DRUG RESISTANCE GENOTYPING AND PHYLOGENETIC ANALYSIS OF HIV IN CHRONICALLY INFECTED ANTIRETROVIRAL NAÏVE PATIENTS

DISSERTATION SUBMITTED IN FULFILMENT OF THE AWARD OF MASTER OF SCIENCE DEGREE IN MICROBIOLOGY

SUBMITTED

BY

BALOYI TLANGELANI (11595307)

TO THE DEPARTMENT OF MICROBIOLOGY,

SCHOOL OF MATHEMATICAL AND NATURAL SCIENCES

UNIVERSITY OF VENDA,

SOUTH AFRICA.

SUPERVISOR: PROF PASCAL OBONG BESSONG

UNIVERSITY OF VENDA

CO-SUPERVISOR: PROF AFSATOU NDAMA TRAORE

UNIVERSITY OF VENDA

MAY 2018

Declaration

I, Baloyi Tlangelani, hereby declare that this dissertation for the award of Master of Science degree in Microbiology at the University of Venda is my original work. It has not been submitted before for any degree examination at this or any other University. It is my own work at execution and all the reference materials contained are therein have been duly acknowledged.

Signature…… Date……

Acknowledgements

First and foremost, I would like to thank the almighty God for giving me strength and wisdom to carry out this work until completion.

I would like to express my gratitude and appreciation to my supervisor Prof P.O. Bessong for giving me the platform to do scientific research and introduced me to the field of molecular biology. This dissertation could not be written to its fullest without his scientific guidance and constructive criticism.

To my co-supervisor Prof A.N. Traore thanks for your personal, professional and scientific guidance. You were always there for me when I needed help.

A special thanks to Dr L.G. Mavhandu for technical and logistic support with sample collection and lab work.

I extend my appreciations to my colleagues of the HIV/AIDS & Global Health Research Programme: Ms D. Matume, Mr B. Ogola, Ms R. Rikhotso, Ms E. Etta, Ms L. Tambe and Ms Munyai for their support.

To my friend Thelma Banda thank you for your encouragement and comforting at times when research was hitting hard on me.

A token of appreciation to my study participants because without them I would not have samples to analyse.

Nobody has been more important to me in the pursuit of this project than members of my family. I would like to thank my sisters (Vutomi and Langutani Baloyi) whose love and guidance was with me throughout my work, they are the ultimate role models.

I extend by gratitude to the National Research Foundation and the University of Venda for financial assistance.

Dedication

I dedicate this dissertation to my late parents Khensani and Thomas Baloyi.

Abstract

Background: Antiretroviral treatment (ART) has grown to be one of the most effective tool in the fight to control HIV/AIDS morbidity and mortality worldwide. However, due to the emergence of drug resistant HIV, ART efficacy can be jeopardized. Drug resistant HIV strain has a potential of becoming a major public threat, as its limit treatment options on people living with HIV. With several findings worldwide reporting drug resistant HIV to be currently being transmitted to ART-naïve persons, measures have been taken to genotype drug resistant HIV prior to treatment initiation. However, in resource limited countries such measures are not executed especially in public sectors due to the costs associated with the required assays for genotyping.

Objective: The objectives of the study was to establish a deep sequencing protocol (Next Generation Sequencing-NGS) using an Illumina MiniSeq Platform and subsequently apply it to genotype HIV in chronically infected drug naïve persons for resistance mutations and viral genotypes

Methods: HIV positive Individuals without any exposure to ART (Treatment-naive) were recruited. Partial pol fragment (complete protease and ~1104bp reverse transcriptase) were amplified and purified. Libraries were prepared using Nextera XT library preparation kit, fragmented, tagmented, pooled and denatured then sequenced with Illumina MiniSeq instrument. Consensus sequences were derived, aligned and phylogenetically analysed. The Stanford HIV Drug Resistance Algorithm was used to infer the presence of drug resistant mutants, at the viral minority and majority population levels.

Results and discussion: An NGS protocol to generate nucleotide sequences for drug resistance inference was established. No major drug resistance mutations were detected against protease, reverse transcriptase inhibitors in the study subjects investigated. Nevertheless, V179D change was observed in one patient (8.3%). V179D has been shown to impact a low-level resistance to NNRTI. On the other hand, several secondary and unusual mutations at known drug sites were detected even at minority threshold level of <20%.

Conclusion: No major drug resistance mutations was detected in the drug naïve study population. This finding suggests that there is no risk of treatment failure to the investigated subjects, however it is important to assess the potential phenotypic

significance of the identified secondary resistance mutations in the context of HIV-1 subtype C. The established NGS protocol should be applied in subsequent HIV drug resistance studies.

Keywords: HIV 1 subtype C; Next generation sequencing; Drug resistance mutations; Treatment naïve patients; South Africa.

Table of Contents

List of abbreviations

AIDS: Acquired Immune Deficiency Virus
ART: Antiretroviral Treatment
ARV: Antiretroviral drug
AZT: Zidovudine
bp: Base pair
CA: Capsid
cDNA: Complementary Deoxyribonucleic acid
CD4: Cluster of differentiation 4 (type of white blood cell that fights infection)
CRFs: Circulating Recombinant Forms
DEPC: Diethyl pyrocarbonate
DHB: District health barometer
DNA: Deoxyribonucleic acid
DNTPs: Deoxynucleotide triphosphate
dsDNA: Double stranded DNA
EAV: Equine anaemia virus
EIA: Enzyme immune assay
EDTA: Ethylene diamine tetra-acetic acid
ELISA: Enzyme-Linked Immnunosorbent Assay
FI's: Fusion inhibitor's
FIV: Feline immune deficiency virus
FDA: Food & drug administration
gp: Glycoprotein
HAART: Highly active antiretroviral therapy
HIV: Human immunodeficiency virus
HIV-1: Human immunodeficiency virus type-1
HIV-2: Human immunodeficiency virus type-2
HIVDR: HIV drug resistance
INSTI: Integrase inhibitor
LTRs: Long Terminal Repeats
MTCT: Mother-To-Child-Transmission

NGS: Next Generation Sequencing
nM: nanomolar
NNRTI: Non-Nucleoside Reverse Transcriptase Inhibitors
NRTI: Nucleoside Reverse Transcriptase Inhibitors
PBMC: Peripheral blood mononuclear cell
PBS: Phosphate buffer saline
PCR: Polymerase chain reaction
PI: Protease Inhibitors
pM: Picomolar
PMTCT: Prevention of Mother to Child Transmission
Pol: Polymerase protein
PR: Protease enzyme
RITA: Recent infection testing algorithm
RNA: Ribonucleic acid
RT: Reverse Transcriptase
SA: South Africa
SDRMs: Surveillance Drug Resistance Mutations
SIV: Simian immunodeficiency virus
SIVcpz: Simian immunodeficiency virus found in chimpanzee
SIVsm: Simian immunodeficiency virus found in Sooty mangabey
SIVgor: Simian immunodeficiency virus found in gorilla
STLV3: Simian T-cell lymphotropic virus3
TDR: Transmitted drug resistance
URFs: Unique Recombinant Forms
UNAIDS: United nation on acquired human immunodeficiency virus
UK: United Kingdom
WHO: World health organization
%: Percentage
ºC: Degree Celsius
µl: Microliter
µM: Micrometre

List of figures

Chapter 1

Chapter 2

Chapter 3

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 1: Introduction and literature review

1.1 Introduction

Human immunodeficiency virus (HIV) is the causative agent of acquired immunodeficiency syndrome (AIDS), one of the most devastating medical condition of all times. In 2015 about 36.9 million people were believed to be living with HIV (UNAIDS, 2017). In the same year an estimated 1.1 million AIDS-related deaths were reported (UNAIDS, 2016). The vast majority of infected people live in low and middle-income countries where there are other socio-economic problems such as food insecurity and malnutrition (Dallar and Karim, 2015; UNAIDS, 2017).

Sub-Saharan Africa remains one of the region most affected, with nearly 25.6 million people living with HIV in 2015 (WHO, 2016). The region accounts for about 70% of the people living with HIV worldwide (WHO, 2015). In Sub-Saharan Africa, South Africa is considered as the epicenter of HIV due to its high burden of the disease with an estimated 6.3 million people living with HIV (UNAIDS, 2015).

Thus far there are no effective vaccine or cure for HIV, although the introduction of highly antiretroviral therapy (HAART) significantly improves prognosis (WHO, 2015). HAART averted 240 000 AIDS related death in South Africa in 2013 (Konstant et al., 2015). This tremendous achievement was observed after the roll out program of antiretroviral therapy (ART) in public sectors. However, ART roll-out has been associated with emergence of HIV drug resistance (HIVDR) in therapy-naïve and treated individuals due to low treatment adherence, tolerability and long-term toxicity that act as limiting factors to treatment efficacy. Drug-resistant strains are archived in viral reservoirs and may persist as minority variants when outgrown by wild-type strains, in the absence of sufficient drug pressure in treatment-naïve patients.

HIV drug resistance can jeopardize the efficacy of ART regimens to reduce HIV-associated morbidity and mortality, so it is important to understand and monitor their

spread in the population. Previous studies done on drug resistance utilizes Sanger sequencing for drug resistance mutations genotyping. However, minority drug resistant variants are not readily detectable with Sanger sequencing when constituting less than 20% of the viral population. Deep sequencing methods such as next generation sequencing is being used to detect resistant mutants existing as a minority population, at a threshold as low as 5%.

Understanding and monitoring the spread of drug resistant viruses is important in terms of implementation of future prevention and control strategies. Understanding of drug resistant variant requires knowledge of their genetic diversity, epidemiology and evolutionary history. Phylogenetic analysis serves to infer the evolutionary dynamics of virus genetic diversity (Holmes and Grenfell, 2009). The focus in phylogenetic approach is to show how phylogenetic analysis have influenced the current understanding of the emergence and evolution of drug resistance, epidemiology and dynamics of HIV.

1.2 Literature review

1.2.1 The origin, and discovery of HIV-1 and HIV-2

HIV 1 and HIV 2 belong to the lentivirus genus (retroviridae family). The two species are distinguished based on their genome organization, virulence factors, clinical manifestation, phylogenetic relationships and geographical distributions.
It is believed that HIV crossed into the human species by zoonotic transmission of simian immunodeficiency virus (SIV) from non-human primates (African green monkeys) into humans in the West African region. The non-human primates said to be carrier of HIV-1-like virus once called simian T-cell lymphotropic virus 3 (STLV3), now called SIV (Kanki et al., 1987). The serum of HIV -1 infected individuals which cross reacted with SLTV-3 proteins was used as an immunological evidence to the zoonotic transmission (Barin et al., 1985). It is believed that the transmission occurred because of hunting and butchering of non-human primates for wild meat as well as capturing, trading and keeping of those animals as pets (Hahn, 2000).

There are several theories which suggest that HIV evolved from SIVs in 1930s, although several studies placed the origin of HIV to be between 1884 and 1924 (Worobey et al., 2008; and Korber et al., 2000). HIV/AIDS was discovered in 1981 due to an increasing number of homosexual individuals suffering from unusual opportunistic infection and type of rare malignancies (Friedman-Kien et al., 1981). The cause of the disease was unknown by the time, until 1983 when a retrovirus which was later termed HIV-1 was identified as the causative agent (Gallo et al., 1984; Barre-sinoussi et al., 1983).

In 1986, three years after the discovery of HIV-1, another virus (HIV-2) was isolated from a hospitalized patient in West Africa (Clavel et al., 1986). The virus was morphologically similar to HIV-1, but antigenically different. The virus was distantly related to HIV-1, but closely related to a simian virus that is believed to cause immunodeficiency in captive macaques (Chakrabarti et al., 1987; Guyader et al., 1987). The virus was later termed HIV-2**.** Figure 1 present the relationship among lentiviruses.

1.2.2 Classification of HIV

Human Immunodeficiency Virus is of the *Retroviridae* family and Orthoretrovirinae subfamily. It belongs to the genus lentivirus. Lentiviruses are characterised by very long incubation period. All members of the lentivirus genus contain a lipid envelope derived from the host cell membrane when the virus buds out of the infected cell. Some members of lentivirus genus include Equine Anaemia virus (EAV), Simian Immunodeficiency Virus (SIV) and Feline Immunodeficiency Virus (FIV) (Coffin et al., 1997).

HIV is genetically classified into two types: HIV-1 and HIV-2. HIV-2 is subdivided into 8 different groups (A-H), with group A and B as the major groups. On the other hand, HIV-1 is further grouped into groups namely M for main, O for outlier, N for Non-M-Non-O and P for putative. Group M viruses are further divided into nine subtypes A to D, F, G, H, J and K, circulating recombinant forms (CRF) and unique recombinant forms (URF) (Ariën, 2005; Gupta et al., 2005; Vallari et al., 2011).

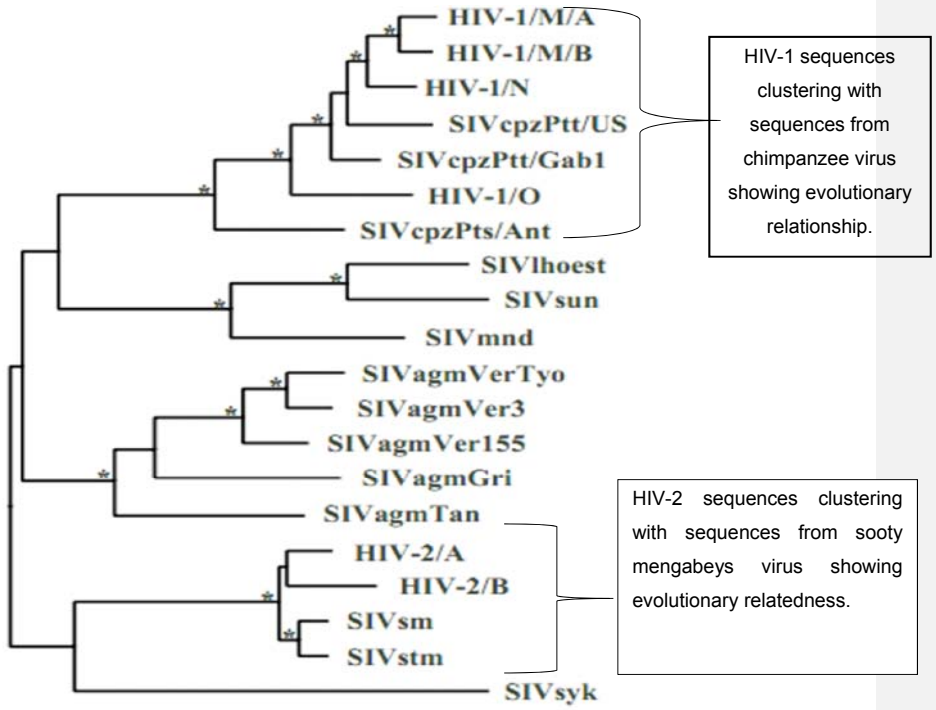


Figure 1: Evolutionary relationships among primate lentiviruses. This Figure shows phylogenetic relationship between HIV-1 and SIVcpz (chimpanzees virus) and HIV-2 and SIVsm (sooty mangabeys virus). Figure reveal HIV-1 and HIV-2 are not adjacent to each other.
(Adapted from Rajarapu, 2013; http://dx.doi.org/10.4172/2329-9002.1000126)

1.2.3 Epidemiology of HIV lineages

HIV-1 has an extensive genetic diversity which resulted from four various lineages which are M, N, O and P. HIV-1 group M was the first to be discovered and has been the most successful in establishing the human pandemic with a global prevalence of more than 98% (Sharp and Hahn, 2011; Ariën, 2005). HIV-1 group N and O are less prevalent, accounting approximately 1 % of HIV cases worldwide and reported in Cameroon and Gabon (Vallari et al., 2010). P is the rarest group, first isolated from a Cameroonian

woman in France. It has an estimated prevalence of 0.06% (Peeters et al., 1997; Vallari et al., 2011; Plantier et al., 2009).

Cases of HIV-2 are mostly reported from West African countries, with Senegal and Guinea Bissau having the highest infection rate. HIV-2 is divided into 8 different groups namely A to H. Group A is generally predominant in Sub Saharan Africa (de Silva et al., 2008), and group Bis reported commonly in Ivory Coast (Ishikawa et al., 2001). Due to their sporadic nature of infection, groups C to H are referred to have “dead- end” transmissions (Sharp and Hahn, 2011; Elena and Sanjuán, 2005 and Smith et al., 2009).

1.2.4 Basic structure of HIV

HIV is spherical in shape, with a diameter of approximately 120 nanometres (nM). The virus consists of two single stranded RNA molecules in a capsid protein (p24). The envelope of the virus consists of two layers of lipids in which various proteins are embedded. The glycoproteins inside the viral envelope (gp120 ang gp41) plays an important role in mediating the process of HIV viral infection. Gp120 is a surface glycoprotein that is located on the surface of the virus. Gp120 plays a significant role in the attachment processes of the virus to the host cell. Gp41 is a trans-membrane protein, important during virus-cell fusion process.

The HIV matrix protein (p17 protein) is located between the envelope and capsid. The viral capsid contains the viral capsid protein p24 which surrounds the two single strands of positive viral RNA and the enzymes required for HIV replication, such as reverse transcriptase, protease and ribonuclease. The HIV genome consists of 9 genes, of which three namely *gag*, *pol* and envelope contain the information needed to make structural proteins for new virus particles. Figure 2. presents the structure of HIV virion.

