Recombination events and epitope prediction in HIV-1 strains from Southwest Cameroon

UNIVERSITY OF VENDA
CALL No.
ACC No.

By

Bixa O. Ogola

Student Number: 11530098

Dissertation submitted in fulfilment of the requirements for the award of Master of Science degree in Microbiology

To the Department of Microbiology,

School of Mathematical and Natural Sciences

University of Venda, South Africa

UNIVERSITY OF VENDA
LIBRARY

Supervisor: Prof. Pascal O. Bessong (University of Venda)

Co-supervisor: Dr. Denis M. Tebit (University of Virginia)

April 2016

Acknowledgements

I would like to express my earnest appreciation to the following people:

My supervisor, Professor Pascal Obong Bessong, Head of the HIV/AIDS and Global Health Research Programme, University of Venda, whose invaluable advice, technical expertise, insights and exceptional scientific knowledge made this work possible. His endearing patience, fortitude and charity made it possible for me to see the toiling worthwhile. There are no words for appreciation for the fruits of his labor.

My co-supervisor, Denis Manga Tebit (Assistant Professor of Research at Myles Thaler Centre for AIDS and Human Retrovirus Research, Department of Microbiology, Immunology and Cancer Biology, University of Virginia, USA), whose dedicated support, camaraderie, perceptions and advice I could not have done without.

I would also like to pass my appreciation to Doctor Helen Kimbi, and Emaculate Lum for sample collection and data capturing. I would also like to extend my gratitude to Professors Marie-Louise Hammarskjold and David Rekosh (Co-Directors of the Myles H. Thaler Center for AIDS and Human Retrovirus Research, University of Virginia) for their invaluable assistance in bridging the knowledge gap that I had to overcome to do this study.

My mother, Joan Akoth Ogola and my father, Professor Emeritus Jason Samuel Ogola, whose sturdy determination and faith have brought about my life's successes. Their diligence and hard work has taught me a lot about the gifts of labor. Their advice and tireless aid has been the wings on which I soar. My brothers and sister, Gerbera Yonga Ogola, Nouvelle Magero Ogola, Zavel Yves Ogola and Loh Aouta Ogola, for their understanding and support. They inspired me to do better, and their encouragement made it possible to persevere.

Thanks to my seniors within the HIV/AIDS and Global Health Research Programme, Dr. Lufuno Mavhandu, Dr. Tracy Masebe, Dr. Maina Gachara and fellow colleagues namely: Ms. Tshifhiwa T, Ms. Matume D, Ms. Tshivhula H, Ms. Etta E., Mr. Magoro T, Ms. Munzhedzi M., Mr. Alayande D.P., Ms. Hlungwane V.M. and Ms. Rikhotso R.R. for their support and motivation with dealing with the grinding and grueling of research work.

Lastly, to the Research and Publications Committee of the University of Venda, South Africa, whose funding partly aided in meeting the research costs.

Declaration

I, **Bixa Okech Ogola**, hereby declare that this dissertation for the award of a Master of Science degree in Microbiology at the University of Venda, hereby submitted by me has not been previously submitted for any degree examination at this or any other University. It is my own work in design and in accomplishment and that all reference materials contained herein have been duly acknowledged.

Signature..... Date. 15/AUGUST/2016

Bixa Okech Ogola

Dedication

I dedicate this work to my father, Professor Emeritus Jason Samuel Ogola.

Abstract

BACKGROUND: HIV has a heterodiploid genome and reverse transcriptase (RT) that has recombinogenic propeties which promotes high diversity. This diversity is illustrated by the types, groups, subtypes and growing number of circulating recombinant viruses. Viral diversity may impact on viral fitness, diagnosis, disease progression, therapy management, as well as vaccine design. Cameroon is known to be a "hotspot" for high HIV diversity with a predominance of the CRF_02 AG strain. This study was conducted on viruses from Southwest Region of Cameroon with the aim of describing novel recombinant viruses. Doing this could reveal information on the complexity of the identified unclassified strains as well as predict what impact they may have on the course of managing them.

METHOD: The study setting was the Mutengene Baptist hospital in the Southwest Region of Cameroon. Study subjects included individuals from different locales in the Southwest and Littoral Regions of Cameroon. Approval was obtained from the Ethics Committee of the Cameroon Baptist Health Board (Cameroon) and the Research Ethics Committee of the University of Venda, South Africa. Whole blood was collected from 107 HIV-1 infected individuals of whom 83 were drug naïve and 24 were drug experienced. Continuous partial p17 and partial p24 subgenomic regions was PCR amplified from proviral DNA, sequenced using the Sanger protocol, genotyped using MEGA and analyzed for recombination patterns using RIP and jpHMM. CTL epitope prediction was done using NETCTLpan and MHC-1 tools using HLA alleles with the highest frequencies among the coastal people of Cameroon. Virus BM189 was selected, due to its unique recombination pattern, for near-full length genome amplification, followed by next generation sequencing, and genotyped using MEGA and analysed for recombination using, Simplot, RIP and jpHMM. Drug resistance profiles of viral populations at >20%, <5% and <1% were determined using different interpretation algorithms. CTL epitopes within *gag* and *env* genes were predicted.

RESULTS: Fifty two of 107 specimens were successfully amplified and 39 were successful sequenced. The majority of the 39 sequences (74.4%) were subtype A-related strains (n=29/39), 12.8% were CRF01_AE (n=5/39), 7.7% were subtype G-related (n=3/39) and 5.1%

were sub-subtype F2-related (n=2/39). From the 39 *gag* sequences, it was observed that most of subtypes A related, G related and CRF01_AE related sequences were possible recombinants based on the high amino acid variation to what was expected from their pure subtypes/CRF. Twenty of the 39 samples had partial *pol* and *env* sequences available from previous analysis and analysis of the subgenomes revealed that the dominant strain (50.0%) was CRF02_AG (n=10/20), 35.0% were URFs (n=7/20), 10.0% were CRF22_01A1 (n=2/20) and 5.0% were subsubtype F2 (n=1/20). One of the URFs, Virus BM189, which showed subtype F2 for partial *gag* and partial *pol* and subtype A1 for partial *env* analysis was selected and succesfully amplified for near-full length genome analysis. The virus was obtained from a 28 year old unmarried female with residence in Mutengene, with no history of anti-retroviral treatment, and classified as AIDS stage 3. The near full length sequence (9077 nucleotides) was determined to be a second generation recombinant of subtypes F2/CRF01_AE/A1/F1. It was made up of mostly sub-subtype F2, CRF01_AE within 407 nucleotides of gp120 of *env*, sub-subtype A1 for rest of *env* and F1 for most of *nef*.

