Drug resistance, evolution and diversity of HIV-1 in Malawi

Thesis submitted for the degree of Ph.D. to the National University of Ireland, Galway by:

Vijay Balasaheb Bansode

Supervisors:
Dr. Grace McCormack, B.Sc., Ph.D. (NUI)
Prof. Simon Travers, B.Sc., Ph.D. (NUI)
School of Natural Sciences,
National University of Ireland,
Galway

2012

Table of Contents

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Grace for your guidance, direction, patience, and understanding. Apart from your encouragement and knowledge, you offered me support in many ways, which helped me to settle well in completely different world than mine. You always had time for me and you were always happy to correct my English! I could not have asked a better supervisor than you. I warmly thank Simon, my co-supervisor; your valuable training of bioinformatics helped me understand various aspects of bioinformatics.

Thanks to the Scientists and collaborators at KPS, Malawi for providing me necessary information from time to time. Thanks also to friends I made at the KPS. I really had one of the best times in my life in Malawi. I would also like to thank to the Indian friends in Malawi, Tushar and Nilesh Adhikari, Mangesh Deshmukh and Sampath.

I would like to express my thanks to my lab mates Ishla, Kelly, Carsten, Conor, Niamh (R), Salla who played an important role in my life during my stay in Ireland. Thanks especially to Ishla for your friendship and for being my awesome lab/office mates Thanks to my lab GF Kelly for all the (non-science) discussions and Carsten for providing me photography knowledge. Thanks to Conor for providing me scripts for my analysis. Special thanks to Zuza, for her friendship and help in the lab, whose work contributed to my project. Also thank you tag rugby teams I played with over the four years, it gave me confidence that a ‘skinny’ guy also can score!.

Thanks to all other Zoology, Marine Science members, especially Peter, Niamh (Q), Cat, Emma, Conall and numerous friends I made in Ireland for Friday, Saturday and mid-week drinking nights. My housemates from 74- Rosan Glas; Conall, Florence, Marcus and lovely Beco who made my stay in Ireland cheerful. I never thought of changing the house because of you guys. Special thanks to Suvarn who saved my life by bringing those precious Eppendorf tubes at 3 AM at Dublin airport without hesitating for a second.

Thanks to technicians John, Albert and Eoin, their constant helping hand made my lab life very easy. Also I would like to express sincere thanks to Anne

Quinn (Zoology) and Louise Coyle (International Affairs Office) who helped out for the necessary paperwork in time for my visas.

I would also like to thank Matteo for offering me a position in your lab and helping me for settling me in Strasbourg quickly. I am lucky for being a part of your dynamic and friendly team where I have new great labmates Sarah, Paola, Daniela, Flore, Marika, Claudia, Romain and Piere. Special thanks to Sarah for her friendship and helping me in the search for house without complaining and finally finding the best house for me! I wish you best of luck for your thesis presentation.

Thanks to my brothers and family members, Anant (Bandu), Varsha vahini and little Arya, Sanju and Pooja for encouragement and support. Thank you mostly to my parents, Aai and Bappa for supporting me throughout the life. All the hard work you did only to see me happy, I will never be able to come out of that debt, thank you.

Last but not the least, my best thanks are reserved for my love Megha, whose love, care, friendship affection and confidence in me made this happen. You never complained, doubted my decisions and happily stayed away from me to support me. It is always hard to stay away from you; I don’t know how I managed it.

Abstract

HIV/AIDS is the world's biggest infectious cause of death and Sub-Saharan Africa being most affected. Although ART has significantly reduced AIDS related deaths in recent years, the development of drug resistance remains a major obstacle in the success of HIV treatment especially in countries from Sub-Saharan Africa such as Malawi, where ART started relatively recently. As there is scale up of ART in such countries in recent years, the limited availability of drug choice and drug resistance tests may expand the problem of drug resistance in the near future. Thus monitoring the genotypes HIV-1 infected individuals is necessary in order to monitor the presence, emergence and transmission of drug resistance. Genotyping is common in the western world and a well-established drug resistance profile for subtype B is available, however, limited data were available for subtype C, which predominates the epidemic in the Sub-Saharan Africa. Various genotypic and phenotypic methods are used to monitor drug resistance, however the most commonly used methods such as bulk sequencing are unable to detect minor drug resistant variants.

This work aimed to monitor the emergence and transmission of drug resistance in HIV-1 in Karonga District, Malawi and to study the diversity of HIV-1 in the Karonga population by using the sequences generated in this study. The initial results of bulk sequencing of proviral DNA showed that polymorphisms associated with drug resistance were present in 15 % of the drug naïve HIV-1 infected individuals. Further clonal sequencing revealed more DRMs in some individuals where initial bulk sequencing failed to detect DRMs. Furthermore, bulk sequencing of RT from HIV-1 infected individuals on ART revealed that some individuals showed DRMs at baseline but no DRMs at subsequent timepoints pointing to the limitations of bulk sequencing.

Subsequently, the work was then focused use of UDPS technology to resolve the discrepancy in DRMs found by bulk sequencing in the RT gene. Although a good sequence depth was yielded which helped in identifying minor DRMs that were undetected by bulk sequencing, further discrepancies in the presence and prevalence of DRMs were found. Furthermore, low abundance resistance mutations against current first line therapy were identified. In order to maximize diversity of

templates subjected to UDPS, multiple primary and secondary (nested) PCRs were carried out. It was found that it is the quality of the initial amplification of templates that is of major importance for characterizing genetic diversity rather than the number of PCR steps.

Phylogenetic analysis of RT and Protease sequences showed that subtype C is the predominant subtype (97% of the sequences) in the Karonga District extending the previous findings subtypes A1 and G were also found at lower percentages. Studies on selection pressure analysis revealed that the RT gene in Karonga dataset is not driven by the antiretroviral therapy (yet at least).

Declaration

This thesis has not been submitted in whole or in part to this or to any other University for any other degree and is, except where otherwise stated, the original work of the author.

Signed……

Vijay Bansode

Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral Therapy
CCR5	Chemokine receptor 5
CD4	Cluster of Differentiation 4
CRF	Circulating Recombinant Form
CXCR4	Chemokine receptor 4
DRM/s	Drug Resistance Mutation/s
gp120	Glycoprotein 120
gp160	Glycoprotein 160
gp41	Glycoprotein 41
GTR	General Time Reversible model of DNA substitution
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immunodeficiency Virus
HIV-1	Human Immunodeficiency Virus type 1
HIV-2	Human Immunodeficiency Virus type 2
IN	Integrase
MA	Matrix
NGS	Next Generation Sequencing
NRTI	Nucleoside/tide Reverse Transcriptase Inhibitor
NNRTI	Non-Nucleoside Reverse Transcriptase Inhibitor
PCR	Polymerase Chain Reaction
PI	Protease Inhibitor
PIC	Pre-integration Complex
PR	Protease
RT	Reverse Transcriptase
SIV	Simian Immunodeficiency Virus
TDR	Transmitted Drug Resistance
UDPS	Ultra-Deep Pyro Sequencing

Index of Figures

Index of Tables

Chapter 1

General Introduction

1.1 Human Immunodeficiency Virus (HIV)

Human Immunodeficiency Virus (HIV) is the causative agent of Acquired Immune Deficiency Syndrome (AIDS), one of the deadliest diseases of modern era. Almost 38 years after its discovery (Barre-Sinoussi et al. 1983), the search for a cure is still going on. HIV is a problem of pandemic proportions. More than 25 million people have died from HIV infection. At the end of 2010, it was estimated that 34 million people had been infected with HIV, with 1.8 million new infections in 2010 (Figure 1.1, (UNAIDS 2011b). The highest prevalence of HIV is within Sub-Saharan Africa, especially in the southern region, compared to the other regions in the world (Figure 1.2), while the prevalence appears low in Northern Africa and the Middle East, although such figures may represent underreporting or lack of data (Figure 1.2). The prevalence within countries may vary, in particular between rural and urban areas. For example the prevalence in Urban Malawi, Lilongwe was 25 % in 2001 (NACO 2001), while the prevalence in the Karonga District in Northern Malawi, where this project is based was between 10 and 15% (Glynn et al. 2001). AIDS related deaths have dropped in recent years due to the introduction of more potent drug cocktails and greater access to antiretroviral therapies (Figure 1.3). Furthermore, new HIV infections have dropped due to key behavioural changes. For example, in Cambodia, from 1994 to 2000, new HIV infections in adults dropped from an estimated >50,000 to approximately 9,000, due to condom use by sex workers (Figure 1.4, (UNAIDS 2011b). However, despite a drop in AIDS related deaths and new infections in recent years, AIDS related illnesses are still a major cause of death worldwide.

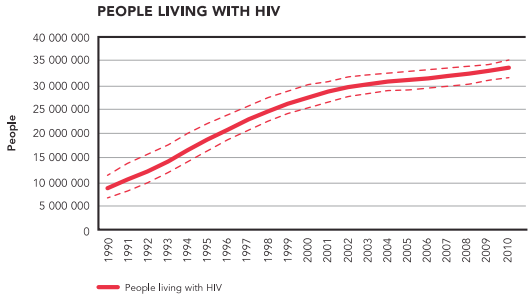


Figure 1.1: The state of the AIDS epidemic: number of people living with HIV (UNAIDS 2011b).

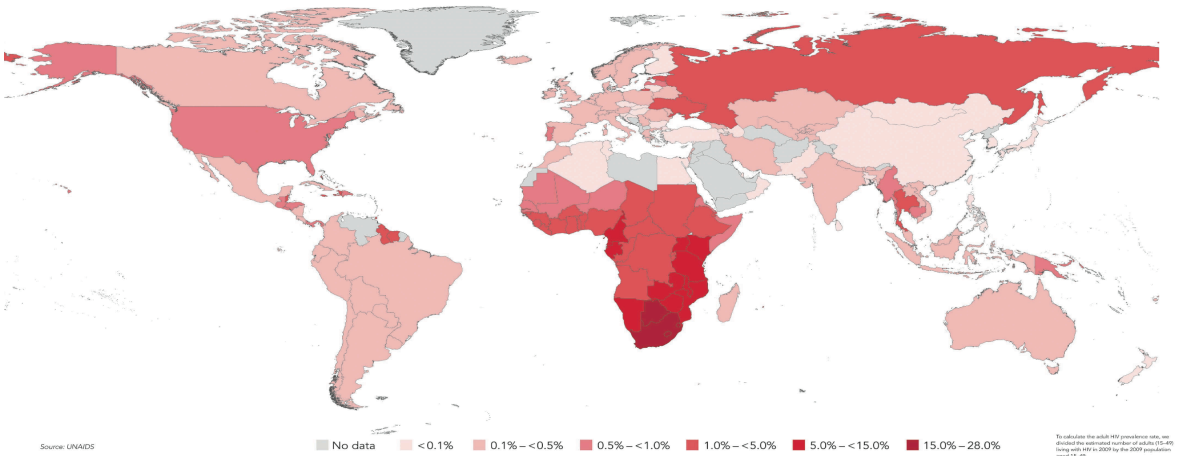


Figure 1.2: A global view of HIV infections: The prevalence of HIV infections in different regions of the world (UNAIDS 2011b).

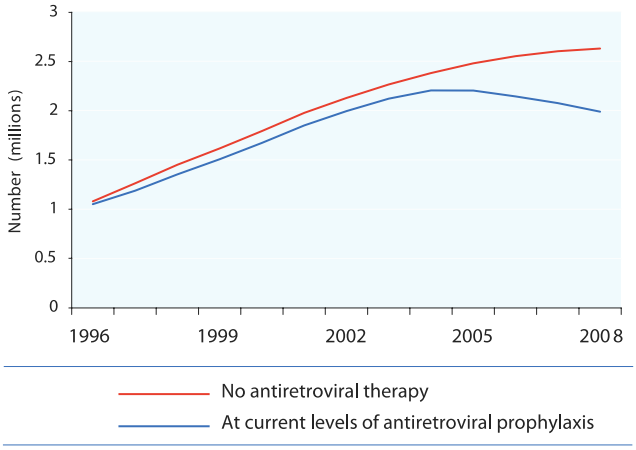


Figure 1.3: Global estimate of the number of AIDS-related deaths from 1996-2008 with and without antiretroviral therapy (UNAIDS 2009a).

HIV infections have been recorded as early as 1959 - 1960 (Zhu et al. 1998; Worobey et al. 2008), however, the first case of AIDS was reported in 1981, when a few opportunistic infections such as Kaposi's Sarcoma and *Pneumocystis* Pneumonia were discovered in homosexual men (CDC 1981a; CDC 1981b). In each patient, it was found that the humoral immunity was intact, while, cellular immunity was impaired with depression in absolute lymphocytes and T cell counts, allowing the infections to occur (Gottlieb et al. 1981; Masur et al. 1981). Later, in 1983, it was discovered that this depression in cellular immunity, was caused by a "micro organism" that was transmitted by sexual contact and blood products, this condition was named as AIDS (Francis, Curran, and Essex 1983). In the same year, the

detection of reverse transcriptase activity was reported in a cell culture from lymph nodes taken from patient with pre-AIDS symptoms. This, along with the morphology of the viruses from cell culture and visualised by electron microscopy indicated that the causative agent might be from the retrovirus group (Barre-Sinoussi et al. 1983; Levy et al. 1984).

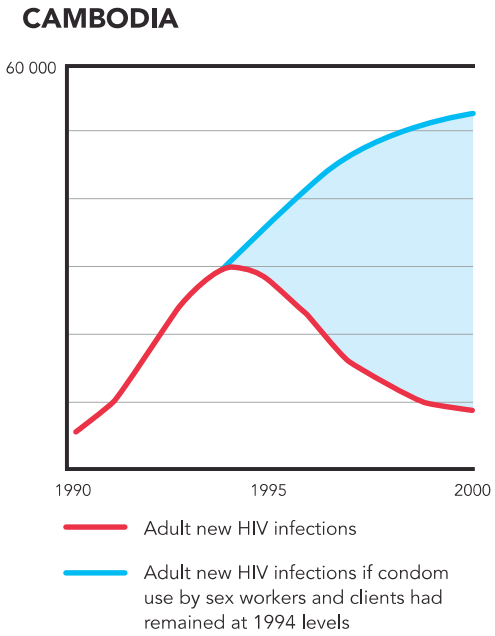


Figure 1.4: An example of Cambodia depicting the comparison of estimated new HIV infections with actual adult infections (UNAIDS 2011b).

In parallel, a retrovirus was isolated, propagated and characterized from various AIDS patients (Popovic et al. 1984). Later studies proved that the virus isolated by both these studies was same, and this virus, named as HIV-1, was unambiguously the causative agent of AIDS. A growing number of researchers focused on HIV research and in 1985, the first HIV genome was cloned and sequenced (Wain-Hobson et al. 1985).

1.2: HIV structure and replication cycle

HIV is a member of the genus *Lentivirus* from the family Retroviridae (Coffin, Hughes, and Varmus 1997).The HIV virion is spherically shaped with a diameter of between 100-250 nm (Figure 1.5). The outer coat of the virus – the viral envelope - consists of two layers of lipids, which are derived from the host cell membrane. The envelope also contains different proteins forming “spikes”. There are approximately 72 spikes; each spike is composed of surface units (gp120, SU) that are linked with trans-membrane units (gp41, TM). The inner surface of the lipid bi-layer is lined with matrix proteins (MA). The matrix protein p17, is attached to the internal face of the envelope. Inside the viral envelope, a cone shaped nucleocapsid is present. The capsid surrounds two copies of RNA genome with the reverse transcriptase (RT) molecules (Figure 5) (Briggs et al. 2003). The nucleocapsid protein, p9, is non-covalently attached to the viral RNA. Other viral enzymes, protease (PR) and integrase (IN) and some accessory and regulatory proteins (p6, *Nef, Vif* and *Vpr*) are also found in the nucleocapsid (Zanetti et al. 2006; Zhu et al. 2006).

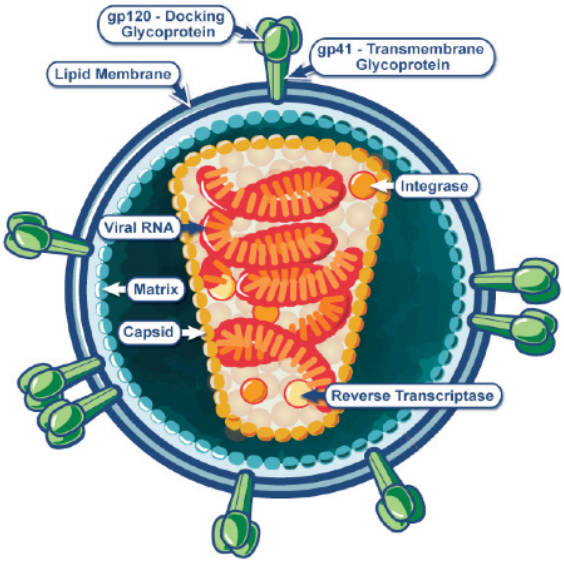


Figure 1.5: An illustrated diagram of single, infective particle of HIV virus.
Source: http://www.niaid.nih.gov/topics/HIVAIDS/Understanding/Biology/Pages/hivVirionLargeImage.aspx. Jan 2009

The HIV replication cycle can be subdivided broadly into seven steps (Figure 1.6) from entry into the host cell to the release of the mature virion outside the cell. The first step is binding and fusion, which starts with interaction between the viral envelope protein (gp120), and the CD4 receptor located on the membrane of T

lymphocytes and macrophages. This process is followed by additional interactions with the CXCR4 or CCR5 chemokine co-receptors, located on the cell surface (Reviewed in (Doms and Trono 2000) (Step 1, Figure 1.6). These interactions lead to a conformational change in the trans-membrane units leading to the fusion of the viral lipid bi-layer and the host plasma membrane. At this step, the viral core penetrates the cell (Cladera, Martin, and O'Shea 2001)(Step 2, Figure 1.6). Following the fusion, the viral nucleoprotein complex, which consists of MA, RT, IN, *Vpr* and RNA, is exposed. (Step 3, Figure 1.6). The viral RT then reverse transcribes the viral RNA into complementary DNA (cDNA) with the help of a cellular lysine tRNA molecule as a primer (Cen et al. 2001) (Step 3, Figure 1.6). The RNAse activity of RT degrades the viral RNA template. The newly synthesized cDNA (double-stranded) along viral and cellular proteins make a protein complex called the pre-integration complex (PIC). This PIC is transported to the nucleus by the nuclear targeting action of Vpr (Andersen and Planelles 2005). As the size of the PIC is larger than a nuclear pore, an active cellular transport mechanism is used for its nuclear import (Reviewed in (Suzuki and Craigie 2007) (Step 4, Figure 1.6).

Once inside the nucleus, the viral cDNA is integrated into the host DNA (Figure 6, Step 4) with the help of viral integrase. Integrase processes the cDNA and attaches to the double stranded staggered cut in the chromosomal DNA with the help of cellular proteins (Vandegraaff and Engelman 2007). After integration, low-level spliced mRNA transcripts are produced with the help of cellular transcription factors. These transcripts encode the regulatory proteins Tat, Rev and Nef. Tat encodes a trans-activating protein, which increases the viral transcription rate (Emerman and Malim 1998). These transcripts are transported from the nucleus to the cytoplasm with the help of the Rev Response Element (RRE) protein encoded by *rev*. The RRE also permits the expression of structural proteins Gag, Pol and Env (Emerman and Malim 1998).

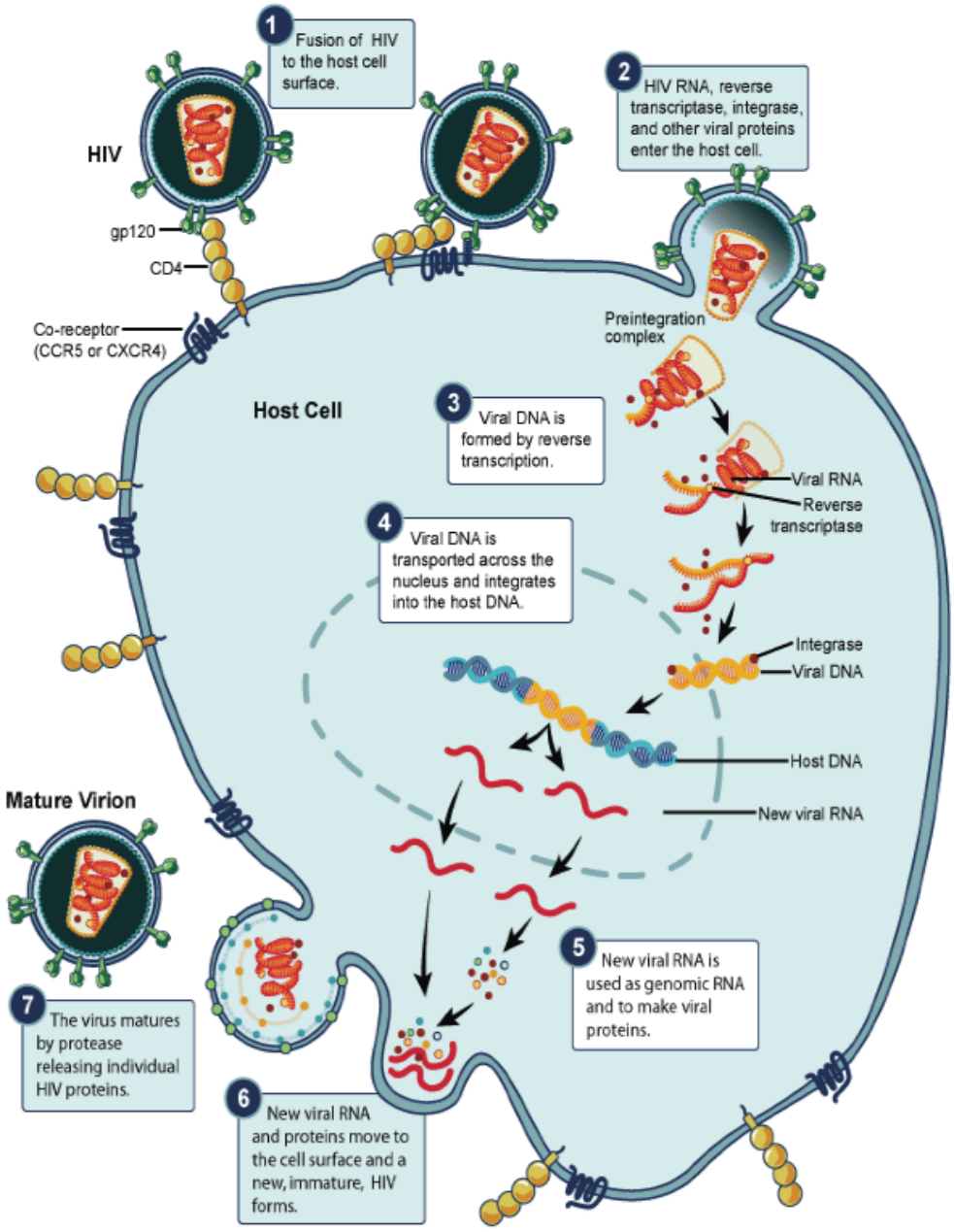


Figure 1.6: An overview of the HIV replication cycle. Image source: http://www.niaid.nih.gov/topics/HIVAIDS/Understanding/Biology/Pages/hivReplicationCycle.aspx. Dec 2011

The mRNA inside the cytoplasm is translated by host cellular mechanisms resulting in the formation of the large pre-proteins Pr55 Gag, Env precursor and Gag-Pol precursor (Figure 1.6, Step 5 and 6). Gag Pr55 and Gag-Pol are the results of two reading frames; first encoding pr55 and second for Gag-Pol. (Jacks et al. 1988). The Env precursor is produced by an mRNA of intermediate length, which, is cleaved during virion maturation to form other regulatory proteins, Vpr, Vpu and Vif along with gp41 and gp120, the envelope proteins. These proteins along with

membrane proteins from the host cell are added into the plasma membrane to make a viral envelope around the capsid containing the viral genome (Figure 1.6, Step 6). The virions emerge from the infected cells by a budding process. These virions are immature, non-infectious virus particles in which intact Gag polyprotein (Briggs et al. 2006). Eventually viral proteases cleave the Gag polyprotein, with the assistance of Vif, resulting in morphological rearrangements that lead to the formation of mature, infectious virus particles (Figure 1.6, Step 7). This completes a single replication cycle starting with one virion, leading to formation of several virions, of which approximately 5 to 11 go on to infect subsequent cells (Ribeiro et al. 2010).

1.3 HIV genome organization, proteins and their functions

The HIV-1 genome is linear, single stranded, plus-sense RNA, which is 9.2 kb in size, (Figure 1.7). Two copies of the RNA genome are present in the virus making the virus diploid. The two copies form a dimer, which is non-covalently linked near the 5' end of both strands. The RNA genome consists of a single protein-coding region, containing the genes *gag, pol,* and *env* that encode the major structural, enzymatic and envelope poly-proteins respectively. In addition to these proteins, six additional proteins are encoded by HIV-1 RNA namely Tat, Rev, Nev, Vif, Vpr and Vpu. On the both sides of the protein-coding region, two un-translated regions (UTRs) consisting of a 5' and 3' unique region (U5 and U3) flanked by two identical repeat regions (R). U5 and U3 are copied during reverse transcription to form the long terminal repeats (LTRs) of the proviral DNA. The protein-coding region of the genome also encodes important RNA structural elements, such as the rev response element (RRE) and the frame- shift sequence that are essential for viral replication (Malim et al. 1989; Watts et al. 2009).

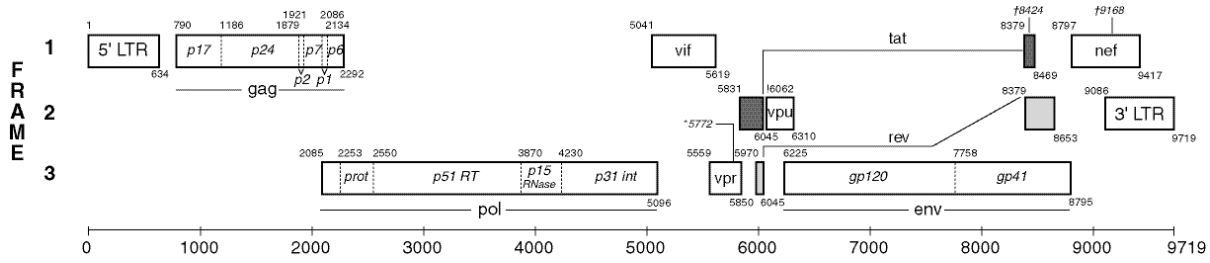


Figure 1.7: Schematic of the HIV-1 genome: (source: http://www.hiv.lanl.gov)

The HIV-1 genes are split into three main categories based on the function of the resulting proteins, structural (*gag, pol* and *env*), regulatory (*tat* and *rev*) and accessory (*vif, vpr ,vpu and nef*).

1.3.1 Structural Proteins

Gag: The *gag* gene (approx 1500 nucleotides) is translated into the 55-kDa Gag precursor protein (p55) from unspliced genomic RNA (Adamson and Freed 2007). This protein is associated with the host cell membrane and recruits the other viral proteins and RNA to the site. The continuous addition of p55 to the site of budding results in an outgrowth of the cell membrane, which, later makes part of HIV particle. When this particle is free of the host cell, p55 is cleaved by the viral protease (PR) during maturation into the virion structural proteins matrix (MA), capsid (CA) and nucleocapsid (NC) in addition to the C-terminal p6 domain and the spacer peptides SP1 and SP2. The matrix protein is responsible for targeting of p55 to the plasma membrane, incorporation of the PIC into the nucleus and also in the incorporation of the viral envelope into the virion surface (Bieniasz 2009). The capsid forms the shell of the virion core, which encapsulates the viral RNA, associated enzymes and nucleocapsid (Gelderblom 1991; Morikawa 2003). The nucleocapsid plays a vital role in viral RNA packaging to make a functional particle (Morikawa 2003). The p6 is responsible for the detachment of the virion from the cell surface and incorporation of Vpr into the virus particle (von Schwedler et al. 2003)

Pol: The *pol* gene encodes the enzymatic proteins: protease (PR), reverse transcriptase (RT) and integrase (IN), in addition to the transframe protein (TF/p6). The *pol* gene overlaps with the 3' end of *gag* and is translated from a -1 reading frame as a fusion protein with Gag. This fusion protein is known as Gag-Pol (Pr160Gag-Pol). Translation of Gag-Pol requires a heptanucleotide slippery sequence (UUUUUUA) and a stable RNA structure, which mediates a ribosomal frame-shift event (Mazauric et al. 2009). The protease enzyme is involved in the cleavage of the *gag -pol* precursor into the structural products of *gag* and enzymatic products of *pol* (Ross et al. 1991; Pettit et al. 2004). Reverse transcriptase is integral to transcribing viral RNA into cDNA during the HIV life cycle (Sarafianos et al. 2009). RNase H subsequently removes the RNA template after reverse transcriptase has finished DNA synthesis (Schultz and Champoux 2008). The last enzyme produced by *pol* is integrase, which is involved in the insertion of proviral cDNA into the host chromatin and also plays a part in the transport of the PIC through the nuclear membrane (Chiu and Davies 2004).

Env: The *env* gene is translated into the Env 160-kDa glycoprotein precursor (gp160) from a singly spliced RNA by ribosomes associated with the rough endoplasmic reticulum (ER). Envgp160 is cleaved by a cellular furin-type protease into the mature gp120 and gp41 Env subunits (Moulard and Decroly 2000). These two proteins are present on the cell surface as a trimeric complex consisting of three gp120 monomers with three gp41 monomers known as the envelope spike (Liu, Liu, and Fu 2008). The primary function of gp120 is the binding of host receptor CD4 on the cell surface followed by interaction with a coreceptor, either CCR5 or CXCR4 (Wang et al. 2008). The gp41, a glycoprotein, plays an important role in the membrane fusion and viral entry (Lay et al. 2011).

1.3.2: Regulatory proteins:

HIV-1 encodes two essential regulatory proteins, regulator of virion expression (Rev) and trans-activator of transcription (Tat), which mediate their effects through conserved RNA structures. The Rev protein is responsible for the nuclear export of the viral mRNA in the correct unspliced form to the cytoplasm (Dayton 2004). This protein contains both a nuclear localization sequence (NLS) and a nuclear export sequence. The Rev response element (RRE) of the *env* gene binds to the RNA binding motif of Rev, helping the export of the HIV genome from the nucleus (Malim et al. 1989; Dayton 2004). The Tat protein helps in promotion of the transcription of proviral DNA in order to ensure the full length transcripts are produced (Debaisieux et al. 2011).

1.3.3: Accessory proteins: HIV-1 also encodes four accessory proteins: Nef, Vif, Vpr, and viral protein U (Vpu). The accessory proteins are not absolutely required for infectivity in cell culture but are critical virulence factors in vivo. The Nef protein interacts with host kinases resulting in arrested cell growth and increased infectivity (James et al. 2004). Nef is responsible for the reduction of the immune response to HIV infection (Hung et al. 2007), an increase in the efficiency of reverse transcriptase (Fournier et al. 2002) and reduction of the levels of CD4 on the cell surface (Schwartz et al. 1995; James et al. 2004). Vif prevents APOBEC3G from blocking HIV replication (Sheehy et al. 2002). As APOBEC3G is present only in peripheral blood lymphocytes and macrophages, the Vif function is limited to those cells only (Navarro and Landau 2004). Vpr is responsible for several roles which include,

ensuring the accuracy of the reverse-transcription, nuclear import of the viral DNA (part of PIC), cell cycle progression (Le Rouzic and Benichou 2005).

The Vpu helps the release and prevents endocytosis of retroviral particles from the cell membrane (Neil et al. 2006). It also down regulates the CD4 on the cell surface thus reducing the incidence of CD4-Env complexes hence increasing the processing of gp160 (Willey et al. 1992a; Willey et al. 1992b). In addition, Vpu also interacts with the invariant chain of MHC-II, which decreases its cell surface presentation and thus interferes with the immune response. (Hussain et al. 2008). Vpu also inhibits the host protein tethrin (Neil, Zang, and Bieniasz 2008), which helps the release of the viral particles from the cell surface without associated cytopathicity (Klimkait et al. 1990).

1.4 Immunopathogenesis of HIV infection

The rate of progression of HIV disease may be different among HIV-infected individuals. The time from initial infection until progression to AIDS in normal progressors is eight to ten years. However some (10 to 15%) progress to AIDS within two to five years (rapid progressors) and less than 5% remain asymptomatic for at least 10 years (long term non-progressors) (Levy 2007). Disease progression following HIV infection includes broadly three phases; primary infection, clinical latency and AIDS.

The primary phase of infection lasts for a few weeks. Most of the HIV-1 infected individuals show flu-like symptoms during this phase, however, some HIV-infected patients remain asymptomatic during the primary infection phase (Schacker et al. 1996). A burst of viremia (up to 10^7 HIV RNA c/ml of plasma) and a high level of infected peripheral blood mononuclear cells (PBMCs) are associated with the primary phase of infection (Clark et al. 1991; Graziosi et al. 1993) and at the immunological level, CD4+ T cells rapidly decrease. Six to eight weeks after the appearance of symptoms, the HIV specific immune responses appear and the acute viral syndrome resolves. The virus-expressing cells are eliminated by HIV specific cytotoxic T lymphocytes and antibodies against different HIV proteins are produced thus contributing to the down regulation of viremia (Clark et al. 1991; Daar et al. 1991) These antibodies form immune complexes, trap the virus in the reticulo-endothelial system, while neutralizing antibodies are only detected several months after the seroconversion (Koup et al. 1994).

The second phase, the clinical latency phase, involves a temporary rise in numbers of CD4+ T lymphocytes and the down regulation of viremia. A persistent decline in the immune system is observed due to the lack of ability to completely regenerate the CD4+ T lymphocytes, while virus replication is still ongoing at a baseline level. When the CD4+ T cell count falls below 200 cells/µl the immune system is no longer capable of controlling the virus and other pathogens, and the patient becomes vulnerable to opportunistic infections. This is the onset of the AIDS, the final stage of HIV infection (Levy 2007), The pathogenesis and evolution of HIV-1 infection is depicted in Figure 1.8.

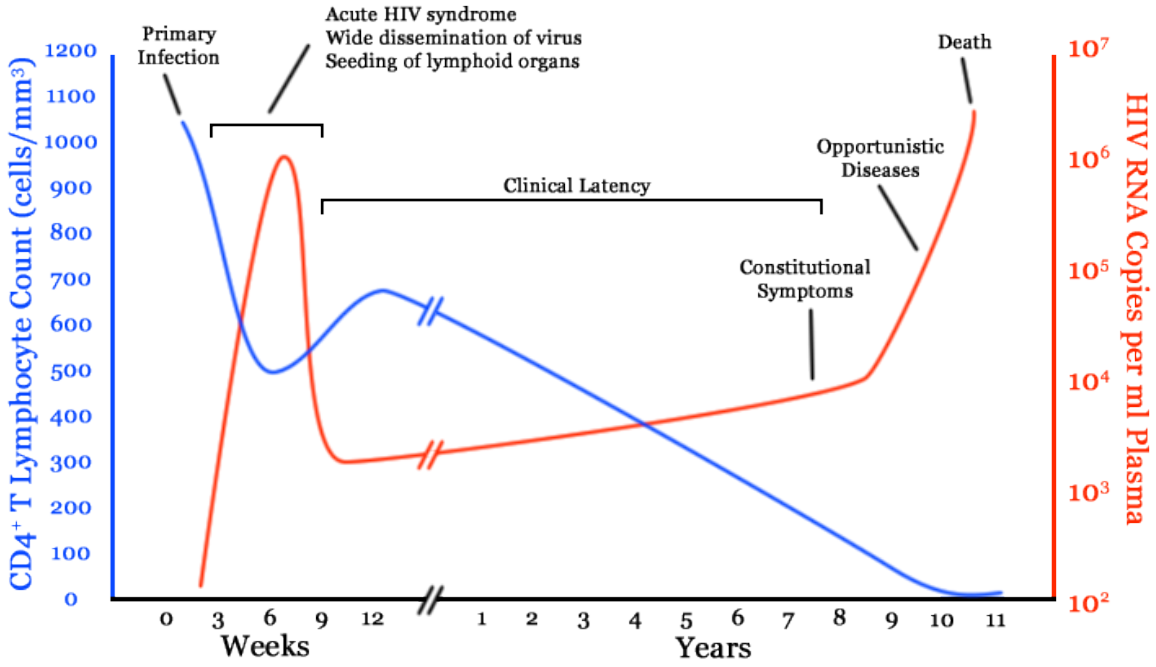


Figure 1.8: The Immunopathogenesis of the HIV-1 infection. (Fauci et al. 1996).

1.5 Classification and Diversity of HIV

HIV is classified as member of the family Retroviridae due to the ability of its RNA genome to be reverse transcribed into DNA (Barre-Sinoussi et al. 1983). HIV is further classified as belonging to the genus *Lentivirus* due to its long asymptomatic period within its host (Buchen-Osmond 1997). HIV is distinguished from other types of retroviruses due to the presence of accessory gens (*tat, rev, nef, vpr, vif and vpu/vpx*), which promote viral infectivity and the modulation of host cell functions. While lentiviruses were known to infect sheep, goats, cows and horses, HIV was the first lentivirus found in primates along with its closest relative Simian Immunodeficiency Virus (SIV) (Buchen-Osmond 1997).

The reverse transcriptase enzyme of HIV is error-prone, i.e. it lacks proofreading activity. This confers a mutation rate of approximately 3.4×10^{5} mutations per base pair per replication cycle (Nowak 1990). As the genome size of HIV-1 is an estimated 10^{4} base pairs in length and the rate of viral production is approximately 10^{10} virions per day (Perelson et al 1996), a large number of viral variants are produced within an infected person per day. Viral diversity also occurs due to recombination events when reverse transcriptase transfers between the two RNA templates within a single virus, generating a daughter DNA that is a recombinant of the parental genomes (Negroni and Buc 2001). Recombination between two separate viruses can lead to further viral diversity when one person is co-infected with two separate strains of the virus. This genetic diversity has significant clinical and public health consequences such as the development of drug resistance and in significantly hampering vaccine development (Butler et al. 2007).

The globally circulating strains of HIV-1 are classified into several major phylogenetic groups: M (main), N (Non-M-Non-O), O (outlier) and P (Plantier et al. 2009). Group M is responsible for the global epidemic, which is widely accepted to be the result of a zoonotic transmission of SIVcpzPtt from chimpanzees (*Pan troglodytes troglodytes*) to humans (Gao et al. 1999; Korber et al. 2000). Comparative sequence analysis suggests that this transmission event occurred at the beginning of the 20th century (Korber et al. 2000; Worobey et al. 2008). Despite it's relatively recent evolutionary history, Group M is so diverse that it can be further subdivided into nine genetically associated subtypes which are equidistantly related

(A, B, C, D, F, G, H, J and K) (HIV Sequence Compendium 2011). Some groups can be further divided into sub-subtypes: F (F1 and F2) and A (A1 and A2). The genetic distance between the subtypes depends on the region of the genome analysed, as different regions of the genome have different rates of evolution. The coding sequences of the envelope protein (Env) are separated by 25-35%, whereas the coding sequences of the structural components (Gag) are more conserved and are separated by approximately 14%, across subtypes (Janssens, Buve, and Nkengasong 1997; Thomson, Perez-Alvarez, and Najera 2002). Nevertheless, the diversity of HIV-1 is so great that even isolates that form the same subtype will be significantly different at both the RNA and protein level.

It was subsequently realized that for some isolates, the difference in their placement on phylogenetic topologies generated from different genomic regions was the result of recombination between viral isolates of different subtypes. These isolates are termed circulating recombinant forms (CRFs)(Robertson et al. 2000). CRFs play an important role in the pandemic and it is estimated that around 20% of global infections are recombinants (Hemelaar et al. 2011). Thus, the HIV-1 epidemic is highly dynamic and continues to rapidly evolve. Figure 1.9 depicts the global distribution of HIV-1 subtypes. We can see that subtype C is almost exclusively responsible for all infections in Southern Africa, Horn of Africa and Southern Asia (Ariën, Vanham, and Arts 2007), including Malawi, where this project is based. The greatest diversity of subtypes is found in the African continent where, along with subtype C, other subtypes such as A1, D, B and various circulating recombinant forms (CRFs) and Unique Recombinant Forms (URFs) are found (Figure 1.9).

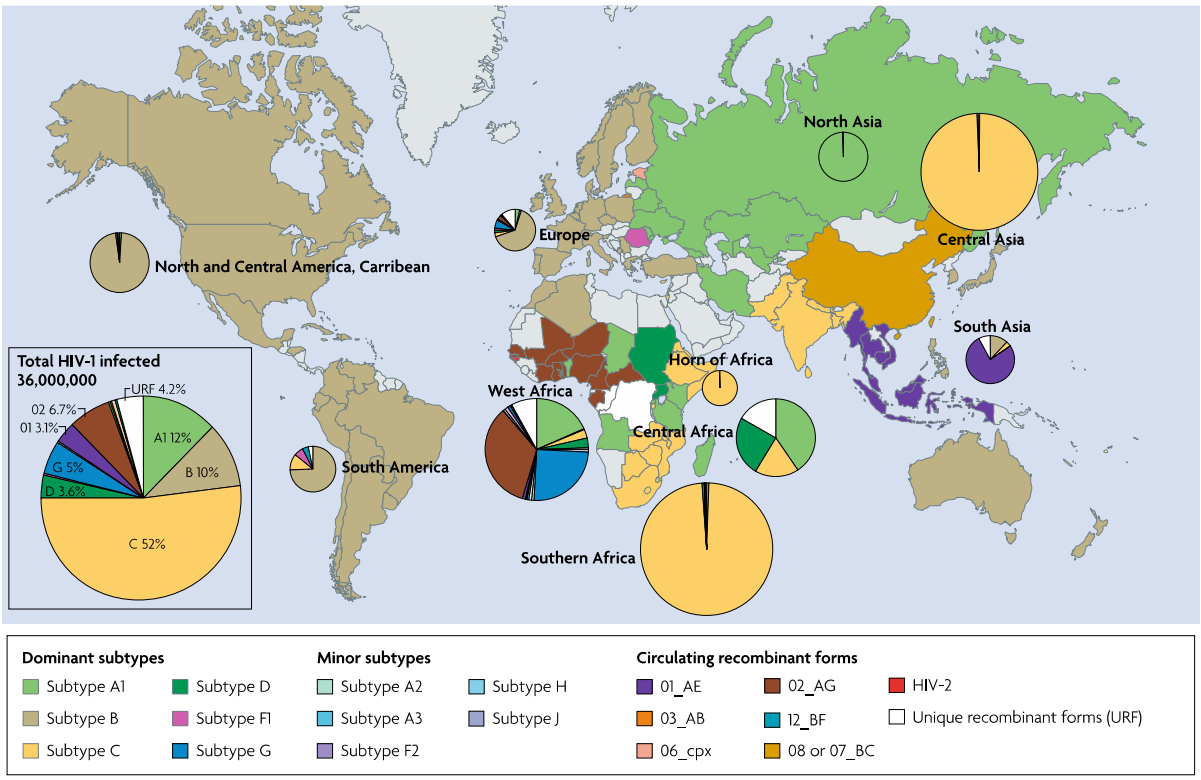


Figure 1.9: The breakdown of subtype prevalence in the worldwide epidemic. Subtype C is the most predominant subtype in the epidemic among all other subtypes and is prevalent in southern Africa, Horn of Africa and South Asia. (Ariën, Vanham, and Arts 2007)

1.6 Antiretroviral therapy and drug resistance

In 1987, the first approval of zidovudine (AZT) was made by the US Food and Drug Administration (FDA) leading to the first breakthrough in antiretroviral therapy. It was soon discovered that HIV quickly developed drug resistance when AZT was given (Larder, Darby, and Richman 1989; Larder and Kemp 1989). Although other RT inhibitors became available in next few years, a major breakthrough trial that was announced by (Gulick et al. 1996) showed the potent antiretroviral activity of the triple combination of two RT inhibitors and a protease inhibitor. Since then it has been shown that this triple therapy, also known as highly active antiretroviral therapy (HAART) can reduce viral load below the detection limit of current assays (less than 50 RNA copies/ml) and restore the number of CD4+ T cells back to normal levels (Gulick et al. 1997; Hammer et al. 1997). This therapy was immediately introduced in clinics and had a big impact on anti-HIV treatment. By the end of 2008, more than 20 antiretroviral drugs had been approved by the FDA. These can be divided into five classes (Figure 1.10) (Palmisano and Vella 2011).

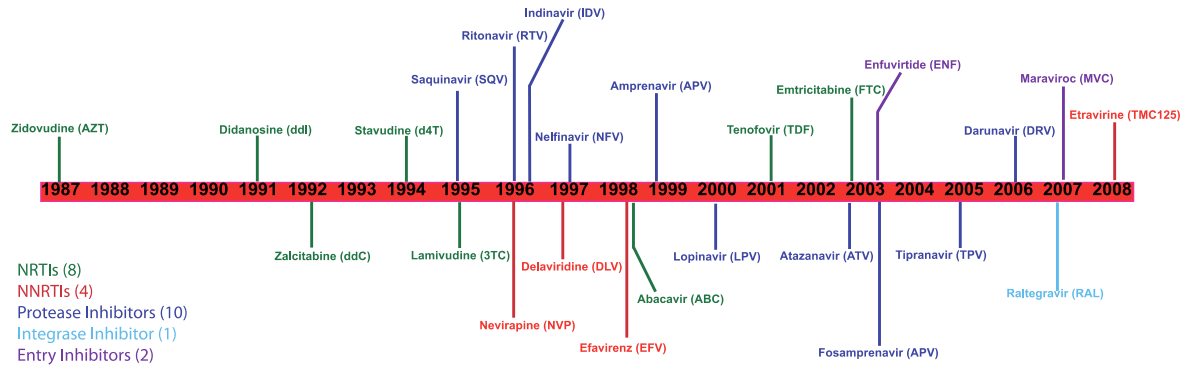


Figure 1.10: Antiretroviral drugs approved by US FDA as of 2008. They are divided into five classes on the basis of their mode of action.

The Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NRTIs) compete with natural dNTPs to bind the RT enzyme. The NRTI acts as chain terminator due to the lack of a 3' hydroxyl group. Currently eight NRTIs are approved by US FDA (Figure 1.10) (Palmisano and Vella 2011). However, despite their central role in HIV therapy, NRTIs have been associated with severe side-effects, especially related to mitochondria damage and dysfunction (Cherry and Wesselingh 2003). The second class of antiretroviral drugs, the NNRTIs (Non-Nucleoside Reverse Transcriptase Inhibitors), inactivate the RT enzyme by binding to the hydrophobic pocket. These inhibitors are HIV-1 specific and will therefore not bind HIV-2 RT or cellular DNA

polymerase (Sluis-Cremer, Arion, and Parniak 2000). Currently there are four NNRTIs approved by the US FDA (Figure 1.10) (Palmisano and Vella 2011). The third class of ARV drugs, the protease inhibitors (PIs), bind to the PR cleavage site and inhibit the enzyme using two different mechanisms: They can bind to the active site, or mimic the transition state during peptide cleaving (Vazquez et al. 1995). These drugs are mainly synthetic analogues of the phenlalanine-profile cleavage site in the *gag -pol* polyprotein (Flexner 1998). A total of ten protease inhibitors were approved by US FDA by 2008 (Figure 1.10) (Palmisano and Vella 2011). Only these three classes of drugs are available in Malawi to date (Health 2007).

Additional antiretroviral drugs available are integrase inhibitors (INIs), fusion and entry inhibitors. The integrase inhibitors target the point of the HIV life cycle where viral DNA is incorporated into the host DNA. By blocking the function of integrase, these drugs block viral genome integration and thus halt the HIV life cycle at this point. Integrase inhibitors bind directly to the active site of integrase and make multiple contacts with the protein in order to inactivate its function (Goldgur et al. 1999).Only one integrase inhibitor (INI) (Raltegravir) is approved by US FDA. One other INI (Elvitegravir) is currently in late stage clinical trials(Palmisano and Vella 2011). Fusion and entry inhibitors work by preventing the virus from infecting a cell, by either inhibiting fusion between the virus and the host cell or by blocking the co-receptors. Only one fusion inhibitor (Enfuvirtide) has been approved by the FDA (Palmisano and Vella 2011). The binding of gp120 to CD4 and CCR5, leads to conformational changes in gp41, which leads to the fusion of viral and cellular membranes. ENF inhibits this fusion by preventing gp41 to undergo a fusogenic conformation (Matthews et al. 2004). Recently, the first co-receptor blocker, Maraviroc, was approved by the FDA (Palmisano and Vella 2011). This drug makes the CCR5 receptor unavailable for 'R5 tropic' HIV, whereby the virus cannot engage with the CD4+ T-cell to initiate an infection (Dorr et al. 2005). A new approach of using the *gag* gene as a target is being studied as the products of this gene are involved in assembly, maturation and release. However, these proteins have not yet been successfully exploited in the development of antiviral therapies (Waheed and Freed 2012).