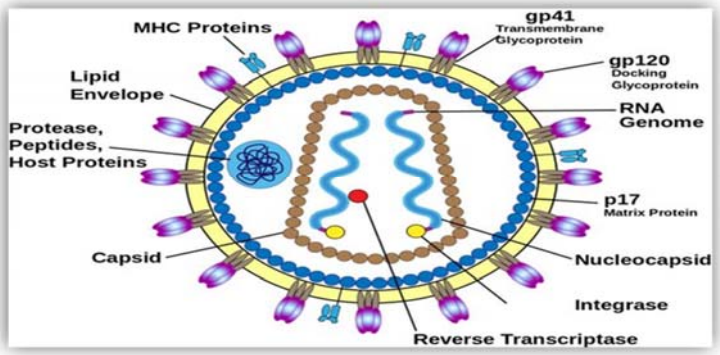


Figure 2: The structure of HIV virion
Adapted from: https://mappingignorance.org/fx/media/2013/01/Fig1.png

1.2.5 Genomic organization of HIV-1

HIV provirus genome (figure 1.2), also known as proviral DNA genome is composed of two identical RNA strands of about 9.7 kilobase (kb) in size. The genome has 9 coding genes coding for various proteins (table 1.2). Each end of the genome is flanked by long terminal repeat (LTR) sequences. The 5′ LTR region codes for transcription promotor of the viral genes. Starting from 5′ to 3′ the reading frame of the gag gene follows, encoding the proteins of the outer core membrane (MA, p17), the capsid protein (CA, p24), the nucleocapsid (NC, p7) and a smaller, nucleic acid-stabilising protein (Lu et al., 2011; GAC ,2016).

Polymerase gene codes for the enzymes protease (PR, p12), reverse transcriptase (RT, p51) and RNase H (p15) and integrase (IN, p32). Adjacent to the pol gene, there is envelope gene reading frame from which the two envelope glycoproteins gp120 and gp41 are derived from. HIV genome codes for several regulatory proteins which includes: Tat and Rev. Tat and Ref are necessary for the initiation of HIV replication, while the other regulatory proteins Nef, Vif, Vpr and Vpu have an impact on viral replication, virus budding and pathogenesis (Levy,2011; Sauter et al., 2012). HIV-2 codes for Vpx which is partially responsible for the reduced pathogenicity of HIV-2 (Vincenzi and Poli, 2013). The

genomic structure of the immunodeficiency viruses of chimpanzees (SIVcpz) and gorillas (SIVgor) is similar to that of that of HIV-1 (Kuiken et al., 2012). Figure 3. presents the genomic overview of HIV-1. Table1 present HIV-1 proteins and their specific functions.

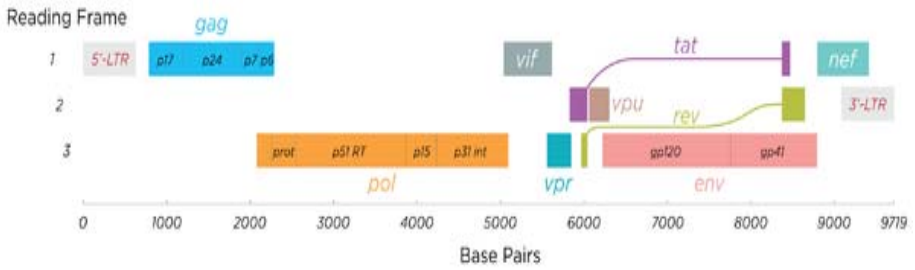


Figure 3: Genomic organisation of HIV (adapted from https://en.wikipedia.org/wiki/).

Table 1: Overview of HIV-1 proteins and their functions

Gene	Protein	Function
***Gag* Pr55Gag protein (precursor of the inner structural proteins)**		
p24	Capsid protein (CA)	Formation of conical capsid
p17	Matrix protein (MA)	Myristilated protein, forming the inner membrane layer
p7	Nucleoprotein (NC)	Formation of the nucleoprotein/RNA complex
P6		Involved in virus particle release
***Pol* Pr160GagPol (precursor of the viral enzymes)**		
p10	Protease (PR)	Proteolytic cleavage of Gag (Pr55) and Gag-Pol (Pr160GagPol) precursor protein; release of structural proteins and viral enzymes
p51	Reverse transcriptase (RT)	Transcription of HIV RNA in proviral DNA
p15 (66)	RNase H	Degradation of viral RNA in the viral RNA/DNA replication complex
p32	Integrase (IN)	Integration of proviral DNA into the host genome
***Env* PrGp160 (precursor of the envelope proteins SU and TM, cleavage by cellular protease)**		
gp120	Surface glycoprotein (SU)	Attachment of virus to the target cell
gp41	Transmembrane protein (TM)	Anchorage of gp120, fusion of viral and cell membrane
Tat p14	Trans activator protein	Activator of transcription of viral genes
Rev p19	RNA splicing regulator	Regulates the export of non-spliced and partially spliced viral mRNA
Nef p27	Negative regulating factor	Myristilated protein, influence on HIV replication, enhancement of infectivity of viral particles, downregulation of CD4 on target cells and HLA cells on target
Vif p23	Viral infectivity protein	Critical for infectious virus production in vivo
Vpr p15	Virus protein r	Component of virus particles, interaction with p6, facilitates virus infectivity, effect on the cell cycle
Vpu p16	virus protein unique	Efficient virus particle release, control of CD4 degradation, modulates intracellular trafficking

1.2.6 HIV-1 replication

HIV Replication or HIV Life Cycle refers to how the HIV virus reproduce itself utilizing the genetic makeup of the Host cell. The replication process of HIV consists of several stages or steps. Each of the steps are target for HIV drugs to inhibit progress of the infection.

1.2.6.1 Attachment and entry

During binding or attachment stage, the virus attaches itself to the host cell co-receptors (CCR5 and CXCR4) on the surface of CD4+ cells by using its receptor known as gp120 (a glycoprotein). The virus uses either CCR5 or CXCR4 CD4+ host cell co-receptor depending on the viral tropism (Markosyan et al., 2003). The virus infects only CD4+ cells because CD4+ cells express the co-receptors that help the virus to enter the host cells. The binding of the of gp120 to co-receptor (CCR5 or CXCR4) leads to the insertion of the transmembrane glycoprotein gp41 fusion peptide which triggers endocytosis (Sierra et al., 2005).

1.2.6.2 Reverse transcription

Reverse transcription is the crucial step in HIV replication, allowing conversion of the single-stranded genomic RNA (ssRNA) into a double-stranded DNA with duplicated long terminal repeats. This is achieved by the viral reverse transcriptase (RT) enzyme package within the viral particle, that possesses an RNA- and DNA- dependent DNA polymerase activity as well as an endonuclease activity (RNase H). Reverse transcriptase enzyme lacks proof-reading mechanism to correct errors accumulated when transcribing viral RNA to DNA and as such it produces high mutation rates resulting in accumulation of new variant (quasispcies) increasing viral genetic diversity (Gupta et al., 2005). DNA synthesis is initiated by the cellular tRNA3 Lys selectively packaged into the virion. The complete dsDNA forms the viral pre-integration complex with other viral proteins such as integrase and Vpr bound to it (Gallay et al., 1995). This process occurs in the cytoplasm and it ends by the time the complex reaches the nucleus of the host cell.

1.2.6.3 Integration and replication

The new viral DNA is transported into the host cell's nucleus, where it is integrated into a host cell chromosome as a provirus by viral integrase (IN). The integration can either be random by means of DNA splicing or stable DNA cycle (Zhang et al., 2002). The inserted provirus replicates as part of the host genome. The transport of spliced HIV mRNA to the cytoplasm is facilitated by viral protein Rev (Martinez-Mariño et al., 2007). In the cytoplasm, translation of HIV proteins occurs within host cell ribosomes.

1.2.6.4 Assembly and Budding

Once the HIV polypeptide proteins are produced, they move out of the cytoplasm to the surface of the CD4+ cell membrane to assemble into an immature virion (non-infectious HIV). Provirus that replicates in latent state may produce new retroviruses after moving to the surface of the CD4+ cell membrane. During budding the viral proteins within the virions are then cleaved into short chains protein (functional form) by viral protease. The short chains of HIV proteins combine to form the mature HIV virus (infectious) and can then infect other CD4+ cells. The Vpu protein facilitates the release of the virus in the late stage of replication (Ganser-Pornillos et al., 2008).

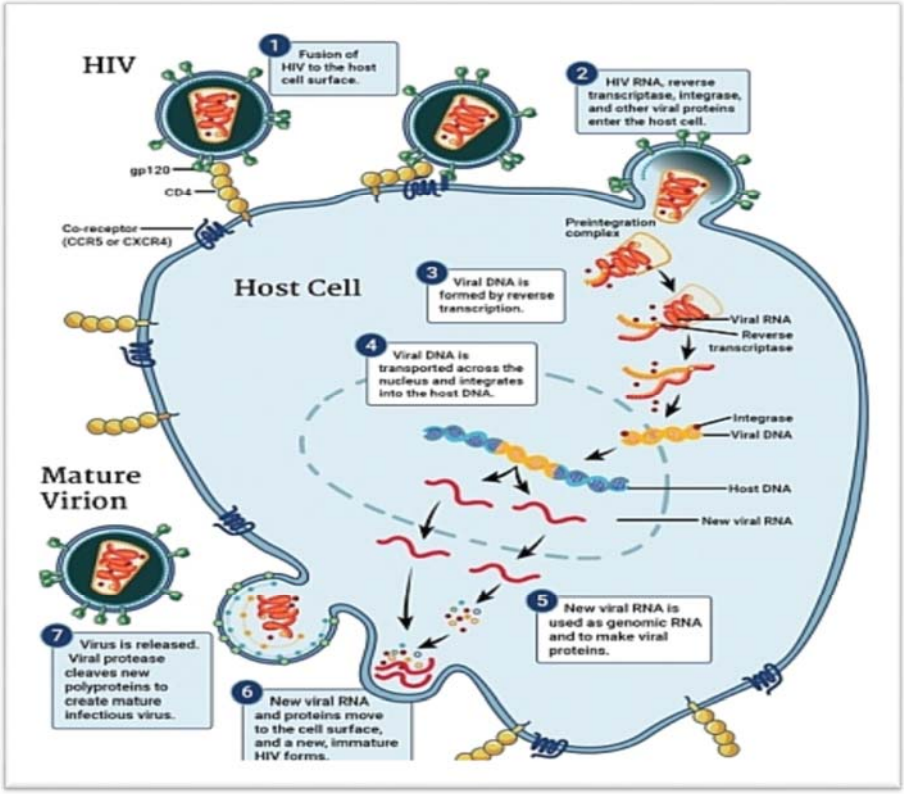


Figure 4: Overview of HIV-1 life cycle

The above figure shows the developmental process of HIV particles, intermediate state in maturation from virion to a mature virus.

(Adapted from: http://www.jotscroll.com/images/forums-concatenated-images/1508065343-HIV-Replication-Life-Cycle.jpg, 2018)

1.2.7 HIV infection

HIV attacks the immune system, especially the CD4+ T-lymphocytes. Once the CD4+ cells are infected the virus overpowers the host's defence mechanism silently and gradually allowing opportunistic infections and cancers to occur. The depletion of CD4+ cells in the peripheral blood results in immune system being compromised (Brenchley et al., 2004). People who are not on treatment or have failed treatment, the decrease in their CD4+ cells count continues over a long period of time until the individual succumbs to AIDS. AIDS is the final stage of the HIV infection (Figure 5.). AIDS can present itself between 2 and 15 years of post-infection (Moss et al., 1988). Figure 5 indicates the timeline of HIV infection progression from stage 1 to stage 4.

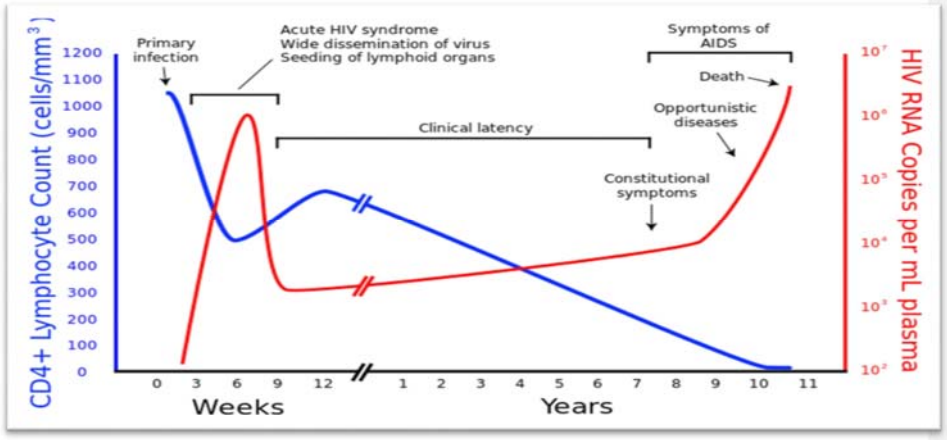


Figure 5: Clinical stages of HIV infection progression
(Adapted from: https://upload.wikimedia.org/wikipedia/commons/thumb/9/9e/HIV-timecourse-de.svg/500px-HIV-timecourse-de.svg.png)

1.2.8 HIV transmission routes

There are many ways in which HIV can be transmitted from one individual to another, which include sexual, blood transfusion, mother to child through birth canal and through sharing of syringes. The most common route of transmission is sexual.

Sexual Route: Through unprotected sexual activities with an infected person (both homosexual and heterosexual relationships).

Blood and Blood Product Route: Transmission can occur through the sharing of contaminated syringes and needles by intravenous drug users or even in health care settings where syringes and needles are re-used for different individuals, or protective wears such as gloves are not used by health care professionals, or where there is the practice of recapping used needles and through the transfusion of contaminated blood.

Mother-to-Child Transmission (MTCT) Route: HIV can be transmitted from an HIV-positive woman to her child during pregnancy, delivery and breastfeeding. Mother-to-child transmission (MTCT), which is also called 'vertical transmission', accounts for a huge number of new infections in children. In the absence of treatment, the transmission rate from mother-to-child is around 15% to 45%. However, where ART, caesarean section and other effective PMTCT are available, the risk can be reduced to as low as between 1-5%. (Cooper et al., 2002; UNAIDS, 2016).

HIV mother-to-child transmission (MTCT) in South Africa declined to just 1.5% in 2015, down from 30% in the early 2000s, surpassing the national target of 1.8% (Massyn et al., 2015). The decrease in new infection was credited to enhancements in antiretroviral treatment (ART) access, with 91% of pregnant mothers living with HIV now receiving ART from government clinics to prevent transmission to their children and for their own wellbeing (Massyn et al., 2015).

The increase in the number of women accessing ART has also led to a slight decline in the number of women who died during pregnancy and childbirth. The rate of maternal mortality caused by HIV is expected to keep on declining as more HIV pregnant women get access to treatment. Despite of the improvements, South Africa still has a fight to control HIV epidemic. The report by UNAIDS in 2012 reported about 469,000 new HIV

infections cases, with a high incidence in young women aged between15-24 years (WHO, 2013).

1.2.9 Diagnosis of HIV

HIV diagnostic tests are used to detect the presence of HIV in serum, blood or urine. These tests detect viral antibodies, viral antigens or HIV viral ribonucleic acid (RNA). Basically, there are three types of tests that can be used in HIV diagnosis.

Antibody only tests

These methods indirectly detect HIV by demonstration of the viral specific antibodies produce in response to an HIV infection in the blood within 6 weeks of exposure to HIV. The most common used antibody test for detecting HIV is rapid test also called Enzyme immuno-assay (EIA), is a common and simple method for testing for HIV antibodies, prefer by most people, because it is not expensive, and it is available from pharmacies to test at home. Enzyme linked immuno-sorbent assay (ELISA) is another antibody test for HIV detection, it is based on the principle of specific antigen-antibody reaction (Connick, 2005. HIV test results can be repeated after three months of testing to rule out the window period status, because if an HIV antibody test is performed during the window period the results may be negative. Another HIV antibody test is western blot. Western blot is often used as secondary or confirmatory test after ELISA or rapid test.

p24 antigen tests and combined antibody/antigen tests

P24 antigen assay is an indirect detection method of HIV used mostly in research studies. It is an antibody/ antigen combination assay detecting HIV antigen, a protein referred to as p24 that shows in the early stages of infection within 2-3 weeks after exposure. P24 antigen tests are often used for early detection of HIV, as p24 antigen elevate soon after infection relative to antibodies, and the test is often used in combination with an antibody test.

Nucleic acid-based tests (NAT)

These methods look directly for HIV usually in a blood rather than the immune response. Polymerase chain reaction (PCR) is one of the NAT tests. PCR is used to detect viral antigens in cases where antibody testing is not conclusive, for example in new born babies where the antibodies they harbour may have come from the mother.

HIV infection diagnostic window period is the time required after potential exposure to HIV infection for the body to produce detectable levels of antibodies (Busch,1997; Gorodin et al., 2013). If an HIV antibody test is performed during the window period to an infected person, the outcome may be negative. However, the person is infectious and could transmit HIV to others during this time. People testing for HIV are advised to return for follow-up testing in 2-3 months if the test outcome is negative. Window period is a major obstacle in the path of early and complete detection of HIV infection. However, modified ELISA tests called RITA (Recent Infection Testing Algorithm) with high sensitivity and specificity were developed to shorten the window period. RITA indicate whether the infection is likely to have been in the previous six months or not.