Drug resistance profile for BM189 showed that viral populations >20% within the quasispecies were susceptible to all known anti-retroviral drugs. Nevertheless, both <5% and <1% viral populations had M184V mutation which is associated with high level resistance to Emtricitabine (FTC) and Lamivudine (3TC), and K103N associated with high level resistance to Efavirenz (EFV) and Nevirapine (NVP). The majority population (>20%) were predicted to be a R5 tropic virus with susceptibility to entry inhibitors, whereas the minority populations had G36S mutation which is associated with resistance to the entry inhibitor Enfuvirtide (INN).

Partial *gag* epitope prediction showed numerous epitopes that were good binders for the HLAs used. P24 epitope TSTLQEQIGW was predicted with a 20% frequency among CRF01_AE related sequences and was restricted by HLA-B*58:01, which has been associated with slow disease progression. P17 epitopes WPFRNSRRM and YLIQQPSIVY, restricted by HLA-B*35:01, was predicted within CRF01_AE related sequences. This allele has been associated with rapid disease progression. BM189 *gag* and *env* epitope prediction showed epitopes that were strong binders for HLA-B*58:01. *Env* epitopes (HSFTCGGEFFY and FAILKCNDAEF) were predicted to be strong binders of HLA-B*35:01 and gag epitope (MYSPVSIL) was predicted to be strong binders for HLA-C*04:01. HLA-B*35:01 and HLA-C*04:01 have been

associated with rapid disease progression. HLA typing was not done on the study participants, hence the presence of these epitopes cannot adequeately tell the nature of disease progression. HLA typing on the participants is recommended to determine if these associations held true for the study participants. Furthermore, other phenotypic studies such as interferon-gamma enzyme-linked immunospot (IFN-γ ELISPOT) need to be done to determine if these predictions will hold true.

In conclusion, the findings from this study supports the on-going genetic diversification of HIV in Cameroon; and that recombinants viruses are dominating the epidemic. Furthermore, it highlights the importance of next-generation sequencing as a genotypic tool for drug resistance testing for minority populations. Numerous epitopes were predicted, which would serve in future phenotypic studies.

Keywords: HIV, recombination events, cytotoxic T-cell epitopes, Southwest Cameroon.

Table of contents

List of figures

List of tables

List of abbreviations and symbols

(-)	Minus or antisense nucleic acid strand
-	Symbol for to (For example: C2 to C3) and symbol for dash (For example: L – 100 bp)
%	Percentage
~ or ≈	approximately
(+)	Plus or sense nucleic acid strand
+	Plus symbol (For example: CD4 plus) or And symbol (For example: DNA and RNA hybrid)
<	Less than
>	Greater than
±	Plus or minus
°C	Degree Celsius
µl	Microliter
µM	Micromolar
10^{-4}	Ten to the exponent of negative four
10^{9}	Ten to the exponent of 9
3'	Three prime
3TC	Lamivudine
5'	Five prime
aa	Amino acid
AIDS	Acquired Immunodeficiency Syndrome
ANN	Artificial neural network
APOBEC3G	Apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3G
ART	Antiretroviral treatment
ARV	Antiretroviral
AZT	Zidovudine
bDNA	branched chain DNA
BLAST	Basic Local Alignment Search Tool
BM	Baptist Mutengene (code used for samples based on the site they were collected - Mutengene Baptist Hospital)

Bp	Base pair
C2	Second constant region of gp120
C3	Third conserved region of gp120
CA	Capsid protein
CCR1	C-C motif chemokine receptor type 1
CCR2b	C-C motif chemokine receptor type 2b
CCR3	C-C motif chemokine receptor type 3
CCR5	C-C motif chemokine receptor type 5
CCR8	C-C motif chemokine receptor type 8
CD4	Cluster of differentiation 4
CD8	Cluster of differentiation 8
CDC	Centre for Disease Control and prevention
cDNA	Complementary DNA
Cells/mm^3	Cells per cubic millitre of blood
Comblib	Combinatorial peptide libraries
cPPT	Central polypurine tract
CRF	Circulating Recombinant Forms
C-terminal	Carboxyl terminal
CTL	Cytotoxic T lymphocytes
CXCR4	C-X-C motif chemokine receptor type 4
ddH_20	double distilled water
DIS	Dimer initiation signal
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease (any enzyme that degrades DNA)
dNTP	Deoxynucleotide triphosphate
dNTPs A, C,G, T	Adenine Cytosine Guanine, Thymidine
ds-DNA	Double stranded-DNA
Dual tropic	viral strains that use CXCR4 and CCR5 co-receptors for entry into cells
Expected (E) value	A parameter that describes the number of hits one can "expect" to see by chance when searching a database of a particular size

e.g.	Example
EDTA	Ethylene diamine tetra-acetic acid
EFV	Efavirenz
ELISA	Enzyme-Linked Immnunosorbent Assay
env	Envelope gene
ER	Endoplasmic reticulum
Et al.	Et alia (and others)
FPR	False positive rate
FTC	Emtricitabine
g	Gram
gag	Group specific antigen gene
gp	Glycoprotein
Gp120	120 kilodalton glycoprotein (surface glycoprotein)
Gp41	41 kilodalton glycoprotein (transmembrane protein)
Group M	Main group
Group N	Non-M and non-O group
Group O	Outlier group
Group P	Putative group
H_20	Water
HIV	Human immunodeficiency virus
HIV-1	Human immunodeficiency virus type-1
HIV-2	Human immunodeficiency virus type-2
HLA	Human leukocyte antigen
IC_{50}	Half maximal inhibitory concentration (indirect indicator of binding affinity)
ICAM1	Intercellular adhesion molecule 1
IEDB	Immune epitope database
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IN	Integrase enzyme
jpHMM	Jumping profile Hidden Markov Model

kb	kilobase
Lab	Laboratory
LANL	Los Alamos national laboratory
LTR	Long terminal repeat
M	Molar
MA	Matrix protein
MEGA	Molecular evolutionary genetic analysis
mer	Part or a specified number of subunits
mg	milligram
$MgCl_2$	Magnesium chloride
MHC-1	Major histocompatablity complex type 1
ml	Milliliter
ml	milliliter
mM	millimolar
mm^3	cubic millimeter
Monomer	Single unit of a protein
n	Number
NA or N/A	Not available
NASBA	Nucleic acid sequence based amplification
NAT	Nucleic acid testing
NC	Nucleocapsid protein
NCBI	National center of biotechnology information
nef	Negative factor gene
NGS	Next generation sequencing
N-linked	Amino linked
nm	Nanometers
nm	Nanometre
nM	Nanomolar
NNRTI	Non-nucleoside analogue Reverse Transcriptase Inhibitor
NRTI	Nucleoside analogue Reverse Transcriptase Inhibitor