1.7 The Importance of Drug Resistance Testing

Despite the success of antiretroviral therapy in HIV management and control of spread, resistance to antiretrovirals has emerged in all regions. Similar to developed countries, with increasing access to antiretroviral drugs, the problems involving drug resistant virus have begun to occur in developing countries (Hirsch et al. 2008). It has been found that approximately 8.4% of the infections in Europe are with transmitted drug-resistant HIV (Vercauteren et al. 2009). If new infections in developing countries where HIV prevalence is high occur at similar proportions, it will have serious consequences by the rapid spread of drug resistant viruses in the population. A resistance test can provide information regarding the presence of resistance mutations, when clinical, immunological or virological parameters indicate that a specific regimen does not have the desired effect of viral suppression (Langford, Ananworanich, and Cooper 2007). With the help of this information, clinicians choose to alter the drug regimen or decide the most appropriate new regimen. As antiretroviral therapy was introduced to resource-limited settings more recently, the experience with the therapy is limited. Because NRTI regimens were standard practice in Europe and the US for many years before the approval of first PI in 1996 (Palmisano and Vella 2011) drug resistance to NRTIs had a longer time to develop. The more recent introduction of ARVs, which includes PIs as well as NRTIs and NNRTIs, to developing countries, may mean that drug resistance will not develop in the same way.

Although antiretroviral therapy was introduced in resource limited settings more recently; the development of drug resistance in inevitable. Resistance testing is not strictly necessary when a first-line regimen is failing if new drugs can replace all drugs in the regimen without possible cross-resistance with the drugs from failing regimen, i.e. drugs from a different class. However, in resource-limited settings, alternative drugs for second line treatment are limited and drugs from a different class are not always available (Keiser et al. 2009). However, drug resistance testing in these countries might be indicative from the first line failure onwards. Furthermore, resistance testing can be a useful tool to prevent unnecessary treatment switches in patients with extremely low adherence. In such patients, virologic failure can be experienced with no evidence of resistance, in which case a switch to an expensive

second-line regimen can be prevented by focusing on the improvement of the patient's adherence (Macias et al. 2005).

Transmitted drug resistance (TDR) is significantly prevalent in countries with established ART programs (Murillo et al. 2010; Wheeler et al. 2010; Avila-Rios et al. 2011) and resistance testing prior to the administration of ARV drugs can thus be useful. The highest prevalence of TDR is seen in the US and Western Europe (6.0% to 24.1%) and some parts of South- America (3.1% to 22.2%) (Rahim et al. 2009). Since ART became more readily available in resource- limited settings, a low TDR prevalence has also been detected in some Sub-Saharan countries (Geretti 2007). These regions are even more at risk because therapy switches are often based on clinical criteria allowing prolonged virus replication under drug pressure, hence promoting the development of drug resistant strains. However, an excellent virological and immunological response is seen in TDR patients when the first-line regimen was based on the baseline resistance results (Oette et al. 2006; Shet et al. 2006) thus underlining the importance of resistance testing. Resistance testing is not only useful before initiation of treatment at an individual patient level, it can also be used in population-based studies to monitor the success of treatment initiatives and the effectiveness of HIV transmission prevention programs between HAART treated patients (Gilks et al. 2006). In countries such as Malawi, where the antiretroviral treatment program started in 2004 (Zijlstra and van Oosterhout 2006), very little data was available on drug resistance. Therefore, monitoring the genotypes of treatment naïve and individuals receiving treatment is necessary in order to monitor the presence, emergence and transmission of drug resistance. One of the aims of this study is to monitor the emergence of drug resistance to first line therapy in Karonga District Malawi.

1.7.1 Types of Drug Resistance Testing

Three major approaches are used to measure drug resistance in HIV-1 against antiretroviral drugs. The most direct method is the one that measures the phenotypic susceptibility of the virus to drugs by directly culturing virus in the presence of increasing concentrations of the drugs of interest. The second type determines the sequence of the genes targeted by the drug. This is the most commonly used method because it is relatively less laborious and time consuming. A third method,

the virtual phenotyping assay, is a mixture of first two methods. In this method the virus is sequenced and changes that are identified are matched to a large dataset of genotype-phenotype pairs of viruses that are previously sequenced and for which the phenotype is known (Clavel and Hance 2004).

1.7.1.1 Phenotypic resistance testing

Phenotypic drug resistance testing measures the susceptibility of the virus to the drug in an *in vitro* culture assay. Initially this test was done by enzyme-based assays (Japour et al. 1993), as they were very expensive and laborious, a new, replication-based assay, was developed. Briefly, the PCR (Polymerase Chain Reaction)-amplified gene of interest is incorporated in a HIV backbone or vector, which lacks the gene that is sequenced from the patient sample. (Hertogs et al. 1998; Petropoulos et al. 2000; Zhang et al. 2004; Garcia-Perez et al. 2007). The susceptibility of the viruses to drugs is measured by the fold difference in drug concentration that is needed to inhibit the recombinant virus compared to the wild-type. Results are expressed in IC_{50} or IC_{90} where the IC_{50} and IC_{90} are the drug concentrations needed to inhibit 50% and 90% of the virus replication respectively. The interpretations of the test employ the biological and clinical cut-offs. The clinical cut-offs allow better interpretation of results to guide the clinical decisions (Harrigan et al. 2001; Swanstrom et al. 2004). The biological cut-off defines the fold changes at which the virus can be assumed to be resistant in vitro compared to the wild-type virus. The clinical cut-off is based on clinical observations from various drug resistant patients in clinical trials and cohorts. Because clinical cut-offs are only available for a limited number of drugs, the clinician depends on 'biological cut-offs' that are derived from *in vitro* susceptibility experiments with clinical isolates from drug naïve patients. These values cannot directly link *in vitro* resistance information with the actual phenotype of the virus (Hirsch et al. 2003). Another disadvantage of the phenotypic assay is the lengthy time scale, with analysis taking up to four weeks to complete, further more, the method is expensive. Phenotypic assays were not available for this project and so resistance testing involved the genotypic approach.

1.7.1.2: Genotypic resistance testing:

A genotypic resistance testing identifies mutations in protease, reverse transcriptase and gp41 at amino acid positions that are related to resistance to

antiretroviral drugs. A list of detected mutations is generated and frequently updated (Johnson et al. 2011). The drug resistance mutations (DRMs) are represented by a code, e.g. M41L, whereby the number refers to the amino acid position in the respective genes, the first letter is the amino acid present in the wild-type amino acid and the letter at the end is the mutated amino acid. Genotypic assays are the most commonly used methods of testing drug resistance because of their relatively low cost (at least in developed countries) and low turnaround time. However, these assays have the disadvantage of just producing a list of mutations. A correlation between a single mutation and resistance towards a specific drug is not always straightforward. Characteristics such as hyper- susceptibility, cross-resistance and re-sensitization are difficult to take into account when interpreting the results of genotypic assays (Vandamme, Van Laethem, and De Clercq 1999). Different algorithms for the interpretation of mutational patterns have been developed (De Luca et al. 2003). Most commonly used are the Rega Institute (Rega) algorithm (Van Laethem et al. 2002), Antiviral Drug Resistance Analysis (ADRA) (http://www.hiv.lanl.gov/content/sequence/ADRA/adra.html), and the HIV & Protease Sequence Database (HIVDB) algorithm, Stanford University (http://hivdb.stanford.edu/index.html). The first two algorithms describe specific mutational patterns associated with a particular antiretroviral drug/s, whereas the HIVDB gives a score to each of the observed mutations in the sequences. The last approach takes patterns of mutations into consideration. Here, the mutation scores are based on published literature with links of specific antiretroviral drugs and mutations. This includes correlations between the genotype and treatment, genotype and phenotype, and genotype and clinical outcome (Rhee et al. 2003; Shafer 2006). It was considered that this latter approach was the best and so was the approach employed throughout this project.

Two major techniques are used to detect DRMs: bulk sequencing and single point mutation assay. Bulk sequencing uses a viral population-based sequencing method that produces a single sequence from the viral pool that is present in the PCR reaction. This method is most commonly used because as it is relatively quick, less laborious and less time consuming. However, it only produces list of mutations and is not able to detect minor variants. Several sequencing methods are

commercially available, though a number of laboratories have developed their own in-house methods to reduce costs (Hoffmann et al. 2007). The single point mutation assays such as LigAmp (Shi et al. 2004) or LiPA (Derdelinckx et al. 2003) are more sensitive for minor variant detection compared to bulk sequencing. These assays use specific probes to detect the single point mutations. But the disadvantage of this test is the limited number of mutations that can be detected (Stuyver et al. 1997) and new probes which need to be designed whenever a new mutation is found.

Various other methods are used to detect DRMs such as selective PCR, which is carried out by performing two separate PCR reactions, one to amplify the wild-type sequence and the other to amplify the mutant sequence. The primers are specifically designed to hybridize with either the mutant or wild-type sequence and amplification will only take place when the primer perfectly matches the target sequence (Anderson et al. 1994). Selective PCRs for specific mutations were used to detect minor mutant variants (Bergroth, Sonnerborg, and Yun 2005; Descamps et al. 2006). Another method, oligonucleotide ligation assay uses two differently labeled oligonucleotides, which are hybridized to a template. Only the oligonucleotide with the perfect match can be detected (Stuyver et al. 1997; Edelstein et al. 1998; Derdelinckx et al. 2003).

All the genotypic assays (except the single point mutation assay) are unable to detect minor variants that represent <20% of the viral population but that can be clinically significant (Johnson et al. 2008; Paredes et al. 2010). Several other methods are used to identify minor drug resistant variants, such as sequencing of multiple clones, sequencing after single genome amplification (Palmer et al. 2005), allele-specific real time PCRs (Roquebert et al. 2006) and microarrays (Kozal et al. 1996). However these methods are time consuming and labour intensive. Recently a new approach, based on pyrosequencing –next generation sequencing, has been used in a number of studies to detect minor drug resistant variants within viral populations (Simen et al. 2009; Hedskog et al. 2010; Delobel et al. 2011). This approach can produce large numbers of sequences from multiple samples in a single run and has been found to be very efficient in detecting drug resistant variants present only up to 0.1% of the viral population (Jabara et al. 2011). Although this approach is currently at a development stage, in future, it can be very useful to

reduce cost and labour, and also to increase the sensitivity of minor variant detection. This project began by using bulk sequencing to monitor the prevalence of DRMs in a HIV-1 positive population in rural Malawi and to monitor the development of resistance in patients receiving ART. Clonal sequencing was also employed for detection of intra-individual variation and moved towards employing the ultra-deep pyrosequencing approach for the identification of minor DRMs in this resource-limited setting.

1.8 HIV/AIDS and access to treatment in resource-limited countries

The number of people living with HIV and the number of AIDS-related deaths continues to rise despite the efforts to counter the epidemic. Unfortunately, resource-limited settings continue to carry the burden of the epidemic with 68% of all HIV-infected adults living in Sub-Saharan Africa. Sub-Saharan Africa also accounted for 70% of the new infections in the year 2010 with the epidemic continuing to be most severe in southern Africa (UNAIDS 2011b).

1.8.1. Increasing access to antiretroviral therapy

In the last decade, access to ART in low and middle-income countries has expanded significantly. When WHO (World Health Organization) and UNAIDS (Joint United Nations Program on HIV/AIDS) started the “3 by 5” initiative of ART in December 2003, only 400, 000 people in these countries received ART at the time (UNAIDS/WHO 2007). By the end of 2010, the percentage of people in need of ART in low and middle-income countries increased from 7% to 47% with a 20% increase between 2009 and 2010. The increase was most dramatic in sub-Saharan Africa resulting in more than 6.7 million people with access to ART today. (UNAIDS 2011b). Figure 1.11 shows the rapid scale up of ART in Africa from 2002 to 2005, but also indicates the differences in progress between countries.

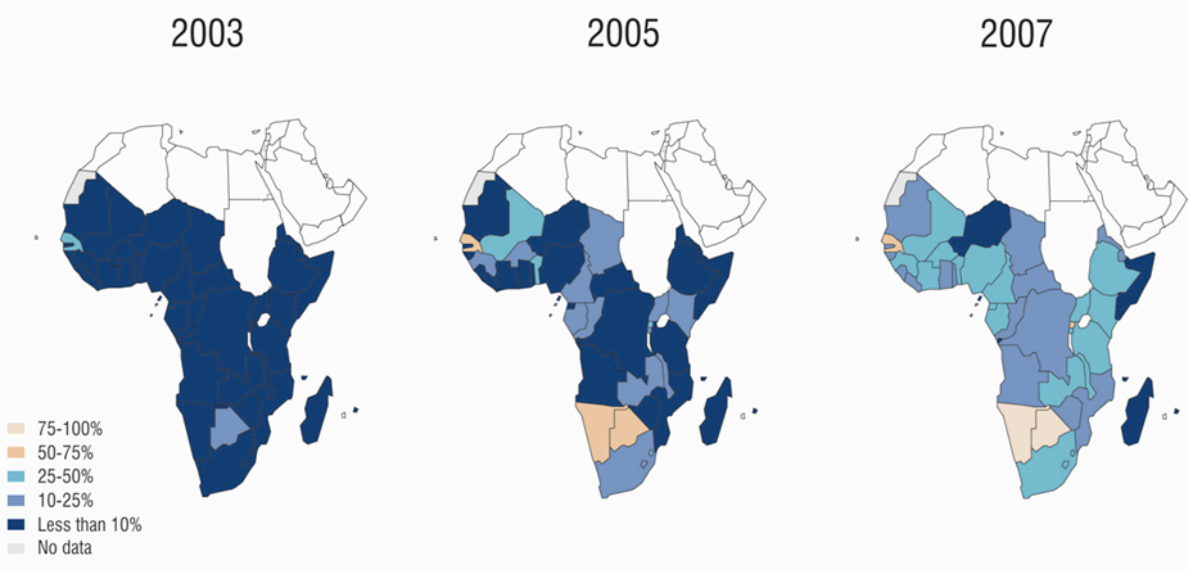


Figure 1. 11: People in sub-Saharan Africa receiving antiretroviral therapy as a percentage of those in need, 2003-2007 (WHO 2006b; UNAIDS/WHO 2007)

1.8.2 Access to laboratory tests to monitor the drug resistance

In resource-limited settings such as countries in Sub-Saharan Africa, understandably, more financial resources were made available for the prevention and treatment of diseases (Bates and Maitland 2006). Thus, very little funding was made available to improve laboratory capacity in order to support the services to monitor these patients (Petti et al. 2006). Viral load tests, CD4+ T cell counts and even resistance tests are considered a standard part of the care in the Western world (DHHS 2006). These assays can help the clinicians to decide when to start ART and when to switch a failing ART regimen to an alternative regimen. However, in some resource-limited settings, the combined cost of a commercial viral load assay and a CD4+ T cell count test exceeds twice the cost of one month ART (Stephenson 2002). According to the WHO guidelines to scale up ART in resource-limited countries, a CD4+ T cell count is expected at baseline and every 6 months when on treatment (DHHS 2006). The CD4 T cell tests are becoming more commonly available although the supply of reagents remains challenging in most places (Steegan 2007). It has been shown that there is a significant imbalance of therapeutic responses based on clinical and immunological criteria only (Hosseinipour, van Oosterhout, and Weigel 2007), therefore monitoring of viral load can be used as a parameter to access treatment responses and to monitor patients' adherence to ART (Boulle et al. 2006). But due to poor lab infrastructure and lack of trained personnel, viral load assays in resource-limited settings are not always possible. Performance data of HIV viral load assays in resource-limited settings are still limited.

In the case of treatment failure, WHO advises to change all components of the regimen (WHO 2006a). However, due to the limited availability of second-line ARV drugs in resource-limited settings, replacing all components of the first-line regimen is very difficult and recycling of drugs for a second-line regimen remains the only option. In this case, the efficiency of the second-line regimen will definitely improve if the choice of drugs can be guided by resistance testing. The need for drug resistance assays will increase even more when the number of available ARV drugs in resource-limited settings increases and third-line regimens become an option. Along with the technical, logistical and economical obstacles to implementing treatment-monitoring assays in resource-limited countries, sample collection, transportation and storage remain as additional challenges. Many sites lack a stable

electricity supply, hence it is often difficult obtain high quality samples that were appropriately processed and stored. In this respect dried blood spots (DBS) are an ideal medium to collect samples in the field (Steegen et al. 2007).

1.9 HIV/AIDS in Malawi

Malawi is a landlocked country in southeast Africa, which is bordered by Zambia, Tanzania and Mozambique. (Figure 1.12) Malawi is over 118,000 KM^2 with an estimated population of 13,900,000. HIV-1 prevalence in Malawi currently stands at approximately 11% (UNAIDS 2010) and Subtype C is the predominant subtype however, two other subtypes (A and D) and three recombinant forms (AC, AD and DC) were also found in the Malawian population (McCormack et al. 2002; McCormack et al. 2003; Petch et al. 2005; Hosseinipour et al. 2009). Although prevalence is higher in urban areas, 80% of the population lives in rural areas, where the epidemic remains a concern and access to treatment more problematic. Approximately 250,000 (49-57% of ART eligible individuals) are currently on treatment (UNAIDS 2011a). Around 95% of them are on first-line therapy, 4% are on alternative first-line therapy, and less than 1% are on second-line therapy (Health 2007).

The Malawi national ART program includes only reverse transcriptase inhibitors as first line therapy (Health 2008). Briefly, a generic fixed-dose combination treatment (Triommune®) is given which consists of: stavudine (d4T) and lamivudine (3TC), which are both Nucleoside Reverse Transcriptase Inhibitors (NRTIs), and nevirapine (NVP) which is Non-nucleoside Reverse Transcriptase Inhibitor (NNRTI). An alternative first line therapy is available for those having unacceptable side effects to Triommune, which entails replacement of d4T with Zidovudine (AZT – a NRTI) and NVP with efavirenz (EFV- NNRTI). If an individual is considered to have failed first line treatment, second line treatment is available on a limited named-patient basis which includes protease inhibitors (lopinavir/ritonavir) and tenofovir, Zidovudine (AZT) (Health 2008).

The Karonga District is the northernmost district of Malawi and is largely rural with the presence of one small town, Karonga. It is bordered by Lake Malawi on the east, bordering Tanzania on the north and the Nyika Plateau to the south and west (Figure 1.12). The district was isolated until 1981, until the national highway was extended to connect Tanzania from the south. The district is the site of the Karonga Prevention Study whose project headquarters are approximately six hours drive from Malawi's capital Lilongwe. The population of Karonga district is approximately

270,000. Mostly it is rural, and dependent upon subsistence agriculture and fishing. Health-wise, pneumonia, diarrhoea and malaria are highly prevalent and HIV is the principal killer of young adults (Glynn et al. 2001; Gordon and Stephen 2006). Tuberculosis incidence increased since the introduction of HIV in the 1980s with about two thirds of TB arising from recent infection (Crampin, Glynn, and Fine 2009). HIV was present by 1983, and approximately 10 % of adults are infected (Crampin, Glynn, and Fine 2009). Government health services are basic, with one District Hospital in Karonga town with two doctors and six clinical officers. There are two rural hospitals (in Chilumba and Kaporo) and ten health centers throughout the district (McGrath et al. 2007).

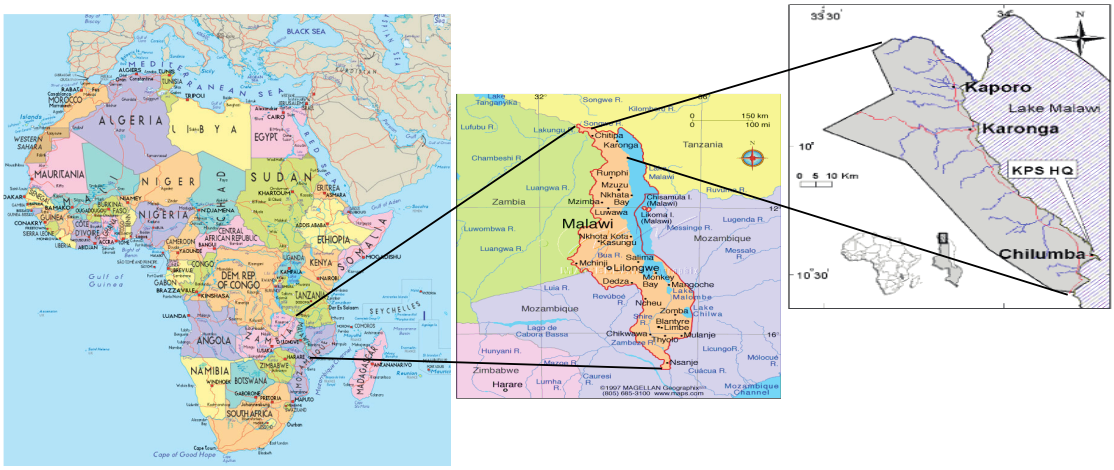


Figure 1.12 Map of African continent showing Malawi and Karonga district. The Study site (KPS-Chilumba) is located at southern part Karonga.
Sources:http://www.map-of-africa.co.uk, http://www.infoplease.com/atlas/country/malawi.html and http://www.lshtm.ac.uk/eph/ide/research/kps/district/index.html

A large research project named the Karonga Prevention study or KPS in brief (originally named as LEPRA) was initiated in 1979 on the epidemiology of mycobacterial infections in the Karonga district. Later, a decline in Leprosy and an increase in HIV and TB infections was observed. The project's focus then diverted towards TB, HIV and other infectious diseases. The work included molecular epidemiology, vaccine and treatment trials as well as demography, immunology and genetics (Crampin, Glynn, and Fine 2009). Thus over the last three decades, the KPS has made major contributions to the understanding and control of infectious diseases in this area.

The KPS includes demographic surveillance of a population of 32,000 in a continuous registration system that started in 2002, involving 135 km^2 near the KPS campus in Chilumba (Figure 1.12). In this surveillance area, there is a continuous registration of births and deaths and demographic changes to whole households. A sexual behavior questionnaire, counseling and testing of HIV (HIV Sero-survey-HSS) is carried out for all adults at baseline and after 4 years, but it is carried out annually for younger age groups i.e. those groups at greatest risk.

Full ART first became available at Karonga district hospital in June 2005. Since then ARV delivery clinics have been established at four smaller health centers in the district. ARV delivery has been largely based on clinical assessment with CD4 counting becoming available in January 2008 at one clinic (McGrath et al. 2007). Viral load assessment and routine drug resistance testing are not available and the base line genotype of the genes, which are the targets for ARVs are unknown. As part of a larger study investigating how the availability and use of ART may change the HIV epidemic and its socio-demographic impact in a rural African population, an ART research cohort was established from a proportion of those individuals attending the ART clinic at Chilumba Rural Hospital (McGrath et al. 2007). The study aimed to study the population level impact of ART along with monitoring the emergence and transmission of drug resistance among the population. The cohort of individuals attending the ART clinic was followed up by KPS every month. Blood samples from HIV-1 infected individuals who were taking ARV were collected and stored every 3 months. There was a need to establish the baseline genotyping of HIV-1 infected individuals (both treatment naïve and those who started the treatment) in Karonga district in order to monitor the presence and transmission of drug resistance. Furthermore, no information on the antiretroviral drug resistance in Malawi was available, until recently (Petch et al. 2005). In this project, along with baseline genotyping, the emergence of drug resistance in the individuals who are on ART is also monitored.

1.10: Research Aim and Objectives

Full ART became available in Karonga district in 2005 resulting in improvement of quality and length of life for majority of ART recipients. However, this benefit can be counter-balanced by survival, prolonging the time available for transmission and increase in risky behavior leading to increase in the number of sexually active HIV infected adults. If drug resistance develops in treated individuals and may thus lead to the transmission of drug resistant viruses. Hence there is a need to monitor the baseline genotyping of HIV-1 infected individuals in Karonga district. Baseline genotyping of treatment naïve individuals and those who are beginning treatment will be monitored by the current ART study. This will provide information of primary drug resistance in the Karonga population. The primary drug resistance in HIV-1 infected individuals can be used to estimate the transmission from ART-treated individuals. Proviral DNA from approximately 1000 blood samples will be extracted from individuals from HSS. Proviral DNA will also be extracted from individuals on ART at baseline and every three months after the start of ART. The HIV-1 RT gene will be amplified using PCR and will be sequenced by conventional bulk sequencing in order to study drug resistance. As protease inhibitors were not part of the current National ART delivery program and second line treatment, protease genotyping will be carried out on a small sample (50 individuals chosen at random). Therefore the overall aim of my work was to monitor the emergence and transmission of drug resistance in HIV-1 in Karonga and to study the diversity of HIV-1 in Karonga population by using the sequences generated in this study.

The specific objectives of my work were:

1) To investigate the presence of polymorphisms associated with drug resistance in reverse transcriptase and protease gene regions in drug naïve HIV-1 infected individuals in Karonga to establish the baseline genotype in this population.
2) To explore the intra-patient diversity of HIV-1 in these gene regions at a time when most individuals seen are drug naïve and transmission of drug resistance from ARV treated individuals is uncommon.
3) To investigate the presence and emergence of drug resistance mutations in HIV-1 infected individuals who were enrolled in the ART study by using a traditional consensus sequence genotyping approach (bulk sequencing).
4) To optimize the Ultra Deep Pyrosequencing (UDPS) approach for detecting

minor variants in RT, by a) optimising the PCR protocol to maximize the diversity of the input templates for UDPS, and b) to study the effect of the PCR approach on the number and diversity of the sequences returned from UDPS.

5) To use UDPS:
    a. To investigate drug resistance mutations in longitudinally collected blood samples from patients on ART.
    b. To gain deeper knowledge of the discrepancy in the presence and DRMs at various timepoints found by conventional bulk sequencing.
    c. To identify DRMs against the drugs other than first line therapy in Malawi.
6) By using sequences generated during this study
    a. To study the diversity of HIV-1 Karonga population by using the sequences generated in this study
    b. To study the selection pressure operating on RT gene in Karonga population.

Chapter 2

Drug Resistance Mutations in Drug-Naive HIV Type 1 Subtype C-Infected Individuals from Rural Malawi

Contents of this chapter were published as

Bansode, V., Z. J. Drebert, S. A. Travers, E. Banda, A. Molesworth, A. Crampin, B. Ngwira, N. French, J. R. Glynn, and G. P. McCormack. **2011**. Drug resistance mutations in drug-naive HIV type 1 subtype C-infected individuals from rural Malawi. AIDS Res Hum Retroviruses **27**:439-444.

2.1 Introduction

Malawi has an HIV-1 prevalence of around 11 % (UNAIDS 2009b). Although the prevalence is higher in urban areas, 80% of the population lives in rural areas, and the epidemic in these areas remains a concern and access to treatment more problematic. Over 250,000 people in Malawi have been on antiretroviral therapy (ART) and approximately 147,000 are currently on treatment (Health 2007). Around 95% of them are on first-line therapy, 4% are on alternative first-line therapy, and less than 1% are on second-line therapy (Health 2007). A fundamental concern of antiretroviral drug delivery programs is the avoidance of the development of widespread drug resistance. Drug resistance mutations (DRMs) have been primarily characterized for subtype B, which is predominant in America, Europe, and Australia where ART has been available for a decade and half, and where approximately 10% of new HIV infections are with drug-resistant isolates (Hemelaar *et al*. 2006). Fewer studies have characterized drug resistance in subtype C, which is responsible for more than 50% of worldwide infections. However, minor mutations associated with drug resistance have been reported in HIV-1 subtype C-infected drug-naive patients in a number of African locations including Malawi, and the percentage of drug-naive individuals harboring HIV with DRMs thus far has been shown to vary between 2% and 10% (Handema *et al*. 2003; Petch *et al*. 2005; Bessong *et al*. 2006; Kassu *et al*. 2007). Further characterization of drug resistance mutations in subtype-C HIV-1 and monitoring for transmission of drug- resistant strains are extremely important both given the more recent wide-scale availability of ART in countries in which subtype C is predominant, and as the prevalence of non-B HIV-1 increases in countries in which antiretroviral drugs are widely used.

Karonga District is the northernmost district of Malawi and HIV-1 subtype C has been reported to be the most predominant subtype in this District (McCormack *et al*. 2002). ART drugs were introduced in Malawi first in the context of prevention of maternal-to-child transmission (PMTCT) through antenatal clinics (using a mother and child nevirapine regime, which is ongoing). Full ART first became available at public clinics in the two major cities in Malawi in 2003, in the regional capital in 2004, and at Karonga district hospital in June 2005, at which time Karonga residents who had been seeking care outside the district were transferred to local services. ART

delivery has been largely based on clinical assessment with CD4 counting becoming available in January 2008 at one clinic. Viral load assessment and routine drug resistance testing are not available and the baseline genotype of the genes, which are the targets for ART, are unknown. The objective of this study was to investigate the presence of polymorphisms associated with drug resistance found in the reverse transcriptase (RT) and protease (PR) regions and to explore the intra-patient diversity of HIV-1 in these gene regions at a time when most individuals seen are drug naive and transmission of drug resistance from ART- treated individuals should be uncommon.

2.2 Materials and Methods

DNA was extracted from blood samples from 71 individuals living in Karonga District, Malawi using the QIAamp DNA Blood Mini Kit (QIAGEN Ltd). Extracted DNA was subjected to nested polymerase chain reaction (PCR) amplification of the HIV-1 PR and RT genes employing the primers previously described (Handema *et al*. 2003). The PCR mixture contained 1.5 mM $MgCl_2$, 1.5 units of *Taq* (Roche Expand High Fidelity PCR system), 0.8 mM each dNTP, and 5 pM primers (Eurofins MWG Operon). The cycling conditions were as follows: ''hot-start'' at 98^0C for 1 min, primary denaturation at 94^0C for 2 min followed by 35 cycles with 1 min denaturation at 94^0C, 1 min annealing at 42^0C, 4 min extension at 72^0C, and final extension for 7 min at 72^0C. For each gene region from 10 samples two PCR amplifications were pooled and the mixture cloned using the TOPO Cloning Kit (Invitrogen). Colony PCR and sequencing were carried out on 30 colonies per sample. Sequences obtained were edited in SeqMan (Lasergene, DNASTAR, Inc) and aligned using MacClade 4.0 (Sinauer Associates).

Preliminary phylogenetic analyses were first carried out to identify the subtype status of all sequences generated. Subsequently all consensus and clonal sequences that were subtype C were assembled into multiple alignments for each gene region using MacClade 4.0. Phylogenetic trees were reconstructed using the LANL (www.hiv.lanl.gov/content/index) subtype C ancestral sequence as outgroup under the GTR + gamma model of DNA substitution implemented RAxML 7.0.3 (Stamatakis 2006) with all parameters optimized by RAxML. Confidence levels in the groupings in the phylogeny were assessed using 1000 bootstrap replicates as part of the RAxML phylogeny reconstruction. To evaluate the effect of DRMs on clustering of individual sequences, analyses were also carried out with all sites showing the presence of DRMs removed. Sequences were submitted for analysis of DRMs to the Stanford Database (http://hivdb.stanford.edu).

Permission for this study was received from the National Health Sciences Research Committee, Malawi and the Ethics Committee of the London School of Hygiene and Tropical Medicine, UK.

2.3 Results

Of the 71 DNA samples, amplification and sequencing were successful from 62 for RT and from 53 for PR. Three RT and six PR sequences were excluded from analysis because of poor quality and four individuals were infected with non-subtype C virus (no DRMs were present in these non-C sequences). For both gene fragments from subtype C strains phylogenetic analyses with and without DRMs present showed identical tree topologies indicating that sharing the same DRM was not, by itself, responsible for individual sequences clustering together. Figure 2.1 shows the relationships between sequences with DRMs included.

2.3.1 Reverse Transcriptase: Reverse transcriptase sequences included sequences from 40 individuals who were drug naive and from 17 who had been on ART. None of the drug-exposed individuals showed any DRMs, however, five (5/40) drug-naïve individuals showed nucleoside reverse transcriptase inhibitor (NRTI) resistance mutations and four (4/40) showed non-NRTI (NNRTI) resistance mutations with both types of drug resistance mutations found in one individual (Table 2.1). The NRTI mutations found were V75LV (one individual) and V118I (four individuals) with five NNRTI mutations identified (A98G, K101Q, V106LV, E138A, and G190R) (Table 2.1).

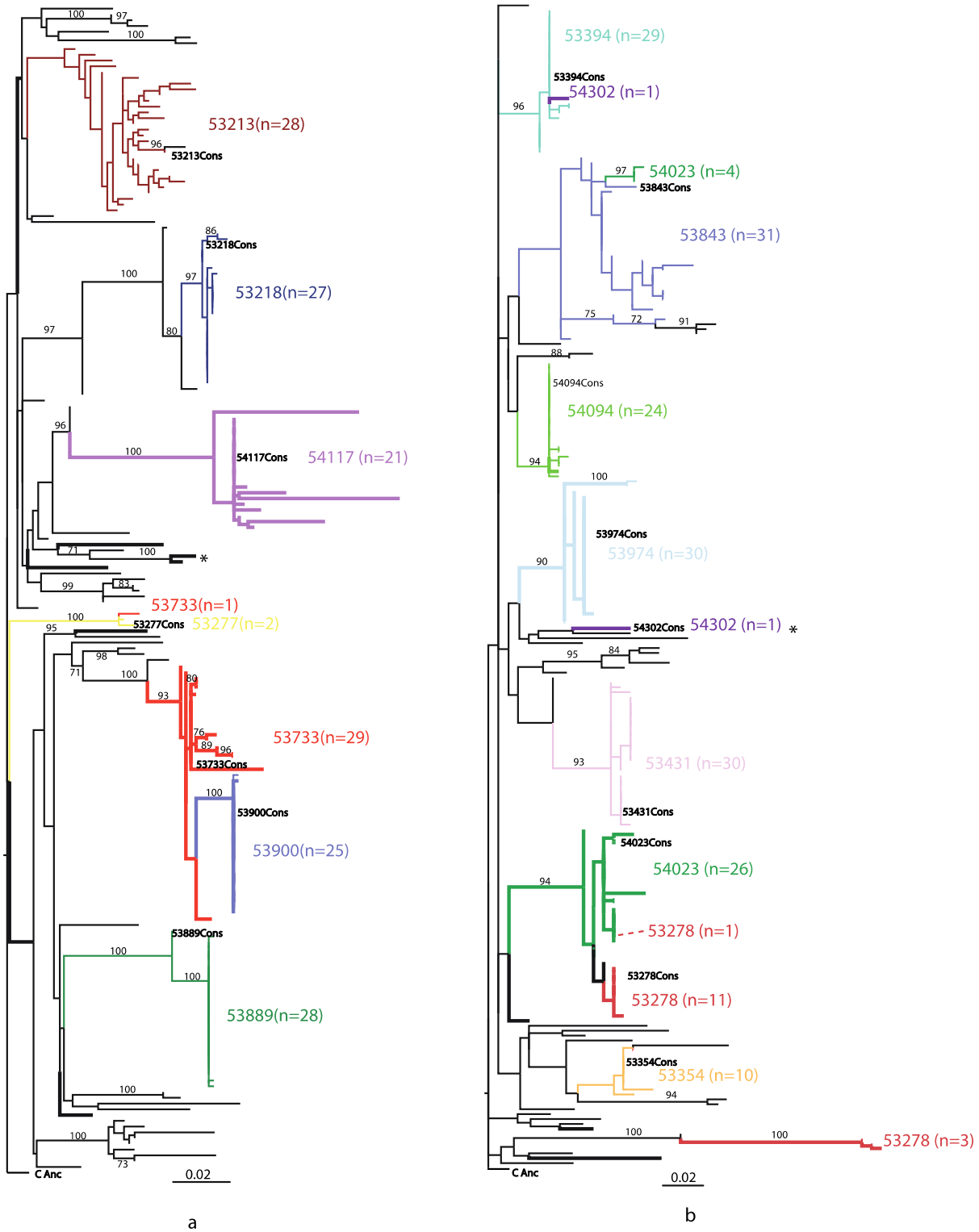


Figure 2.1: Maximum likelihood trees generated from reverse transcriptase (a) and protease (b) genes from HIV-1-positive individuals from Malawi. Branches in color represent sequences from multiple clones of different individuals. Branches in bold are those showing drug resistance mutations. The numbers on the branches are bootstrap values. *(a) Two consensus sequences showing the presence of drug resistance mutations grouping together with high support. *(b) Consensus sequences from a husband and wife pair that group together with high bootstrap support.

Table 2.1: Mutations associated with antiretroviral drug resistance found in consensus sequences from HIV positive individuals from Karonga District Malawi. Mutations involved in resistance to NRTI and NNRTI drugs and minor mutations involved in resistance to protease inhibitors are shown. The first four individuals in the table (shown in bold) were previously drug exposed prior to enrolling in the study.

	Reverse Transcriptase		Protease
Sample No	**NRTI**	**NNRTI**	
Drug exposed			
54302	**No sequence**		**T74S**
53972	**No sequence**		**L76S**
53278	**None**	**None**	**T74S**
53974	**None**	**None**	**L10V, T74S**
Drug naive			
54208	No sequence		T74S
53394	V118I	None	None
53436	V118I	None	No sequence
53733	V118I	None	No sequence
53900	V118I	None	None
54434	V75LV	V106LV	None
53977	None	None	T74S
54023	None	K101Q, E138A	T74S
54117	None	G190R	None
54357	None	A98G	No sequence

Cloning of reverse transcriptase was successful for seven samples (Table 2.2). From individuals whose consensus sequence showed neither NRTI nor NNRTI drug resistance mutations, none of the clonal sequences generated showed drug resistance mutations (Table 2.2). All cloned sequences obtained from individual 53733 contained the NRTI V118I mutation that was found in the consensus sequence and one clonal sequence showed an additional L74K NRTI-resistant mutation. The consensus sequence retrieved from individual 54117 showed a G190R (NNRTI) mutation and also had four stop codons at positions 71, 88, 212, and 239 (according to HXB2 RT gene). All clones obtained from this individual showed the same four stop codons and the same G190R mutation (Table 2.2). The phylogenetic relationships of the RT sequences from most individuals were straightforward with both the clonal and consensus sequences grouping together in monophyletic groups, with high bootstrap support (Fig. 2.1a). The clones and the consensus sequence from 53733 appear to be ancestral to the virus infecting individual 53900 with both of these individuals infected with viral strains containing the same NRTI DRM (V118I).

Table 2.2. Summary of amino acid diversity in multiple sequences retrieved from individuals showing evidence of NRTI and NNRTI resistance mutations in the original consensus sequence and from those that did not. The first four individuals in the table (shown in bold) were previously drug exposed prior to enrolling in the study.

* Consensus sequence was different from each of the clones.

** All clones and consensus contained four stop codons.

Sample No.	Mutation in Original Seq		No of cloned Sequences obtained	No. Of New Amino Acid Variants (n)	Variants (V_n) [number of clones in that variant]	NRTI mutations found in clones	NNRTI mutations found in clones
	NRTI	NNRTI					
Drug Exposed							
53277	**None**	**None**	**1**	**1**	**V_1[1]**	**None**	**None**
53213	**None**	**None**	**27**	**27**	**$V_{1\text{-}27}$ [All are different]**	**None**	**None**
53218	**None**	**None**	**26**	**2**	**V_1[1]**	**None**	**None**
					V_2[1]	**None**	**None**
53889*	**None**	**None**	**27**	**1**	**V_1[27]**	**None**	**None**
Drug Naive							
53733	V118I	None	20	2	V_1 [1]	V118I	None
					V_2 [1]	L74K,V118I	None
53900	V118I	None	24	None	None	V118I	None
54117**	None	G190R	13	10	$V_{1\text{-}10}$ [All 10 are different]	None	G190R

2.3.2 Protease:

Of 45 subtype C protease sequences, 16 were from drug-exposed individuals and 29 were from drug-naive individuals. Seven consensus sequences showed minor PI mutations including three from drug-naive individuals (Table 2.1). Cloning was successful for eight samples, three for which the consensus had shown minor resistance sites and five for which the consensus showed no drug resistance mutations. Three drug-exposed individuals (53278, 54302, and 53974) showed minor PR resistance mutations in clonal sequences that were not seen in the consensus sequences, whereas 53278 also showed evidence of major PR resistance mutations in the sequenced clones (Table 2.3). Furthermore, this pattern was repeated among the drug-naive individuals with major and minor DRMs present in clones that were not present in the consensus sequence (Table 2.3). Clones of 53278 showed major mutations D30N, M46I, and I84T.

For most individuals phylogenetic relationships between the sequences were uncomplicated; however, while 10 clones and the consensus sequence from individual 53278 clustered closely and all showed the T74S minor PI mutation, sequences of three variants clustered together in the tree with sequences retrieved from a different individual with whom she has no known epidemiological links (Fig. 2.1b) rather than with other 53278 sequences. These three sequence variants were those that showed major PR inhibitor resistance mutations. Furthermore, None of these three variants showed the original T74S mutation but did show another minor PI mutation (G48R). One additional clone sequence was found clustering with clonal sequences from another female, 54023, again with whom she had no known link.

Table 2.3: Summary of amino acid diversity in multiple sequences retrieved from individuals showing evidence of Protease inhibitor resistance mutations and from those that didn't. Drug exposed individuals are highlighted in bold.

Sample No.	Mutation in Original Seq		No of cloned Sequences obtained	No of new amino acid variants (n)	Variants (V_n) [number of clones in that variant]	PI resistance major mutations found in clones	PI resistance minor mutations found in clones
	Major	Minor					
Drug Exposed							
53278	**None**	**T74S**	**14**	**4**	**V_1 [1]**	**D30N,M46I**	**G48R**
					V_2 [1]	**D30N,M46I**	**G48R**
					V_3 [1]	**D30N,M46I,I84T**	**G48R**
					V_4 [1]	**None**	**T74S**
54302	**None**	**T74S**	**1**	**1**	**V_1[1]**	**None**	**L10F**
53354	**None**	**None**	**9**	**2**	**V_1 [1]**	**None**	**None**
53974	**None**	**L10V, T74S**	**29**	**4**	**V_1[1]**	**None**	**L10V, T74S**
					V_2[1]	**None**	**L10V,G48R,T74S**
					V_3[1]	**None**	**None**
					V_4[1]	**None**	**None**
					V_2 [1]	**None**	**None**
Drug Naive							
54023	None	T74S	29	3	V_1 [1]	None	M46T,T74S
					V_2[1]	None	T74S
					V_3[4]	None	None
53394	None	None	28	1	V_1[9]	None	None
53431	None	None	30	1	V_1[15]	None	None
53843	None	None	30	5	V_1 [1]	None	None
					V_2[1]	None	None
					V_3[1]	None	None
					V_3[7]	None	None
					V_4[14]	None	None
54094	None	None	23	2	V_1[1]	I47M	None
					V_2[1]	None	None

2.4 Discussion

In this study, we showed that in 2008, 3 years after antiretroviral therapy was introduced into the Karonga district, Malawi, a greater than expected number of drug-naïve individuals have been infected with HIV-1 subtype C virus harboring major and minor DRMs. From a sample size of 40 RT sequences from drug naïve individuals, five individuals showed NRTI and four showed NNRTI mutations with one individual showing both. From 29 PR sequences, again from drug naïve individuals, we found evidence of minor DRMs in three. Additional major and minor DRMs were found in clonal sequences from a number of individuals that were not present in the original consensus sequences. This study illustrates the importance of sequencing multiple HIV-1 variants from individuals to fully assess drug resistance.

In reverse transcriptase, both NRTI and NNRTI resistance mutations were observed in drug naïve individuals. V118I occurs in 2% of untreated persons infected with subtype C and with increased frequency in persons receiving multiple NRTIs (Rhee *et al*. 2003). It causes low-level resistance to 3TC and possibly to other NRTIs when present with other mutations (Rhee *et al*. 2006). This was the only mutation found in both previous studies of drug resistance mutations in Malawi, being present in 4 of 21 individuals studied by (Petch *et al*. 2005) and 15 of 96 individuals with confirmed virologic failure by (Hosseinipour *et al*. 2009). Of the other drug resistance mutations found, K101Q minimally reduces susceptibility to each of the NNRTIs, whereas mutation at position 75 (V75T/M/A/I) is associated with reduced NRTI susceptibility (http://hivdb.stanford.edu). According to the Stanford database G190R is a highly unusual mutation and is flagged as an NNRTI resistance mutation. Little seems to be known about the effects of this mutation as yet in subtype C but G190A/S/E/Q/T/V/C are NNRTI resistance mutations. Similarly, although V106A causes high-level resistance to NVP and DLV and low to intermediate resistance to EFV, V106L is a rare polymorphism and its association with NNRTI resistance is less clear; E138A is responsible for decreased ETR response and A98G reduces NVP susceptibility by two- to threefold (http://hivdb.stanford.edu).

In protease, all DRMs found in consensus sequencing were minor DRMs as per Stanford database. However, some major mutations were found in clones generated from protease gene. D30N causes high-level resistance to NFV and potential low-level resistance to ATV. M46I decreases susceptibility to IDV, NFV, FPV, LPV, and

ATV when present with other mutations. I84V causes intermediate to high-level resistance to various PIs. Minor mutations at position 10 (L10I/V/F/R/Y) are associated with resistance to most PIs when present with other mutations and mutation at this site occurs in 5–10% of untreated persons including subtype C-infected individuals in Africa (Bessong *et al*. 2006; Kassu *et al*. 2007; Vidal *et al*. 2007; Hosseinipour *et al*. 2009). Mutation at position 74 (T74S) is associated with reduced NFV susceptibility whereas mutation at position 76 (L76V) reduces susceptibility to FPV, IDV, LPV, and DRV and increases susceptibility to SQV, ATV, and TPV (Vidal *et al*. 2007; Somi *et al*. 2008).

This study utilized proviral DNA to generate RT and PR sequences due to the difficulty of working with RNA in this context (i.e., getting samples from a large population study in rural Africa). HIV-1 with a G–A hypermutation at positions 88 and 212 has previously been found in resting CD4 cells and the resulting DNA is reported to be degraded or defective (Goff 2003; Kieffer *et al*. 2005). It is possible that the plasma of this individual would yield functional HIV- 1. It is also possible that the individual may be infected with a defective virus and that this might affect disease progression. There is no known epidemiological link between these individuals. The laboratory work on these samples was carried out by different people and therefore contamination is not likely in this case. Of the individuals for whom the PCR products were not cloned there is only one instance of clustering of sequences containing DRMs with high bootstrap support (see the asterisk in Fig. 2.1a). One of these individuals is a male and the other is a female. These individuals are in the same reporting group and thus may live geographically close together.

This work clearly shows the presence of both NRTI and NNRTI drug resistance mutations and the presence of minor and major PR resistance mutations in a number of drug-naive individuals in rural Malawi in 2008, 3 years after ART became widely available in the district. We have also shown the importance of sequencing multiple HIV-1 variants from individuals to fully assess drug resistance, as a number of individuals have shown major and/or minor drug resistance mutations in sequenced clones that were not present in the consensus sequences. Sequencing multiple clones of RT and PR from 16 individuals showed that one of them (female, aged 52, 53278 in Fig. 2.1b) was dual infected and showed evidence of being infected with viruses containing DRMs. The dual infection was confirmed by

sequencing of multiple clones of the gag gene as part of another project, which showed a similar pattern of multiple lineages for this individual (Seager, unpublished). In terms of identifying possible transmission events involving HIV-1 with DRMs, apart from the dually infected individual mentioned above, there are very few cases of clustering of sequences from individuals with DRMS. Our analyses showed that the presence of DRMs alone was not responsible for the pattern of relationships shown on the trees indicating common ancestry rather than convergence. However, given the short length of gene fragments used to explore drug resistance and in many cases having sequence evidence from only one gene, such results are not reliable reports of transmission and sequencing of additional loci is underway to confirm transmission in each case. The presence of drug-resistant mutations in drug-naive individuals may prevent successful treatment of certain individuals. This knowledge may also inform policymakers in planning future therapy strategies as this indicates that additional therapy combinations may be required. Although antiretroviral drugs were made more widely available in Karonga District in June 2005, prior to this NVP was available for the PMTCT in antenatal clinics and full ART would have been available to a number of individuals elsewhere in Malawi in 2002. Thus, it is also possible that some individuals assumed to be drug naïve did not disclose (or in the case of PMTCT were not aware of) ART exposure, and did not appear in the ART study cohorts. Furthermore, the presence of DRMs in drug-naive persons may represent natural polymorphisms and may not be suggestive of transmission of DRMs. The impact of such natural polymorphisms on the development of drug resistance in those people at commencement of ART is unknown and such individuals need to be monitored. Indeed the impact of such polymorphisms on the HIV-1 subtype C-infected population and the speed of more widespread drug resistance are also unknown. Therefore it may be important to further examine some of the individuals and mutations found during this study in view of a long-term treatment strategy in the district.

Chapter 3

Reverse transcriptase drug resistance mutations in HIV-1 subtype C infected patients on ART in Karonga District, Malawi

Contents of this chapter were published as

Bansode, V. B., S. A. Travers, A. C. Crampin, B. Ngwira, N. French, J. R. Glynn, and G. P. McCormack. 2011. Reverse transcriptase drug resistance mutations in HIV-1 subtype C infected patients on ART in Karonga District, Malawi. AIDS Res Ther. **8**:38.

