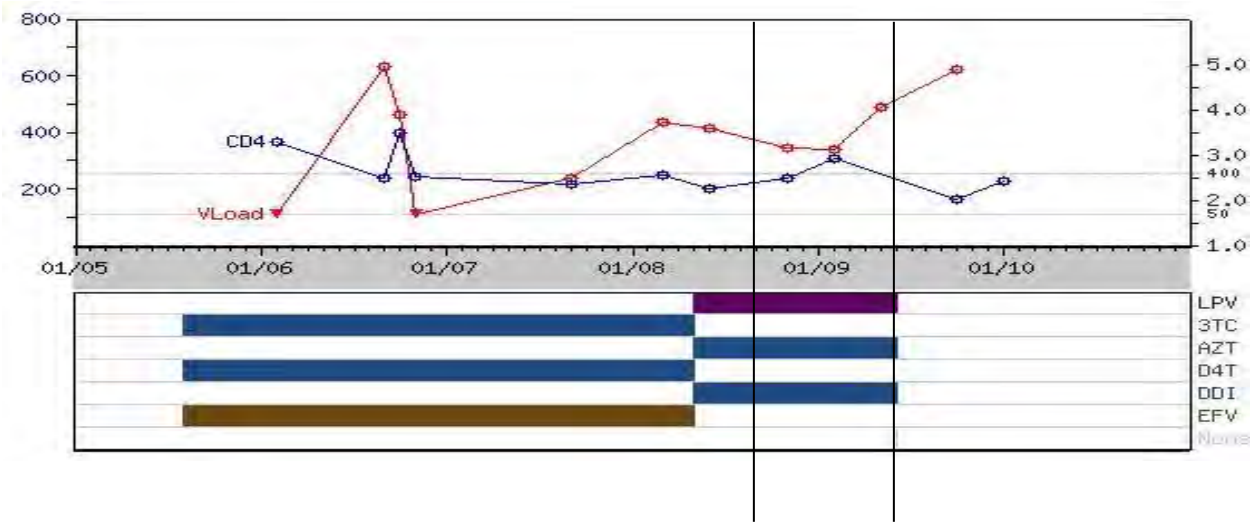
Figure 2.6: Graphical summary of drug resistance history and the resistance profile of PCSM004. The CD4 cell count (blue) is represented on the left y-axis and the viral load (red) is on the right y-axis. The PIs are represented by the purple bars, NRTIs by the blue bars and NNRTIs by the brown bars. The x-axis represents time. The boxes show the mutations detected by standard genotyping and cloning during the two time points. Red mutations are additional mutations detected only by cloning.

Table 2.12: The frequency of mutations at the 1st and 2nd time points for PCSM004.

Mutations	**Frequency of mutation (%)**	
PI	**1st time point**	**2nd time point**
-	-	-
NRTI		
M41L	73	0
M184V	97	0
NNRTI		
V106M	97	4
V179D	97	14
K103R	93	0
F227L	33	0

2.3.1.1.2 Longitudinal analysis of PCSM009

For PCSM009, a total of 53 HIV-1 Pol gene sequence were analyzed: 23 sequences at the 1st and 30 sequences at the 2nd time point. At the 1st time point, the RT drug resistance mutations detected by standard genotyping included the major NRTI mutations (D67N and M184V) and NNRTI mutations (K101E, V106M, E138G/K and G190A). No additional mutations were detected by cloning. The frequency of most of these mutations were also high (> 65%) except for E138G which had a frequency of 9% (Table 2.13) at this time point. After twelve months, a 2nd genotype was performed. No mutations were detected by standard genotyping at the 2nd time point (Figure 2.7). The mutations only detected by cloning included the NRTI mutation D67G and NNRTI mutations (K101E, V106M, E138G and G190A) from the 1st time point (Figure 2.7). However, the frequency of these mutations was low (<10%) (Table 2.13). No mutations were detected in PR at either time point for both standard genotyping as well as cloning.



1st time point
PR: None
NRTI: D67N/G, M184V
NNRTI: K101E, V106M, E138G/K/, G190A

2nd time point
PR: None
NRTI: D67G
NNRTI: K101E, V106M, E138G, G190A

Figure 2.7: Graphical summary of drug resistance history and the resistance profile of PCSM009.

Table 2.13: The frequency of mutations at the 1st and 2nd time points for PCSM009

Mutations	Frequency of mutations (%)	
PI	**1st time point**	**2nd time point**
-	-	-
NRTI		
D67G	0	3
D67N	65	0
M184V	100	0
NNRTI		
K101E	100	10
V106M	100	10
E138G	9	7
E138K	91	0
G190A	100	10

2.3.1.1.3 Longitudinal analysis of PCSM012

For PCSM012, a total of 62 sequences were analyzed: 35 sequences at the 1st and 29 sequences at the 2nd time point. At the 1st time point, the RT drug resistance mutations detected by standard genotyping included the major NNRTI mutation K103N (figure 2.8). Additional mutations detected by cloning included the NRTI mutation T69A and NNRTI mutations: V106M, E138K and G190A. The frequencies of most of these additional mutations were low (≥ 3%) (Table 2.14). After eight months, a 2nd genotype was performed. The NRTI mutation T69A and NNRTI mutation K103N were still detected, with an increase in frequency from 14%- 62% (T69A) and 26% to 66% for K103N (Table 2.14). The NRTI mutation T69D was detected by standard genotyping at this time point, however the frequency of this mutation in the clones was low (3%) (Table 2.14). The V106M, E138K and G190A mutations were not detected at this time point. No mutations were detected in PR at either time points for both standard genotyping as well as cloning. .

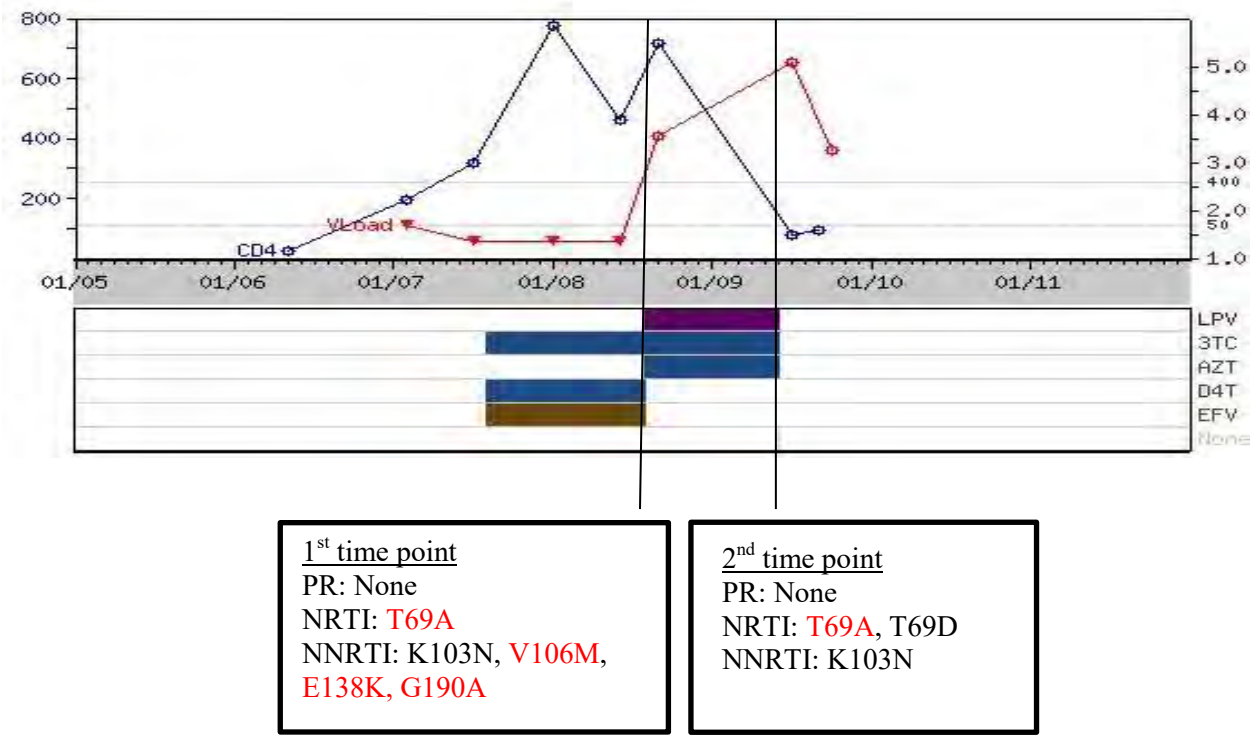


Figure 2.8: Graphical summary of drug resistance history and the resistance profile of PCSM012.

Table 2.14: The frequency of mutations at the 1st and 2nd time points for PCSM012.

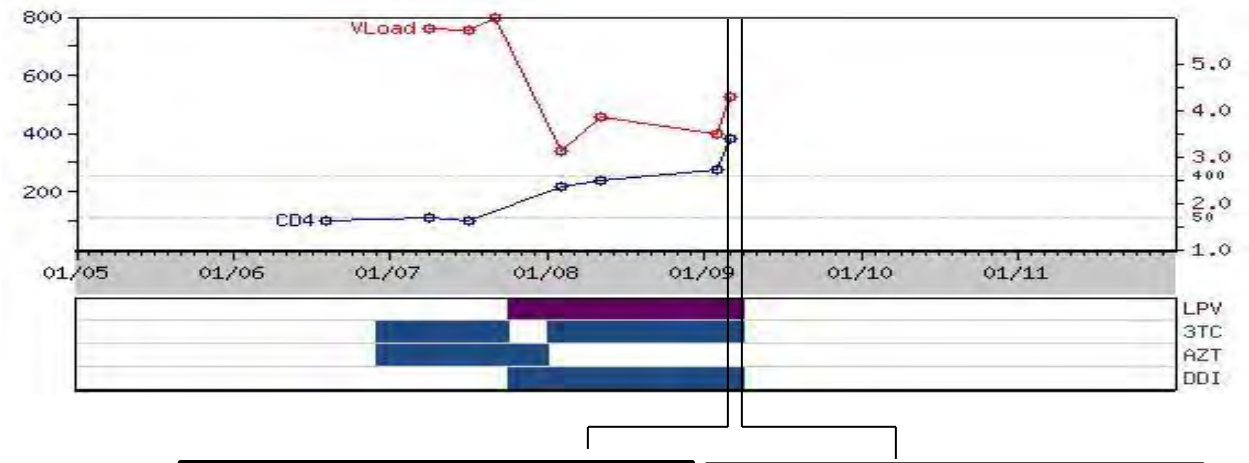
Mutations	Frequency of mutations (%)	
PI	**1st time point**	**2nd time pint**
-	-	-
NRTI		
T69A	14	62
T69D	0	3
NNRTI		
K103N	26	66
V106M	9	0
E138K	3	0
G190A	3	0

2.3.1.2 Longitudinal analysis of patients that remained on 2nd line therapy.

All these patients had several PR and RT while on PI inclusive regimen, with the exception of PCSM002 which only harbored M148V. Additional mutations were detected by cloning in the PR as well as in the RT region.

2.3.1.2.1 Longitudinal analysis of PCSM001

For PCSM001, a total of 57 HIV-1 Pol sequences were analyzed: 29 sequences were analyzed at the 1st and 28 sequences at the 2nd time point. At the 1st time point, the PR drug resistance mutations detected by standard genotyping included the mutations M46I, I54V, L76V, I84V and the minor (secondary) mutations L10F and K20R. The frequency of most of these mutations was > 55% (Table 2.15). Additional mutations detected by cloning included the major mutations I47V and I54L. The frequency of mutation I47V was low (3%) while for I54L it was high (66%) (Table 2.15). The RT drug resistance mutations included the NRTI mutations T69N, L74V, M184I/V, T215F, and K219Q and NNRTI mutations L100I, K103N, H221Y and M230L. The frequency of these mutations was high (≥ 93%). Additional mutations detected by cloning included the NRTI mutation D67N and the NNRTI mutation V106M. The frequency of this mutation was also high (67%). After one month, a 2nd genotype was performed. The majority of the PR and RT mutations that were detected at the 1st time point were still present except for PR mutations I54V and I47V (Figure 2.9 and table 2.15). The frequency of most of these remained the same (<100%). However, the frequency of PR mutations L10F and the additional mutation I54L which was detected by cloning increased to 97% while the frequency of RT Mutations M184I and K103N decreased to 3% and 21% respectively (Table 2.15)



1st time point
PR: L10F, M46I, I47V, K20R, I54V, I54L, L76V, V82A, I84V,
NRTI: D67N, T69N, L74V, M184I/V, T215F, K21Q
NNRTI: L100I, K103N, V106M, H22IY, M230L

2nd time point
PR: L10F, M46I, K20R, I54V, I54L, L76V, I84V,
NRTI: T69N, L74V, M184I/V, T215F, K219Q
NNRTI: L100I, K103N, H22IY, M230L

Figure 2.9: Graphical summary of drug resistance history and the resistance profile of PCSM001.

Table 2.15: The frequency of mutations at the 1st and 2nd time point for PCSM001.

Mutations	Frequency of mutations (%)	
PI	**1st time point**	**2nd time point**
L10F	67	97
K20R	100	93
M46I	100	100
I47V	3	0
I54L	66	97
I54V	55	0
L76V	97	97
V82A	68	0
I84V	97	97
NRTI		
D67N	67	0
T69N	93	97
L74V	93	93
M184I	62	3
M184V	100	90
T215F	93	93
K219Q	93	97
NNRTI		
L100I	93	93
K103N	86	21
V106M	67	0
H221Y	93	97
M230L	93	93

2.3.1.2.2 Longitudinal analysis of PCSM002

For PCSM002, a total of 55 sequences HIV-1 Pol were successfully analyzed: 28 sequences at the 1st and 27 sequences at the 2nd time point. At the 1st time point, the PR drug resistance mutations detected by standard genotyping included the major PI mutations M46I, I54V, T74S, V82A and the minor mutation K20R. The frequency of most of these mutations was high (>71%) (Table 2.16). Additional mutations detected by cloning included the minor mutation L10F and the frequency of this mutation was low (4%). The RT resistance mutations detected by standard genotyping included the NRTI mutation M184V. The frequency of this mutation was also high (89%) (Table 2.16). No NNRTI mutations were detected by standard genotyping and cloning. After two months, a 2nd genotype was performed. The PR resistance mutations detected at 1st time point was still detected (Figure 2.10 and Table 2.16). The frequency of these mutations remained the same while the frequency of L10F increased from 4% to 93% at this time point (Table 2.16). No additional mutations were detected by cloning at the 2nd time point.

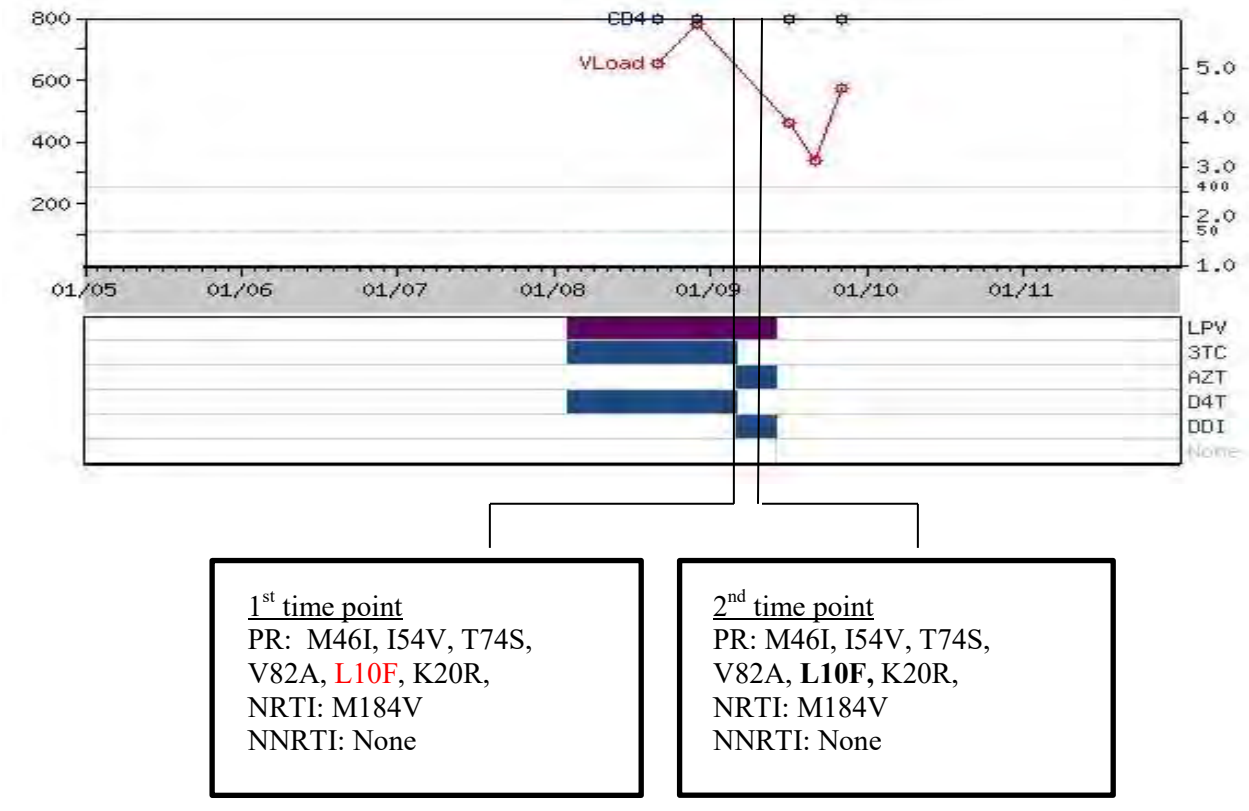


Figure 2.10: Graphical summary of drug resistance history and the resistance profile of PCSM002.

Table 2.16: The frequency of mutations at the 1st and 2nd time point for PCSM002.

Mutations	Frequency of mutations (%)	
PR	**1st time point**	**2nd time point**
L10F	4	93
K20R	100	93
M46I	100	93
I54V	71	93
T74S	100	93
V82A	89	93
NRTI		
M184V	89	93
NNRTI	-	-

2.3.1.2.3 Longitudinal analysis of PCSM007

For PCSM007, a total of 58 HIV-1 Pol sequences were analyzed: 30 sequences at the 1st and 28 sequences at the 2nd time point were successfully analyzed. At the 1st time point, the PR drug resistance mutations detected by standard genotyping included the major PI mutations M46I, I54V, A71V, L76V, V82A and the minor mutation L10F (Figure 2.11; Table 2.17). The frequency of these mutations was >50% at this time point (Table 2.17). Additional mutation detected by cloning included the major mutation K20R. The frequency of this mutation was 50% (Table 2.17). The RT drug resistance mutations detected by standard genotyping included the NRTI mutations D67N, M184V and NNRTI mutations, K103N, V106M. The frequency of the D69N, K103N and V106M mutations was high (>93%) while M184V was low (43%). Additional mutations detected by cloning included the NRTI mutation K70R and NNRTI mutations K101E, E138K and G190A. The frequency of K70R was 53%, while the frequency of K101E, E138K and G190A was the same (14 %) as they occurred in the same clones. After two months, a 2nd genotype was performed. The PR resistance mutations detected at the 1st time point was still detected. However, there was an increase in the frequency of these mutations e.g. L10F (82%), A71V (79%) and L76V (89%). An additional mutation K20R was still detected as well.

However, the frequency of this mutation was low (18%) (Table 2.17) and the RT resistance mutations were still detected. The frequency of the RT mutations D69N, K103N and V106M did not change dramatically, while the frequency of mutations K70R, E138K and G190A increased to 57% (Table 2.17). Additional mutation detected by cloning at this time point was M184V. However, the frequency this mutation was low (14%).

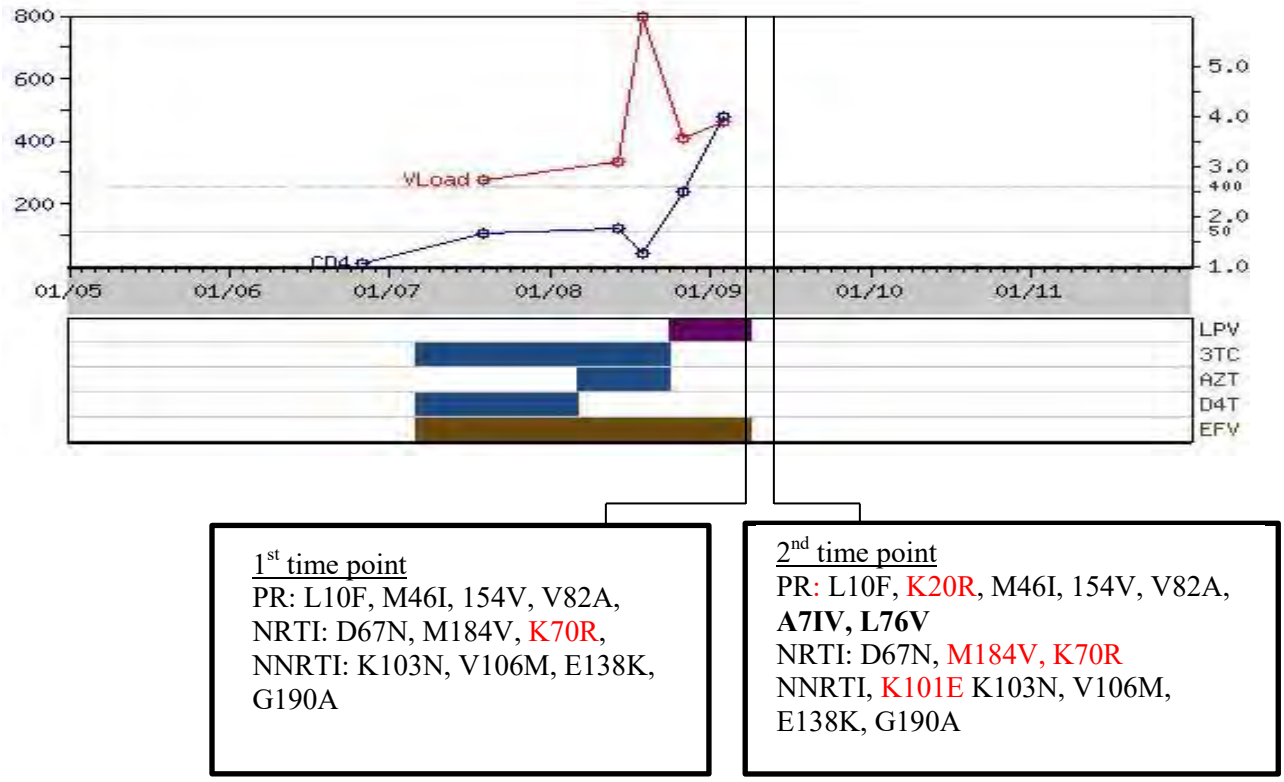


Figure 2.11: Graphical summary of drug resistance history and the resistance profile of PCSM007.