1.2.10 Treatment

Thus far there is no successful vaccine or cure for HIV/AIDS. However, effective management of HIV can be possible using various combinations of antiretroviral drugs. This method of treatment is known as antiretroviral therapy (ART). Standard ART consists of a combination of at least three medications (called “highly active antiretroviral therapy” or HAART) (Brass et al., 2008). HAART helps control or prevent the virus from multiplying and increases CD4 cells count, thus, prolonging the non-symptomatic phase of infection, slowing the progression of the disease, and helps in reducing the risk of transmission (WHO, 2013). As of June 2017, approximately 21.7 million people (59%) living with HIV were accessing ART worldwide (UNAIDS, 2017). An estimated 65.8% and 39.3% HIV/AIDS infected individuals in Eastern & Southern Africa and West & Central Africa respectively had access to ART, making it the first regions with more distribution of ART (UNAID, 2017). Africa region has high access to ART in 2013 (figure 6).

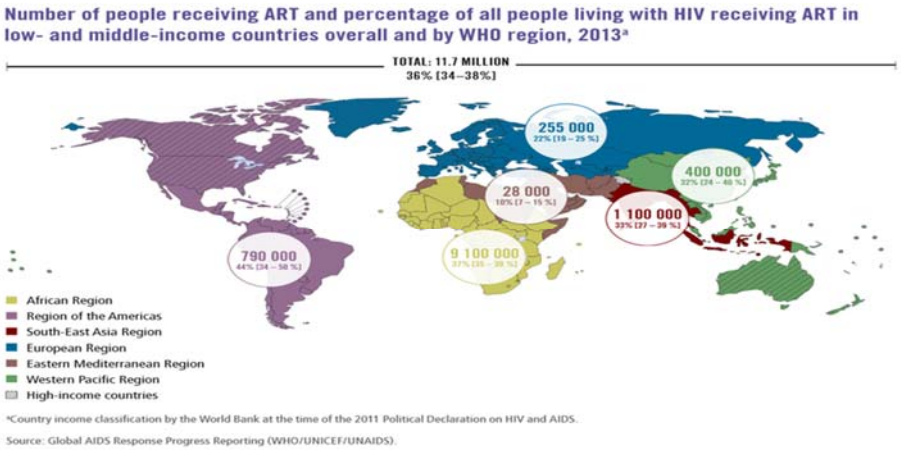


Figure 6: Map of ART coverage by WHO region in 2013 (UNAIDS, 2013)

Currently there are six classes of HAART approved by FDA which have been found to induce more than 200 mutations in the viral genome based on the mode of action namely, Nucleoside reverse transcriptase inhibitors (NRTIs), a-nucleoside reverse transcriptase inhibitors (NNRTIs), Protease inhibitors (PIs), Integrase inhibitors (INSTIs), fusion/entry inhibitors (FIs) and Chemokine receptor antagonists (CCR5 antagonists) (Table 2.). The six approved antiretroviral drugs classes comprised of 19 antiretroviral drugs,1 nucleotide and 7 nucleoside reverse transcriptase inhibitors (NRTIs), 7 PIs, 3 NNRTIs,1 fusion/entry inhibitor and 3 integrase inhibitors.

Table 2: Classification of Antiretroviral Agents and their mode of action

Class of ARV's	Example of drug	Mode of action
Nucleoside reverse transcriptase inhibitors	Tenofovir, lamivudine, stavudine, abacavir, didanosine, emtricitabine, zaicitabine, zidovudine	Inhibit replication cycle of HIV via competitive inhibition of RT and termination of the DNA chain.
Non-nucleoside reverse transcriptase inhibitors	Delaviridine, efavirenz, etravirine, nevirapine, rilpvirine	Prevent RT enzyme to transcribe from viral RNA to viral DNA.
Protease inhibitors	Atazanavir, darunavir, fosomprenavir, indinavir, copinavir, nelfinavir, ritonavir, saquinavir, tipranavir	Competitively prevent the proteolytic cleavage of polypeptides precursors into mature enzymes.
Fusion inhibitor/ entry inhibitor	Enfavirtide	Act extracellularly preventing the HIV fusion to the CD4 or another target cell.
Integrase inhibitors	Raltegravir, dulutegravir, elvitegravir	Prevents insertion of genetic materials into human genome.
Chemokine receptor antagonists (CCR5 antagonists)	Maraviroc (Selzentry)	Binds to human co-receptors preventing viral entry.

Currently in South Africa HIV treatment is recommend for all patients, irrespective of their CD4+ cell count. The first line regimens for HIV treatment in South Africa consist of two nucleoside reverse transcriptase inhibitors (NRTIs) and one non-nucleoside reverse transcriptase inhibitor (NNRTIs) inhibitor abacavir + lamivudine and efavirenz for children, while stavudine/tenofovir + lamivudine and efavirenz/nevirapine is recommended for adolescents and adults (South Africa HIV Treatment Guidelines, 2013; Meintjies et al., 2017).

1.2.11 HIV-1 drug resistance

Resistance of HIV to ARVs is one of the major causes of therapeutic failure. The emergence of drug-resistant HIV variants is a common incident, even under the best of circumstances. The first emergence of antiretroviral drug resistant HIV was reported in 1989 (Larder et al., 1989). In this case, the patients began to respond poorly to AZT therapy after a period of treatment. A situation in which there is a rebound in HIV replication during antiretroviral therapy is considered as a major cause of treatment failure (Hammer et al., 1996). Drug resistance mutations occur in a gene that is the target of the drug, which is exposed to sub optimal drug concentrations. Protease and reverse transcriptase inhibitors provoke mutations in the protease and reverse transcriptase genes respectively, which are the molecular targets of the drug (De-Jong et al., 1996; Verger et al., 2002; Menéndez-Arias., 2008; and Daiz et al., 2008).

Drug resistance mutations are categorized into two, the major (also called primary) and minor (secondary) mutations. Major or primary drug resistance mutations are mutations which can cause resistance to one or more drugs on their own (e.g. M184V) (Clavel et al., 2004; 2010). Minor or secondary drug resistance mutations are mutations which can only cause resistance if present in combination with major drug resistance mutations (e.g. E138A) (Clavel et al., 2004). Major resistance mutations affect the phenotype of the target gene whereas the minor resistance mutations do not have significant effects on the phenotype (Clavel et al., 2010).

Various methods are available for drug-resistance testing to improve the ability of clinicians to deal intelligently with HIV drug resistance. These methods include; genotypic and phenotypic resistance testing. Genotypic testing determines the resistance-related mutation pattern of the virus population. Drug-resistance testing has enabled researchers to develop novel studies as well as therapeutics to come up with treatment for patients with differing resistance profiles. All these aspects require first the knowledge on the mechanism of HIV drug resistance.

1.2.12 Mechanism of HIV drug resistant viruses

There are various methods in which HIV confers resistance to antiretroviral drugs depending on the class of the drug and its mode of action. The following are the mechanism of HIV resistance in each drug class:

1.2.12.1 Resistance to Nucleoside Reverse Transcriptase Inhibitors (NRTIs)

There are two important mechanisms by which resistance to nucleoside reverse transcriptase inhibitors (NRTIs) happen. The first mechanism involves mutations (e.g., M184V, K65R, Q151M) that emerges at or close to the drug-binding site of the reverse transcriptase gene, increasing drug discrimination by this gene. This is the primary mechanism of resistance to most of the NRTIs (Clavel et al., 2004.

The second mechanism that is of concern when talking about NRTIs includes key mutations that basically work to undo the action of these drugs, regardless of whether they do manage to bind correctly inside the RT gene. NRTIs apply a blocking effect by inserting a non-extendable nucleoside analogue monophosphate to the 3' end of the growing proviral DNA chain (Clavel et al., 2004; 2010). This process effectively terminates chain extension and inhibits viral replication. However, the process can be changed by a reverse transcriptase reaction that separate the chain-terminating residue and restore an extendable primer. This reverse reaction of DNA polymerization is referred to as pyrophosphorolysis, enables reverse transcription and DNA synthesis to resume.

1.2.12.2 Resistance to Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)

Resistance to NNRTIs occur when the hydrophobic pocket located near an active site of RT enzyme opens in the presence of the drug, inhibiting the functioning of the enzyme (Ren et al., 2008). When the enzyme is not functioning, DNA cannot be synthesized. Interestingly, the mechanism of K103N is different from the one above, even though the mutation results from NNRTIs. K103N mutation create a hydrogen bond in the entrance

of the hydrophobic pocket in un-liganded RT, which helps to close the pocket making it difficult for NNRTIs to penetrate the pocket (Clavel et al., 2004).

1.2.12.3 Resistance to protease inhibitors (PIs)

Resistance to PIs arise because of structural changes that reduce the binding of the protease to the inhibitor (Clavel et al., 2010). Many protease inhibitors resistance mutations involve amino acids that are nowhere near the binding sites for these drugs, so they end up changing the overall structure of the enzyme for them to cause resistance. The use of PIs has led to a marked reduction in mortality and morbidity in patients with advanced HIV infection. The unavoidable widespread use of PIs has led to the emergence of drug-resistant HIV variants, most of which display cross-resistance to the inhibitors.

1.2.13 Epidemiology of transmitted HIV-1 drug resistance

Infection with drug resistant viruses jeopardise the efficiency and outcome of antiretroviral therapy (ART) (Hamers et al., 2013). Viruses associated with drug resistance mutations can be transmitted to newly infected individuals, meaning there is acquisition of viruses which already have drug resistance conferring mutations (Baxter et al., 2015).

The prevalence of TDR (transmitted drug resistance) is lower in developing countries when compared to developed countries. Moreover, TDR is emerging in countries where access to treatment is being scaled up, including sub-Saharan Africa (Geretti, 2007). In different regions of the world, there are distinctive resistance patterns to the different classes of drugs. In Africa for instance, the utilization of nevirapine to anticipate mother-to-child transmission, could have expanded the predominance of transmitted resistance to NNRTIs. In Europe, there is an expanded TDR predominance in NRTIs, because of long haul presentation to Zidovudine, which was given as monotherapy before 1996 (Frentz et al., 2012).

The highest prevalence of TDR continues to be observed in North America (12.9%), Western Europe (10.9%) and areas of South America (6.3%). The prevalence of TDR in

Africa and Asia was observed to be >5% during 2004-2012 (Frentz et al., 2012). In South Africa, quite several studies have been conducted on TDR, which had a low prevalence at the start, but has been seen to gradually increase with time (Manasa et al., 2014; Gupta et al., 2011). As reported by Manasa et al., (2012), TDR has increased in most regions of South Africa, especially in the Gauteng and Kwazulu-Natal provinces, with a prevalence of >5% in studies carried out either in young primigravida or in treatment naïve individuals commencing ART (Kiepiela et al., 2014).

This increase in drug resistant viruses can be attributed to the absence of resistance genotyping before commencement of treatment. This is because resistance testing is only available to a small portion of patients, mainly in the private sector or in research settings. (Parboosing et al., 2011). In rural settings, such as those of Northern South Africa (Limpopo province), it is important to continuously monitor the prevalence of transmitted drug resistance in the population to contribute to the dataset needed to inform policy on treatment management, and research direction. However, from the few studies which have been carried out several years ago, it was shown that there was a low prevalence of transmitted drug resistance in the region (Nwobegahay et al., 2011; Bessong et al., 2006). It is important to note that this observation was when the number of individuals on ART was still very low.

The World Health Organization developed a “threshold survey” method which is suitable for use in resource-limited settings, where treatment is being expanded. This WHO method is based on lot quality acceptance sampling in recently infected, ARV-naïve antenatal mothers. However, since determining “recent” infection is challenging, newly diagnosed individuals are considered instead, as this will provide a much more reasonable chance of including a high proportion of recently infected persons among the newly diagnosed (Bennet et al., 2008).

1.2.14 Viral Phylodynamics

Grenfell and co-authors (2004) define “Phylodynamics as the study of how epidemiological and evolutionary processes act and interact to shape the viral

phylogeny". The term was coined in 2004 to shift its focus to RNA viruses in particular. The reason is that RNA viruses like influenza and HIV have very high genetic variability due to their rapid evolution which is observable over the time scale of human observation allowing phylodynamic inferences to be made. Phylodynamics uses sequence data and demographic data (sample data) incorporated in computational method like Bayesian inferences implemented in BEAST software to infer the evolutionary dynamics and epidemiology of the viruses (Shiino, 2012). A better understanding of RNA virus phylodynamics also helps with pathogen surveillance and facilitate more accurate predictions of the impact of epidemiology of newly emerged viruses and help with the control of viruses that exhibit complex patterns of antigenic variation such as dengue, influenza and HIV viruses (Volz et al., 2013; Shiino, 2012; Holmes, 2009; Holmes and Grenfell, 2009). In the current study, phylogenetic analysis was employed to help reconstruct the evolutionary history of the detected HIV drug resistant viruses and check if their transmission dynamics play a role in their genetic variation.

1.2.15 Molecular phylogenetic

Phylogenetics is referred to as the study of phylogenies—that is, the study of the evolutionary relationships of biological entities (species or genes). Molecular phylogenetics is the field in biology that uses the structure and function of molecules and how they change over time to infer the evolutionary relationships (Dowell, 2008). This field of study emerged in the early 20th century but began to be applied in 1960s, with the advent of protein sequencing, PCR, electrophoresis, and other molecular biology techniques. Molecular approach in determining phylogenetic relationships are often expressed in the form of 'trees' in which the positions and lengths of the 'branches' depict the relatedness between organisms (Brown, 2002; Dowell, 2008). Molecular phylogenetics has grown in stature largely because of the development of more rigorous methods for tree building, combined with the explosion of DNA sequence information obtained initially from sequencing techniques such as sanger and NGS. The importance of molecular phylogenetics has also been improved by the successful application of tree reconstruction and other phylogenetic techniques.

Retrovirus genomes such as HIV genomes accumulate mutations very quickly due to reverse transcriptase enzyme which lack an efficient proofreading activity during replication and so tends to make errors when it carries out RNA-dependent DNA synthesis (Steinhauer, 1992; Elena, 2005). This means that molecular clock runs fast in HIV genome that diverged quite recently display sufficient nucleotide dissimilarity for a phylogenetic analysis to be carried out. HIV genomes have sufficient data for their relationships to be inferred by phylogenetic analysis

1.2.16 Next generation sequencing

Next generation sequencing (NGS), also called deep sequencing, allows for the generation of millions of reads from a template. NGS allows generating larger volumes of sequencing data through massive parallel approach without the need to clone each molecule. All NGS technologies require DNA library preparation, sequencing and imaging, and then sequence data analysis. The advantage of NGS is that all the fragments are sequenced without the need for cloning each fragment (Aralaguppe et al., 2016). Capillary sequencing requires the knowledge of the gene to be sequenced. However, NGS does not require such information and it is used to study genomes or discover mutations (Flynn et al., 2015).

NGS has many useful applications, ranging from measuring gene expression levels to discovering rare viruses or metagenomic profiling. During replication, Deep sequencing allows the study of the genetic diversity of HIV and the clinical implications thereof. Another important advantage of NGS is the capability of detecting mutations that occur in low variant frequency that cannot be detected by traditional Sanger sequencing. This is important in drug resistance studies to get a deeper insight in the development of viral resistance within individuals and at the population level. In addition of revealing insights into intra host diversity, deep sequencing of samples from populations of infected individuals can be used epidemiologically to study transmission patterns. (Ansorge, 2009).

1.2.17 Study rationale

The introduction of antiretroviral therapy (ART) as one of the major strategies for controlling HIV associated morbidity and mortality is of paramount importance. Because of ART more than 50% of HIV infections in low and middle- income countries since its implementation was averted (UNAIDS, 2013). However, the emergence of drug resistance, mostly transmitted drug resistance (TDR), remains a concern globally. The scaling up of ART in South Africa coupled with the occurrence of new infections may lead to an increase and spread of TDR in treatment-naïve individuals.

Genotypic drug resistance assays are required to predict drug susceptibility of the virus before initiation of treatment. Sanger sequencing method has been utilized for a long period as one of the approaches for drug resistance genotyping. However, sanger sequencing is unable to detect drug resistant variants that occur as a minority population in the viral quasispecies (Quiñones-Mateu et al., 2014). Deep sequencing methods such as NGS are therefore important in the detection of minority variants that constitute as much as 20% of the viral quasispecies (Zagordi et al., 2010).

Drug resistant variants can jeopardize the efficacy of ART aimed at reducing HIV-associated morbidity and mortality. As a result, it is important to understand and monitor the spread of TDR in treatment-naive population. Understanding of the viruses harbouring TDR mutations requires knowledge of their genetic diversity, epidemiology and evolutionary history. Phylogenetic analysis serves to infer the evolutionary dynamics of virus genetic diversity (Holmes and Grenfell, 2009). The focus in phylogenetic approach is to show how phylogenetic analysis have influenced the current understanding of the emergence and evolution of drug resistance, epidemiology and dynamics of HIV.

Study Hypothesis: Chronically infected ART-naïve persons harbor HIV drug resistance viruses before initiation of treatment.

1.2.17.1 Study objectives

General objective of the study: The general objective of the study was to describe the prevalence of HIV-1 drug resistant viruses in chronically infected antiretroviral-naïve patients entering an anti-retroviral treatment programme in Limpopo Province, South Africa.

Specific objectives

The specific objectives were to:

1. To establish a next-generation sequencing protocol on the Illumina MiniSeq Platform to infer genetic drug resistance mutations.
2. Determine the prevalence and types of mutations associated with drug resistance among HIV infected treatment-naïve patients.

Chapter 2: Materials and Methods

2.1 Ethical considerations

The study was approved by the University of Venda Research Ethics Committee (SMNS/15/MBY/23/0710). Permission to access public hospitals and clinics in Limpopo province to recruit study subjects and collect specimens was obtained from Limpopo Provincial Department of Health, Polokwane. Authorities of Rethabile Community Health Centre, Seshego Health Centre, Thohoyandou Health Centre, and Donald Frazer Hospital gave permission to use their facilities to conduct the study. Figure 7 shows the geographical location of the health facilities from where study samples were obtained.
Each study participant provided a signed informed consent before demographic and clinical data were collected. Consent forms were checked for completeness stored away securely. Consent form used for collection of demographic and clinical data is attached in Appendix A. Research codes were used to tag specimens and the subsequently derived data in order to preserve the confidentiality of the study subjects.