3.1 Introduction:

It has been estimated that in sub-Saharan Africa, approximately 3.9 million people have started antiretroviral treatment (ART) since its introduction (UNAIDS, 2010). Given the large population on treatment, viral diversity coupled with low adherence could lead to the emergence and large-scale transmission of drug resistant strains. Rates of drug resistance among patients who received ART in sub-Saharan Africa range from 3.7%–49% after 24–163 weeks of HAART (Hamers et al. 2008). Various factors contribute to this large range in resistance among African cohorts such as variation in available healthcare systems and practices, adherence, and access to monitoring (Shafer, Rhee, and Bennett 2008). Development of DRMs to Trioimmune®, the drug combination used as first line therapy in Karonga District, Malawi, has been reported in Zambia (Gupta et al. 2010), South Africa (Hoffmann et al. 2009), Cameroon (Burda et al. 2010), Kenya (Steegen et al. 2009) and Uganda (Weidle et al. 2003). Previous studies on drug resistance in Malawi showed various DRMs to both NRTIs and NNRTIs in both drug naïve individuals (Bansode et al. 2011a) and those failing therapy (Hosseinipour et al. 2009). However, very little data are yet available on the emergence of drug resistance to ongoing treatment and the transmission of drug resistant variants in subtype C infected countries (Weidle et al. 2003; Steegen et al. 2009; Burda et al. 2010; Gupta et al. 2010; Towler et al. 2010).
The Malawi antiretroviral treatment (ART) program started in 2004, and between then and the end of June 2010 over 225,000 patients had initiated first-line antiretroviral therapy (ART) through 396 ART clinics (Ministry of Health 2010). As part of the Karonga Prevention Study (KPS), investigating how the availability and use of ART may change the HIV epidemic and its socio-demographic impact in the rural Karonga District, (northern Malawi), an ART research cohort was established from those attending the ART clinic at Chilumba Rural Hospital. HIV-1 subtype C is the predominant subtype in this District (McCormack et al. 2002). The objective of this overall study was to investigate the success of the current ART delivery programme in a rural population, and, as a component of this, to investigate the evolution of drug resistance using a traditional consensus sequence genotyping approach.

3.2 Materials and Methods:

3.2.1 Study Participants and Treatment schedules:

At the Ministry of Health ART clinic at Chilumba Rural Hospital all those attending for screening for ART suitability and who are resident in a geographically defined area adjacent to the clinic, are invited to take part in an observational cohort study. Every three months participants are clinically assessed by KPS research staff. Blood samples are collected at the their first visit (baseline) and at every follow-up visit. A CD4 count is performed at baseline, 6, 12 and 24 months or at the time of a clinical failure, defined by a new WHO stage 3 or 4 event after six months of therapy (WHO, 2006). First line therapy is a generic fixed-dose combination treatment (Triommune®), which consists of: stavudine (d4T), lamivudine (3TC), and nevirapine (NVP). All individuals were on first line therapy only.

DNA Extraction, PCR and Sequencing:

Whole blood samples were collected in 4.5 ml vacutainer tubes. Samples were centrifuged and plasma and cell pellet were stored separately at -70^{0}C. DNA was extracted from whole blood cell pellet samples using the QIAamp DNA Blood Mini Kit (QIAGEN ltd). Extracted DNA was subjected to nested PCR amplification of the HIV-1 reverse transcriptase as described in Bansode et al (Bansode et al. 2010). All PCR amplicons were gel purified and automatically sequenced.

3.2.2 Sequence Analyses:

Sequence chromatographs were edited in SeqMan (DNASTAR, Inc) and all sites that showed ambiguities (two or more peaks of equal, or almost equal, height) were noted. Multiple alignments were assembled of all subtype C sequences generated with the 57 reverse transcriptase sequences generated from (Bansode et al. 2010) using MacClade 4.0 (Sinauer Assoc). Sequences were submitted for analysis of DRMs to the Stanford Database (Rhee et al. 2003). To check for transmission of DRMs, phylogenetic trees were reconstructed using the LANL subtype C ancestral sequence as outgroup under the GTR + gamma model of DNA substitution implemented RAxML7.0.3 (Stamatakis 2006) with all parameters optimised by RAxML. Confidence levels in the groupings in the phylogeny were assessed using 1000 bootstrap replicates as part of the RAxML phylogeny reconstruction.

Permission for this study was received from the National Health Sciences Research Committee, Malawi and the Ethics Committee of the London School of Hygiene and Tropical Medicine, UK.

3.3 Results

One hundred and forty nine subtype C sequences were generated from 75 individuals, 65 of which were from blood samples collected at baseline (and reported to be ART naïve). DRMs were found in sequences from 15 individuals (20%) overall, and for 10 individuals (15.4%) the mutations were found in sequences from baseline samples (drug naïve). Details of observed drug resistance mutations are summarized in Table 3.1. Seven individuals showed DRMs (or ambiguities that suggest the presence of DRMs) to NRTIs used in Karonga with 6/7 showing the mutation V118I. While ten individuals showed the presence of DRMs to NNRTIs only five showed DRMs against therapies used in Karonga, the most common being Y181C and G190AE.

Some individuals showed a discrepancy in the presence and type of DRMs over time. Three patients (Pt2, Pt12 and Pt66) did not show any DRMs at baseline but showed DRMs at subsequent time points (Table 3.1). Patient 2 also showed a significant drug resistance-related ambiguity (K103KN) in the consensus sequence at 8 months while a different DRM was seen at 11 months (Y181C). No DRM was seen in the sequence from the 14-month sample. Patient 12 showed a similar pattern, where an NNRTI associated mutation (Y181NY) was present in the sequence collected at 5 months, while the sequences at baseline and 7 months did not show any DRMs (Table 3.1). Both individuals (patient 2 and patient 12) showed immune failure (their CD4 count did not rise over 200 cells/mm^3 after 12 months on ART). Three patients (Pt32, Pt61 and Pt76) showed DRMs at baseline but different DRMs at later time-points (Table 3.1) with patient 32 showing a high variation of DRMs across timepoints. The baseline sequence from patient 61 showed V118I and K219R, the latter of which was not found in the sequence at 9 months. In Patient 76, the baseline sequence showed the ambiguity Y181CY, with two additional NNRTI mutations (V90IV and H221HY), the 6 months sequence showed the full DRM at position 181 while sequences retrieved from 3 month and 9 month samples showed no DRMs (Table 3.1).

Table 3.1: Mutations associated with antiretroviral drug resistance found in sequences from HIV-1 subtype C infected individuals from Karonga District Malawi. Mutations in bold are against current ART drugs in use in Karonga District.
* Patients showing discrepancies in DRMs between different timepoints

Patient	Comments	Sex	Time point (month)	NRTI	NNRTI
Patient 2 *	Immune Failure	F	0	No DRMs	No DRMs
			8	No DRMs	**K103KN**
			11	No DRMs	**Y181C**
			14	No DRMs	No DRMs
Patient 5		F	0	No DRMs	E138A
			13	No DRMs	E138A
Patient 12 *	Immune Failure	F	0	No DRMs	No DRMs
			5	No DRMs	**Y181NY**
			7	No DRMs	E138A
Patient 14		M	12	No DRMs	E138A
Patient 20	Immune Failure	F	0	No DRMs	E138A
			3	No DRMs	E138A
			6	No DRMs	E138A
			9	No DRMs	E138A
Patient 32 *		M	0	No DRMs	V90I
			7	**M41MR, T215ST**	No DRMs
			14	No DRMs	No DRMs
			16	No DRMs	V108AV
			24	No DRMs	No DRMs
Patient 42 *		M	0	No DRMs	No DRMs
			0	**V118IV**	No DRMs
			5	**V118I**	No DRMs
Patient 45		F	0	No DRMs	V106I, E138A, **G190A**
			8	No DRMs	V106I, E138A, **G190A**
Patient 61 *		F	0	**V118I**, K219R	No DRMs
			9	**V118I**	No DRMs
Patient 66		M	0	No DRMs	No DRMs
			9	**V118IV**	No DRMs
Patient 76 *		M	0	No DRMs	V90IV, **Y181CY**, H221HY
			3	No DRMs	No DRMs
			6	No DRMs	**Y181C**
			9	No DRMs	No DRMs
Patient 77		F	0	**V118I**	No DRMs
			12	**V118I**	No DRMs
Patient 91			0	**V118I**	No DRMs
Patient 93		M	7	**V118I**	E138A
			10	**V118I**	E138A
Patient 95			0	No DRMs	E138A

There was no evidence of transmission of drug resistant HIV between the individuals examined here. Sequences retrieved from each individual grouped monophyletically in all cases. Few individuals showed their sequences clustering with other patients with high bootstrap support but DRMs were not present in both individuals, e.g. sequences from patient 47 and 77 formed a cluster together and are from the same geographical area but while patient 77 showed DRMs, patient 47 did not (data not shown).

3.4 Discussion:

Through genotyping RT from HIV-1 subtype C infected individuals on ART using a consensus sequencing approach, we have shown the presence of mutations associated with drug resistance to the therapy used in Karonga District. Drug resistance to Trioimmune® occurred at an overall rate of 20% of individuals (both drug naïve and drug exposed), which is comparable to rates found in other African countries (Weidle *et al*. 2003; Laurent *et al*. 2005; Seyler *et al*. 2007; Hoffmann *et al*. 2009; Johannessen *et al*. 2009; Steegen *et al*. 2009; Burda *et al*. 2010) but, as expected, greater than that described in our previous study (7.5%) (Bansode et al. 2011a) which did not include individuals currently on therapy.

Patients 2 and 12, both females, had immune failure prior to ART initiation and continued to exhibit immune failure while on ART (i.e. their CD4 counts did not rise above 200 cells/mm^3 after 12 months on therapy). While neither showed DRMs from baseline samples they subsequently showed the DRMs Y181C and Y181NY respectively, which is responsible for high-level resistance to NVP, the NNRTI used in 1^{st} line therapy in Karonga. Patient 2 also showed a DRM (K103KN), after 6^{th} months of ART, which also causes high-level resistance to NVP. For these, and a third individual who also exhibited immune failure, it will be important to monitor the individuals and DRMs at subsequent timepoints in case of continued immune failure and development of clinical failure.

Three drug-naïve individuals (Pt 61, Pt 77, Pt 91) showed V118I while another (Pt 42) showed an ambiguity at this position (V118IV). According to the Stanford HIV drug resistance database, V118I is responsible for low-level resistance to 3TC and possibly to other NRTIs when present with other mutations. The mutation has been reported to occur in ~2% of untreated persons infected with subtype C and with increased frequency in persons receiving multiple NRTIs (Pillay *et al*. 2002) and so it may not be unexpected to find it in this cohort. It was the only DRM found in all three previous studies of drug resistance in Malawi (Petch et al. 2005; Hosseinipour et al. 2009; Bansode et al. 2010) and was also reported in subtype C infected drug naïve patients from Zambia (Handema *et al*. 2003), Zimbabwe (Kantor *et al*. 2002) and South Africa (Pillay *et al*. 2002). However, considering the fact that this mutation is naturally present in untreated individuals infected with subtype C, presence of this

mutation does not seem to affect drug resistance in this cohort. It has been suggested that along with drug resistance, the V118I mutation alone is a marker of advanced HIV infection and disease progression (Zaccarelli *et al*. 2007). As no associated mutations were found in the three individuals, and they all had a satisfactory response to treatment, this mutation is probably not significant but may become important if a second mutation were to arise.

Mutation G190A (shown in a female patient 45) according to the Stanford drug resistance database, causes high-level resistance to NVP and intermediate resistance to EFV. The mutation was present at baseline in this individual and could indicate acquisition of drug resistant HIV. However, although all individuals participating in the ART cohort study were reported to be ART naïve, we cannot exclude the possibility that some individuals had received some form of ART previously, (e.g. received prevention of mother to child transmission treatment) and did not disclose this fact. The DRM does not appear to have had any major effect on treatment to date, as this individual also has had a satisfactory response.

Drug resistance mutations were found to emerge in some individuals during ART. Patient 32 showed a number of NNRTI mutations and a number of ambiguities at sites important in susceptibility to NRTIs (e.g. the mutation T215S is one of many transitions between wild type and the mutations Y and F (Bennett *et al*. 2009)). Most of the ambiguities do not reduce NRTI susceptibility but their presence may suggest that the DRM may also be present (Shafer *et al*. 2007). This patient had made additional visits to the clinic outside of the routine ART cohort study because of diabetic complications. Additional sequences produced from samples taken at those additional visits showed further mutations associated with drug resistance to NVP and AZT (M41L, M184I, G190E- data not shown), however he has had a satisfactory response to treatment to date.

This study was based on a consensus sequencing approach from provirus due to the difficulty of amplifying HIV RNA from individuals on ART. While provirus may not provide as clear picture of the genotype of the circulating virus as would be retrieved from RNA in individuals who have been infected for long periods of time, it has been shown in patients with virological failure that archived resistance mutations previously detected in the proviral DNA were observed in the sequences obtained

from the plasma viruses at the time of virological failure (Boucher *et al*. 2005). When the selective pressure provided by ART compromises the current plasma population, archived viruses can re-emerge (Joos *et al*. 2008). Therefore documenting the proviral population is also important in monitoring the emergence of drug resistance. Despite the presence of DRMs to current therapy in some individuals, and immune failure in three, no signs of clinical failure were seen during this study. This cohort will continue to be monitored as part of the Karonga Prevention Study so that the long-term impact of these mutations can be assessed.

Chapter 4

Ultra-deep pyrosequencing and HIV-1 diversity

Contents of this chapter contributed to the following article

Vijay Bansode, Grace P McCormack, Amelia Crampin, Bagrey Ngwira Neil French, Judith R Glynn, Simon A A Travers. 2012. Characterizing the Emergence and Persistence of Drug Resistant Mutations in HIV-1 Subtype C Infections using 454 Ultra Deep Pyrosequencing. BMC infectious diseases. Submitted

4.1 Introduction

An accurate and sensitive detection of drug resistant HIV strains is important for the treatment of HIV infection. Antiretroviral therapy (ART) failure can occur if a person is infected with drug resistant HIV strains (Menendez-Arias 2010). This can be prevented by early detection of resistant HIV strains in an individual (Hoffmann et al. 2007). Various genotypic and phenotypic methods are commonly used to detect antiretroviral resistance in clinical settings. Phenotypic methods rely on cloning the Reverse Transcriptase (RT) and Protease (PR) coding regions into a standard HIV plasmid backbone which allows viral stock generation and functional analysis of viral drug sensitivity in short-term culture (Clavel and Hance 2004). The genotypic methods include conventional bulk sequencing of RT and PR, which report predominantly circulating viral variants (Hoffmann et al. 2007). The resistance profiles of HIV-1 to the commonly used drug classes (Protease inhibitors-PI, Nucleoside Reverse Transcriptase Inhibitors – NRTI and Non Nucleoside Reverse Transcriptase Inhibitors- NNRTI) are well characterized (Soriano and de Mendoza 2002; van Maarseveen and Boucher 2006). Such characterization allows predicting the probable resistance patterns in HIV-1 and helps in planning the current and future treatment strategies. However, conventional bulk sequencing methods have limitations in their inability to find the low abundance drug resistance mutations (with less than 20% prevalence) (Korn et al. 2003). An ideal method to detect low abundance drug resistance mutations would yield many sequences and allow analysis of many samples of HIV populations in a single experiment.

Very little genotype data were available, until very recently, for the targeted genes of first and second line antiretroviral therapies in Malawi (Petch et al. 2005; Hosseinipour et al. 2009; Bansode et al. 2011a; Bansode et al. 2011b). As described in chapter 2 and 3, two hundred and six consensus sequences were produced from the RT gene from subtype C infected individuals, 105 of which were from drug naïve individuals. Fifteen of these 106 drug naïve individuals (14%) showed DRMs while five (5%) drug exposed individuals showed DRMs (Bansode et al. 2011a; Bansode et al. 2011b). In chapter 2, it has also been shown that when multiple variants were sequenced from selected individuals, via molecular cloning, additional drug resistance mutations in protease and reverse transcriptase were present in drug

naïve individuals. This underlined the importance of cloning to reveal a better picture of viral variants within individuals, which maybe helpful to detect minor drug resistant variants (Bansode et al. 2011a). It was also found in chapter 3 that while some individuals appeared to develop DRMs over time on treatment, some individuals showed DRMs at baseline but no DRMs at subsequent timepoints (Chapter 3). This discrepancy further points to the limitations of bulk sequencing for determining drug resistance. Bulk sequencing is likely to produce a sequence only from the dominant circulating virus. However, it is also possible due to the Polymerase Chain Reaction (PCR) method that a consensus sequence produced in bulk sequencing is not from the dominant viral population but rather from a minor variant from the population. Either way only one virus is sequenced, thus limiting studies to consensus sequences is likely to miss the viral diversity, which influences fitness and drug resistance.

Resistant viruses that make up as little as 1% of viral population are clinically important as they can grow rapidly under the selective pressure introduced by drugs (Johnson et al. 2008; Paredes et al. 2010). Although consensus/bulk sequencing approaches have been a major advancement in the understanding and management of HIV drug resistance in clinical practice, a major limitation is the inability to detect low abundance drug resistant mutations at levels <20% of the viral quasispecies (Korn et al. 2003). Various methods are used to identify multiple drug resistant variants, such as sequencing of multiple clones, sequencing after single genome amplification (Palmer et al. 2005), allele-specific real time PCRs (Roquebert et al. 2006), and microarrays (Kozal et al. 1996). However, these methods can be labor intensive, expensive and time consuming.

A new approach, the use of pyrosequencing was used for detection of minor DRMs in the protease gene. This pyrosequencing approach was not enough as it was able to detect upto 100 nucleotides in length and could identify the minor variants present in 25% of the viral population. Later, based on pyrosequencing, various sequencing platforms were available commercially which were used obtain better coverage of the gene (Reviewed in (Metzker 2010; Barzon et al. 2011). An adapted version of the pyrosequencing approach was used to characterize rare DRMs in many samples in parallel (Hoffmann et al. 2007). This method used a DNA

bar coding system (Binladen et al. 2007) combined with pyrosequencing (Margulies et al. 2005) to produce 118,093 sequences from ~100 bp segments of the PR and RT coding regions from seven viral populations. This study identified minor drug resistant alleles in patient samples, which were of potential clinical significance, and demonstrated the feasibility of using pyrosequencing for efficient DRM analysis. Since then several other studies used this method employing the next generation sequencing (NGS) platform, the Genome Sequencer 20 (GS20) from 454 Life Sciences (Margulies et al. 2005). Since the release of GS20, several other NGS platforms have been introduced and, in addition, several updates have been released, which have increased the throughput and length. The major platforms used today include the 454 FLX and 454 Titanium platforms from Roche, the SOLID platform from Applied Biosystems and the Solexa platform from Illumina. The SOLID and Solexa approaches produce reads that are 50- 100 bp long; the FLX and Titanium approaches produce reads up to 500 bp (Mardis 2008) while newly released Roche FLX Plus can produce reads upto 1000 bp (Roche 2011). The advantage of amplicon sequencing on 454 platform is the capacity to sequence many individual template molecules to study minority sequence variants and haplotype composition with longer read lengths.

This ultra-deep pyrosequencing (UDPS) technology has wide range of applications in the field of genetics such as the study of microbial diversity within a heterogeneous population (Andersson et al. 2008), in the field of cancer biology to study cancer-associated genes in humans (De Grassi et al. 2010) or in the field of virology to study low frequency variants within a single host (Hoffmann et al. 2007; Margeridon-Thermet et al. 2009) or within a population of individuals (Ji et al. 2010). UDPS has also been used to study viral quasispecies in order to determine low abundance drug resistant viral variants in Hepatitis B Virus (Margeridon-Thermet et al. 2009). The UDPS technology is now being used in the study of HIV minority variants within population including drug resistance (Wang et al. 2007; Mitsuya et al. 2008; Le et al. 2009; Simen et al. 2009; Varghese et al. 2009; Hedskog et al. 2010; Ji et al. 2010; Delobel et al. 2011), co-receptor use (Archer et al. 2009; Rozera et al. 2009; Tsibris et al. 2009), co-evolution in the *nef* gene (Poon et al. 2009) and to reconstruct the date of HIV infection (Poon et al. 2011).

PCR is widely used in the UDPS approach. Template amplification by PCR increases the number of available molecules for sequencing but can introduce random errors, which can complicate HIV genotyping (Kozarewa et al. 2009). An additional problem in the amplification and detection of HIV proviral DNA is the presence of high amounts of human DNA in the sample where non-specific annealing of primers can result in increased background amplification leading to suppressed amplification of actual target sequence (Boni et al. 2004). While amplifying low copy number templates, the PCR error problem worsens because, if a mutation arises in an early PCR cycle it can lead to a high proportion of mutated reads. On the other hand, some variants can be missed if sequence reads are all derived from one allele compared to alternate alleles due to bias in PCR amplification. The initial PCR step is crucial, as it needs to amplify each genome variant present. It is important to maximize the amplification of initial templates. This can be attempted by performing multiple primary and secondary (nested) PCR amplifications, which are then combined to help avoid the PCR bottleneck . Up to 40 amplification cycles and a sufficient amount of the initial DNA template is required to amplify low abundance (less than 1%) genotype templates (Simen et al. 2009; Solmone et al. 2009; Tsibris et al. 2009). The number of DNA sequences generated by the first PCR defines the sensitivity of UDPS for low abundance amplifications and approximately 10,000 reads per amplicon are necessary to detect amplification products presenting 0.1-1% of the initial template (Eriksson et al. 2008).

The aim of this study was to use UDPS to study the discrepancies in presence and absence of DRMs found in bulk sequences (Chapter 3, (Bansode et al. 2011b) from patients on ART. As PCR is widely used in UDPS, the initial PCR reactions play crucial role in the detecting the viral diversity present within population. Various other studies on HIV drug resistance using UDPS used different approaches such as use of one (Hoffmann et al. 2007), three (Codoner et al. 2011) or five (Hedskog et al. 2010) initial PCR reactions on the viral template. We sought to increase heterogeneity by avoiding PCR bottleneck, by optimizing the methodology at PCR stage. The specific aims of the present study were,

1) To optimize the PCR protocol to maximize the diversity of the input templates for UDPS.
2) To study the effect of PCR on the number and diversity of the sequences returned from UDPS

4.2 Materials and Methods

4.2.1 Patients

Five individuals were selected on the basis of discrepancies in drug resistance mutation patterns from blood samples collected at different time points. The individuals and the patterns of DRMs in the bulk consensus sequences are described in chapter 3 and are summarized in Table 4.1. From these five individuals, fifteen samples were taken for further PCR and deep sequencing.

Table 4.1. Samples subjected to ultra-deep pyrosequencing. Consensus sequencing of reverse transcriptase from five patients had previously revealed discrepancies in the presence/absence of DRMs. Baseline describes the sampling timepoint prior to ART initiation with subsequent samples being described as the number of months post ART initiation. Mutations shown in the table are those observed in the consensus sequencing of each sampling timepoint. Samples selected for subsequent ultra-deep pyrosequencing are highlighted. The remaining samples were excluded due to low success in further multiple PCR reactions.

Patient	Month	NRTI	NNRTI
Patient 2	Baseline		
	8 Months		K103KN
	11 Months		Y181C
	14 Months		
Patient 32	Baseline		V90I
	7 Months	M41MR, T215ST	
	14 Months	M41I, M184I	
	16 Months		
	16 Months		G190AE
	19 Months		V108AV
	24 Months		
Patient 42	Baseline 1 (2008)		
	Baseline 2 (2009)	V118IV	
	5 Months	V118I	
Patient 45	Baseline		V106I, E138A, G190A
	8 Months		V106I, E138A, G190A
	13 Months		Sequence Failed
Patient 76	Baseline		V90I, Y181CY, H221HY
	3 Months		
	6 Months		Y181C
	9 Months		Sequence Failed

4.2.2 DNA extraction and PCR amplification for ultra deep sequencing:

Whole blood cell pellets, which had been stored at -70°C were re-suspended in 1 ml of sterile Phosphate Buffered Saline (PBS). In each case 45 µl of this volume was transferred to a new eppendorf and a further 155 µl of PBS added to bring the volume up to 200 µl. DNA was extracted from each sample using the QIAamp DNA Blood Mini Kit (QIAGEN ltd) as per manufacturers instructions.

First round PCR amplifications were carried out as described in chapter 2; for the second round of PCR, each primer was modified to include 5' extensions. These extensions included, directional GS FLX Titanium Primers A and B (which included a four base 'key' sequence) for forward and reverse directions respectively, followed by a 10-base Multiplex Identifier (MID) sequence (Figure 1).

Forward primer (Primer A-Key):

5'-CGTATCGCCTCCCTCGCGCCATCAG-{MID}-{template-specific-sequence}-3'

Reverse primer (Primer B-Key):

5'-CTATGCGCCTTGCCAGCCCGCTCAG-{MID}-{template-specific-sequence}-3'

Figure .1 The Forward and Reverse primers (with A and B Key) sequences used for 454 sequencing. The MID sequence and template specific sequences for each patient are listed in Appendix Table 1.

Reproduced from 454 Sequencing Technical Bulletin: 2009, No. 013-2009

The multiplex identifier assures the correct identification of each sample by automated software after pooling and sequencing. The primers were designed as per the guidelines of amplicon fusion primer design guidelines for GS FLX Titanium Series Lib-A chemistry (Roche Diagnostics) with the list of MIDs provided by LGC Genomics (Germany). The nucleotide sequences of both forward and reverse Titanium Fusion Primers with MID and forward and reverse PCR primers that were designed are listed in Appendix Table 1

The designed primers were ordered from Eurofins MWG Operon (Germany). The lyophilized primers obtained were rehydrated in nuclease free water to make a stock of 100 pm/µL. As the second round PCR primers were modified (extended length), affecting melting temperature, a temperature gradient reaction was carried out with an annealing temperature range from 55 to 70 °C. Amplification was optimal at 57.5 °C, which was chosen as the annealing temperature in subsequent PCRs.

Considerable effort was invested in optimization of the PCR amplification with the aim of maximizing the genetic heterogeneity sequenced. While the standard PCR volume employed for both primary and secondary PCR is 100 µL; various combinations of 10 µl, 25 µL and 50 µl amplification reactions were attempted for first and second round PCR, to enable multiple start points for PCR. Only reaction volumes of 100 µl were successful routinely and thus this volume was retained. For each sample, 10 parallel first round PCR reactions were carried out using 5 µl of the template. Five µl of the product of each first round was used as starting material for two second round reactions, giving a total of 20 amplification reactions to be used for sequencing (Figure 4.2).

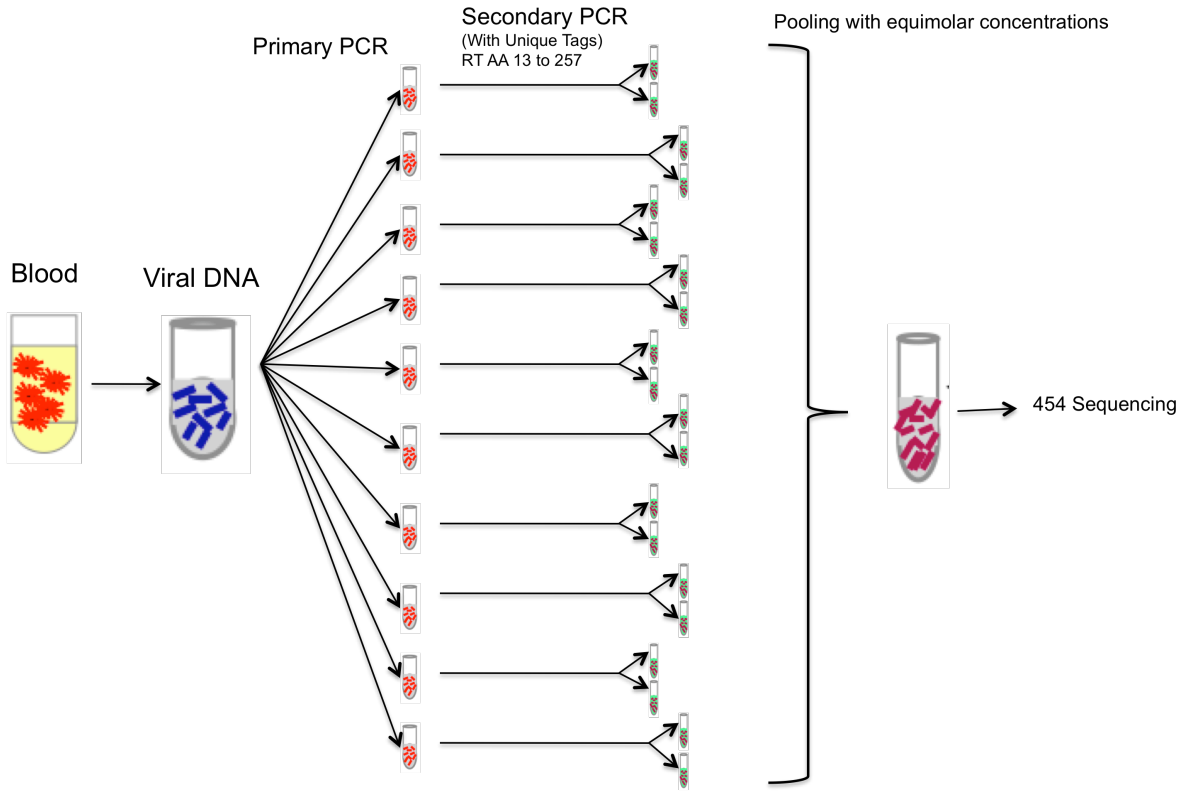


Figure 4.2: Workflow of optimized protocol per sample. Viral DNA was extracted and subjected to 10 primary PCR amplifications followed by two secondary amplifications from each primary. The PCR products from secondary were pooled in equimolar concentrations and subjected for 454 sequencing.

4.2.3 Sample preparation for ultra deep sequencing: Five micro liters of the 20 PCR products obtained from each DNA sample were run on 1% agarose gels alongside the Hyperladder 1 molecular marker (Bioline) to enable quantification and an assessment of the quality of amplification. All 20 PCR products were then mixed to yield a fixed amount of product (1400 ng). A volume yielding 67 ng of the mixed products arising from each DNA sample were then electrophoresed side by side for comparison and further confirmation of quality and quantity, then the combined PCR products from all samples were pooled in equimolar amounts and sent to LGC Genomics (Germany) for deep sequencing. The deep sequencing was carried out in both forward and reverse direction on one half of a picotiter plate, on a GS FLX Titanium Platform by LGC Genomics (Germany).

4.2.4 Data processing and cleaning

The region targeted for amplification was 760 nucleotides in length, covering amino acids positions 13 through 257 of RT (HXB2 Numbering). The sequence reads obtained were separated based on their MIDs and quality control and subsequent analyses were performed independently for the reads corresponding to each sample. The reads obtained were in the Standard Flowgram Format (SFF, files containing flowgrams), which shows the intensity of the light signal when a nucleotide is incorporated (Figure 4.3). To qualify of being a ‘high quality’ read, each sequence has to pass through diverse filters where poor quality reads are removed on the basis of Phred score (Ewing et al. 1998). The data was trimmed using the modified-Mott algorithm implemented in Geneious 5.4.3 (Drummond et al. 2011) with an 0.01 error probably limit.

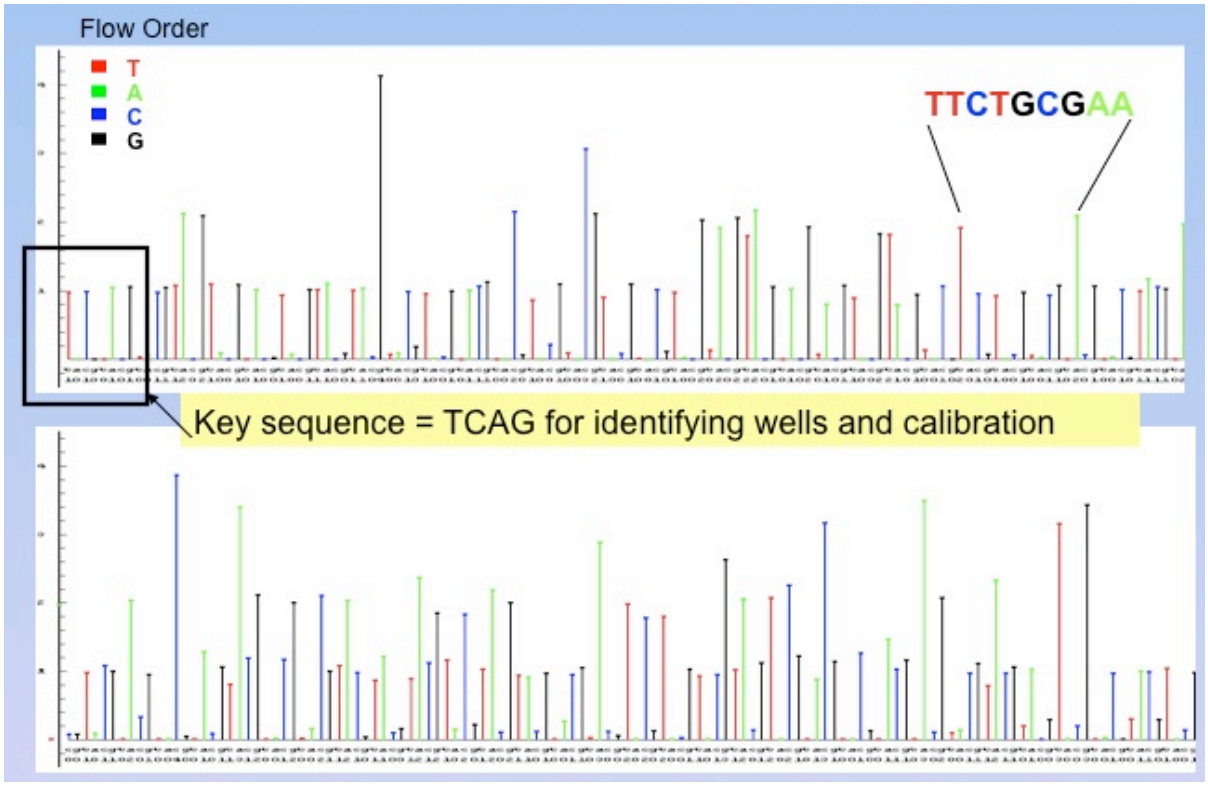


Figure 4.3: Example of Standard Flowgram Format (SFF). The four based "key" sequence identifies the well and calibration, the intensity of light signal is showed when a nucleotide is incorporated.

Reproduced from http://mammoth.psu.edu/howToSeqMammoth.html

The data cleaning strategy was designed to remove reads with probable sequencing errors as follows:

a) The identity of the resulting reads to their corresponding consensus sequence was determined and all reads with low similarity to the corresponding consensus sequence were removed. A modified BLAST word matching tool implemented in Segminator 1.3.2 (Archer et al. 2010) with a word size of five and a read quality of two was used to assemble the sequence reads to their corresponding template consensus sequence (the consensus sequence for each timepoint). For two timepoints (Pt45_13 and Pt76_9), the consensus sequences were not available in which case a consensus sequence from a previous timepoint (Pt45_8 and Pt76_6 respectively) from same individual was used. During this analysis, the reads that did not cover the region of interest (Amino acids 13 to 257 in RT) were removed

b) The reads with significant identity to their corresponding consensus sequence were aligned. The alignment excluded any reads less than 22 nucleotides in length.

c) The reads containing ambiguous bases and reads with out-of-frame indels or stop codons were removed. A subset of reads (6,323 of 14,609) removed for low identity with their respective template from one sample (Pt 42_0_09) were subjected to BLAST analysis to compare their identity with a consensus subtype C genome.

4.2.5 Analysis of viral variants:

The datasets obtained after data cleaning were separated on the basis of forward and reverse directions. The length of sequences that were retained ranged from 22 bp to 350 bp. Therefore each set (forward and reverse) was separated into six categories namely <50 bp, 50-100 bp, 100-150 bp, 150-200 bp, 200-250 bp and >250 bp by using an in-house python script (Developed by Conor Meehan). Forward and reverse sequence reads that were >250 bp were assembled into two alignments using MacClade 4.08 (Sinauer assoc). An amino acid alignment was created from both nucleotide alignments obtained from above. Both nucleotide and amino acid alignments were subjected to an analysis that determines the number of unique nucleotide or amino acid variants. An in-house Python script was used to perform this analysis (Developed by Conor Meehan), which recognizes unique sequences within an alignment and counts the number of each unique sequence.

DNA quantity in each of the 15 amplicon samples was compared to the number of reads returned after quality control by using linear regression. Furthermore, the numbers of variants obtained were correlated with the number of primary PCR reactions by performing linear regression analysis. Similarly the reads returned after quality control, were compared with the numbers of nucleotide variants obtained.

4.3 Results

All fifteen DNA samples from the cell pellets of five HIV-1 positive individuals on ART yielded PCR amplicons of the RT gene. The UDPS approach was successful generating an average of 24,742 RT sequences per sample. The number of successful PCRs and the quality of amplicons varied, as did the numbers of quality RT sequences produced by UDPS. The number of sequence variants within a sample showed direct correlation to PCR success.

4.3.1 PCR Amplifications:

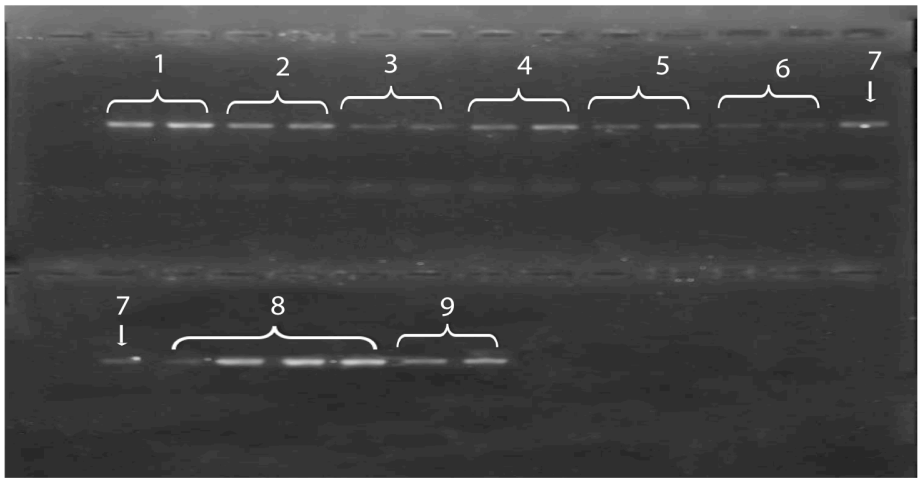
For optimization of smaller volume PCR reactions, five samples were subjected to 10, 25 and 50 µL reactions (both primary and nested secondary). None of the five samples subjected to small volume reactions, showed successful amplification. Furthermore, of five samples from 10µL and 25µL reactions, which stemmed from 100 µL primary reactions, three showed successful amplifications, but the PCR success was sporadic hence it was decided not to use these reaction volumes. A distinct amplification band was observed at an annealing temperature of 57.5^{o}C. Although brighter bands were observed at lower annealing temperatures, because of the presence of non-specific amplification bands, the annealing temperature of 57.5 was considered to be optimal and used in subsequent secondary PCR amplifications.

The number of successful PCR amplifications for primary and secondary PCR reactions are shown in Table 4.2. For six of 15 samples, all ten primary amplicons were positive in subsequent secondary amplifications, while the remaining nine samples showed amplifications from fewer than 10 primary amplicons, (especially all samples from patients 32 and 42 (Table 2). Of the 15 samples included, 10 showed successful amplification in all 20 secondary PCR reactions while for the remaining five, amplification was less successful, with one sample from patient 42 showing only 6 successful secondary PCR reactions stemmed from two primary reactions.

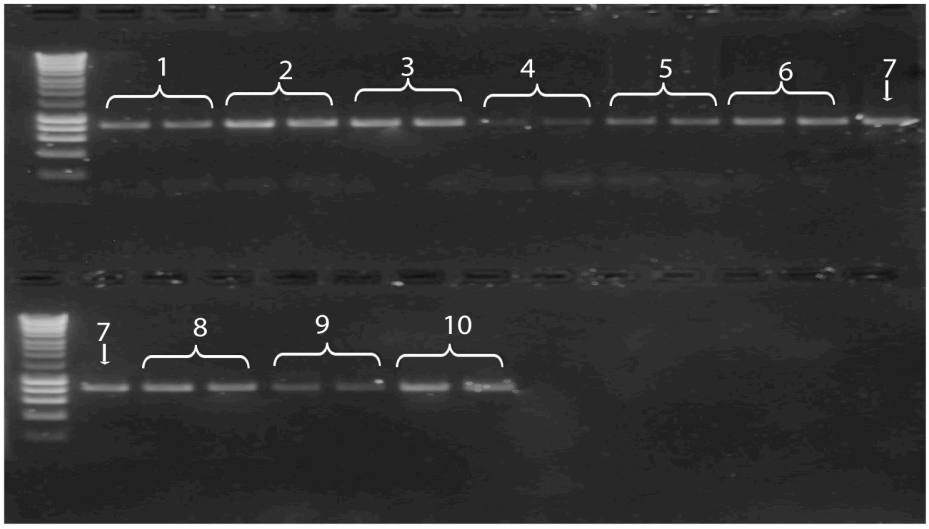
Table 4.2: Number of successful primary and secondary PCR amplifications of the samples subjected to ultra deep sequencing.

Patient	Month	Number of Primary amplicons	Number of Secondary amplicons
Patient 2	Baseline	9	9
	11 Months	10	20
	14 Months	3	20
Patient 32	Baseline	2	8
	7 Months	2	20
	16 Months	3	20
	24 Months	2	20
Patient 42	0 (2008)	2	6
	0 (2009)	5	20
Patient 45	Baseline	9	20
	8 Months	10	20
	13 Months	10	20
Patient 76	Baseline	10	20
	6 Months	10	15
	9 Months	10	19

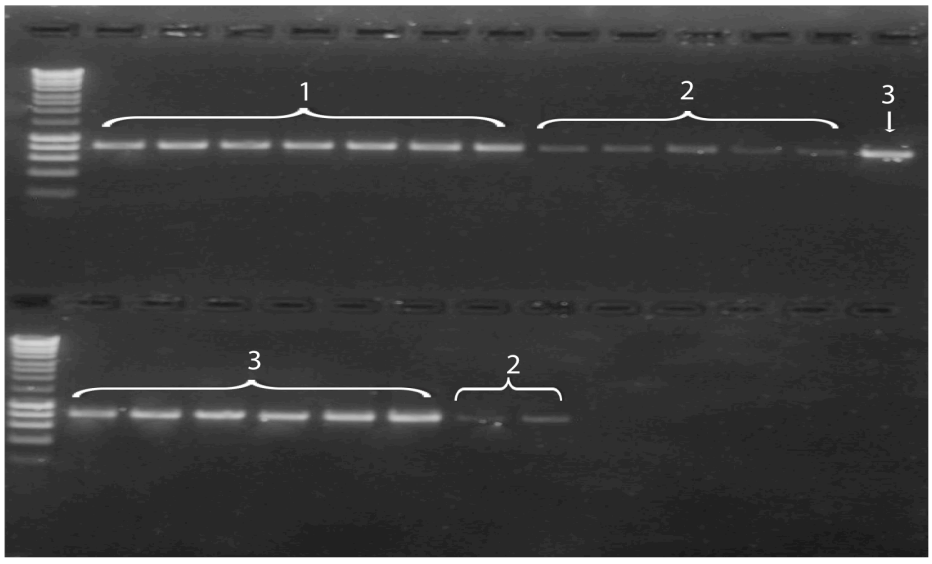
The PCR amplification products for patient 2 are shown in Figure 4.4. DNA quantity was the same in all PCR products from secondary reactions that stemmed from a single primary PCR, while the concentration of DNA differed in the secondary PCRs that originated from different primary reactions. Similar results were observed in all samples from patient 32 and the first baseline sample from patient 42 (Figure 4.5, Figure 4.6). However, two samples from patient 32 (Pt32_7; Figure 4.5b and Pt32_16; Figure 4.5c) and all 20 amplicons from the second baseline sample from patient 42 showed very weak amplification (Figure 4.6b). All 20 amplicons from this second baseline sample from patient 42 stemmed from same primary amplicon. However, in samples from patients 45 and 76, there was also variation in the quantity of DNA in amplicons stemming from the same primary PCR (Figure 4.7 and Figure 4.8)



a. Pt2_0



b. Pt2_11



c. Pt2_14

Figure 4.4: 1% agarose gel of secondary PCR amplifications from three samples of patient 2. Each number represents the primary PCR from which secondary PCR product/s were generated.

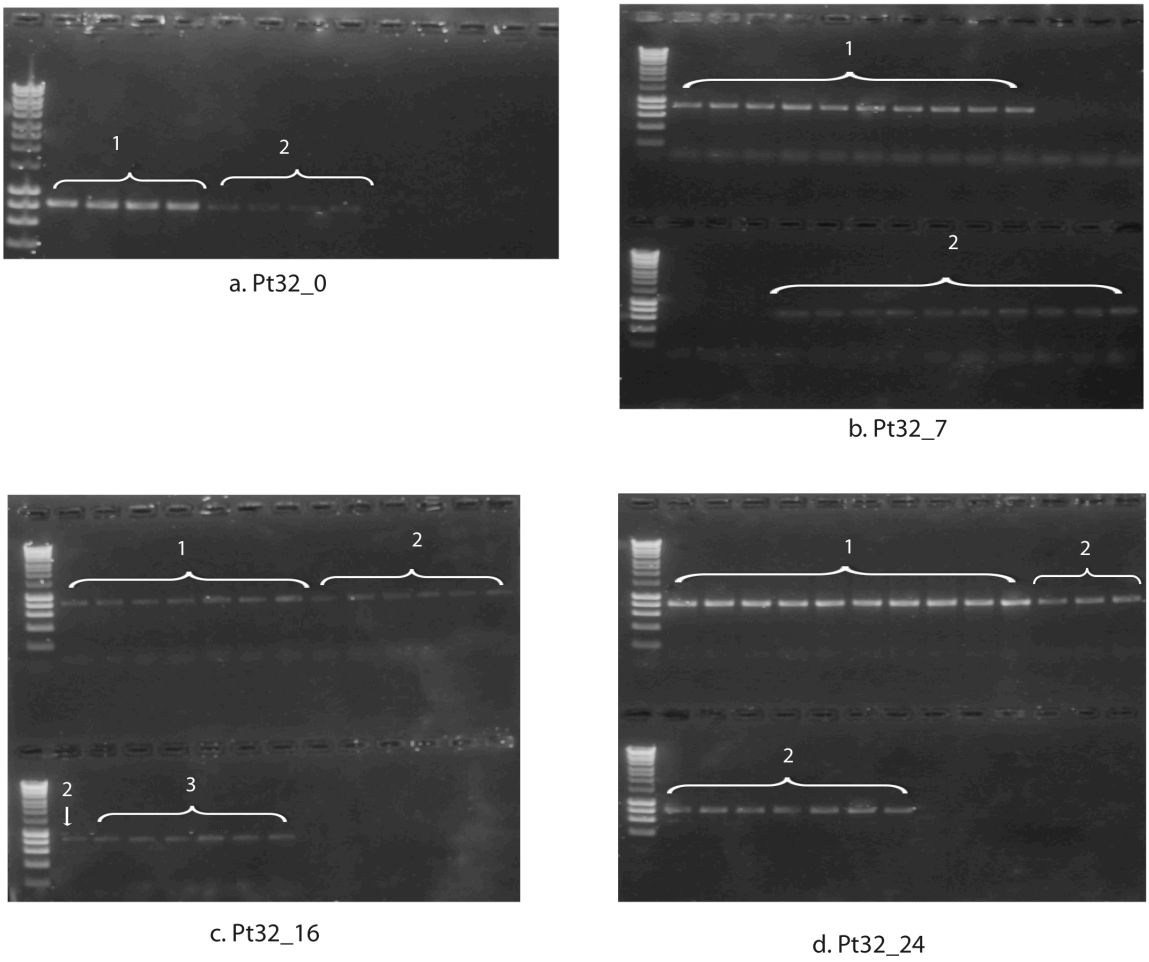


Figure 4.5: 1% agarose gel of secondary PCR amplifications from four samples of patient 32. Each number represents the primary PCR from which secondary PCR product/s were generated.