NSI	Non-syncytium inducing
nt	Nucleotide
N-terminal	Amino terminal
NVP	Nevirapine
P17	17 kilodalton protein (monomer of matrix protein)
P24	24 kilodaton protein (monomer of capsid proteins)
P31	31 kilodalton protein (monomer of integrase protein)
P51	51 kilodalton protein (portion of reverse transcriptase heterodimer with RNase H activity)
P6	6 kilodalton protein (portion of nucleocapsid protein)
P66	66 kilodalton protein (portion of reverse transcriptase heterodimer with polymerase activity)
P7	7 kilodaton protein (monomer of nucleocapsid protein)
PBS	Primer binding site
PCR	Polymerase Chain Reaction
PI	Protease inhibitor
Pol II	Polymerase 2 (Human RNA polymerase 2)
Pol	Polymerase enzyme
PR	Protease enzyme
Pr160gagpol	160 kilodalton precursor gagpol polyprotein
Pr55gag	55 kilodalton precursor gag polyprotein
PSSM	Position specific scoring matrices
QC	Quality Control
R5 tropic	viral strains that only use CCR5 co-receptor for entry into cells
Rev	Regulator of expression of viral protein
RIP	Recombination identification program
RNA	Ribonucleic acid
RNase	Ribonuclease (nucleases that catalyzes the degradation of RNA)
RNase H	Ribonuclease H
RT	Reverse transcriptase enzyme
RT-PCR	Real-time polymerase chain reaction

SI	Syncytium inducing
SIV	Simian immunodeficiency virus
SIVcpz	Simian immunodeficiency virus found in chimpanzee
SIVgsm	Simian immunodeficiency virus found in greater spot-nosed monkeys
SIVrcm	Simian immunodeficiency virus found in red capped mangabeys
SIVsmm	Simian immunodeficiency virus found in Sooty mangabey
SMM	Stabilized matrix method
SU	Surface unit (surface glycoprotein)
SW	Southwest
TAE	Tris acetate EDTA
TAP	Transporter associated with antigen processing
Taq	Thermus aquaticus
Tat	Transcriptional transactivator
TDF	Tenofovir
TM	Transmembrane protein
tRNA	Transfer ribonucleic acid
U	Units
U/μL	Units per microloter
U5	Unique 5 region
UNAIDS	Joint United Nations Programme on HIV and AIDS
URF	Unique recombinant forms
USA	United State of America
UV	Ultraviolet
v	volts
V3 loop	Loop formed at third variable region (of gp120)
V5	fifth variable region of gp120
Vif	Virion infectivity factor
Vpr	Viral protein R
Vpu	Viral protein U
Vpx	Viral protein X

UNIVERSITY OF VENDA
LIBRARY

WHO	World Health Organisation
X4 tropic	viral strains that only use CXCR4 co-receptor for entry into cells
ZTE	zoonotic transmission events

List of amino acid symbols

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

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

1.1. Introduction

Human immunodeficiency virus (HIV) the etiologic agent for aquired immunodeficency syndrome (AIDS) (Barré-Sinoussi et al., 1983; Gallo et al., 1983) is responsible for 25.3 million deaths globally since 2000 with an estimated 36.9 million living with the virus as of June 2015 (UNAIDS, 2015).

Since the earliest documented reports of AIDS in 1981 in the United States of America (CDC, 1981), there has been the discovery of two types of HIV, namely HIV-1 and HIV-2 as well as subtypes, sub-subtypes and recombinant forms (Robertson et al., 1999). The pandemic seems to have stabilised as incidence rates are falling across regions of the world. This is due, in large part, to HIV prevention campaigns and the scale-up in treatment in developing countries with the largest HIV burden (as reviewed by Lau and Wong, 2013).

Despite the low prevalence rates in Cameroon (4.8%), the country has the highest diversity with all known types, groups, subtypes and recombinants being reported (Tebit and Arts, 2011; UNAIDS, 2015). Recombination plays a vital role in the high degree of viral diversity and also a significant role in immune response, therapy, diagnosis, as well as vaccine development. Therefore, the arising recombinant forms of the virus have to be detected and classified.

1.2 Literature review

1.2.1 Morphology of HIV-1

HIV is a spherically shaped enveloped retrovirus of about 120 nanometers in diameter (Figure 1.1). The envelope, that is acquired from the host's cell membrane, is studded with approximately 8-10 trimeric viral envelope (*Env*) proteins. Each *env* protein is made up of a trimer of exposed surface glycoprotein (SU/gp120), anchored to the virus by a trimer of transmembrane proteins (TM/gp41) (Zhu et al., 2003). The SU and TM are bound through non-covalent interactions. A matrix shell (MA/p17) lines the inner surface of the envelope with a conical shaped capsid core (CA/p24) making up the center of the virus where two plus strand genomic RNA are housed as a stabilized ribonucleoprotein complex with a nucleocapsid protein (NC/p7) (as reviewed by Hutchinson, 2001).