2.2 Participating health care facilities

Study participants were recruited from Rethabile Community Health Centre and Shesego Health Centre both in the Capricorn district, and from Donald Fraser Hospital, Thohoyandou Health Centre from Vhembe Distict. Capricorn and Vhembe districts are two of the five districts constituting the Limpopo Province of South Africa.

Figure 7: Limpopo province map indicating the study sites which are Capricorn and Vhembe district respectively. Each district consists of four municipalities as indicated in the map. The study site is marked by triangles with colours red and blue respectively. Adapted from: https://en.wikipedia.org/wiki/File:Map_of_Limpopo_with_municipalities_named_and_districts_shaded_(2016).svg

2.3 Study population, and collection of samples

Individuals recruited for the study were persons who were about to enter the HIV anti-retroviral treatment programme in Limpopo Province. These individuals are prepared according to the National Treatment Guidelines. A tube of 5ml of whole blood in EDTA for viral RNA isolation, and CD4+ cell count measurement, and another tube of 5ml of whole blood in EDTA as an anticoagulant for viral load measurement was collected from each consented subject. Specimens were collected between April 2016 and April 2017.

Specimens were processed for subsequent RNA isolation or CD4+ cell count and viral load measurement with 48 hours after collection.

2.4 HIV viral load and CD4+ cell count measurements

The viral load and CD4+ cell counts are used in determining the immune competence and viral burden of each patient in the cohort. Upon samples arrival in the lab the CD4+ cell counts were determined using BD FACS-Presto flow cytometer Aquious CLTM (Beckman Coulter, Inc.) according to the manufacturer's instructions. Specimens were analysed for viral load at the Lancet Laboratories (Johannesburg) using COBAS AmpliPrep/COBAS TaqMan. The assay has been calibrated against the first WHO International standard for nucleic acid amplification techniques.

2.5 Plasma and PBMC's preparation from whole blood

Plasma was separated from total cells by centrifugation of the whole blood at 3000rpm for 5 minutes using eppendorf centrifuge. Approximately 200µl Plasma were aspirated aseptically in 2ml DNase and RNase free cryovials and stored at -80°c for subsequent RNA extraction. Peripheral Blood Mononuclear Cells (PBMC's) were isolated from the total cells using histopaque (ficoll) gradient centrifugation method. Gradient centrifugation method uses an equal ratio of ficoll histopaque and whole blood, followed by centrifugation for 30 minutes at 2800rpm. Differential migration of cells during centrifugation results in the formation of three layers containing different cell types. The interface between the plasma and the ficoll histopaque layer contains PBMC's in buffy coat. The buffy coat was washed using 1X phosphate buffer saline (PBS) and resuspended in 200µl PBS and stored at -80°c for subsequent DNA extraction. All the preparations were done strictly in aseptic conditions under level 2 biosafety conditions.

2.5.1 Total RNA, and DNA extraction

Viral RNA was extracted using an in-house TRIzol RNA extraction method. The method is based on using guanidinium-thiocyanate and isopropanol to inactivate RNases and precipitate RNA. In this method, 100µl of plasma was aspirated into a sterile 2ml

eppendorf tube and 1400µl of 5mM Tris-HCL containing 150mM of NaCl was then added to give a final volume of 1500µl. The mixture was spun for 10 minutes at 4ºc at the speed of 5300rpm. The supernatant was discarded, and the pellet was resuspended in 50µl of 5mMTris-HCL. Ten microliters of proteinase K was added to the solution and briefly mixed and incubated at 55ºc for 30 minutes. A volume of 200ml of Guanidium isothiocyanate was added, followed by the addition of 10µl of glycogen and properly mixed. The solution was incubated at room temperature for five minutes. After incubation, 270µl of 100% isopropanol was added, mixed and spun at room temperature for 20 minutes at 21000xg. The resulting supernatant was then washed with 500µl 70% ethanol, vortexed briefly and spun at 21000xg for 5 minutes at room temperature. The supernatant was discarded, and the pellet (containing RNA) was then allowed to dry for 2-3 minutes, followed by resuspension in 40µl 5mM RNase-free Tris-HCL, and then stored at -80ºC until used.
Viral DNA was isolated from PBMC in 200 µl of PBS from each sample using the Qiagen Blood DNA Mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Purified DNA was stored at -80°c until used.

2.5.2 Synthesis of complementary DNA (cDNA)

The significance of cDNA synthesis step is to reverse transcribe viral RNA into complementary DNA (cDNA) to use as template in PCR. The principle of the technique is based on using oligo $(dt)_{20}$ as primer and superscript iv reverse transcriptase enzyme to transcribe RNA. To synthesize cDNA; 10µl viral RNA, 10mM dNTP, 4.2µM Oligo $(dt)_{20}$ and DEPC-treated water was added together in sterile PCR tube. The mixture was short spun and heated in thermocycler at 65ºc for 5 minutes. While heating a concoction of the following reagents at final concentration was prepared, 4µl of 5x superscript buffer, 200U superscript IV reverse transcriptase, 20U RNase inhibitor and 5mM DTT in final volume of 20µl. Both solutions were combined in 1 PCR tube vortexed and briefly centrifuge then heated in the thermocycle at 55ºc for 10 minutes and 80ºc for 10 minutes. Rnase H of 1 µl was added into the mixture and incubated at 37ºc for 20 minutes in the thermocycler. The product was stored in -40ºc for subsequent PCR amplification.

2.5.3 First round and nested polymerase chain reaction (PCR)

The transcribed RNA (cDNA) was used as template for first round PCR carried out in the final volume of 50µl tube containing, 1X PCR buffer, 0.2µM of each of the primers, 80µM dNTP mix, 0.0125U platinum Taq polymerase, 1.5mM magnesium chloride (MgCl2) and 5µl of cDNA. Nuclease free water was added to make up the final volume to 50µl. Amplifications were carried within the Proflex PCR system thermocycler with the following cycling conditions: initial step of 95℃ for 2minutes, followed by 30 cycles of 95℃ for 1minute, 60℃ for 1minute and 72℃ for 2minutes and a final extension time of 10 min at 72℃. The concentrations and final volume for First round and nested PCR was the same and the cycling conditions were the same as well except for the annealing temperature for nested PCR which was 57ºc. A partial polymerase fragment of about 1.650 bp was the target gene and was generated using the following primer pairs:

Table 3: Primers for amplifying HIV-1 partial pol region

Primer name	**Primer direction**	**Primer sequence**
Primer sets for first round PCR		
1395	Forward	5'-tggcaaggaagggcacatagccaaaaaattg-3'
1353	Reverse	5'-ttaggagtctttccccatattactatgcttt-3'
Primer sets for Nested PCR		
1389	Forward	5'-aaattgcagggcccctagg-3'
1396	Reverse	5'-ctctgttaactgttttacatcattagtgtggg-3'

2.6 Visualization of nested PCR product with agarose gel electrophoresis

The products of PCR were analysed using 1% molecular grade agarose gel (Sigma-Aldrich) in 100ml of 1X TAE buffer. The gel was stained with 5µl ethidium bromide (10mg/µl) and 5 µl PCR product was loaded into the gel together with 1kb DNA marker. Electrophoresis was conducted for 30 minutes in electric field strength of 80V then

visualized under UV transilluminator (Syngene, Germany). This was performed to verify the correct size of the expected products.

2.7 Measures taken to eliminate contamination.

Elimination measures to prevent contamination were considered during all processes of samples handling and amplification. That was achieved by working under a validated biosafety cabinet (hood) all the time. The working space in the hood was sterilized by 70% ethanol together with RNase and DNase free reagents. All tubes used were RNase and DNase free.

2.8 Establishment of a next-generation sequencing protocol for drug resistant studies

Next generation sequencing (NGS) protocol was the main protocol of sequencing to determine minority drug resistance variant in the study population. The protocol was established in-house with the guidance from the guideline manuals (Nextera XT work flow) provided by the whitehead scientific company and other reading materials retrieved from internet. The flow of the protocol from the first step to the last is as follows:

2.8.1 Step 1: Purification of PCR products

PCR products were purified as required for next generation sequencing protocol using AMPure XP magnetic beads. The method is based on the use of magnetic beads unto which DNA binds and prevent the DNA from being washed away during the subsequent washing step.

Consumables involves in the method such as resuspension buffers (RBS), AMPure XP beads and 80% ethanol were prepared according to the manufacturer's protocol. The ratio of the beads and the PCR products (3:2) was maintained as the standard requirement. The beads were vortexed and briefly centrifuge at 280xg at 20ºc for a minute. About 45µl of each PCR products and 27µl of the beads were mixed together in a PCR tube and shaken at 1800 rpm for 2 minutes, then incubated at room temperature for 5 minutes. The mixture was placed on a magnetic stand to clear, and the supernatant

was discarded. Two-hundred microliters of 80% ethanol was added to wash the impurities. Ethanol was then discarded, and the beads coated with DNA allowed to dry in air for 15 minutes. The beads were removed from the magnetic stand and resuspended in 30µl RBS. The solution was shaken at 1800rpm for 2 minutes and incubated at room temperature for 2 minutes then placed again on the magnetic stand to clear. Thirty microliters of the solution (DNA) was transferred into clean PCR tube and stored at -40ºC.

2.8.2 Step 2: DNA library preparation and sequencing

Sequencing of the purified DNA was done using the Illumina Mini-Seq instrument following the manufacturer's protocol. The purified DNA samples were quantified using Qubit 3.0 instrument with dsDNA high sensitivity kit, with a detection range of 10pg/µl to 100ng/µl. The samples were then normalized to a final concentration of 2 ng/µl by diluting with EB buffer. The Nextera XT DNA Library prep workflow was followed as prescribed by the Nextera XT DNA Kit for library preparation. The workflow consists of the following processes:

1.Tagmentation of genomic DNA: In this step, DNA is being fragmented randomly by the Nextera transposase and tagged with the adapters sequences in one step.
2. Library Amplification: After tagmentation with Nextera kit, molecular tags were added in via short PCR. The PCR step added Illumina Index 1 and 2 adapters (i7 and i5 adapters) using the TruSeq Index Plate Fixture.
3.Clean up libraries: This process uses Size exclusion bead purification. AMPure XP beads are used to purify the DNA libraries and provides a size selection step that removes unadded tags short library fragments. DNA library was then quantified using Qubit high sensitivity kit to determine the concentration of the amplified DNA fragments. The resulted size of the libraries was confirmed on E-gel.
4. Normalized Libraries: Samples were pooled together in this step at an equal molar ratio of 1nM per sample. This was based on the size of the library of each library to ensure more equal library representation in each pooled sample.
5. Pool Libraries: This step combines all normalized libraries of equal volumes in a single tube. After pooling of the libraries, they were diluted and denatured using 0.1N of NaOH.

Phix was diluted as well using the same concentration of NaOH as the libraries and used as control during sequencing. The denatured library (1.8pM) was spiked with 25% of Phix then loaded on the sequencing Output cartridge 300 cycle and sequenced.

2.8.3 Step 3: Sequenced data clean up and analysis

After completion of the run, FastaQC sequences were retrieved by copying into a USB drive. The sequences were then subjected to FastaQC software to determine the quality of the generated reads per sample. The sequences together with HIV-1 reference sequence (GenBank accession number AY585267.1) were imported into the geneious software (version 11.2.1). The beginning and end of each sequence (complete HIV protease (PR) (297bp) and ~1104bp of RT) was annotated on the reference sequence. Incorrect base calling generated during sequencing were filtered and trimmed out. The trimmed fragments of PR and partial RT were assembled and mapped with the reference sequence and used to generate consensus for each sample. Consensus per sample were generated at frequency threshold of <20% with <1% off. Generated consensus for each sample was copied into a text file format and used for viral subtype analysis.

2.9 Drug resistance genotyping

HIV drug resistance genotyping was performed using Stanford HIV Drug Resistance database Interpretation Algorithm tool http://hivdb.Stanford.edu. The tests sequences consensus were submitted to Stanford HIV resistance database for analysis and interpretation of DRM's and transmitted drug resistance mutation. Variant (SNP) were called with geneious software (version 11.1.2) using frequency threshold of <20% with < 1% cut off. The Stanford HIV resistance database offers HIV-1 genotyping and antiretroviral test. The tests evaluate HIV-1 reverse transcriptase and protease genes respectively and provides inferred resistance information for 19 common prescribed RT and PR inhibitors. The database also shows the percentage of resistance and the significance of the resistance found as well as the HIV subtype. The mutations obtained from the database were compared with WHO surveillance transmitted drug resistance list.

2.10 Phylogenetic analysis (viral subtyping) and Recombination analysis

HIV-1 subtyping was done by phylogenetic analysis of partial *pol*. The entire protease (297bp) and partial reverse transcriptase fragment (1104bp) were aligned with reference sequences (HIV-1 subtypes A-D, F-H, J and K) obtained from GenBank using muscle Clustal W incorporated in Mega 7 software. Sequences from previous studied conducted in Limpopo province were included in phylogenetic analysis. Neighbour-joining phylogenetic trees rooted with group O strain were generated to infer the evolutionary relationship between sequences using PHYLIP programme. The Bootstrapping of 1000 replicates was used to assess the tree reliability. The mean genetic difference was calculated using Kimura 2 parameter method which takes into consideration the transition and transversion substitution rates while assuming that the four nucleotide frequencies are the same and that rates of substitution do not differ among sites.

Recombination analysis tools (REGA and JpHMM) were used for further analysis of the subtypes. To confirm the subtypes obtained from phylogenetic analysis and Stanford drug resistance tool sequences were submitted online to REGA subtyping tool version 2.0 and Jumping profile Hidden Markov Model (JpHMM) database. Both tools determine the subtypes and reveal any recombination when presents. JpHMM relates to nucleic acid sequences of various alignment to a sequence family in which a classification of sub classes is in existence (Alcantara et al., 2009 and Zhang et al., 2006). JpHMM was used in detecting recombination break points in test sequences based on HXB2 and REGA was used as confirmatory tool for subtypes.

2.11 Summary of the methodology

Summarised methodology including all the steps from plasma preparation to viral DNA amplification and sequence, as well as the technology used for sequencing is presented in figure 8.

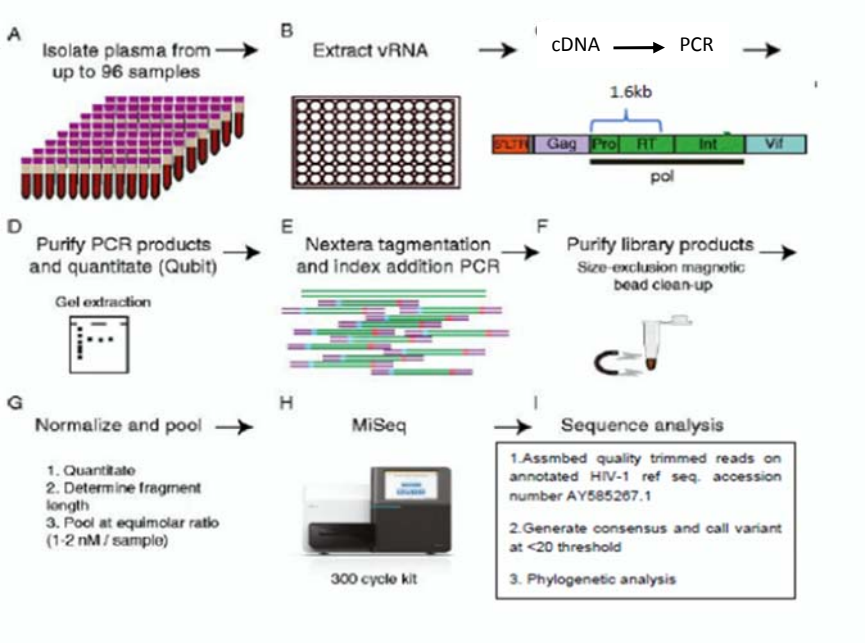


Figure 8: Overview flow diagram of the NGS methodology

(A)Plasma is isolated from whole blood. (B) Viral RNA is isolated from up 200 µl of plasma. (C) Viral RNA is used in PCR amplification of 1.6 kb region of the partial pol gene (D) PCR products were visualized by gel electrophoresis followed by purification using the magnetic beads and then quantified by the Qubit system. (E) Purified products were randomly fragmented and subjected to cycle PCR to add sequencing adaptors and indices used for multiplexing samples. (F) Libraries were purified by size-exclusion magnetic beads to remove short fragments. (G) The average size of the library fragments estimated by E gel visualization. The final concentration of the libraries calculated by Qubit were normalize and pooled together at equimolar ratios. (H) Libraries were sequenced on the Illumina Mini-Seq machine. (I) Geneious Pro Software (version11.1.2) was used to trim sequencing reads based on QC scores and assemble the reads to a reference sequence, then generate consensus sequences.