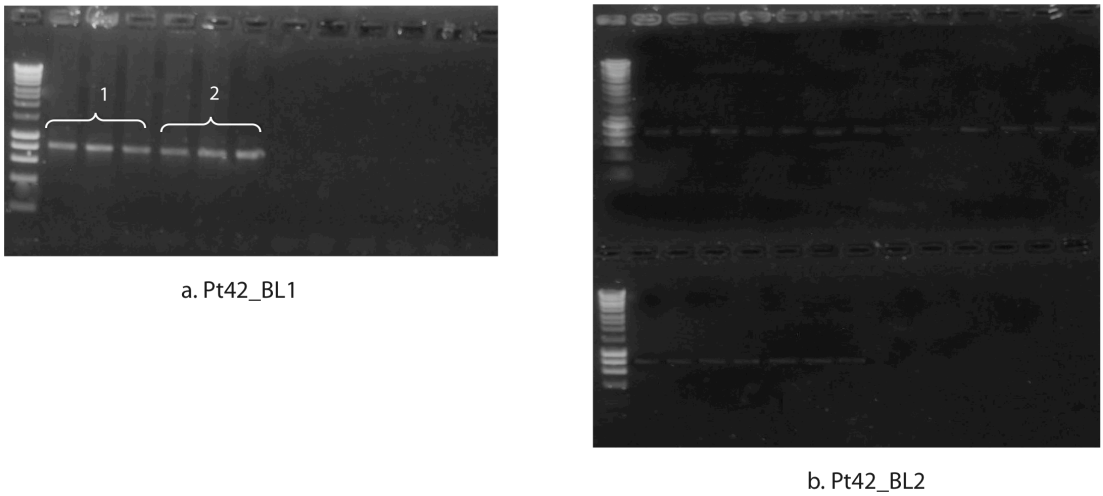
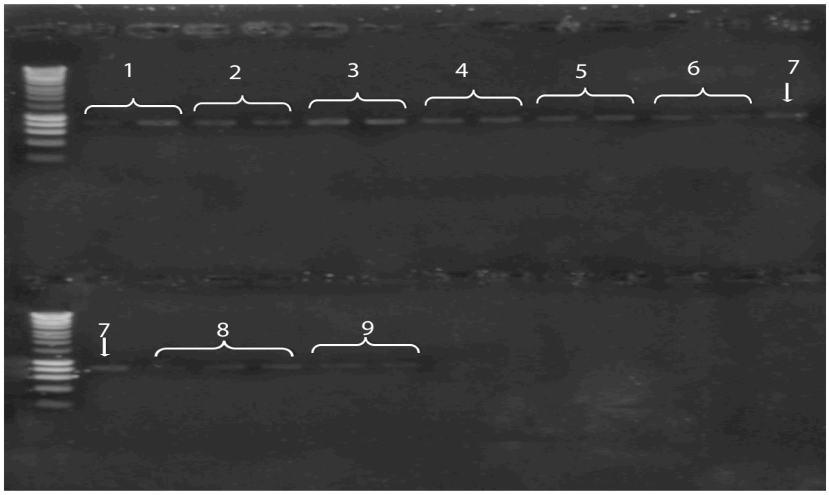
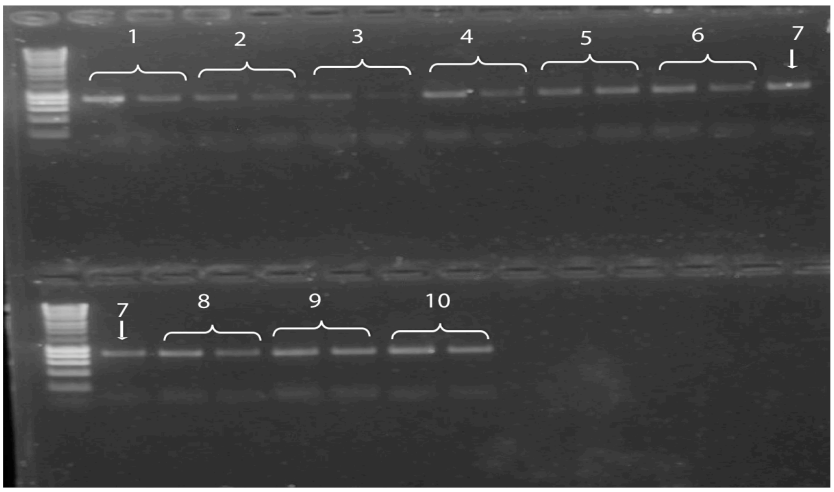


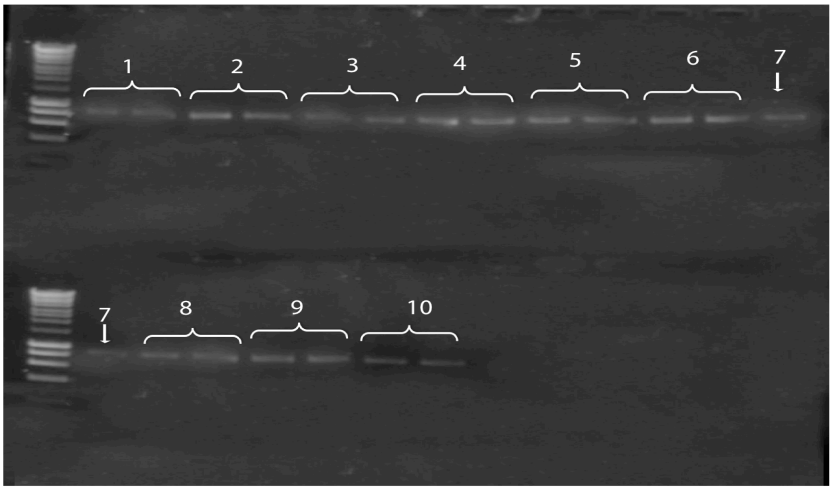
Figure 4.6: 1% agarose gel of secondary PCR amplifications with modified primers from two samples of patient 42. In a. the number represents the primary PCR from which the secondary PCR products were generated, In b. all 20 secondary PCR amplifications stemmed from one primary PCR.



a. Pt45_0



b. Pt45_8



c. Pt45_13

Figure 4.7: 1% agarose gel of secondary PCR amplifications from three samples of patient 45. Each number on the top of the band represents the primary PCR from which secondary PCR product/s were generated

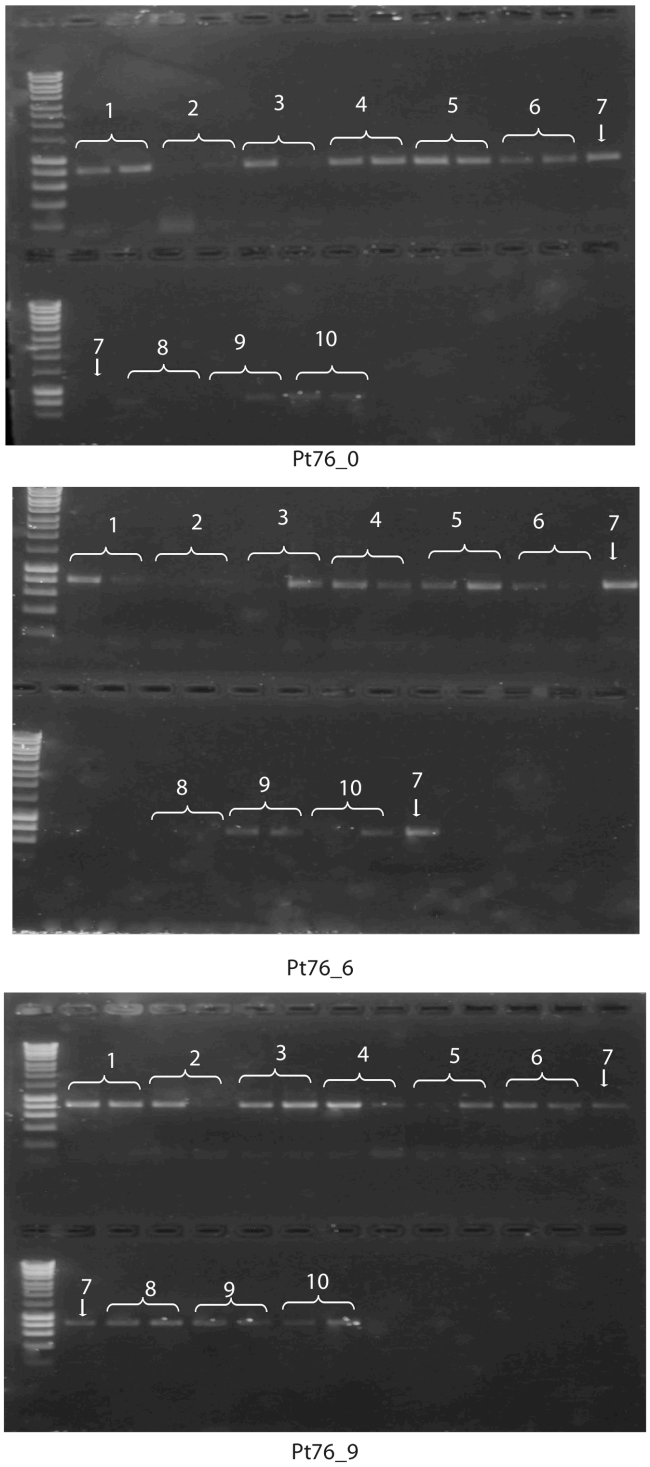


Figure 4.8: 1% agarose gel of secondary PCR amplifications from three samples of patient 76. Each number on the top of the band represents the primary PCR from which secondary PCR product/s were generated

For equimolar pooling of PCR products, visual inspection of each band was carried out. Each band was assigned a quantity based on comparison with the standard molecular marker. Due to the amount of combined product required by LGC Genomics for the pyrosequencing reaction (20µg) appropriate amounts in µL of each amplicon was combined to make 1400 ng DNA per sample. An appropriate aliquot of each pooled sample was then taken in such a way that it will contain 67 ng DNA (leaving 1333 ng for the sequencing reaction). Each 67 ng aliquot was visualised on a 1% agarose gel, with same molecular marker. As shown in Figure 4.9, band intensity was not equal across the fifteen sets of pooled amplicons. The amount of product was much lower in the wells containing the pooled amplicons from the second baseline sample from patient 42 (Pt42_0_09) and the baseline sample from patient 45 (Pt45_0). Pooled amplicons from Pt45_8, Pt45_13 were also relatively weak while those from Patient 76 were among the brightest.

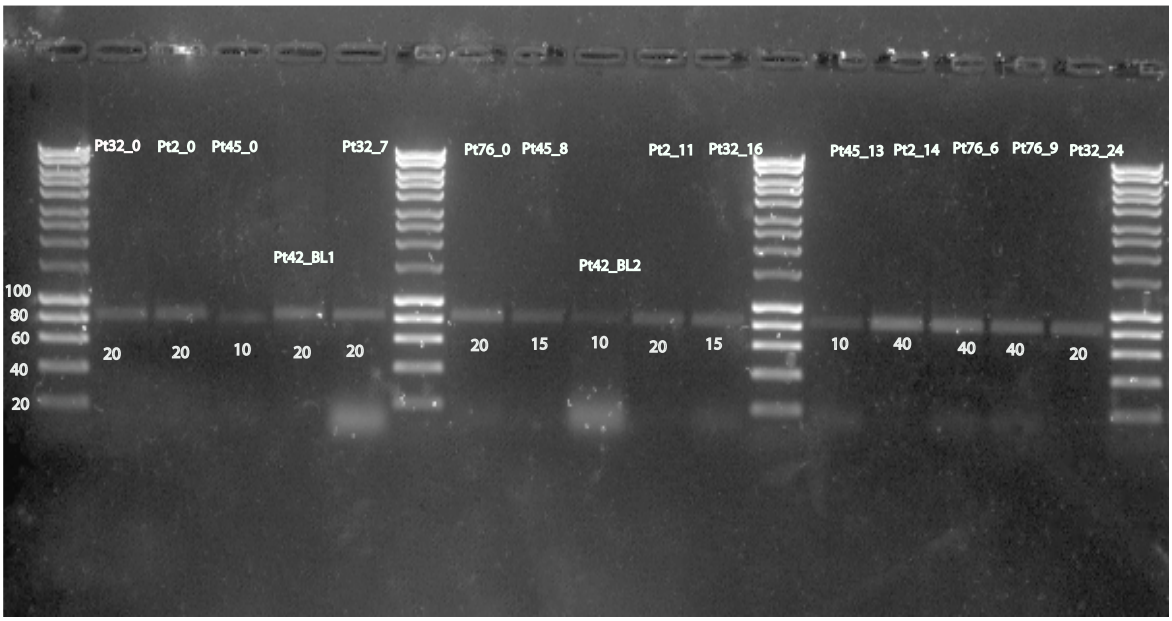


Figure 4.9: 1% agarose gel electrophoresis of all 15 samples from five patients loaded after equimolar pooling. Representative bands from hyper ladder I are labeled with the quantity of the DNA (ng) (Source: Bioline Reagents Ltd, UK). The assigned amount of DNA as per the Hyperladder I (Bioline) (in ng) to each pooled sample is given below each band

4.3.2 Deep Sequencing and data cleaning

A total of 372,169 sequence reads were returned from ultra deep sequencing from the 15 samples from 5 patients, of which 1038 reads could not be sorted on the basis of their MIDs, hence 371,131 reads were taken for further analysis, where, the number of reads per sample ranged from 12,926 to 38,514 (Table 4.3, Figure 4.10A). The average length of reads obtained prior to quality control analysis was 541 bp. Further processing of reads included trimming on the basis of quality with a 0.01 error probability limit. This means each trimmed read had an average base call accuracy of 99.9%. This trimming step did not result in the removal of any reads (as no minimum length requirement was imposed in this step), however the average length of all reads was significantly decreased from 541 bp for reads before trimming to 176 bp reads following trimming (Figure 4.10B). Furthermore, all reads less than 22 nucleotides were excluded, which resulted in the removal of 22,607 reads (range 677 to 3007) and between 1,324 and 14,609 reads per patient (total 68,894) were excluded because of low identity with template (Figure 4.10A).

Table 4.3. : Numbers of reads obtained from deep sequencing and number and percentage of reads obtained after data cleaning from 15 samples from 5 patients. The highlighted figures represent the three samples where highest numbers of reads were removed

Patient	Month	Reads Before data cleaning	Reads retained	% Of reads retained	Reads Removed	% Of reads removed	Too short Reads	Reads with low identity with template
Patient 2	0	33022	26493	80.25	6522	19.75	2180	4342
	11	22000	18053	82.07	3945	17.93	1173	2772
	14	31034	25496	82.17	5532	17.83	3007	2525
Patient 32	0	20565	16314	79.37	4240	20.63	1160	3080
	7	23961	19866	82.92	4092	17.08	1282	2810
	16	12926	4987	**38.58**	7938	61.42	677	**7261**
	24	23902	19862	83.12	4035	16.88	1657	2378
Patient 42	0 (2008)	17614	15160	86.08	2451	13.92	1127	1324
	0 (2009)	24452	8471	**34.65**	15978	65.35	1369	**14609**
Patient 45	0	24509	19782	80.72	4724	19.28	1017	3707
	8	20516	15438	75.26	5076	24.74	1273	3803
	13	23418	13359	**57.06**	10053	42.94	1172	**8881**
Patient 76	0	32563	23911	73.45	8643	26.55	1509	**7134**
	6	38514	33874	88.08	4585	11.92	2734	1851
	9	22135	18447	83.34	3687	16.66	1270	2417

After data cleaning the number of sequence reads ranged from 4,987 to 33,874 per sample (Table 4.3). The percentage of the original number of reads per sample taken for further analysis after data cleaning ranged from 34.65 to 88.08 (Table 4.3, Figure 4.10A). Three of the fifteen samples (Pt45_13, Pt32_16 and Pt42_BL2) had a large number of reads removed during quality control with 43%, 61% and 65% of the reads removed respectively. In each of the cases, the majority of reads were removed due to their low identity with template (Table 4.3). The coverage of reads across the RT gene after quality control is shown in Figure 4.11. The numbers of reads obtained at the 5' end of RT are higher than at the 3' end for both primers. RT amino acids 103 to 141 had the lowest number of reads. From one timepoint of patient 42 (Pt 42_0_09), of the 14609 reads rejected because of low identity with the template, a subset (6323) was subjected to BLAST search to investigate what the identity of the sequences. It was found that 469 reads matched to a region of the HIV-1 genome outside of region of interest while the remaining 5,854 did not match with any region of HIV-1 subtype C.

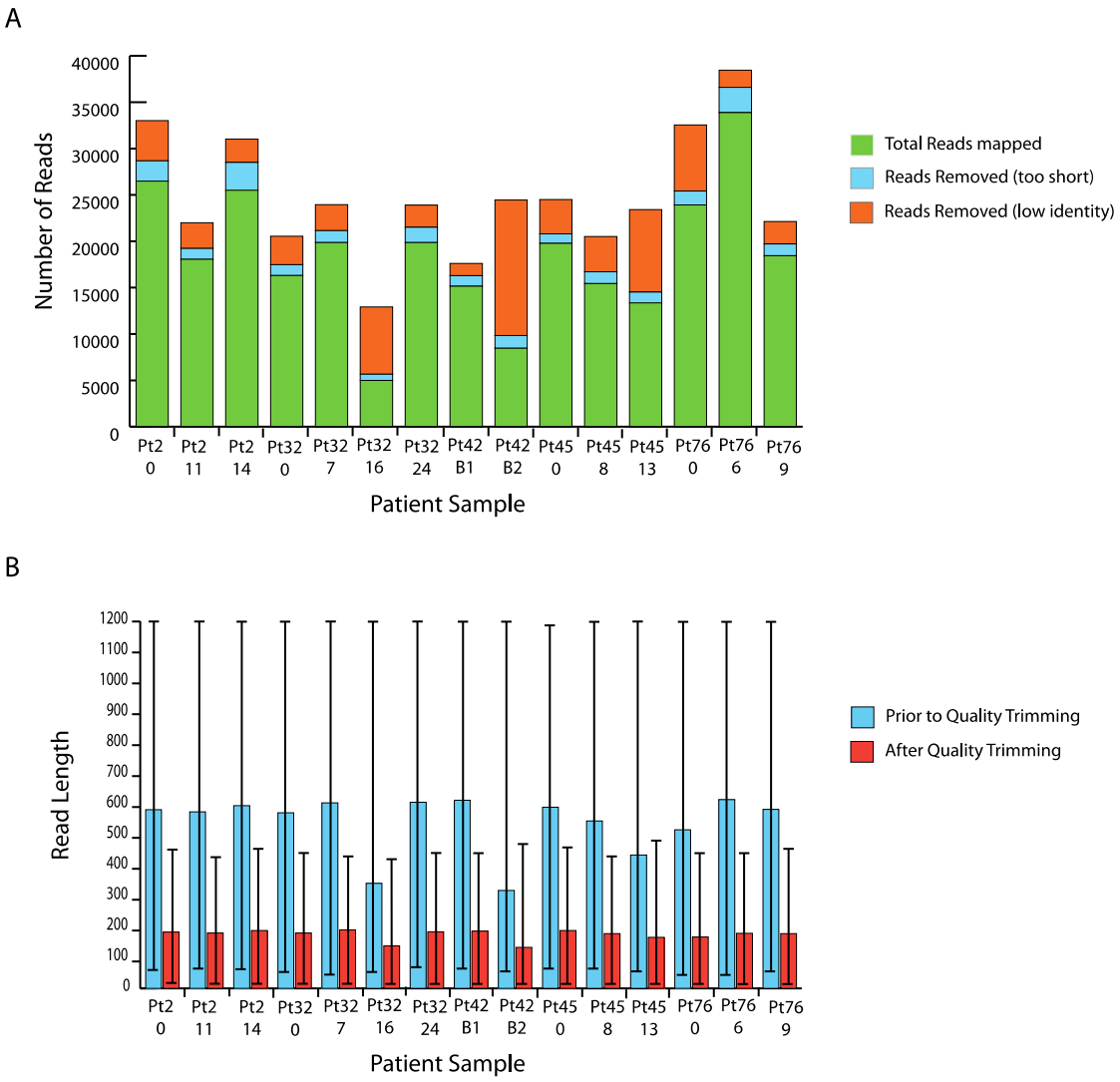


Figure 4.10: The number of nucleotide reads obtained from deep sequencing

(A) The total number of sequenced reads obtained for each sampling time-point is shown as the total size of the bar representing each time-point. The number of reads mapped to the reference sequence and removed as a result of length and identity cutoffs are also shown. (B) The mean and range of read lengths observed for each sequenced sample are shown both before and after quality trimming. In all cases the mean and range of read lengths decreases significantly following trimming.

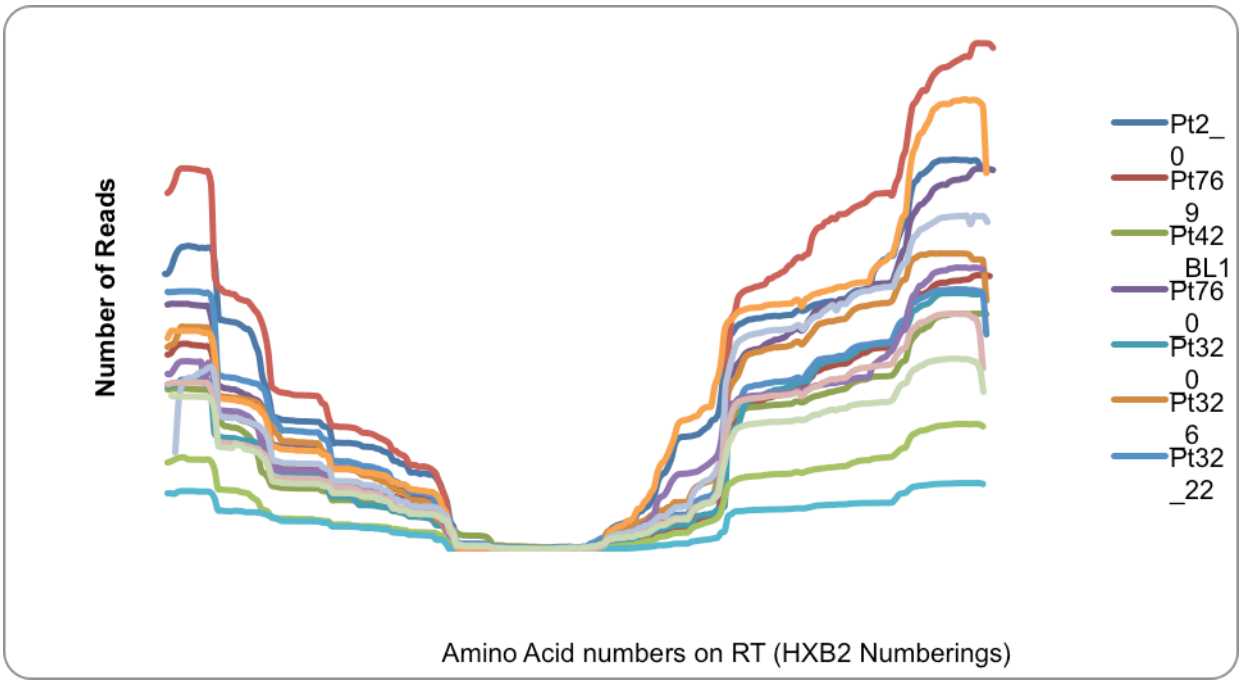


Figure 4.11: Coverage of UDPS reads across the RT gene (HXB2 AA 13-257)

4.3.3 Analysis of viral variants:

Reads obtained after data cleaning were separated into six different categories based on the nucleotide length, their distribution in these six categories is outlined in Appendix Table 2. Most reads were in the nucleotide length range 50-99, followed by 200-249 and then >250 nucleotides (Appendix Table 2, Figure 4.12). The number of reads greater than 250 ranged from 961 to 8,112 with forward and reverse reads together, the lowest being those from Pt32_16 while highest were from Pt2_14. The second lowest number of sequences (>250 bp) was 1757 (Pt42_0_1), followed by 3131 (Pt45_13). The three samples that had the lowest number of reads > 250 bp were also those that had a large numbers of reads removed during quality control analysis (Table 4.3) and the lowest mean read length (Fig 4.10).

The numbers of unique variants obtained from each sample are outlined in Appendix Table 3 and Figure 4.13. Overall, the highest number of nucleotide and amino acid variants were found in samples from patients 76 and 2 with the baseline sample from patient 76 being the highest (2,794 Nt and 1,553 AA variants). The lowest number of variants was observed in 16th month sample

from patient 32 (301 Nt and 238 AA variants) followed by second baseline sample from patient 42. Patients 2 and 76 showed uniformity in the number of variants across all time points. In patient 45, the number of variants increased with the baseline sample showing the lowest number of variants than subsequent timepoints. In patient 42, first baseline sample showed more variants than second baseline while samples from patient 32 showed variations across all timepoints (Appendix Table 3, Figure 4.13).

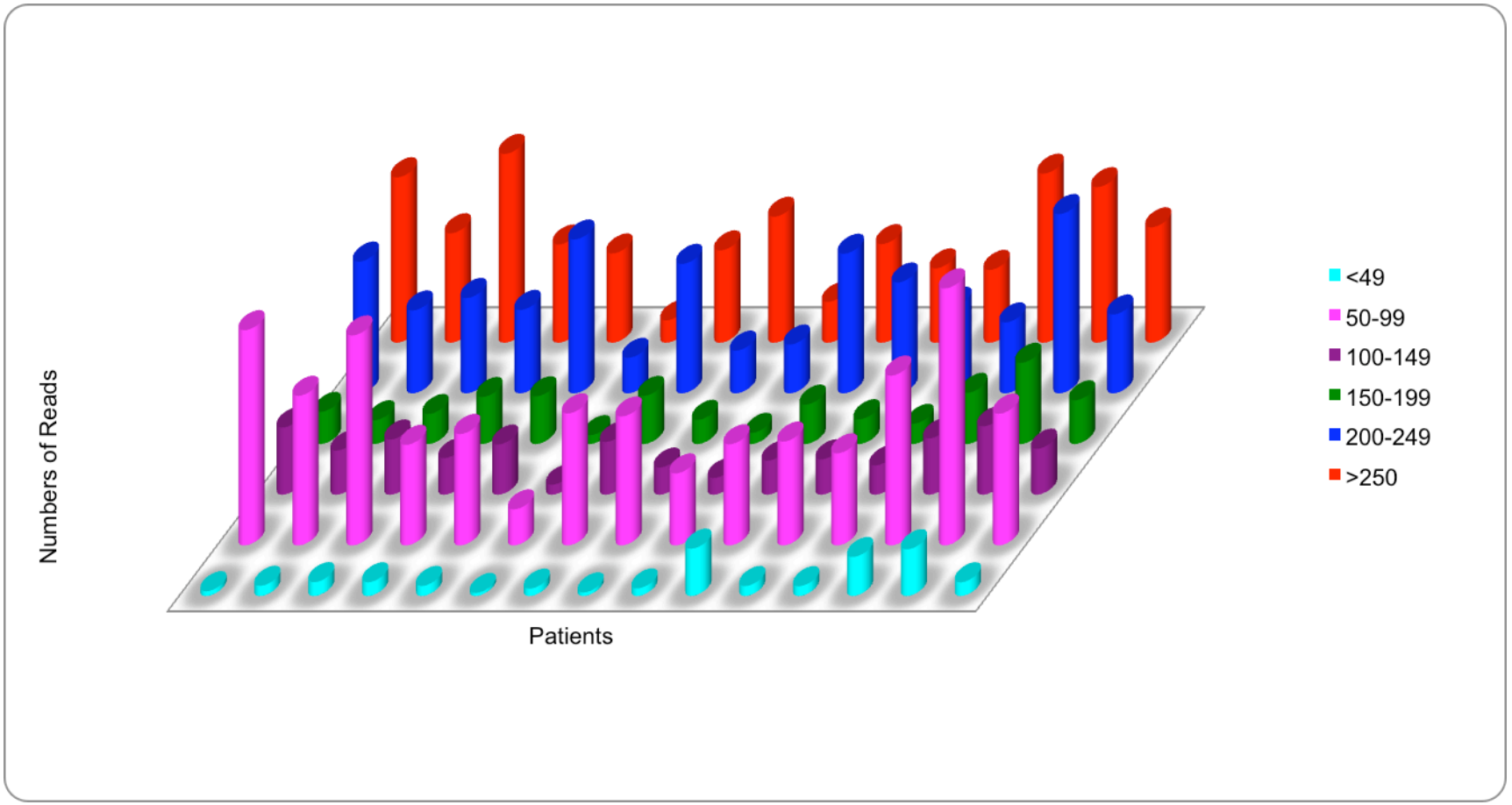


Figure 4.12: Reads after segregation into different categories on the basis of length.

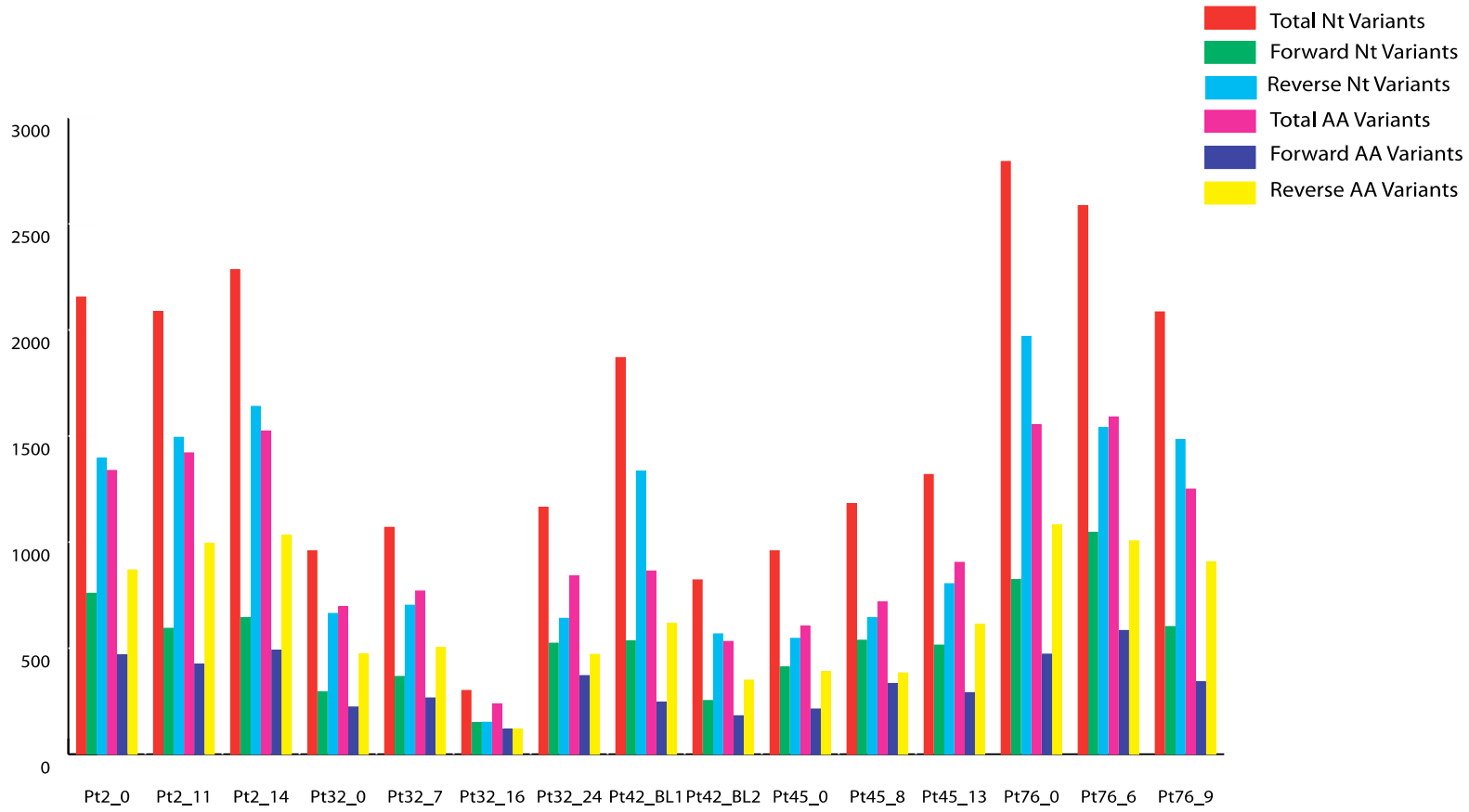


Figure 4.13: Nucleotide and amino acid variants in sequence sets of >250 nucleotide in 15 samples from five patients.

The number of nucleotide and amino acid variants showed a direct correlation to the numbers of reads returned after quality control (Figure 4.14). The nucleotide reads and the nucleotide and amino acid variants obtained were also viewed in light of the success of PCR amplification and the quantity of PCR amplicons. The 16th month sample from Patient 32 (Figure 4.5c), along with the 2nd baseline sample from patient 42 (figure 4.6b) showed the weakest amplification in secondary PCR reactions and also showed the lowest numbers of reads obtained.

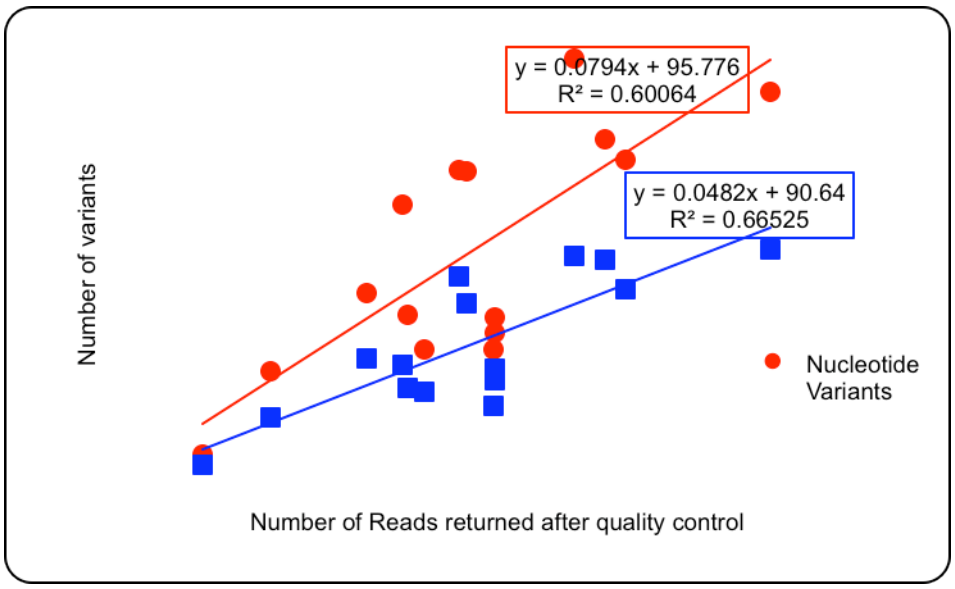
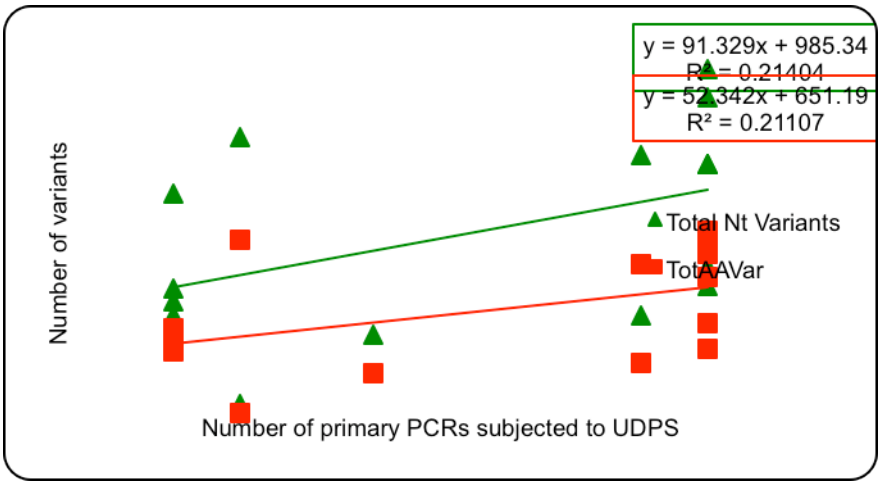


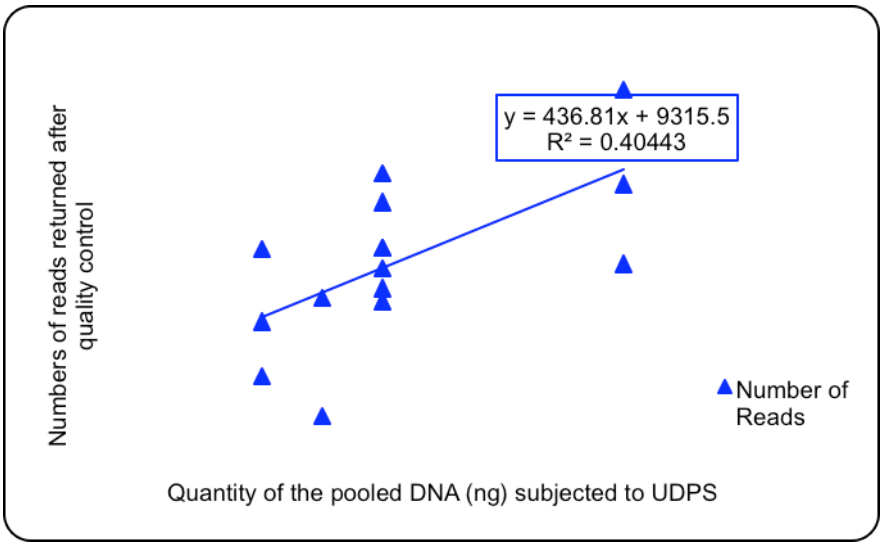
Figure 4.14: Comparison of number of reads obtained after quality control with numbers of nucleotide and amino acid variants obtained in sequences greater than 250 basepairs by linear regression analysis.

In general, the numbers of variants were higher in the samples where the PCR reactions were successful from 9 or 10 primary reactions but the correlation was not significant (Figure 4.15a). The number of primary PCRs subjected to secondary PCRs does not seem to have good correlation with the numbers of variants obtained (Figure 4.15a). However, some cases, despite only three successful primary PCR reactions, high numbers of nucleotide and amino acid variants were obtained (e.g. In Pt2_14, 2284 Nt variants and 1564 AA variants). On the other hand, low number of variants was observed in some samples with 9 or 10 successful PCR reactions, (e.g. Pt 45, 959 Nt variants and 604 AA Variants) (Figure 4.15a).

To analyze further, a quantity was assigned as per Hyperladder I to the gel electrophoresis band obtained after running the pooled samples (Figure 4.10) and was correlated with the numbers of reads obtained after quality control by performing linear regression analysis. It can be seen from Figure 15a, there was a significant correlation between the quantity of the pooled DNA and the numbers of reads returned from UDPS.



a



b

Figure 4.15: Comparison of a) The quantity of pooled DNA subjected to UDPS and b) Numbers of primary PCRs to number of variants by linear regression analysis.

4.5 Discussion:

The aim of this chapter was to develop an experimental protocol for maximizing the diversity of viral templates subjected to UDPS and further explore the effect of PCR on the number and diversity of sequences returned from UDPS. Here, multiple PCR reactions were performed, followed by equimolar pooling of samples subjected to UDPS. The genetic diversity obtained after UDPS was analyzed. It was expected that the approach of using multiple PCRs would have an effect on viral diversity found in the sequences returned. It has been found that the number of multiple primary PCR reactions that are subjected to UDPS does not affect the number and diversity of sequences returned from UDPS. Hence, this approach does not seem to have any effect on the viral diversity. However, the quality of amplification, of the PCR reactions played a crucial role in generation of number of sequence reads and diversity.

The number of initial viral templates that can be successfully extracted and amplified limits the sensitivity of UDPS to detect variants. To estimate the genetic diversity within a sample, an adequate sample size that represents the viral diversity within the patient is required. Several previous studies used plasma with initial volumes ranging from 140 µL to 1000 µL (mostly 200 µL). This was due to two reasons; i) variation in initial viral load, different initial volumes were used (Hedskog et al. 2010) or ii) in order to compare the 454 sequencing with initial bulk sequencing which used a particular amount of starting volume (Le et al. 2009). The sample quantity equivalent to 200 µL of blood was used as a starting volume to extract proviral DNA, as the viral load was not available. The bulk sequencing was performed on the same sample. This was followed by multiple primary PCR reactions to maximize the genetic heterogeneity in UDPS sequences returned. The combination of multiple first and second round PCRs and equimolar pooling enabled the reduction of “founder effect” of the PCR. Conventional bulk sequencing and deep sequencing was performed on the DNA obtained from same extraction, allowing us to have direct comparison of the two methods. The numbers of multiple PCRs did not show any positive correlation with the numbers of variants obtained as few samples with just three numbers of PCRs yielded higher numbers of variants (See Figure 4.15a).

Initial attempts to use smaller volumes of PCR reactions resulted in limited success probably due to small reaction space. Most of the samples showed successful amplification of the RT gene in the 100µL reactions with the exception of patients 32 and 42, where there was limited success in the primary PCR reactions. This might due to low viral load in these patients but as the viral load is not available, it cannot be verified. In the secondary PCRs, there was weak amplification in three samples (Pt 32 month 16, patient 45 month 13 and the second baseline sample of patient 42). Therefore for these samples, higher amount of PCR product had to be added while doing equimolar pooling to reach the final quantity for the sequencing. For these three samples, more sequences were excluded from further analysis as they had low identity with their templates. Two (Pt32_16 and Pt42_BL1 of these three samples returned with less than 10,000 reads. According to (Eriksson et al. 2008) this number is not sufficient to detect low abundance DRMs representing 0.1 to 1% of the population. On the other hand, for few successful PCRs with strong amplification (e.g. Pt2_14), less numbers of sequences were excluded from the quality control analysis. Thus by just adding more products from multiple amplifications that were weak has not been found to be optimal. The recommendation here is to ensure that all PCR reactions are optimized.

Previous studies have shown that 454-pyrosequencing technology can reliably detect rare variations over cloning/consensus sequencing (Hedskog et al. 2010; Codoner et al. 2011; Liang et al. 2011). Our findings are consistent with those reports. Here, ultra deep pyrosequencing has been used to investigate the presence of low abundance viral variants present within the proviral DNA of individuals on ART. This experiment generated thousands of sequences from proviral DNA in a single experiment. Despite the data loss (25%) upon filtering for quality (From 371,131 to 279,513 reads) and a short mean read length (from 576 to 176 bps) in this study, we have achieved sequence reads in the numbers that allowed to detect nucleotide variants (ranging from 301 to 2586 variants), which is not easily achievable by SGA or cloning (even though for two samples we returned less than 10,000 sequence reads). Furthermore, due to use of GS FLX platform and careful optimization of initial PCR reactions, this experiment has sequenced considerably deeper (median 18,447, range 4,900 to 33874 reads after quality control) compared

to the previous studies where the median of sequence reads obtained ranged from 3900 to 16,016 (Wang et al. 2007; Mitsuya et al. 2008; Le et al. 2009; Varghese et al. 2009; Hedskog et al. 2010). This allowed us to gain better coverage of sequences representing the viral population. Although stringent quality control criteria were used, the amount of data loss in our study is 25%. This is low compared to another study on proviral DNA (Liang et al. 2011) maybe due to the use of a better sequencing platform (GS FLX vs GS 20). Although it was expected to get full coverage of RT by using primers from both ends of RT gene and given the average length of reads promised by LGC genomics (around 400 bp each primer), due to variations in PCR success, the quality trimming step reduced the average length of sequences from 576 to 176 base pairs.

In conclusion, ultra-deep pyrosequencing has proven to be a powerful method for characterizing genetic diversity of HIV-1 quasispecies, especially for detection of low abundance variants. But considerations should be paid to quality of the amplification of initial templates. It is also recommended here to use multiple overlapping amplicons that cover the entire RT so that uniform coverage can be obtained.

5

Characterization of the emergence and Persistence of Drug Resistant Mutations in HIV-1 Subtype C Infections using Ultra Deep Pyrosequencing

Contents of this chapter contributed to the following article

Vijay Bansode, Grace P McCormack, Amelia Crampin, Bagrey Ngwira Neil French, Judith R Glynn, Simon A A Travers. 2012. Characterizing the Emergence and Persistence of Drug Resistant Mutations in HIV-1 Subtype C Infections using 454 Ultra Deep Pyrosequencing. BMC infectious diseases. Submitted

5.1 Introduction

It was shown in chapter 2 that, of 40 HIV-1 subtype C infected individuals on ART studied, 14% contained drug resistance mutations (DRMs) in the reverse transcriptase (RT) gene of the proviruses (Bansode et al. 2011a). Bulk sequencing of latent viruses during HAART showed discrepancies between different timepoints in the presence of DRMs in the proviral DNA during treatment (Bansode et al. 2011b) (Chapter 3). These discrepancies point to a limitation of bulk sequencing for determining drug resistance in viruses contained within the proviral DNA. Furthermore, while the role of minor variants in HIV-1 RNA in the emergence of resistance in persons on highly active antiretroviral therapy (HAART) has been widely documented, less is known about the implications of historical viruses in the proviral DNA. Following treatment interruption, latent viruses stored in the provirus have been shown to be responsible for the rapid rebound of viral load (Garcia et al. 1999; Boucher et al. 2005; Joos et al. 2008). However, Palmisano and colleagues suggested that the mutational archive stored in proviral DNA remains unchanged during HAART (Palmisano et al. 2009), with higher levels of DRMs observed in RNA extracted from circulating virus than those present in the provirus (Wirden et al. 2011). The suggested reason for these discrepancies is that standard bulk sequencing cannot fully access the spectrum of viral variants stored in the proviral DNA (Wirden et al. 2011), and resistant viruses may be present in low abundance in proviral DNA.

Several studies have found positive correlations between the presence of low abundance viruses and clinical outcomes (Peuchant et al. 2008; Jakobsen et al. 2010). To my knowledge, no studies have been undertaken to use ultra deep pyrosequencing (UDPS) to quantify the prevalence of low abundance drug resistant viral variants in the proviral DNA. Here, we undertake such a study and endeavor to correlate the presence/absence of such viruses with treatment outcome. The aims of this chapter are;

i) To use UDPS to investigate drug resistance mutations in longitudinally collected blood samples from patients on ART.

ii) To gain deeper knowledge of the discrepancy in the presence of DRMs at various timepoints found by conventional bulk sequencing.

iii) To identify DRMs against the drugs other than first line therapy in Malawi.

5.2 Materials and Methods:

5.2.1 Patients, ultra-deep sequencing and analysis of sequences

Five individuals infected with HIV-1 subtype C recruited to an antiretroviral cohort between 2007 and 2009 were selected on the basis of discrepancies or ambiguities in drug resistance mutations found through standard bulk sequencing (Chapter 3, table 3.1, (Bansode et al. 2011b). Fifteen samples representing an average of three sampling points during treatment were subjected to ultra deep sequencing, the data obtained was subjected to quality control analysis (Chapter 4). The sequence reads, that covered 13 to 257 AA of the RT gene (HXB2 numbering) generated after quality control analysis were employed in drug resistance analysis (Chapter 4).

5.2.2 Drug resistance analysis:

For all sequence reads obtained after data cleaning, the nucleotide data was translated into amino acids sequence reads. These sequences were analyzed for the presence of drug resistance mutations (DRMs) based on the Stanford Drug Resistance Database (Rhee et al. 2003). The frequency of amino acids at each position was determined using Segminator 1.3.2 (Archer et al. 2010). The prevalence of DRMs at each amino acid position was calculated. The prevalence was categorized into three categories; First, DRMs detected in less than 1 % of the reads sequenced from the amplified viral population for that position; this category was discounted from further analyses to account for potential errors due to the error rate of PCR and UDPS. The second category of mutations consisted of those with a prevalence rate greater than 1% and less than 20 %; these mutations correspond to those that cannot be determined using bulk sequencing. The third category consisted of those DRMs with prevalence greater than 20%, comprising mutations that can, in theory, be observed using traditional bulk sequence genotyping.

5.3 Results

Thirteen of 15 samples from five individuals on ART showed DRMs against current 1st line therapy in Malawi (Lamuvidine, Stavudine and Nevirapine). Although the majority of the mutations detected were at a prevalence level <1%, various low abundance DRMs (with a prevalence level ranging from 1 to 20%) and high abundance DRMs (prevalence >20%) were detected by UDPS. Furthermore, additional mutations against other antiretroviral drugs (e. g. Rilpivirine, Etravirine) were also detected.

5.3.1 Prevalence of Drug Resistance Mutations identified with UDPS

Drug resistance mutations shown by deep sequencing were classified into three categories namely; <1 %, 1 to 20% and > 20%. In all samples the majority (between 74% and 96%) of drug resistance mutations identified by deep sequencing were detected at prevalence levels less than 1% of the sequenced viral population (Figure 5.1A) and were, thus, excluded from any subsequent analysis. An average of 11% (range 2.5-21.21%) were detected at prevalence levels between 1% and 20% of the sequenced viral population with an average of 6% (range 0-12.5%) identified at greater than 20% prevalence in the sequenced viral population (Figure 5.1A).

Table 5.1. Prevalence of DRMs in pyrosequencing data: Only DRMs observed at >1% in the sequence data are shown. DRMs relevant to each patient's current therapy regime are shown in bold. CD4 counts ($x10^{-6}$/L) are shown. When a CD4 count was not available for a sampling time-point the count from the sample taken at a date closest to that time-point is shown. DRMs that were also found in bulk sequencing are underlined.