Table 2.17: The frequency of mutations at the 1st time point and 2nd time point for PCSM007

Mutations	Frequency of mutations (%)	
PR	**1st time point**	**2nd time point**
L10F	57	82
K20R	50	18
M46I	87	96
I54V	87	18
A71V	50	79
L76V	50	89
V82A	87	96
NRTI		
D69N	96	96
K70R	53	14
M184V	43	14
NNRTI		
K101E	0	57
K103N	93	96
V106M	100	96
E138K	14	57
G190A	14	57

2.4 Phylogenetic analysis

Phylogenetic analysis was performed to assess the evolution of the virus as well as to assess the relationship between the sequences with PI resistance mutations at the 1st and the 2nd time point. Maximum likelihood (ML) trees for each patients were constructed in PAUP*4.0b10 using the appropriated evolutionary model selected by the AIC in MODELTEST 3.7. In the patients that

remained on 2nd line therapy in both time points, the HKY+ I+G model was selected in two patients (PCSM001 and PCSM007) while GTR +I +G model was selected for PCSM002. In patients that were switched from 1st to 2nd line therapy, TVM +I+G model was selected for PCSM009 and PCSM012 while K81uf +I+G was selected for PCSM004. The trees were rooted with the _best fit root' as determined by Path-O-Gen V.1.4. Trees were viewed with Figtree v1.1.4.

2.4.1 Phylogenetic analysis of the patients that remained on 2nd line therapy.

There was intermingling between the sequences at the 1st and the 2nd time point, with the exception of PCSM007 where there was no intermingling of the sequences between the time points. Collectively, the tree topologies of these patients were not only due to the effect of drug of drug resistance mutations as the analysis of the trees where the drug resistance mutations were removed showed similar analysis (data not shown).

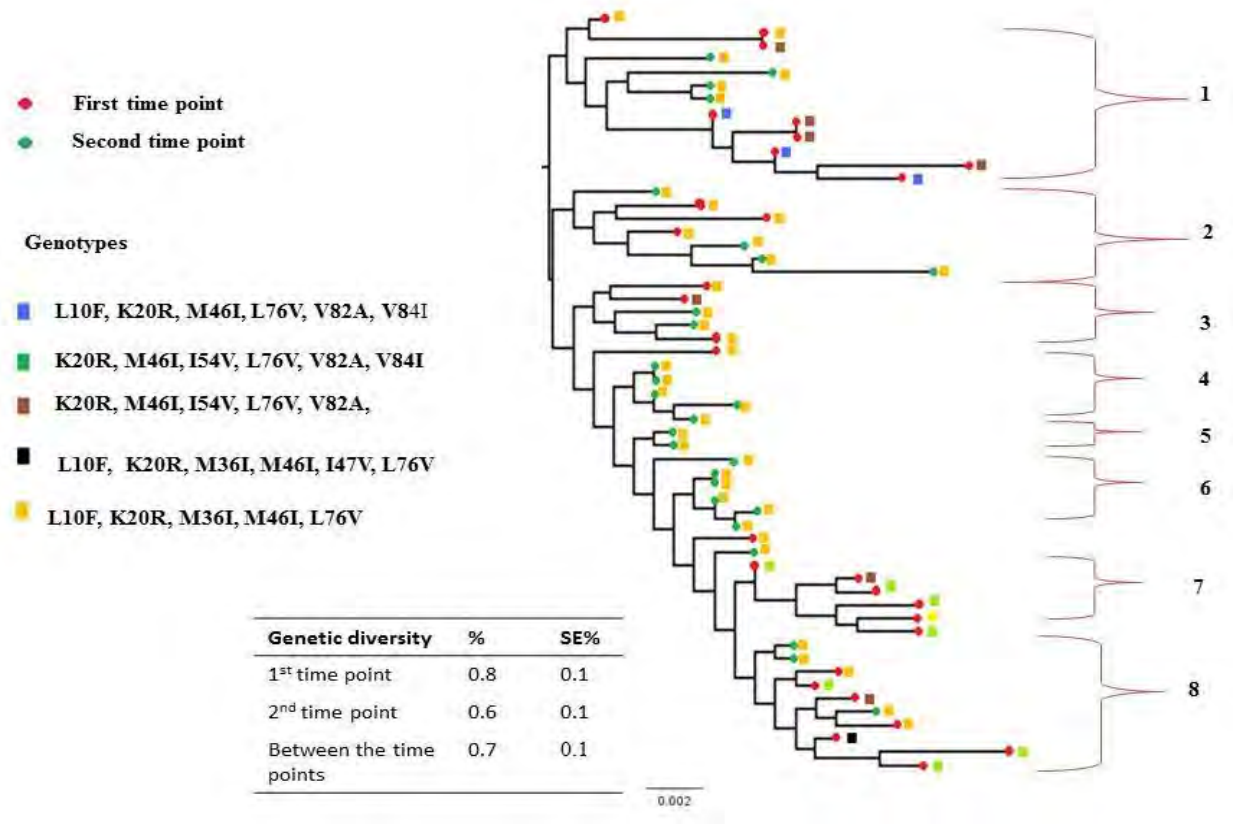
2.4.1.1 Phylogenetic analysis and intra sequence diversity of PCSM001

For PCSM001, the phylogenetic tree had eight clusters: four showing the intermingling between the sequences at the 1st (red circles) and the 2nd time point (green circles) (Clusters 1-3 and 8) and three clusters only with sequences from the 2nd time point (cluster 4-6) and one cluster only with sequences from the 1st time point (cluster 7) (Figure 2.12). Clusters 1, 2 and 8 had a diverse pattern characterized by both long and short branch lengths, suggesting that there is distinctive evolutionary pressure acting on the different virus thus leading to this diverse branch lengths pattern (Fischer et al., 2012). This could be amino acid sites selected under drug selective pressure. Clusters 3 and 7 showed long branches of similar lengths, while clusters 4-6 had short branch lengths suggesting less diversity in these sequences.

The intra -sequence diversity at each time point and between the time points is represented in the table alongside the phylogenetic tree (Figure 2.12). The intra sequence diversity at the 1st time point was 0.8% (SE ±0.1%) (Figure 2.12) and decreased slightly at the 2nd time point to 0.6% (SE ± 0.1%) while the diversity between the time points was 0.7% (SE ± 0.01%).

2.4.1.1.1 Phylogenetic analysis of PCSM001 investigating the evolution of PI resistance mutations at the 1st and the 2nd time point.

In this patient the sequences harboring the same resistance mutations clustered together. These clusters included both the 1st and 2nd time point sequences, suggesting that the viral population at the 2nd time point evolved from the mutant population detected at the 1st time point (Halvas et al., 2010). Specifically, the 1st time point sequences harboring the genotypes: L10F, K20R, M46I, L76V, V82A, V84I (blue squares); K20R, M46I, I54V, L76V, V82A, V84I (green squares); K20R, M46I, I54V, L76V, V82A (brown squares), L10F, K20R, M46I, I47V, L76V (black squares) and L10F, K20R, M46I, L76V(orange squares) intermingled with the mutant sequences at the 2nd time point harboring the genotype L10F, K20R, M46I, L76V (orange square) (Figure 2.12). Interestingly, there was only one genotype observed at the 2nd time point (orange square), with the variants with the I47V, I54V, V82A and V84I mutations no longer detectable.



Genetic diversity	%	SE%
1st time point	0.8	0.1
2nd time point	0.6	0.1
Between the time points	0.7	0.1

Figure 2.12: **Maximum likelihood (ML) trees of the HIV-1 pol region of PCSM001**.The ML was constructed using the HKY+ I+ G model of evolution selected by AIC implemented in the MODELTEST 3.7. The scale bar at the bottom of the tree indicates the number of nucleotide substitutions per site. Red circles represent the sequences at the 1st time point while the green circles represent the sequences at the 2nd time point. The squares represent the genotypes for the PR region. The intra sequence diversity and the standard error (SE) at the 1st and the 2nd time point and between the time points are represented in the table alongside the tree. The numbers the right-hand-side of tree represent the phylogenetic clusters.

Additionally, the same ML tree was also analyzed to investigate the evolution of NNRTI and NNRTI mutation overtime. Again, the sequences harboring the same RT resistance mutations clustered together (Figure 2.13). This clusters included the 1st and the 2nd time point sequences. In cluster 1, the 1st time point sequences harboring the genotypes: D67N, T69N, L74V, L100I, K103N/S,V106M, M184V/I, T215F, K219Q, H221Y, M230L (blue squares) and T69N, L74V, L100I, K103N, M184V, T215F, K219Q, H221Y, M230L (green squares) intermingled with the sequences at the 2nd time point harboring the genotypes: T69N, L74V, L100I, K103S, M184V, T215F, K219Q, H221Y, M230L and one sequence with genotype T69N, L74V, L100I, K103N,

M184I, K219Q, H221Y, M230L (pink squares). In clusters 2, the 1st time point sequences harboring the genotype T69N, L74V, L100I, K103N, M184V, T215F, K219Q, H221Y, M230L (green squares) intermingled with the 2nd time point sequences harboring the genotype T69N, L74V, L100I, K103S, M184V, T215F, K219Q, H221Y, M230L (purple squares) and one sequence with no mutations (orange squares). Cluster 3 also showed intermingling of the 1st time point sequences with the genotype: T69N, L74V, L100I, K103S, M184V, T215F, K219Q, H221Y, M230L (purple squares); D67N, T69N, L74V, L100I, K103N/S, V106M, M184V/I, T215F, K219Q, H221Y, and M230L (blue squares). Again cluster 8 showed intermingling of the 1st and 2nd time point sequences where the 1st time point harbored the genotype D67N, T69N, L74V, L100I, K103N/S,V106M, M184V/I, T215F, K219Q, H221Y, M230L (blue squares); T69N, L74V, L100I, K103N, M184V, T215F, K219Q, H221Y, M230L (green squares) and one sequences with genotype D67N, K103N, V106M, M184V (brown squares). Clusters 4-6 had sequences from the 2nd time point only harboring the genotypes: T69N, L74V, L100I, K103S, M184V, T215F, K219Q, H221Y, M230L (purple squares); T69N, L74V, L100I, K103N, M184V, T215F, K219Q, H221Y, M230L (green squares) and T69N, L74V, L100I, K103N, M184V, T215F, K219Q, H221Y, M230L (red squares), while cluster 7 had only sequences from the 1st time point only with the genotype D67N, T69N, L74V, L100I, K103N/S,V106M, M184V/I, T215F, K219Q, H221Y, M230L (blue squares).

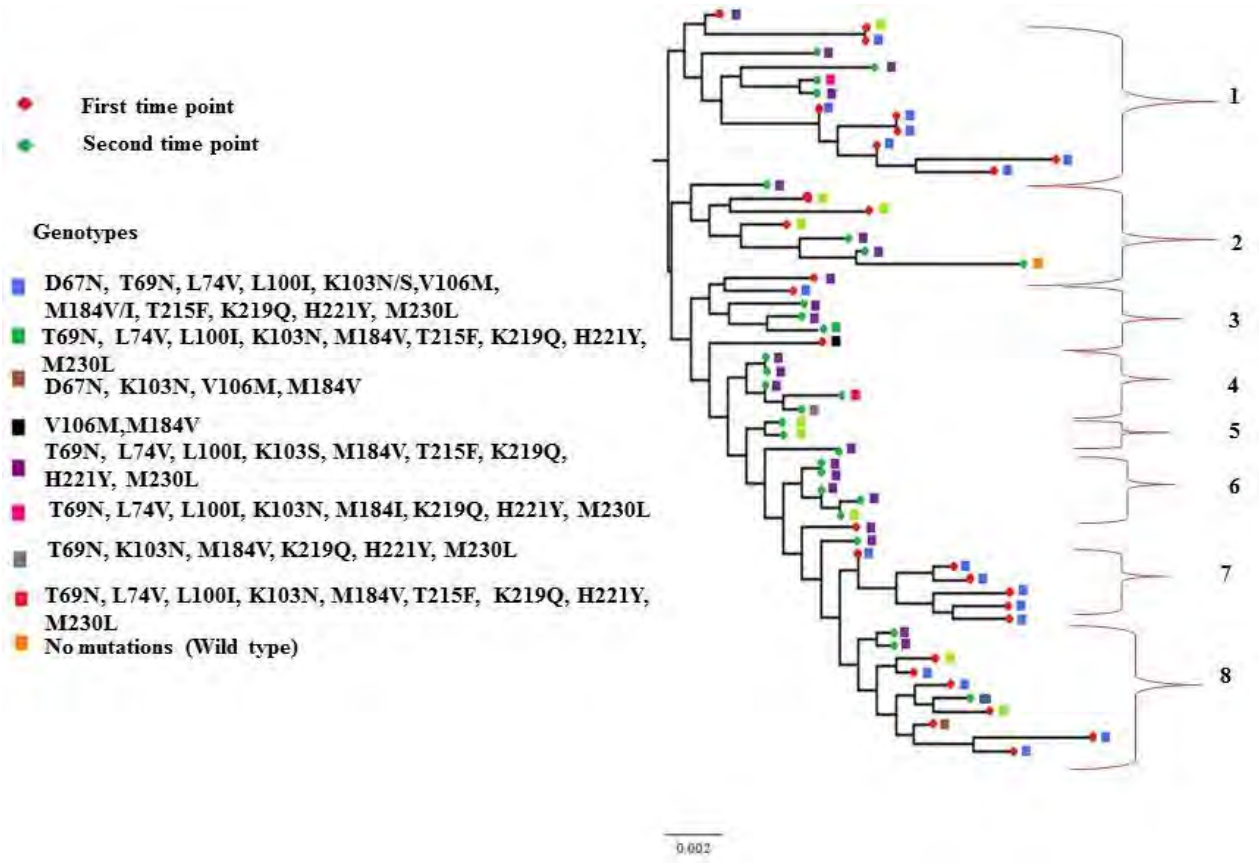


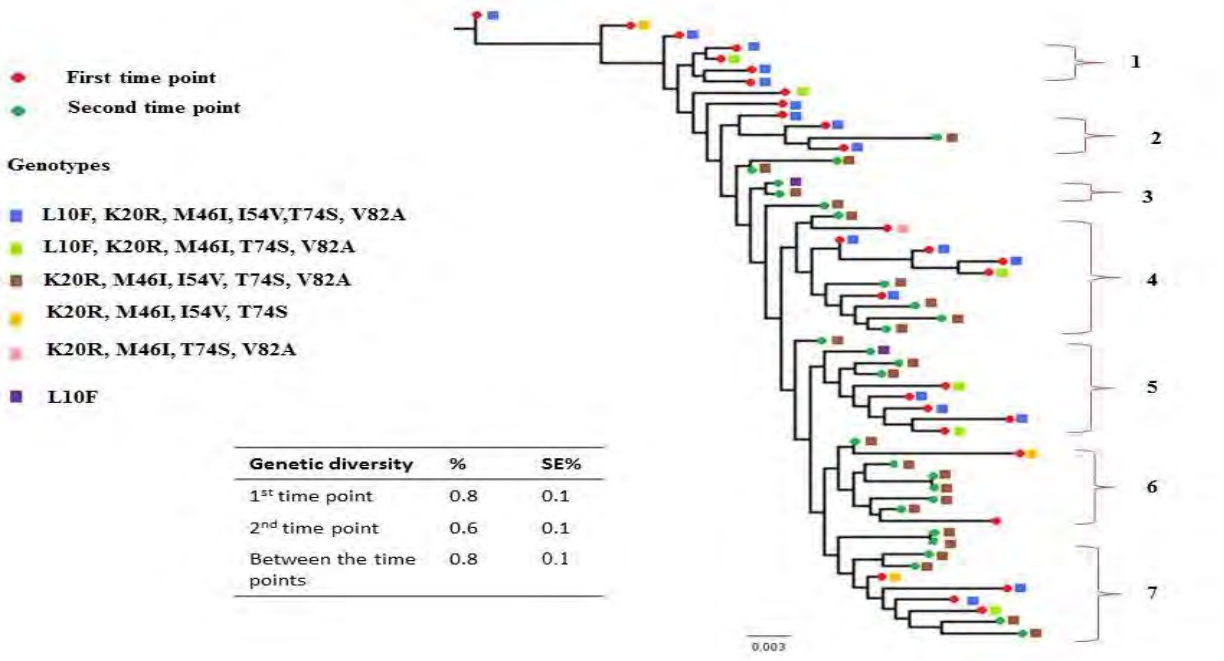
Figure 2.13: Maximum likelihood (ML) trees of the HIV-1 pol region of PCSM001 showing the evolution of RT resistance overtime. Red circles represent the sequences at the 1st time point while the green circles represent the sequences at the 2nd time point. The squares represent the genotypes at for the RT region.

2.4.1.2 Phylogenetic analysis and intra sequence diversity of PCSM002

In PCSM002, the phylogenetic tree had seven clusters: five showing the intermingling between the sequences at the 1st and the 2nd time point (Clusters 2, 4-7) and two clusters only with sequences from the 1st time point (clusters 1 and 3). Clusters 2, 4-6 and 7 showed both long and short branch lengths. Cluster 1 showed branches of similar lengths, while cluster 3 had short branch lengths compared to the other branches, suggesting less diversity in these sequences. The intra -sequence diversity at each time point and between the time points is represented in the table alongside the phylogenetic tree (Figure 2.14). The intra sequence diversity at the 1st time point was 0.8% (SE ± 0.1%) (Figure 2.14) but decreased slightly at the 2nd time point to 0.6% (SE ± 0.1%) while the diversity between the time points was 0.8% (SE ± 0.1%).

2.4.1.2.1 Phylogenetic analysis of PCSM002 investigating the evolution of PI resistance mutations at the 1st and the 2nd time point.

Again, the sequences harboring the same resistance mutations clustered together. These clusters included both 1st time point and 2nd time point sequences. The 1st time point sequences harboring the genotypes: L10F, K20R, M46I, I54V, T74S, V82A (blue squares); K20R, M46I, I54V, T74S (orange squares); L10F, K20R, M46I, T74S, V82A (green squares) and K20R, M46I, T74S, V82A (pink squares) intermingled with the sequences at the 2nd time point. All but two sequences at the 2nd time point harbored the K20R, M46I, I54V, T74S, V82A genotype (brown squares). The remaining two sequences only showed the L10F mutation in the PR (purple squares) (Figure 2.14 A).



Genetic diversity	%	SE%
1st time point	0.8	0.1
2nd time point	0.6	0.1
Between the time points	0.8	0.1

Figure 2.14: Maximum likelihood (ML) trees of the HIV-1 pol region of PCSM002.The ML was constructed using the GTR + I+ G model of evolution selected by AIC implemented in the MODELTEST 3.7. The scale bar at the bottom of the tree indicates the nucleotide substitution per site. Red circles represent the sequences at the 1st time point while the green circles represent the sequences at the 2nd time point. The squares represent the genotypes at the PR. The intra sequence diversity and the standard error (SE) at the 1st and the 2nd time point and between the time points are represented in the table alongside the tree. The numbers alongside the tree represent the phylogenetic clusters.

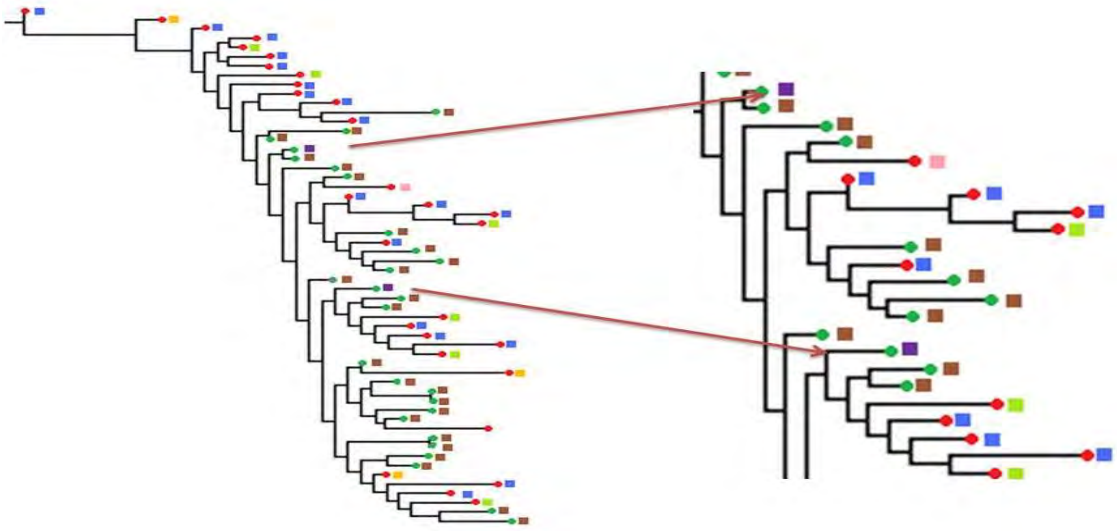


Figure 2.14 A. ML tree indicating the sub tree showing the two sequences with the L10F mutation in PR (Purple).