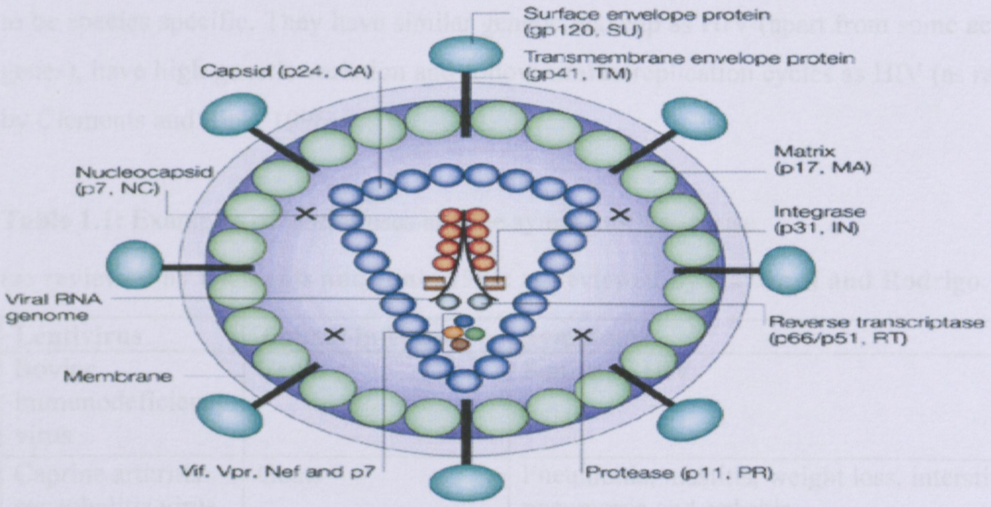


Figure 1.1: HIV-1 mature virion

Monomers that make up capsid, nucleocapsid, matrix, surface envelope proteins, transmembrane envelope proteins, reverse transcriptase and protease are shown in brackets. The captialised letters within the brackets indicate abbreviations for the proteins. For example CA is an abbreviation for capsid (Robinson, 2002)

Three virally encoded functional proteins are found within the capsid: namely, protease (PR); reverse transcriptase (RT); and integrase (IN). Furthermore, three accessory viral proteins: Negative Factor (*nef*), Viral Infectivity Factor (*vif*) and viral protein U (*vpu*) are present within

the virus. The accesory proteins are not essential, however their expression aids in the virus virulence (as reviewed by Turner and Summers, 1999; as reviewed by Briggs and Kräusslich, 2011). There are several host proteins that may be present on the viral envelope such as Intercellular Adhesion Molecule 1 (ICAM 1). Other host proteins may be present within a mature virion such as Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G (APOBEC3G), ubiquitin, actin, human leukocyte antigens and cyclophilin A. Some of these proteins, like cyclophilin A, have been shown to be useful for viral activity but not essential, whereas others like APOBEC3G are inhibitory to viral activity (Stopak et al., 2003; as reviewed by Zhou et al., 2012).

1.2.2 HIV taxonomy

According to the International Committee on Taxonomy of Viruses (Serio et al., 2012), HIV falls under the *Retroviridae* family, subfamily Orthoretrovirinae and the lentivirus genus. Some of the known lentiviruses, with exclusion of HIV, are listed in Table 1.1. These viruses seem to be species specific. They have similar genetic makeup as HIV (apart from some accessory genes), have high genetic variation and follow similar replication cycles as HIV (as reviewed by Clements and Zink, 1996).

Table 1.1: Examples of lentiviruses and the symptoms they cause

(as reviewed by Clements and Zink, 1996; as reviewed by Hayward and Rodrigo, 2010)

Lentivirus	Animal infected	Symptoms
Bovine immunodeficiency virus	Cattle	Similar to HIV
Caprine arthritis encephalitis virus	Goats	Pneumonia, mastitis, weight loss, interstitial pneumonia and arthritis
Equine infectious anemia virus	Horses	Periodic episodes of fever, anemia, thrombocytopenia, leukopenia and weight loss.
Feline immunodeficiency virus	Domesticated cats and in some wild cats like lions and cougars	AIDS-like symptoms in domesticated cats but less pathogenic in non-domesticated cats like lions and cougars
Simian immunodeficiency virus	African monkeys	Similar to HIV and can cause AIDS-like illness in some species. However it can have longer incubation periods in many species
Visna-maedi virus	Sheep	Dyspnea, chronic active inflammation, interstitial pneumonitis (maedi), loss of weight, mastitis and arthritis

1.2.3 HIV-1 replication cycle

The replication cycle of HIV-1 can be divided into two phases (Figure 1.2). The early phase is the infection stage which involves entry into the host cell and integration of the viral cDNA into the host's genome (as reviewed by Nisole and Saïb, 2004). The late phase involves the expression of the integrated provirus to yield complete mature virion (Freed and Martin, 2001).

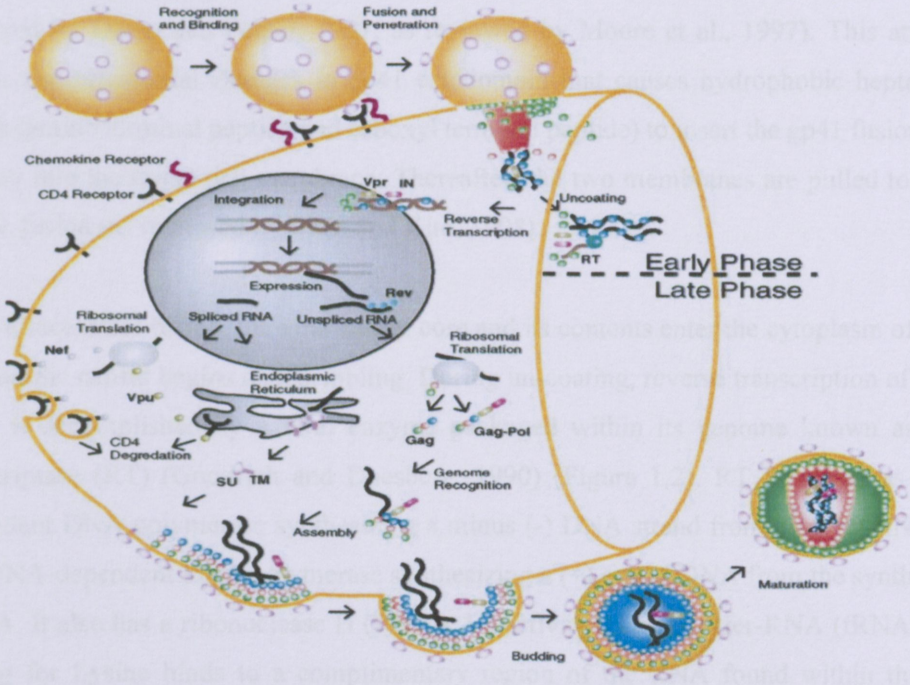


Figure 1.2: Steps in the HIV-1 replication cycle

The densely small orange "blobs" at the top illustrate viral particle involved in binding, fusion and penetration. The various stages involved in production of progeny virions are shown within the host cell (the large "egg shaped" structure with orange outline) (as reviewed by Turner and Summers, 1999)