Patient Number	Sampling Time-point	CD4 Count	1-20% Prevalence	>20% Prevalence
Patient 2	Baseline	64		
	11 Months	N/A (9 months: 110)	**K65R**, **M184V**, **G190E**	
	14 Months	153	**K65R**	
Patient 32	Baseline	271	V90I	
	7 Months	384	**K65R**	V90I
	16 Months	N/A (19 months: 443)	**K65R**	
	24 Months	N/A (19 months: 443)	**K65R**	
Patient 42	Baseline 1 (2008)	N/A	**K65R**	
	Baseline 2 (2009)	362	**K65R**, A98S, **G190A**	
Patient 45	Baseline	55	A98S, **K101E**, K103E, **K103N**, K103R, **M184I**	V106I, E138A, **G190A**
	8 Months	216	A98S, **K101E**, K103E, K103R	V106I, **G190A**
	13 Months	258	**K65R**, **K101E**, K103E, **K103N**, K103R	V106I, E138A, **G190A**
Patient 76	Baseline	118	**K65R**, T69I, **G190A**, H221Y	V90I, A98S, **Y181C**
	6 Months	N/A (8 months: 269)	**K103N**, H221Y	V90I, A98S, K103E, **Y181C**, H221Y
	9 Months	N/A (12 months: 268)	V90I, K103E, **G190A**, H221Y	A98S, K103R, **Y181C**

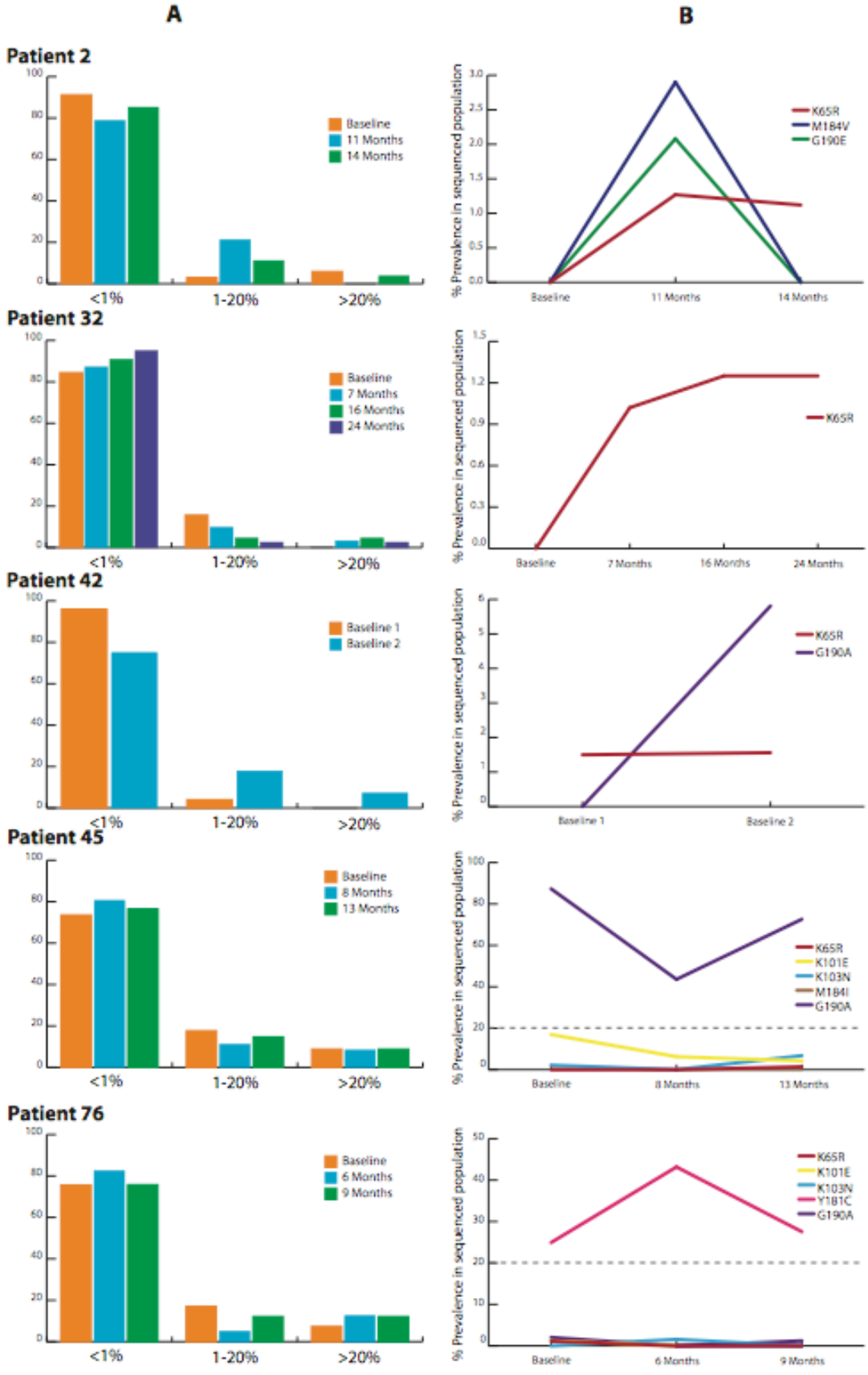


Figure 5.1: A) DRMs found in UDPS classified into three categories on the basis of prevalence levels
B) DRMs against current therapy in Malawi found in UDPS at >1% prevalence level

5.3.2 Comparison of DRMs found in bulk sequencing and UDPS

Of the 15 DRMs (five to current therapy) seen in bulk sequencing of samples subjected to UDPS (Chapter 3, Table 3.1), 10 of these (four against current therapy) were also observed in the UDPS sequencing with the vast majority of these (8 of the 10) being seen at a prevalence greater than 20% in the sequenced viral population (Table 5.1). The two DRMs observed at less than 20% of the sequenced viral population were V90I (9.5% prevalence) and H221Y (7.7% prevalence) in the samples from patients 32 and 76 respectively. Three DRMs identified in bulk sequencing (Y181C in patient 2 at 11 months and T215ST and V108AV in patient 32 at 7 and 16 months respectively, (Chapter 4, Table 4.1) were not observed at any level of prevalence in UDPS. Sequence chromatographs from bulk sequencing were re-examined to confirm the presence of the DRMs recorded. Two DRMs, V118I and E138A, from patients 42 (baseline 2) and 45 (8 months) respectively, were observed in the UDPS data but sequence coverage at these positions was below the cutoff. However, samples from patients 32 (16^{th} month) and 42 (Second baseline) had high numbers of reads removed during quality control process due to low identity with the template.

Bulk sequencing was not successful for the 13^{th} and 9^{th} month samples from patients 45 and 76 (Chapter 4) respectively, however UDPS identified DRMs at both 1-20% (9 DRMs, 4 against current therapy) and greater than 20% prevalence (6 mutations, 2 against current therapy) within the sequenced viral population (Table 5.1). These DRMs included those that were present in the consensus sequences from the earlier sampling times as well as others. For the remaining samples, whose bulk sequencing was successful, UDPS identified an additional 32 DRMs (17 against current therapy) that had not been observed by bulk sequencing (Table 1). The vast majority of these (26 DRMs) were observed at prevalence between 1-20% of the sequenced viral population (Table 5.1).

5.3.3 DRMs against current 1^{st} line ART in Malawi:

Of the mutations detected by UDPS against current 1^{st} line therapy, only K65R was observed, albeit at varying levels, in all of the patients studied (Figure 5.1B). In all patients the prevalence of K65R in the sequenced viral population was observed at less than 20%, however the prevalence in each individual increased

over time. Only in two patients were DRMs to current 1st line therapy observed at prevalence greater than 20% of the sequenced viral population with G190A and Y181C seen in patients 45 and 76 respectively (Figure 5.1B). In both of these cases the DRM was already present in the patient's viral population prior to ARV exposure and was observed to fluctuate during exposure (Figure 5.1B). A similar fluctuating pattern was seen in patient 2 and patient 76. In patient 2, M184V and G190E were not identified by UDPS at baseline or at 14 months yet represented 3% and 2% respectively of the sequenced viral population in the sample at 11 months (Figure 5.1B). In patient 76, K103N was observed at a prevalence of 1.6% at 6 months but was not present at baseline or 9 months. On the other hand, K103N, in patient 45, was not present at 6 months but was present at baseline at prevalence level of 2% and with an increasing prevalence to 7% at 12 months (Figure 5.1B). Two other mutations from patient 45, K101E and M184I, were observed at baseline at prevalence levels of 17% and 1.6% respectively. In subsequent months, K101E was found at prevalence levels decreasing over time (6% and 4% at 6 and 12 months respectively) and M184I was not observed in any subsequent months. Mutation G190A also observed in the 2nd baseline sample from patient 42 with a prevalence level of 6%.

5.3.4 DRMs found against other ARV drugs

Fifteen DRMS against other ARVs were identified across five patients (Figure 5.2). These mutations were mostly polymorphisms that have limited effect on drug resistance. One exception, however, was E138A (conveys resistance to both Raltegravir and Elvitegravir Integrase inhibitors), which predominated the viral population at all sampling timepoints in patient 45. Multiple mutations at position 103 (K103E, K103R, K103T) were also found in three of five patients (Pt 2, Pt 45 and Pt 76).

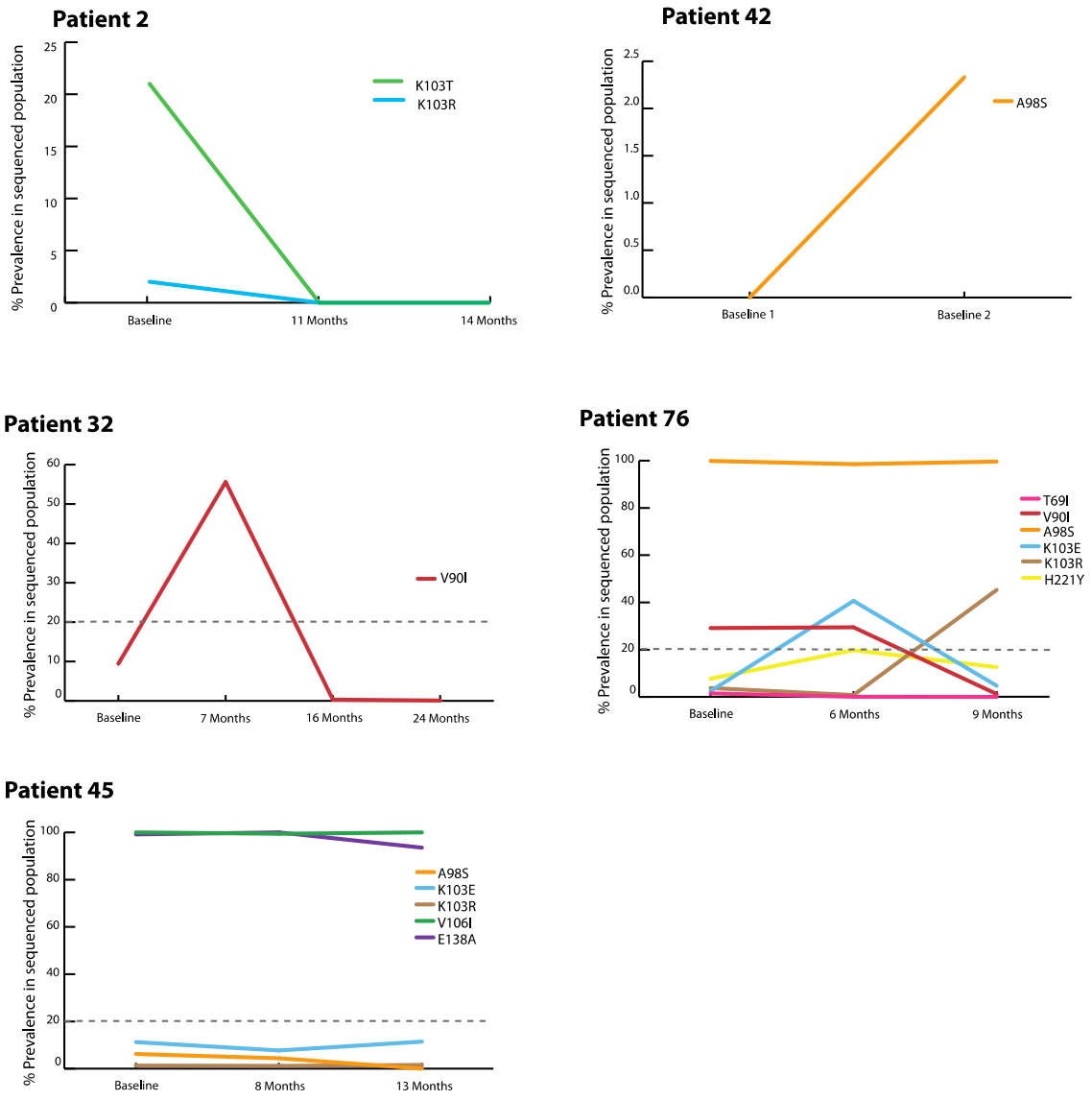


Figure 5.2: DRMs against other ARV drugs with prevalence greater than 1% found in UDPS of reverse transcriptase from five patients

5.4 Discussion:

As can be seen from chapter 3, there was discrepancy in the presence of DRMs in HIV-1 infected individuals on ART. Fifteen individuals showed DRMs of whom three individuals in whom no DRMs were observed at baseline showed emergence of DRMs during ART exposure. Four individuals showing DRMs at baseline showed additional DRMs at subsequent timepoints. The aim here was to get higher sequence depth to see the presence of minor DRMs and address the discrepancy. This study also provides a comparison of bulk sequencing and ultra deep sequencing of patients on ART as the same DNA sample was used for bulk sequencing and UPDS. In this study, a good sequence depth was yielded which helped in identifying minor DRMs that were undetected by bulk sequencing. However, further discrepancies in the presence and prevalence of DRMs were found.

It is shown here that bulk sequencing was unable to detect some DRMs, which are against current first line ART in Malawi (e.g. K65R, K101E). Resistant viruses that make up as little as 1% of the viral population of the circulating virus within an individual have suggested to be clinically important as they can expand rapidly under selective pressure exerted by exposure to HAART (Palmer et al. 2006; Hirsch et al. 2008; Johnson et al. 2008; Metzner et al. 2009; Simen et al. 2009; Halvas et al. 2010; Paredes et al. 2010). In this study, through UDPS, low abundance resistance mutations against current first line therapy were identified through UDPS in the provirus. The DRM, K65R, was not detected in any sample via bulk sequencing, yet it was detected in all patients at low abundance via UDPS. This mutation, according to the Stanford drug resistance database, is responsible for high-level resistance to abacavir (ABA), didanosine (DDI), emtricabine (FTC), lamivudine (3TC) and tenofovir (TFV) and low-level resistance to stavudine, of which, 3TC is being used as a part of current first line therapy in Malawi. This mutation is strongly associated with development of virological failure in subtype C infected individuals (Doualla-Bell et al. 2006; Coutsinos et al. 2011) and (Brenner et al. 2006) showed that K65R emerges more rapidly in subtype C than in subtype B in the presence of tenofovir. Previous studies in Malawi also showed the emergence of K65R in patients with virologic failure in subtype C infected individuals.

(Hosseinipour et al. 2009). Here, although none of the patients showed virologic failure, the presence of K65R at low prevalence may be an indication of future virologic failure. Also, however, it can not be ruled out that the observation of low abundance K65R at various timepoints in all patients could be a result of the previously reported propensity for PCR error at this position in subtype C viruses (Varghese et al. 2010). Mutations at position 190 (G190A or G190E) were shown in four of the five patients via UDPS, at varying levels of prevalence, two of whom showed the presence of this DRM prior to ARV exposure (patients 45 and 76). These DRMs have also been reported in patients with virologic failure who were infected with subtype C in Malawi. (Hosseinipour et al. 2009). These mutations cause high-level resistance to NVP and intermediate resistance to EFV, which are part of first and alternative first line therapy in Malawi.

A strong correlation was found between DRMs observed in bulk sequencing and those at greater than 20% prevalence in UDPS. However, there were further DRMs observed at greater than 20% prevalence that had not been identified by bulk sequencing. This observation is likely to be the result of my approach successfully accessing a greater level of the viral diversity present in the proviral DNA of individuals on ART. Furthermore, in patient 76, a drug resistance related ambiguity at position 181 (Y181CY) was shown at baseline by bulk sequencing while the DRM Y181C was shown in the 6 months sample. One of the aims of employing UDPS was to explore whether the DRM (Y181C) was prevalent at the earlier time point and not seen via bulk sequencing. UDPS found Y181C at all three timepoints with a prevalence > 20%. As this DRM causes high level resistance to NVP, DLV and EFV and found in 55% of patients infected with HIV-1 subtype C from south Malawi who are in virologic failure (Hosseinipour et al. 2009).

It was expected that all DRMs found by bulk sequencing would be detected by UDPS, which would thus resolve the discrepancies found in presence of DRMs present in bulk sequencing (Chapter 4). However, five mutations present in bulk sequencing were not found in UDPS at greater than 20% prevalence. Furthermore, bulk sequencing showed different mutations at different timepoints in patient 32 (Chapter 3) (V90I at baseline, M41MR, T215ST at 7^{th} months, M184I at 14^{th} months), however, none of those mutations were found in sequences returned by

UDPS. This might be due to the PCR reactions that yielded poor amplification and in turn due to the low numbers of reads obtained after quality control analysis on UDPS data. Despite utilizing multiple primary and secondary PCRs to maximize the diversity sampled, PCR bottleneck still cannot be ruled out as a cause of the absent DRMs. However, as PCR amplifications were from proviral DNA it is perhaps not surprising that some differences would be found between amplicons deriving from different aliquots of DNA, as latent viruses may be more diverse than circulating virus. It was similar for patient 42, where there was poor PCR amplification, and only two DRMs were obtained with a prevalence > 1%. Along with this discrepancy between bulk sequencing and UDPS, a greater discrepancy was found by UDPS in the presence and prevalence of DRMs within patients. In UDPS, the mutation at M184V was found in the sample from patient 2, at 11^{th} months but not at baseline or at 14^{th} months. The prevalence of this mutation also fluctuated in patients 45 and 32. This discrepancy in UDPS might be due to the initial PCR bottleneck despite the optimization and use of proviral DNA. The use of RNA as starting material and prior knowledge of viral load would have helped to reveal the discrepancies by using known numbers of template. Thus analysis of further samples from these patients will be helpful in evaluating the clinical significance of the DRMs found in UDPS.

Despite the observation of low and high abundance drug resistant variants in all individuals, these viruses appear to have minimal effects on treatment outcome. Due to geographically isolated setting resulting in logistical difficulties, viral loads are not routinely assessed. The RNA samples and viral load were not available for this study, the comparison of provirus and circulating virus was not possible but is recommended in future studies to determine future treatment strategies. The treatment success for patients in this ART cohort has been measured until recently using WHO stages and CD4 counts, although viral loads are currently being introduced. Using these criteria, patient 2 is defined as having immune failure (CD4 cell count <200 cells/mm^3 after at least 12 months on ART) with the remaining patients showing satisfactory treatment response. Thus, all of the patients were on first line therapy up to 24 months after treatment initiation. While treatment outcome appears to be unaffected in the short term, previous work suggests that HIV rebounds from latently infected cells rather than continuing low-level replication, and

that the low abundance resistant variants could emerge to dominate from the proviral DNA following treatment interruption (Joos et al. 2008).

The detection of low abundance DRMs in this study that are undetected by conventional bulk sequencing extend the findings in various other studies, which, showed that UDPS detected low abundance mutations that are undetected by conventional bulk sequencing (Mitsuya et al. 2008; Le et al. 2009; Codoner et al. 2011). Learning from reported issues in other studies, larger sample volumes are used here (Wang et al. 2007; Le et al. 2009; Varghese et al. 2009), high depth of coverage (Wang et al. 2007; Le et al. 2009; Simen et al. 2009; Varghese et al. 2009) and somewhat conservative cut off of 1% for identifying low prevalence variants. However, it should be noted that the presentation of prevalence in our work as with all other similar studies, should be interpreted as the prevalence observed in the amplified and sequenced viral population as opposed to being a direct measure of the prevalence of a variant in an individual's viral population The discrepancy between these two interpretations has yet to be elucidated, however the development of degenerative primer ID approaches (Jabara et al. 2011) will enable such quantification of these potential biases in the near future.

Thus, it is observed, that in these individuals on continuous antiretroviral therapy, low abundance drug resistant viral variants present in the proviral DNA do not appear to play an immediate role in facilitating the emergence of drug resistance through emergence to dominance. However, further study on subsequent samples from these patients is recommended. The increased sensitivity of deep sequencing could reassure clinicians about the absence of additional genotypic resistance when making clinical management decisions. This is particularly important for the management of subjects with suboptimal adherence.

Chapter 6

Molecular Evolution and genetic diversity of HIV-1 in Karonga District Malawi

6.1 Introduction

The African continent is home to the largest number of all HIV infections in the world with approximately 68% of people living with HIV in Sub-Saharan Africa (UNAIDS 2011a). This continent also seems to be the epicentre of HIV diversity as all subtypes along with some circulating recombinant forms (CRFs) and unique recombinant forms (URFs) are reported to be circulating in this continent (Figure 1) (Ariën, Vanham, and Arts 2007) (Taylor and Hammer 2008). Subtype C, which is responsible for approximately 52% of worldwide infections, currently dominates Southern Africa where it accounts for more than 95% of all HIV infections (Ariën, Vanham, and Arts 2007). Due to the importance of subtype C in this region, over the past few years, this subtype has been extensively studied (Arroyo et al. 2004; Bessong et al. 2005; Bell et al. 2007; Dalai et al. 2009; Thomson and Fernandez-Garcia 2011).

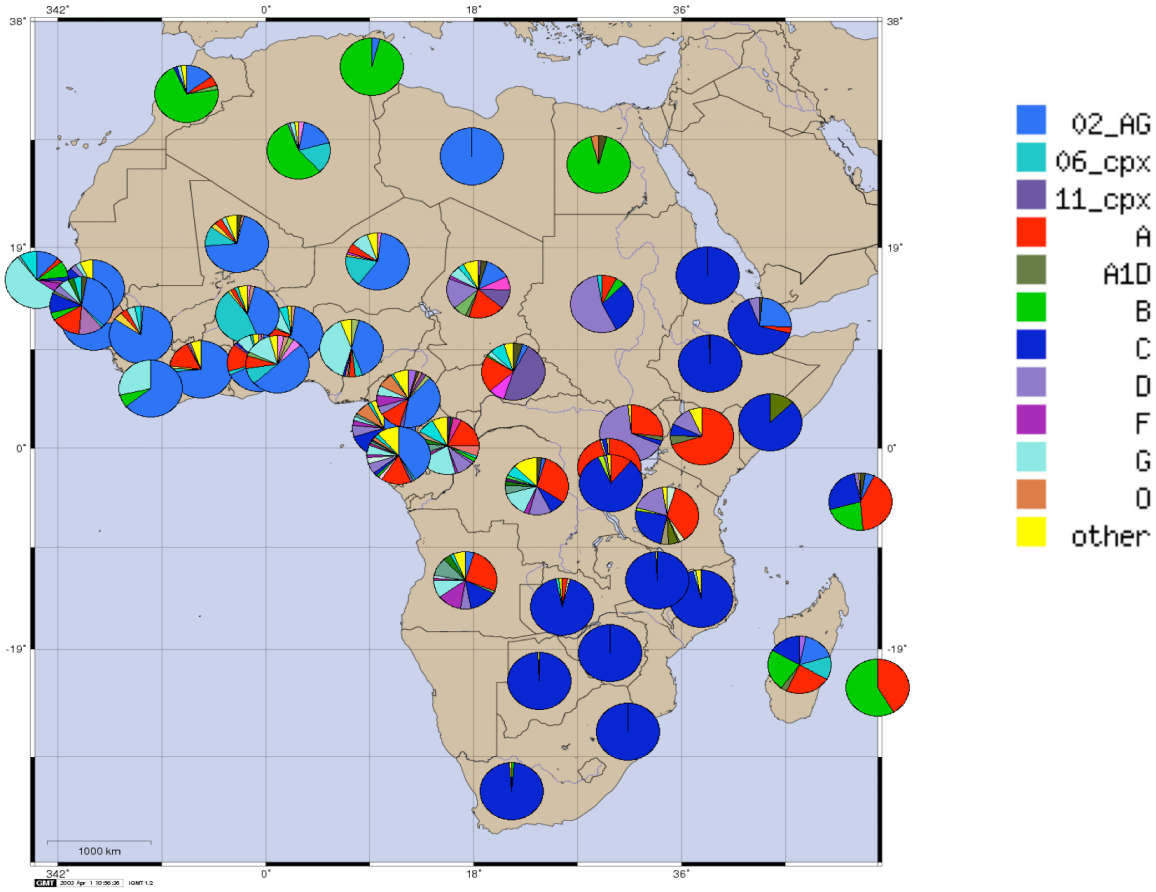


Figure 6.1: HIV-1 Subtype Distribution in Africa (LANL 2012)

The HIV epidemic gradually spread to Sub-Saharan Africa from Zaire (now Democratic Republic of Congo), Rawanda and Uganda where the first few cases of AIDS were reported (Piot et al. 1984; Van de Perre et al. 1984; Serwadda et al. 1985; Walker et al. 2005). It has been estimated that subtype C was first introduced in Ethiopia in the early 1980s (Abebe et al. 2001). The first documented case of subtype C infection comes from a sample taken from a Malawian patient in 1983 (McCormack et al. 2002). It has been suggested that the most recent common ancestor of subtype C dates back to the late 1960s (Travers et al. 2004). This is consistent with the theory that HIV-1 group M originated in the late 1930s (Korber et al. 2000) and with estimates of the origins of subtype B in the range from 1915 to 1945) (Korber et al. 2000; Lukashov and Goudsmit 2002).

One contributing factor to subtype C predominance has been suggested to be the efficiency by which subtype C is transmitted from one person to other (Arien et al. 2005). The unique genetic characteristics of this subtype that differentiates it from other subtypes are i) the presence of an extra NF-κβ binding site in the LTR along with premature truncated proteins (Tat and Rev) ii) a 15-bp insertion at the 5' end of the *vpu* reading frame and iii) a prematurely truncated Rev protein (Peeters and Sharp 2000; Shankarappa et al. 2001; Huang, Giesler, and Bremer 2003). It has been suggested that the presence of the extra NF-κβ site may enhance gene expression and hence the rapid expansion of subtype C and its pathogenesis compared to other subtypes. The enhanced gene expression alters the transmissibility and pathogenesis of subtype C making it an easily transmissible virus. (Jeeninga et al. 2000; Shankarappa et al. 2001). The 15 bp insertion at the 5' end of the *vpu* region along with the prematurely truncated Rev protein may influence viral gene expression and alter the virulence of subtype C (Tatt et al. 2001). Another reason suggested for the increased level of viral transmission in subtype C is that a substantial proportion of the population of HIV-1 infected individuals have high HIV RNA levels (Novitsky et al. 2010).

Malawi is one of the countries that lie in the centre of HIV epidemic with a prevalence of 11% (UNAIDS 2010). Along with subtype C, two other subtypes (A and D), three recombinant forms (AC, AD and DC) and one unclassifiable strain have been reported in Malawian population; however, subtype C is found to be

dominant subtype in Malawi (McCormack et al. 2002; McCormack et al. 2003; Petch et al. 2005; Hosseinipour et al. 2009). The available data on subtype C in Karonga suggest that the Karonga District might have had multiple introductions of HIV-1 with limited spread in the early 1980s and later diversified (McCormack et al. 2003). The molecular epidemiological data on Malawian sequences is mainly focused on *env* and *gag* genes (McCormack et al. 2002; McCormack et al. 2003; Kwiek et al. 2008). However, nowadays in Malawi, protease (PR) and reverse transcriptase (RT) regions of *pol* gene are being sequenced for genotypic drug resistance testing and a large amount of data is now being generated (Hosseinipour et al. 2009). The *pol* gene holds sufficient genetic variability to permit phylogenetic reconstructions of transmission (Hue et al. 2004).

The action of antiretroviral drugs and the host immune system creates selective pressures that may lead to mutations in favour of HIV replication. Thus, HIV replication may continue during antiretroviral therapy resulting in viral evolution along with the emergence of drug resistance (Napravnik et al. 2005; Shi et al. 2010). Measurement of positive selection at sites conferring drug resistance can be used to study the mechanism of evolution of drug resistance and to discover novel mutations. The selection pressure operating on a gene is measured by calculating the ratio of observed non-synonymous (nucleotide mutations that change the encoded amino acid, dN) over synonymous mutations (mutations that do not change the amino acid, dS) (Nei and Gojobori 1986; Li 1993). If the ratio of dN and dS (ω) is greater than 1, it is considered that the gene is under positive selection. Initially the methods measuring selection pressure were either distance based (Yang et al. 2000) or maximum likelihood based (Goldman and Yang 1994). Subsequently (Pond and Frost 2005b) classified the methods into three approaches:

1) Single Likelihood Ancestral Counting (SLAC): These methods are based on counting methods that employ either single most likely ancestral reconstruction, weighing across all possible ancestral reconstructions or sampling from ancestral reconstructions (Suzuki and Gojobori 1999; Suzuki 2004)

2) Random Effect Likelihood (REL): These methods model the differences in dN and dS according to a pre-defined distribution and the selection pressure is inferred using empirical Bayes approach (Nielsen and Yang 1998)

3) Fixed Effect Likelihood (FEL): These methods directly measure dN and dS rates at each sites (Suzuki 2004). SLAC methods are recommended for larger datasets (over 40 sequences), as they are very fast and conservative methods (Pond and Frost 2005b). REL methods give similar results as counting based methods with the exceptions of the extent to which the test statistics is conservative or liberal but are computationally complex, while FEL is recommended for intermediate sized datasets (20-40 sequences)(Pond and Frost 2005b).

Recent research carried out in the Karonga District, Malawi aims to understand the risk factors, trends and dynamics of HIV and how treatment of HIV alters HIV transmission. In brief, the survey area is divided into 21 reporting groups that are visited in sequence once per year and annual HIV sero-surveys in adult members of this study population have been undertaken since 2007. Blood samples that were collected from HIV positive individuals between 2007 and 2010 were employed in this work to sequence PR and RT genes (Chapter 2 and 3). The aims of this chapter are to study diversity of HIV-1 in the Karonga population using *pol* sequences and to explore differential selection pressure operating on RT in antiretroviral naïve and exposed individuals.

6.2 Materials and Methods:

6.2.1 Sequences and subtyping

A combined dataset of 156 reverse transcriptase (RT) sequences from Karonga was generated which consisted of 132 sequences generated from chapters 2 and 3 along with 8 sequences from seroconverters and additional 16 unpublished sequences. The seroconverter samples were collected within one year of seroconversion. The seroconverter sequences and unpublished sequences in this dataset were submitted for analysis of Drug Resistance Mutations (DRMs) to the Stanford HIV Drug Resistance Database (http://hivdb.stanford.edu). The protease sequence dataset comprising 47 sequences from chapter 2 was also utilised.

For subtyping of RT and PR sequences, subtype reference alignments for both PR and RT were downloaded from LANL (http://www.hiv.lanl.gov/content/index). The reference alignment contained approximately four representative sequences of each subtype. Alignments of both RT (132 Sequences) and PR (47 Sequences) from the Karonga dataset along with subtype reference sequences were prepared in MacClade 4.0.8 (Maddison and Maddison 2005.). Phylogenetic trees of both PR and RT sequences were reconstructed under the GTR+ gamma model of DNA substitution implemented in RAxML 7.0.3 with all parameters optimized by RAxML (Stamatakis 2006). Confidence levels in the groupings in the phyogeny were assessed using 1000 bootstrap replicates as part of the RAxML phylogeny reconstruction. A bootstrap value equal to or greater than 70% was considered significant. Recombination patterns in RT sequences were identified using the jumping profile Hidden Markov Model using jpHMM (http://jphmm.gobics.de/jphmm.html) (Zhang et al. 2006; Schultz et al. 2009).

6.2.2 Phylogenetic analysis of subtype C RT sequences from Malawi.

An alignment of 286 RT sequences was created which comprised of, i) Subtype C sequences from Karonga dataset including sequences generated from the patients on ART (more than one sequence/individual) (Chapter 3) iii) Subtype C sequences from another study from southern Malawi (22 Sequences) (Petch et al. 2005) and iv) Subtype C sequences representing India (14 sequences) and Africa (17 sequences). The sequences in India and Africa were downloaded form LANL

website. Although more sequences from African continent were downloaded, only 17 were taken further due to the quality and length of sequences. A phylogenetic tree was reconstructed by using RAxML as described above using the LANL subtype C (www.lanl.gov) ancestral sequence as outgroup.

6.2.3 Analysis of selection pressure operating on RT and protease:

For the analysis of selection pressure operating on the RT gene, datasets of drug naïve and drug exposed sequences were generated for both subtype B and C (amino acid positions 13 to 257 of RT gene (HXB2 numbering). For subtype C, two datasets were generated, one comprising of sequences from Karonga (128 drug naïve 105 exposed) with the second dataset generated using subtype C RT sequences downloaded from LANL HIV database (126 naïve and 394 exposed) (Global Subtype C- Table 6.1). Similarly, for subtype B, a sequence dataset of drug naïve (492 sequences) and drug exposed (500 sequences) RT sequences was downloaded from NCBI. Manual examination of the sequence information of individual sequences was done in order to confirm the drug naïve/exposed status of the sequences. Sequence alignments were generated for all datasets using MacClade.

Table 6.1: Sequence datasets generated for selection pressure analysis. Drug naïve and drug exposed sequences are the sequences generated from sample collected prior to and after antiretroviral treatment respectively

	Reverse Transcriptase		Total
	Drug Naïve	Drug Exposed	
Karonga dataset (Subtype C)	128	105	233
Global Subtype C	126	394	520
Global Subtype B	492	500	992

The datasets were submitted to HyPhy online web server (www.datamonkey.org) (Pond and Frost 2005a; Delport et al. 2010). A codon based maximum likelihood method - Single Likelihood Ancestral Counting (SLAC) (Pond and Frost 2005b) was used with General Reversible Model (REV) model as model of DNA substitution. The SLAC was chosen because it is fast and conservative method

and is suitable for analysis of larger datasets (Over 40 sequences). For all three datasets, drug naïve and exposed datasets were submitted separately as well as pooled (naïve plus exposed). They were pooled to see the effect of the event of fixation of mutations in one or both of these datasets. The alignment restrictions for the SLAC method on the online web server allow only 500 sequences. Therefore, for global subtype B sequences, ten datasets of 500 sequences were generated by random selection of 250 sequences each from drug naïve and drug exposed sequence datasets.

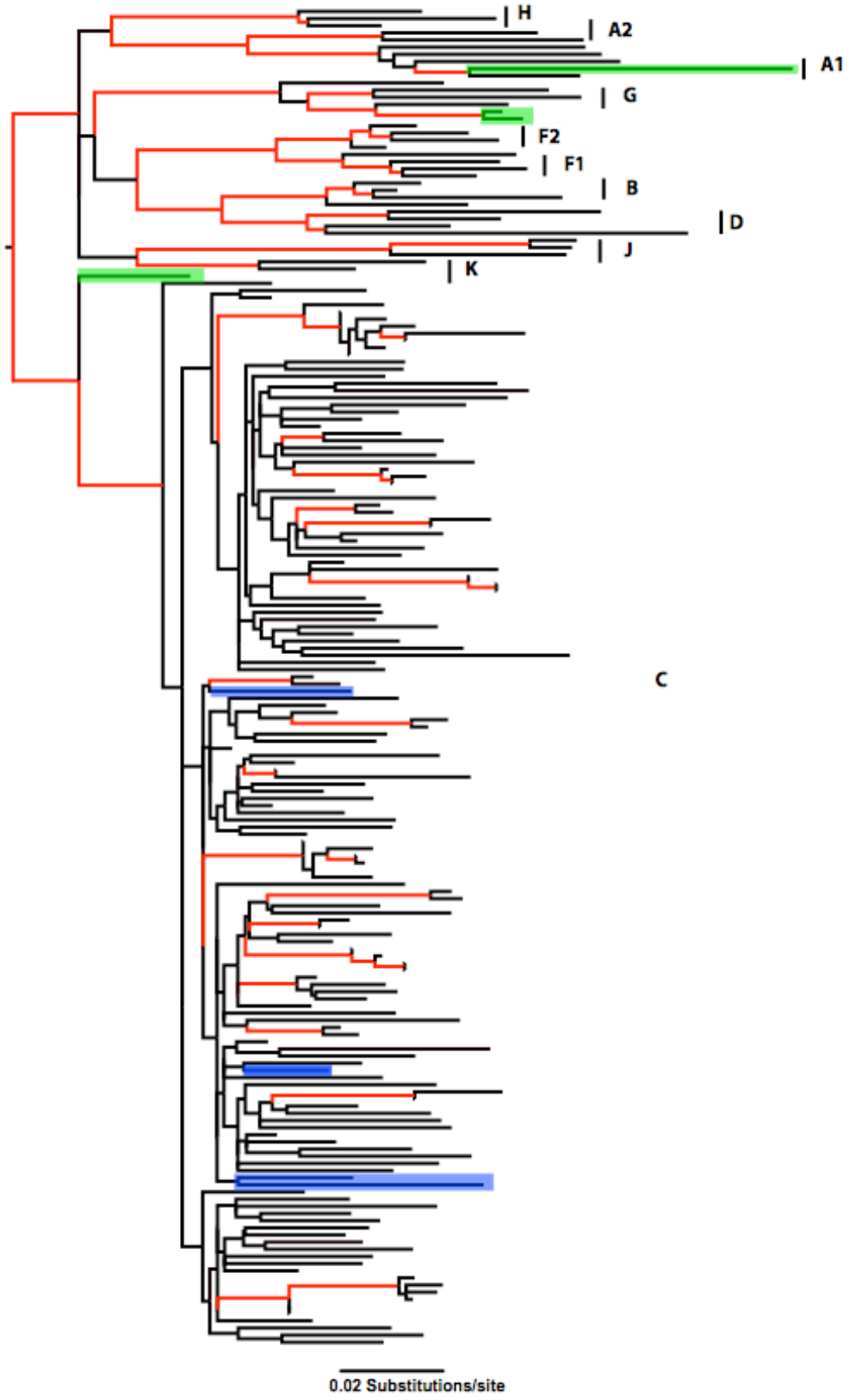
6.3 Results:

6.3.1 Subtyping of PR and RT gene regions in Karonga dataset:

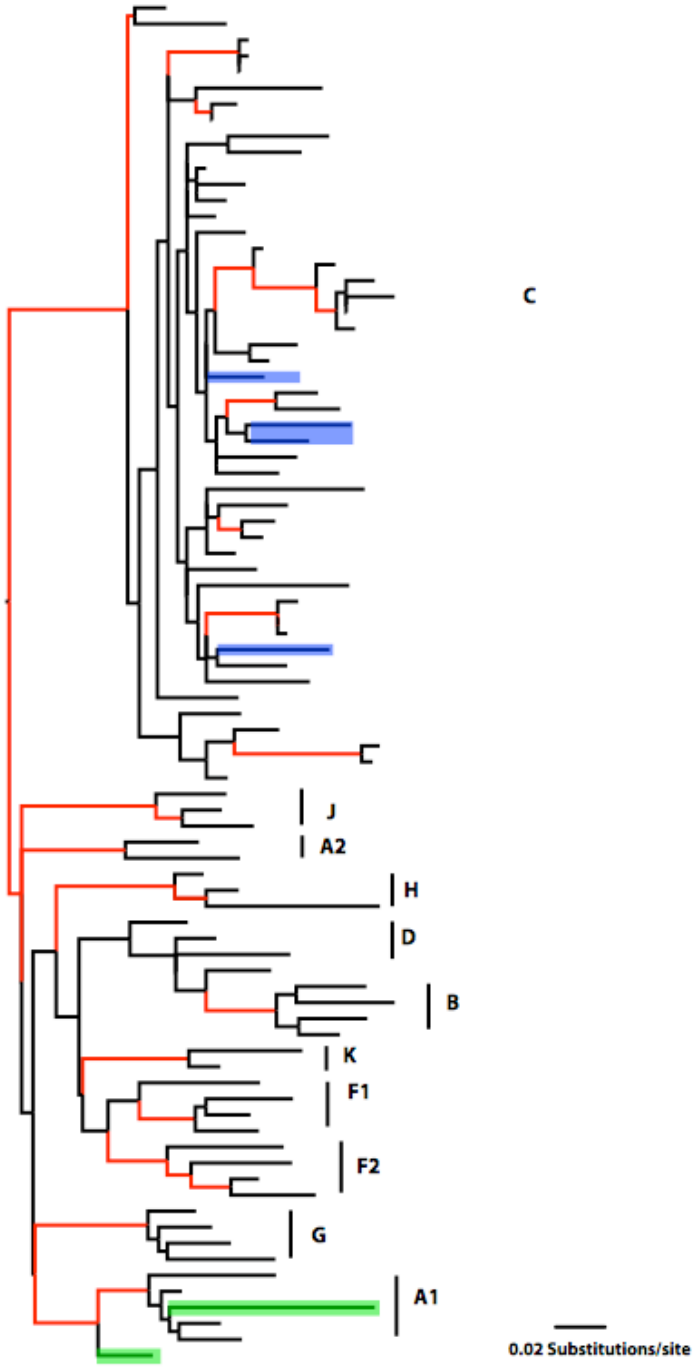
Of the 156 RT sequences from Karonga, two sequences (54026KarongaMW and 54028KarongaMW) clustered with subtype G, one sequence (Pt50_6) clustered with Subtype A1 while one sequence (Pt3) did not cluster with any of the reference sequences (Figure 6.2A). Upon submitting these non subtype C sequences to REGA HIV-1 automated subtyping tool, one RT sequence (Pt50_6) was predicted to be recombinant A1_D while all three sequences from one patient (Pt3) were predicted to be recombinant A1_C (Table 6.2). Of 47 PR sequences, two (54332a and 54535a) sequences clustered with subtype A1 (Figure 6.2B). RT sequence was available for one of these two sequences (54535a), which clustered with subtype C (Table 6.2, Figure 6.2A).

Table 6.2: Non Subtype C sequences found in Karonga population. The recombination break points were obtained from REGA HIV-1 automated subtyping tool. For Pt3, three sequences were obtained from samples collected at three timepoints. * Sequence not retrieved.

Sequence Name	HIV-1 Subtypes			
	PR	RT	Intersubtype recombination break points (positions according to HXB2)	
54026	n/a*	G		
54028	C	G		
Pt50_6	n/a*	A1_D	2586-2846 A1	2847-3323 D
Pt3_6	n/a*	A1_C	2586-2759 A1	2760-3323 C
Pt3_9	n/a*	A1_C	2586-2899 A1	2900-3323 C
Pt3_12	n/a*	A1_C	2586-2765 A1	2766-3323 C
54332	A1	n/a*		
54535	A1	C		



A



B

Figure 6.2: Maximum likelihood trees for A) reverse transcriptase and B) protease genes from HIV-1 positive individuals from Karonga. The branches in red color represent the bootstrap support of 70% or higher. The branches highlighted in light blue are subtype C reference sequences. Non-subtype C sequences from this the Karonga dataset are shaded in light green

.

6.3.2 Phylogenetic reconstruction of the Karonga subtype C RT sequences

The RT sequences from Karonga were spread across the phylogeny reconstructed (Figure 6.3). Sequences from the Karonga dataset were found to be intermixed with subtype C sequences from other African countries. All sequences from India clustered together along with a sister cluster consisting sequences from Karonga and southern Malawi but with no support from bootstrapping (Figure 6.3). Three large clusters were seen although with no bootstrap support (Clusters highlighted as I, II and III in Figure 6.3). Multiple sequences generated from samples collected at different timepoints from individuals on ART clustered together with high bootstrap support. For all other individuals included, multiple sequences from a few individuals were found to be clustering with sequences from other individuals with high bootstrap support (labelled as clusters 1 through 14 in Figure 6.3). Known epidemiological links were available only for clusters 2, 6, 8 and 13. In cluster 2, which is comprised of sequences from four individuals, two come from the same reporting group while two other belong to an adjacent reporting group indicating they may therefore live in relatively close geographical proximity. While in cluster 6 which is made up of sequences from two individuals, both belong to the same reporting group. Similarly sequences generated from both individuals in cluster 8, come from the same reporting group. Finally in cluster 13, three out of five sequences, are from individuals who belong to the same reporting group.

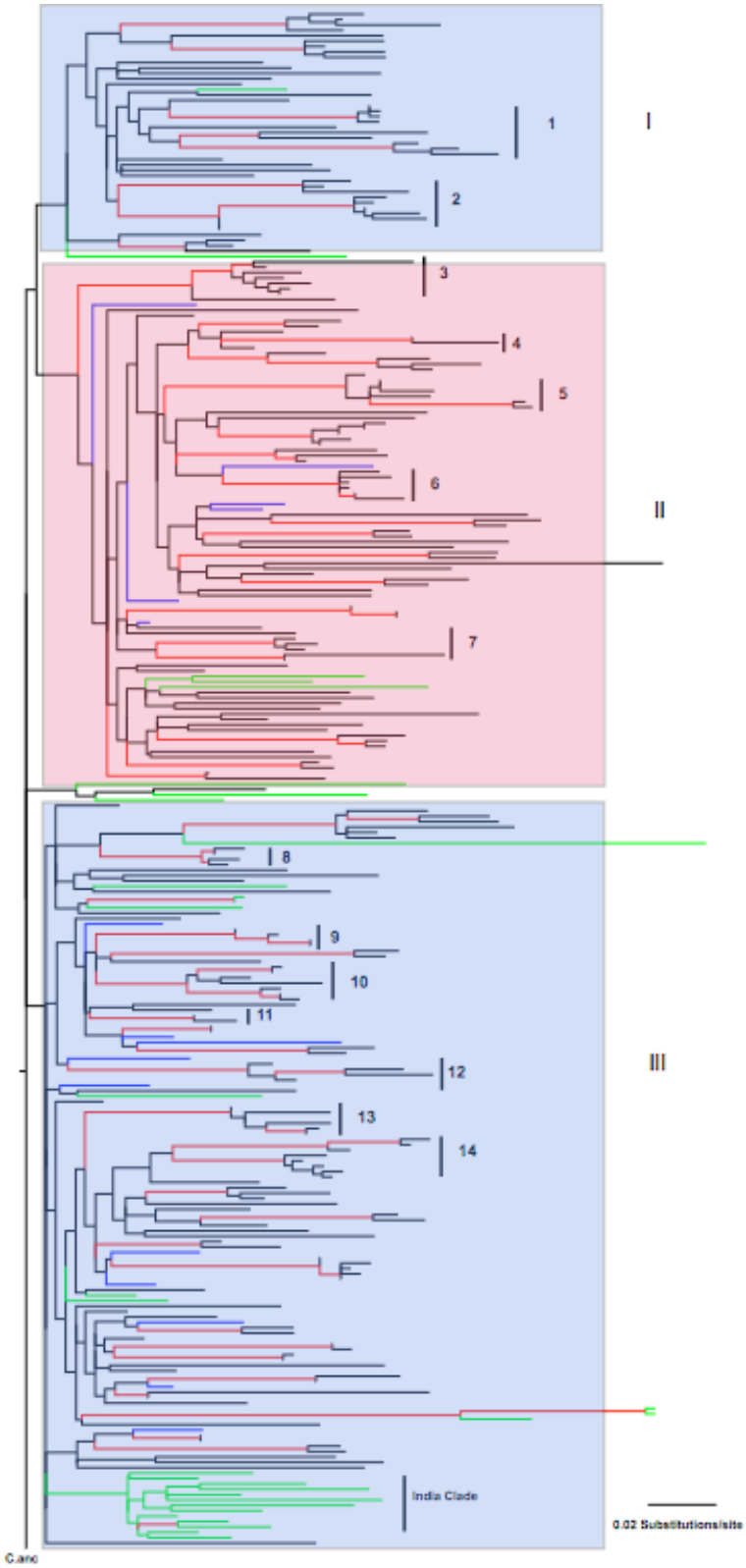


Figure 6.3: Maximum likelihood tree for HIV-1 Subtype C RT sequences. Branches in red color represent the bootstrap support 70% or higher. Sequences from Karonga dataset are colored in black; sequences from another study from Malawi are colored in blue (Petch et al. 2005) while subtype C sequences representing Africa and India are colored in Green. The three distinct clusters of Karonga sequences are highlighted (I, II and III). The cluster numbers are given for each cluster where the sequences from different patients clustered together with bootstrap support 70% or higher.