Again, the same ML tree was also analyzed to investigate the evolution of NNRTI and NNRTI mutation overtime. In this patient, the majority of the sequences in both time points had the mutation M184V with a few sequences which were wild type (Figure 2.15). This suggests M184V was the fit mutation and might have increased treatment failure. Additionally, there was no evolution of RT resistance overtime suggesting that the M184V at the 1st time point persisted to 2nd time point.

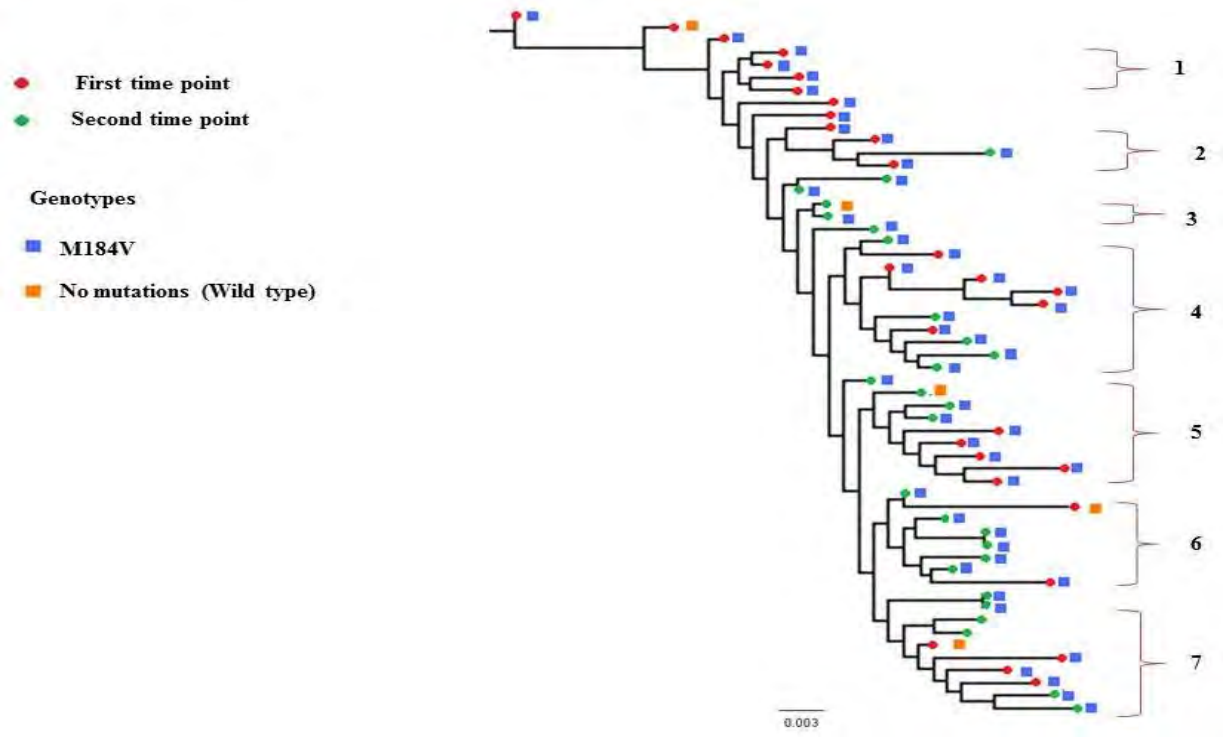


Figure 2.15: Maximum likelihood (ML) trees of the HIV-1 pol region of PCSM001 showing the evolution of RT resistance overtime. Red circles represent the sequences at the 1st time point while the green circles represent the sequences at the 2nd time point. The squares represent the genotypes at for the RT region.

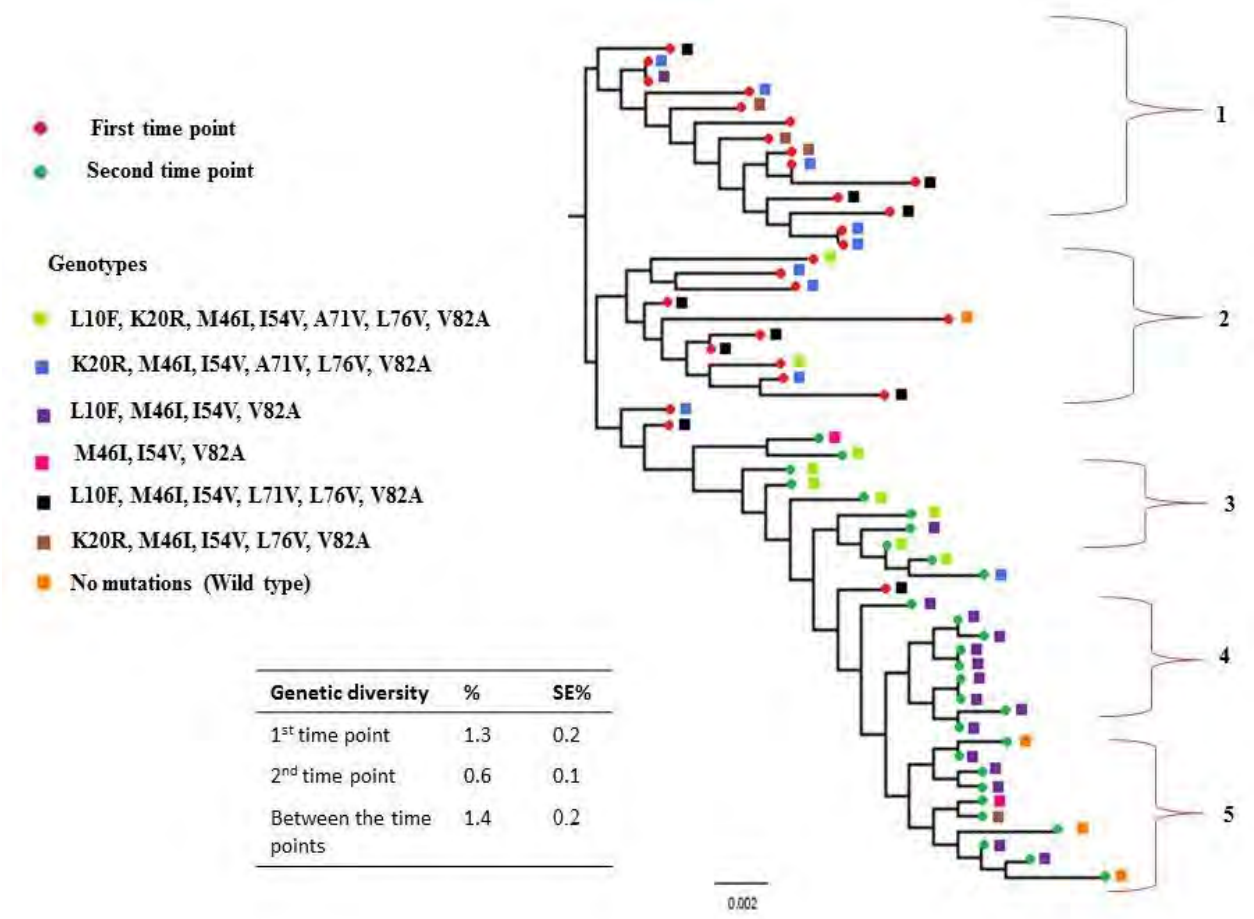
2.4.1.2 Phylogenetic analysis and intra sequence diversity of PCSM007

For PCSM007, the phylogenetic tree had five clusters: two showing the sequences from the 1st time point only (clusters 1 and 2) and three clusters showing the sequences from the 2nd time point (Clusters 3-5) (Figure 2.16). Only three sequences from the 1st time point clustered with the sequences from the 2nd time point suggesting that the sequences at the 2nd time point evolved from the sequences detected at the 1st time point. Clusters 1 and 2 had longer branch lengths indicating high diversity in those sequences. Cluster 3-5 showed short branch lengths of similar lengths, specifically in clusters 4 and 5, with only a few sequences with longer branch lengths (clusters 3 and 5).

The intra sequence diversity at the 1st time point was high: 1.3 % (SE ± 0.2%) but decreased at the 2nd time point to 0.6 % (SE ± 0.1), while the diversity between the time points was 1.4% (SE ±0.2%).

2.4.1.2.1 Phylogenetic analysis of PCSM007 investigating the evolution of PI resistance mutations at the 1st and the 2nd time point.

Again, the sequences harboring the same resistance mutations clustered together (Figure 2.16). The 1st time point sequences harbored the genotypes: L10F, K20R, M46I, I54V, A71V, L76V, V82A (green squares), L10F, M46I, I54V, L71V, L76V, V82A (black squares), K20R, M46I, I54V, A71V, L76V, V82A (blue squares), K20R, M46I, I54V, L76V, V82A (brown squares), and one sequences which had no mutations (orange squares). The majority of the sequences at the 2nd time point harbored the genotype: L10F, M46I, I54V, V82A (purple squares). The remaining sequences harbored the genotypes: L10F, K20R, M46I, I54V, A71V, L76V, V82A (green squares), K20R, M46I, I54V, A71V, L76V, V82A (blue squares), M46I, I54V, V82A (pink squares) and few which were wild types (orange squares). Interestingly, only two genotypes detected at the 1st time point were observed at the 2nd time point i.e. genotypes: L10F, K20R, M46I, I54V, A71V, L76V, V82A (green squares) and K20R, M46I, I54V, A71V, L76V, V82A (blue squares) (Figure 2.16).



Genetic diversity	%	SE%
1st time point	1.3	0.2
2nd time point	0.6	0.1
Between the time points	1.4	0.2

Figure 2.16: Maximum likelihood (ML) trees of the HIV-1 pol region of PCSM007. The ML was constructed using the HKY+ I+G model of evolution selected by AIC implemented in the MODELTEST 3.7. The scale bar at the bottom of the tree indicates the nucleotide substitution per site. Red circles represent the sequences at the 1st time point while the green circles represent the sequences at the 2nd time point. The squares represent the genotypes at the PR. The intra sequence diversity and the standard error (SE) at the 1st and the 2nd time point and between the time points are represented in the table alongside the tree. The numbers alongside the tree represent the phylogenetic clusters.

Again, the same ML tree was also analyzed to investigate the evolution of NNRTI and NNRTI mutation overtime. Again, the sequences harboring the same RT resistance mutations clustered together (Figure 2.17). The majority of the sequences at the 1st time point harbored the genotype D67N, K103N, V106M (pink squares) with a few sequences harboring the genotype D67N, K70R, K103N, V106M, E138K, M184V, G190A (purple squares) and one sequences which was wild type (orange squares). This patient harbored more diverse genotypes at the 2nd time point:

D67N, K70R, K101E, K103N, V106M, E138K, G190A (blue squares), D67N, K103N, V106M, M184V (green squares), D67N, K103N, V106M (pink squares) and one sequence harboring the genotype K70R, K101E, V106M, E138K, M184V, G190A (brown squares).

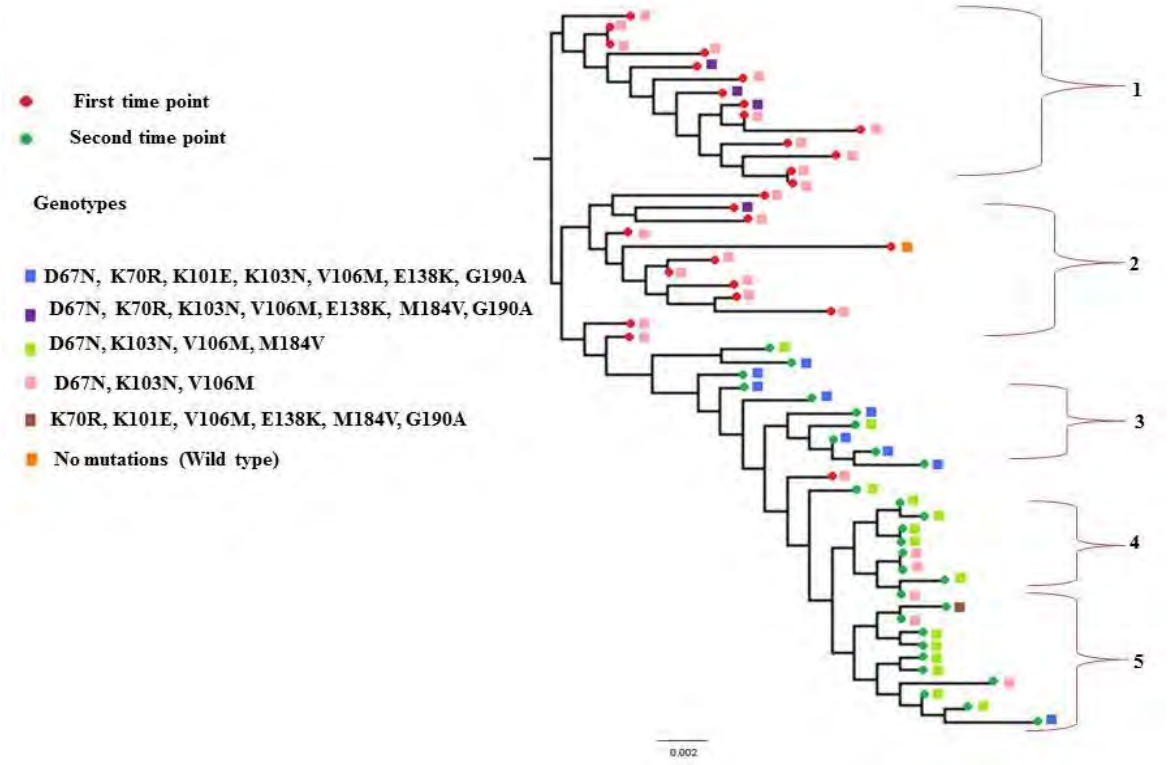


Figure 2.17: Maximum likelihood (ML) trees of the HIV-1 pol region of PCSM007 showing the evolution of RT resistance overtime. Red circles represent the sequences at the 1st time point while the green circles represent the sequences at the 2nd time point. The squares represent the genotypes at for the RT region.

2.4.2 Phylogenetic analysis of the patients that were switched from 1st to 2nd line therapy.

There was no intermingling between the sequences at the 1st and the 2nd time point, with the exception of PCSM012 where a few sequencing from 1st time point intermingled with the sequences at the 2nd time point. Again, the tree topologies of these patients were not only due to the effect of drug resistance mutations as the analysis of the trees where the drug resistance mutations were removed showed similar analysis (data not shown).

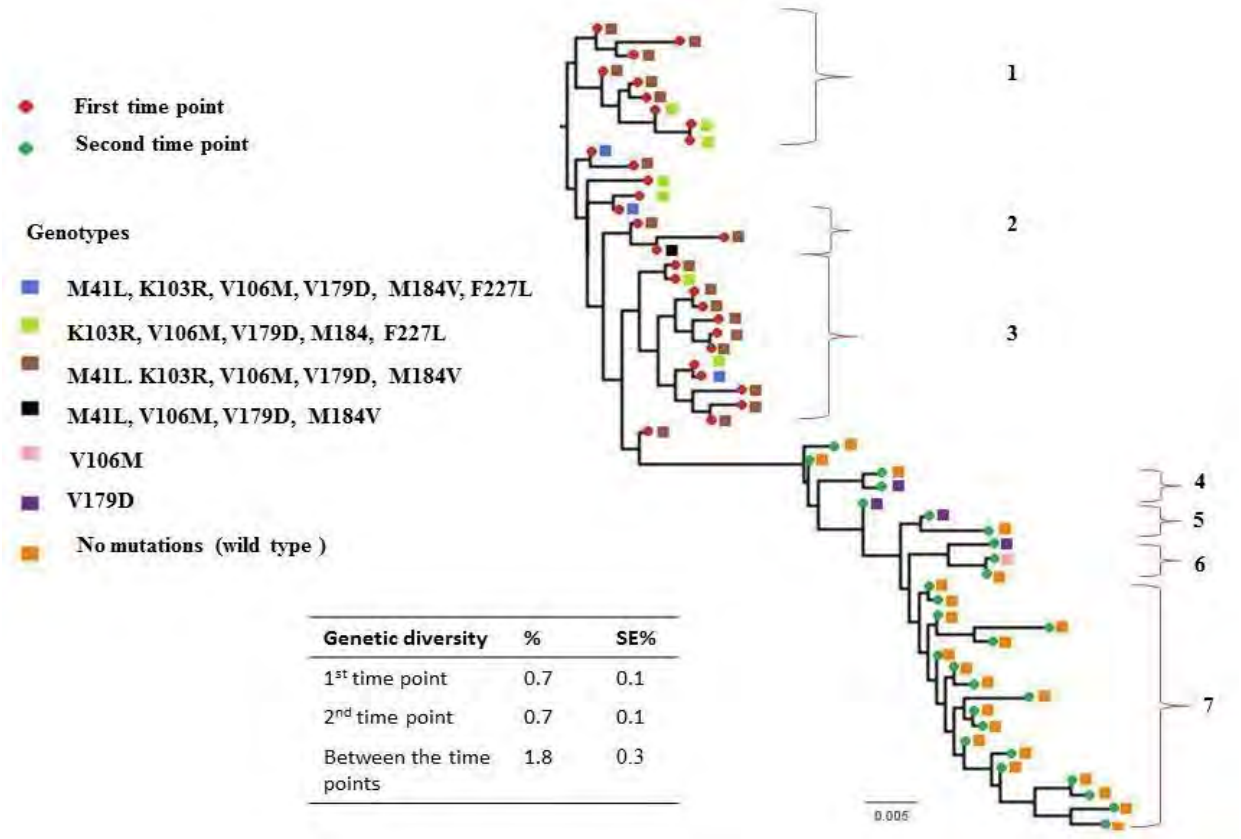
2.4.2.1 Phylogenetic analysis and intra sequence diversity of PCSM004

PCSM004 showed two distinct populations of sequences, one with the sequences at the 1st time point and the other with the sequences at the 2nd time point (Figure 2.18). The phylogenetic tree had seven clusters: three showing the sequences from the 1st time point (clusters 1-3) and four clusters showing the sequences from the 2nd time point (clusters 4-7). There was no distinguishing pattern between these clusters as they are characterized by both long and short branch lengths, again suggesting that there is distinctive evolutionary pressure acting on the different virus thus leading to this diverse branch lengths pattern (Fischer et al., 2012).

The intra sequence diversity at the 1st and the 2nd time point was low i.e. 0.7% (SE= 0.1%) while the diversity between the time point was high 1.8 % (SE = 0.3%) (Figure 2.18).

2.4.2.1.1 Phylogenetic analysis of PCSM004 investigating the evolution of RT resistance mutations at the 1st and the 2nd time point.

In this patient the sequences harboring the same resistance mutations clustered together. However, there was no intermingling between sequences from the 1st time point (clusters 1-3) and the 2nd time point (clusters 4-7) (Figure 2.18). This patient harbored a more diverse genotypes at the 1st time point: M41L, K103R, V106M, V179D, M184V, F227L (blue squares); K103R, V106M, V179D, M184V, F227L (green squares); M41L, K103R, V106M, V179D, M184V (brown squares) and M41L, V106M, V179D, M184V (black squares) compared to the 2nd time point which mainly harbored the wild type virus (orange squares) and those with the V106M (pink squares) and V179D (purple squares) mutations. This suggests that this patient was non-adherent thereby allowing the wild type variants to become the dominant strains.



Genetic diversity	%	SE%
1st time point	0.7	0.1
2nd time point	0.7	0.1
Between the time points	1.8	0.3

Figure 2.18: Maximum likelihood (ML) trees of the HIV-1 pol region of PCSM004.The ML was constructed using the K81uf +I+G model of evolution selected by AIC implemented in the MODELTEST 3.7. The scale bar at the bottom of the tree indicates the number of nucleotide substitutions per site. Red circles represent the sequences at the 1st time point while the green circles represent the sequences at the 2nd time point. The squares represent the genotypes for the PR region. The intra sequence diversity and the standard error (SE) at the 1st and the 2nd time point and between the time points are represented in the table alongside the tree. The numbers the right-hand-side of tree represent the phylogenetic clusters.

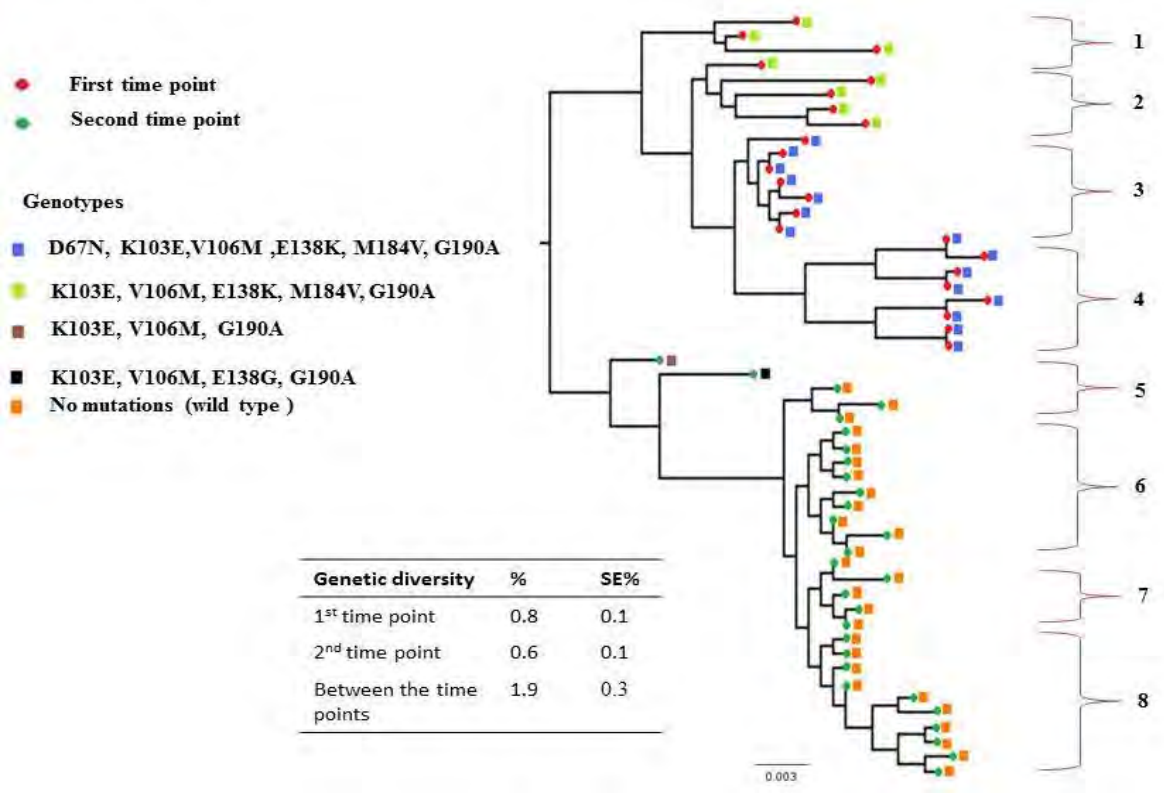
2.4.2.2 Phylogenetic analysis and intra sequence diversity of PCSM009

PCSM009 also showed two distinct populations of sequences, one with the sequences at the 1st and the other with the sequences at the 2nd time point (Figure 2.19). The phylogenetic tree had eight clusters: four showing the sequences from the 1st time point (clusters 1-4) and four clusters showing the sequences from the 2nd time point (clusters 5-8). Clusters 1 and 2 had longer branch lengths compared to clusters 3 and 4. Clusters 5-8 showed short branch lengths of similar length with only a few sequences with longer branch lengths (in cluster 5-7). However, there are two sequences at the 2nd time point that form part of the cluster but also fall outside the cluster. This

could be possibly due to recombination. The intra sequence diversity at the 1st time point was 0.8% (SE = 0.1 %) and decreased to 0.6% (SE = 0.1%) at the 2nd time point, while the diversity between the time point was high 1.9 % (SE =0.3%) (Figure 2.19).