Chapter 3: Results

3.1 Demographic information

Overall, 29 HIV positive patients without prior exposure to antiretroviral therapy (treatment-nave), were enrolled into this study between 2016-2017. The sociodemographic and clinical characteristics of the participants are presented in Table 3.1. In regard to gender and age, 79.3% were female and 20.7% were males. The mean age of the participants was 46 years, ranging from 23 to 56 years old. Most participants attended school, however only 75.9% of the participants passed grade 12. The proportion of unemployed (58.6%) participants were a bit high when compare to employed (41.4%). Only 31% of the participants were married. Among all 29 participants, 89.7% were infected through sexual intercourse and 10.3% were unware of the possible mode of infection. Sexual activity was the major pathway for HIV-1 transmission within the cohort. Viral load and CD4+ counts was available for only 18 out of 29 of the samples due to cost of the techniques required for these test, lack of the equipment especially for viral loads on site and period in which samples were transported from collection site to the lab. For viral loads and CD4+ counts ranges and values refer to table 4. Over 89.7% of the participants were in stage 1 of the disease (AIDS) and 10.3% were in stage 2.

Table 4: Demographic and clinical data of the study participants

Characteristics	Number of study participants
Gender	
Females	23 (79.3%)
Males	6 (20.7%)
Age (years)	
≤ 35 years	11 (37.9%)
≥ 35 years	18 (62.1%)
Age mean and range	46.2 (23-56 years)
Education/qualification	
Grade 1-7 (Primary)	3 (10.3%)
Grade 8-12 (Secondary and tertiary)	22 (75.9%)
Unknown & illiterate	4 (13.8%)
Employment status	
Unemployed	17 (58.6%)
Employed	12 (41.4%)
Marital status	
Single	19 (65.5%)
Married	9 (31.0%)
Widow	1 (3.4%)
Possible mode of infection	
Sexual	26 (89.7%)
Unknown	3 (10.3%)
CD4+ counts (cells/mm^3) prior to treatment initiation	
≤ 350 cells/mm^3	16 (55.2%)
≥ 350 cells/mm^3	2 (6.9%)
Unknown	11 (37.9%)
Viral load (copies/ml) prior to treatment initiation	
≤ 1 000 copies	1 (3.4%)
≥ 1000 copies	16 (55.2%)
Unknown	12 (41.4%)
AIDS stages (WHO)	
Stage 1	26 (89.7%)
Stage 2	3 (10.3%)

3.2 Polymerase chain reaction results

HIV-1 gag/pol gene (positions 1998–3628 corresponding to HXB2) was successfully amplified from 22 (75.9%) participate out of 29. The gene is comprised of the whole protease (297bp) and a fragment of reverse transcriptase (1104bp). Polymerase gene is the location for enzymes which partake in HIV replication, thus making the site target for the antiretroviral drugs. The successfully amplified samples were purified using Ampure XP beads following manufacture's protocol. Representative gel photos of PCR amplicons before and after purification are shown in figure 9 and 10 respectively.

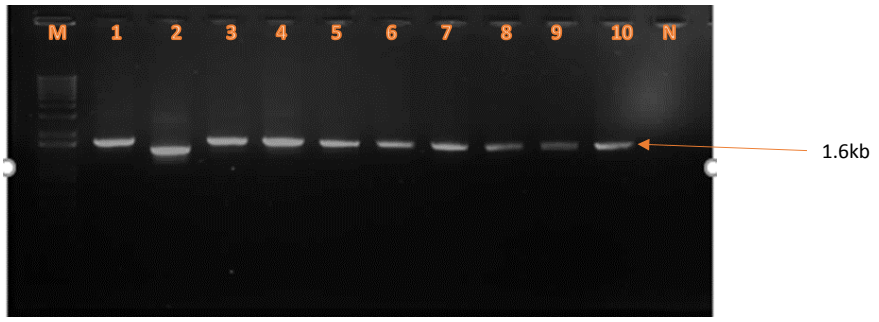


Figure 9: Representative amplicons for first and nested PCR loaded in 1% agarose gel. Lane M is the DNA track 1kb plus marker. Lane 1-10 are positive amplicons. Lane N 3 is the negative control (nuclease free water). The arrow indicates the target size.

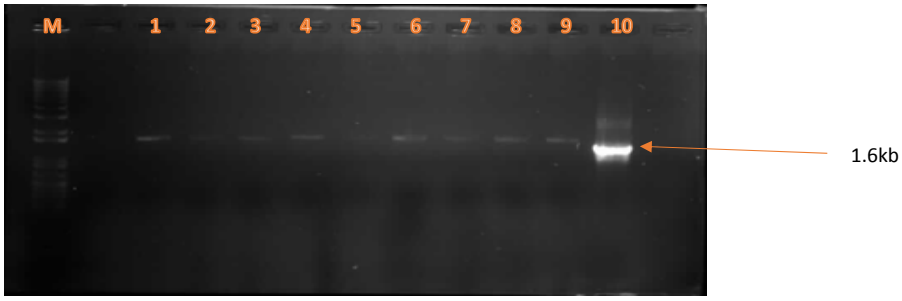


Figure 10: One microliter of purified amplicons loaded on 1% agarose gel stained with ethidium bromide and run at 80 volts and 100 Ampere of electric field strength. Lane M is

the DNA track 1kb plus marker. Lane 1-10 are positive amplicons. Lane N 3 is the negative control (nuclease free water). The arrow indicates the target size.

3.3 NGS validation

The NGS protocol used was validated based on the matrices obtained at the end of sequencing. The MiniSeq run generated about 4.2 Gb of data at the quality score of 97% not bad when considering that MiniSeq system normally yield data between 6.6-7.5 Gb. From 25% Phix loaded on the run, 24.38% was recovered after the run. The amount of Phix recovered after the run indicate that the error rate on Phix was 0.62% which was very low, indicating that the pipetting was done properly.

3.4 Sequences generated during sequencing

Sequences were available for only 14 samples in total of 22 sequenced samples. There were no data generated during sequencing in 10 samples, this might be due to the low input DNA concentration of these samples during sequencing. From 14 generated sequences 2 of them were ruled out due to bad quality after subjecting them to QC report. Overall, a total of 12 sequences were analyzed for drug resistance mutation and subtype by phylogenetic analysis.

3.5 Drug resistance mutations detection and interpretation

PR and RT genes of 12 viral isolates from ART-naïve persons were analysed to determine the presence of drug resistance mutations. The Stanford sequence database analysis did not detect any major DRM's in PR in all isolates, however polymorphisms (**Q58E, K20R, K20M, M36I, L89M, I13V, I93M, L63V/P/S, V771, T74S**) occurred at sites associated with secondary drug resistance at frequency threshold of <20%. The polymorphisms were classified by Stanford HIIV database as Accessory-selected PI DRM's. The most predominant in all 12 isolates were **M36I, L89M, I93M, L63V/P/S** and **V771**, with almost 50-90% of the isolates having them (**Table 5**). None of the isolates harboured major DRM's towards NRTI in RT gene. Sample PK089D was the only one with 1 NNRTI (**V179D**) mutation in all 12 samples. Other polymorphism near the drug sites of RT gene were detected (Table 6) with no impact to both NRTI and NNRTI

Table 5: Secondary (accessory) drug resistance mutations detected in PR gene of each isolates. No major (primary) PR drug resistance was detected

Sample code	PI secondary (Accessory / Minor) resistance mutations		Comments
	Polymorphic PI	Non-polymorphic PI	
DF003	T74S, M36I, L63V, I93L, L89M	NONE	Polymorphic mutations detected in DF003, they increase replication fitness in subtype B viruses. They natural subtype C consensus.
PK005	V771, L89M, L63P, I93L,	NONE	
PK012	K20R, M36I, L63V, T74A, L89M, I93L	NONE	K20R is a highly polymorphic PI-selected accessory mutation that increases replication fitness in viruses with PI-resistance mutation.
PK026T	I13V, L63P, V77I, I93L,	NONE	I13V is a potential contraindication to the use of PI inhibitors.
PK069	M36I, L89M, I93L,	K20M	K20M/V is not well studied but is a potential contraindication to the use of PI inhibitors.
PK070	M36I, L63T, I93L	Q58E	Q58E It is likely associated with low-level resistance to TPV and other PIs
PK071	V77I, L89M, I93L, L63P	NONE	
PK087	T74S, V77I, L89M, L63V, I93L	NONE	V77I is a potential contraindication to the use of PI inhibitors.
PK089D	M36I, L89M, L63A, I93L, I93L,	NONE	
PK092	I13V, L63S, V771, I72V	NONE	I13V, L63P
PK199	M36I, L89M, I93L	NONE	
U002	V771I, L89M, I93L, L639	NONE	
M36I, L89M, I93L and T74S are Polymorphic in non- subtype B viruses and occur in almost 15% of ART-naïve persons. They increase replication fitness in subtype B viruses.			

Table 6: NNRTI and NTRI drug resistances with other detected unusual mutations. No major DRM mutation detected.

Sample code	NNRTI resistance mutations	NRTI resistance mutations	Other Mutation detected in both NNRTI and NRTI
Df003	NONE	NONE	V35T, E36A, T39E, S48T, V60I, D121Y, K122E, K173T, Q174K, D177E, G196E, T200A, Q207E, R211K, F214L, V245N, D250E, S251D, A272P, K275Q, K281R, T286A, E291D, V292I, I293V, E312Q, Q334N, G335D, R356K, G359T
PK005	NONE	NONE	G15D, V35T, T39A, S48T, K122E, D123G, K166R, K173T, D177E, T200A, Q207N, R211K, V245Q, E248T, A272P, K277R, E291D, I293V, T296N, G335D, R356K, G359T
PK012	NONE	NONE	K30R, V35T, T39E, S48T, P95S, K122E, D123N, K173T, D177E, I178M, E194G, T200A, E204K, Q207E, V245K, A272P, K277R, T286A, E291D, V292I, I293V, K311R, D324E, Q334E, R356K, G359S
PK026T	NONE	NONE	V8I, V35T, E36A, T39D, S48T, K122E, D123N, I135T, S162C, E169D, K173A, Q174K, D177E, T200A, Q207E, R211Q, V245Q, A272P, K277R, T286A, E291D, V292I, I293V, Q334H, G335D, R356K, G359T
PK069	NONE	NONE	V35T, T39D, S48T, K122E, K173A, Q174K, G196E, T200A, E204Q, Q207D, R211K, P243A, V245Q, E248D, D250E, A272S, T286A, E291D, V292I, I293V, E297A, A304S, N306P, R307I, E308K, I309D, K311I, E312A, P313E, V314I
PK070	NONE	NONE	V35T, E36A, T39D, K102Q, K173A, D177G, T200A, Q207E, R211K, V245Q, A272P, K275R, K277R, Q278H, E291D, V292I, I293V, V314I
PK071	NONE	NONE	V35T, T39A, K43E, S48T, K122E, D123G, K166R, K173T, D177E, T200A, Q207N, R211K, V245Q, E248T, A272P, K277R, E291D, I293V, T296N
PK087	NONE	NONE	V35T, T39E, K122E, D123S, I135V, K166R, K173A, Q207D, R211K, V245K, A272P, K277R, Q278H, K281R, T286A, E291D, V292I, I293V, L325I, E328D, Q334N, G359T

PK089D	**V179D**	NONE	V35T, T39E, K122E, D123S, I135V, K166R, K173A, Q207D, R211K, V245K, A272P, K277R, Q278H, K281R, T286A, E291D, V292I, I293V, L325I, E328D, Q334N, G359T
PK092	NONE	NONE	V35T, K122E, D123E, Q174H, D177E, T200A, I202V, Q207E, R211K, F214L, A272P, K277R, I293V, I329V, G333E, G335N, T338S, R356K, G359S
PK199	NONE	NONE	V35T, E36A, T39D, S48E, I159V, K173A, D177G, T200A, Q207E, R211K, V245Q, A272P, K275R, E291D, V292I, I293V, G335D, R356K, G359T
U002	NONE	NONE	V35T, E36A, T39E, S48T, K122E, D123N, I135T, S162C, K173A, Q174K, D177E, I195K, G196P, Q197L, T200A, Q207A, R211K, F214L, V245Q, A272P, K275Q, Q278N, T286A, E291D, I293V, P294S, E297K, R307K, D324E, Q334H, G335D, R356K, M357T, G359T
V179D is a polymorphic accessory NNRTI-selected mutation that contributes low-levels reductions in susceptibility to each of the NNRTIs (EFV, ERT, NVP and RPV). V179D do not reduce the virologic response to first-line EFV containing regimen.			

3.6 Viral subtyping and recombinant analysis

REGA version 2.0 HIV subtyping tool and Jumping profile Hidden Markov Model (JpHMM) were used to determine the viral subtype and recombinants analysis of each isolates. The results displayed on the software after analysis revealed the subtype assigned to each sequence and the precise location or position on the genome with reference to HXB2. It further displays a bootscan analysis for recombination analysis of each sample. Both software shows the majority of the sequences (11 (91.7%)) as subtype C viruses and 1 (8.3%) as subtype B (Figure 11, 12, 13, 14, 15 and 16). Recombination were observed in all samples at very low probability, the results are displayed in Figure 12, 14, 15 and 16 respectively.

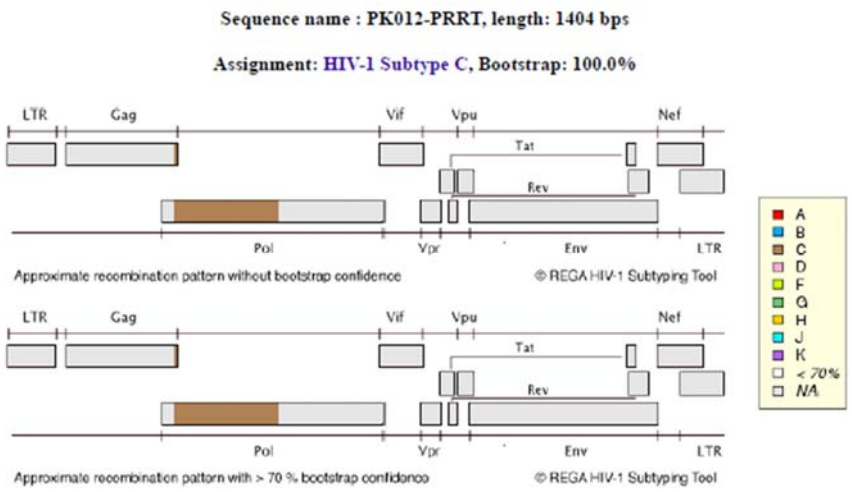


Figure 11: Genome map based on HXB2 numbering for the PR-RT region of sample PK012 showing Subtype C as the assigned subtype based on REGA analysis. All assigned subtype C samples have the same schematic and coloration as the above figure.

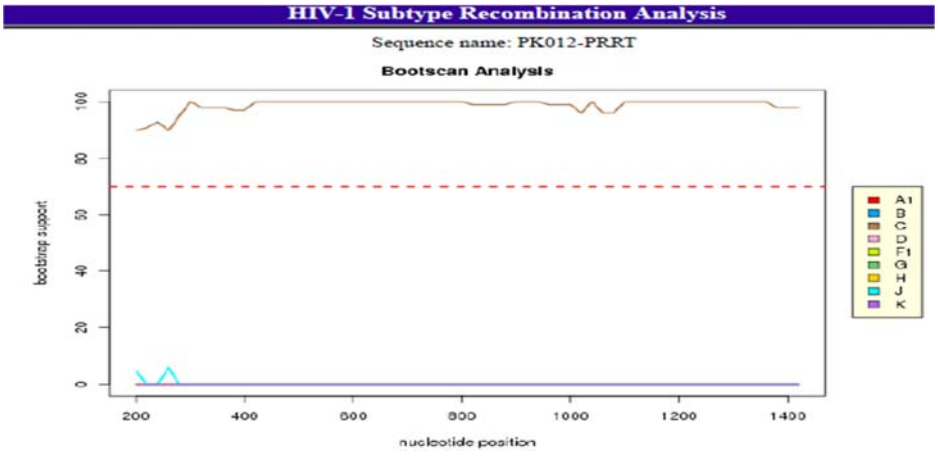


Figure 12: Bootscan recombination analysis for sample PK012 base on REGA. Sample PK012 recombine at a very low probability with subtype K throughout the gene and subtype J on the beginning of the gene.

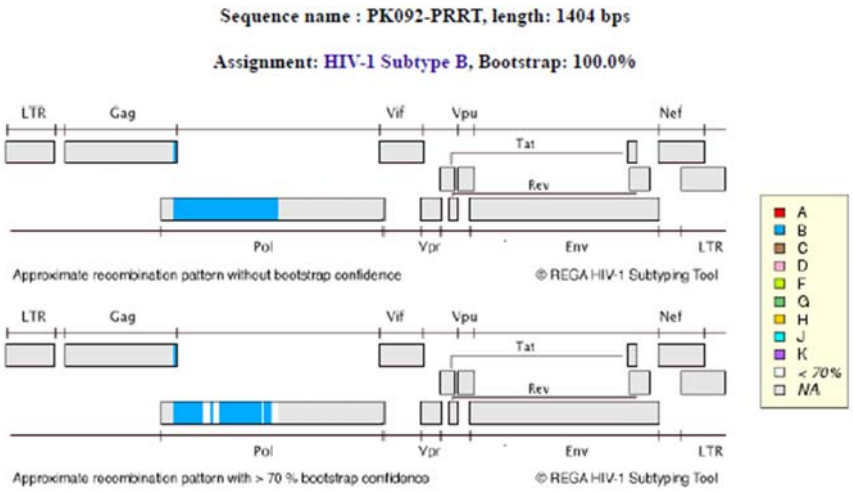


Figure 13: Genome map based on HXB2 numbering for the PR-RT region of sample PK092 showing Subtype B as the assigned subtype as indicated by the blue coloration.