6.3.3 Selection pressure analysis on RT gene of subtype B and subtype C.

The positively selected sites found in RT of subtype C (Karonga and Global) and subtype B (Global) are outlined in Table 3. In subtype C, the drug naïve Karonga dataset showed the presence of 9 positively selected sites while the exposed showed presence of 5 sites that were positively selected. The combined naïve and exposed dataset showed 6 positively selected sites. While in the global subtype C dataset, 10 and 11 positively selected sites were found in naïve and exposed datasets respectively. The combined dataset of global subtype C showed 14 positively selected sites. None of the positively selected sites in subtype C datasets were associated with drug resistance; however, a number of positively selected sites with their reported association with HLA were found (Sites in bold in Table 6.3). In the combined dataset of naïve and exposed (N+E) of global subtype C, four additional sites (102, 123, 158 and 166) were found that were not present in individual naïve or exposed datasets.

In the subtype B dataset, 9 and 11 positively selected sites were found in drug naïve and exposed datasets respectively while in the combined dataset of naïve and exposed showed 18 positively selected sites. In subtype B, two positively selected sites (98 and 190) were found that confer drug resistance. Similar to subtype C, three additional sites (58, 117 and 182) were found in N+E dataset (Table 6.3).

Table 6.3, HXB2 amino acid numbers of positively selected sites in different RT sequences. Cells shaded in N+E datasets are the additional sites found that are not present in either naïve or exposed individual datasets. Numbers in brackets in N+E Subtype B dataset are the frequency of the positively selected site in ten datasets. Positions highlighted in red are the sites that confer drug resistance. Sites in bold are amino acid sites that are sites associated with HLA (Moore et al. 2002; Chui et al. 2007).

Subtype C						Subtype B		
Karonga			Global			Global		
Naïve	Exposed	N+E*	Naïve	Exposed	N+E*	Naïve	Exposed	N+E*
23	23	23	36	36	36	14	14	14 (3)
35	36	36	**121**	38	48	73	35	35 (5)
48	**173**	48	**135**	111	**102**	77	45	45 (3)
123	**245**	**123**	**162**	113	**121**	87	77	58 (6)
162	250	**173**	164	**162**	**123**	110	97	73 (6)
164		**174**	**173**	**165**	**158**	152	98	77 (9)
173			**174**	**173**	**162**	153	159	97 (10)
174			**207**	**174**	**165**	175	175	**98 (5)**
245			**211**	**200**	**166**	186	**177**	110 (10)
			245	**207**	**173**		186	**117 (3)**
				245	**174**		190	152 (5)
					200			153 (4)
					207			159 (6)
					245			175 (10)
								177 (2)
								182 (5)
								186 (9)
								190 (9)
9	5	6	10	11	14	9	11	18

6.4 Discussion

Studies on *gag* and *env* genes revealed that subtype C has emerged as the most prevalent HIV-1 strain in Karonga District (McCormack et al. 2002). The proportion of subtype C found in sequence datasets in the region increased from 55% in early 1980s to 90% by late 1980s (McCormack et al. 2002) and to 94% by 2009 (Seager, Unpublished). A high proportion of subtype C has also been reported in countries neighbouring Malawi such as Tanzania (Arroyo et al. 2004), Zambia (Trask et al. 2002), Mozambique (Abreu et al. 2008; Lahuerta et al. 2008) and Zimbabwe (Dalai et al. 2009). The high prevalence of subtype C in Karonga can be due to several factors, one being that this subtype is the most prevalent subtype in the countries surrounding Malawi (Trask et al. 2002; Arroyo et al. 2004; Abreu et al. 2008; Lahuerta et al. 2008) thus limiting the possible introduction of new subtypes. Another reason is strong founder effect, similar to the transmission of subtype B in Europe (Paraskevis et al. 2009) and America (Junqueira et al. 2011). The available data on subtype C in Karonga suggest that the initial subtype C in Karonga district might have had multiple introductions of HIV-1 with limited spread in the early 1980s and later diversified (McCormack et al. 2002). The initial number of subtype C viruses introduced in Karonga population are likely to have been transmitted through unprotected sex, which is the primary mode of HIV transmission in Malawi. It has also been reported that subtype C viruses are less virulent than other subtypes (Ariën, Vanham, and Arts 2007) resulting longer asymptomatic periods and thus providing an opportunity for efficient transmission and making subtype C as the dominant subtype in the population. The presence of non-subtype C in northern Malawi has been reported as early as 1982, although at low prevalence (McCormack et al. 2003). In this study it has been found that 97% of sequences belonged to subtype C along with the presence of subtypes A1 and G. This is the first report of subtype G in Malawi, although it has been previously reported in neighbouring Zambia (Trask et al. 2002). Subtype D, which was previously reported in the Malawian population was not found in this study. It has been reported that the local epidemics are dominated by one or few subtypes, e.g. subtype B in Caribbean (Nadai et al. 2009) and Bolivia (Guimaraes et al. 2012) or subtype A1 and B in Latvia (Balode et al. 2004).

Phylogenetic analysis of subtype C sequences showed that all sequences from Africa were distributed across the tree with no specific cluster representing any specific geographical region, Sequences from India formed a separate cluster, although with low bootstrap support similar to the results obtained in a previous study (Agnihotri et al. 2006). Previous work using *env* and *gag* gene fragments from Karonga district identified four clusters (McCormack et al. 2002) and it was hypothesised that one cluster may have been introduced in Karonga district from a single event sometime in the early 1980s (McCormack et al. 2002). This subsequently spread and accounts for over 40% of infections seen in the district in the late 1980s. This group ('Malawi cluster') appeared to be largely restricted to the sequences from Malawi. The Malawian cluster was still present in *gag* phylogenetic tree in the sequences generated from 1990 and 2000s, while in the *env* gene, the cluster was split into two (Seager, unpublished). In both gag and env genes, very few sequences from different countries were found in Malawian cluster (Seager, unpublished). In this study, the sequences from Karonga were distributed throughout the tree with three large clusters (although with no significant bootstrap support). Within those three large clusters, small distinct clusters of multiple sequences from single individuals were observed. Few clusters (14) with sequences derived from more than one individual have been found with some of them having known epidemiological links. However, considering geographical isolation of the Karonga district, it can be possible that multiple transmissions from one individual might have occurred.

The selection analysis showed the presence of positively selected sites in both subtype B and subtype C datasets. However, none of the positively selected sites in subtype C sequences (either drug exposed or drug naïve) are associated with drug resistance. It seems that the reverse transcriptase gene in Karonga dataset is not driven by antiretroviral therapy (yet at least). Similar results were obtained in subtype C by (Varella, Schrago, and Zalis 2009) who analysed sequences from drug naïve and treatment failure individuals separately by using maximum-likelihood based methods. They showed that none of the sites under positive selection from drug naïve individuals were associated with drug resistance. However, in the same study, individuals failing treatment were showing presence of

positively selected sites associated with drug resistance. Individuals in our work are on ART for relatively short period of time (up to 24 months) and are responding well to the treatment with none of these individuals showed treatment failure. Further sampling from these individuals, and from individuals failing therapy, with sequences from subsequent timepoints would be helpful in studying the selection pressure operating on RT in the Karonga population. In the global subtype C sequences also did not show any positively selected sites, this might be due the fact that the sequences are from the samples collected at the earlier stage of the epidemic. Furthermore, the positively selected sites in the Karonga sequences are different than global subtype C datasets. Although the reason behind this is unknown but generally the local epidemics are dominated by certain strains that are introduced and later diversified thus generating a 'local' pool of variants. Furthermore, this study is based on sequences generated from provirus, which might not represent the currently circulating virus and hence might not be under drug pressure.

It has been reported that in subtype B, a positive selection at codons conferring drug resistance was found only in drug treated individuals and positive selection detected in drug naïve individuals was unrelated to drug resistance (de S. Leal, Holmes, and Zanotto 2004), however, it is not clear what stage of drug treatment the exposed individuals were. Similarly, in this study, we found two positively selected drug resistance sites in subtype B drug exposed datasets. A larger study, aimed at studying positive selection in 40,000 HIV-1 subtype B sequences showed that, out of 110 positively selected sites, twenty corresponded to known drug resistant mutations (Chen, Perlina, and Lee 2004). As antiretroviral drugs were first introduced in developed countries in 1990s, where subtype B is predominant, this subtype is being exposed to drug pressure since then. On the contrary, for countries such as Malawi, where subtype C is prevalent, the antiretroviral drugs were introduced relatively recently (in 2000s) (Brenner 2007; McGrath et al. 2007). This might be the reason in this study to see the positively selected sites in drug exposed individuals from subtype C are unrelated to drug resistance. In this study, in all three datasets, positively selected sites that are reported to have association with HLA have been found. These sites are the

epitopes of cell-mediated immunity (Moore et al. 2002) indicating the selection pressure is directed by immune pressure.

The sequence generated during the early seroconversion stage represents the currently transmitted strain. These viruses that are transmitted, could be originated from either ARV treated individuals or drug naïve individuals. One of the aims of this study was to study the transmission of drug resistance by analysis of recent seroconverters. None of the sequences from the recent seroconverters showed the presence of drug resistance indicating that newly transmitted viruses in Karonga don't carry resistance. However, the sample number for the recent seroconverters was very low so it is difficult to draw such conclusions regarding transmitted drug resistant strains in this instance.

From this study, it can be concluded that subtype C is the dominant subtype in Karonga and may continue to dominate the Karonga population, as the chances of introduction of new subtype are limited in this region because of predominance of subtype C in the countries surrounding Malawi. However, an increase in the prevalence of intra-subtype recombinants is possible because of presence of other subtypes in the population although at low levels. This study also suggests that selection pressure on HIV in Malawi is not yet driven by antiretroviral therapy. However, this analysis is based on proviral DNA thus limiting the study results to the proviral DNA. HIV-1 RNA sequence represents the circulating virus that is constantly exposed to immune and drug pressure, a comparative analysis of selection pressure on sequences generated from RNA and proviral sequences is recommended.

Chapter 7

General Discussion

The worldwide epidemic of HIV remains a challenge despite the fact that the first case of HIV/AIDS was identified more than 25 years ago and it has been more than 10 years since the first use of HAART. HAART has revolutionised HIV treatment and control as AIDS related deaths have dropped in recent years due to the introduction of more potent drug cocktails and greater access to antiretroviral therapies. However, the development of drug resistance remains a major obstacle in the success of HIV treatment. The majority of HIV-positive individuals reside in Sub-Saharan Africa and Asia where antiretroviral treatment started relatively recently (Keiser et al. 2008). Various factors influence the drug resistance in HIV-1 such as the availability and continuity of drug supply, the ability of ART regimens to suppress replication completely, the fitness of the resistant variant, and adherence to treatment (Clavel and Hance 2004). Therefore the efforts to prevent drug resistance should be focused on these factors.

ART became available in the Karonga District in 2005, and was comprised of only RT inhibitors as a first line therapy. Subtype C is the most prevalent subtype in the district, consistent with the previous reports on HIV-1 in Malawi (McCormack et al. 2002; McCormack et al. 2003; Petch et al. 2005; Hosseinipour et al. 2009). Despite the fact that other subtypes were also found, (and this study describes the first report of subtype G present), they are far fewer in number and thus it is largely the interaction of HIV-1 subtype C with available therapies that are the focus of this project. The main aims of this project at the outset were threefold; 1) to sequence RT from approximately 1000 and PR from 50 HIV-1 infected individuals in order to monitor the baseline drug resistance in the Karonga District, 2) to genotype newly infected individuals to detect possible transmission of drug resistant strains and 3) to monitor the development of drug resistance in individuals attending ART clinic. As RNA stabilization was not feasible considering the scale of the sampling programme involved and proviral DNA in newly infected individuals could be a reasonable measure of active virus present, it was decided to genotype the provirus for the project. The initial part of this study was based on conventional bulk sequencing of RT and PR and chapter 3 showed the presence of DRMs in the drug naïve population at a rate of 15%. However, additional drug resistance mutations were present in clones that were sequenced from selected individuals and for some

individuals where no DRMS were detected in the consensus sequence, DRMS were present in clones from the same individual. This underlined the importance of sequencing multiple variants in order to detect minor drug resistant variants and showed the limitations of bulk sequencing for detecting DRMs from provirus at least. Furthermore in chapter 3, using bulk sequencing to detect DRMs in individuals on ART, some individuals showed DRMs at baseline but no DRMs at subsequent timepoints further pointing to the limitations of bulk sequencing as this technique yield sequence from only one virus.

Due to this limitation, further bulk sequencing was not performed. To detect low abundant variants, traditionally, methods such as single genome amplification, and real-time PCR are used. However, these methods are labour intensive, time consuming and restricted to the detection of few variants. The study was then focused on establishing UDPS technology in the laboratory and using it to resolve the discrepancy in the presence and absence of DRMs found by bulk sequencing at different timepoints in individuals on ART. UDPS technology, which has wide range of applications in the field of genetics, is now being used to study viral quasispecies and drug resistance. PCR is used to amplify the templates that are subjected to UDPS. Though PCR optimization was carried out in order to maximize the diversity of templates subjected to UDPS, this work suggests that it is the quality of the initial amplification of templates that is of major importance for characterizing genetic diversity of HIV-1 quasispecies rather than the number of PCR steps. For those samples that had weak PCR amplification, fewer and shorter sequences were returned by UDPS than samples with strong amplification, and many more reads had low identity to the consensus sequence. Despite multiple first and secondary PCRs being used to maximise diversity, there were some DRMs absent in the UDPS data that were present in the consensus sequence generated through bulk sequencing. For some of these PCR bottleneck cannot be ruled out as a cause despite multiple PCR products being combined. To overcome this, a number of approaches can be used i) mixture of two primer sets in the outer PCR reaction (Messiaen et al. 2012) ii) use of quantitative real-time PCR to measure directly the concentration of DNA followed by use of bioinformatics applications to guide the dilution of DNA to use for

PCR (Butler et al. 2009) and more accurate method iii) use of unique primer id for each template subjected to the initial PCR reactions (Jabara et al. 2011)

This study showed the presence of DRMs in RT and PR genes in drug naïve individuals in Karonga population. Prior to the availability of ART in the Karonga district since June 2005, Nevirapine was available for the PMTCT (Prevention of Mother to Child Transmission) and full ART would have been available to number of individuals. Thus some individuals that are assumed to be drug naïve might have been exposed to drugs (either they did not disclose the fact or they were not aware of having ART, e.g. in the case of PMTCT). This might be one of the possible reasons of finding DRMs in naïve individuals but it cannot be ruled out that these DRMs in drug naïve individuals may also represent natural polymorphisms or transmitted strains. The recent seroconverters included in this study did not show any DRMs but they are a small number of individuals. However, given the lack of confidence in knowing absolutely if an individual was indeed drug naïve and given the lack of DRMs in newly infected individuals DRMs found in drug naïve individuals cannot be taken as strong evidence for transmission of drug resistant strains in Karonga District. This study suggests that selection pressure on HIV in Malawi is not yet driven by antiretroviral therapy. Further follow-up of these individuals can reveal a better picture of the drug resistance in these individuals. Natural polymorphisms may affect ART in different ways in different populations (Martinez-Cajas et al. 2008) hence it is difficult to know the effect of the polymorphisms on drug resistance in the Karonga population. This study also showed that DRMs in RT associated with current antiretroviral therapy (i.e. d4T, 3TC and nevirapine) were found in 20% of the individuals (both drug naïve and drug exposed). Furthermore several mutations that cause resistance Nevirapine are reported in this study (A98G, K103KN, Y181C, G190A). This suggests that the resistance is acquired either during current therapy or due to possibility of receiving Nevirapine as PMTCT before they enrolled in the ART, although transmission remains a possibility.

As ART is becoming increasingly available in resource-limited settings, the need for genotyping tests and viral load monitoring is increasing (Calmy et al. 2007). Treatment failure can be monitored by viral load assays. Regular viral load assessment allows the early recognition of treatment failure and an associated

change of therapy in time, which can avoid the accumulation of resistance mutations and the development of cross-resistance. Along with viral load assays the genotyping in order to assess drug resistance is helpful for the selection of optimal therapy. One study in southern Malawi showed a high rate of drug resistance in individuals showing treatment failure (Hosseinipour et al. 2009). The results of this study, based on sequencing of viral RNA, were supported by the availability of CD4 count and viral load monitoring and underlined the need for improvement of monitoring ART failure in resource poor setting (Hosseinipour et al. 2009). However, in the case of rural Malawi such as Karonga District viral load monitoring is not a routine practice. The lack of storage and transport capacity to keep RNA stable hampers viral load testing. Instead, treatment failure is measured by clinical progression or immunologic failure. Although DRMs were found in the Karonga population, it is unlikely that this data can be used immediately in changing the treatment strategy of any individual because of the limited availability of alternate regimens permits very little change based on genotyping results. Thus clinicians will run out of choices if resistance is developed to an alternative drug. Furthermore the effect of the polymorphisms in subtype C is unknown; hence it is not known that these DRMs will cause resistance in the respective person. So these initial studies are important to document genotype and further monitoring of these individuals is very important so that these DRMs can be linked to further treatment failure.

It has been suggested that in resource limited settings, whole blood cells or dried blood spots (DBS) could be useful for HIV-1 subtyping or examining overall distribution of drug resistance in a population (Steegen et al. 2007). The results from our study also suggest that genotyping based on whole blood samples could be helpful in drug resistance. However, various factors can affect the emergence of drug resistance in resource-limited settings. The effect of ART can lead to less infectious HIV in the individuals on ART due to lowering of viral load during successful treatment. This benefit can be hampered by an increase in the time available for transmission and thus increasing the probability of more sexually active HIV-1 infected individuals. Furthermore, as treatment success increases, it can lead to behavioural changes such as low adherence to the drugs and thus increasing chances of development of resistance. This makes drug resistance testing at

baseline important in order to identify any DRMs, which may develop rapidly under suboptimal adherence or increase in risky behaviour. Genotyping is common in the western world and a well-established drug resistance profile for subtype B is available. In the case of resource-limited settings, the allocation of resources for laboratory testing has not been a priority. As ART continues to be scaled up rapidly, it is necessary to have an assessment of transmitted and acquired drug resistance. Population-based genotyping assays such as ViroSeq® and TRUGENE® are affordable genotyping methods and are available easily in the western world (Saravanan et al. 2009). However these assays were primarily designed for subtype B viruses thus limiting their utility in countries where non-B subtypes are prevalent (Beddows et al. 2003; Aghokeng et al. 2011). A low-cost alternative to the current genotyping method based on UDPS was successfully used on 48 plasma samples in a single sequencing run. This method was found to be 4-fold more sensitive and 3 to 5 times less expensive than bulk sequencing (Dudley et al. 2012). This method, although was performed on relatively small number of samples, can be exploited using bench top next generation sequencer (Roche/454GS Junior) and can be implemented in resource poor settings. However, in the case of Karonga district, there is a need to have increased focus on providing basic laboratory facilities that can provide good storage for the plasma RNA. Alternatively, a tagged pooled pyrosequencing (TPP) that was effectively used on dried blood spots (Ji et al. 2011) can also be successfully implemented on proviral DNA at places where RNA storage facilities are not available.

Subtype C is found to be predominant subtype in the Karonga district and it can be concluded that this subtype C may continue to dominate in the near future due to limited chances of introduction of new subtype in the region. Furthermore, analysis of selection pressure on the sequences generated in this study showed that selection pressure on HIV in Karonga District is not yet driven by antiretroviral therapy although the study was limited to proviral DNA. A comparative analysis of selection pressure on sequences generated from RNA is recommended.

Thus, improved sample collection and processing is important and is recommended for detailed study on drug resistance in resource-limited settings such as Karonga District, Malawi. To achieve this, more investment in the basic laboratory

facilities and training is recommended so that it would lead to more accurate sample processing and storage. Furthermore such investment can give training and encouragement to local researchers ultimately benefiting those who are mostly affected by the epidemic.

.

References

Abebe, A., V. V. Lukashov, G. Pollakis, A. Kliphuis, A. L. Fontanet, J. Goudsmit, and T. F. de Wit. 2001. Timing of the HIV-1 subtype C epidemic in Ethiopia based on early virus strains and subsequent virus diversification. Aids **15**:1555-1561.

Abreu, C. M., P. A. Brindeiro, A. N. Martins, M. B. Arruda, E. Bule, S. Stakteas, A. Tanuri, and R. de Moraes Brindeiro. 2008. Genotypic and phenotypic characterization of human immunodeficiency virus type 1 isolates circulating in pregnant women from Mozambique. Arch Virol **153**:2013-2017.

Adamson, C. S., and E. O. Freed. 2007. Human immunodeficiency virus type 1 assembly, release, and maturation. Adv Pharmacol **55**:347-387.

Aghokeng, A. F., E. Mpoudi-Ngole, J. E. Chia, E. M. Edoul, E. Delaporte, and M. Peeters. 2011. High failure rate of the ViroSeq HIV-1 genotyping system for drug resistance testing in Cameroon, a country with broad HIV-1 genetic diversity. J Clin Microbiol **49**:1635-1641.

Agnihotri, K. D., S. P. Tripathy, A. P. Jere, S. M. Kale, and R. S. Paranjape. 2006. Molecular analysis of gp41 sequences of HIV type 1 subtype C from India. J Acquir Immune Defic Syndr **41**:345-351.

Andersen, J. L., and V. Planelles. 2005. The role of Vpr in HIV-1 pathogenesis. Curr HIV Res **3**:43-51.

Anderson, B. D., T. Shirasaka, E. Kojima, R. Yarchoan, and H. Mitsuya. 1994. Identification of drug-related genotypic changes in HIV-1 from serum using the selective polymerase chain reaction. Antiviral Res **25**:245-258.

Andersson, A. F., M. Lindberg, H. Jakobsson, F. Backhed, P. Nyren, and L. Engstrand. 2008. Comparative analysis of human gut microbiota by barcoded pyrosequencing. PLoS One **3**:e2836.

Archer, J., M. S. Braverman, B. E. Taillon, B. Desany, I. James, P. R. Harrigan, M. Lewis, and D. L. Robertson. 2009. Detection of low-frequency pretherapy chemokine (CXC motif) receptor 4 (CXCR4)-using HIV-1 with ultra-deep pyrosequencing. Aids **23**:1209-1218.

Archer, J., A. Rambaut, B. E. Taillon, P. R. Harrigan, M. Lewis, and D. L. Robertson. 2010. The evolutionary analysis of emerging low frequency HIV-1 CXCR4 using variants through time--an ultra-deep approach. PLoS Comput Biol **6**:e1001022.

Arien, K. K., A. Abraha, M. E. Quinones-Mateu, L. Kestens, G. Vanham, and E. J. Arts. 2005. The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. J Virol **79**:8979-8990.

Ariën, K. K., G. Vanham, and E. J. Arts. 2007. Is HIV-1 evolving to a less virulent form in humans? Nature Reviews Microbiology **5**:141-151

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

Avila-Rios, S., C. Garcia-Morales, D. Garrido-Rodriguez, C. E. Ormsby, R. Hernandez-Juan, J. Andrade-Villanueva, L. A. Gonzalez-Hernandez, I. Torres-Escobar, S. Navarro-Alvarez, and G. Reyes-Teran. 2011. National prevalence and trends of HIV transmitted drug resistance in Mexico. PLoS One **6**:e27812.

Balode, D., A. Ferdats, I. Dievberna, L. Viksna, B. Rozentale, T. Kolupajeva, V. Konicheva, and T. Leitner. 2004. Rapid epidemic spread of HIV type 1 subtype A1 among intravenous drug users in Latvia and slower spread of subtype B among other risk groups. AIDS Res Hum Retroviruses **20**:245-249.

Bansode, V., Z. J. Drebert, S. A. Travers, E. Banda, A. Molesworth, A. Crampin, B. Ngwira, N. French, J. R. Glynn, and G. P. McCormack. 2011a. Drug resistance mutations in drug-naive HIV type 1 subtype C-infected individuals from rural Malawi. AIDS Res Hum Retroviruses **27**:439-444.

Bansode, V., Z. J. Drebert, S. A. Travers, E. Banda, A. Molesworth, A. Crampin, B. Ngwira, N. French, J. R. Glynn, and G. P. McCormack. 2010. Drug Resistance Mutations in Drug-Naive HIV Type 1 Subtype C-Infected Individuals from Rural Malawi. AIDS Res Hum Retroviruses.

Bansode, V. B., S. A. Travers, A. C. Crampin, B. Ngwira, N. French, J. R. Glynn, and G. P. McCormack. 2011b. Reverse transcriptase drug resistance mutations in HIV-1 subtype C infected patients on ART in Karonga District, Malawi. AIDS Res Ther **8**:38.

Barre-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum, and

L. Montagnier. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science **220**:868-871.

Barzon, L., E. Lavezzo, V. Militello, S. Toppo, and G. Palu. 2011. Applications of next-generation sequencing technologies to diagnostic virology. Int J Mol Sci **12**:7861-7884.

Bates, I., and K. Maitland. 2006. Are laboratory services coming of age in sub-Saharan Africa? Clin Infect Dis **42**:383-384.

Beddows, S., S. Galpin, S. H. Kazmi, A. Ashraf, A. Johargy, A. J. Frater, N. White, R. Braganza, J. Clarke, M. McClure, and J. N. Weber. 2003. Performance of two commercially available sequence-based HIV-1 genotyping systems for the detection of drug resistance against HIV type 1 group M subtypes. J Med Virol **70**:337-342.

Bell, C. M., B. J. Connell, A. Capovilla, W. D. Venter, W. S. Stevens, and M. A. Papathanasopoulos. 2007. Molecular characterization of the HIV type 1 subtype C accessory genes vif, vpr, and vpu. AIDS Res Hum Retroviruses **23**:322-330.

Bennett, D. E., R. J. Camacho, D. Otelea, D. R. Kuritzkes, H. Fleury, M. Kiuchi, W. Heneine, R. Kantor, M. R. Jordan, J. M. Schapiro, A. M. Vandamme, P. Sandstrom, C. A. Boucher, D. van de Vijver, S. Y. Rhee, T. F. Liu, D. Pillay, and R. W. Shafer. 2009. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. PLoS ONE **4**:e4724.

Bergroth, T., A. Sonnerborg, and Z. Yun. 2005. Discrimination of lamivudine resistant minor HIV-1 variants by selective real-time PCR. J Virol Methods **127**:100-107.

Bessong, P. O., C. Larry Obi, T. Cilliers, I. Choge, M. Phoswa, C. Pillay, M. Papathanasopoulos, and L. Morris. 2005. Characterization of human immunodeficiency virus type 1 from a previously unexplored region of South Africa with a high HIV prevalence. AIDS Res Hum Retroviruses **21**:103-109.

Bessong, P. O., J. Mphahlele, I. A. Choge, L. C. Obi, L. Morris, M. L. Hammarskjold, and D. M. Rekosh. 2006. Resistance mutational analysis of HIV type 1 subtype C among rural South African drug-naive patients prior to large-scale availability of antiretrovirals. AIDS Res Hum Retroviruses **22**:1306-1312.

Bieniasz, P. D. 2009. The cell biology of HIV-1 virion genesis. Cell Host Microbe **5**:550-558.

Binladen, J., M. T. Gilbert, J. P. Bollback, F. Panitz, C. Bendixen, R. Nielsen, and E. Willerslev. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS One **2**:e197.

Boni, J., C. Shah, M. Flepp, R. Luthy, and J. Schupbach. 2004. Detection of low copy numbers of HIV-1 proviral DNA in patient PBMCs by a high-input, sequence-capture PCR (Mega-PCR). J Med Virol **72**:1-9.

Boucher, S., P. Recordon-Pinson, D. Neau, J. M. Ragnaud, K. Titier, M. Faure, H. Fleury, and B. Masquelier. 2005. Clonal analysis of HIV-1 variants in proviral DNA during treatment interruption in patients with multiple therapy failures. J Clin Virol **34**:288-294.

Boulle, A., G. Van Cutsem, D. Coetzee, K. Hilderbrand, E. Goemaere, and J. Maartens. 2006. Regimen Durability and Tolerability to 36-month Duration on ART in Khayelitsha, South Africa (Abstract 66). . Conference on Retrovirusses and Opportunistic Infections. Denver, Colorado.

Brenner, B. G. 2007. Resistance and viral subtypes: how important are the differences and why do they occur? Curr Opin HIV AIDS **2**:94-102.

Brenner, B. G., M. Oliveira, F. Doualla-Bell, D. D. Moisi, M. Ntemgwa, F. Frankel, M. Essex, and M. A. Wainberg. 2006. HIV-1 subtype C viruses rapidly develop K65R resistance to tenofovir in cell culture. Aids **20**:F9-13.

Briggs, J. A., M. C. Johnson, M. N. Simon, S. D. Fuller, and V. M. Vogt. 2006. Cryo-electron microscopy reveals conserved and divergent features of gag packing in immature particles of Rous sarcoma virus and human immunodeficiency virus. J Mol Biol **355**:157-168.

Briggs, J. A., T. Wilk, R. Welker, H. G. Krausslich, and S. D. Fuller. 2003. Structural organization of authentic, mature HIV-1 virions and cores. EMBO J **22**:1707-1715.

Buchen-Osmond, C. 1997. Further progress in ICTVdB, a universal virus database. Arch Virol **142**:1734-1739.

Burda, S. T., R. Viswanath, J. Zhao, T. Kinge, C. Anyangwe, E. T. Tinyami, B. Haldar, R. L. Powell, V. Jarido, I. K. Hewlett, and P. N. Nyambi. 2010. HIV-1

reverse transcriptase drug-resistance mutations in chronically infected individuals receiving or naive to HAART in Cameroon. J Med Virol **82**:187-196.

Butler, D. M., M. E. Pacold, P. S. Jordan, D. D. Richman, and D. M. Smith. 2009. The efficiency of single genome amplification and sequencing is improved by quantitation and use of a bioinformatics tool. J Virol Methods **162**:280-283.

Butler, I. F., I. Pandrea, P. A. Marx, and C. Apetrei. 2007. HIV genetic diversity: biological and public health consequences. Curr HIV Res **5**:23-45.

Calmy, A., N. Ford, B. Hirschel, S. J. Reynolds, L. Lynen, E. Goemaere, F. Garcia de la Vega, L. Perrin, and W. Rodriguez. 2007. HIV viral load monitoring in resource-limited regions: optional or necessary? Clin Infect Dis **44**:128-134.

CDC. 1981a. A Cluster of Kaposi's Sarcoma and Pneumocystis carinii Pneumonia among omosexual Male Residents of Los Angeles and range Counties, California. Morbidity and Mortality Weekly Report **31**:305-307.

CDC. 1981b. Pneumocystis Pneumonia --- Los Angeles. Morbidity and Mortality Weekly Report **30**:1- 3.

Cen, S., A. Khorchid, H. Javanbakht, J. Gabor, T. Stello, K. Shiba, K. Musier-Forsyth, and L. Kleiman. 2001. Incorporation of lysyl-tRNA synthetase into human immunodeficiency virus type 1. J Virol **75**:5043-5048.

Chen, L., A. Perlina, and C. J. Lee. 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**:3722-3732.

Cherry, C. L., and S. L. Wesselingh. 2003. Nucleoside analogues and HIV: the combined cost to mitochondria. J Antimicrob Chemother **51**:1091-1093.

Chiu, T. K., and D. R. Davies. 2004. Structure and function of HIV-1 integrase. Curr Top Med Chem **4**:965-977.

Chui, C. K., Z. L. Brumme, C. J. Brumme, B. Yip, E. J. Phillips, J. S. Montaner, and P. R. Harrigan. 2007. A simple screening approach to reduce B*5701-associated abacavir hypersensitivity on the basis of sequence variation in HIV reverse transcriptase. Clin Infect Dis **44**:1503-1508.

Cladera, J., I. Martin, and P. O'Shea. 2001. The fusion domain of HIV gp41 interacts specifically with heparan sulfate on the T-lymphocyte cell surface. EMBO J **20**:19-26.

Clark, S. J., M. S. Saag, W. D. Decker, S. Campbell-Hill, J. L. Roberson, P. J. Veldkamp, J. C. Kappes, B. H. Hahn, and G. M. Shaw. 1991. High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection. N Engl J Med **324**:954-960.

Clavel, F., and A. J. Hance. 2004. HIV drug resistance. N Engl J Med **350**:1023-1035.

Codoner, F. M., C. Pou, A. Thielen, F. Garcia, R. Delgado, D. Dalmau, M. Alvarez-Tejado, L. Ruiz, B. Clotet, and R. Paredes. 2011. Added value of deep sequencing relative to population sequencing in heavily pre-treated HIV-1-infected subjects. PLoS One **6**:e19461.

Coffin, J. M., S. H. Hughes, and H. E. Varmus. 1997. Retroviral Taxonomy, Protein Structures, Sequences, and Genetic Maps. Retroviruses **Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1997.**

Coutsinos, D., C. F. Invernizzi, D. Moisi, M. Oliveira, J. L. Martinez-Cajas, B. G. Brenner, and M. A. Wainberg. 2011. A template-dependent dislocation mechanism potentiates K65R reverse transcriptase mutation development in subtype C variants of HIV-1. PLoS One **6**:e20208.

Crampin, A. C., J. R. Glynn, and P. E. Fine. 2009. What has Karonga taught us? Tuberculosis studied over three decades. Int J Tuberc Lung Dis **13**:153-164.

Daar, E. S., T. Moudgil, R. D. Meyer, and D. D. Ho. 1991. Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection. N Engl J Med **324**:961-964.

Dalai, S. C., T. de Oliveira, G. W. Harkins, S. G. Kassaye, J. Lint, J. Manasa, E. Johnston, and D. Katzenstein. 2009. Evolution and molecular epidemiology of subtype C HIV-1 in Zimbabwe. Aids **23**:2523-2532.

Dayton, A. I. 2004. Within you, without you: HIV-1 Rev and RNA export. Retrovirology **1**:35.

De Grassi, A., C. Segala, F. Iannelli, S. Volorio, L. Bertario, P. Radice, L. Bernard, and F. D. Ciccarelli. 2010. Ultradeep sequencing of a human ultraconserved

region reveals somatic and constitutional genomic instability. PLoS Biol **8**:e1000275.

De Luca, A., A. Antinori, S. Di Giambenedetto, A. Cingolani, M. Colafigli, C. F. Perno, and R. Cauda. 2003. Interpretation systems for genotypic drug resistance of HIV-1. Scand J Infect Dis Suppl **106**:29-34.

de S. Leal, E., E. C. Holmes, and P. M. Zanotto. 2004. Distinct patterns of natural selection in the reverse transcriptase gene of HIV-1 in the presence and absence of antiretroviral therapy. Virology **325**:181-191.

Debaisieux, S., F. Rayne, H. Yezid, and B. Beaumelle. 2011. The Ins and Outs of HIV-1 Tat. Traffic **13**:355-363.

Delobel, P., A. Saliou, F. Nicot, M. Dubois, S. Trancart, P. Tangre, J. P. Aboulker, A. M. Taburet, J. M. Molina, P. Massip, B. Marchou, and J. Izopet. 2011. Minor HIV-1 variants with the K103N resistance mutation during intermittent efavirenz-containing antiretroviral therapy and virological failure. PLoS One **6**:e21655.

Delport, W., A. F. Poon, S. D. Frost, and S. L. Kosakovsky Pond. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics **26**:2455-2457.

Derdelinckx, I., K. Van Laethem, B. Maes, Y. Schrooten, K. De Schouwer, S. De Wit, K. Fransen, S. Garcia Ribas, M. Moutschen, D. Vaira, G. Zissis, M. Van Ranst, E. Van Wijngaerden, and A. M. Vandamme. 2003. Performance of the VERSANT HIV-1 resistance assays (LiPA) for detecting drug resistance in therapy-naive patients infected with different HIV-1 subtypes. FEMS Immunol Med Microbiol **39**:119-124.

Descamps, D., M. Ait-Khaled, C. Craig, S. Delarue, F. Damond, G. Collin, and F. Brun-Vezinet. 2006. Rare selection of the K65R mutation in antiretroviral-naive patients failing a first-line abacavir/ lamivudine-containing HAART regimen. Antivir Ther **11**:701-705.

DHHS. 2006. Guidelines for the use of antiretroviral agents in HIV-infected adults and adolescents. In. Edited by DHHS.

Doms, R. W., and D. Trono. 2000. The plasma membrane as a combat zone in the HIV battlefield. Genes Dev **14**:2677-2688.

Dorr, P., M. Westby, S. Dobbs, P. Griffin, B. Irvine, M. Macartney, J. Mori, G. Rickett, C. Smith-Burchnell, C. Napier, R. Webster, D. Armour, D. Price, B. Stammen, A. Wood, and M. Perros. 2005. Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. Antimicrob Agents Chemother **49**:4721-4732.

Doualla-Bell, F., A. Avalos, B. Brenner, T. Gaolathe, M. Mine, S. Gaseitsiwe, M. Oliveira, D. Moisi, N. Ndwapi, H. Moffat, M. Essex, and M. A. Wainberg. 2006. High prevalence of the K65R mutation in human immunodeficiency virus type 1 subtype C isolates from infected patients in Botswana treated with didanosine-based regimens. Antimicrob Agents Chemother **50**:4182-4185.

Drummond, A., B. Ashton, S. Buxton, M. Cheung, r. A. Coope, C. Duran, M. Field, J. Heled, M. Kearse, S. Markowitz, R. Moir, S. Stones-Havas, S. Sturrock, T. Thierer, and A. Wilson. 2011. Geneious v5.4, Available from http://www.geneious.com/.

Dudley, D. M., E. N. Chin, B. N. Bimber, S. S. Sanabani, L. F. Tarosso, P. R. Costa, M. M. Sauer, E. G. Kallas, and D. H. O.,ÄôConnor. 2012. Low-Cost Ultra-Wide Genotyping Using Roche/454 Pyrosequencing for Surveillance of HIV Drug Resistance. PLoS One **7**:e36494.

Edelstein, R. E., D. A. Nickerson, V. O. Tobe, L. A. Manns-Arcuino, and L. M. Frenkel. 1998. Oligonucleotide ligation assay for detecting mutations in the human immunodeficiency virus type 1 pol gene that are associated with resistance to zidovudine, didanosine, and lamivudine. J Clin Microbiol **36**:569-572.

Emerman, M., and M. H. Malim. 1998. HIV-1 regulatory/accessory genes: keys to unraveling viral and host cell biology. Science **280**:1880-1884.

Eriksson, N., L. Pachter, Y. Mitsuya, S. Y. Rhee, C. Wang, B. Gharizadeh, M. Ronaghi, R. W. Shafer, and N. Beerenwinkel. 2008. Viral population estimation using pyrosequencing. PLoS Comput Biol **4**:e1000074.

Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res **8**:175-185.

Fauci, A. S., G. Pantaleo, S. Stanley, and D. Weissman. 1996. Immunopathogenic mechanisms of HIV infection. Ann Intern Med **124**:654-663.

Flexner, C. 1998. HIV-protease inhibitors. N Engl J Med **338**:1281-1292.

Fournier, C., J. C. Cortay, C. Carbonnelle, C. Ehresmann, R. Marquet, and P. Boulanger. 2002. The HIV-1 Nef protein enhances the affinity of reverse transcriptase for RNA in vitro. Virus Genes **25**:255-269.

Francis, D. P., J. W. Curran, and M. Essex. 1983. Epidemic acquired immune deficiency syndrome: epidemiologic evidence for a transmissible agent. J Natl Cancer Inst **71**:1-4.

Gao, F., E. Bailes, D. L. Robertson, Y. Chen, C. M. Rodenburg, S. F. Michael, L. B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature **397**:436-441.

Garcia, F., M. Plana, C. Vidal, A. Cruceta, W. A. O'Brien, G. Pantaleo, T. Pumarola, T. Gallart, J. M. Miro, and J. M. Gatell. 1999. Dynamics of viral load rebound and immunological changes after stopping effective antiretroviral therapy. Aids **13**:F79-86.

Garcia-Perez, J., S. Sanchez-Palomino, M. Perez-Olmeda, B. Fernandez, and J. Alcami. 2007. A new strategy based on recombinant viruses as a tool for assessing drug susceptibility of human immunodeficiency virus type 1. J Med Virol **79**:127-137.

Gelderblom, H. R. 1991. Assembly and morphology of HIV: potential effect of structure on viral function. AIDS **5**:617-637.

Geretti, A. M. 2007. Epidemiology of antiretroviral drug resistance in drug-naive persons. Curr Opin Infect Dis **20**:22-32.

Gilks, C. F., S. Crowley, R. Ekpini, S. Gove, J. Perriens, Y. Souteyrand, D. Sutherland, M. Vitoria, T. Guerma, and K. De Cock. 2006. The WHO public-health approach to antiretroviral treatment against HIV in resource-limited settings. Lancet **368**:505-510.

Glynn, J. R., J. Ponnighaus, A. C. Crampin, F. Sibande, L. Sichali, P. Nkhosa, P. Broadbent, and P. E. Fine. 2001. The development of the HIV epidemic in Karonga District, Malawi. AIDS **15**:2025-2029.

Goff, S. P. 2003. Death by deamination: a novel host restriction system for HIV-1. Cell **114**:281-283.

Goldgur, Y., R. Craigie, G. H. Cohen, T. Fujiwara, T. Yoshinaga, T. Fujishita, H. Sugimoto, T. Endo, H. Murai, and D. R. Davies. 1999. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. Proc Natl Acad Sci U S A **96**:13040-13043.

Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol **11**:725-736.

Gordon, S., and G. Stephen. 2006. Epidemiology of Respiratory Disease in Malawi. Malawi Medical Journal **18**:134-146.

Gottlieb, M. S., R. Schroff, H. M. Schanker, J. D. Weisman, P. T. Fan, R. A. Wolf, and A. Saxon. 1981. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N Engl J Med **305**:1425-1431.

Graziosi, C., G. Pantaleo, L. Butini, J. F. Demarest, M. S. Saag, G. M. Shaw, and A. S. Fauci. 1993. Kinetics of human immunodeficiency virus type 1 (HIV-1) DNA and RNA synthesis during primary HIV-1 infection. Proc Natl Acad Sci U S A **90**:6405-6409.

Guimaraes, M. L., K. G. Velarde-Dunois, D. Segurondo, and M. G. Morgado. 2012. The HIV-1 epidemic in Bolivia is dominated by subtype B and CRF12_BF "family" strains. Virol J **9**:19.

Gulick, R., J. Mellors, D. Havlir, J. Eron, C. Gonzalez, and D. McMahon. 1996. Potent and sustained antiretroviral activity of indinavir (IDV) in combination with zidovudine (ZDV) and lamivudine (3TC) (abstract LB7). Conference on Retroviruses and Opportunistic Infections Washington DC, USA, 28 January - 1 February 1996. .

Gulick, R. M., J. W. Mellors, D. Havlir, J. J. Eron, C. Gonzalez, D. McMahon, D. D. Richman, F. T. Valentine, L. Jonas, A. Meibohm, E. A. Emini, and J. A. Chodakewitz. 1997. Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. N Engl J Med **337**:734-739.

Gupta, R. K., D. Ford, V. Mulenga, A. S. Walker, D. Kabamba, M. Kalumbi, P. R. Grant, A. Ferrier, D. Pillay, D. M. Gibb, and C. Chintu. 2010. Drug resistance

in human immunodeficiency virus type-1 infected Zambian children using adult fixed dose combination stavudine, lamivudine, and nevirapine. Pediatr Infect Dis J **29**:e57-62.

Halvas, E. K., A. Wiegand, V. F. Boltz, M. Kearney, D. Nissley, M. Wantman, S. M. Hammer, S. Palmer, F. Vaida, J. M. Coffin, and J. W. Mellors. 2010. Low frequency nonnucleoside reverse-transcriptase inhibitor-resistant variants contribute to failure of efavirenz-containing regimens in treatment-experienced patients. J Infect Dis **201**:672-680.

Hamers, R. L., I. Derdelinckx, M. van Vugt, W. Stevens, T. F. Rinke de Wit, and R. Schuurman. 2008. The status of HIV-1 resistance to antiretroviral drugs in sub-Saharan Africa. Antivir Ther **13**:625-639.

Hammer, S. M., K. E. Squires, M. D. Hughes, J. M. Grimes, L. M. Demeter, J. S. Currier, J. J. Eron, Jr., J. E. Feinberg, H. H. Balfour, Jr., L. R. Deyton, J. A. Chodakewitz, and M. A. Fischl. 1997. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. N Engl J Med **337**:725-733.

Handema, R., H. Terunuma, F. Kasolo, H. Kasai, M. Sichone, A. Yamashita, X. Deng, G. Mulundu, K. Ichiyama, M. Munkanta, T. Yokota, N. Wakasugi, F. Tezuka, N. Yamamoto, and M. Ito. 2003. Prevalence of drug-resistance-associated mutations in antiretroviral drug-naive Zambians infected with subtype C HIV-1. AIDS Res Hum Retroviruses **19**:151-160.

Harrigan, P. R., J. S. Montaner, S. A. Wegner, W. Verbiest, V. Miller, R. Wood, and B. A. Larder. 2001. World-wide variation in HIV-1 phenotypic susceptibility in untreated individuals: biologically relevant values for resistance testing. AIDS **15**:1671-1677.

Health, M. o. 2007. Antiretroviral Treatment Programme in MALAWI. Annual report.

Health, M. o. 2008. Guidelines for the use of antiretroviral therapy in Malawi, Third edition. .

Hedskog, C., M. Mild, J. Jernberg, E. Sherwood, G. Bratt, T. Leitner, J. Lundeberg, B. Andersson, and J. Albert. 2010. Dynamics of HIV-1 quasispecies during antiviral treatment dissected using ultra-deep pyrosequencing. PLoS One **5**:e11345.

Hemelaar, J., E. Gouws, P. D. Ghys, and S. Osmanov. 2011. Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS **25**:679-689.

Hemelaar, J., E. Gouws, P. D. Ghys, and S. Osmanov. 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. Aids **20**:W13-23.

Hertogs, K., M. P. de Bethune, V. Miller, T. Ivens, P. Schel, A. Van Cauwenberge, C. Van Den Eynde, V. Van Gerwen, H. Azijn, M. Van Houtte, F. Peeters, S. Staszewski, M. Conant, S. Bloor, S. Kemp, B. Larder, and R. Pauwels. 1998. A rapid method for simultaneous detection of phenotypic resistance to inhibitors of protease and reverse transcriptase in recombinant human immunodeficiency virus type 1 isolates from patients treated with antiretroviral drugs. Antimicrob Agents Chemother **42**:269-276.

Hirsch, M. S., F. Brun-Vezinet, B. Clotet, B. Conway, D. R. Kuritzkes, R. T. D'Aquila, L. M. Demeter, S. M. Hammer, V. A. Johnson, C. Loveday, J. W. Mellors, D. M. Jacobsen, and D. D. Richman. 2003. Antiretroviral drug resistance testing in adults infected with human immunodeficiency virus type 1: 2003 recommendations of an International AIDS Society-USA Panel. Clin Infect Dis **37**:113-128.

Hirsch, M. S., H. F. Gunthard, J. M. Schapiro, F. Brun-Vezinet, B. Clotet, S. M. Hammer, V. A. Johnson, D. R. Kuritzkes, J. W. Mellors, D. Pillay, P. G. Yeni, D. M. Jacobsen, and D. D. Richman. 2008. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. Clin Infect Dis **47**:266-285.

Hoffmann, C., N. Minkah, J. Leipzig, G. Wang, M. Q. Arens, P. Tebas, and F. D. Bushman. 2007. DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. Nucleic Acids Res **35**:e91.

Hoffmann, C. J., S. Charalambous, J. Sim, J. Ledwaba, G. Schwikkard, R. E. Chaisson, K. L. Fielding, G. J. Churchyard, L. Morris, and A. D. Grant. 2009. Viremia, resuppression, and time to resistance in human immunodeficiency virus (HIV) subtype C during first-line antiretroviral therapy in South Africa. Clin Infect Dis **49**:1928-1935.

Hosseinipour, M., J. van Oosterhout, and R. Weigel. 2007. Validating clinical and immunological definitions of antiretroviral treatment failure in Malawi.

International AIDS Society Conference on HIV Pathogenesis, Treatment and Prevention (Abstract WEAB101). Sydney, Australia.

Hosseinipour, M. C., J. J. van Oosterhout, R. Weigel, S. Phiri, D. Kamwendo, N. Parkin, S. A. Fiscus, J. A. Nelson, J. J. Eron, and J. Kumwenda. 2009. The public health approach to identify antiretroviral therapy failure: high-level nucleoside reverse transcriptase inhibitor resistance among Malawians failing first-line antiretroviral therapy. Aids **23**:1127-1134.

Huang, D. D., T. A. Giesler, and J. W. Bremer. 2003. Sequence characterization of the protease and partial reverse transcriptase proteins of the NED panel, an international HIV type 1 subtype reference and standards panel. AIDS Res Hum Retroviruses **19**:321-328.

Hue, S., J. P. Clewley, P. A. Cane, and D. Pillay. 2004. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. Aids **18**:719-728.

Hung, C. H., L. Thomas, C. E. Ruby, K. M. Atkins, N. P. Morris, Z. A. Knight, I. Scholz, E. Barklis, A. D. Weinberg, K. M. Shokat, and G. Thomas. 2007. HIV-1 Nef assembles a Src family kinase-ZAP-70/Syk-PI3K cascade to downregulate cell-surface MHC-I. Cell Host Microbe **1**:121-133.