2.4.2.2. Phylogenetic analysis of PCSM009 investigating the evolution of RT resistance mutations at the 1st and the 2nd time point

In this patient the sequences harboring the same resistance mutations clustered together (Figure 2.14). This patient harbored two genotypes at the 1st time point: D67N, K103E, V106M, E138K, M184V, G190A (blue squares) and K103E, V106M, E138K, M184V, G190A (green squares). The sequences at the 2nd time point were mostly wild type (orange squares) and with only two sequences with K103E, V106M, E138G, G190A (black squares) and the other with K103E, V106M, G190A (brown squares), with the variants with D67N, M184V, and E138K no longer detectable.



Genetic diversity	%	SE%
1st time point	0.8	0.1
2nd time point	0.6	0.1
Between the time points	1.9	0.3

Figure 2.19: Maximum likelihood (ML) trees of the HIV-1 pol region of PCSM009. The ML was constructed using the TVM +I+G model of evolution selected by AIC implemented in the MODELTEST 3.7. The scale bar at the bottom of the tree indicates the number of nucleotide substitutions per site. Red

circles represent the sequences at the 1st time point while the green circles represent the sequences at the 2nd time point. The squares represent the genotypes for the PR region. The intra sequence diversity and the standard error (SE) at the 1st and the 2nd time point and between the time points are represented in the table alongside the tree. The numbers the right-hand-side of tree represent the phylogenetic clusters.

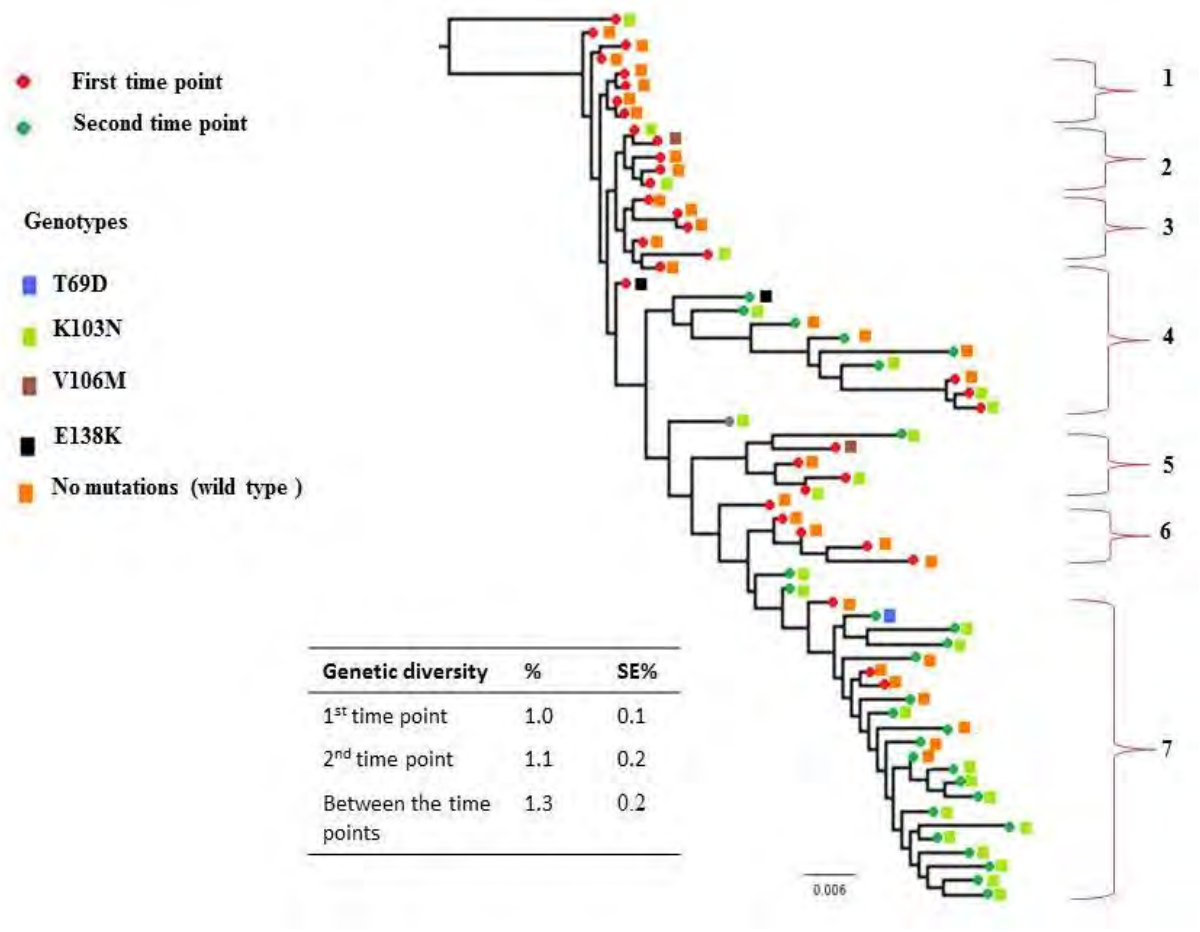
2.4.2.3 Phylogenetic analysis and intra sequence diversity of PCSM012

In PCSM012, the phylogenetic tree had seven clusters: three showing the intermingling between the sequences at the 1st (red circles) and at the 2nd time point (green circles) (clusters 4, 5 and 7) and four clusters only with sequences from the 1st time point (1-3 and 6). Cluster 4, 5 and 7 had a diverse pattern characterized by short branch lengths at the 1st time point and by both short and longer branch lengths at the 2nd time point, suggesting less diversity at the 1st time point. Clusters 1-3 and 6 showed short branch lengths with only a few sequences with longer branch lengths (clusters 3 and 6).

The intra sequence diversity at the 1st time point was 1.0% (SE = 1.0%) and remained more or less the same at the 2nd time point: 1.1% (SE = 0.2%) (Figure 2.20), while the diversity between the time points was high: 1.3% (SE =0.2%).

2.4.2.2 Phylogenetic analysis of PCSM012 investigating the evolution of RT resistance mutations at the 1st and the 2nd time point

In this patient the majority of the sequences at the 1st time point were wild type (orange squares) with a few sequences harboring the K103N (green squares) mutation and one sequence with V106M (brown) which intermingled with the K103N mutation at the 2nd time point (Figure 2.20a). The sequences at the 2nd time point harbored the K103N mutation with a few wild type viruses and one sequence with E138K mutation (black squares). Interestingly, all the sequences harbored only one of these mutations, suggesting independent selection of the viruses with each of the mutations. Collectively, the tree topologies of these patients were not only due to the effect of drug resistance mutations as the analysis of the trees where the drug resistance mutations were removed showed similar analysis (data not shown).



Genetic diversity	%	SE%
1st time point	1.0	0.1
2nd time point	1.1	0.2
Between the time points	1.3	0.2

Figure 2.20: Maximum likelihood (ML) trees of the HIV-1 pol region of PCSM0012. The ML was constructed using the TVM +I+G model of evolution selected by AIC implemented in the MODELTEST 3.7. The scale bar at the bottom of the tree indicates the number of nucleotide substitutions per site. Red circles represent the sequences at the 1st time point while the green circles represent the sequences at the 2nd time point. The squares represent the genotypes for the PR region. The intra sequence diversity and the standard error (SE) at the 1st and the 2nd time point and between the time points are represented in the table alongside the tree. The numbers the right-hand-side of tree represent the phylogenetic clusters.

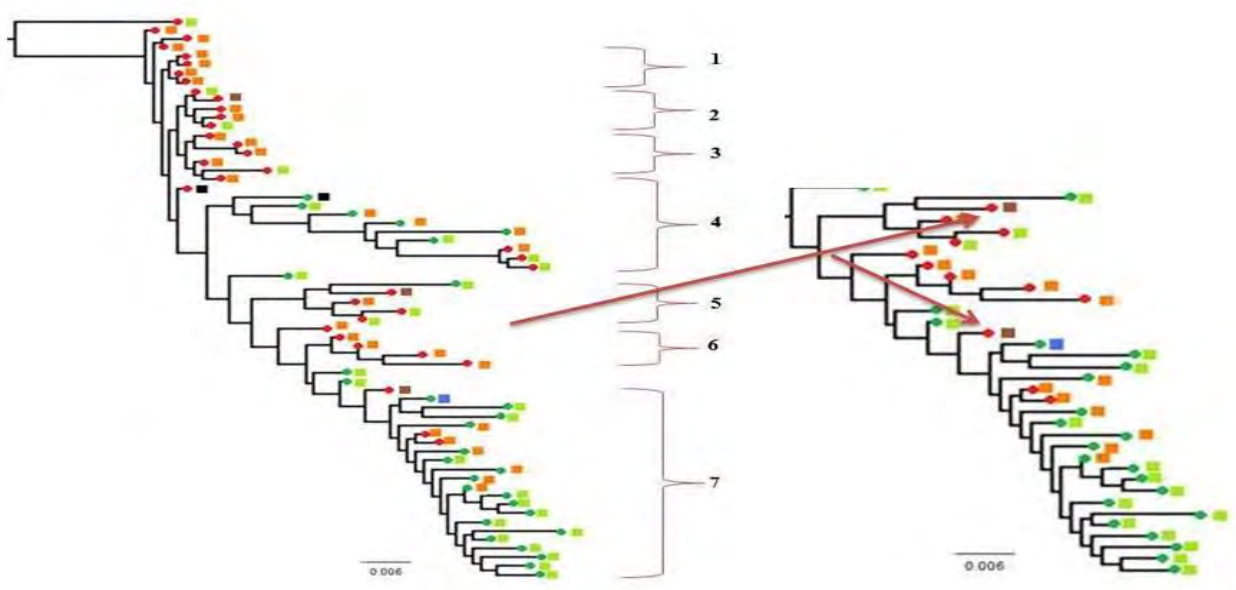


Figure 2.20A: Phylogenetic tree indicating the sub tree showing the sequences at the 1st time point harboring the K103N (green) mutation and two sequences with V106M (brown) which intermingled the K103N (green) mutation at the 2nd time point.

2.6 Positive selection pressure

The sequences of the two groups were then analyzed to investigate the sites under drug selection pressure. Analysis by the IFEL method showed that the positively selected sites were found mostly in the RT region in all the patients. More sites under positive selections were found in the patients that remained on 2nd line therapy compared to the patients that were switched from a 1st to 2nd line therapy. In the patients that remained on 2nd line therapy, ten sites were positively selected (Table 2.18). Of these sites, two were found at the PR positions (codons 10 and 76) (Table 2.18). These sites are associated with PIs resistance i.e. minor mutations L10F and major mutation L76V. Eight positively selected sites were found in RT positions (codons 57, 70, 103, 135, 184, 286, 297 and 376) (Table 2.18). Of these sites, four sites were associated with RT resistance i.e. K70E/R/S, K103K/N/S/E, I135L/M/T and M184V. In addition, most of the sites under positive selection (60%) were found in putative CTL epitopes: PR codon 76 and RT codons 57, 135, 184, 297 and 397 (Table 2.18).

In the patients that were switched from a 1st to 2nd line therapy, all five sites under positive selection pressure were in RT: codons 106, 173, 179, 210 and 272 (Table 2.3.5). Of these sites

two were associated with RTI resistance (V106M and V179D/E/F). All the sites under positive selective pressure were found at the putative CTL epitopes (Table 2.18).

Table 2.18: Positive selection pressure HIV-1 pol region

Patients that remained on 2nd line therapy			**Patients that were switched from a 1st to 2nd line therapy**		
Codon position	AA	Epitope position (hxb2)	Codon position	AA	Epitope position (hxb2)
PR-10	**F/P**	-	-	-	-
PR-76	**V/L**	68-76	-	-	-
57	T/S	57-66	-	-	-
RT-70	**K/R**	-	-	-	-
RT-103	**K/S/N/E**	-	-	-	-
			RT-106	**V/M**	105-113
RT-135	**I/V/T/M**	127-135	-	-	-
		-	RT-173	K/E/Q	173-183
			RT-179	**V/D/I**	179-187
RT-184	**V/I/M**	175-184		-	-
		-	RT-210	K/R	202-210
		-	RT-272	P/A/S/P	269-277
RT-285	T/A/V/I	-	-	-	-
RT-297	E/K	292-300	-	-	-
RT-376	V/Q/L	374-383	-	-	-

-Underline and bold indicate sites that are associated with drug resistance.

2.7 Identification of signature differences within sequences at two time points.

Signature differences between the sequences at the 1st and the 2nd time point were generated in VESPA using a 50% threshold (sequence difference were in > 50% of the sequences). In the patients that remained on 2nd line therapy, while there were individual differences between the time points, there were no common signature mutations seen. For PCSM001, the signature

difference was noted in one position i.e. codon position 201 which is at the RT position (Table 2.19). For PCSM002, the signature difference was found at 1 PR position (codon 10) and at two RT positions (codon 233 and 301). For PCSM007, signature difference was found at six PR positions (codons 38, 45, 53, 70, 75 and 81) while three were found at the RT position (codon 156, 345 and 395) suggesting that in this patient the genetic diversity was high at the PR region (Table 2.19).

Table: 2.19: Signature differences positions of patients that remained on 2nd line therapy.

	Signature difference		
AA Position	**PCSM001**	**PCSM002**	**PCSM007**
PR-10		L/F	
PR-38			Q/P
PR-45			M/I
PR-53			I/V
PR-70			A/V
PR-75			L/V
PR-81			V/A
RT-156			S/T
RT-201	S/N		
RT-233		I/T	
RT-301		K/E	
RT-384			A/V
RT-395			E/K

For the patients that were switched from a 1st to 2nd line therapy, there were three signature differences that were common in RT: 200, 201 and 237 (Table 2.20). Overall, more signature differences was seen between the time points than in the patients that remained on 2nd therapy. For PCSM004, signature differences were found at 11 RT positions (codons 139, 200, 201, 204, 237, 271, 277, 282, 312, 384 and 395). For PCSM009, this was noted at 2 PR positions (codon 39 and 89) and 15 RT positions (codons 119, 135, 166, 200, 205, 237, 242, 265, 278, 289, 310,

336, 426, 433 and 455). For PCSM012, this was found at six RT positions (codon 125, 167, 201, 272, 370 and 484).

Table 2.20: Signature difference positions of patients that were switched from the 1st to the2nd line therapy.

	Signature difference		
AA Position	PCSM004	PCSM009	PCSM012
PR-39		Q/P	
PR-89		L/M	
RT-119		R/K	
RT-125			S/T
RT-135		E/A	
RT-139	L/M		
RT-166		N/D	
RT167			T/A
RT-200	**R/K**	**E/K**	
RT201	**R/K**		K/N
RT-204	M/V		
RT-205		M/V	
RT-237	**K/T**	**K/E**	
RT-241		I/V	
RT-265		K/R	
RT-271	A/V		
RT-272			K/E
RT277	D/V		
RT-278		V/M	
RT-282	V/M		
RT-289		A/G	
RT-310		K/R	
RT-312	L/F		
RT-336		N/D	

RT-370				S/A
RT-384	A/T			
RT-395	E/A			
RT-426		A/V		
RT-433		D/N		
RT-455		H/Y		
RT-484				T/I

-Underline and bold indicate signature differences that are common in the patients that are associated with drug resistance.

2.8 Discussion

The study of minority variants in HIV-1 infection is important as the impact of minority variants related to drug resistance on clinical outcomes remains unresolved although there is evidence that these minority variants can influence treatment efficacy. Detection of minority variants can assist in predicting the virological failure especially in the HIV-1 individuals that didn‘t have mutations detected by standard genotyping (Paredes i Deiros, 2009). Among the developed methods used to detect minority variants, cloning is one of the method that can allow the detection of minority variants to as low as ≥ 10 % of the viral population (Paredes i Deiros, 2009).

In this study, the minority variants in the HIV-1 Pol that were not detected by standard genotyping in patient failing HAART were assed using cloning. Different indicators of evolutionary forces were used to explain the evolution of the virus overtime, signature difference and putative sites under drug selection pressure. This also includes the phylogenetic analysis. First, the minority variants were identified at two time points in these patients. None of the patients that were switched from 1st to 2nd line therapy, which were studied to investigate the resistance profile before and after starting a PI inclusive regimen developed PI mutations when switched to PI inclusive regimen. This could be due to the high genetic barrier of PIs as the development of PI mutations takes time to develop (Charpentier et al., 2004) or it could be due to adherence. In PCSM004 and PCSM009 the additional NNRTI mutations detected at the 2nd line failure by cloning illustrate that these mutations can persist in the viral quasispecies after the drug exposure although at low frequencies. In PCSM009, the NNRTI mutations E138G (9%)

and E138K (91%) were detected by both methods at the 1st time point but only E138G was detected at the 2nd time point although the frequency was low (7%). This suggests that the mutation E138G was fit at the 2nd time point compared to E138K. PCSM012 had additional NRTI and NNRTI mutations at the 1st line failure detected by cloning also at low frequencies. Again, this suggests that this patient was not adherent therefore allowing the reemergence of the wild type viruses from the reservoir when they stop taking their treatment following the emergence of this drug resistance mutations (Hué et al., 2009).

For the patients that remained on 2nd line therapy, which was studied to investigate the resistance profile with prolonged PI therapy, all the patients harbored both the major and the minor PI resistance mutations. There was an increase in the number and the frequency of the PI mutations and this could be attributed to the prolong time on a failing regimen. For PCSM001 the additional major mutation I47V was detected by cloning at the 1st time point, however the frequency was low (3%) suggesting that this variant might have been stored in the viral population and decayed to levels that cannot be detected by standard genotyping (Johnson and Geretti, 2010). In the same patient, the additional mutation I54L detected only by cloning in both time points and the detection of I54V only at the 1st time point by both methods suggests that at the 1st time point this mutations occurred as a mixture and due to the results of fitness cost i.e. high fitness cost of I54V, the mutation I54L was overgrown by I54V at the 2nd time point (Johnson et al., 2008). In PCSM001, the additional NNRTI mutations detected at the 1st line failure, again suggest that this patient was not adherent allowing these drug resistance mutations to decay to levels that cannot be detected by standard genotyping (Johnson and Geretti, 2010). In PCSM007, the detection of additional NNRTI mutations in both time points suggests that the detected mutations at the earlier time point persisted and could have contributed to treatment failure at the 2nd time point.

Positive selection was the dominant selective force in the RT region in all the patients and more in the patients that remained on 2nd line therapy. The selection of these sites indicates that they increased viral fitness under drug selective pressure (Banke et al., 2009). Previous studies have suggested that the sites under positive selection also play a role in CTL response (Banke et al., 2009). In this study, all the sites under positive selection in the patients that were switched from a 1st to 2nd line therapy were found in putative CTL epitopes. This suggests that these sites under

positive selection could also be involved in CTL response. Four novel sites (sites 57, 285, 297 and 376) in the group that remained on 2nd line therapy and two novel sites (210 and 272) in the patients that were switched from the 1st to 2nd line therapy were detected in the RT region. The functional implication in positive selection at these sites is unknown; therefore these sites could be accessory mutations that improve the virus fitness rather than playing a role in drug resistance (Chen et al., 2004).

Signature difference at the PR and RT in these groups was also characterized. Interestingly, the PR positions 10 (PCSM001) and 70 (PCSM007) that were positively selected were also a signature difference in these patients. This suggests that there was a drug selective pressure driving selection of these mutations at these sites (Huang et al., 2012). There was no signature difference at drug resistance positions in the patients that were switched from 1st to 2nd line therapy. This result adds to the study by Huang et al., 2012 where they also found no signature difference at drug resistance position. Thus the suggestion of these results on the evolution of drug resistance needs further study.