1.2.3.1 Binding, membrane fusion, reverse transcription and integration

The gp120 protein binds to CD4 receptor commonly found on CD4+ T lymphocytes, macrophages and dendritic cells of the host. The interaction between viral gp120 and CD4 receptors depends on chance proximity and mainly anchors the virus to the host cell (as reviewed by Freed, 2002). The high affinity between the host CD4 receptor and the viral gp120 is primarily due to the interaction of the conserved domains (CD3 and CD4) within the gp120 and the N-terminal extracellular domain of the CD4 receptor. The CD4/gp120 interaction leads

to a conformational change in the structure of the gp120 resulting in increased affinity for V3 loop within the gp120 to interact with the seven transmembrane domain chemokine receptor (commonly CCR5 or CXCR4) on the host's CD4 cell. Differences in the V3 sequences determine whether a particular gp120 will interact with CCR5 or CXCR4 receptor. The binding of gp120 to the co-receptor leads to a stable two-pronged attachment (as reviewed by Bour et al., 1995; as reviewed by Doms and Peiper, 1997). Other chemokine receptors like CCR1, CCR3, CCR5, CCR2b and CCR8 can act as co-receptors for a sub-set of viral strains (as reviewed by Doms and Peiper, 1997; as reviewed by Moore et al., 1997). This attachment causes conformational changes in gp41 ectodomain that causes hydrophobic heptad repeat motifs (amino terminal peptide and caboxyl terminal peptide) to insert the gp41 fusion peptide directly into the target cell membrane. Thereafter, the two membranes are pulled together to induce fusion (as reviewed by Chan and Kim, 1998).

After a successful fusion, the viral capsid core and its contents enter the cytoplasm of the host cell and the capsid begins disassembling. During un-coating, reverse transcription of the viral RNA is accomplished by a viral enzyme packaged within its genome known as reverse transcriptase (RT) (Goodrich and Duesberg, 1990) (Figure 1.2). RT functions as a RNA-dependent DNA polymerase synthesizing a minus (-) DNA strand from plus (+) RNA strand and DNA-dependent DNA polymerase synthesizing a (+) strand DNA from the synthesized (-) DNA. It also has a ribonuclease H (RNAse H) activity. Host transfer-RNA (tRNA) usually coding for Lysine binds to a complimentary region of the RNA found within the Primer Binding Site (PBS) within the 3' Long Terminal Repeat (LTR). RT adds nucleotides to the primer (tRNA) in a complimentary manner in the 5' to 3' direction whereas RNAse H cleaves the phosphodiester bonds of the genomic RNA within the DNA+RNA hybrid at about 18 base pairs (bp) upstream of the RNA (Hurwitz and Leis, 1972; Basu et al., 2008).

Once a complimentary DNA (cDNA) also known as a minus strand strong stop DNA, has been synthesized to the 3' LTR end, RT makes a template jump (known as first strand transfer) to the 5' end of the genomic RNA. The generated (-) strand DNA binds its complimentary R region within the 5' LTR and reverse transcription continues. This strand transfer event can happen on the same (+) RNA that was being transcribed or on the other (+) RNA of the viral genome. Upon complete transcription of the (-) cDNA, the complimentary (+) DNA strand is synthesized by using small fragments of RNA on the cDNA strand that is slightly resistant to

RNAse H cleavage known as the central polypurine tract (cPPT). RT transcribes the second DNA strand from the 5' end of RNA on the cPPT. RNAse H degrades the tRNA bound to the (–) DNA strand before a second strand transfer occurs, whereby the transcribed tRNA sequence on the (+) strand binds to the complimentary PBS on the (–) DNA strand and the (+) strand is transcribed to the 5' end of the negative strand. The end result is a complete double stranded viral DNA. (as reviewed by Temin, 1993; as reviewed by Pathak and Hu, 1997; as reviewed by Basu et al., 2008).

RT in comparison to human DNA polymerases, has a slow synthesis rate, is highly inaccurate, lacks proofreading mechanism, and has a low fidelity for the RNA template leading to an accumulation of mutations, which account for some of the virus' high diversity. Furthermore, RT's low fidelity causes it to jump from one RNA template to the other during reverse transcription (as reviewed by Preston et al., 1988; Freed, 2002). The reasons or models to explain how and why these template jumps occur are explained in section 1.2.10.

Entry of the viral DNA into the nucleus is followed by the integration of its DNA to the host's genome by the viral enzyme integrase, present within the viral pre-integration complex. Integration is a three step process. The first is 3' processing which occurs in the cytoplasm whereby integrase removes a dinucleotide from the 3' end of each LTR of the reverse transcribed DNA. The second step is strand transfer that occurs in the nucleus after nuclear import whereby integrase mediates a nucleophilic attack by the 3'-hydroxyl residues of the viral DNA on phosphodiester backbone of the target DNA, leading to a transesterification reaction. Lastly, the unpaired nucleotides are removed and the single strand gaps are repaired by host cell DNA repair mechanisms resulting in an integrated provirus (Engelman et al., 1991; as reviewed by Engelman and Cherepanov, 2008).

1.2.3.2 Transcription, translation, assembly, budding and maturation

In actively dividing cells, the provirus acts as a template for the synthesis of a large number of viral viral messenger RNA's (mRNA) (as reviewed by Joshi and Joshi, 1996; as reviewed by Pereira et al., 2000; as reviewed by Wu, 2004). The basal RNA Polymerase II transcriptional activity for the HIV LTR is very low due to the association of the enzyme with negative transcription elongation factors bound to proviral LTR. Productive and processive transcription is greatly increased with the presence of viral protein known as transcriptional transactivator

(*tat*) (as reviewed by Hetzer et al., 2005; as previewed by Rice, 2010). *Tat* binds at the trans-activation response region (TAR) bulge of the Long terminal repeat (LTR) on a nascent RNA transcript and recruits transcription elongation factors as well as other co-factors to form a complex that releases the RNA Pol II from pausing leading to promoter clearance and processive transcription (as reviewed by Wu, 2004; as reviewed by Ott et al., 2011).

Transcription generates a large number of viral pre-mRNA that undergo complex alternative splicing to end up producing more than 40 different spliced mRNA species. These species can be collectively grouped as completely spliced mRNAs (about 1.7 to 2.0 kilobases length); incompletely spliced mRNAs (about 5 kilobases in length) and unspliced mRNA (as reviewed by Freed, 2002; as reviewed by Karn and Stoltzfus, 2012). Nuclear export of most intron-containing viral mRNA is carried out by a viral protein known as regulator of expression of viral proteins (*rev*). *Rev* disrupts intron splicing and interact with nuclear export receptors to facilitate exit through the nuclear membrane (as reviewed by Yedavalli and Jeang, 2011). Translation of viral polycistronic mRNAs requires the virus to use the host's translation mechanisms including mechanisms like leaky scanning for expressing *env* and *vpu* and programmed ribosomal frame shifting for expressing *pol* (Biswas et al., 2004; as reviewed by Bolinger and Boris-Lawrie, 2009).