Hussain, A., C. Wesley, M. Khalid, A. Chaudhry, and S. Jameel. 2008. Human immunodeficiency virus type 1 Vpu protein interacts with CD74 and modulates major histocompatibility complex class II presentation. J Virol **82**:893-902.

Jabara, C. B., C. D. Jones, J. Roach, J. A. Anderson, and R. Swanstrom. 2011. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci U S A **108**:20166-20171.

Jacks, T., M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr, and H. E. Varmus. 1988. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature **331**:280-283.

Jakobsen, M. R., M. Tolstrup, O. S. Sogaard, L. B. Jorgensen, P. R. Gorry, A. Laursen, and L. Ostergaard. 2010. Transmission of HIV-1 drug-resistant variants: prevalence and effect on treatment outcome. Clin Infect Dis **50**:566-573.

James, C. O., M. B. Huang, M. Khan, M. Garcia-Barrio, M. D. Powell, and V. C. Bond. 2004. Extracellular Nef protein targets CD4+ T cells for apoptosis by interacting with CXCR4 surface receptors. J Virol **78**:3099-3109.

Janssens, W., A. Buve, and J. N. Nkengasong. 1997. The puzzle of HIV-1 subtypes in Africa. AIDS **11**:705-712.

Japour, A. J., D. L. Mayers, V. A. Johnson, D. R. Kuritzkes, L. A. Beckett, J. M. Arduino, J. Lane, R. J. Black, P. S. Reichelderfer, R. T. D'Aquila, and et al. 1993. Standardized peripheral blood mononuclear cell culture assay for determination of drug susceptibilities of clinical human immunodeficiency virus type 1 isolates. The RV-43 Study Group, the AIDS Clinical Trials Group Virology Committee Resistance Working Group. Antimicrob Agents Chemother **37**:1095-1101.

Jeeninga, R. E., M. Hoogenkamp, M. Armand-Ugon, M. de Baar, K. Verhoef, and B. Berkhout. 2000. Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. J Virol **74**:3740-3751.

Ji, H., Y. Li, M. Graham, B. B. Liang, R. Pilon, S. Tyson, G. Peters, S. Tyler, H. Merks, S. Bertagnolio, L. Soto-Ramirez, P. Sandstrom, and J. Brooks. 2011. Next-generation sequencing of dried blood spot specimens: a novel approach to HIV drug-resistance surveillance. Antivir Ther **16**:871-878.

Ji, H., N. Masse, S. Tyler, B. Liang, Y. Li, H. Merks, M. Graham, P. Sandstrom, and J. Brooks. 2010. HIV drug resistance surveillance using pooled pyrosequencing. PLoS One **5**:e9263.

Johannessen, A., E. Naman, S. L. Kivuyo, M. J. Kasubi, M. Holberg-Petersen, M. I. Matee, S. G. Gundersen, and J. N. Bruun. 2009. Virological efficacy and emergence of drug resistance in adults on antiretroviral treatment in rural Tanzania. BMC Infect Dis **9**:108.

Johnson, J. A., J. F. Li, X. Wei, J. Lipscomb, D. Irlbeck, C. Craig, A. Smith, D. E. Bennett, M. Monsour, P. Sandstrom, E. R. Lanier, and W. Heneine. 2008. Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naive populations and associate with reduced treatment efficacy. PLoS Med **5**:e158.

Johnson, V. A., V. Calvez, H. F. Gunthard, R. Paredes, D. Pillay, R. Shafer, A. M. Wensing, and D. D. Richman. 2011. 2011 update of the drug resistance mutations in HIV-1. Top Antivir Med **19**:156-164.

Joos, B., M. Fischer, H. Kuster, S. K. Pillai, J. K. Wong, J. Boni, B. Hirschel, R. Weber, A. Trkola, and H. F. Gunthard. 2008. HIV rebounds from latently infected cells, rather than from continuing low-level replication. Proc Natl Acad Sci U S A **105**:16725-16730.

Junqueira, D. M., R. M. de Medeiros, M. C. Matte, L. A. Araujo, J. A. Chies, P. Ashton-Prolla, and S. E. Almeida. 2011. Reviewing the history of HIV-1: spread of subtype B in the Americas. PLoS One **6**:e27489.

Kantor, R., L. S. Zijenah, R. W. Shafer, S. Mutetwa, E. Johnston, R. Lloyd, A. von Lieven, D. Israelski, and D. A. Katzenstein. 2002. HIV-1 subtype C reverse transcriptase and protease genotypes in Zimbabwean patients failing antiretroviral therapy. AIDS Res Hum Retroviruses **18**:1407-1413.

Kassu, A., M. Fujino, M. Matsuda, M. Nishizawa, F. Ota, and W. Sugiura. 2007. Molecular epidemiology of HIV type 1 in treatment-naive patients in north Ethiopia. AIDS Res Hum Retroviruses **23**:564-568.

Keiser, O., K. Anastos, M. Schechter, E. Balestre, L. Myer, A. Boulle, D. Bangsberg, H. Toure, P. Braitstein, E. Sprinz, D. Nash, M. Hosseinipour, F. Dabis, M. May, M. W. Brinkhof, and M. Egger. 2008. Antiretroviral therapy in resource-limited settings 1996 to 2006: patient characteristics, treatment regimens and monitoring in sub-Saharan Africa, Asia and Latin America. Trop Med Int Health **13**:870-879.

Keiser, O., H. Tweya, A. Boulle, P. Braitstein, M. Schecter, M. W. Brinkhof, F. Dabis, S. Tuboi, E. Sprinz, M. Pujades-Rodriguez, A. Calmy, N. Kumarasamy, D. Nash, A. Jahn, P. MacPhail, R. Luthy, R. Wood, and M. Egger. 2009. Switching to second-line antiretroviral therapy in resource-limited settings: comparison of programmes with and without viral load monitoring. AIDS **23**:1867-1874.

Kieffer, T. L., P. Kwon, R. E. Nettles, Y. Han, S. C. Ray, and R. F. Siliciano. 2005. G->A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting CD4+ T cells in vivo. J Virol **79**:1975-1980.

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

Korber, B., M. Muldoon, J. Theiler, F. Gao, R. Gupta, A. Lapedes, B. H. Hahn, S. Wolinsky, and T. Bhattacharya. 2000. Timing the ancestor of the HIV-1 pandemic strains. Science **288**:1789-1796.

Korn, K., H. Reil, H. Walter, and B. Schmidt. 2003. Quality control trial for human immunodeficiency virus type 1 drug resistance testing using clinical samples reveals problems with detecting minority species and interpretation of test results. J Clin Microbiol **41**:3559-3565.

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

Kozal, M. J., N. Shah, N. Shen, R. Yang, R. Fucini, T. C. Merigan, D. D. Richman, D. Morris, E. Hubbell, M. Chee, and T. R. Gingeras. 1996. Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. Nat Med **2**:753-759.

Kozarewa, I., Z. Ning, M. A. Quail, M. J. Sanders, M. Berriman, and D. J. Turner. 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat Methods **6**:291-295.

Kwiek, J. J., E. S. Russell, K. K. Dang, C. L. Burch, V. Mwapasa, S. R. Meshnick, and R. Swanstrom. 2008. The molecular epidemiology of HIV-1 envelope diversity during HIV-1 subtype C vertical transmission in Malawian mother-infant pairs. AIDS **22**:863-871.

Lahuerta, M., E. Aparicio, A. Bardaji, S. Marco, J. Sacarlal, I. Mandomando, P. Alonso, M. A. Martinez, C. Menendez, and D. Naniche. 2008. Rapid spread and genetic diversification of HIV type 1 subtype C in a rural area of southern Mozambique. AIDS Res Hum Retroviruses **24**:327-335.

Langford, S. E., J. Ananworanich, and D. A. Cooper. 2007. Predictors of disease progression in HIV infection: a review. AIDS Res Ther **4**:11.

LANL. 2012. Distribution of all HIV-1 sequences: AFRICA HIV Sequence Database.

Larder, B. A., G. Darby, and D. D. Richman. 1989. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. Science **243**:1731-1734.

Larder, B. A., and S. D. Kemp. 1989. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). Science **246**:1155-1158.

Laurent, C., N. F. Ngom Gueye, C. T. Ndour, P. M. Gueye, M. Diouf, N. Diakhate, N. C. Toure Kane, I. Laniece, A. Ndir, L. Vergne, I. Ndoye, S. Mboup, P. S. Sow, and E. Delaporte. 2005. Long-term benefits of highly active antiretroviral therapy in Senegalese HIV-1-infected adults. J Acquir Immune Defic Syndr **38**:14-17.

Lay, C. S., L. E. Ludlow, D. Stapleton, A. K. Bellamy-McIntyre, P. A. Ramsland, H. E. Drummer, and P. Poumbourios. 2011. Role for the terminal clasp of HIV-1 gp41 glycoprotein in the initiation of membrane fusion. J Biol Chem **286**:41331-41343.

Le Rouzic, E., and S. Benichou. 2005. The Vpr protein from HIV-1: distinct roles along the viral life cycle. Retrovirology **2**:11.

Le, T., J. Chiarella, B. B. Simen, B. Hanczaruk, M. Egholm, M. L. Landry, K. Dieckhaus, M. I. Rosen, and M. J. Kozal. 2009. Low-abundance HIV drug-resistant viral variants in treatment-experienced persons correlate with historical antiretroviral use. PLoS One **4**:e6079.

Levy, J. A. 2007. HIV and the Pathogenesis of AIDS, 3rd ed. Washington DC. ASM Press.

Levy, J. A., A. D. Hoffman, S. M. Kramer, J. A. Landis, J. M. Shimabukuro, and L. S. Oshiro. 1984. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. Science **225**:840-842.

Li, W. H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol **36**:96-99.

Liang, B., M. Luo, J. Scott-Herridge, C. Semeniuk, M. Mendoza, R. Capina, B. Sheardown, H. Ji, J. Kimani, B. T. Ball, G. Van Domselaar, M. Graham, S. Tyler, S. J. Jones, and F. A. Plummer. 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.

Liu, S. Q., S. X. Liu, and Y. X. Fu. 2008. Molecular motions of human HIV-1 gp120 envelope glycoproteins. J Mol Model **14**:857-870.

Lukashov, V. V., and J. Goudsmit. 2002. Recent evolutionary history of human immunodeficiency virus type 1 subtype B: reconstruction of epidemic onset based on sequence distances to the common ancestor. J Mol Evol **54**:680-691.

Macias, J., J. C. Palomares, J. A. Mira, M. J. Torres, J. A. Garcia-Garcia, J. M. Rodriquez, S. Vergera, and J. A. Pineda. 2005. Transient rebounds of HIV plasma viremia are associated with the emergence of drug resistance mutations in patients on highly active antiretroviral therapy. J Infect **51**:195-200.

Maddison, W. P., and D. R. Maddison. 2005. Analysis of phylogeny and character evolution. Version 4.08a. . http://macclade.org.

Malim, M. H., J. Hauber, S. Y. Le, J. V. Maizel, and B. R. Cullen. 1989. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature **338**:254-257.

Mardis, E. R. 2008. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet **9**:387-402.

Margeridon-Thermet, S., N. S. Shulman, A. Ahmed, R. Shahriar, T. Liu, C. Wang, S. P. Holmes, F. Babrzadeh, B. Gharizadeh, B. Hanczaruk, B. B. Simen, M. Egholm, and R. W. Shafer. 2009. Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naive patients. J Infect Dis **199**:1275-1285.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and

J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**:376-380.

Martinez-Cajas, J. L., N. Pant-Pai, M. B. Klein, and M. A. Wainberg. 2008. Role of genetic diversity amongst HIV-1 non-B subtypes in drug resistance: a systematic review of virologic and biochemical evidence. AIDS Rev **10**:212-223.

Masur, H., M. A. Michelis, J. B. Greene, I. Onorato, R. A. Stouwe, R. S. Holzman, G. Wormser, L. Brettman, M. Lange, H. W. Murray, and S. Cunningham-Rundles. 1981. An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. N Engl J Med **305**:1431-1438.

Matthews, T., M. Salgo, M. Greenberg, J. Chung, R. DeMasi, and D. Bolognesi. 2004. Enfuvirtide: the first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes. Nat Rev Drug Discov **3**:215-225.

Mazauric, M. H., Y. Seol, S. Yoshizawa, K. Visscher, and D. Fourmy. 2009. Interaction of the HIV-1 frameshift signal with the ribosome. Nucleic Acids Res **37**:7654-7664.

McCormack, G. P., J. R. Glynn, A. C. Crampin, F. Sibande, D. Mulawa, L. Bliss, P. Broadbent, K. Abarca, J. M. Ponnighaus, P. E. Fine, and J. P. Clewley. 2002. Early evolution of the human immunodeficiency virus type 1 subtype C epidemic in rural Malawi. J Virol **76**:12890-12899.

McCormack, G. P., J. R. Glynn, A. C. Crampin, F. Sibande, D. Mulawa, P. E. Fine, and J. P. Clewley. 2003. Highly divergent HIV type 1 group M sequences evident in Karonga District, Malawi in early 1980s. AIDS Res Hum Retroviruses **19**:441-445.

McGrath, N., K. Kranzer, J. Saul, A. C. Crampin, S. Malema, L. Kachiwanda, B. Zaba, A. Jahn, P. E. Fine, and J. R. Glynn. 2007. Estimating the need for antiretroviral treatment and an assessment of a simplified HIV/AIDS case definition in rural Malawi. Aids **21 Suppl 6**:S105-113.

Menendez-Arias, L. 2010. Molecular basis of human immunodeficiency virus drug resistance: an update. Antiviral Res **85**:210-231.

Messiaen, P., C. Verhofstede, I. Vandenbroucke, S. Dinakis, V. Van Eygen, K. Thys, B. Winters, J. Aerssens, D. Vogelaers, L. J. Stuyver, and L. Vandekerckhove.

2012. Ultra-deep sequencing of HIV-1 reverse transcriptase before start of an NNRTI-based regimen in treatment-naive patients. Virology **426**:7-11.

Metzker, M. L. 2010. Sequencing technologies - the next generation. Nat Rev Genet **11**:31-46.

Metzner, K. J., S. G. Giulieri, S. A. Knoepfel, P. Rauch, P. Burgisser, S. Yerly, H. F. Gunthard, and M. Cavassini. 2009. Minority quasispecies of drug-resistant HIV-1 that lead to early therapy failure in treatment-naive and -adherent patients. Clin Infect Dis **48**:239-247.

Ministry of Health, G. o. M. 2010. Quarterly HIV Programme Report of HIV testing and counseling, Prevention of Mother to Child Transmission, Antiretroviral Therapy, Treatment of Sexually Transmitted Infections with results upto June 2010.

Mitsuya, Y., V. Varghese, C. Wang, T. F. Liu, S. P. Holmes, P. Jayakumar, B. Gharizadeh, M. Ronaghi, D. Klein, W. J. Fessel, and R. W. Shafer. 2008. Minority human immunodeficiency virus type 1 variants in antiretroviral-naive persons with reverse transcriptase codon 215 revertant mutations. J Virol **82**:10747-10755.

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

Morikawa, Y. 2003. HIV capsid assembly. Curr HIV Res **1**:1-14.

Moulard, M., and E. Decroly. 2000. Maturation of HIV envelope glycoprotein precursors by cellular endoproteases. Biochim Biophys Acta **1469**:121-132.

Murillo, W., G. Paz-Bailey, S. Morales, E. Monterroso, M. Paredes, T. Dobbs, B. S. Parekh, J. Albert, and I. L. Rivera. 2010. Transmitted drug resistance and type of infection in newly diagnosed HIV-1 individuals in Honduras. J Clin Virol **49**:239-244.

NACO. 2001. Estimating National HIV Prevalence in Malawi from Sentinel Surveillance Data The National AIDS Control Programme, The POLICY Project, Lilongwe, Malawi

.

Nadai, Y., L. M. Eyzaguirre, A. Sill, F. Cleghorn, C. Nolte, M. Charurat, S. Collado-Chastel, N. Jack, C. Bartholomew, J. W. Pape, P. Figueroa, W. A. Blattner,

and J. K. Carr. 2009. HIV-1 epidemic in the Caribbean is dominated by subtype B. PLoS One **4**:e4814.

Napravnik, S., D. Edwards, P. Stewart, B. Stalzer, E. Matteson, and J. J. Eron, Jr. 2005. HIV-1 drug resistance evolution among patients on potent combination antiretroviral therapy with detectable viremia. J Acquir Immune Defic Syndr **40**:34-40.

Navarro, F., and N. R. Landau. 2004. Recent insights into HIV-1 Vif. Curr Opin Immunol **16**:477-482.

Negroni, M., and H. Buc. 2001. Mechanisms of retroviral recombination. Annu Rev Genet **35**:275-302.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol **3**:418-426.

Neil, S. J., S. W. Eastman, N. Jouvenet, and P. D. Bieniasz. 2006. HIV-1 Vpu promotes release and prevents endocytosis of nascent retrovirus particles from the plasma membrane. PLoS Pathog **2**:e39.

Neil, S. J., T. Zang, and P. D. Bieniasz. 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature **451**:425-430.

Nielsen, R., and Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148**:929-936.

Novitsky, V., R. Wang, H. Bussmann, S. Lockman, M. Baum, R. Shapiro, I. Thior, C. Wester, C. W. Wester, A. Ogwu, A. Asmelash, R. Musonda, A. Campa, S. Moyo, E. van Widenfelt, M. Mine, C. Moffat, M. Mmalane, J. Makhema, R. Marlink, P. Gilbert, G. R. Seage, 3rd, V. DeGruttola, and M. Essex. 2010. HIV-1 subtype C-infected individuals maintaining high viral load as potential targets for the "test-and-treat" approach to reduce HIV transmission. PLoS One **5**:e10148.

Nowak, M. 1990. HIV mutation rate. Nature **347**:522.

Oette, M., R. Kaiser, M. Daumer, R. Petch, G. Fatkenheuer, H. Carls, J. K. Rockstroh, D. Schmaloer, J. Stechel, T. Feldt, H. Pfister, and D. Haussinger. 2006. Primary HIV drug resistance and efficacy of first-line antiretroviral

therapy guided by resistance testing. J Acquir Immune Defic Syndr **41**:573-581.

Palmer, S., V. Boltz, F. Maldarelli, M. Kearney, E. K. Halvas, D. Rock, J. Falloon, R. T. Davey, Jr., R. L. Dewar, J. A. Metcalf, J. W. Mellors, and J. M. Coffin. 2006. Selection and persistence of non-nucleoside reverse transcriptase inhibitor-resistant HIV-1 in patients starting and stopping non-nucleoside therapy. Aids **20**:701-710.

Palmer, S., M. Kearney, F. Maldarelli, E. K. Halvas, C. J. Bixby, H. Bazmi, D. Rock, J. Falloon, R. T. Davey, Jr., R. L. Dewar, J. A. Metcalf, S. Hammer, J. W. Mellors, and J. M. Coffin. 2005. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. J Clin Microbiol **43**:406-413.

Palmisano, L., M. Giuliano, C. M. Galluzzo, R. Amici, M. Andreotti, L. E. Weimer, M. F. Pirillo, V. Fragola, R. Bucciardini, and S. Vella. 2009. The mutational archive in proviral DNA does not change during 24 months of continuous or intermittent highly active antiretroviral therapy. HIV Med **10**:477-481.

Palmisano, L., and S. Vella. 2011. A brief history of antiretroviral therapy of HIV infection: success and challenges. Ann Ist Super Sanita **47**:44-48.

Paraskevis, D., O. Pybus, G. Magiorkinis, A. Hatzakis, A. M. Wensing, D. A. van de Vijver, J. Albert, G. Angarano, B. Asjo, C. Balotta, E. Boeri, R. Camacho, M. L. Chaix, S. Coughlan, D. Costagliola, A. De Luca, C. de Mendoza, I. Derdelinckx, Z. Grossman, O. Hamouda, I. Hoepelman, A. Horban, K. Korn, C. Kucherer, T. Leitner, C. Loveday, E. Macrae, I. Maljkovic-Berry, L. Meyer, C. Nielsen, E. L. Op de Coul, V. Ormaasen, L. Perrin, E. Puchhammer-Stockl, L. Ruiz, M. O. Salminen, J. C. Schmit, R. Schuurman, V. Soriano, J. Stanczak, M. Stanojevic, D. Struck, K. Van Laethem, M. Violin, S. Yerly, M. Zazzi, C. A. Boucher, and A. M. Vandamme. 2009. Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. Retrovirology **6**:49.

Paredes, R., C. M. Lalama, H. J. Ribaudo, B. R. Schackman, C. Shikuma, F. Giguel, W. A. Meyer, 3rd, V. A. Johnson, S. A. Fiscus, R. T. D'Aquila, R. M. Gulick, and D. R. Kuritzkes. 2010. Pre-existing minority drug-resistant HIV-1 variants, adherence, and risk of antiretroviral treatment failure. J Infect Dis **201**:662-671.

Peeters, M., and P. M. Sharp. 2000. Genetic diversity of HIV-1: the moving target. Aids **14 Suppl 3**:S129-140.

Petch, L. A., I. F. Hoffman, C. S. Jere, P. N. Kazembe, F. E. Martinson, D. Chilongozi, S. A. Fiscus, and M. S. Cohen. 2005. Genotypic analysis of the protease and reverse transcriptase of HIV type 1 subtype C isolates from antiretroviral drug-naive adults in Malawi. AIDS Res Hum Retroviruses **21**:799-805.

Petropoulos, C. J., N. T. Parkin, K. L. Limoli, Y. S. Lie, T. Wrin, W. Huang, H. Tian, D. Smith, G. A. Winslow, D. J. Capon, and J. M. Whitcomb. 2000. A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. Antimicrob Agents Chemother **44**:920-928.

Petti, C. A., C. R. Polage, T. C. Quinn, A. R. Ronald, and M. A. Sande. 2006. Laboratory medicine in Africa: a barrier to effective health care. Clin Infect Dis **42**:377-382.

Pettit, S. C., L. E. Everitt, S. Choudhury, B. M. Dunn, and A. H. Kaplan. 2004. Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. J Virol **78**:8477-8485.

Peuchant, O., R. Thiebaut, S. Capdepont, V. Lavignolle-Aurillac, D. Neau, P. Morlat, F. Dabis, H. Fleury, and B. Masquelier. 2008. Transmission of HIV-1 minority-resistant variants and response to first-line antiretroviral therapy. Aids **22**:1417-1423.

Pillay, C., H. Bredell, J. McIntyre, G. Gray, and L. Morris. 2002. HIV-1 subtype C reverse transcriptase sequences from drug-naive pregnant women in South Africa. AIDS Res Hum Retroviruses **18**:605-610.

Piot, P., T. C. Quinn, H. Taelman, F. M. Feinsod, K. B. Minlangu, O. Wobin, N. Mbendi, P. Mazebo, K. Ndangi, W. Stevens, and et al. 1984. Acquired immunodeficiency syndrome in a heterosexual population in Zaire. Lancet **2**:65-69.

Plantier, J. C., M. Leoz, J. E. Dickerson, F. De Oliveira, F. Cordonnier, V. Lemee, F. Damond, D. L. Robertson, and F. Simon. 2009. A new human immunodeficiency virus derived from gorillas. Nat Med **15**:871-872.

Pond, S. L., and S. D. Frost. 2005a. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics **21**:2531-2533.

Pond, S. L., and S. D. Frost. 2005b. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Molecular Biology and Evolution **22**:1208-1222.

Poon, A. F., R. A. McGovern, T. Mo, D. J. Knapp, B. Brenner, J. P. Routy, M. A. Wainberg, and P. R. Harrigan. 2011. Dates of HIV infection can be estimated for seroprevalent patients by coalescent analysis of serial next-generation sequencing data. Aids.

Poon, A. F., L. C. Swenson, W. W. Dong, W. Deng, S. L. Kosakovsky Pond, Z. L. Brumme, J. I. Mullins, D. D. Richman, P. R. Harrigan, and S. D. Frost. 2009. Phylogenetic analysis of population-based and deep sequencing data to identify coevolving sites in the nef gene of HIV-1. Mol Biol Evol **27**:819-832.

Popovic, M., M. G. Sarngadharan, E. Read, and R. C. Gallo. 1984. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science **224**:497-500.

Rahim, S., L. M. Fredrick, B. A. da Silva, B. Bernstein, and M. S. King. 2009. Geographic and temporal trends of transmitted HIV-1 drug resistance among antiretroviral-naive subjects screening for two clinical trials in North America and Western Europe. HIV Clin Trials **10**:94-103.

Rhee, S. Y., M. J. Gonzales, R. Kantor, B. J. Betts, J. Ravela, and R. W. Shafer. 2003. Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res **31**:298-303.

Rhee, S. Y., R. Kantor, D. A. Katzenstein, R. Camacho, L. Morris, S. Sirivichayakul, L. Jorgensen, L. F. Brigido, J. M. Schapiro, and R. W. Shafer. 2006. HIV-1 pol mutation frequency by subtype and treatment experience: extension of the HIVseq program to seven non-B subtypes. Aids **20**:643-651.

Ribeiro, R. M., L. Qin, L. L. Chavez, D. Li, S. G. Self, and A. S. Perelson. 2010. Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection. J Virol **84**:6096-6102.

Robertson, D. L., J. P. Anderson, J. A. Bradac, J. K. Carr, B. Foley, R. K. Funkhouser, F. Gao, B. H. Hahn, M. L. Kalish, C. Kuiken, G. H. Learn, T.

Leitner, F. McCutchan, S. Osmanov, M. Peeters, D. Pieniazek, M. Salminen, P. M. Sharp, S. Wolinsky, and B. Korber. 2000. HIV-1 nomenclature proposal. Science **288**:55-56.

Roche. 2011. GS FLX+ System

Sanger-like read lengths - the power of next-gen throughput

. www.454.com.

Roquebert, B., I. Malet, M. Wirden, R. Tubiana, M. A. Valantin, A. Simon, C. Katlama, G. Peytavin, V. Calvez, and A. G. Marcelin. 2006. Role of HIV-1 minority populations on resistance mutational pattern evolution and susceptibility to protease inhibitors. Aids **20**:287-289.

Ross, E. K., T. R. Fuerst, J. M. Orenstein, T. O'Neill, M. A. Martin, and S. Venkatesan. 1991. Maturation of human immunodeficiency virus particles assembled from the gag precursor protein requires in situ processing by gag-pol protease. AIDS Res Hum Retroviruses **7**:475-483.

Rozera, G., I. Abbate, A. Bruselles, C. Vlassi, G. D'Offizi, P. Narciso, G. Chillemi, M. Prosperi, G. Ippolito, and M. R. Capobianchi. 2009. Massively parallel pyrosequencing highlights minority variants in the HIV-1 env quasispecies deriving from lymphomonocyte sub-populations. Retrovirology **6**:15.

Sarafianos, S. G., B. Marchand, K. Das, D. M. Himmel, M. A. Parniak, S. H. Hughes, and E. Arnold. 2009. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. J Mol Biol **385**:693-713.

Saravanan, S., M. Vidya, P. Balakrishnan, N. Kumarasamy, S. S. Solomon, S. Solomon, R. Kantor, D. Katzenstein, B. Ramratnam, and K. H. Mayer. 2009. Evaluation of two human immunodeficiency virus-1 genotyping systems: ViroSeq 2.0 and an in-house method. J Virol Methods **159**:211-216.

Schacker, T., A. C. Collier, J. Hughes, T. Shea, and L. Corey. 1996. Clinical and epidemiologic features of primary HIV infection. Ann Intern Med **125**:257-264.

Schultz, A. K., M. Zhang, I. Bulla, T. Leitner, B. Korber, B. Morgenstern, and M. Stanke. 2009. jpHMM: improving the reliability of recombination prediction in HIV-1. Nucleic Acids Res **37**:W647-651.

Schultz, S. J., and J. J. Champoux. 2008. RNase H activity: structure, specificity, and function in reverse transcription. Virus Res **134**:86-103.

Schwartz, O., A. Dautry-Varsat, B. Goud, V. Marechal, A. Subtil, J. M. Heard, and O. Danos. 1995. Human immunodeficiency virus type 1 Nef induces accumulation of CD4 in early endosomes. J Virol **69**:528-533.

Serwadda, D., R. D. Mugerwa, N. K. Sewankambo, A. Lwegaba, J. W. Carswell, G. B. Kirya, A. C. Bayley, R. G. Downing, R. S. Tedder, S. A. Clayden, and et al. 1985. Slim disease: a new disease in Uganda and its association with HTLV-III infection. Lancet **2**:849-852.

Seyler, C., C. Adje-Toure, E. Messou, N. Dakoury-Dogbo, F. Rouet, D. Gabillard, M. Nolan, S. Toure, and X. Anglaret. 2007. Impact of genotypic drug resistance mutations on clinical and immunological outcomes in HIV-infected adults on HAART in West Africa. Aids **21**:1157-1164.

Shafer, R. W. 2006. Rationale and uses of a public HIV drug-resistance database. J Infect Dis **194 Suppl 1**:S51-58.

Shafer, R. W., S. Y. Rhee, and D. E. Bennett. 2008. Consensus drug resistance mutations for epidemiological surveillance: basic principles and potential controversies. Antivir Ther **13 Suppl 2**:59-68.

Shafer, R. W., S. Y. Rhee, D. Pillay, V. Miller, P. Sandstrom, J. M. Schapiro, D. R. Kuritzkes, and D. Bennett. 2007. HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. Aids **21**:215-223.

Shankarappa, R., R. Chatterjee, G. H. Learn, D. Neogi, M. Ding, P. Roy, A. Ghosh, L. Kingsley, L. Harrison, J. I. Mullins, and P. Gupta. 2001. Human immunodeficiency virus type 1 env sequences from Calcutta in eastern India: identification of features that distinguish subtype C sequences in India from other subtype C sequences. J Virol **75**:10479-10487.

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

Shet, A., L. Berry, H. Mohri, S. Mehandru, C. Chung, A. Kim, P. Jean-Pierre, C. Hogan, V. Simon, D. Boden, and M. Markowitz. 2006. Tracking the prevalence of transmitted antiretroviral drug-resistant HIV-1: a decade of experience. J Acquir Immune Defic Syndr **41**:439-446.

Shi, B., C. Kitchen, B. Weiser, D. Mayers, B. Foley, K. Kemal, K. Anastos, M. Suchard, M. Parker, C. Brunner, and H. Burger. 2010. Evolution and

recombination of genes encoding HIV-1 drug resistance and tropism during antiretroviral therapy. Virology **404**:5-20.

Shi, C., S. H. Eshleman, D. Jones, N. Fukushima, L. Hua, A. R. Parker, C. J. Yeo, R. H. Hruban, M. G. Goggins, and J. R. Eshleman. 2004. LigAmp for sensitive detection of single-nucleotide differences. Nat Methods **1**:141-147.

Simen, B. B., J. F. Simons, K. H. Hullsiek, R. M. Novak, R. D. Macarthur, J. D. Baxter, C. Huang, C. Lubeski, G. S. Turenchalk, M. S. Braverman, B. Desany, J. M. Rothberg, M. Egholm, and M. J. Kozal. 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.

Sluis-Cremer, N., D. Arion, and M. A. Parniak. 2000. Molecular mechanisms of HIV-1 resistance to nucleoside reverse transcriptase inhibitors (NRTIs). Cell Mol Life Sci **57**:1408-1422.

Solmone, M., D. Vincenti, M. C. Prosperi, A. Bruselles, G. Ippolito, and M. R. Capobianchi. 2009. Use of massively parallel ultradeep pyrosequencing to characterize the genetic diversity of hepatitis B virus in drug-resistant and drug-naive patients and to detect minor variants in reverse transcriptase and hepatitis B S antigen. J Virol **83**:1718-1726.

Somi, G. R., T. Kibuka, K. Diallo, T. Tuhuma, D. E. Bennett, C. Yang, C. Kagoma, E. F. Lyamuya, R. O. Swai, and S. Kassim. 2008. Surveillance of transmitted HIV drug resistance among women attending antenatal clinics in Dar es Salaam, Tanzania. Antivir Ther **13 Suppl 2**:77-82.

Soriano, V., and C. de Mendoza. 2002. Genetic mechanisms of resistance to NRTI and NNRTI. HIV Clin Trials **3**:237-248.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**:2688-2690.

Steegan, K. 2007. Evaluation of HIV-1 viral load and genotypic resistance assays for resource- limited settings

. Doctoral Thesis, Ghent University.

Steegen, K., S. Luchters, K. Dauwe, J. Reynaerts, K. Mandaliya, W. Jaoko, J. Plum, M. Temmerman, and C. Verhofstede. 2009. Effectiveness of antiretroviral

therapy and development of drug resistance in HIV-1 infected patients in Mombasa, Kenya. AIDS Res Ther **6**:12.

Steegen, K., S. Luchters, E. Demecheleer, K. Dauwe, K. Mandaliya, W. Jaoko, J. Plum, M. Temmerman, and C. Verhofstede. 2007. Feasibility of detecting human immunodeficiency virus type 1 drug resistance in DNA extracted from whole blood or dried blood spots. J Clin Microbiol **45**:3342-3351.

Stephenson, J. 2002. Cheaper HIV drugs for poor nations bring a new challenge: monitoring treatment. JAMA **288**:151-153.

Stuyver, L., A. Wyseur, A. Rombout, J. Louwagie, T. Scarcez, C. Verhofstede, D. Rimland, R. F. Schinazi, and R. Rossau. 1997. Line probe assay for rapid detection of drug-selected mutations in the human immunodeficiency virus type 1 reverse transcriptase gene. Antimicrob Agents Chemother **41**:284-291.

Suzuki, Y. 2004. New methods for detecting positive selection at single amino acid sites. J Mol Evol **59**:11-19.

Suzuki, Y., and R. Craigie. 2007. The road to chromatin - nuclear entry of retroviruses. Nat Rev Microbiol **5**:187-196.

Suzuki, Y., and T. Gojobori. 1999. A method for detecting positive selection at single amino acid sites. Mol Biol Evol **16**:1315-1328.

Swanstrom, R., R. J. Bosch, D. Katzenstein, H. Cheng, H. Jiang, N. Hellmann, R. Haubrich, S. A. Fiscus, C. V. Fletcher, E. P. Acosta, and R. M. Gulick. 2004. Weighted phenotypic susceptibility scores are predictive of the HIV-1 RNA response in protease inhibitor-experienced HIV-1-infected subjects. J Infect Dis **190**:886-893.

Tatt, I. D., K. L. Barlow, A. Nicoll, and J. P. Clewley. 2001. The public health significance of HIV-1 subtypes. Aids **15 Suppl 5**:S59-71.

Taylor, B. S., and S. M. Hammer. 2008. The challenge of HIV-1 subtype diversity. N Engl J Med **359**:1965-1966.

Thomson, M. M., and A. Fernandez-Garcia. 2011. Phylogenetic structure in African HIV-1 subtype C revealed by selective sequential pruning. Virology **415**:30-38.

Thomson, M. M., L. Perez-Alvarez, and R. Najera. 2002. Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. Lancet Infect Dis **2**:461-471.

Towler, W. I., L. Barlow-Mosha, J. D. Church, D. Bagenda, P. Ajuna, M. Mubiru, P. Musoke, and S. H. Eshleman. 2010. Analysis of drug resistance in children receiving antiretroviral therapy for treatment of HIV-1 infection in Uganda. AIDS Res Hum Retroviruses **26**:563-568.

Trask, S. A., C. A. Derdeyn, U. Fideli, Y. Chen, S. Meleth, F. Kasolo, R. Musonda, E. Hunter, F. Gao, S. Allen, and B. H. Hahn. 2002. Molecular epidemiology of human immunodeficiency virus type 1 transmission in a heterosexual cohort of discordant couples in Zambia. J Virol **76**:397-405.

Travers, S. A., J. P. Clewley, J. R. Glynn, P. E. Fine, A. C. Crampin, F. Sibande, D. Mulawa, J. O. McInerney, and G. P. McCormack. 2004. Timing and reconstruction of the most recent common ancestor of the subtype C clade of human immunodeficiency virus type 1. J Virol **78**:10501-10506.

Tsibris, A. M., B. Korber, R. Arnaout, C. Russ, C. C. Lo, T. Leitner, B. Gaschen, J. Theiler, R. Paredes, Z. Su, M. D. Hughes, R. M. Gulick, W. Greaves, E. Coakley, C. Flexner, C. Nusbaum, and D. R. Kuritzkes. 2009. Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. PLoS One **4**:e5683.

UNAIDS. 2009a. AIDS epidemic update.

UNAIDS. 2010. Report on the global AIDS epidemic: Global facts and figures.

UNAIDS. 2011a. Global HIV/AIDS response, Epidemic update and health sector progress towards Universal Access. Progress Report.

UNAIDS. 2009b. Report on the global AIDS epidemic: Global facts and figures.

UNAIDS. 2011b. World AIDS Day Report 2011.

UNAIDS/WHO. 2007. AIDS Epidemic update. In Geneva.

Van de Perre, P., D. Rouvroy, P. Lepage, J. Bogaerts, P. Kestelyn, J. Kayihigi, A. C. Hekker, J. P. Butzler, and N. Clumeck. 1984. Acquired immunodeficiency syndrome in Rwanda. Lancet **2**:62-65.

Van Laethem, K., A. De Luca, A. Antinori, A. Cingolani, C. F. Perna, and A. M. Vandamme. 2002. A genotypic drug resistance interpretation algorithm that significantly predicts therapy response in HIV-1-infected patients. Antivir Ther **7**:123-129.

van Maarseveen, N., and C. Boucher. 2006. Resistance to protease inhibitors.

Vandamme, A. M., K. Van Laethem, and E. De Clercq. 1999. Managing resistance to anti-HIV drugs: an important consideration for effective disease management. Drugs **57**:337-361.

Vandegraaff, N., and A. Engelman. 2007. Molecular mechanisms of HIV integration and therapeutic intervention. Expert Rev Mol Med **9**:1-19.

Varella, R. B., C. G. Schrago, and M. G. Zalis. 2009. Differential evolution of Human Immunodeficiency Virus type 1 Protease and Reverse Transcriptase genes between HAART-failing and naive-treated individuals. Curr HIV Res **7**:601-605.

Varghese, V., R. Shahriar, S. Y. Rhee, T. Liu, B. B. Simen, M. Egholm, B. Hanczaruk, L. A. Blake, B. Gharizadeh, F. Babrzadeh, M. H. Bachmann, W. J. Fessel, and R. W. Shafer. 2009. Minority variants associated with transmitted and acquired HIV-1 nonnucleoside reverse transcriptase inhibitor resistance: implications for the use of second-generation nonnucleoside reverse transcriptase inhibitors. J Acquir Immune Defic Syndr **52**:309-315.

Varghese, V., E. Wang, F. Babrzadeh, M. H. Bachmann, R. Shahriar, T. Liu, S. J. Mappala, B. Gharizadeh, W. J. Fessel, D. Katzenstein, S. Kassaye, and R. W. Shafer. 2010. Nucleic acid template and the risk of a PCR-Induced HIV-1 drug resistance mutation. PLoS One **5**:e10992.

Vazquez, M. L., M. L. Bryant, M. Clare, G. A. DeCrescenzo, E. M. Doherty, J. N. Freskos, D. P. Getman, K. A. Houseman, J. A. Julien, G. P. Kocan, and et al. 1995. Inhibitors of HIV-1 protease containing the novel and potent (R)-(hydroxyethyl)sulfonamide isostere. J Med Chem **38**:581-584.

Vercauteren, J., A. M. Wensing, D. A. van de Vijver, J. Albert, C. Balotta, O. Hamouda, C. Kucherer, D. Struck, J. C. Schmit, B. Asjo, M. Bruckova, R. J. Camacho, B. Clotet, S. Coughlan, Z. Grossman, A. Horban, K. Korn, L. Kostrikis, C. Nielsen, D. Paraskevis, M. Poljak, E. Puchhammer-Stockl, C. Riva, L. Ruiz, M. Salminen, R. Schuurman, A. Sonnerborg, D. Stanekova, M. Stanojevic, A. M. Vandamme, and C. A. Boucher. 2009. Transmission of drug-resistant HIV-1 is stabilizing in Europe. J Infect Dis **200**:1503-1508.

Vidal, N., T. Niyongabo, J. Nduwimana, C. Butel, A. Ndayiragije, J. Wakana, M. Nduwimana, E. Delaporte, and M. Peeters. 2007. HIV type 1 diversity and

antiretroviral drug resistance mutations in Burundi. AIDS Res Hum Retroviruses **23**:175-180.

von Schwedler, U. K., M. Stuchell, B. Muller, D. M. Ward, H. Y. Chung, E. Morita, H. E. Wang, T. Davis, G. P. He, D. M. Cimbora, A. Scott, H. G. Krausslich, J. Kaplan, S. G. Morham, and W. I. Sundquist. 2003. The protein network of HIV budding. Cell **114**:701-713.

Waheed, A. A., and E. O. Freed. 2012. HIV type 1 Gag as a target for antiviral therapy. AIDS Res Hum Retroviruses **28**:54-75.

Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon. 1985. Nucleotide sequence of the AIDS virus, LAV. Cell **40**:9-17.

Walker, P. R., O. G. Pybus, A. Rambaut, and E. C. Holmes. 2005. Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic growth. Infect Genet Evol **5**:199-208.

Wang, C., Y. Mitsuya, B. Gharizadeh, M. Ronaghi, and R. W. Shafer. 2007. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res **17**:1195-1201.

Wang, J., J. Sen, L. Rong, and M. Caffrey. 2008. Role of the HIV gp120 conserved domain 1 in processing and viral entry. J Biol Chem **283**:32644-32649.

Watts, J. M., K. K. Dang, R. J. Gorelick, C. W. Leonard, J. W. Bess, Jr., R. Swanstrom, C. L. Burch, and K. M. Weeks. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature **460**:711-716.

Weidle, P. J., R. Downing, C. Sozi, R. Mwebaze, G. Rukundo, S. Malamba, R. Respess, K. Hertogs, B. Larder, D. Ochola, J. Mermin, B. Samb, and E. Lackritz. 2003. Development of phenotypic and genotypic resistance to antiretroviral therapy in the UNAIDS HIV Drug Access Initiative--Uganda. Aids **17 Suppl 3**:S39-48.

Wheeler, W. H., R. A. Ziebell, H. Zabina, D. Pieniazek, J. Prejean, U. R. Bodnar, K. C. Mahle, W. Heneine, J. A. Johnson, and H. I. Hall. 2010. Prevalence of transmitted drug resistance associated mutations and HIV-1 subtypes in new HIV-1 diagnoses, U.S.-2006. AIDS **24**:1203-1212.

WHO. 2006a. Antiretroviral therapy for HIV infection in adults and adolescents: towards universal access. Recommendations for a public health approach. In. Geneva: World Health Organization; .

WHO. 2006b. Progress on Global Access to HIV Antiretroviral Therapy. A report on "3 by 5" and beyond.

Willey, R. L., F. Maldarelli, M. A. Martin, and K. Strebel. 1992a. Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. J Virol **66**:7193-7200.

Willey, R. L., F. Maldarelli, M. A. Martin, and K. Strebel. 1992b. Human immunodeficiency virus type 1 Vpu protein regulates the formation of intracellular gp160-CD4 complexes. J Virol **66**:226-234.

Wirden, M., C. Soulie, M. A. Valantin, S. Fourati, A. Simon, S. Lambert-Niclot, M. Bonmarchand, C. Clavel-Osorio, A. G. Marcelin, C. Katlama, and V. Calvez. 2011. Historical HIV-RNA resistance test results are more informative than proviral DNA genotyping in cases of suppressed or residual viraemia. J Antimicrob Chemother **66**:709-712.

Worobey, M., M. Gemmel, D. E. Teuwen, T. Haselkorn, K. Kunstman, M. Bunce, J. J. Muyembe, J. M. Kabongo, R. M. Kalengayi, E. Van Marck, M. T. Gilbert, and S. M. Wolinsky. 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature **455**:661-664.

Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155**:431-449.

Zaccarelli, M., V. Tozzi, P. Lorenzini, F. Forbici, P. Narciso, F. Ceccherini-Silberstein, M. P. Trotta, A. Bertoli, G. Liuzzi, P. Marconi, S. Mosti, C. F. Perno, and A. Antinori. 2007. The V118I mutation as a marker of advanced HIV infection and disease progression. Antivir Ther **12**:163-168.

Zanetti, G., J. A. Briggs, K. Grunewald, Q. J. Sattentau, and S. D. Fuller. 2006. Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ. PLoS Pathog **2**:e83.

Zhang, H., Y. Zhou, C. Alcock, T. Kiefer, D. Monie, J. Siliciano, Q. Li, P. Pham, J. Cofrancesco, D. Persaud, and R. F. Siliciano. 2004. Novel single-cell-level phenotypic assay for residual drug susceptibility and reduced replication capacity of drug-resistant human immunodeficiency virus type 1. J Virol **78**:1718-1729.

Zhang, M., A. K. Schultz, C. Calef, C. Kuiken, T. Leitner, B. Korber, B. Morgenstern, and M. Stanke. 2006. jpHMM at GOBICS: a web server to detect genomic recombinations in HIV-1. Nucleic Acids Res **34**:W463-465.

Zhu, P., J. Liu, J. Bess, Jr., E. Chertova, J. D. Lifson, H. Grise, G. A. Ofek, K. A. Taylor, and K. H. Roux. 2006. Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature **441**:847-852.

Zhu, T., B. T. Korber, A. J. Nahmias, E. Hooper, P. M. Sharp, and D. D. Ho. 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature **391**:594-597.

Zijlstra, E. E., and J. J. van Oosterhout. 2006. [The introduction of antiretroviral therapy in Malawi]. Ned Tijdschr Geneeskd **150**:2774-2778.

Appendices

Appendix

Appendix Table 1: The nucleotide sequences of GS FLX Titanium Fusion Primers with MID and template specific PCR primers. For each sample, Primer A represents a forward primer and Primer B represents reverse primer.

Patient		GS FLX Primers (Last 4 bases are 'key')	MID	Template specific Primers
Pt2_0	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	ACGAGTGCGT	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	ACGAGTGCGT	TCCCACTAACTTCTGTATATC
Pt2_11	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	TCTCTATGCG	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	TCTCTATGCG	TCCCACTAACTTCTGTATATC
Pt2_14	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	TACTGAGCTA	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	TACTGAGCTA	TCCCACTAACTTCTGTATATC
Pt32_0	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	ATCAGACACG	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	ATCAGACACG	TCCCACTAACTTCTGTATATC
Pt32_ 7	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	ATATCGCGAG	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	ATATCGCGAG	TCCCACTAACTTCTGTATATC
Pt32_16	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	TGATACGTCT	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	TGATACGTCT	TCCCACTAACTTCTGTATATC
Pt32_24	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	CGTGTCTCTA	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	CGTGTCTCTA	TCCCACTAACTTCTGTATATC
Pt42_0_08	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	ACGCTCGACA	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	ACGCTCGACA	TCCCACTAACTTCTGTATATC
Pt42_0_09	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	TAGTATCAGC	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	TAGTATCAGC	TCCCACTAACTTCTGTATATC
Pt45_0	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	CATAGTAGTG	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	CATAGTAGTG	TCCCACTAACTTCTGTATATC
Pt45_8	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	CGAGAGATAC	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	CGAGAGATAC	TCCCACTAACTTCTGTATATC
Pt45_13	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	TCACGTACTA	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	TCACGTACTA	TCCCACTAACTTCTGTATATC
Pt76_0	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	AGCACTGTAG	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	AGCACTGTAG	TCCCACTAACTTCTGTATATC
Pt76_6	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	CTCGCGTGTC	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	CTCGCGTGTC	TCCCACTAACTTCTGTATATC
Pt76_9	Primer A	CGTATCGCCTCCCTCGCGCCATCAG	AGACGCACTC	CCAGTAAAATTAAAGCCA
	Primer B	CTATGCGCCTTGCCAGCCCGCTCAG	AGACGCACTC	TCCCACTAACTTCTGTATATC

Appendix Table 2: Number of segregated reads obtained after processing the reads obtained after data cleaning. Reads obtained after quality control, were segregated into five categories on the basis of their length.

Nt Length	Pt2_0	Pt2_11	Pt2_14	Pt32_0	Pt32_7	Pt32_16	Pt32_24	Pt42BL1	Pt42BL2	Pt45_0	Pt45_8	Pt45_13	Pt76_0	Pt76_6	Pt76_9
<49	188	439	596	591	433	134	318	157	328	2039	430	422	1680	2008	590
50-99	9244	6428	9021	4318	4786	1548	5670	5540	3092	4343	4460	3955	7296	11042	5655
100-149	2890	1893	2373	1573	2157	430	2271	1170	716	1465	1518	1257	2421	2939	1975
150-199	1384	1022	1291	2023	2053	381	2073	1041	480	1697	1069	872	2200	3492	1879
200-249	5662	3566	4099	3587	6614	1529	5564	1837	2098	6000	4765	3722	3048	7706	3389
>250	7125	4705	8116	4222	3839	965	3966	5415	1757	4238	3196	3131	7266	6687	4959
Total	26493	18053	25496	16314	19882	4987	19862	15160	8471	19782	15438	13359	23911	33874	18447

Appendix Table 3: The numbers of unique sequence and amino acid variants in the sequences obtained by all 15 samples from five patients. The first column represents the numbers of sequence reads in a sequence set of greater than 250 nucleotides followed by details of sequences with forward and reverse strands including number of reads, length after making a uniform sequence length alignment, nucleotide variants, percentage variants in total number of reads.