Phylogenetic analysis was performed to investigate the evolution of the PR and the RT resistance mutations at the 1st and the 2nd time point. For most of the patients that remained on 2nd line therapy, there was intermingling between the sequences at the 1st and the 2nd time point with the exception of PCSM007 where three clones from the 1st time point clustered with the sequences at the 2nd time point (Figure 2.16). This suggests that the sequences at the 2nd time point evolved from the sequences at the 1st point (Halvas et al., 2010). In contrast, there was no intermingling between the sequences at the 1st and the 2nd time point in the patients that were switched from 1st to 2nd line therapy except for PCSM012 which had a few sequences intermingling with the sequences at the 2nd time point. Similarly, the phylogenetic clusters specifically in the patients that remained on 2nd line therapy showed both the sequences from the 1st and the 2nd time point while for the PI naïve patients, the phylogenetic clusters either showed sequences from the 1st or 2nd time point (Figure 2.13 and 2.14), except for PCSM0012. The tree topologies in each phylogenetic cluster could possibly be due to the presence of common drug resistant mutation and/or compensatory mutations that are associated with the increase in the viral fitness of the drug resistant virus (Hué et al., 2009). However, the presence of drug resistance mutations alone was not responsible for the tree topologies as the analysis of the trees where the drug resistance

mutations were removed showed similar results (data not shown). This result also adds to what has been reported by Kijak et al., 2002 and Bansode et al., 2011.

The diverse pattern characterized by different branch lengths observed in the phylogenetic clusters could be the results of distinct evolutionary pressures within the virus such as positive selection (Fischer et al., 2012). Ideally, the sequences with less positive selection suggest that there is less diversity and the high positive selection suggest high diversity. Thus the difference in the branch lengths is not associated with viral fitness; only the drug resistance mutations are associated with viral fitness (Hué et al., 2009). Similarly, the distinct viral population of sequence, one with the sequences at the 1st time point and the other with the sequences at the 2nd time point observed in PCSM004 and PCSM009 could be due to the evolutionary pressures that occurred at each time point.

Furthermore, the intra sequence diversity (genetic distances) in all the patients did not increase overtime and the overall diversity was low in all the isolates. However, the genetic diversity between the time points was higher than the genetic distances observed within the time points in all the isolates. Again, suggesting that the evolutionary pressure might have influenced this differences observed between the time points.

The phylogenetic trees were further analyzed to assess the evolution of the RT and the PR resistance mutations at the 1st and the 2nd time point. This analysis showed a diverse pattern of genotypes in the RT and the PR region, specifically in the patients that remained on 2nd line therapy. The sequences harboring the same drug resistance mutations clustered together. There are several explanations for the presence of the different genotypes in each of the sequences. HIV-1 has a large diversity due to presence quasispecies (viral variants of related but distinctly different virus). This makes it possible for the selections of different genotypes in the sequences (Charpentier et al., 2004). In the patients that remained on 2nd line therapy, some of the genotypes were still detected at the 2nd time point with a few drug resistance mutations no longer detected. This finding could be due to fact that the time difference between the time points was short i.e. < 3 months thus no difference was seen over time. In the case of the patients that were switched from a 1st to 2nd line therapy, most of the drug resistance mutations detected at the 1st time point were not observed at the 2nd time point except for PCSM012 which had a few drug resistance mutation still detected at the 2nd time point. In this event, the drug resistant mutations

reverted back to the wild type at the 2nd time point. This indicates that the patients were not drugs thus the wild type viruses reemerged from the viral population after the emergence of drug resistance mutations that occurred while they were on treatment (Hué et al., 2009). Interestingly, these wild type viruses were not observed in a single cluster but rather in sub clusters. taking their This suggests that this viruses originated by reemergence of the several viral clones (Kijak et al., 2002). This findings also adds to what has been reported by other studies that in the ARV treated patients who stops taking their treatment, the wild type viruses may be reemerging from the original viral variants that had occurred before the drug resistant mutations emerge (Kijak et al., 2002).

These findings that the patients that were switched from a 1st to 2nd line therapy (with the exception of PCSM012) showed the reemergence of the wild type at the 2nd time point can also be explained by the different occurrence of these wild type viruses in different patients, related with the time since the patients were on treatment. The observation that the drug resistant mutations were not detected at the 2nd time point shows that the fitness of the wild type viruses was higher at this time point (Kijak et al., 2002). In the case of PCSM012 at the 2nd time point, the sequences harboring the K103N mutation as those at the 1st time point coexisted with the sequences which were wild type. Again, this suggests that the wild type viruses at this time point could be drug resistant variants that were selected during treatment and remerged at the 2nd time point.

Chapter 3

Minority HIV-1 Drug resistance mutations by Ultra Deep Pyrosequencing

3.1 Introduction

Detection of minority variants is important in clinical practice for effective treatment strategies. More sensitive methods which have been used to detect minority variants have shown that the preexisting minority variants that are not detect by standard genotyping can influence treatment efficacy both in treatment naïve and experienced patients (Avidor et al., 2013). In this study, minority variants were detected by cloning as described in chapter 2 and these minority variants were related to resistance to their 1st line regimen, especially among the patients that were switched from a 1st to 2nd line therapy. Thus these results are consistence with other studies that the minority variants that occur at low frequency of the HIV-1 viral population do contribute to treatment failure. However, the cloning method have many disadvantages such as being extremely tedious, very expensive when trying to get a significant number of clones to detect minority variants to as low as 10 % and time consuming (Avidor et al., 2013, Paredes i Deiros, 2009, Ramirez et al., 2013).

Ultra Deep Pyro-Sequencing (UDPS) has been shown to be more effective in detecting minority HIV-1 drug (Avidor et al., 2013). UDPS can overcome the disadvantages of clonal sequencing as it involves the direct clonal sequencing of mixed samples and resulting in more than 10000 reads per base (Ramirez et al., 2013). It is also less time consuming and its sensitivity to detect minority variants is higher compared to cloning (Ramirez et al., 2013). In this study, UDPS was performed to compare the drug resistance mutations detected by cloning and the standard genotyping.

3.2. Materials and methods

3.2.1 Generation of Amplicons for UDPS

Amplicons were generated using a protocol for Amplicon Sequencing of HIV RT and PR yellow plate (Roche 454 life science). Briefly, the RNA was extracted from the plasma of samples using the QIAMP viral RNA mini kit (Qiagen) according to the kit protocol as described in Chapter 2.2.3. Then the extracted RNA was stabilized using a 10 ng/µl of MS2 carrier RNA. The cDNA synthesis was performed in cDNA synthesis 96 well plate and the amplicon generation was performed with primers as described in the protocol (Roche 454 life science).

3.2.2 Amplification purification

The PCR products were purified using the AMPure XP kit (Agencourt). Using a multi-channel pipette, 22.5 µl of the molecular grade water and 22.5 µl of the PCR products was added to each well on the 96 well plate according to the plate layout. The AMPURE beads bottle was mixed for 30 seconds, 45 µl of the beads were added to each well and mixed at least 12 times. The plate was incubated at room temperature for 10 minutes. Then the plate was placed on a 96 well magnetic ring and incubated for 5 minutes at room temperature. After 5 minutes the supernatant was removed and discarded without disturbing the beads. Subsequently, the plate was removed from the 96 well magnetic ring and the beads were washed with 70% ethanol. The plate was placed back on the magnetic ring for 1 minute and the supernatant was removed again. The wash step was repeated. After, the plate was placed back on the 96 magnetic ring stand and placed on a heat block set at 40 ˚C for 10 - 20 minutes. Ten µl of Tris Ethylediaminetet ethylenediaminetetraacetic acid (TE) was added to each well to elute the PCR products from the beads. The plate was incubated for 2 minutes and after the supernatant was transferred into a new 96 well plate. The plate was covered with plate seal and stored at -20˚C until ready to be quantified.

3.2.4 Amplicon Quantification

The amplicons were quantified by flourometry using a Quant-it Pico Green dsDNA Assay kit (Invitrogen). Firstly, a DNA serial dilution standard was prepared. The DNA standard was thawed. 1 × TE was transferred to 1.5 microcentrifuge tube as follows: 441 µl TE to tube 1 while 225 µl was transferred to tubes 2-8. Then a 50 × dilution was prepared by transferring 9 µl of DNA standard to tube 1 and vortex for 30 seconds. Subsequently, 225 µl from tube 1 was transferred to tube 2 and the serial dilution was completed by transferring 225 µl from tube to tube until to tube 7. Tube 8 served as a control DNA. One hundred µl of TE was transferred into each well of the 96 –well black flourometer.

Using a multi-channel pipette, 99 µl of 1 × TE buffer was transferred to the remaining wells and 1 µl of the DNA sample was transferred to each well containing the DNA sample. 100 µl of the diluted PicoGreen solution was added to each well and mixed by pipetting up and down. The flourometer plate was analyzed on the Glomax flourometer (Promega). The data was saved as an

excel file. The dilution was calculated using the rapid library calculator available on www.my454.com . This plots the standard curve and calculates the dilution volumes.

[DNA]	200	100	50	25.00	12.50	6.25	3.13	0	
Value	75222	38286	19045	9584	4942	2547	1386	212	R^2= 1.000 good
[Calc]	199.4	101.2	50.0	24.8	12.5	6.1	3.0	-0.1	

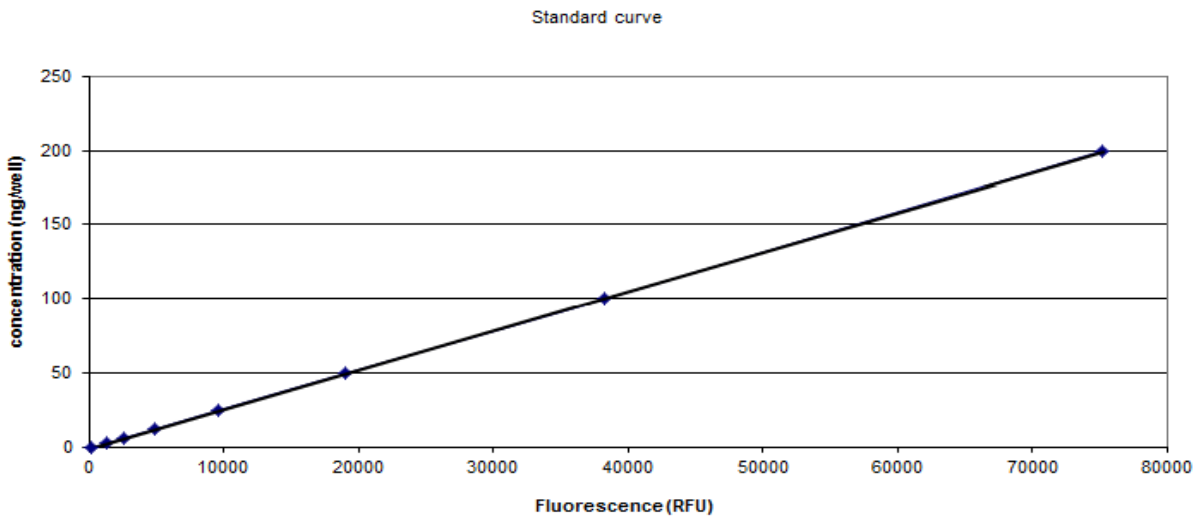


Figure 3.1: Quantification of the amplicons using the fluorometry. www.my454.com

All amplicons with a concentration ≤ 5 ng/µl were analyzed using the Agilent 2100 Bioanalyzer (Agilent Technologies) to further asses their quality. One µl of each purified amplicons were loaded on a DNA chip and then a DNA 1000 series II assay was run to quantify the PCR products with the required size. If the size of the band was ≥ 3:1 molar ratio relative to non – specific products, that amplicons was chosen and diluted according to the calculator and the amplicons were pooled equimolarly.

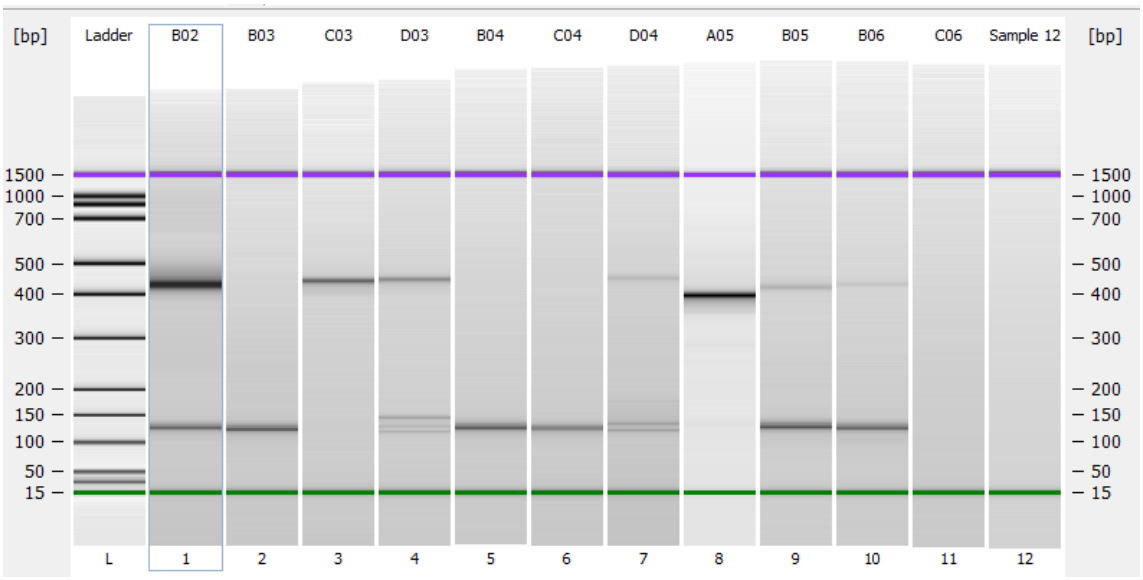


Figure 3.2: This is representative of the electrophoresis on the Agilent 2100 Bioanalyzer. Sample 2 and 3 showed primer dimer at a size of ~100bp. Sample 5 showed primer dimer in primer B only. Sample 4 didn‘t work due to insufficient amplicon.

3.2.5 Amplicon pooling and dilution for emPCR amplification.

After the dilutions and pools were prepared, 22.5 µl from the pool was transferred to a new well of the clean round bottom plate. 22.5 µl of the molecµlar grade water and 45 µl AMPure beads were added to the wells and mixed thoroughly by pitting up and down at least 12 times. The plate was incubated at room temperature for 10 minutes. The plate was then placed on the 96 well magnetic ring stand and incubated for 5 minutes at room temperature. After 5 minutes, the supernatant was carefully removed. The plate was then removed from the magnetic ring stand and washed twice with 100 µl of freshly prepared 70% ethanol. The plate was then placed again on the magnetic ring stand and incubated for 1 minute and the supernatant removed. The washing step was repeated once more. The plate was then placed back on the magnetic ring stand and the pellet was dried at 40 ˚C on a heat block for 20 minutes. 22.5 µl of 1 ×TE was added to each well to elute the PCR product from the magnetic beads. The plate was then incubated for 2 minutes while still on the magnetic ring. The purified PCR products (supernatant) was removed and transferred to 1.5 ml tube and stored at -20 ˚C. A further dilution of the intermediate pool of 2×10^6 molecules was prepared by adding 1 µl of the 1×10^9 molecules to 499 µl TE in a 1.5ml tube and this final pool was stored was stored at – 20 ˚C.

3.2.6 Emulsion PCR amplification (emPCR).

Emulsion PCR which is the emulsion based clonal amplification was performed using emPCR kit (Lib A) following the manufactures protocol (Roche, Applied Biosystem). Briefly, during the emPCR, 2.0 ml of 1 × Mock mix was added to the Turrax stirring tube containing the emulsion oil. The stirring tube was placed in the Ultra Turrax Tube Drive (UTTD) and run at 4000 rpm for 5 min to mix the emulsion. The DNA library capture beads were prepared by adding 1 ml of the prepared 1 × wash buffer to each tube of the capture beads. The capture beads were placed in a bench top minifuge to pellet the capture beads. The supernatant was then removed. The capture beads were washed again with 1 ml of 1 × wash buffer and placed in a bench top minifuge to pellet. The amplicon DNA library to be amplified was thawed and the required volume was calculated as follows:

$$5\ \mu l \text{ of library} = \frac{2 \text{ molecules per bead} \times 5 \text{ million beads}}{2 \text{ million molecules}}$$

Therefore, 10 µl of DNA library was added to 30 µl of molecular grade water to make 0.5 copies of the DNA library bead. The DNA library bead was added to washed capture beads A and B. Six hundred µl of the Live Amp Mix A and B were added to the tube of the capture library A and B respectively and the mix transferred into the Turrax string tube and placed in the UTTD set at 200 rpm for 5 minutes. One hundred µl of the emulsion reaction mix was transferred into a 96 well plate and the plate was sealed. The plate was placed in the thermocycler and the amplification was carried out following this conditions:

Temperature ˚C	Time	Cycles
94	4 minutes	1
94	30 seconds	50
58	4.5 minutes	
68	30 seconds	
10	For hold	

3.2.7 Bead washes and recovery

The emulsion was aspirated from the plate into a 50 ml tube using GS Junior Titanium emPCR oil and breaking kit (Roche 454 life science). The plate was washed twice with 100 µl of isopropanol. Additionally, 5 ml of isopropanol was added to collect any beads that may remain in the tubing. For the bead recovery, the beads were washed several times using centrifugation to complete the breaking procedure (Roche 454 life science).

3.2.8 DNA library and enrichment

The enrichment beads tubes were vortex for 1 minute to resuspend its contents. The tubes were then placed in the Magnetic Particle Collector (MPC) and incubated for 3 minutes to pellet the enrichment beads and the supernatant removed. 500 µl of the enhancing buffer was added to the enrichment bead and pellet the enrichment beads using an MPC. The latter step was repeated once more. The beads were resuspended in 80 µl of the enhancing buffer and vortex.

During the enrichment of the DNA carrying beads, 80 µl of washed enrichment beads were mixed with DNA beads and placed on the rotator at room temperature for 5 minutes. The tube was placed in the MPC for 3-5 minutes to pellet the enrichment beads and the supernantant removed. The beads were then washed 6 to 10 times with 1ml of enhancing buffer and the tubes were placed into the MPC to pellet the beads and the supernatant removed.

Following this, the collection of the beads was done as follows: the tube of the enriched beads was removed from the MPC and resuspended in 700 µl of melt solution, vortex for 5 seconds and placed again in the MPC to pellet the beads. The supernatant containing the enriched DNA beads was transferred into a new 1.7 ml and resuspended once again with 700 µl of melt solution and the supernatant transferred to same tube. The DNA beads were then washed with 1 ml of the annealing buffer and vortex for 5 seconds, spin- rotate-spin and the supernatant was discarded. This washing step was repeated two times. Finally, 100 µl of the annealing buffer was added to the beads and vortex.

For the sequencing primer annealing, 15 µl of Seq Primer A and Primer B were added to the DNA capture beads, vortex and the tube was placed in a heat block at 65 ˚C for 5 minutes and

subsequently cooled on ice for 2 minutes. The beads were then washed twice with 1 ml annealing buffer, vortex for 5 seconds and spin-rotate-spin and the supernatant was removed.

The amount of enriched beads was evaluated using the GS junior beads counter. According to the kit protocol the recommended input bead number for UDPS is 500, 000 enriched DNA beads (Avidor et al., 2013).

3.2.9 Ultra deep pyro sequencing (UDPS).

The UDPS was performed using the GS Junior Titanium sequencing kit together with the matching GS junior Titanum PicoTiterPlate (PTP) kit following the kits protocol (Roche 454 life science). Briefly, the initial steps involved in the GS junior sequencing procedure includes the pre wash of the GS junior instruments with pre wash buffer, washing step of the PTP plate with the Bead Buffer 2 (BB2). Four bead layers were prepared: Enzyme bead pre-layer, DNA and packing beads, enzyme beads post layer and the PPiase Beads. Firstly, an aliquot of the enriched DNA beads was mixed with 6 µl of the control beads XLTF and 500 µl of BB2 was added to the beads, vortex briefly and incubated for 20 minutes at room temperature on the lab rator. While waiting for the 20 minutes, the packing beads, enzyme beads post layer were washed three times with 1 ml with BB2 and centrifuge at 9.300 RCF for 5 minutes for each wash. 175 µl of packing beads were added to DNA mixture containing the polymerase, polymerase cofactor and BB2, vortex at low speed and incubated on the lab rotator at room temperature for 5 minutes. The PTP was loaded first with 350 µl of the bead suspension (enzyme bead pre layer) and this was centrifuge for 5 minutes at 1, 620 RCF. After 5 minutes the supernatant was removed and 350 µl of the second layer was loaded to the PTP. This was centrifuge for 10 minutes at 1.6220 RCF. Again after 10 minutes the supernatant was removed and the enzyme post layer was loaded and then centrifuge at 1,620 RCF for 10 minutes. After centrifugation, the supernatant was removed and subsequently 350 µl of the last layer (PPiase beads) was loaded to the plate and centrifuged in BDD for 5 minutes at 1,620 RCF. Finally, the reagent cassettes were loaded onto the GS junior instrument, the PTP plate was loaded and the sequencing programme started.

3.2.10 Data analysis

The sequence analysis was performed using the Amplicon Variant GS Amplicon Variant (AVA) software (Roche 454 life science). Briefly, the AVA software detects and quantitates the known variants as well as the novel DNA variants. It computes the alignment of the read sequences and identifies variation between the reads and the reference sequence. The identified variants are represented by histograms that shows the variant position and multiple alignments that are color coded highlighting the regions and bases that are different from the reference sequence (Figure 3.3 and 3.4).