Assembly, in large part involves the action of the major structural protein called *gag* poly-protein (Pr55*Gag*) (Ono et al., 2000). The matrix domain within Pr55*Gag* is co-translationally modified at its N-terminus with myristic acid which targets and binds to the plasma membrane by the insertion of the myristoyl group into the lipid membrane. This binding is further stabilized by *gag* multimerisation (Figure 1.2). Genomic RNA is assembled by complementary base-pairing at a site known as the dimer initiation signal (DIS) to form a dimer structure between the two single stranded genomic RNA. Thereafter, the nucleocaspid (NC) domain of Pr55*Gag* binds and randomly coats the viral RNA to package and further stabilize the RNA structure to its most energetically favorable conformation (Lodmell et al., 2001; Nikolaitchik et al., 2013). *Gag-gag* oligomerization occurs simultaneously during RNA/NC binding via the C-terminal of the capsid domain of Pr55*Gag*. Stability of the RNA dimer is completely achieved during maturation. *Vif* has chaperone activity and may be involved in the formation of RNA dimer as it has binding sites at the 5' untranslated region (as reviewed by Bukrinskaya, 2004).

HIV-1 *env* precursor glycoprotein (gp160) is translated by the ribosomes studded on the rough endoplasmic reticulum (ER) and co-translationally inserted into the ER through the N-terminal ER signal peptide located in gp41 portion of gp160. The gp120 domain is heavily glycosylated with N-linked oligosaccharide side chains in the ER (as reviewed by Checkley et al., 2011). Three monomers of Gp160 form intramolecular disulfide bonds between them before they are transported to the Golgi complex where the side chains undergo complex modifications. To inhibit interaction between newly synthesized gp160 and newly synthesized CD4 receptors, viral protein *vpu* downregulates newly synthesized CD4 receptors by binding its C-terminal to the cytoplasmic tail of CD4. *Vpu* transports the bound CD4 to be degraded possibly by cytoplasmic proteasomes (as reviewed by Frankel and Young, 1998; as reviewed by Murakami, 2012). Gp160 is cleaved by cellular furin-like proteases into a mature gp120 and gp41 prior to being transported to the cell surface through the host secretory pathway. Once at the cell membrane, gp41 is non-covalently bound to gp120 and rapidly internalized by host cell recognition of the endocytosis motif within gp41 cytoplasmic tail. *Env* is then incorporated into the virion possibly through interaction with MA (as reviewed by Checkley et al., 2011).

Specific viral sequences called late domains promote budding by interacting with the host cellular ubiquitination and endosomal sorting machinery. The interaction results in viral membrane fission and release (as reviewed by Bieniasz, 2006; as reviewed by Pincetic and Leis, 2009). Upon release, the immature non-infectious viral particle rearranges its structure to produce a mature infectious virion. This is largely driven by viral protease (PR) cleaving the Pr55*Gag* and Pr160*GagPol* polyprotein precursors to generate the mature MA, CA, NC, reverse transcriptase and integrase proteins. Protease protein is liberated from Pr160*GagPol* by auto-proteolysis. PR cleaves each site with a differing efficiency, as a result, PR-mediated *Gag* and *GagPol* processing takes place as an ordered, stepwise cascade of cleavage reactions (as reviewed by Freed, 2002; as reviewed by Ganser-Pornillos et al., 2008).

1.2.4 HIV genes and gene products

Table 1.2 lists all HIV-1 genes and the proteins they make. HIV-1 has a 9.1 kb RNA genome that has 9 genes expressing 15 different proteins. *Gag* and *env* genes expresses structural

proteins whereas *pol* expresses functional genes. *Tat, rev, vif, vpu, vpr* and *nef* express accesory proteins.

Table 1.2: HIV-1 proteins and functions

(as reviewed by Frankel and Young, 1998; as reviewed by Freed, 2002; as reviewed by HIV-LANL, 2008)

Genes	Gene Products	Functions of the product
Gag	MA/ P17	Interacts with host's plasma membrane and incorporates *env* glycoprotein during viral assembly
	CA/ P24	Forms viral shell and plays vital role in viral assembly
	NC/ P7	Interacts with unspliced RNA leading to effective encapsidation
	P6	Forms part of nucleocapsid and binds vpr
Pol	Protease	Autocatalytic cleavage of gag/pol and catalytic cleavage of viral proteins
	Reverse transcriptase	Viral enzyme that catalyzes viral cDNA formation from two single stranded RNA
	Integrase	Viral enzyme used for integration of viral cDNA into host's chromosome
Env	SU/ Gp120	External glycoprotein essential for viral attachment to host's receptors
	TM/ Gp41	Vital for viral and host membrane fusion enabling viral core penetration into host's cell
Tat	Tat	Binds to LTR and enables transcription of long and whole length viral transcripts
Rev	Rev	Protein that enables nuclear export of incompletely spliced viral mRNA
Vif	Vif	Inhibits RNA editing enzymes from acting on viral RNA during reverse transcription
Vpu	Vpu	Promotes viral release and degrades CD4
Vpr	Vpr	Induces arrest of mitotic cell division at G_2 phase in vpr expressing cells and translocate viral PIC into nucleus
Nef	Nef	Downregulates CD4 and MHC-1

1.2.5 Stages of HIV-1 Clinical infection and pathogenesis

World Health Organization guidelines state that there are four stages of HIV-1 infection (W.H.O., 2005). Clinical stage 1 or eclipse phase normally lasts for about a week to two weeks. During this phase, the virus replicates and spreads from the site of infection to tissues and organs with a large number of susceptible cells. Viremia is undetectable and there is no immune response or symptoms (as reviewed by Coffin and Swanstrom, 2013). Clinical stage 2 also known as the acute phase, lasts for about two to four weeks. Early in the stage, CD4+ T cells

LIBRARY

in the blood drop and viremia is high. Levels of serum antibodies to HIV proteins are usually not sufficiently elevated to permit detection of infection by enzyme-linked immunosorbent assay (ELISA) and immunoblot, but high levels of viral RNA are readily detectable in plasma (as reviewed by Lackner et al., 2012; as reviewed by Coffin and Swanstrom, 2013). Later in the stage, seroconversion occurs and viral replication is lowered as an immune response to the virus is mounted. There are usually no symptoms at this stage though certain individuals may experience acute sero-conversion illnessess which present as influenza-like illness such as fevers, malaise, diarrhea, swelling of lymph nodes, muscle and joint pain and headache. These symptoms are nonspecific and often attributed to nonspecific viral infections (Lackner et al., 2012). There is sufficient serum antibodies after seroconversion for detection by ELISA. Clinical stage 2 ends with a lowered viral load and a rebounding of CD4+ T cells as a stable equilibrium (also known as viral set point) of viral production and elimination is reached.