Pt76_9	4959	1406	252	602	342	3553	246	1483	908	2085	1250
Patient and Time point	**Reads (>250nt)**	**Forward**				**Reverse**				**Total nucleotide variants**	**Total amino acid variants**
		Reads	**Length (nt)**	**Nt Variants**	**AA Variants**	**Reads**	**Length (nt)**	**Nt Variants**	**AA Variants**		
Pt2_0	7107	2155	258	759	469	4952	240	1396	869	2155	1338
Pt2_11	4701	1174	258	594	426	3527	237	1494	995	2088	1421
Pt2_14	8112	2129	252	644	491	5983	237	1640	1033	2284	1524
Pt32_0	4222	950	258	295	223	3272	237	664	473	959	696
Pt32_7	3839	1109	255	367	266	2730	237	703	504	1070	770
Pt32_16	965	449	255	150	119	516	243	151	119	301	238
Pt32_24	3966	1815	240	524	371	2148	237	641	471	1165	842
Pt42_BL1	5415	1267	252	535	246	4148	237	1335	618	1870	864
Pt42_BL2	1757	560	255	254	182	1197	240	568	350	822	532
Pt45_0	4285	1191	222	412	214	3094	246	547	390	959	604
Pt45_8	3192	1332	243	538	334	1860	222	644	384	1182	718
Pt45_13	3131	1320	249	515	291	1811	258	804	614	1319	905
Pt76_0	7265	1952	249	824	472	5313	237	1970	1081	2794	1553
Pt76_6	6686	2300	255	1046	584	4386	237	1540	1006	2586	1590

Publications

AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 26, Number 00, 2010

DOI: 10.1089/aid.2010.0203

Drug Resistance Mutations in Drug-Naive HIV Type 1 Subtype C-Infected Individuals from Rural Malawi

Vijay Bansode,[1] Zuzanna J. Drebert,[1] Simon A.A. Travers,[1,2] Emmanuel Banda,[3] Anna Molesworth,[3,4] Amelia Crampin,[3,4] Bagrey Ngwira,[3,4] Neil French,[3,4] Judith R. Glynn,[3] and Grace P. McCormack[1]

Abstract

In this preliminary study we show that in 2008, 3 years after antiretroviral therapy was introduced into the Karonga District, Malawi, a greater than expected number of drug-naive individuals have been infected with HIV-1 subtype C virus harboring major and minor drug resistance mutations (DRMs). From a sample size of 40 reverse transcriptase (RT) consensus sequences from drug-naive individuals we found five showing NRTI and four showing NNRTI mutations with one individual showing both. From 29 protease consensus sequences, again from drug-naive individuals, we found evidence of minor DRMs in three. Additional major and minor DRMs were found in clonal sequences from a number of individuals that were not present in the original consensus sequences. This clearly illustrates the importance of sequencing multiple HIV-1 variants from individuals to fully assess drug resistance.

Malawi has an HIV-1 prevalence of around 12% (UNAIDS 2009). Although the prevalence is higher in urban areas, 80% of the population lives in rural areas, and the epidemic in these areas remains a concern and access to treatment more problematic. Over 200,000 people in Malawi have been on antiretroviral therapy (ART) and approximately 147,000 are currently on treatment.[1] Around 95% of them are on first-line therapy, 4% are on alternative first-line therapy, and less than 1% are on second-line therapy.[1] A fundamental concern of antiretroviral drug delivery programs is the avoidance of the development of widespread drug resistance. Drug resistance mutations (DRMs) have been primarily characterized for subtype B, which is predominant in America, Europe, and Australia where ART has been available for a decade and half, and where approximately 10% of new HIV infections are with drug-resistant isolates.[2] Fewer studies have characterized drug resistance in subtype C, which is responsible for more than 50% of worldwide infections. However, minor mutations associated with drug resistance have been reported in HIV-1 subtype C-infected drug-naive patients in a number of African locations including Malawi, and the percentage of drug-naive individuals harboring HIV with DRMs thus far has been shown to vary between 2% and 10%.[3–6] Further characterization of drug resistance mutations in subtype-C HIV-1 and monitoring for transmission of drug-resistant strains are extremely important both given the more recent wide-scale availability of ART in countries in which subtype C is predominant, and as the prevalence of non-B HIV-1 increases in countries in which antiretroviral drugs are widely used.

Karonga District is the northernmost district of Malawi and HIV-1 subtype C has been reported to be the most predominant subtype in this District.[7] ART drugs were introduced in Malawi first in the context of prevention of maternal-to-child transmission (PMTCT) through antenatal clinics (using a mother and child nevirapine regime, which is ongoing). Full ART first became available at public clinics in the two major cities in Malawi in 2003, in the regional capital in 2004, and at Karonga district hospital in June 2005, at which time Karonga residents who had been seeking care outside the district were transferred to local services. ART delivery has been largely based on clinical assessment with CD4 counting becoming available in January 2008 at one clinic. Viral load assessment and routine drug resistance testing are not available and the baseline genotype of the genes, which are the targets for ART, are unknown. The objective of this study was to investigate the presence of polymorphisms associated with drug resistance found in the reverse transcriptase (RT) and protease (PR) regions and to explore the intrapatient diversity of HIV-1 in these gene regions at a time when most individuals seen are

[1]Zoology Department, Martin Ryan Institute, School of Natural Sciences, National University of Ireland, Galway, Ireland.
[2]South African National Bioinformatics Institute, University of the Western Cape, Bellville, South Africa.
[3]Karonga Prevention Study, Chilumba, Malawi.
[4]Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK.

drug naive and transmission of drug resistance from ART-treated individuals should be uncommon.

DNA was extracted from blood samples from 71 individuals living in Karonga District, Malawi using the QIAamp DNA Blood Mini Kit (QIAGEN Ltd). Extracted DNA was subjected to nested polymerase chain reaction (PCR) amplification of the HIV-1 PR and RT genes employing the primers previously described.[6] The PCR mixture contained 1.5 mM $MgCl_2$, 1.5 units of *Taq* (Roche Expand High Fidelity PCR system), 0.8 mM each dNTP, and 5 pM primers (Eurofins MWG Operon). The cycling conditions were as follows: "hotstart" at 98°C for 1 min, primary denaturation at 94°C for 2 min followed by 35 cycles with 1 min denaturation at 94°C, 1 min annealing at 42°C, 4 min extension at 72°C, and final extension for 7 min at 72°C. For each gene region from 10 samples two PCR amplifications were pooled and the mixture cloned using the TOPO Cloning Kit (Invitrogen). Colony PCR and sequencing were carried out on 30 colonies per sample. Sequences obtained were edited in SeqMan (Lasergene, DNASTAR, Inc) and aligned using MacClade 4.0 (Sinauer Associates).

Preliminary phylogenetic analyses were first carried out to identify the subtype status of all sequences generated. Subsequently all consensus and clonal sequences that were subtype C were assembled into multiple alignments for each gene region using MacClade 4.0. Phylogenetic trees were reconstructed using the LANL (www.hiv.lanl.gov/content/index) subtype C ancestral sequence as outgroup under the GTR + gamma model of DNA substitution implemented RAxML 7.0.3[8] with all parameters optimized by RAxML. Confidence levels in the groupings in the phylogeny were assessed using 1000 bootstrap replicates as part of the RAxML phylogeny reconstruction. To evaluate the effect of DRMs on clustering of individual sequences, analyses were also carried out with all sites showing the presence of DRMs removed. Sequences were submitted for analysis of DRMs to the Stanford Database (http://hivdb.stanford.edu).

Of the 71 DNA samples, amplification and sequencing were successful from 62 for RT and from 53 for PR. Three RT and six PR sequences were excluded from analysis because of poor quality and four individuals were infected with non-subtype C virus (no DRMs were present in these non-C sequences). For both gene fragments from subtype C strains phylogenetic analyses with and without DRMS present showed identical tree topologies indicating that sharing the same DRM was not, by itself, responsible for individual sequences clustering together. Figure 1 shows the relationships between sequences with DRMS included.

RTsequences included sequences from 40 individuals who were drug naive and from 17 who had been on ART. None of the drug-exposed individuals showed any DRMs, however, five (5/40) drug-naïve individuals showed nucleoside reverse transcriptase inhibitor (NRTI) resistance mutations and four (4/40) showed non-NRTI (NNRTI) resistance mutations with both types of drug resistance mutations found in one individual (Table 1). The NRTI mutations found were V75LV (one individual) and V118I (four individuals) with five NNRTI mutations identified (A98G, K101Q, V106LV, E138A, and G190R) (Table 1). V118I occurs in ~2% of untreated persons infected with subtype C and with increased frequency in persons receiving multiple NRTIs.[9] It causes low-level resistance to 3TC and possibly to other NRTIs when present with other mutations.[10] This was the only mutation found in both previous studies of drug resistance mutations in Malawi, being present in 4 of 21 individuals studied by Petch *et al*.[3] and 15 of 96 individuals with confirmed virologic failure by Hosseinipour *et al*.[11] Of the other drug resistance mutations found K101Q minimally reduces susceptibility to each of the NNRTIs, whereas mutation at position 75 (V75T/M/A/I) is associated with reduced NRTI susceptibility (http://hivdb.stanford.edu).

Table 1. Mutations Associated with Antiretroviral Drug Resistance Found in Consensus Sequences from HIV$^+$ Individuals from Karonga District Malawi[a]

	Reverse transcriptase		
Sample no.	*NRTI*	*NNRTI*	*Protease*
Drug exposed			
54302	**No sequence**		**T74S**
53972	**No sequence**		**L76S**
53278	**None**	**None**	**T74S**
53974	**None**	**None**	**L10V, T74S**
Drug naïve			
54208	No sequence		T74S
53394	V118I	None	None
53436	V118I	None	No sequence
53733	V118I	None	No sequence
53900	V118I	None	None
54434	V75LV	V106LV	None
53977	None	None	T74S
54023	None	K101Q, E138A	T74S
54117	None	G190R	None
54357	None	A98G	No sequence

[a]Mutations involved in resistance to NRTI and NNRTI drugs and minor mutations involved in resistance to protease inhibitors are shown. The four shown in bold were previously drug exposed prior to enrolling in the study.

According to the Stanford database G190R is a highly unusual mutation and is flagged as an NNRTI resistance mutation. Little seems to be known about the effects of this mutation as yet in subtype C but G190A/S/E/Q/T/V/C are NNRTI resistance mutations. Similarly, although V106A causes high-level resistance to NVP and DLV and low to intermediate resistance to EFV, V106L is a rare polymorphism and its association with NNRTI resistance is less clear; E138A is responsible for decreased ETR response and A98G reduces NVP susceptibility by two- to threefold (http://hivdb.stanford.edu).

Cloning was successful for seven samples (Table 2). From individuals whose consensus sequence showed neither NRTI nor NNRTI drug resistance mutations, none of the clonal sequences generated showed drug resistance mutations (Table 2). All cloned sequences obtained from individual 53733 contained the NRTI V118I mutation that was found in the consensus sequence and one clonal sequence showed an additional L74K NRTI-resistant mutation. The consensus sequence retrieved from individual 54117 showed a G190R (NNRTI) mutation and also had four stop codons at positions 71, 88, 212, and 239 (according to HXB2 RT gene). All clones obtained from this individual showed the same four stop codons and the same G190R mutation. This study utilized proviral DNA to generate RT and PR sequences due to the

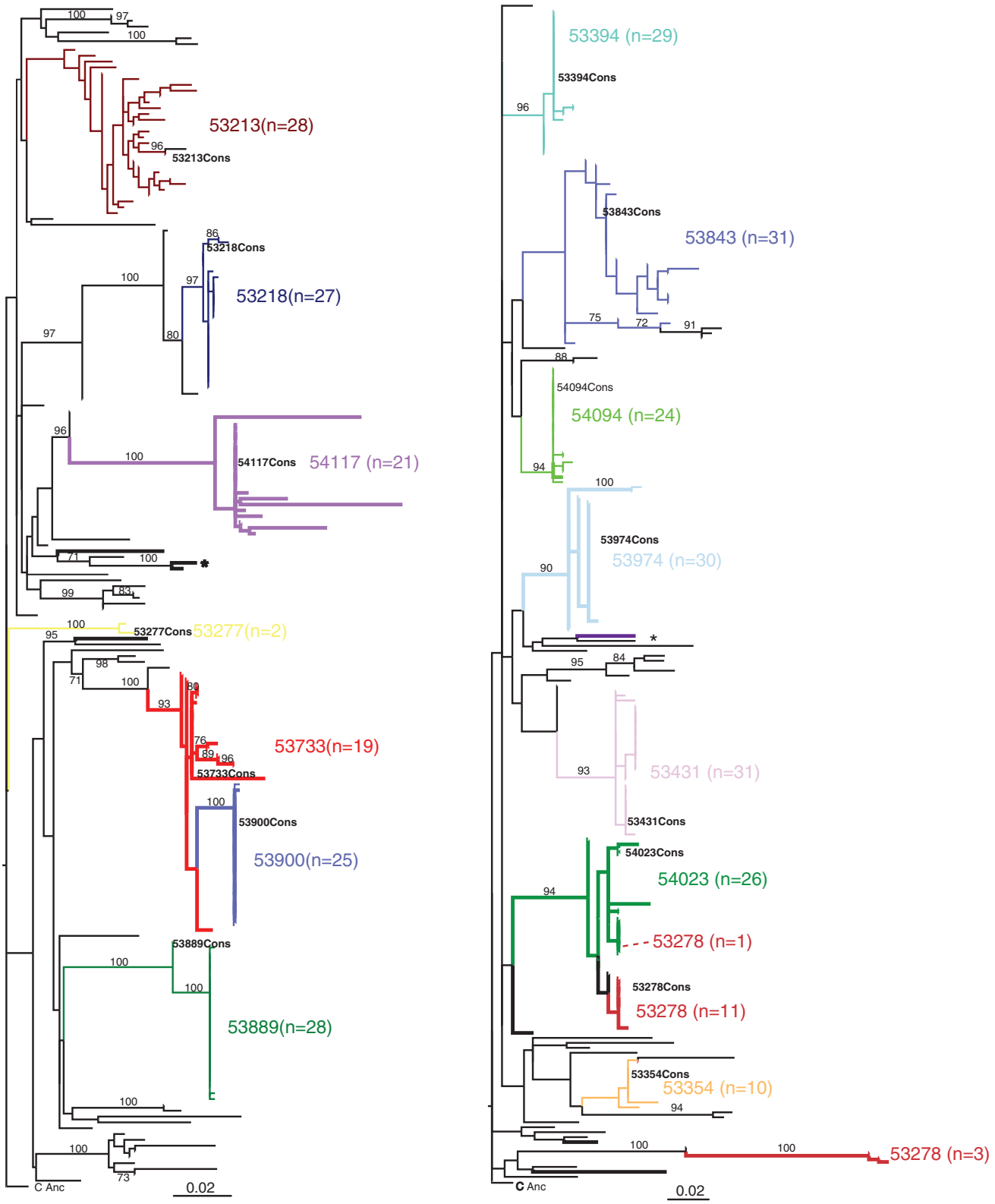


FIG. 1. Maximum likelihood trees generated from reverse transcriptase **(a)** and protease **(b)** genes from HIV-1-positive individuals from Malawi. Branches in color represent sequences from multiple clones of different individuals. Branches in bold are those showing drug resistance mutations. The numbers on the branches are bootstrap values. ***(a)** Two consensus sequences showing the presence of drug resistance mutations grouping together with high support. ***(b)** Consensus sequences from a husband and wife pair that group together with high bootstrap support.

Table 2. Summary of Amino Acid Diversity in Multiple Sequences Retrieved from Individuals Showing Evidence of NRTI and NNRTI Resistance Mutations in the Original Consensus Sequence and from Those That Did Not[a]

Sample no.	Mutation in original sequence		Number of cloned sequences obtained	NRTI mutations found in clones	NNRTI mutations found in clones
	NRTI	NNRTI			
53277	**None**	**None**	1	**None**	**None**
53213	**None**	**None**	27	**None**	**None**
53218	**None**	**None**	26	**None**	**None**
					None
53889[b]	**None**	**None**	27	**None**	**None**
53733	V118I	None	18	L74K, V118I	None
53900	V118I	None	24	V118I	None
54117[c]	None	G190R	20	None	G190R

[a]Drug-exposed individuals are in bold.
[b]Consensus sequence was different from each of the clones.
[c]All clones and consensus contained four stop codons.

difficulty of working with RNA in this context (i.e., getting samples from a large population study in rural Africa). HIV-1 with a G–A hypermutation at positions 88 and 212 has previously been found in resting CD4 cells and the resulting DNA is reported to be degraded or defective.[12,13] It is possible that the plasma of this individual would yield functional HIV-1. It is also possible that the individual may be infected with a defective virus and that this might affect disease progression.

The phylogenetic relationships of the RT sequences from most individuals were straightforward with both the clonal and consensus sequences grouping together in monophyletic groups, with high bootstrap support (Fig. 1a). The clones and the consensus sequence from 53733 appear to be ancestral to the virus infecting individual 53900 with both of these individuals infected with viral strains containing the same NRTI DRM (V118I). There is no known epidemiological link between these individuals. The laboratory work on these samples was carried out by different people and therefore contamination is not likely in this case. Of the individuals for whom the PCR products were not cloned there is only one instance of clustering of sequences containing DRMs with high bootstrap support (see the asterisk in Fig. 1a). One of these individuals is a male and the other is a female. These individuals are in the same reporting group and thus may live geographically close together.

Of 45 subtype C protease sequences, 16 were from drug-exposed individuals and 29 were from drug-naive individuals. Seven consensus sequences showed minor PI mutations including three from drug-naive individuals (Table 1). Cloning was successful for eight samples, three for which the consensus had shown minor resistance sites and five for which the consensus showed no drug resistance mutations. Three drug-exposed individuals (53278, 54302, and 53974) showed minor PR resistance mutations in clonal sequences that were not seen in the consensus sequences, whereas 53278 also showed evidence of major PR resistance mutations in the sequenced clones (Table 3). Furthermore, this pattern was repeated among the drug-naive individuals with major and minor DRMs present in clones that were not present in the consensus sequence (Table 3). Clones of 53278 showed major mutations D30N, M46I, and I84T. D30N causes high-level resistance to NFV and potential low-level resistance to ATV. M46I decreases susceptibility to IDV, NFV, FPV, LPV, and ATV when present with other mutations. I84V causes intermediate to high-level resistance to various PIs. Minor mutations at position 10 (L10I/V/F/R/Y) are associated with

Table 3. Summary of Amino Acid Diversity in Multiple Sequences Retrieved from Individuals Showing Evidence of Protease Inhibitor Resistance Mutations in the Original Consensus Sequence and from Those That Did not[a]

Sample no.	Mutation in original sequence		Number of cloned sequences obtained	PI resistance major mutations found in clones	PI resistance minor mutations found in clones
	Major	Minor			
53278	**None**	**T74S**	**14**	**D30N, M46I, I84T**	**G48R, T74S**
53354	**None**	**None**	**9**	**None**	**None**
53974	**None**	**L10V, T74S**	**29**	**None**	**L10V, G48R, T74S**
54023	None	T74S	25	None	M46T, T74S
53394	None	None	28	None	None
53431	None	None	30	None	None
53843	None	None	30	None	None
54094	None	None	23	I47M	None

[a]Drug-exposed indivudals are in bold.

resistance to most PIs when present with other mutations and mutation at this site occurs in 5–10% of untreated persons including subtype C-infected individuals in Africa.[4,5,11,14] Mutation at position 74 (T74S) is associated with reduced NFV susceptibility whereas mutation at position 76 (L76V) reduces susceptibility to FPV, IDV, LPV, and DRV and increases susceptibility to SQV, ATV, and TPV.[14,15]

For most individuals phylogenetic relationships between the sequences were uncomplicated; however, while 10 clones and the consensus sequence from individual 53278 clustered closely and all showed the T74S minor PI mutation, sequences of three variants clustered together in the tree with sequences retrieved from a different individual with whom she has no known epidemiological links (Fig. 1b) rather than with other 53278 sequences. These three sequence variants were those that showed major PR inhibitor resistance mutations. Furthermore, none of these three variants showed the original T74S mutation but did show another minor PI mutation (G48R). One additional clone sequence was found clustering with clonal sequences from another female, 54023, again with whom she had no known link.

This work clearly shows the presence of both NRTI and NNRTI drug resistance mutations and the presence of minor and major PR resistance mutations in a number of drug-naive individuals in rural Malawi in 2008 3 years after ART became widely available in the district. We have also shown the importance of sequencing multiple HIV-1 variants from individuals to fully assess drug resistance, as a number of individuals have shown major and/or minor drug resistance mutations in sequenced clones that were not present in the consensus sequences. Sequencing multiple clones of RT and PR from 16 individuals showed that one of them (female, aged 52, 53278 in Fig. 1b) was dual infected and showed evidence of being infected with viruses containing DRMs. The dual infection was confirmed by sequencing of multiple clones of the *gag* gene as part of another project, which showed a similar pattern of multiple lineages for this individual (Seager, unpublished). In terms of identifying possible transmission events involving HIV-1 with DRMs, apart from the dually infected individual mentioned above, there are very few cases of clustering of sequences from individuals with DRMS.

Our analyses showed that the presence of DRMs alone was not responsible for the pattern of relationships shown on the trees indicating common ancestry rather than convergence. However, given the short length of gene fragments used to explore drug resistance and in many cases having sequence evidence from only one gene, such results are not reliable reports of transmission and sequencing of additional loci is underway to confirm transmission in each case.

The presence of drug-resistant mutations in drug-naive individuals may prevent successful treatment of certain individuals. This knowledge may also inform policymakers in planning future therapy strategies as this indicates that additional therapy combinations may be required. Although antiretroviral drugs were made more widely available in Karonga District in June 2005, prior to this NVP was available for the PMTCT in antenatal clinics and full ART would have been available to a number of individuals elsewhere in Malawi in 2002. Thus, it is also possible that some individuals assumed to be drug naïve did not disclose (or in the case of PMTCT were not aware of) ART exposure, and did not appear in the ART study cohorts. Furthermore, the presence of DRMs in drug-naive persons may represent natural polymorphisms and may not be suggestive of transmission of DRMs. The impact of such natural polymorphisms on the development of drug resistance in those people at commencement of ART is unknown and such individuals need to be monitored. Indeed the impact of such polymorphisms on the HIV-1 subtype C-infected population and the speed of more widespread drug resistance are also unknown. Therefore it may be important to further examine some of the individuals and mutations found during this study in view of a long-term treatment strategy in the district.

Sequence Data

All sequences have been deposited into GenBank, accession numbers HQ159410–HQ159841, and alignments employed in this work are available from the authors on request.

Acknowledgments

This material is based upon works supported by Science Foundation Ireland under Grant No. 07/RFP/EEEOBF424 and 08/UR/B1350. The Karonga Prevention Study is funded primarily by the Wellcome Trust, with contributions from LEPRA. Permission for the study was received from the National Health Sciences Research Committee, Malawi, and the Ethics Committee of the London School of Hygiene and Tropical Medicine, UK.

Author Disclosure Statement

No competing financial interests exist.

References

1. Ministry of Health: Antiretroviral Treatment Programme in MALAWI. Annual report.
2. Hemelaar J, Gouws E, Ghys PD, and Osmanov S: Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS 2006;20:W13–23.
3. Petch LA, Hoffman IF, Jere CS, *et al*.: Genotypic analysis of the protease and reverse transcriptase of HIV type 1 subtype C isolates from antiretroviral drug-naive adults in Malawi. AIDS Res Hum Retroviruses 2005;21:799–805.
4. Kassu A, Fujino M, Matsuda M, Nishizawa M, Ota F, and Sugiura W: Molecular epidemiology of HIV type 1 in treatment-naive patients in north Ethiopia. AIDS Res Hum Retroviruses 2007;23:564–568.
5. Bessong PO, Mphahlele J, Choge IA, *et al*.: Resistance mutational analysis of HIV type 1 subtype C among rural South African drug-naive patients prior to large-scale availability of antiretrovirals. AIDS Res Hum Retroviruses 2006;22: 1306–1312.
6. Handema R, Terunuma H, Kasolo F, *et al*.: Prevalence of drug-resistance-associated mutations in antiretroviral drug-naive Zambians infected with subtype C HIV-1. AIDS Res Hum Retroviruses 2003;19:151–160.
7. McCormack G, Glynn JR, Crampin AC, *et al*.: Early evolution of the human immunodeficiency virus type 1 subtype C epidemic in rural Malawi. J Virol 2002;76:12890–12899.
8. Stamatakis A: RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006;22:2688–2690.
9. Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J, and Shafer RW: Human immunodeficiency virus reverse

transcriptase and protease sequence database. Nucleic Acids Res 2003;31:298–303.
10. Rhee SY, Kantor R, Katzenstein DA, *et al*.: HIV-1 pol mutation frequency by subtype and treatment experience: Extension of the HIVseq program to seven non-B subtypes. AIDS 2006;20:643–651.
11. Hosseinipour MC, van Oosterhout JJ, Weigel R, *et al*.: The public health approach to identify antiretroviral therapy failure: High-level nucleoside reverse transcriptase inhibitor resistance among Malawians failing first-line antiretroviral therapy. AIDS 2009;23:1127–1134.
12. Kieffer TL, Kwon P, Nettles RE, Han Y, Ray SC, and Siliciano RF: G → A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting CD4+ T cells in vivo. J Virol 2005;79:1975–1980.
13. Goff SP: Death by deamination: A novel host restriction system for HIV-1. Cell 2003;114:281–283.
14. Vidal N, Niyongabo T, Nduwimana J, *et al*.: HIV type 1 diversity and antiretroviral drug resistance mutations in Burundi. AIDS Res Hum Retroviruses 2007;23:175–180.
15. Somi GR, Kibuka T, Diallo K, *et al*.: Surveillance of transmitted HIV drug resistance among women attending antenatal clinics in Dar es Salaam, Tanzania. Antivir Ther 2008;13(Suppl 2):77–82.

Address correspondence to:
Grace McCormack
Molecular Evolution & Systematics Laboratory
Zoology Department
Martin Ryan Institute
School of Natural Sciences
National University of Ireland
Galway, Ireland

E-mail: Grace.mccormack@nuigalway.ie

AIDS RESEARCH AND THERAPY

SHORT REPORT Open Access

Reverse transcriptase drug resistance mutations in HIV-1 subtype C infected patients on ART in Karonga District, Malawi

Vijay B Bansode[1], Simon AA Travers[2], Amelia C Crampin[3,4], Bagrey Ngwira[3,4], Neil French[3,4], Judith R Glynn[4] and Grace P McCormack[1*]

Abstract

Background: Drug resistance testing before initiation of, or during, antiretroviral therapy (ART) is not routinely performed in resource-limited settings. High levels of viral resistance circulating within the population will have impact on treatment programs by increasing the chances of transmission of resistant strains and treatment failure. Here, we investigate Drug Resistance Mutations (DRMs) from blood samples obtained at regular intervals from patients on ART (Baseline-22 months) in Karonga District, Malawi. One hundred and forty nine reverse transcriptase (RT) consensus sequences were obtained via nested PCR and automated sequencing from blood samples collected at three-month intervals from 75 HIV-1 subtype C infected individuals in the ART programme.

Results: Fifteen individuals showed DRMs, and in ten individuals DRMs were seen from baseline samples (reported to be ART naïve). Three individuals in whom no DRMs were observed at baseline showed the emergence of DRMs during ART exposure. Four individuals who did show DRMs at baseline showed additional DRMs at subsequent time points, while two individuals showed evidence of DRMs at baseline and either no DRMs, or different DRMs, at later timepoints. Three individuals had immune failure but none appeared to be failing clinically.

Conclusion: Despite the presence of DRMs to drugs included in the current regimen in some individuals, and immune failure in three, no signs of clinical failure were seen during this study. This cohort will continue to be monitored as part of the Karonga Prevention Study so that the long-term impact of these mutations can be assessed. Documenting proviral population is also important in monitoring the emergence of drug resistance as selective pressure provided by ART compromises the current plasma population, archived viruses can re-emerge

Keywords: HIV-1, drug resistance, subtype C, ART, Malawi, Reverse transcriptase

Introduction

It has been estimated that in sub-Saharan Africa, approximately 3.9 million people have started antiretroviral treatment (ART) since its introduction (UNAIDS, 2010). Given the large population on treatment, viral diversity coupled with low adherence could lead to the emergence and large-scale transmission of drug resistant strains. Rates of drug resistance among patients who received ART in sub-Saharan Africa range from 3.7%-49% after 24-163 weeks of HAART [1]. Various factors contribute to this large range in resistance among African cohorts such as variation in available healthcare systems and practices, adherence, and access to monitoring [2]. Development of DRMs to Trioimmune®, the drug combination used as first line therapy in Karonga District, Malawi, has been reported in Zambia [3], South Africa [4], Cameroon [5], Kenya [6] and Uganda [7]. Previous studies on drug resistance in Malawi showed various DRMs to both NRTIs and NNRTIs in both drug naïve individuals [8] and those failing therapy [9]. However, very little data is yet available on the emergence of drug resistance to ongoing treatment and the transmission of drug resistant variants in subtype C infected countries [3,5-7,10].

* Correspondence: grace.mccormack@nuigalway.ie
[1]Molecular Evolution and Systematics Laboratory, Zoology, Ryan Institute, School of Natural Sciences, National University of Ireland, Galway, Ireland
Full list of author information is available at the end of the article

The Malawi antiretroviral treatment (ART) program started in 2004, and between then and the end of June 2010 over 225, 000 patients had initiated first-line antiretroviral therapy (ART) through 396 ART clinics [11]. As part of the Karonga Prevention Study (KPS), investigating how the availability and use of ART may change the HIV epidemic and its socio-demographic impact in the rural Karonga District, (northern Malawi), an ART research cohort was established from those attending the ART clinic at Chilumba Rural Hospital. HIV-1 subtype C is the predominant subtype in this District [12]. The objective of this overall study was to investigate the success of the current ART delivery programme in a rural population, and, as a component of this, to investigate the evolution of drug resistance using a traditional consensus sequence genotyping approach.

Materials and methods

Study Participants and Treatment schedules

At the Ministry of Health ART clinic at Chilumba Rural Hospital all those attending for screening for ART suitability and who are resident in a geographically defined area adjacent to the clinic, are invited to take part in an observational cohort study. Every three months participants are clinically assessed by KPS research staff. Blood samples are collected at the their first visit (baseline) and at every follow-up visit. A CD4 count is performed at baseline, 6, 12 and 24 months or at the time of a clinical failure, defined by a new WHO stage 3 or 4 event after six months of therapy (WHO, 2006). First line therapy is a generic fixed-dose combination treatment (Triomune®), which consists of: stavudine (d4T), lamivudine (3TC), and nevirapine (NVP). All individuals were on first line therapy only.

DNA Extraction, PCR and Sequencing

Whole blood samples were collected in 4.5 ml vacutainer tubes. Samples were centrifuged and plasma and cell pellet were stored separately at -70°C. DNA was extracted from whole blood cell pellet samples using the QIAamp DNA Blood Mini Kit (QIAGEN ltd). Extracted DNA was subjected to nested PCR amplification of the HIV-1 reverse transcriptase as described in Bansode et al [13]. All PCR amplicons were gel purified and automatically sequenced.

Sequence Analyses

Sequence chromatographs were edited in SeqMan (DNASTAR, Inc) and all sites that showed ambiguities (two or more peaks of equal, or almost equal, height) were noted. Multiple alignments were assembled of all subtype C sequences generated with the 57 reverse transcriptase sequences generated from [13] using MacClade 4.0 (Sinauer Assoc). Sequences were submitted for analysis of DRMs to the Stanford Database [14]. To check for transmission of DRMs, phylogenetic trees were reconstructed using the LANL subtype C ancestral sequence as outgroup under the GTR + gamma model of DNA substitution implemented RAxML7.0.3 [15] with all parameters optimised by RAxML. Confidence levels in the groupings in the phylogeny were assessed using 1000 bootstrap replicates as part of the RAxML phylogeny reconstruction.

Permission for the study was received from the National Health Sciences Research Committee, Malawi, and the Ethics Committee of the London School of Hygiene and Tropical Medicine, UK

Results

One hundred and forty nine subtype C sequences were generated from 75 individuals, 65 of which were from blood samples collected at baseline (and reported to be ART naïve). DRMs were found in sequences from 15 individuals (20%) overall, and for 10 individuals (15.4%) the mutations were found in sequences from baseline samples (drug naïve). Details of observed drug resistance mutations are summarized in Table 1. Seven individuals showed DRMs (or ambiguities that suggest the presence of DRMs) to NRTIs used in Karonga with 6/7 showing the mutation V118I. While ten individuals showed the presence of DRMs to NNRTIs only five showed DRMs against therapies used in Karonga, the most common being Y181C and G190AE.

Some individuals showed a discrepancy in the presence and type of DRMs over time. Three patients (Pt2, Pt12 and Pt66) did not show any DRMs at baseline but showed DRMs at subsequent time points (Table 1). Patient 2 also showed a significant drug resistance-related ambiguity (K103KN) in the consensus sequence at 6 months while a different DRM was seen at 9 months (Y181C). No DRM was seen in the sequence from the 12-month sample. Patient 12 showed a similar pattern, where an NNRTI associated mutation (Y181NY) was present in the sequence collected at 3 months, while the sequences at baseline and 12 months did not show any DRMs (Table 1). Both individuals (patient 2 and patient 12) showed immune failure (their CD4 count did not rise over 200 cells/mm^3 after 12 months on ART).

Three patients (Pt32, Pt61 and Pt76) showed DRMs at baseline but different DRMs at later time-points (Table 1) with patient 32 showing a high variation of DRMs across timepoints. The baseline sequence from patient 61 showed V118I and K219R, the latter of which was not found in the sequence at 9 months. In Patient 76, the baseline sequence showed the ambiguity Y181CY, with two additional NNRTI mutations (V90IV and H221HY), the 6 months

Table 1 Mutations associated with antiretroviral drug resistance found in sequences from HIV-1 subtype C infected individuals from Karonga District Malawi

Patient	Comments	Sex	Time point (month)	NRTI	NNRTI
Patient 2 *	Immune Failure	F	0	No DRMs	No DRMs
			6	No DRMs	**K103KN**
			9	No DRMs	**Y181C**
			12	No DRMs	No DRMs
Patient 5		F	0	No DRMs	E138A
			6	No DRMs	E138A
Patient 12 *	Immune Failure	F	0	No DRMs	No DRMs
			3	No DRMs	**Y181NY**
			12	No DRMs	E138A
Patient 14		M	12	No DRMs	E138A
Patient 20	Immune Failure	F	0	No DRMs	E138A
			3	No DRMs	E138A
			6	No DRMs	E138A
			9	No DRMs	E138A
Patient 32 *		M	0	No DRMs	V90I
			6	**M41MR, T215ST**	No DRMs
			12	No DRMs	No DRMs
			15	No DRMs	V108AV
			22	No DRMs	No DRMs
Patient 42 *		M	0	No DRMs	No DRMs
			0	**V118IV**	No DRMs
			3	**V118I**	No DRMs
Patient 45		F	0	No DRMs	V106I, E138A, **G190A**
			6	No DRMs	V106I, E138A, **G190A**
Patient 61 *		F	0	**V118I**, K219R	No DRMs
			9	**V118I**	No DRMs
Patient 66		M	0	No DRMs	No DRMs
			9	**V118IV**	No DRMs
Patient 76 *		M	0	No DRMs	V90IV, **Y181CY**, H221HY
			3	No DRMs	No DRMs
			6	No DRMs	**Y181C**
			9	No DRMs	No DRMs
Patient 77		F	0	**V118I**	No DRMs
			12	**V118I**	No DRMs
Patient 91			0	**V118I**	No DRMs
Patient 93		M	6	**V118I**	E138A
			12	**V118I**	E138A
Patient 95			0	No DRMs	E138A

Mutations in bold are against current ART drugs in use in Karonga District.

* Patients showing discrepancies in DRMs between different timepoints

sequence showed the full DRM at position 181 while sequences retrieved from 3 month and 9 month samples showed no DRMs (Table 1).

There was no evidence of transmission of drug resistant HIV between the individuals examined here. Sequences retrieved from each individual grouped monophyletically in all cases. Few individuals showed their sequences clustering with other patients with high bootstrap support but DRMs were not present in both individuals, e.g. sequences from patient 47 and 77 formed a cluster together and are from the same geographical area but while patient 77 showed DRMs, patient 47 did not (data not shown).

Discussion

Through genotyping RT from HIV-1 subtype C infected individuals on ART using a consensus sequencing approach, we have shown the presence of mutations associated with drug resistance to the therapy used in Karonga District. Drug resistance to Trioimmune® occurred at an overall rate of 20% of individuals (both drug naïve and drug exposed, which is comparable to rates found in other African countries [4-7,16-18] but, as expected, greater than that described in our previous study (7.5%) [8] which did not include individuals currently on therapy.

Patients 2 and 12, both females, had immune failure prior to ART initiation and continued to exhibit immune failure while on ART (i.e. their CD4 counts did not rise above 200 cells/mm^3 after 12 months on therapy). While neither showed DRMs from baseline samples they subsequently showed the DRMs Y181C and Y181NY respectively, which is responsible for high-level resistance to NVP, the NNRTI used in 1^{st} line therapy in Karonga. Patient 2 also showed a DRM (K103KN), after 6^{th} months of ART, which also causes high-level resistance to NVP. For these, and a third individual who also exhibited immune failure, it will be important to monitor the individuals and DRMs at subsequent timepoints in case of continued immune failure and development of clinical failure.

Three drug-naïve individuals (Pt 61, Pt 77, Pt 91) showed V118I while another (Pt 42) showed an ambiguity at this position (V118IV). According to the Stanford HIV drug resistance database, V118I is responsible for low-level resistance to 3TC and possibly to other NRTIs when present with other mutations. The mutation has been reported to occur in ~2% of untreated persons infected with subtype C and with increased frequency in persons receiving multiple NRTIs [19] and so it may not be unexpected to find it in this cohort. It was the only DRM found in all three previous studies of drug resistance in Malawi [9,13,20] and was also reported in subtype C infected drug naïve patients from Zambia [21], Zimbabwe [22] and South Africa [19]. It has been suggested that along with drug resistance, the V118I mutation alone is a marker of advanced HIV infection and disease progression [23]. As no associated mutations were found in the three individuals, and they all had a satisfactory response to treatment, this mutation is probably not significant but may become important if a second mutation were to arise.

Mutation G190A (shown in a female patient 45) according to the Stanford drug resistance database, causes high-level resistance to NVP and intermediate resistance to EFV. The mutation was present at baseline in this individual and could indicate acquisition of drug resistant HIV. However, although all individuals participating in the ART cohort study were reported to be ART naïve, we cannot exclude the possibility that some individuals had received some form of ART previously, (e.g. received prevention of mother to child transmission treatment) and did not disclose this fact. The DRM does not appear to have had any major effect on treatment to date, as this individual also has had a satisfactory response.

Drug resistance mutations were found to emerge in some individuals during ART. Patient 32 showed a number of NNRTI mutations and a number of ambiguities at sites important in susceptibility to NRTIs (e.g. the mutation T215S is one of many transitions between wild type and the mutations Y and F [24]). Most of the ambiguities do not reduce NRTI susceptibility but their presence may suggest that the DRM may also be present [25]. This patient had made additional visits to the clinic outside of the routine ART cohort study because of diabetic complications. Additional sequences produced from samples taken at those additional visits showed further mutations associated with drug resistance to NVP and AZT (M41L, M184I, G190E- data not shown), however he has had a satisfactory response to treatment to date.

This study was based on a consensus sequencing approach from provirus due to the difficulty of amplifying HIV from RNA from individuals on ART. While provirus may not provide as clear a picture of the genotype of the circulating virus as would be retrieved from RNA in individuals who have been infected for long periods of time, it has been shown in patients with virological failure that archived resistance mutations previously detected in the proviral DNA were observed in the sequences obtained from the plasma viruses at the time of virological failure [26]. When the selective pressure provided by ART compromises the current plasma population, archived viruses can re-emerge [27]. Therefore documenting the proviral population is also important in monitoring the emergence of drug resistance. Despite the presence of DRMs to current therapy in some individuals, and immune failure in three, no signs of clinical failure were seen during this study. This cohort will continue to be monitored as part of the Karonga Prevention Study so that the long-term impact of these mutations can be assessed.

Acknowledgements

This material is based upon works supported by Science Foundation Ireland under Grant No. 07/RFP/EEEOBF424. The Karonga Prevention Study is funded primarily by the Wellcome Trust, with contributions from LEPRA.

Author details

[1]Molecular Evolution and Systematics Laboratory, Zoology, Ryan Institute, School of Natural Sciences, National University of Ireland, Galway, Ireland. [2]South African National Bioinformatics Institute, University of the Western Cape, Bellville, South Africa. [3]Karonga Prevention Study, Chilumba, Malawi. [4]Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK.

Authors' contributions

VB carried out the molecular biology work and subsequent analysis; AC, BN, NF and JG participated in design of the study; ST and GM conceived and supervised the study. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 22 June 2011 Accepted: 13 October 2011
Published: 13 October 2011

References

1. Hamers RL, Derdelinckx I, van Vugt M, Stevens W, Rinke de Wit TF, Schuurman R: **The status of HIV-1 resistance to antiretroviral drugs in sub-Saharan Africa.** *Antivir Ther* 2008, **13**:625-639.
2. Shafer RW, Rhee SY, Bennett DE: **Consensus drug resistance mutations for epidemiological surveillance: basic principles and potential controversies.** *Antivir Ther* 2008, **13(Suppl 2)**:59-68.
3. Gupta RK, Ford D, Mulenga V, Walker AS, Kabamba D, Kalumbi M, *et al*: **Drug resistance in human immunodeficiency virus type-1 infected Zambian children using adult fixed dose combination stavudine, lamivudine, and nevirapine.** *Pediatr Infect Dis J* 2010, **29**:e57-62.
4. Hoffmann CJ, Charalambous S, Sim J, Ledwaba J, Schwikkard G, Chaisson RE, *et al*: **Viremia, resuppression, and time to resistance in human immunodeficiency virus (HIV) subtype C during first-line antiretroviral therapy in South Africa.** *Clin Infect Dis* 2009, **49**:1928-1935.
5. Burda ST, Viswanath R, Zhao J, Kinge T, Anyangwe C, Tinyami ET, *et al*: **HIV-1 reverse transcriptase drug-resistance mutations in chronically infected individuals receiving or naive to HAART in Cameroon.** *J Med Virol* 2010, **82**:187-196.
6. Steegen K, Luchters S, Dauwe K, Reynaerts J, Mandaliya K, Jaoko W, *et al*: **Effectiveness of antiretroviral therapy and development of drug resistance in HIV-1 infected patients in Mombasa, Kenya.** *AIDS Res Ther* 2009, **6**:12.
7. Weidle PJ, Downing R, Sozi C, Mwebaze R, Rukundo G, Malamba S, *et al*: **Development of phenotypic and genotypic resistance to antiretroviral therapy in the UNAIDS HIV Drug Access Initiative–Uganda.** *Aids* 2003, **17(Suppl 3)**:S39-48.
8. Bansode V, Drebert ZJ, Travers SA, Banda E, Molesworth A, Crampin A, *et al*: **Drug Resistance Mutations in Drug-Naive HIV Type 1 Subtype C-Infected Individuals from Rural Malawi.** *AIDS Res Hum Retroviruses* 2011, **27**:439-444.
9. Hosseinipour MC, van Oosterhout JJ, Weigel R, Phiri S, Kamwendo D, Parkin N, *et al*: **The public health approach to identify antiretroviral therapy failure: high-level nucleoside reverse transcriptase inhibitor resistance among Malawians failing first-line antiretroviral therapy.** *Aids* 2009, **23**:1127-1134.
10. Towler WI, Barlow-Mosha L, Church JD, Bagenda D, Ajuna P, Mubiru M, *et al*: **Analysis of drug resistance in children receiving antiretroviral therapy for treatment of HIV-1 infection in Uganda.** *AIDS Res Hum Retroviruses* 2010, **26**:563-568.
11. Ministry of Health GoM: **Quarterly HIV Programme Report of HIV testing and counseling, Prevention of Mother to Child Transmission, Antiretroviral Therapy, Treatment of Sexually Transmitted Infections with results upto June 2010.** 2010.
12. McCormack GP, Glynn JR, Crampin AC, Sibande F, Mulawa D, Bliss L, *et al*: **Early evolution of the human immunodeficiency virus type 1 subtype C epidemic in rural Malawi.** *J Virol* 2002, **76**:12890-12899.
13. Bansode V, Drebert ZJ, Travers SA, Banda E, Molesworth A, Crampin A, *et al*: **Drug Resistance Mutations in Drug-Naive HIV Type 1 Subtype C-Infected Individuals from Rural Malawi.** *AIDS Res Hum Retroviruses* 2010.
14. Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW: **Human immunodeficiency virus reverse transcriptase and protease sequence database.** *Nucleic Acids Res* 2003, **31**:298-303.
15. Stamatakis A: **RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.** *Bioinformatics* 2006, **22**:2688-2690.
16. Seyler C, Adje-Toure C, Messou E, Dakoury-Dogbo N, Rouet F, Gabillard D, *et al*: **Impact of genotypic drug resistance mutations on clinical and immunological outcomes in HIV-infected adults on HAART in West Africa.** *Aids* 2007, **21**:1157-1164.
17. Laurent C, Ngom Gueye NF, Ndour CT, Gueye PM, Diouf M, Diakhate N, *et al*: **Long-term benefits of highly active antiretroviral therapy in Senegalese HIV-1-infected adults.** *J Acquir Immune Defic Syndr* 2005, **38**:14-17.
18. Johannessen A, Naman E, Kivuyo SL, Kasubi MJ, Holberg-Petersen M, Matee MI, *et al*: **Virological efficacy and emergence of drug resistance in adults on antiretroviral treatment in rural Tanzania.** *BMC Infect Dis* 2009, **9**:108.
19. Pillay C, Bredell H, McIntyre J, Gray G, Morris L: **HIV-1 subtype C reverse transcriptase sequences from drug-naive pregnant women in South Africa.** *AIDS Res Hum Retroviruses* 2002, **18**:605-610.
20. Petch LA, Hoffman IF, Jere CS, Kazembe PN, Martinson FE, Chilongozi D, *et al*: **Genotypic analysis of the protease and reverse transcriptase of HIV type 1 subtype C isolates from antiretroviral drug-naive adults in Malawi.** *AIDS Res Hum Retroviruses* 2005, **21**:799-805.
21. Handema R, Terunuma H, Kasolo F, Kasai H, Sichone M, Yamashita A, *et al*: **Prevalence of drug-resistance-associated mutations in antiretroviral drug-naive Zambians infected with subtype C HIV-1.** *AIDS Res Hum Retroviruses* 2003, **19**:151-160.
22. Kantor R, Zijenah LS, Shafer RW, Mutetwa S, Johnston E, Lloyd R, *et al*: **HIV-1 subtype C reverse transcriptase and protease genotypes in Zimbabwean patients failing antiretroviral therapy.** *AIDS Res Hum Retroviruses* 2002, **18**:1407-1413.
23. Zaccarelli M, Tozzi V, Lorenzini P, Forbici F, Narciso P, Ceccherini-Silberstein F, *et al*: **The V118I mutation as a marker of advanced HIV infection and disease progression.** *Antivir Ther* 2007, **12**:163-168.
24. Bennett DE, Camacho RJ, Otelea D, Kuritzkes DR, Fleury H, Kiuchi M, *et al*: **Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update.** *PLoS One* 2009, **4**:e4724.
25. Shafer RW, Rhee SY, Pillay D, Miller V, Sandstrom P, Schapiro JM, *et al*: **HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance.** *Aids* 2007, **21**:215-223.
26. Boucher S, Recordon-Pinson P, Neau D, Ragnaud JM, Titier K, Faure M, *et al*: **Clonal analysis of HIV-1 variants in proviral DNA during treatment interruption in patients with multiple therapy failures.** *J Clin Virol* 2005, **34**:288-294.
27. Joos B, Fischer M, Kuster H, Pillai SK, Wong JK, Boni J, *et al*: **HIV rebounds from latently infected cells, rather than from continuing low-level replication.** *Proc Natl Acad Sci USA* 2008, **105**:16725-16730.

doi:10.1186/1742-6405-8-38
Cite this article as: Bansode *et al.*: Reverse transcriptase drug resistance mutations in HIV-1 subtype C infected patients on ART in Karonga District, Malawi. *AIDS Research and Therapy* 2011 **8**:38.