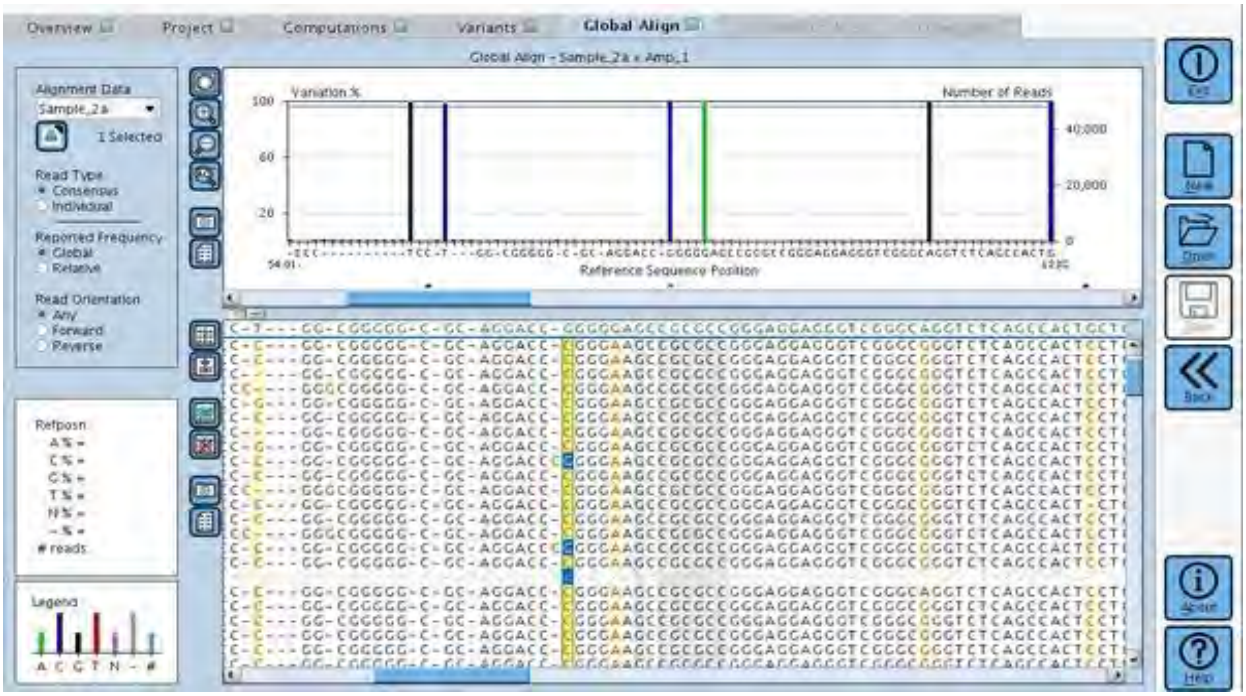


Figure 3.3: The GS junior Amplicon variant analyzer showing the base substitution point that is different from the reference sequence. The top graph shows the mutation frequency of the sample. www.roche-biochem. 3/03/2014.

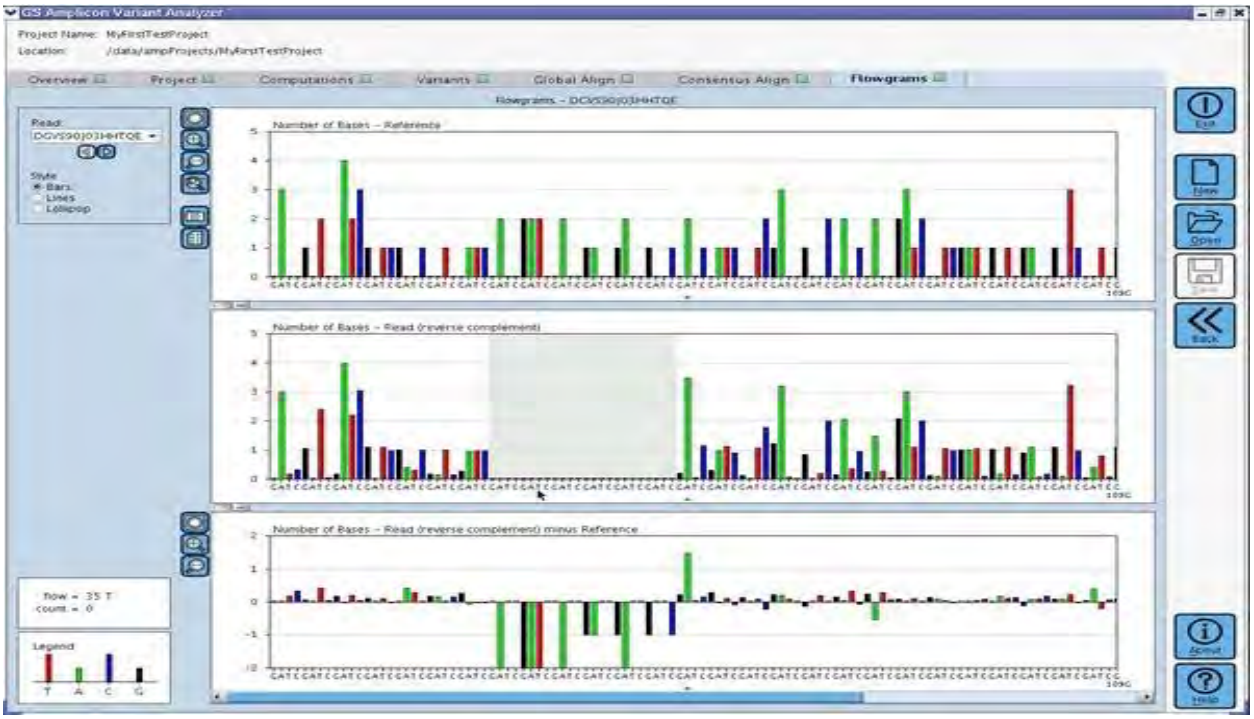


Figure 3.4: GS junior Amplicon variant analyzer showing multiple alignments. The top graph represents the reference sequence, the 2nd graph represent the difference between the reference sequence and the sample sequences while the 3rd graph shows the difference between the two graph above such that if the sample sequences is the same as the reference sequence the histogram bar height is zero. Source: (Roche 454 life science, www.roche-biochem.jp)

This AVA software therefore assigns each read to the proper amplicon and patient sample. It also aligns the generated sequences and reports the frequency of identified drug resistance mutations in a summary table according to the international AIDS Society USA (ISA-USA) and the Stanford algorithm (Avidor et al., 2013). Using the AVA project manager, a data file of the results was created and exported into excel where the data was analyzed.

3.3 Results

The patients that were switched from a 1st to 2nd line therapy were selected for UDPS because they had no PI resistance mutations detected by standard genotyping and cloning when they were switched from a 1st line to 2nd line therapy.

Despite considerable effort in performing UDPS for HIV RT and PR, PCR amplification was not successful for all the DNA samples. For PCSM004, only the 2nd time point PCR amplification was successful while for PCSM009 and PCSM012 the amplicons from both time points were successfully sequenced. Although amplification was successful for PCSM012 in both time points, sequence of the 1st time point amplicon was unsuccessful.

In all the samples, most of the minority variants that were detected by UDPS were also detected by cloning (Table 3.1- 3.3). For PCSM004, at the 2nd time point, UDPS detected two additional minority variants that were not detected by cloning: NRTI mutations K65N and K65R with a frequency of 2.32 % and 0.73% respectively. The NNRTI mutation V179D which was detected by cloning was also detected by UDPS but the frequency was low (1.54%) compared to the one detected buy cloning (14%) (Table 3.1). No drug resistance mutations were detected in PR at this time point for both cloning as well as UDPS. However, the only mutation detected in PR was L10Y which is described as being a non-polymorphic accessory PI mutation (Table 3.1).

Tables 3.1-3.3 shows the comparison of the mutations detected by the three methods used in this study i.e. Standard genotyping, Cloning and UDPS. The red marked mutations shows additional mutations detected by cloning as described in chapter 2 (Tables 2.12- 2.14) while the blue marked mutations shows the additional mutations detected by UDPS.

Table 3.1: Comparison of mutations detected by standard genotyping, cloning and UDPS for PCSM004. The red marked mutations are additional mutations detected by Cloning while the blue marked mutations are additional mutations detected by UDPS.

Mutations	Mutations detected by standard genotyping		Frequency (%) of mutations detected by cloning		Frequency (%) of mutations detected UDPS	
PI	1st time point	2nd time point	1st time Point	2nd time point	1st time Point	2nd time point
p-L10Y	no	no	0	0	NA	**0.21**
NRTI						
M41L	yes	no	73	0	NA	0
K65R	no	no	0	0	NA	**2.32**
K65N	no	no	0	0	NA	**0.73**
M184V	yes	no	97	0	NA	0
NNRTI						
K103R	yes	no	93	0	NA	0
V106M	yes	no	97	**4**	NA	0
V179D	yes	no	97	**14**	NA	**1.54**
F227L	yes	no	33	0	NA	0

***NA- not applicable**

For PCSM009, again the minority variants that were detected by UDPS were also detected by cloning (Table 3.2). At the 1st time point, the RT drug resistance detected by all the methods included major NRTI mutations D67N and M184V and NNRTI mutations K101E, V106M, E138G/K and G190A. The frequencies of these mutations were very similar in both methods (Table 3.2). The additional RT mutations detected by UDPS included the NRTI mutations K65R (0.65 %), and the NNRTI mutation F227L (16.68%) (Table 3.2). At the 2nd time point, the NRTI mutations D67G, D67N, K65R, and NNRTI mutations K101E, V106M, E138G, E138K, G190A and F227L were also detected by UDPS. The frequencies of these mutations decreased at this time point but they were higher compared to the frequencies detected by cloning (Table 3.2).

Again, no major mutations were detected in PR at either time points for both cloning and UDPS. However, accessory mutation A71T (1.19%), was detected at this time point by UDPS.

Table 3.2: Comparison of mutations detected by standard genotyping, cloning and UDPS for PCSM009. The red marked mutations are additional mutations detected by Cloning while the blue marked mutations are additional mutations detected by UDPS.

Mutations	**Mutations detected by standard genotyping**		**Frequency (%) of mutations detected by Cloning**		**Frequency (%) of mutations detected by UDPS**	
PI	1st time point	2nd time point	1st time point	2nd time point	1st time Point	2nd time point
p-A71T	no	no	0	0	0	**1.19**
NRTI						
D67G	no	no	0	**3**	0	**0.24**
D67N	yes	no	65	0	0	**0.53**
K65R	no	no	0	0	**0.64**	**0.75**
M184V	yes	no	100	0	99.8	0
NNRTI						
K101E	yes	no	100	**10**	92.07	**24.54**
V106M	yes	no	100	**10**	100	**24.61**
E138G	yes	no	9	**7**	4.19	**24.79**
E138K	yes	no	91	0	87.88	**0.46**
G190A	yes	no	100	**10**	99.85	**98.08**
F227L	no	no	0	0	**16.68**	**1.8**
p-A98S	no	no	0	0	0	**0.46**
p- V90I	no	no	0	0	0	**0.14**
p-V179I	no	no	0	0	**4.63**	**2.06**

For PCSM012, the minority variants detected by cloning at the 1st time point were also detected at the 2nd time point by UDPS (Table 3.3). The RT drug resistance mutations included the NRTI mutation T69A and the NNRTI mutations K103N, V106M, E138K and G190A. V106M, E138K

and G190A were only detected by UDPS at the 2nd time point. For mutations T69A and K103N, the frequency was low compared to the frequency detected by cloning. For cloning, the frequency for T69A was 62% while for UDPS it was 24.95%. For the mutation K103N, frequency detected by cloning was 66% by cloning while for UDPS it was 25.13% (Table 3.3).

Table 3.3: Comparison of mutations detected by standard genotyping, cloning and UDPS for PCSM012. The red marked mutations are additional mutations detected by Cloning while the blue marked mutations are additional mutations detected by UDPS.

Mutations	Mutations detected by Standard genotyping		Frequency (%) of mutations detected by Cloning		Frequency (%) of mutations detected UDPS	
PI	1st time point	2nd time point	1st time point	2nd time point	1st time point	2nd time point
-	no	no	0	0	NA	0
NRTI						
T69A	no	no	**14**	**62**	NA	**24.95**
T69D	no	yes	0	0	NA	0
NNRTI						
K103N	yes	yes	26	66	NA	**25.13**
V106M	no	no	**9**	0	NA	**24.61**
E138K	no	no	**3**	0	NA	**24.79**
G190A	no	no	**3**	0	NA	**0.46**

***NA- not applicable**

In addition, to analyze the mean number of minority drug resistance mutations detected specifically by cloning and UDPS, a paired student Test using Graph Pad prism was done. This analysis was performed on the drug resistance mutations detected at the 2nd time point as the samples at the 1st time point was not successful for all the samples. The mean number of difference of mutations between the methods was 2.667 with 95 % CI -1.128- 6.461. The differences in mutations detected per drug class are shown in Figure 3.1.

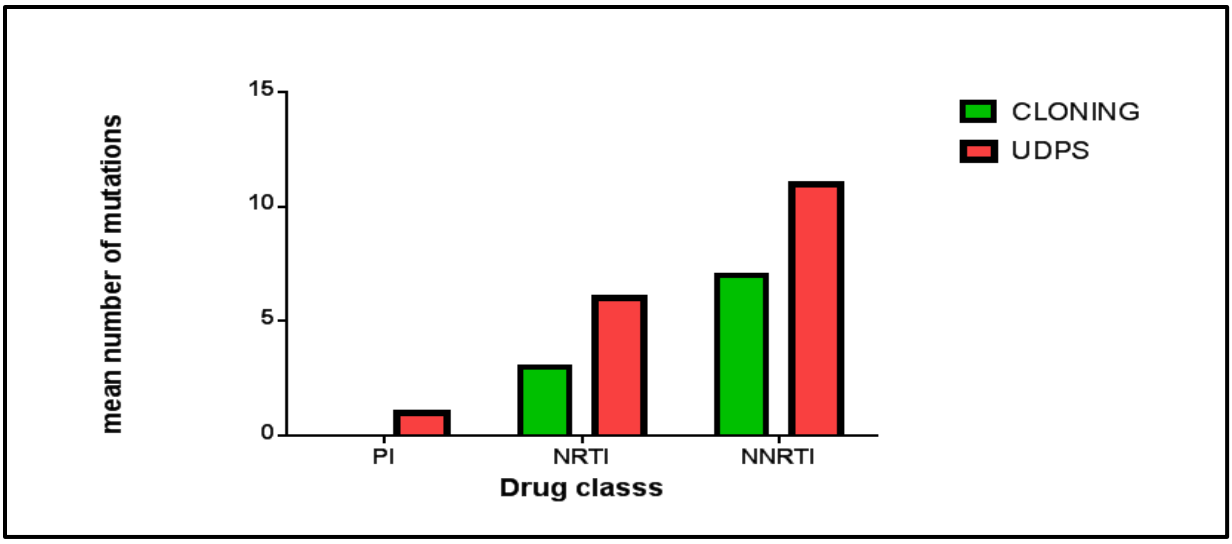


Figure 3.5 Mean number of drug resistance mutations (Drug class) detected by cloning and UDPS.

In addition, one sample from the patients that remained on 2nd line therapy (PCSM007) was also sequenced by UDPS as described above. This was the only sample available from this group and only the 1st time point sample was available for sequencing.

In this patient, at this time point, the PR drug resistance mutations detected by all the methods included L10F, M46I, I54V and V82A. The frequency of this mutations detected by UDPS was high compared to the frequency detected by cloning (Table 3.4). No additional PR mutations were detected by UDPS at this time point. The RT mutations detected by all the methods included the NRTI mutations D67N, M184V and NNRTI mutations K103N and V106M. The frequency detected by UDPS for the mutation D67N was low (91.87%) compared to the frequency detected by cloning (96%) while the frequency for M184V was high (97.42 %) compared to the frequency detected by cloning (43%). For the mutation K103N, the frequency detected by cloning was 93 % while for the UDPS it was 97.71%. For the mutation V106M, the frequency detected by cloning (99.42%) and UDPS (100%) were very similar (Table 3.4). Additional RT mutations (blue marked mutations) detected by UDPS included the NRTI mutation K65R (0.75%), accessory mutations F77L (0.75%), V75I (0.23%) and major NNRTI mutation K103S (1.3%) (Table 3.4). However, some of the mutations that were detected by standard genotyping and additional mutations detected by cloning were not detected by UDPS. This includes the PR mutations K20R, A71V, L76V, NRTI mutations K70R and NNRTI mutations K101E, E138K, G190A (Table 3.4).

Table 3.4: Comparison of mutations detected by standard genotyping, cloning and UDPS for PCSM007. The red marked mutations are additional mutations detected by Cloning while the blue marked mutations are additional mutations detected by UDPS.

Mutations	Mutations detected by standard genotyping		Frequency (%) of mutations detected by cloning		Frequency (%) of mutations detected UDPS	
PI	1st time point	2nd time point	1st time point	2nd time point	1st time point	2nd time point
L10F	yes	yes	57	82	86.78	NA
K20R	no	no	**50**	**18**	0	NA
M46I	yes	yes	87	96	97.7	NA
I54V	yes	yes	87	18	100	NA
A71V	yes	yes	50	79	0	NA
L76V	yes	yes	50	89	0	NA
V82A	yes	yes	87	96	100	NA
NRTI						
D67N	yes	yes	96	96	91.87	NA
K65R	no	no	0	0	**0.75**	NA
K70R	no	no	**53**	**14**	0	NA
V75I	no	no	0	0	**0.23**	NA
F77L	no	no	0	0	**0.34**	NA
M184V	yes	no	43	14	97.42	NA
NNRTI						
K101E	no	no	0	**57**	0	NA
K103N	yes	yes	93	96	97.71	NA
K103S	no	no	0	0	**1.3**	NA
V106M	yes	yes	100	96	99.42	NA
E138K	no	no	**14**	**57**	0	NA
G190A	no	no	**14**	**57**	0	NA
F227L	no	no	0	0	**0.34**	NA

3.4 Discussion

HIV-1 drug resistance testing has been based on standard genotyping and the detection limitation of standard genotyping is that it detects mutations >20% of the viral population but not the minority variants which occur at low levels of the viral population (Charpentier et al., 2004). To overcome this, more sensitive methods have been developed to detect minority variants (Paredes i Deiros, 2009). In the previous chapter, cloning, which has a sensitivity of >10% in detecting minority variants (Paredes i Deiros, 2009), was used to described resistance mutations. A newly developed method, UDPS is now being used to detect minority variants as an alternative to cloning. UDPS is more sensitive than cloning in detecting minority variants and it has a sensitivity of 0.5-1% (Paredes i Deiros, 2009). In this chapter the minority variants detected by UDPS are compared to those detected by cloning and standard genotyping.

Recent studies applying this method have found that UDPS identified a significantly larger proportion of minority variants compared to standard genotyping and these minority variants increased the risk of virological failure (Simen et al., 2009). In another study by Le et al., 2009, they compared standard genotyping and UDPS in treatment experienced patients, they found that minority variants were detected in all the patients by UDPS and non by standard genotyping and also these minority variants correlated with virological failure. In addition they stated that additional minority variants increased the patients‘ resistance to at least one or two ARVs (Le et al., 2009).

In this study, samples from the 2nd time point of the patients that were switched from a 1st to 2nd line therapy were sequenced by UDPS, and where available. This group was of particular interest, as these patients were switched from their 1st line regimen (PI naive) to their 2nd line regimen (PI inclusive), allowing us to investigate the presence of baseline minority PI mutations, as well as the development of minority PI mutations once on the PI. However, no mutations were detected, even using UDPS in these patients, suggesting that this patients might have poor adherent to PI regimen or they required more time to develop PI mutations as the development of PI resistance is a gradual process and require more mutations to be resistance to PI (Charpentier et al., 2004, Shafer, 2002).

UDPS did detect additional NRTI mutations (D67G, D67N and K65R) when compared to cloning, but this difference was not significant. Only one isolate was sequenced from the patients

that remained on 2nd line therapy. Here, additional NRTI mutations were detected (K65R, V75I and F77L). No additional PI mutations were detected; this could be likely due to that more time was required for development of additional PI mutations as these results were from the 1st time point.

Interestingly, the K65R mutation, which causes resistance to NNRTIs abacavir (ABC), didanosine (ddi), emtricibine (FTC), lamivudine (3TC), tenofovir (TDF) and low level to stavudine (d4T), was found at low levels in 3 isolates (0.75.% - 2.32 %). This mutation has been strongly related with the development of virological failure in Subtype C infected individuals (Bansode et al., 2013, Fischer et al., 2012). It has been reported that in HIV subtype C, the RT KKK nucleotide motif at codons 64-66 seems to results to template pausing that increase the selection of K65R (Recordon-Pinson et al., 2012). However, K65R mutation has been reported to be an UDPS artifact due to PCR errors occurring at this position. This error depends on the template codon in the encoding nucleotide sequence (64 -66) of the RT gene which is different among subtypes B and C (Bansode et al., 2013, Fischer et al., 2012). In Subtype C the encoding nucleotide sequence is either AAA-AAG-AAG or AAA-AAG-AAA while for subtype B is either AAG-AAA-AAA or AAG-AAA-AAG (Fischer et al., 2012). To prevent this error from occurring, a high fidelity enzyme was used in the amplicon generation.

In the group that was switched from a 1st to 2nd line therapy, UDPS also detected additional NNRTI variants in two patients (PCSM004 and PCSM012) with at a higher frequency than cloning. In contrast, the additional NNRTI variants (K103S and F227L) detected in PCSM007 was low and they were not detected by cloning.

In a study where they were comparing the minority variants detected by cloning and UDPS, as expected UDPS yielded more mutations than cloning (Ramirez et al., 2013) but they also stated that cloning included insertions and deletions which was not detected by UDPS (Ramirez et al., 2013). In contrast, in a study by Lanier et al., 2007 where they also compared UDPS and cloning found out that UDPS yield similar data to cloning and their frequency was not significantly different. Both methods showed low levels of L74V and K65R. This finding is consistence with the results we found in this study. Almost all the mutations found by UDPS were also detected by cloning and the frequency of some of the minority variants detected by both methods was very similar. However, some of the minority variants detected by UDPS had a low frequency

compared to cloning. This could be due to that the depth of coverage at those variant positions was low.

There are many factors that could contribute to the detections of minority variants by UDPS. This could be due to the resistant viral population decaying to levels that cannot be detected by cloning and standard genotyping in the absence of drug pressure (Fischer et al., 2012). This usually applies to the patients with poor adherence allowing resistance variants not to reach the threshold (Fischer et al., 2012). This is also seen in our patients that were switched from their 1st to 2nd line therapy that were investigated. We expected UDPS to detect mutations in the PR region other than the accessory mutations, but this was not the case. This suggests that these patients were not adherent.