Clinical stage 3 also known as chronic or clinical latency phase, is characterised by a slow increase in levels of viremia and slow decrease in levels of CD4+ T cells. The shift in the stable equiblrium is due to an immune response that cannot clear the infection completely and large numbers of CD4+ T cells becoming infected and dying every day (as reviewed by Levy, 1993; as reviewed by Coffin and Swanstrom, 2013). This stage can last for about 1 to 20 years. At the beginning of the stage, the infected individual may be asymptomatic however, as the immune system deteriorates major symptoms and opportunistic infections become obvious. Clinical stage 4 or the AIDS stage occurs about 18 months after stage 3. It is characterised by a decline in CD4+ T cells to lower than 200 cells per millilitre of blood and complete failure to control the viral infection. Persistent and untreatable opportunistic infections like *Pneumocystis carinii pneumonia*, *Mycobacteria*, cytomegalovirus and *Cryptococcus* are more common and cancers like Kaposi's Sarcoma begin to appear as the immune system is completely deteriorated and antibiotics seem to be ineffective in curtailing infections. Without anti-retrovirals, death normally occurs within two years (as reviewed by Lackner et al., 2012; as reviewed by Coffin and Swanstrom, 2013).

1.2.6 Laboratory diagnosis of HIV-1

HIV-1 diagnosis relies on detecting the virologic or/and immunologic markers present in bodily fluids. Serologic diagnostic methods rely on detecting antibodies for a particular viral antigen (as reviewed by Iweala, 2004). A rapid test is a screening method that is cost effective,

easy and quick to perform and can be as accurate as the Enzyme linked immunosorbent assay (ELISA). This is due to its high specificity and sensitivity. Commercial rapid tests rely on conserved viral *env* gp41 immunodominant region proteins as epitopes for antibody binding. Current fourth generation kits can detect both anti-HIV IgG and anti-HIV IgM as well as the viral capsid protein p24 antigen detect viral loads exceeding 10,000 virions per millilitre of plasma. Hence the kits have reduced the diagnostic window to about a week or two after infection and offer 100% sensitivity and > 98% specificity (Ketema et al., 2001; as reviewed by Iweala, 2004; as reviewed by Shaw and Hunter, 2012). In contrast, the Nucleic acid testing (NAT), unlike rapid testing, is labour intensive, more costly and takes longer to perform. Real-Time Polymerase Chain Reaction (RT-PCR) assay is a method that is used to quantify viral DNA (as reviewed by Iweala, 2004). Viral genes from amplified HIV DNA can be sequenced to determine viral subtype as well as monitor resistance to anti-retrovirals prior and during therapy. Current HIV-1 NAT based on different amplification platforms, can detect viral sequences qualitatively at 1–5 copies per millilitre of plasma and quantitatively at levels exceeding 50 copies per millilitre plasma (as reviewed by Iweala, 2004; as reviewed by Shaw and Hunter, 2012).

1.2.7 HIV genetic nomenclature

There are two main types of HIV, namely HIV-1 and HIV-2. HIV-1 is responsible for the global pandemic and has four distinct lineages namely: Major (M) group, Outlier (O) group 'Non-M, Non O' (N) group and P group; each lineage arising due to independent cross species transmission events (Figure 1.3). HIV-2 has 8 distinct lineages group A, B, C, D, E, F, G and H; all endemic to West Africa (Robertson et al., 1999; as reviewed by Hemelaar, 2012). Within the four HIV-1 groups, Group M is the major participator in the pandemic and has 9 different subtypes, namely, A, B, C, D, F, G, H, J, K as well as 72 different circulating recombinant forms (CRFs) plus a number of unique recombinant forms (URFs) (as reviewed by Aldrich and Hemelaar, 2012; HIV-LANL, 2016). A strain is classified as a CRF when an intersubtype recombinant has been identified to have identical recombination breakpoints in three or more infected people who are not epidemiologically related, whereas URFs are recombinants found in an infected individual (Hemelaar et al., 2011; as reviewed by Santoro and Perno, 2013). There is genetic variation of sequences across subtypes depending on region of genome studied with a 9% difference across subtypes within *pol* gene (as reviewed by Santos and Soares, 2010). Within some of the subtypes are sub-subtypes. Subtype A has two sub-subtypes, namely, A1

and A2 and subtype F has two: F1 and F2 (Triques et al., 1999; Gao et al., 2001). HIV-1 Group M and N are closely related to the Simian Immunodeficiency Virus (SIV) infecting chimpanzees found in eastern Cameroon (Gao et al., 1999) whreas HIV-1 Group O and P are closely related to the Simian Immunodeficiency Virus (SIV) infecting gorillas (D'arc et al., 2015).

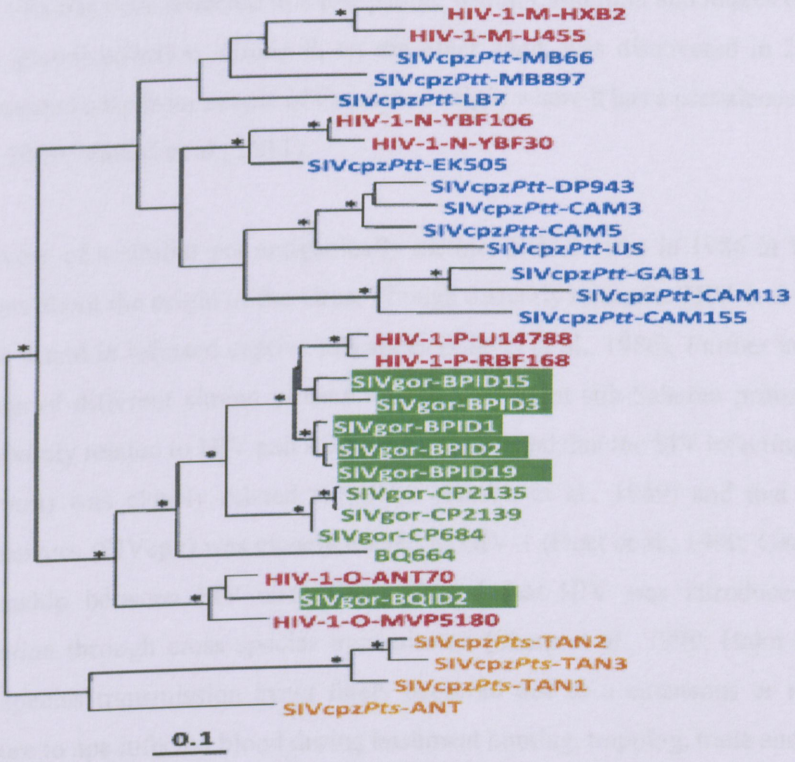


Figure 1.3: Maximum likelihood tree showing evolutionary relationships between HIV-1 and non-human primate retroviruses (SIV) based on partial *gag* gene