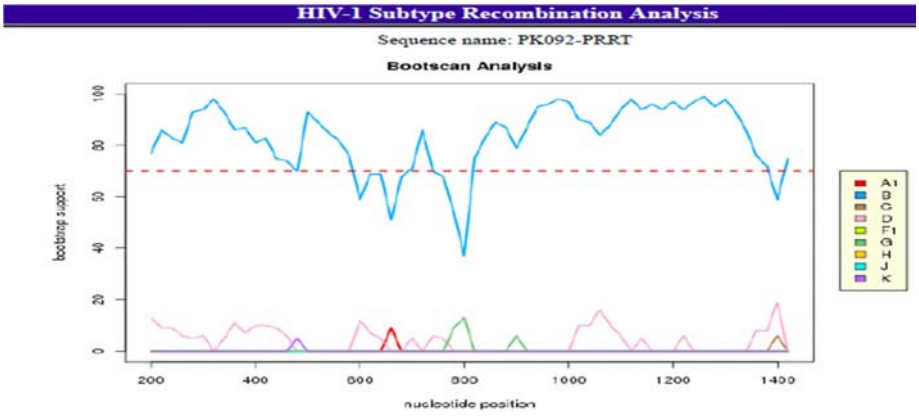


Figure 14: Recombinant analysis of sample PK092 using REGA shows the sample recombining with various subtypes (A1, C, D, G and K) at a very low rate, as indicated by different colors that represent each subtype on the above figure.

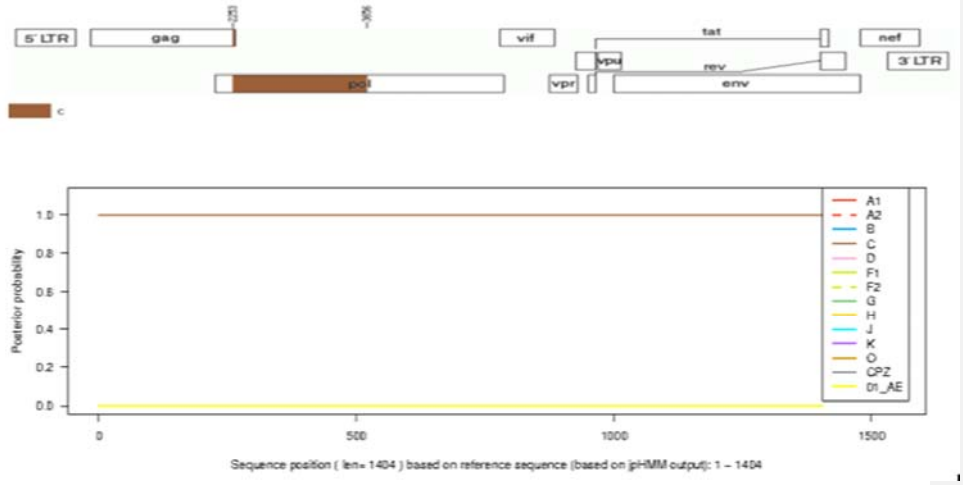


Figure 15: Genome map based on HXB2 numbering for the PR-RT region of sample PK012 showing Subtype C as the assigned subtype based on jpHMM analysis. All assigned subtype C samples have the same schematic and coloration as the above. The sample is also seen to recombine at a very low probability with CRF01_AE, as shown by the yellow coloration.

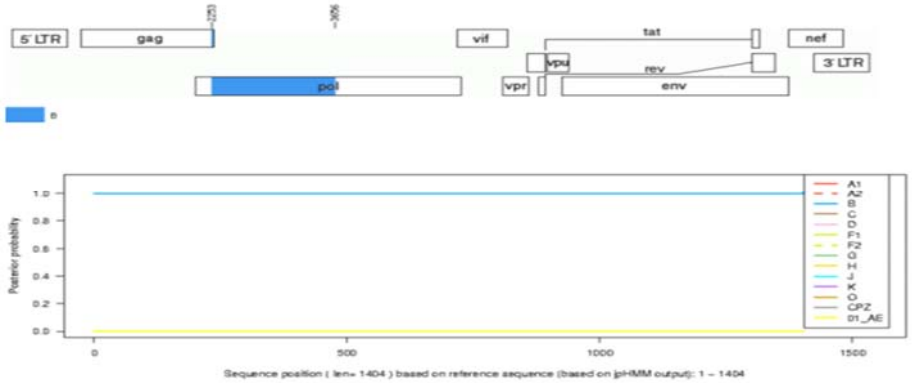


Figure 16: Genome map based on HXB2 numbering for the PR-RT region of sample PK092 showing Subtype B as the assigned subtype based on jpHMM analysis. The subtype is indicated by blue coloration. The sample is also seen to recombine at a very low probability with CRF01_AE, as shown by the yellow coloration.

3.7 Phylogenetic analysis

Phylogenetic analysis of 12 PR and RT sequences with reference sequences from different subtypes confirmed that 11 (91.7%) isolate clustered with subtype C (Figure 17and 19) and 1 (8.3%) isolate clustered with subtype B (Figure 18 and 19). Pairwise nucleotide distance analysis indicated a 2.1–10.2% variation for the PR and 2.4–7.2% for the RT region. The majority of test samples cluster with subtype C virus from South Africa and other countries like Botswana, UK, and USA. The same profile pattern was observed in PR-RT fragment (protease and reverse transcriptase fragments) (Figure 19). Phylogenetic tree to infer evolutionary relationship generated for each fragment of partial pol (PR and RT) and combined of both PR-RT was to check if the isolates in each fragment can intermingle with other subtypes.

Translated amino acids of PR and RT genes were aligned respectively with subtype C reference sequence from South Africa (Accession no. AY585267.1) obtained from Genbank using Multiple alignment on Bio-edit software Figure 20: (A) and (B). Matrix plot represented by dots was used to determine agreement between reference sequence and the test sequences samples. Dots indicate agreement, where there is changes in amino acids is indicated by bold letters.

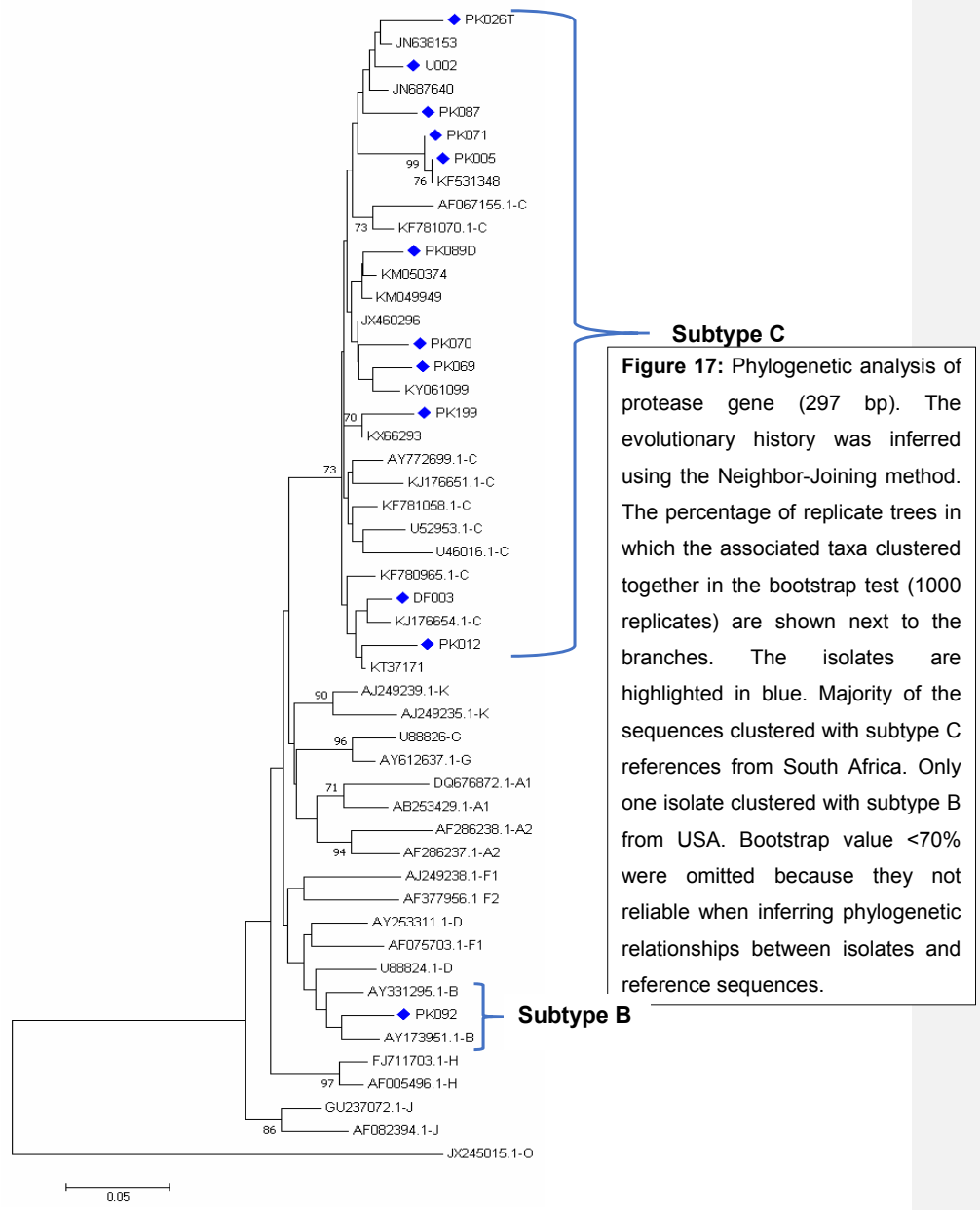


Figure 17: Phylogenetic analysis of protease gene (297 bp). The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The isolates are highlighted in blue. Majority of the sequences clustered with subtype C references from South Africa. Only one isolate clustered with subtype B from USA. Bootstrap value <70% were omitted because they not reliable when inferring phylogenetic relationships between isolates and reference sequences.

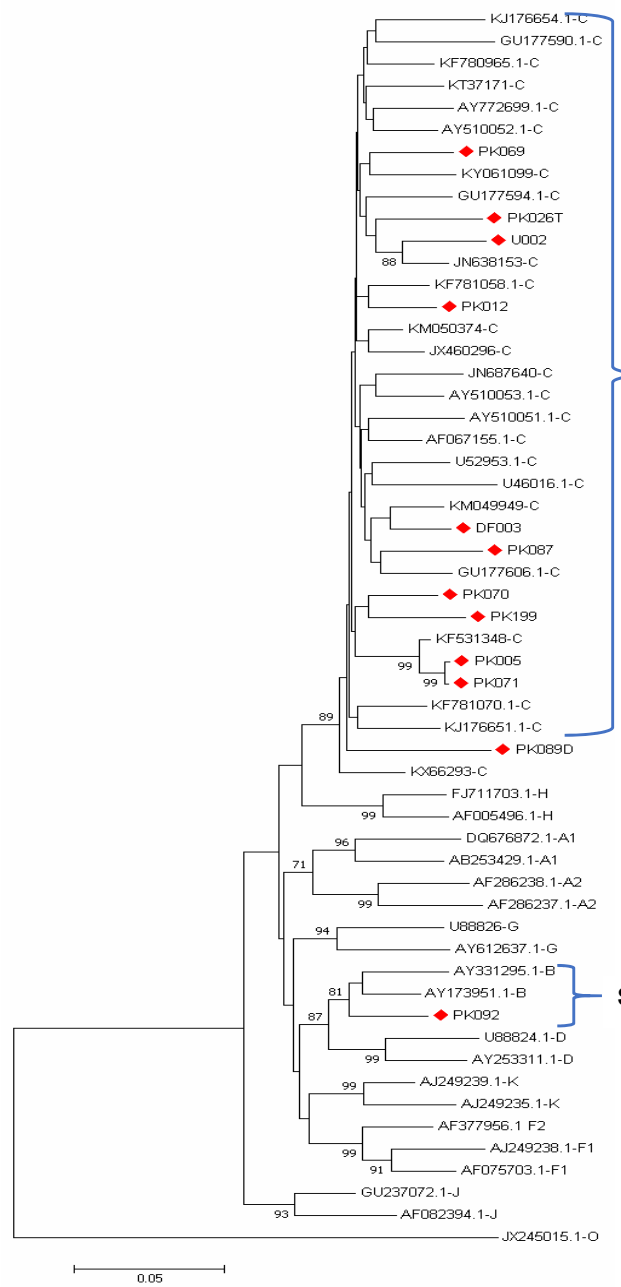


Figure 18: Phylogenetic analysis of reverse transcriptase gene (1104bp). The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The isolates are highlighted in red. Majority of the sequences clustered with subtype C references from South Africa. Only one isolate clustered with subtype B. Bootstrap value <70% were excluded in the tree because they are not reliable in phylogenetic relationship.

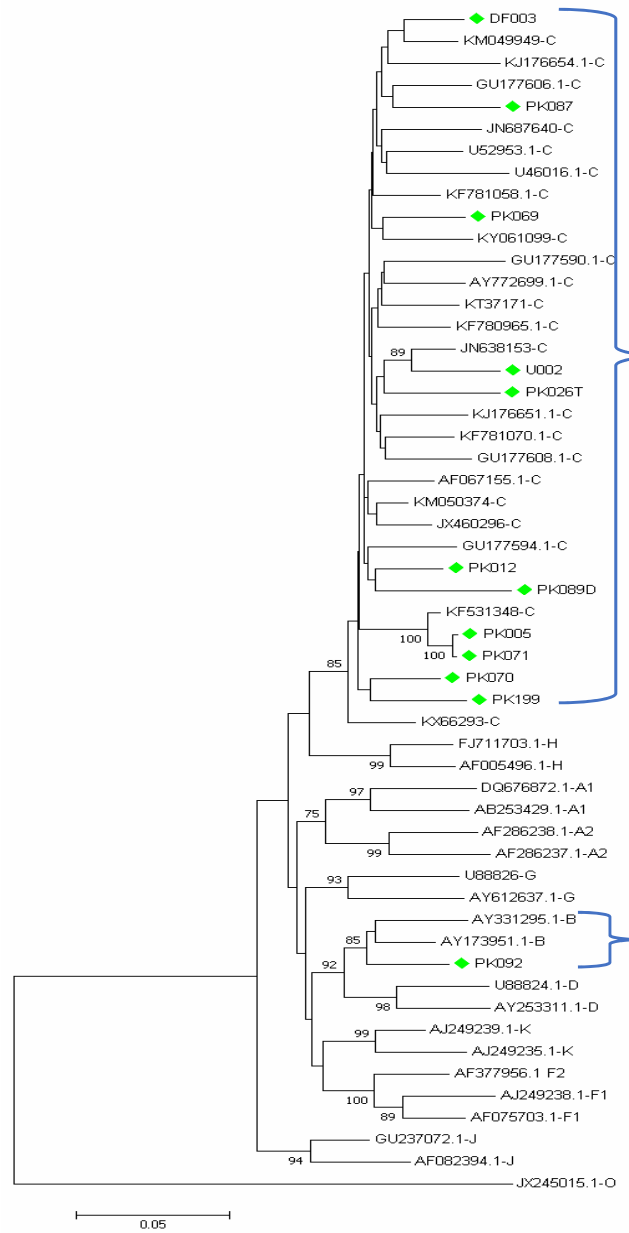
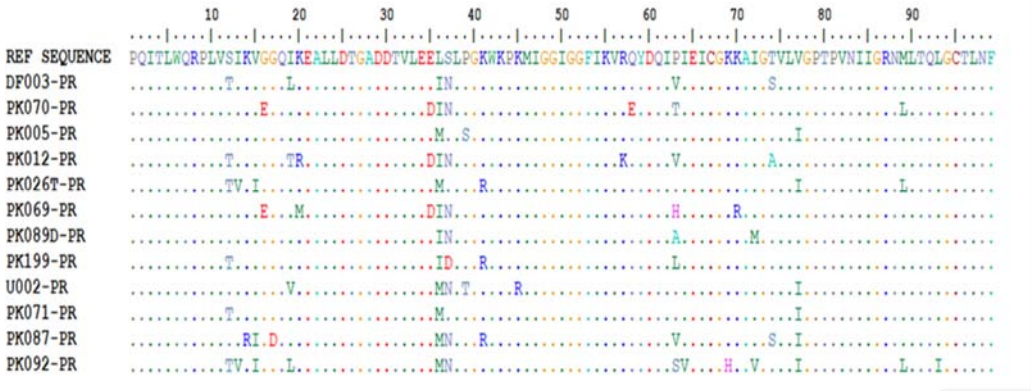


Figure 19: Phylogenetic analysis of partial pol gene (PR~297bp and RT~1104bp). The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The isolates are highlighted in green. Majority of the sequences clustered with subtype C references from South Africa. Only one isolate clustered with subtype B. Bootstrap value of <70% were omitted on the phylogenetic tree due to their insignificance.

A



B

Figure 20: Predicted amino acid sequences of protease **(A)** and reverse transcriptase **(B)** of 12 drug-naïve patients. The sequences were aligned against a Subtype C reference sequence from South Africa (Accession no. AY585267.1) obtained from Genebank. There were deletions and insertions in reverse transcriptase gene of 3 samples (PK069, PK070 and PK071). The matrix plots represented by dots indicate agreement between reference sequence and the isolates, where there is changes in amino acids is indicated by bold letters**.**

.

Chapter 4: Discussion, limitations and conclusion

4.1 Discussion

Antiretroviral treatment (ART) has grown to be one of the most effective tools in the fight to control HIV morbidity and mortality rate worldwide. However, due to the emergence of drug resistant HIV, ART efficacy can be jeopardized. Drug resistant HIV has a potential of becoming a major public threat, as its limit treatment options on people living with HIV. With different scientific finding worldwide reporting drug resistant HIV to be currently being transmitted to ART-naïve persons, measures have been taken to genotype drug resistant HIV prior to treatment initiation. However, in resource limited settings such measures are not executed especially in public sectors due to the costs associated with the required assays for genotyping.