In other studies where they were also investigating the minority variants in patients who were switched to 2nd line therapy, UDPS did detect minority variants at 2nd line failure (Fischer et al., 2012). Based on their results they stated that UDPS may be a better approach in detecting the development of PI resistance before it become fixed in the viral population (Fischer et al., 2012). It has been suggested that most of the minority variants would first occur as a single variant and would represent the first step in the development of PI resistance. However, more mutations would be required to develop PI resistance (Charpentier et al., 2004, Fischer et al., 2012, Shafer, 2002). In patients with poor adherence the wild type viruses are favored and the resistant variants are less fit in the absence of drug selective pressure. This also explains the absence of PI resistance by both UDPS and cloning which could be due to poor adherence to PI regimens in our cohort. In addition, it is challenging to investigate the evolution of PI resistance as it would require several time points to study the multiple pathways to resistance, as PI resistance is a gradual process and it has a high genetic barrier (Charpentier et al., 2004, Fischer et al., 2012, Shafer, 2002). Therefore to achieve this, a large number of samples of patients failing PI therapy would be advantageous.

The limitations of this study include the small sample size, as only 3 patients from the patients that were switched from a 1st to 2nd line therapy were available for UDPS and only the samples at the 2nd time point were available for two patients (PCSM004 and PCSM012). In addition, one sample from the patients that remained on 2nd line therapy was sequenced and also with only one time point available (1st time point). It would have been informative if UDPS was performed in

all the samples in both time points. Secondly, UDPS is very expensive and therefore, samples that failed during the sequencing were not repeated. Thirdly, UDPS is also subjected to PCR and sequencing artifacts which may affect its specificity for example in the case of the detection of mutation K65R which is likely to be PCR artifact (Fischer et al., 2012).

Finally, while UDPS is expensive, cloning is very time consuming, as well as expensive when trying to get a significant number of clones to detect minority variants to as low as 10 % (Paredes i Deiros, 2009). Therefore, UDPS would be a valuable tool for resistance genotyping, especially for the detection of minority variants.

Chapter 4

Contribution of Gag Mutations to Protease Inhibitor Resistance

4.1 Introduction

HIV PR is responsible for the formation of enzymes and proteins of the mature virus through the process of post translational processing of the viral Gag and Gag-pol (Clavel and Mammano, 2010). The development of PI resistance is generally thought to be a gradual process and the increase in the level of resistance requires the presence of many mutations in the gene itself (Charpentier et al., 2004, Shafer, 2002). Resistance mutations that causes changes in the PR has also been shown to decrease the viral replication capacity due to decreased processing of the PR natural substrate (Dam et al., 2009). Apart from the mutations directly affecting the PR, gag mutations have been reported to contribute to the development of PI resistance (Clavel and Mammano, 2010, Dam et al., 2009). Many studies have shown that gag mutations can also contribute to PI resistance in the absence of PR mutations and mostly this was observed in the protease cleavage site (CS) (Ghosn et al., 2011, Nijhuis et al., 2007).

In this study, to further investigate the contribution of gag mutations to protease inhibitor resistance the gag-protease region of the patients that were genotyped at 1st line failure (1st time point), as well as at 2nd line failure (2nd time point) and those that were genotyped at 2nd line failure only (both time points) was sequenced and analyzed.

4.2. Materials and methods

4.2.1 RNA extraction from plasma sample.

RNA was extracted from the plasma of samples using the QIAMP viral RNA mini kit (Qiagen) according to the manufactures protocol (Chapter 2.2.3). RNA was stored at -80C until used.

4.2.2 Reverse transcription and amplification of HIV Gag –protease.

Reverse transcription (RT –PCR) was performed using a superscript III one step RT PCR kit (Invitrogen, life technologies). For the 1st round PCR, primers Gag+1 and 3‘ rvp were used while for the 2nd round PCR, 100- mer forward and reverse were used (Wright et al., 2010). The primers for the 1st round are shown in the table below:

Table 4.1: The HXB2 position and primers used to amplify the gag region.

Primer name	HXB2 position	Primer sequence
Gag+1 (10uM)	675- 697 F	GAGGAGATCTCTCGACGCAGGAC
3'rvp (10uM)	2725- 2696 R	GGAGTGTTATATGGATTTTCAGGCCCAATT

The Primers used for the 2nd round were as follows:

1. Forward primer

ACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTAC
GCCAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGG

2. Reverse primer

GGCCCAATTTTTGAAATTTTTCCTTCCTTTTCCATTTCTGTACAAATTTCTACTAAT
GCTTTTATTTTTTCTTCTGTCAATGGCCATTGTTTAACTTTTG

The RT mix was prepared in the master mix room. The 2 × reaction mix buffer and primers, were thawed and vortexed for 5 seconds. The enzyme superscript III was not vortexed. Care was taken to be quick when preparing the RT PCR master mix for the 1st round as superscript III is unstable at room temperature. The RT master mix was prepared in a 1.5 ml micro centrifuge as follows:

Reagents	Volume (µl)	Final concentration
2 × Reaction mix buffer	12.5	× 1
Gag +1 (10µM)	0.5	0.2 Um
3rvp (10µM)	0.5	0.2µM
DEPC	3	-
Superscript III (200U/µl)	1	4U/µl
RNA template	7.5	-
Total	25	

A volume of 17.5 µl was added to the labeled 0.2 ml reaction tube 7.5 µl of RNA was added and mixed properly. The conditions for the 1st round PCR product were as follows:

Temperature ˚C	Time	Cycles
55	30 minutes	1
94	2 minutes	1
94	15 seconds	35
55	30 seconds	
68	2 minutes	
68	5 minutes	1
4	Hold	

The master mix for the 2nd round PCR product was prepared as follows:

Reagents	Volume (µl)	Final concentration
DEPC	37	-
10 × Taq Buffer	5	×
dNTPs (2.5mM)	4	0.2µM
Forward long primer (10µM)	0.8	0.16µM
Reverse long primer (10µM)	0.8	0.16µM
Ex taq (250U)	0.25	1.25U/µl
1st round PCR product	2	-
Final volume	49.85	

A volume 47.8 µl of PCR master mix was added to each labeled 0.2 ml tube and 2 µl of the 1st round PCR product was added to the tubes. The PCR was run on thermocycler using the following condition:

Temperature ˚C	Time	Cycles
99	2 minutes	1
94	30 seconds	40
60	30 seconds	
72	2 minutes	
72	7 minutes	1
4	Hold	

4.2.3 Amplicon purification and quantification

The PCR products were purified using a Qiagen PCR clean up kit following the manufacturer's instructions (Qiagen) as described in Chapter 2.2.2. The PCR products were quantified by gel electrophoresis using a low DNA mass ladder to determine the concentration of the PCR products (Figure 4.1).

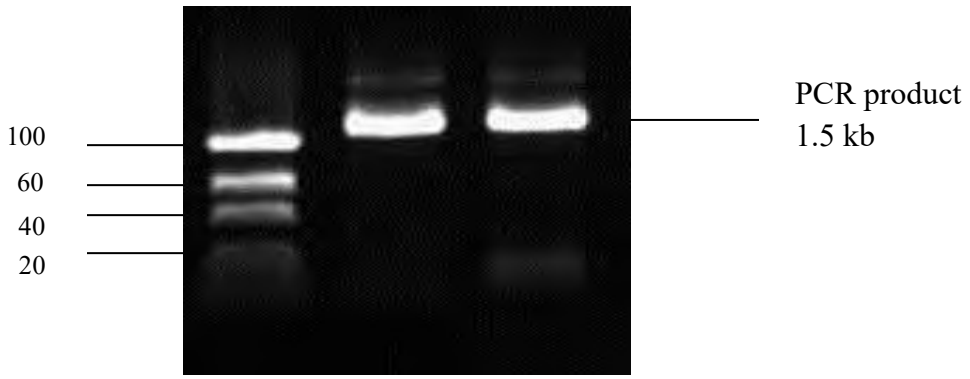


Figure 4.1 An Agarose gel electrophoresis after the PCR cleanup. The Low DNA mass ladder (left) shows the amount of DNA fragment in ng. The size of the PCR product is approximately 1.5 kb.

4.2.4 Sequencing of the PCR products

The PCR products were diluted to 2 ng/µl which is an optimal concentration for sequencing of PCR products. The HXB2 position and sequences of the 5 primers used for the sequencing of the protease-gag region are represented in the table below:

Table 4.2: The HXB2 position and primers used to sequence the gag region.

Primer name	HXB2 Position	Primer Sequence
SQ2FC	991- 1010	CTTCAGACAGGAACAGAGGA
GF100	1817- 1834	TAG AAG AAA TGA TGA CAG
SQ16RC	1098- 1078	CTTGTCTAGGGCTTCCTTGGT
GAS4R	1481- 1462	GGTTCTCTCATCTGGCCTGG
GR1981	1981- 1960	CCT TGC CAC AGT TGA AAC ATT T

Sequencing was carried out as described in chapter 2.2.7 expect that the master mix for each primer was prepared as follows:

Reagents	Volume (µl)	Final concentration
Big Dye ready reaction mix	0.4	-
5×sequencing buffer	2	1×
Primer (3.2 pmol/µl)	1	0.32 pmol/µl
Water	1.6	-
Final Volume	5	-

The sequencing was set up in a micro Amp optical 96 well reaction plates as described in chapter 2.2.7, except that a volume of 5 µl of each reaction primer mix was added to the selected well and of 5 µl of diluted PCR product.

4.2.5 DNA Sequence analysis

Sequence analysis was done as described in chapter 2.2.8. The gag-protease mutations were analyzed based on literature studies and those mutations that have been reported to be associated with drug resistance. This includes the V35I, G123E, H219Q, G381S, V390I, A431V, L449P, and P453L (Fun et al., 2012, Gatanaga et al., 2002, McKinnon et al., 2011).

4.2.6 Statistical analysis

The correlation between the PR and gag mutations was determined using the spearman rank order correlation using the GraphPad Prism 6.

4.3 Results

In order to investigate the contribution of gag mutations to protease inhibitor resistance in these patients, the gag-protease gene was analyzed.

4.3.1 Mutations occurring in the gag-protease region in patients that were switched from a 1st to 2nd line therapy.

None of these patients that were switched from a 1st to 2nd line therapy harbored mutations in the PR region. However gag mutations V35I, G123E, H219Q and gag cleavage site (CS) mutation G381S were detected in some of these patients (Table 4.1).

For PCSM004, at the 1st time point, two gag mutations were observed: mutations G123E and V35I which are located at the MA/p17. These mutations are associated with drug resistance (Gatanaga et al., 2002). However these mutations are rarely described in the literature. After nine months, a 2nd genotype was performed and the gag mutations that were detected at the 1st time point were still detected with no additional mutations in gag.

For PCSM009, at the 1st time point the mutations G123E located at the MA/p17, H219Q at the capsid (CA)/p24 and CS mutation G381S were detected. Again these mutations are associated with drug resistance (Gatanaga et al., 2002). After twelve months, a 2nd genotype was performed. Again, the gag mutations that were detected at the 1st time point were still detected. Interestingly, this patient had no PI mutations in both time points but harbored H219Q at the 1st and 2nd time point and it has been reported that this mutation is selected during PI exposure (Fun et al., 2012). This suggests that this mutation also play a role in PI resistance even in the absence of PI resistance mutations (Ghosn et al., 2011, Nijhuis et al., 2007).

For PCSM012, at the 1st time point only the gag CS mutation G318S was observed. Again this mutation is associated with drug resistance (Gatanaga et al., 2002) but it has been rarely

described in the literature. After eight months, a 2nd genotype was performed. Again for these patients, the gag mutations that were detected at the 1st time point were still detected.

Table 4.3 Mutations occurring in the gag and PR of patients that were switched from a 1st to 2nd line therapy.

Patient	PR mutations		Gag mutations	
	1st time point	2nd time point	1st time point	2nd time point
PCSM004	None	None	G123E, V35I	G123E,V35I
PCSM009	None	None	G123E, H219Q, G381S	G123E, H219Q, G381S
PCSM012	None	None	G381S	G381S

4.3.2 Mutations occurring in the gag-protease region in patients that remained on 2nd line therapy.

All these patients had several gag and PR mutations. The most common gag mutations found in all the patients is A431V, located at the NC/p1 CS. This is consistent with other reports (Fun et al., 2012).

For PSCM001, at the 1st time point the gag mutations that were detected included the gag CS mutations G381S, A431V, L449P and P453L. Mutations L449P and P453L are located at the p1/p6 CS and have also been described to have an impact on PI resistance (Fun et al., 2012). Some of these mutations are observed in the presence of specific PR mutations: mutations A431V, L449P and P453L occurred in the presence of mutations M46I, V82A and I84V (Table 4.2). This suggests that these mutations appear in the presence of PR mutations thus they serve as compensatory mutations to PI resistance (Nijhuis et al., 2007, Roquebert et al., 2006a). The gag CS mutation G381S has been reported to be associated with drug resistance but its role it unclear (Gatanaga et al., 2002). However, since we also found this mutation in the patients that were switched from the 1st line to the 2nd line therapy, this suggests that it is not selected for by PIs. After one month, a 2nd genotype was performed; the gag mutations that were detected at the 1st time point were still detected with no additional mutations in gag.

For PSCM002, at the 1st time point the gag mutations that were observed were G123E, V390I (McKinnon et al., 2011) and gag CS mutation G381S. Again, although these mutations have been rarely described they are associated with PI resistance (Fun et al., 2012, McKinnon et al., 2011). These mutations can appear in the absence of PR mutations selected under PI exposure. After two months, a 2nd genotype was performed. The mutations G123E, G318S and V390I were still detected and during this time point additional mutations A431V and P453L were detected. Again, from the literature, these latter mutations usually appeared in the presence of specific mutations: A431V appears in the presence of V82A and/or M46I, and this was also the case for this patient. P453L has been described to appear in the presence of I50V, I84V, D30N, N88D and L90M (Fun et al., 2012). However, none of these PR mutations were detected in this patient. Interestingly, mutations A431V and P453L were acquired when the L10F mutation in the PR was acquired.

For PCSM007, at the 1st time point the gag CS mutations included G381S and A431V. Again in this patient A431V appeared in the presence PR mutations V82A and M461 (Table 4.2). After two months, a 2nd genotype was performed and these gag mutations were still detected.

Table 4.4: Mutations occurring in the gag and PR of the patients that remained on 2nd line therapy.

Patient	PR mutations		Gag mutations	
	1st time point	2nd time point	1st time point	2nd time point
PCSM001	L10F, K20R, M46I, V82A, I84V	L10F, M46I, I54V, L76V, V82A, I84V	G381S, A431V, P453L, L449P	G381S, A431V, L449P, P453L
PCSM002	K20R, M46I, I54V, T74S, V82A,	L10F,K20R, M46I,I54V, T74S, V82A,	G123E, G381S V390I	G123E, G318S, V390L, A431V, P453L
PCSM007	L10F, M46I, I54V, V82A	L10F, M46I, I54V, V82A, A71V, L76V	G381S, A431V	G381S, A431V

4.2.3 Correlation of gag mutations with Protease resistance mutations

To further investigate the association of gag mutations with PR mutations, an analysis of the correlations between mutations in gag and PR was done. Unfortunately an association could not be made in these patients due to the small sample size, and because most of these patients didn't harbor PR resistance mutations at either time points. It is expected that with a large number of patients a correlation between gag and the PR resistance mutations will be seen.

4.4 Discussion

HIV resistance to PIs generally results in impaired PR function and also decreased replication capacity of the virus and this loss of replication capacity has been reported to be recovered by compensatory mutations that occur in the PR natural substrate i.e. the gag region (Dam et al., 2009). These mutations have also been reported to play a role in development of PI resistance (Clavel and Mammano, 2010, Dam et al., 2009). To investigate this, the mutations occurring in the gag region were also analyzed and the association of these mutations with PR resistance mutations was examined, particularly because some of the patients did not have any PI mutations.

In this study, we found that all the patients harbored gag mutations at both time points. These mutations have been described to be associated with PI resistance (Clavel and Mammano, 2010, Dam et al., 2009, Fun et al., 2012, Gatanaga et al., 2002). However, mutations V35I, G123E and G381S are rarely described in the literature and were seen in isolates before PI exposure. H219Q observed in PCSM009, has been described to be selected during PI exposure (Fun *et al* 2012). In addition, some studies have reported that this mutation is associated with PI resistance mutations I54V/L and I84V (Giandhari 2010) unpublished thesis.

For the patients that were switched from a 1st to 2nd line therapy, none of the patients harbored PI resistance mutations when switched on to 2nd line therapy, however they harbored gag mutations. It has been reported that the increased polyprotein processing as a results of mutations occurring in the gag rather than in the PR itself shows a novel mechanism by which HIV-1 could cause PIs resistance in the absence of PR resistance mutations (Ghosn et al., 2011, Nijhuis et al., 2007). In a study by Nijhuis et al., 2007 showed that some of the patients that did not harbor mutations in the PR region; harbored gag mutations (K48E and I437T/V). Similarly, in our study the gag

mutations V35I, G123E and G381S were present in the absence of PI mutations. However their role in subtype C is unknown.

It has been shown that PI resistance mutations hardly develop in patients on initial treatment with PI therapy (Gardner et al., 2009). This could be due to the high genetic barrier of Lopinavir as the development of PI resistance requires more than one drug resistance mutation to develop. This could explain the absence of PI mutations in our cohort. Gardner et al., 2009 has shown that the development of PI resistance mutations are more frequent in patients with suboptimal adherence and had failure to previous PIs (Gardner et al., 2009).

For the patients that remained on 2nd line therapy, most of the mutations found in gag (A431V, L449P and P453L) have been previously described to be associated with PI resistance mutations. This suggests that these mutations could be compensatory mutations to PI resistance mutations (Nijhuis et al., 2007, Roquebert et al., 2006a). Mutation A431V was common in all the patients and it has been reported to be the most frequent gag mutation in patient failing PI therapy (Fun et al., 2012). In addition, this mutation has been reported to confer PI resistance in the presence of specific PI resistance mutations M46I, L76V, V82A and I84V (Dam et al., 2009, Fun et al., 2012). These results are consistence with our observations in the PI experienced group investigated in this study. The other commonly described mutations were L449P and P453L. Again these mutations were reported to occur in the presence of other PI resistance mutations (D30N, N88D, I50V, I84V and I54V) (Fun et al., 2012). This suggests that these mutations have no effect on PI resistance when they occur alone, thus they act as compensatory mutations (Fun et al., 2012).

Chapter 5

Conclusion

Drug resistance mutations are usually detected by standard genotyping which only detects mutations >20% of the viral population but not the minority variants which occur at low levels of the viral population. The impact of these minority variants related to drug resistance still remains unresolved although there is evidence that these minority variants influence treatment efficacy. Thus more sensitive methods such as cloning and UDPS have been used to detect minority variants.

Additional drug resistance mutations were detected by cloning in all the patients studied. Where patients switched from 1st to 2nd line regimens, most of the minority variants were detected at the 2nd time point and these were related to their 1st line therapy, mostly NNRTI mutations, indicating that these mutations persist even in the absence of drug selection pressure. In addition, where patients remained on a failing PI regimen, PI minority variants were seen at both time points. Of note was the increase in L10F frequency in all the patients in this group.

Furthermore, we also compared the minority variants that were detected by cloning and UDPS. While UDPS is more sensitive than cloning from our findings, UDPS yielded similar results to cloning with very few additional minority variants detected. In the isolates where no PI resistance mutations were detected by Sanger, cloning or UDPS, this could be due to non-adherence, in which case this could indicate that minority PI resistance mutations do not persist in the absence of drug selection pressure. Alternatively, this could be due to the high genetic barrier of PIs as the development of PI resistance takes time to develop. In addition, many studies have reported that gag mutations can confer PI resistance in the absence of PI resistance mutations. In this study we also identified gag mutations in all the patients, however many were present before isolates not exposed to PIs. The A431V, P453L and L449P mutations were only seen in the patients that remained on 2nd line therapy PI, with A431V the most frequent.

Despite the limitations of sample size and availability for UDPS, this study illustrates the importance of detecting minority variants in patients failing treatment, to fully assess the drug resistance mutations spectrum and their impact on future therapy options. Future studies on

associations of gag and PR resistance mutations with increase sample size as well as detecting gag minority variants is essential.

REFERENCES

ABRAM, M. E., FERRIS, A. L., SHAO, W., ALVORD, W. G. & HUGHES, S. H. 2010. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol,* 84**,** 9864-78.

ALI, A., REDDY, G. S., NALAM, M. N., ANJUM, S. G., CAO, H., SCHIFFER, C. A. & RANA, T. M. 2010. Structure-based design, synthesis, and structure-activity relationship studies of HIV-1 protease inhibitors incorporating phenyloxazolidinones. *J Med Chem,* 53**,** 7699-708.

AMIEL, C., CHARPENTIER, C., DESIRE, N., BONNARD, P., LEBRETTE, M. G., WEISS, L., PIALOUX, G. & SCHNEIDER, V. 2011. Long-term follow-up of 11 protease inhibitor (PI)-naïve and PI-treated HIV-infected patients harbouring virus with insertions in the HIV-1 protease gene. *HIV medicine,* 12**,** 138-144.

ARTS, E. J. & HAZUDA, D. J. 2012. HIV-1 antiretroviral drug therapy. *Cold Spring Harbor perspectives in medicine,* 2**,** a007161.

AVIDOR, B., GIRSHENGORN, S., MATUS, N., TALIO, H., ACHSANOV, S., ZELDIS, I., FRATTY, I. S., KATCHMAN, E., BROSH-NISSIMOV, T., HASSIN, D., ALON, D., BENTWICH, Z., YUST, I., AMIT, S., FORER, R., VULIH SHULTSMAN, I. & TURNER, D. 2013. Evaluation of a benchtop HIV ultradeep pyrosequencing drug resistance assay in the clinical laboratory. *J Clin Microbiol,* 51**,** 880-6.