HIV drug resistance genotyping prior to treatment initiation in treatment-naïve persons help to identify mutations associated with HIV drug resistance. In this study deep sequencing (NGS) was utilized to identify minority drug resistant variants in treatment-naïve population, that can consequently lead to potentially treatment failure in future. Drug-resistant HIV are archived in viral reservoirs and may persist as minority variants when outgrown by wild-type strains, in the absence of sufficient drug pressure in treatment-naïve patients.

From this study, no major drug resistance mutations against protease (PI) was detected in both majority and minority variant, however polymorphism leading to secondary (accessory) mutations were detected at drug binding sites at frequency range of >1 to <20 (Table 3.5.1). The detected PI secondary mutations pose little effect on drug susceptibility; however, a phenotypic change in any of them could have relevance to the affinity to one or more PIs. These mutations, in combination with PI major resistance mutations, might have an effect, apparently, they need to be strictly monitored to prevent them from becoming major mutations. Most of those occurred in almost every sample in

the study (**M36I**, **L89M, I93L and L63V**) were located in polymorphic positions observed in non-B subtypes

No major mutations were detected against NRTI in reverse transcriptase region (RT) of all isolates. On the other hand, one mutation (8,3%) against NNRTI was detected (**V179D**) in sample PK089D with low weight in the Tiotec ETR genotypic susceptibility score. V179D is a polymorphic accessory NNRTI-selected mutation that contributes low-levels reductions in susceptibility to each of the NNRTIs. The combination of V179D with K103R act synergistically to reduce NVP and EFV susceptibility. V179D appears similar to NNRTI mutation V179E in its acts on NNRTIs. V179D/E do not appear to reduce the virologic response to a first-line EFV containing regimen.

The detection of secondary PI mutations and NNRTI mutation at low frequency threshold of <20% with <1% cut off clearly indicates that minority population carry variants which show the presence of drug resistance to the various classes of ARVs. With such results being displayed, it indicates that HIV resistance genotypic testing is important prior to treatment initiation for identification of variants harboring resistant, to a better treatment outcome in patient. Furthermore, NGS proves that basic population-based sequencing method is more unlikely to pick the minority variant as they appear in much lower frequency, so NGS is the recommended genotype resistant assay. Several studies have reported on the efficacy of NGS for the inference of drug resistance, due to its constant ability to detect minor variants responsible for drug resistance (Ji et al., 2011; Nanfack et al., 2017). NGS wide sequencing approach offers the opportunity of cost effectiveness, especially in resource-limited countries, to sequence a multitude of samples in one array.

Phylogenetic analysis of all 12 isolates from drug-naïve patients showed 11 samples (91.7%) clustering with Subtype C virus from South Africa and other countries. This findings is in agreement with other studies previously done in Limpopo and other provinces in South Africa (Bessong et al., 2005; Bessong and Nwobegahay, 2012; Iweriebor et al., 2012). Subtype C viruses are predominant worldwide and in Southern countries such as Botswana, Mozambique and Zimbabwe. Further, one isolate was

subtype B; 1 (8.3%) and two isolates from samples PK092 and PK012 showed a very low probability of recombination with various subtype like C, K, J, D, B and CRF-AE respectively (Figures 12,14,15, and 16). Since these CRFs have not been reported in the region and only the partial pol was analysed a full length genomic analysis is required to reveal its complete nature. The detection of Subtype B virus is interesting since the virus is known to be dominate in Europe and America.

4.2 Limitations of the study

This study applied a state-of-the art technology (NGS) to infer drug resistance mutations. However, the number of sequences for individuals available for analyses was relative small. This makes it difficult to draw a wider interpretation on the prevalence of drug resistant mutation in the chronically infected persons.

4.3 Conclusion

In conclusion, the current study has established a next-generation sequencing protocol to investigate genetic drug resistance mutations. The applied protocol did not find major resistance mutations against PI, NRTI and NNRTI in the drug-naïve study subjects. However secondary (accessory) and unusual mutations detected on multiple drug target sites at low frequency threshold might confer resistance when complied with major mutation. It appears, drug resistance genotyping may not be required prior to treatment initiation. The study also identified HIV subtype B and some low probability recombinant forms in the study population. This is an interesting finding since C-viruses have been shown to dominate the epidemic in northern South Africa.

Commented [PB1]: You mentioned in the results section that one sample had V179D, so why do you say there was no major mutation.

References

Commented [PB2]: Ensure that the references are all consistent.

Alcantara, L.C.J., Cassol, S., Libin, P., Deforche, K., Pybus, O.G., Van Ranst, M., Galvao-Castro, B., Vandamme, A.M. and De Oliveira, T., 2009. A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. *Nucleic acids research*, *37*(suppl_2), pp. W634-W642.

Ansorge, W.J., 2009. Next-generation DNA sequencing techniques. *New Biotechnology*, 25(4), pp.195-203.

Aralaguppe, S.G., Siddik, A.B., Manickam, A., Ambikan, A.T., Kumar, M.M., Fernandes, S.J., Amogne, W., Bangaruswamy, D.K., Hanna, L.E., Sonnerborg, A. and Neogi, U., 2016. Multiplexed next-generation sequencing and de novo assembly to obtain near full-length HIV-1 genome from plasma virus. *Journal of Virological Methods*, 236, pp.98-104.

Ariën, K.K., Abraha, A., Quinones-Mateu, M.E., Kestens, L., Vanham, G. and Arts, E.J., 2005. The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. *Journal of Virology*, *79*(14), pp.8979-8990.

Barin, F., Denis, F., Allan, J.S., M'Boup, S., Kanki, P., Lee, T.H. and Essex, M., 1985. Serological evidence for virus related to simian T-lymphotropic retrovirus III in residents of West Africa. *The Lancet*, *326*, pp.1387-1389

Barre-Sinoussi, F., Chermann 1-C., Rey F., Nugeyre M.T., Chamaret S., Gruest J., Dauguet C., Axler-Blin C., Vénizet-Brun F., Rouzioux C., Rozenbaum W., and Montagnier L., 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, *220*, pp.868-871.

Baxter J., Dunn D., White E., Sharma S., Geretti A., Kozal M., and Lundgren J., 2015. Global HIV-1 transmitted drug resistance in the INSIGHT Strategic Timing of Antiretroviral Treatment (START) trial. *HIV Medicine*; 16, pp.77–87.

Bennett, D.E., Myatt, M., Bertagnolio, S., Sutherland, D. and Gilks, C.F., 2008. Recommendations for surveillance of transmitted HIV drug resistance in countries scaling up antiretroviral treatment. *Antiviral Therapy*, *13*, pp.25.

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

Bessong, P.O., Mphahlele, J., Choge, I.A., Obi, L.C., Morris, L., Hammarskjold, M.L. and Rekosh, D.M., 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 Research and Human Retroviruses*, *22*(12), pp.1306-1312.

Blood, G.A.C., 2016. Human Immunodeficiency Virus (HIV). *Transfusion Medicine and Hemotherapy*, *43*(3), p.203.

Brown, T.A. Molecular Phylogenetics. In *Genomes*, 2nd ed.; Garland Science: New York, NY, USA, 2002.

Busch M.P., and Sattern G.A., 1997. Time course of viremia and antibody seroconversion following human immunodeficiency virus exposure. T*he American Journal of Medicine*; 102:117124

Brass, A.L., Dykxhoorn, D.M., Benita, Y., Yan, N., Engelman, A., Xavier, R.J., Lieberman, J. and Elledge, S.J., 2008. Identification of host proteins required for HIV infection through a functional genomic screen. *Sciences*, 319, pp. 921–926.

Brenchley, J.M., Schacker, T.W., Ruff, L.E., Price, D.A., Taylor, J.H., Beilman, G.J., Nguyen, P.L., Khoruts, A., Larson, M., Haase, A.T. and Douek, D.C., 2004. CD4+ T cell

depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *Journal of Experimental Medicine*, *200*(6), pp.749-759.

Chakrabarti, L., Guyader, M., Alizon, M., Daniel, M.D., Desrosiers, R.C., Tiollais, P. and Sonigo, P., 1987. Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. *Nature*, *328*(6130), pp.543-547.

Clavel, F., 1986. Isolation of a new human retrovirus from West African patients with AIDS. *Science*, *233*, pp.343-347.

Clavel, F. and Hance, A.J., 2004. HIV drug resistance. *New England Journal of Medicine*, *350*(10), pp.1023-1035.

Clavel, F. and Mammano, F., 2010. Role of Gag in HIV resistance to protease inhibitors. *Viruses,* 2 (7), pp.1141-1426.

Coffin, J.M., Hughes, S.H. and Varmus, H.E., 1997. Retroviruses. *Plainview.*

Connick, E., 2005. Incomplete antibody evolution and seroreversion after treatment of primary HIV type 1 infection: what is the clinical significance? *Clinical Infectious Diseases*, *40*(6), pp.874-875.

Copper, E.R., Charura,t M., Mofenson, L., Hanson, I.C., Pitt, J., Diaz, C., Hayani, K., Handelsman, E., Smeriglio, V., Hoff, R. and Blattner, W., 2002. Combination antiretroviral strategies for the treatment of pregnant HIV-1- infected women and prevention of perinatal HIV-1 transmission. *Journal of Acquired Immune Deficiency Syndrome*. 29, pp. 484-494.

Dallar, R. and Karim, Q.A., 2015. HIV/AIDS, food insecurity, and undernourishment: amplifying cycles of risk in vulnerable populations. In *Handbook of public health in natural*

disasters: Nutrition, food, remediation and preparation. pp. 273-280. Wageningen Academic Publishers.

De Jong, M.D., Veenstra, J., Stilianakis, N.I., Schuurman, R., Lange, J.M., De Boer, R.J. and Boucher, C.A., 1996. Host-parasite dynamics and outgrowth of virus containing a single K70R amino acid change in reverse transcriptase are responsible for the loss of human immunodeficiency virus type 1 RNA load suppression by zidovudine. *Proceedings of the National Academy of Sciences*, *93*(11), pp.5501-5506.

de Silva, T.I., Cotten, M. and Rowland-Jones, S.L., 2008. HIV-2: the forgotten AIDS virus. *Trends in Microbiology*, *16*(12), pp.588-595.

Department of Health (2013). *National Antiretroviral Treatment Guidelines.* South Africa.

Daiz, R.S., Accetturi, C.A. and Sucupira M.C., 2008. Emergence of resistance mutations preceding virologic failure in patients receiving antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndrome,* 49, PP.111-113.

Dowell, K., 2008. Molecular Phylogenetics An Introduction To Computational Methods And Tools For Analyzing Evolutionary Relationships. *Math*, *500*, p.18.

Elena, S.F. and Sanjuán, R., 2005. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *Journal of virology*, *79*(18), pp.11555-11558.

Frentz, D., Boucher C.A.B., and Van DeVijver D.A.M.C., 2012. Temporal Changes in the Epidemiology of Transmission of Drug-Resistant HIV-1 across the World. *AIDS Reviews*; 14: 17–27.

Friedman-Kien, A.E., Laubenstein, L., Marmor, M., Hymes, K., Green, J., Ragaz, A., Gottleib, J., Muggia, F., Demopoulos, R. and Weintraub, M., 1981. Kaposis sarcoma and

Pneumocystis pneumonia among homosexual men--New York City and California. *MMWR. Morbidity and Mortality Weekly Report*, *30*(25), pp.305-8.

Flynn, W.F., Chang, M.W., Tan, Z., Oliveira, G., Yuan, J., Okulicz, J.F., Torbett, B.E. and Levy, R.M., 2015. Deep sequencing of protease inhibitor resistant HIV patient isolates reveals patterns of correlated mutations in Gag and protease. *PLoS Computational Bbiology*, 11(4), p.e1004249.

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

Gallo, R.C., Salahuddin, S.Z., Popovic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Palker, T.J., Redfield, R., Oleske, J., Safai, B. and White, G., 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science*, *224*(4648), pp.500-503.

Ganser-Pornillos, B.K., Yeager, M. and Sundquist, W.I., 2008. The structural biology of HIV assembly. *Current Opinion in Structural Biology*, 18(2), pp.203-217.

Geretti, A.M., 2007. Epidemiology of antiretroviral drug resistance in drug-naive persons. *Current Opinion in Infectious Diseases*, 20(1), pp.22-32.

Gorodin, S., Unal, S., Wang, Y., Mikhaylov, M.I., Bigbulatova, L. and Jehuda-Cohen, T., 2013. New tools in HCV diagnosis, in light of the enhanced awareness and the new drugs for treatment: SMARTube and stimmunology. *The Scientific World Journal*, 2013.

Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L., Daly, J.M., Mumford, J.A. and Holmes, E.C., 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, *303*(5656), pp.327-332.

Gupta, R.M., Sahni, A.K., Jena, J. and Nema, S.K., 2005. Genomic diversity of human immunodeficiency viruses. *Medical Journal Armed Forces India*, 61(3), pp.267-270.

Gupta R.K., Jordan R.M., Sultan B.J., Hill A., Davis D.H.J., Gregson J., Sawyer A.W., Hammers R.L., Ndembi N., Pillay D., and Bertagnolio S., 2011. Global trends in antiretroviral resistance in treatment-naïve individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis. *Lancet;* 380: pp.1250-1258.

Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. and Alizon, M., 1987. Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature*, *326*(6114), pp.662.

Hahn, B.H., Shaw, G.M., De, K.M. and Sharp, P.M., 2000. AIDS as a zoonosis: scientific and public health implications. *Science*, *287*(5453), pp.607-614.

Hamers, R.L., Sigaloff, K.C., Kityo, C., Mugyenyi, P. and de Wit, T.F.R., 2013. Emerging HIV-1 drug resistance after roll-out of antiretroviral therapy in sub-Saharan Africa. *Current Opinion in HIV and AIDS*, *8*(1), pp.19-26.

Hammer, S.M., Katzenstein, D.A., Hughes, M.D., Gundacker, H., Schooley, R.T., Haubrich, R.H., Henry, W.K., Lederman, M.M., Phair, J.P., Niu, M. and Hirsch, M.S., 1996. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, *335*(15), pp.1081-1090.

Holmes, E.C., 2009. The evolution and emergence of RNA viruses. *Oxford University Press*. New York. pp. 131-135.

Holmes, E.C. and Grenfell, B.T., 2009. Discovering the phylodynamics of RNA viruses. *PLoS Computational Biology*, *5*(10), p.e1000505.

Ishikawa, K., Janssens, W., Banor, J.S., Shinno, T., Piedade, J., Sata, T., Ampofo, W.K., Brandful, J.A., Koyanagi, Y., Yamamoto, N. and Canas-Ferreira, W.F., 2001. Genetic analysis of HIV type 2 from Ghana and Guinea-Bissau, West Africa. *AIDS Research and Human Retroviruses*, *17*(17), pp.1661-1663.

Iweriebor, B.C., Obi, C.L., Akinyemi, O., Ramalivhana, N.J., Hattori, T. and Okoh, A.I., 2012. Uropathogens isolated from HIV-infected patients from Limpopo Province, South Africa. *African Journal of Biotechnology*, *11*(46), pp.10598-10604.

Ji, H., Li, Y., Graham, M., Liang, B.B., Pilon, R., Tyson, S., Peters, G., Tyler, S., Merks, H., Bertagnolio, S. and Soto-Ramirez, L., 2011. Next-generation sequencing of dried blood spot specimens: a novel approach to HIV drug-resistance surveillance. *Antiviral Therapy*, *16*(6), p.871.

Kanki, Phyllis J., Souleymane M'boup, Dominique Ricard, Francis Barin, Francois Denis, Chiekh Boye, Lasana Sangare et al. "Human T-lymphotropic virus type 4 and the human immunodeficiency virus in West Africa." *Science* 236 (1987): 827-832.

Kiepiela P., Manasa J., Moosa M. Y., Moodley P., Gordon M., Parikh U. M., Sunpath H., de Oliveira T., and Ramjee G., 2014. HIV drug resistance patterns at the epicenter of the HIV-1 epidemic in Kwazulu-Natal, South Africa 2003-2013. *Journal of AIDS & Clinical Research;* 5: pp.2155-6113

Konstant, T.L., Rangasami, J., Stacey, M.J., Stewart, M.L. and Nogoduka, C., 2015. Estimating the number of sex workers in South Africa: Rapid population size estimation. *AIDS and Behavior*, *19*(1), pp.3-15.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B.H., Wolinsky, S. and Bhattacharya, T., 2000. Timing the ancestor of the HIV-1 pandemic strains. *Science*, *288*(5472), pp.1789-1796.

Kuiken C., Leitner T., Hahn B., Mullins J., Wolinsky S., Foley B., Apetrei C., Mizrachi I., Rambaut A., Korber B., 2012. HIV Sequence Compendium. Los Alamos, NM*: Los Alamos National Lab*;*www.hiv.lanl.gov*.

Larder, B.A., Darby, G. and Richman, D.D., 1989. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. *Science*, *243*(4899), pp.1731-1735.

Lu, K., Heng, X. and Summers, M.F., 2011. Structural determinants and mechanism of HIV-1 genome packaging. *Journal of Molecular Biology*, *410*(4), pp.609-633.

Levy, J.A., 2011. Virus-host interactions in HIV pathogenesis: directions for therapy. *Advances in Dental Research*, *23*(1), pp.13-18.

Manasa, J., Katzenstein, D., Cassol, S., Newell, M.L. and de Oliveira, for the Southern Africa Treatment and Resistance Network, T., 2012. Primary drug resistance in South Africa: data from 10 years of surveys. *AIDS Research and Human Retroviruses*, 28(6), pp.558-565.