AVERTING HIV AND AIDS, HIV & AIDS in South Africa, 2013. [online] http://www.avert.org/aidsSouthAfrica [Accesed 5 August 2013].

BALDAUF, S. L. 2003. Phylogeny for the faint of heart: a tutorial. *Trends Genet,* 19**,** 345-51.

BANKE, S., LILLEMARK, M. R., GERSTOFT, J., OBEL, N. & JORGENSEN, L. B. 2009. Positive selection pressure introduces secondary mutations at Gag cleavage sites in human immunodeficiency virus type 1 harboring major protease resistance mutations. *J Virol,* 83**,** 8916-24.

BANSODE, V., MCCORMACK, G., CRAMPIN, A., NGWIRA, B., SHRESTHA, R., FRENCH, N., GLYNN, J. & TRAVERS, S. 2013. Characterizing the emergence and persistence of drug resistant mutations in HIV-1 subtype C infections using 454 ultra deep pyrosequencing. *BMC Infectious Diseases,* 13**,** 52.

BOS, D. H. & POSADA, D. 2005. Using models of nucleotide evolution to build phylogenetic trees. *Dev Comp Immunol,* 29**,** 211-27.

BROWN, A. J. & CLELAND, A. 1996. Independent evolution of the env and pol genes of HIV-1 during zidovudine therapy. *AIDS,* 10**,** 1067-73.

CHARPENTIER, C., DWYER, D. E., MAMMANO, F., LECOSSIER, D., CLAVEL, F. & HANCE, A. J. 2004. Role of minority populations of human immunodeficiency virus type 1 in the evolution of viral resistance to protease inhibitors. *J Virol,* 78**,** 4234-47.

CHEN, L., PERLINA, A. & LEE, C. J. 2004. Positive selection detection in 40,000 human immunodeficiency virus (HIV) type 1 sequences automatically identifies drug resistance and positive fitness mutations in HIV protease and reverse transcriptase. *J Virol,* 78**,** 3722-32.

CLAVEL, F. & HANCE, A. J. 2004. HIV drug resistance. *N Engl J Med,* 350**,** 1023-35.

CLAVEL, F. & MAMMANO, F. 2010. Role of Gag in HIV Resistance to Protease Inhibitors. *Viruses,* 2**,** 1411-1426.

COHEN, M. S., HELLMANN, N., LEVY, J. A., DECOCK, K. & LANGE, J. 2008. The spread, treatment, and prevention of HIV-1: evolution of a global pandemic. *The Journal of Clinical Investigation,* 118**,** 1244-1254.

CRAIGIE, R. & BUSHMAN, F. D. 2012. HIV DNA Integration. *Cold Spring Harbor perspectives in medicine,* 2.

DAM, E., QUERCIA, R., GLASS, B., DESCAMPS, D., LAUNAY, O., DUVAL, X., KRÄUSSLICH, H.-G., HANCE, A. J., CLAVEL, F. & GROUP, A. S. 2009. Gag Mutations Strongly Contribute to HIV-1 Resistance to Protease Inhibitors in Highly Drug-Experienced Patients besides Compensating for Fitness Loss. *PLoS Pathog,* 5**,** e1000345.

DE OLIVEIRA, T., ENGELBRECHT, S., JANSE VAN RENSBURG, E., GORDON, M., BISHOP, K., ZUR MEGEDE, J., BARNETT, S. W. & CASSOL, S. 2003. Variability at Human Immunodeficiency Virus Type 1 Subtype C Protease Cleavage Sites: an Indication of Viral Fitness? *Journal of Virology,* 77**,** 9422-9430.

DELPORT, W., POON, A. F., FROST, S. D. & KOSAKOVSKY POND, S. L. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics,* 26**,** 2455-7.

DEPARTMENT OF HEALTH 2013, The South Africa Department of Health Antiretroviral therapy guidelines 2013, acceseed 20 November 2013, www.auruminstitute.org/phocadownload/guidelines-short.pdf

ESHLEMAN, S. H., JONES, D., FLYS, T., PETRAUSKENE, O. & JACKSON, J. B. 2003. Analysis of HIV-1 variants by cloning DNA generated with the ViroSeq HIV-1 Genotyping System. *Biotechniques,* 35**,** 614-8, 620, 622.

FISCHER, W., APETREI, C., SANTIAGO, M. L., LI, Y., GAUTAM, R., PANDREA, I., SHAW, G. M., HAHN, B. H., LETVIN, N. L., NABEL, G. J. & KORBER, B. T. 2012. Distinct evolutionary pressures underlie diversity in simian immunodeficiency virus and human immunodeficiency virus lineages. *J Virol,* 86**,** 13217-31.

FREED, E. O. 2001. HIV-1 replication. *Somat Cell Mol Genet,* 26**,** 13-33.

FUN, A., WENSING, A. M., VERHEYEN, J. & NIJHUIS, M. 2012. Human Immunodeficiency Virus Gag and protease: partners in resistance. *Retrovirology,* 9**,** 63.

GARDNER, E. M., BURMAN, W. J., STEINER, J. F., ANDERSON, P. L. & BANGSBERG, D. R. 2009. Antiretroviral medication adherence and the development of class-specific antiretroviral resistance. *AIDS,* 23**,** 1035-46.

GATANAGA, H., SUZUKI, Y., TSANG, H., YOSHIMURA, K., KAVLICK, M. F., NAGASHIMA, K., GORELICK, R. J., MARDY, S., TANG, C., SUMMERS, M. F. & MITSUYA, H. 2002. Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. *J Biol Chem,* 277**,** 5952-61.

GATTO, L., CATANZARO, D. & MILINKOVITCH, M. C. 2006. Assessing the applicability of the GTR nucleotide substitution model through simulations. *Evol Bioinform Online,* 2**,** 145-55.

GHOSN, J., DELAUGERRE, C., FLANDRE, P., GALIMAND, J., COHEN-CODAR, I., RAFFI, F., DELFRAISSY, J. F., ROUZIOUX, C. & CHAIX, M. L. 2011. Polymorphism in Gag gene cleavage sites of HIV-1 non-B subtype and virological outcome of a first-line lopinavir/ritonavir single drug regimen. *PLoS One,* 6**,** e24798.

GIANELLA, S. & RICHMAN, D. D. 2010. Minority Variants of Drug-Resistant HIV. *Journal of Infectious Diseases,* 202**,** 657-666.

GIANDHARI, J 2010. The Role Of The Protease Cleavage Site In Viral Fitness And Drug Resistance In Hiv-1 Subtype C. Unplubished

HALVAS, E. K., WIEGAND, A., BOLTZ, V. F., KEARNEY, M., NISSLEY, D., WANTMAN, M., HAMMER, S. M., PALMER, S., VAIDA, F., COFFIN, J. M. & MELLORS, J. W. 2010. Low Frequency Nonnucleoside Reverse-Transcriptase Inhibitor—Resistant Variants Contribute to Failure of Efavirenz-Containing Regimens in Treatment-Experienced Patients. *Journal of Infectious Diseases,* 201**,** 672-680.

HIRSCH, M. S., GUNTHARD, H. F., SCHAPIRO, J. M., BRUN-VEZINET, F., CLOTET, B., HAMMER, S. M., JOHNSON, V. A., KURITZKES, D. R., MELLORS, J. W., PILLAY, D., YENI, P. G., JACOBSEN, D. M. & RICHMAN, D. D. 2008. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. *Clin Infect Dis,* 47**,** 266-85.

HOFFMAN, C & ROCKSTROH, J.K. HIV 2012/13. www.hivbook.com. [online] http://hivbook.files.wordpress.com/2011/10/hivbook-2012.pdf [Accessed 16 April 2014]

HOLMES, E. C., ZHANG, L. Q., SIMMONDS, P., LUDLAM, C. A. & BROWN, A. J. 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc Natl Acad Sci U S A,* 89**,** 4835-9.

HUANG, A., HOGAN, J. W., ISTRAIL, S., DELONG, A., KATZENSTEIN, D. A. & KANTOR, R. 2012. Global analysis of sequence diversity within HIV-1 subtypes across geographic regions. *Future Virol,* 7**,** 505-517.

HUÉ, S., GIFFORD, R. J., DUNN, D., FERNHILL, E., PILLAY, D. & RESISTANCE, O. B. O. T. U. C. G. O. H. D. 2009. Demonstration of Sustained Drug-Resistant Human Immunodeficiency Virus Type 1 Lineages Circulating among Treatment-Naïve Individuals. *Journal of Virology,* 83**,** 2645-2654.

IMAZ, A., FALCO, V. & RIBERA, E. 2011. Antiretroviral salvage therapy for multiclass drug-resistant HIV-1-infected patients: from clinical trials to daily clinical practice. *AIDS Rev,* 13**,** 180-93.

JOHNSON, J. A. & GERETTI, A. M. 2010. Low-frequency HIV-1 drug resistance mutations can be clinically significant but must be interpreted with caution. *J Antimicrob Chemother,* 65**,** 1322-6.

JOHNSON, J. A., LI, J.-F., WEI, X., LIPSCOMB, J., IRLBECK, D., CRAIG, C., SMITH, A., BENNETT, D. E., MONSOUR, M., SANDSTROM, P., LANIER, E. R. & HENEINE, W. 2008. Minority HIV-1 Drug Resistance Mutations Are Present in Antiretroviral Treatment–Naïve Populations and Associate with Reduced Treatment Efficacy. *PLoS Med,* 5**,** e158.

JÜLG, B. & GOEBEL, F. D. 2005. HIV Genetic Diversity: Any Implications for Drug Resistance? *Infection,* 33**,** 299-301.

KIJAK, G. H., SIMON, V., BALFE, P., VANDERHOEVEN, J., PAMPURO, S. E., ZALA, C., OCHOA, C., CAHN, P., MARKOWITZ, M. & SALOMON, H. 2002. Origin of human immunodeficiency virus type 1 quasispecies emerging after antiretroviral treatment interruption in patients with therapeutic failure. *J Virol,* 76**,** 7000-9.

KOSAKOVSKY POND, S. L. & FROST, S. D. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol,* 22**,** 1208-22.

KOZISEK, M., HENKE, S., SASKOVA, K. G., JACOBS, G. B., SCHUCH, A., BUCHHOLZ, B., MULLER, V., KRAUSSLICH, H. G., REZACOVA, P., KONVALINKA, J. & BODEM, J. 2012. Mutations in HIV-1 gag and pol compensate for the loss of viral fitness caused by a highly mutated protease. *Antimicrob Agents Chemother,* 56**,** 4320-30.

LARROUY, L., CHAZALLON, C., LANDMAN, R., CAPITANT, C., PEYTAVIN, G., COLLIN, G., CHARPENTIER, C., STORTO, A., PIALOUX, G., KATLAMA, C., GIRARD, P. M., YENI, P., ABOULKER, J. P., BRUN-VEZINET, F., DESCAMPS, D. & GROUP, O. B. O. T. A. S. 2010. Gag Mutations Can Impact Virological Response to Dual-Boosted Protease Inhibitor Combinations in Antiretroviral-Naïve HIV-Infected Patients. *Antimicrobial Agents and Chemotherapy,* 54**,** 2910-2919.

LANIER, R., MOFFATT, A., STONE, C., IRLBECK, D., ROUSE, E., HORTON, J., ROSS, L., DU, L., TAILLON, B., EGHOLM, M. & CRAIG, C. 2007. Comparison of Traditional Plasmid-Based Clonal Sequencing to 454 Ultra-Deep Sequencing For Hiv Clinical Isolates With Reverse Transcriptase (RT) Mutations at K65R And L74V.[Online] www.aegis.org/displayconf/abstract. [Accessed 9 June 2012]

LE, T., CHIARELLA, J., SIMEN, B., HANCZARUK, B., EGHOLM, M., LANDRY, M., DIECKHAUS, K., ROSEN, M. & KOZAL, M. 2009. Low-abundance HIV drug-resistant

viral variants in treatment-experienced persons correlate with historical antiretroviral use. *PLoS One,* 4**,** e6079.

LIANG, B., LUO, M., SCOTT-HERRIDGE, J., SEMENIUK, C., MENDOZA, M., CAPINA, R., SHEARDOWN, B., JI, H., KIMANI, J., BALL, B. T., VAN DOMSELAAR, G., GRAHAM, M., TYLER, S., JONES, S. J. M. & PLUMMER, F. A. 2011. A Comparison of Parallel Pyrosequencing and Sanger Clone-Based Sequencing and Its Impact on the Characterization of the Genetic Diversity of HIV-1. *PLoS ONE,* 6**,** e26745.

LIHANA, R. W., SSEMWANGA, D., ABIMIKU, A. & NDEMBI, N. 2012. Update on HIV-1 diversity in Africa: a decade in review. *AIDS Rev,* 14**,** 83-100.

LIO, P. & GOLDMAN, N. 1998. Models of molecular evolution and phylogeny. *Genome Res,* 8**,** 1233-44.

LOS ALAMOS NATIONAL LABORATORY. *HIV Sequence Database* [Online] http://www.hiv.lanl.gov/content/sequence/VESPA/vespa.html [Accessed 23 March 2013].

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y. J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L., JARVIE, T. P., JIRAGE, K. B., KIM, J. B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W., NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J., SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. & ROTHBERG, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-80.

MARTINEZ-CAJAS, J. L. & WAINBERG, M. A. 2008. Antiretroviral therapy : optimal sequencing of therapy to avoid resistance. *Drugs,* 68**,** 43-72.

MCKINNON, J. E., DELGADO, R., PULIDO, F., SHAO, W., ARRIBAS, J. R. & MELLORS, J. W. 2011. Single genome sequencing of HIV-1 gag and protease resistance mutations at virologic failure during the OK04 trial of simplified versus standard maintenance therapy. *Antivir Ther,* 16**,** 725-32.

METZNER, K. J., RAUCH, P., WALTER, H., BOESECKE, C., ZOLLNER, B., JESSEN, H., SCHEWE, K., FENSKE, S., GELLERMANN, H. & STELLBRINK, H. J. 2005. Detection of minor populations of drug-resistant HIV-1 in acute seroconverters. *AIDS,* 19**,** 1819-25.

NIJHUIS, M., VAN MAARSEVEEN, N. M., LASTERE, S., SCHIPPER, P., COAKLEY, E., GLASS, B., ROVENSKA, M., DE JONG, D., CHAPPEY, C., GOEDEGEBUURE, I. W., HEILEK-SNYDER, G., DULUDE, D., CAMMACK, N., BRAKIER-GINGRAS, L., KONVALINKA, J., PARKIN, N., KRÄUSSLICH, H.-G., BRUN-VEZINET, F. & BOUCHER, C. A. B. 2007. A Novel Substrate-Based HIV-1 Protease Inhibitor Drug Resistance Mechanism. *PLoS Med,* 4**,** e36.

NYOMBI, B. M., HOLM-HANSEN, C., KRISTIANSEN, K. I., BJUNE, G. & MULLER, F. 2008. Prevalence of reverse transcriptase and protease mutations associated with antiretroviral drug resistance among drug-naive HIV-1 infected pregnant women in Kagera and Kilimanjaro regions, Tanzania. *AIDS Res Ther,* 5**,** 13.

PALMER, S., KEARNEY, M., MALDARELLI, F., HALVAS, E., BIXBY, C., BAZMI, H., ROCK, D., FALLOON, J., DAVEY, R. & DEWAR, R. 2005. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J Clin Microbiol,* 43**,** 406 - 413.

PARASCHIV, S., OTELEA, D., DINU, M., MAXIM, D. & TINISCHI, M. 2007. Polymorphisms and resistance mutations in the protease and reverse transcriptase genes of HIV-1 F subtype Romanian strains. *Int J Infect Dis,* 11**,** 123-8.

PAREDES I DEIROS, R. 2009. Clinical Implications of Minority HIV-1 Resistant Variants.

POMERANTZ, R. J. & HORN, D. L. 2003. Twenty years of therapy for HIV-1 infection. *Nat Med,* 9**,** 867-73.

POSADA, D. & CRANDALL, K. A. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics,* 14**,** 817-8.

RAMIREZ, C., GREGORI, J., BUTI, M., TABERNERO, D., CAMOS, S., CASILLAS, R., QUER, J., ESTEBAN, R., HOMS, M. & RODRIGUEZ-FRIAS, F. 2013. A comparative study of ultra-deep pyrosequencing and cloning to quantitatively analyze the viral quasispecies using hepatitis B virus infection as a model. *Antiviral Res,* 98**,** 273-83.

RECORDON-PINSON, P., PAPUCHON, J., REIGADAS, S., DESHPANDE, A. & FLEURY, H. 2012. K65R in Subtype C HIV-1 Isolates from Patients Failing on a First-Line Regimen Including d4T or AZT: Comparison of Sanger and UDP Sequencing Data. *PLoS ONE,* 7**,** e36549.

REGADB SEQUENCE ANALYSIS TOOLS [Online] (http://jose.med.kuleuven.be/sequencetools) [Accessed 29 July 2013].

RICHMAN, D. D. 2001. HIV chemotherapy. *Nature,* 410**,** 995-1001.

RIZZO, J. & ROUCHKA, E. C. 2007. Review of Phylogenetic Tree Construction. *University of Louisville Bioinformatics Laboratory Technical Report Series*.

ROQUEBERT, B., MALET, I., WIRDEN, M., TUBIANA, R., VALANTIN, M., SIMON, A., KATLAMA, C., PEYTAVIN, G., CALVEZ, V. & MARCELIN, A. 2006a. Role of HIV-1 minority populations on resistance mutational pattern evolution and susceptibility to protease inhibitors. *AIDS (London, England),* 20**,** 287 - 289.

ROQUEBERT, B., MALET, I., WIRDEN, M., TUBIANA, R., VALANTIN, M. A., SIMON, A., KATLAMA, C., PEYTAVIN, G., CALVEZ, V. & MARCELIN, A. G. 2006b. Role of HIV-1 minority populations on resistance mutational pattern evolution and susceptibility to protease inhibitors. *AIDS,* 20**,** 287-9.

SALEMI, M., VANDAMME, A.-M. & LEMEY, P. 2009. *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*, Cambridge University Press.

SAN MAURO, D. & AGORRETA, A. 2010. Molecular systematics: A synthesis of the common methods and the state of knowledge. *Cell Mol Biol Lett,* 15**,** 311-41.

SCHOFFELEN, A. F., WENSING, A. M. J., TEMPELMAN, H. A., GEELEN, S. P. M., HOEPELMAN, A. I. M. & BARTH, R. E. 2013. Sustained Virological Response on Second-Line Antiretroviral Therapy following Virological Failure in HIV-Infected Patients in Rural South Africa. *PLoS ONE,* 8**,** e58526.

SHAFER, R. W. 2002. Genotypic testing for human immunodeficiency virus type 1 drug resistance. *Clin Microbiol Rev,* 15**,** 247-77.

SHAFER, ROBERT W. 2006. Rationale and Uses of a Public HIV Drug-Resistance Database. *Journal of Infectious Diseases,* 194**,** S51-S58.

SHAFER, R. W. 2009. Low-abundance drug-resistant HIV-1 variants: finding significance in an era of abundant diagnostic and therapeutic options. *J Infect Dis,* 199**,** 610-2.

SIERRA, S., KUPFER, B. & KAISER, R. 2005. Basics of the virology of HIV-1 and its replication. *Journal of Clinical Virology,* 34**,** 233-244.

SIMEN, B., SIMONS, J., HULLSIEK, K., NOVAK, R., MACARTHUR, R., BAXTER, J., HUANG, C., LUBESKI, C., TURENCHALK, G. & BRAVERMAN, M. 2009. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. *J Infect Dis,* 199**,** 693 - 701.

SIMMONDS, P., BALFE, P., LUDLAM, C. A., BISHOP, J. O. & BROWN, A. J. 1990. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J Virol,* 64**,** 5840-50.

SINGH, K., MARCHAND, B., KIRBY, K. A., MICHAILIDIS, E. & SARAFIANOS, S. G. 2010. Structural Aspects of Drug Resistance and Inhibition of HIV-1 Reverse Transcriptase. *Viruses,* 2**,** 606-638.

SOARES, E. A. J. M., SANTOS, A. F. A., SOUSA, T. M., SPRINZ, E., MARTINEZ, A. M. B., SILVEIRA, J., TANURI, A. & SOARES, M. A. 2007. Differential Drug Resistance Acquisition in HIV-1 of Subtypes B and C. *PLoS ONE,* 2**,** e730.

SOARES, M.A & VEGA, P, C. 2008. Drug resistance diference among HIV types and subtypes: Classification & global distribution of HIV. [Online] http://www.medscape.org/viewarticle/583361_2 [Accessed 2 March 2014].

TAMURA, K., DUDLEY, J., NEI, M. & KUMAR, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol,* 24**,** 1596-9.

TANG, M. W. & SHAFER, R. W. 2012. HIV-1 antiretroviral resistance: scientific principles and clinical applications. *Drugs,* 72**,** e1-25.

UNAIDS 2013, SPECIAL REPORT UPDATE. [online] http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2013/20130521_Update_Africa.pdf [Accessed 7 April 2014].

WAINBERG, M. A. 2004. HIV-1 subtype distribution and the problem of drug resistance. *AIDS,* 18**,** S63-S68.

WRIGHT, J. K., BRUMME, Z. L., CARLSON, J. M., HECKERMAN, D., KADIE, C. M., BRUMME, C. J., WANG, B., LOSINA, E., MIURA, T., CHONCO, F., VAN DER STOK, M., MNCUBE, Z., BISHOP, K., GOULDER, P. J., WALKER, B. D., BROCKMAN, M. A. & NDUNG'U, T. 2010. Gag-protease-mediated replication capacity in HIV-1 subtype C chronic infection: associations with HLA type and clinical parameters. *J Virol,* 84**,** 10820-31.

ZHANG, L. Q., SIMMONDS, P., LUDLAM, C. A. & BROWN, A. J. 1991. Detection, quantification and sequencing of HIV-1 from the plasma of seropositive individuals and from factor VIII concentrates. *AIDS,* 5**,** 675-81.