Manasa, J., Danaviah, S., Pillay, S., Padayachee, P., Mthiyane, H., Mkhize, C., Lessells, R.J., Seebregts, C., de Wit, T.F.R., Viljoen, J. and Katzenstein, D., 2014. An affordable HIV-1 drug resistance monitoring method for resource limited settings. *Journal of Visualized Experiments*: JoVE, (85).

Markosyan, R.M., Cohen, F.S. and Melikyan, G.B., 2003. HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation. *Molecular Biology of the Cell*, 14(3), pp.926-938.

Martinez-Mariño, B., Foster, H., Hao, Y. and Levy, J.A., 2007. Differential gene expression in CD8+ cells from HIV-1-infected subjects showing suppression of HIV replication. *Virology*, 362(1), pp.217-225.

Meintjes, G., Maartens, G., Boulle, A., Conradie, F., Goemaere, E., Hefer, E., Johnson, D., Mathe, M., Moosa, Y., Osih, R. and Rossouw, T., 2017. Guidelines for antiretroviral therapy in adults: guidelines. *Southern African Journal of HIV Medicine*, 13(3), pp.114-133.

Massyn N., Peer N., Padarath A., Barron P., Day C., editors. 2015. District Health Barometer 2014/15. Durban: *Health Systems Trust*.

Menéndez-Arias, L., 2008. Mechanisms of resistance to nucleoside analogue inhibitors of HIV-1 reverse transcriptase. *Virus Research*, 134(1-2), pp.124-146.

Moss, A.R., Bacchetti, P., Osmond, D., Krampf, W., Chaisson, R.E., Stites, D., Wilber, J., Allain, J.P. and Carlson, J., 1988. Seropositivity for HIV and the development of AIDS or AIDS related condition: three year follow up of the San Francisco General Hospital cohort. *British Medical Journal of Clinical Research*, *296*(6624), pp.745-750.

Nanfack, A.J., Redd, A.D., Bimela, J.S., Ncham, G., Achem, E., Banin, A.N., Kirkpatrick, A.R., Porcella, S.F., Agyingi, L.A., Meli, J. and Colizzi, V., 2017. Multimethod Longitudinal HIV Drug Resistance Analysis in ART Naïve Patients. *Journal of Clinical Microbiology*, pp.JCM-00634.

Nwobegahay, J.M., Bessong, P.O., Masebe, T.M., Mavhandu, L.G., Iweriebor, B.C. and Selabe, G., 2011. Prevalence of Antiretroviral Drug Resistance Mutations and HIV-1 Subtypes among Newly-diagnosed Drugnaïve Persons Visiting A Voluntary Testing and Counselling Centre in Northeastern South Africa. *Journal of Health, Population and Nutrition*, pp.303-309

Number of people receiving antiviral therapy (ART) and percentage of all people living with HIV receiving ART in low- and middle-income countries overall and by WHO region, 2013. Accessed: February 28, 2017: http://www.who.int/hiv/data/artmap2014.png?ua=1.

Nwobegahay, J., Selabe, G., Ndjeka, N.O., Manhaeve, C. and Bessong, P.O., 2012. Low prevalence of transmitted genetic drug resistance in a cohort of HIV infected naïve patients entering antiretroviral treatment programs at two sites in northern South Africa. *Journal of Medical Virology*, 84(12), pp.1839-1843.

Parboosing R., Naidoo A., Gordon M., Taylor M., and Vella V., 2011. Resistance to antiretroviral drugs in newly diagnosed, young treatment-naïve HIV-positive pregnant women in the province of KwaZulu-Natal, South Africa. *Journal of Medical Virology* 83: pp. 1508-1513.

Plantier, J.C., Leoz, M., Dickerson, J.E., De Oliveira, F., Cordonnier, F., Lemée, V., Damond, F., Robertson, D.L. and Simon, F., 2009. A new human immunodeficiency virus derived from gorillas. *Nature Medicine*, *15*(8), p.871.

Peeters, M., Gueye, A., Mboup, S., Bibollet-Ruche, F., Ekaza, E., Mulanga, C., Ouedrago, R., Gandji, R., Mpele, P., Dibanga, G. and Koumare, B., 1997. Geographical distribution of HIV-1 group O viruses in Africa. *AIDS*, *11*(4), pp.493-498.

Quiñones-Mateu, M.E., Avila, S., Reyes-Teran, G. and Martinez, M.A., 2014. Deep sequencing: becoming a critical tool in clinical virology. *Journal of Clinical Virology*, 61(1), pp.9-19.

Rajarapu, G., 2013. Genes and Genome of HIV-1. *Journal of Phylogenetics & Evolutionary Biology*, pp.1-7.

Ren, J. and Stammers, D.K., 2008. Structural basis for drug resistance mechanism for non-nucleoside inhibitors of HIV reverse transcriptase. *Virus Research.* 134, pp.157-70.

Sauter, D., Unterweger, D., Vogl, M., Usmani, S.M., Heigele, A., Kluge, S.F., Hermkes, E., Moll, M., Barker, E., Peeters, M. and Learn, G.H., 2012. Human tetherin exerts strong

selection pressure on the HIV-1 group N Vpu protein. *PLoS Pathogens*, *8*(12), p.e1003093.

Sharp, P.M. and Hahn, B.H., 2011. Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspectives in Medicine*, *1*(1), p.a006841.

Shiino, T., 2012. Phylodynamic analysis of a viral infection network. *Genomics and Computational Science for Virus Research*, pp.162.

Sierra, S., Kupfer, B. and Kaiser, R., 2005. Basics of the virology of HIV-1 and its replication. *Journal of Clinical Virology*, 34(4), pp.233-244.

Smith, A.D., Tapsoba, P., Peshu, N., Sanders, E.J. and Jaffe, H.W., 2009. Men who have sex with men and HIV/AIDS in sub-Saharan Africa*. The Lancet*, 374(9687), pp.416-422.

Steinhauer, D.A., Domingo, E. and Holland, J.J., 1992. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene*, *122*(2), pp.281-288.

Stuyer, L. and Rombout, A., 1997. Line probe assay for raped detection of drug-selected mutations in the HIV-1 reverse transcriptase gene. *Journal of Antimicrobial Agents and Chemotherapy* 41, pp. 284-91.

Sheet, U.G.S.F., 2016. UNAIDS [online] http://www. unaids. org/sites/default/files/media_asset. UNAIDS_FactSheet_en. pdf.

UNAIDS (2017) fact sheet: HIV/AIDS. www.unaids.org/sites/default/files/media_asset/20170720_Data_book_2017_en.pdf

UNAIDS (2015) Global fact sheet: HIV/AIDS. www.unaids.org/sites/default/files/media_asset/20150901_FactSheet_2015_en.pdf

Global report: UNAIDS report on the global AIDS epidemic 2013 www.unaids.org/sites/default/files/media.../UNAIDS_Global_Report_2013_en_1.pdf
.

Vallari, A., Bodelle, P., Ngansop, C., Makamche, F., Ndembi, N., Mbanya, D., Kaptué, L., Gürtler, L.G., McArthur, C.P., Devare, S.G. and Brennan, C.A., 2010. Four new HIV-1 group N isolates from Cameroon: Prevalence continues to be low. *AIDS Research and Human Retroviruses*, *26*(1), pp.109-115.

Vallari, A., Holzmayer, V., Harris, B., Yamaguchi, J., Ngansop, C., Makamche, F., Mbanya, D., Kaptué, L., Ndembi, N., Gürtler, L. and Devare, S., 2011. Confirmation of putative HIV-1 group P in Cameroon. *Journal of Virology*, *85*(3), pp.1403-1407.

Vicenzi, E. and Poli, G., 2013. Novel factors interfering with human immunodeficiency virus-type 1 replication in vivo and in vitro. *HLA*, *81*(2), pp.61-71.

Verger, L., Malonga-Louellet, G., Mistou,l I., Mavoungou, R., Mansaray, H. and Peeters, M., 2002. Resistance to antiretroviral in Gabon. Guidelines on antiretroviral therapy and HIV-1 drug resistance monitoring in developing countries. *Journal on Acquired Immune Deficiency Syndrome*, **29,** 165-8.

Volz, E.M., Koelle, K. and Bedford, T., 2013. Viral phylodynamics. *PLoS Computational Biology*, *9*(3), p.e1002947.

World Health Organization, 2013. Global update on HIV treatment 2013: results, impact and opportunities.

WHO (2015) Gap report.

WHO (2016) Gap report.

WHO guidelines: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, *9*(7), pp.409-417.

Worobey, M., Gemmel, M., Teuwen, D.E., Haselkorn, T., Kunstman, K., Bunce, M., Muyembe, J.J., Kabongo, J.M.M., Kalengayi, R.M., Van Marck, E. and Gilbert, M.T.P., 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature*, *455*(7213), pp.661-664.

Zagordi, O., Klein, R., Däumer, M. and Beerenwinkel, N., 2010. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids Research*, 38(21), pp.7400-7409.

Zhang, L., Yu, W., He, T., Yu, J., Caffrey, R.E., Dalmasso, E.A., Fu, S., Pham, T., Mei, J., Ho, J.J. and Zhang, W., 2002. Contribution of human α-defensin 1, 2, and 3 to the anti-HIV-1 activity of CD8 antiviral factor. *Science*, *298*(5595), pp.995-1000.

Zhang, M., Schultz, A.K., Calef, C., Kuiken, C., Leitner, T., Korber, B., Morgenstern, B. and Stanke, M., 2006. jpHMM at GOBICS: a web server to detect genomic recombinations in HIV-1. *Nucleic Acids Research*, 34(suppl_2), pp. W463-W465.

https://en.wikipedia.org/wiki/HIV#/media/File:Hiv-timecourse_copy.svg

Appendix

Appendix A

University of Venda

UNIVERSITY OF VENDA
DEPARTMENT OF MICROBIOLOGY
HIV/AIDS & Global Health Research Programme

RESEARCH CONSENT FORM AND QUESTIONNAIRE

Research Topic: Studies on HIV/AIDS in northern South Africa: transmitted and acquired drug resistance, adherence to treatment, oncoviruses, blood pathogens, and host genetics

Principal Investigator: Professor Pascal O. Bessong
Tel: 015 962 8301; 073 798 5920;
Email: bessong@univen.ac.za

Investigators: Dr Lufuno G. Mavhandu
Tel: 015 962 8962, 071 2933 842
Email: lufuno.mavhandu@univen.ac.za

Dr. Denis Tebit
Tel: +1 216 8324757;
Email: dmt6u@virginia.edu

You are being asked to volunteer as a subject in a research study. This form is designed to provide you with information about this study that you should know and understand, as well as to answer any questions that you may have.

THE PURPOSE OF THE RESEARCH
The HIV/AIDS & Global Health Research Program, University of Venda, is investigating HIV biology, HIV drug resistance and related cancer associated viruses, in the Limpopo Province. The research team will be looking at antiretroviral treatment outcomes, the development and transmission of resistant viruses in adults and children, and the prevalence, risk factors and genetic variants of cancer associated viruses such as HBV, HHV8 and HPV.

In order to understand how viruses develop resistance once treatment is initiated, persons who are about to start antiretroviral treatment will be recruited and followed up for a maximum of 36 months while they are under ARV treatment. To find out if resistant viruses are being transmitted persons going for voluntary testing and counselling will be assessed.

THE FOLLOWING PROCEDURES WILL BE INVOLVED.

Acquired HIV drug resistance:

1. You will complete a questionnaire regarding your age, sex, marital status, probable year and place of infection.
2. 5 ml of venous blood or a needle prick (50 µl) will be collected from you by a qualified health professional at the time when you present yourself at an ARV treatment hospital or clinic **to initiate ARV treatment**. Virus will be isolated from the blood and the genetic resistance profile studied. At the time of blood collection, CD4+ cell count and viral load measurements will be done.

Version 1, August 2015

Appendix A

3. While you are on ARV, 5ml of blood will be taken and your CD4 count and viral load will be measured every 3 months for a maximum of 36 months. For children, head circumference, weight and length (height) will be measured every month. At each CD4+ cell count and viral load measurement the current ART regimen and use of complementary medicine will be noted, the resistance profile of the virus will be determined. The samples collected will also be studied for cancer related viruses.
4. About 20mg of hair samples will also be collected (from the back of your head) prior to treatment initiation (baseline) and every 6 months for a maximum period of 36 months. The hair samples will be studied to monitor adherence to ARVs.
5. A questionnaire will be administered focusing on adherence behavior, stigma, mental health and substance use once every 3 months for 36 months. Additionally, face to face audio-recorded interview lasting 10-15 minutes will be conducted to gather information about your personal experiences and behavior outcomes.

Oncoviruses:

For the detection of cancer viruses, a once-off mouthwash rinse or saliva and cervical cells will be collected from you by a qualified health professional.

1. For the detection of Kaposi Sarcoma-associated Herpes Virus (KSHV), about 10ml of mouthwash rinse or saliva will be collected from you. You will rinse your mouth thoroughly for 30 seconds, then spit into a 50ml nalgene collection tube provided.
2. If you cannot use mouthwash due to medical conditions, you are requested to spit about 5ml of saliva into the 50ml nalgene collection tube.
   It is essential that you do not eat or smoke before collection of mouthwash rinse or saliva.
3. For the detection of Human Papilloma Virus (HPV), cervical cells will be collected from you into a 4.3mL cervical specimen transport tube.

POTENTIAL RISKS

The potential risk associated with 5 ml of blood collection or needle prick is pain and abrasion at the site of collection. Risk associated with mouthwash and cervical cells collection will be pain or soreness during mouthwash rinse when there are blisters and body discomfort, respectively. No other risks are foreseen.

USE OF RESEARCH RESULTS AND POTENTIAL BENEFITS

Your participation in this study will enable us determine whether HIV resistant variants are being transmitted; and whether individuals failing first and second line treatments are infected with viruses bearing resistant mutants. This information may be useful in understanding the impact HAART rollout in Limpopo Province and the prevalence of other related viruses. In general, the data that will be generated may impact policies on the management of HIV/AIDS.

There are no potential patient benefits, however, the results may provide knowledge of specific resistant mutations that may be present in viruses circulating in HAART subjects. The direct benefits from this research for participants may be improved knowledge on the biology of HIV, periodic CD4 counts and viral load assessments, and the understanding that you are contributing in the generation of information required for a better clinical management of HIV/AIDS patients. The results obtained may be published. The specimen and data generated in this study such as age, sex, CD4 counts, viral load measurements, and resistance mutations may be used for future virologic and genetic studies.

Version 1, August 2015

Appendix A

CONFIDENTIALITY

Should you consent to participate in this research project your identity will be kept confidential. Your coded sample will be used for research purposes without your personal identifiers. As such, results analysis will deal only with demographic, viral, immunologic and clinical data and no identifiable markers linked to your person. Consequently, the results obtained may be published but your identity will not be revealed.

GENERAL CONSIDERATIONS

1. Should you agree to participate in this research, you may change your mind at any time without prejudice. Refusal to continue participation shall not be objected to.
2. This study has been approved by the Health, Safety and Research Ethics Committee of the University of Venda, and permission has been obtained from the Provincial Department of Health, Polokwane.

YOUR RIGHTS AS A RESEARCH SUBJECT

I have read or have had read/interpreted to me all of the above. The investigator/interviewer has explained the study to me and answered all my questions. I have been told of the potential risks or discomfort and possible benefits of the study.

I understand that I do not have to take part in this study, and if I agree to participate I may withdraw from the study at any time.

The results from this study may be published, but my identity will not be revealed. I understand my rights as a research subject. I understand what the study is about and why it is being done, and I voluntarily agree to participate in this study.

Name and signature or thumb print of subject/Parent/ legal guardian _____

Date _____ Research code _____

Name of Interpreter (if required) _____Signature _____Date _____

Name of investigator / Interviewer _____Signature _____ Date_____

Version 1, August 2015

Appendix A

University of Venda

UNIVERSITY OF VENDA

Department of Microbiology
HIV/AIDS & Global Health Research Programme

Research Topic: Studies on HIV/AIDS in northern South Africa: transmitted and acquired drug resistance, adherence to treatment, oncoviruses, blood pathogens, and host genetics

Principal Investigator: Professor Pascal O. Bessong
Tel: 015 962 8301; 073 798 5920
Email: bessong@univen.ac.za)

Investigators: Dr Lufuno G. Mavhandu
Tel: 015 962 8962, 071 2933 842
Email: lufuno.mavhandu@univen.ac.za

Dr. Denis Tebit
Tel: +1 216 8324757
Email: dmt6u@virginia.edu

RESEARCH QUESTIONNAIRE

Research code: _____

Sex: _____

Age or date of birth: _____

Marital status: _____

Highest level education: _____

Occupation: _____

Income level (Rand/month) : <R3, 000☐; R3, 000 -10,000 ☐ >R10, 000 ☐

Probable place of infection: _____

Probably year of infection: _____

Risk factor for infection: _____

Current ARV regimen: _____

Start date of treatment: _____

Viral load and CD4+ cell count at treatment initiation: _____

Version 1, August 2015

Appendix A

Other diagnosis at start of treatment (TB etc): _____

Head circumference (for children, in cm): _____

Weight (for children, in kg): _____

Height or length (for children, in cm): _____

WHO stage of AIDS: _____

Use of complementary medicine: _____

Oncoviruses

Do you smoke? _____

Age at first sexual intercourse _____

How many sexual partners do you have? _____

Is your sexual partner male/female? _____

Have you ever been vaccinated against HPV? _____

How often do you go for pap smear testing? _____

Comments: _____

Name of health facility: _____

Version 1, August 2015