The tree shows HIV-1 group O (HIV-1-O-ANT70) forming a monophyletic clade with SIV infecting gorillas (SIVgorBQID2) whereas HIV-1 group P (HIV-1-P-U14788 and HIV-1-P-R8F168) is shown clustering with SIV infecting gorillas (SIVgor-BPID15). HIV-1 group M (HIV-1-M-HXB2) and group N (HIV-1-N-YBF106) are shown to form a clade with SIV infecting chimpanzees *Pan troglodytes troglodytes* (SIVcpzPtt-MB66 and SIVcpzPtt-EK505). Asterisks above branches correspond to nodes supported by bootstrap values over 70%. Newly identified SIVgor strains (highlighted in green boxes) are compared with previously characterized SIVgor (green), SIVcpzPtt (blue), SIVcpzPts (orange), and HIV-1 (red) strains. HIV-1 groups are represented in red color (D'arc et al., 2015)

1.2.8 HIV molecular epidemiology

Within HIV-1, group M is the most predominant and is responsible for up to 98.2% of the global infections (Sharp and Hahn, 2011). Group O, which was discovered in 1990 (Leys et al., 1990), is responsible for about 0.4% of the global infection and it is largely restricted to Cameroon, and its neighboring countries. Group N, which was discovered in 1998 (Simon et al., 1998), has been detected in a few people within Cameroon and makes up less than 0.001% of the global infection. Group P, on the other hand, was discovered in 2009, and has been documented only from people of Cameroon origin where it has a prevalence of 0.06% (Plantier et al., 2009; Vallari et al., 2011).

Discovery of a similar yet antigenically distinct HIV-2 virus in 1986 in West Africa fueled interests about the origin of the virus. Though distantly related to HIV-1, it was closely related to SIV found in infected captive macaques (Clavel et al., 1986). Further investigations led to findings of different simian viruses infecting different sub-Saharan primates. These viruses were closely related to HIV and it was later discovered that the SIV infecting sooty mangabeys (SIVsmm) was closely related to HIV-2 (Hirsch et al., 1989) and that the SIV infecting chimpanzees (SIVcpz) was closely related to HIV-1 (Huet et al., 1990; Gao et al., 1999). The relationship between SIV and HIV indicated that HIV was introduced into the human population through cross species transmission (Sharp et al., 1999; Hahn et al., 2000). The cross-species transmission event likely occurred due to a cutaneous or mucous membrane exposure to ape infected blood during bushmeat hunting, trapping, trade and keeping monkeys as pets (Hahn et al., 2000; Peeters et al., 2002; Wolfe et al., 2004).

Based on phylogenetic and statistical analysis, the common ancestor to group M is dated around 1910 to 1930, indicating that it spread for 50 to 70 years before being discovered. Blood and tissue samples collected from Kinshasa residents in 1959 and 1960 reveal that group M was present and had diversified into different subtypes by that time (Korber et al., 2000; Bailes et al., 2003; Worobey et al., 2008). Migration pathways of some subtypes and CRFs are known. A common hypothesis is that the virus spread along the Congo river in Zaire and began spreading globally. Subtypes A and D originated in central Africa and spread to establish epidemics in eastern Africa. Subtype C was introduced to and predominates in southern Africa where it spread to India and other Asian countries. Subtype B rose from a single African strain that spread to Haiti in the 1960s, before acquiring diversity and thereafter spread to United

States of America before spreading to Europe where the subtype is dominant. The recombination event that lead to the rise of CRF01_AE probably occurred in central Africa before it spread and became epidemic in southeastern Asia (as reviewed by Taylor et al., 2008; as reviewed by Sharp and Hahn, 2011; as reviewed by Lihana et al., 2012). Other recombinants seem to have become more prevalent through human migration and activity such as subtype A in Asia and subtype B in Europe recombined to form a new circulating recombinant form CRF03_AB within Russia and Eastern Europe. South America has seen a rise in BF recombinants despite subtype F having low prevelance globally or in central Africa (Taylor et al., 2008; Rubio et al., 2014).

Figure 1.4 shows the geographic distribution of subtypes and recombinants. It illustrates how HIV-1 subtypes are not randomly distributed globally but have a propensity for particular geographic distribution due to population bottleneck and founder effects. Despite this, there has been co-circluation of subtypes across geographic regions due to human migration which has resulted in recombinant forms of the virus arising and even becoming established regionally. About 48% of the people living with HIV-1 have Group M subtype C with a majority of them in southern Africa (Hemelaar et al., 2011). Only 11% of HIV-infected people globally have group M subtype B, the most studied subtype; 12% with subtype A, 5% for subtype G, 2% for subtype D. Subtypes F, H, J and K together caused fewer than 1% of infections worldwide. Recombinants (CRFs and URFs) are responsible for 20% of the global infections, with CRFs making 16% of global infections. There was a 50% increase in CRFs from 2004 to 2007 and a 39% decrease in URFs in the same time period. This growth is not only in numbers infected but with complexity as well (Hemelaar et al., 2011). An example of complexity is CRF11_cpx, a recombinant made up of subtypes A, G, J, and CRF01_AE (Montavon et al., 2002). The commonly known circulating recombinants are CRF01_AE (a recombinant virus made up of subtypes A and subtype E) in Southeast Asia with about 5% infected globally. The other is CRF02_AG (a recombinant virus made up of subtypes A and subtype G) in west central Africa and makes up 8% of global HIV-1 infections (Carr et al., 1996; Gao et al., 1996; Carr et al., 1998; Tebit et al., 2002).

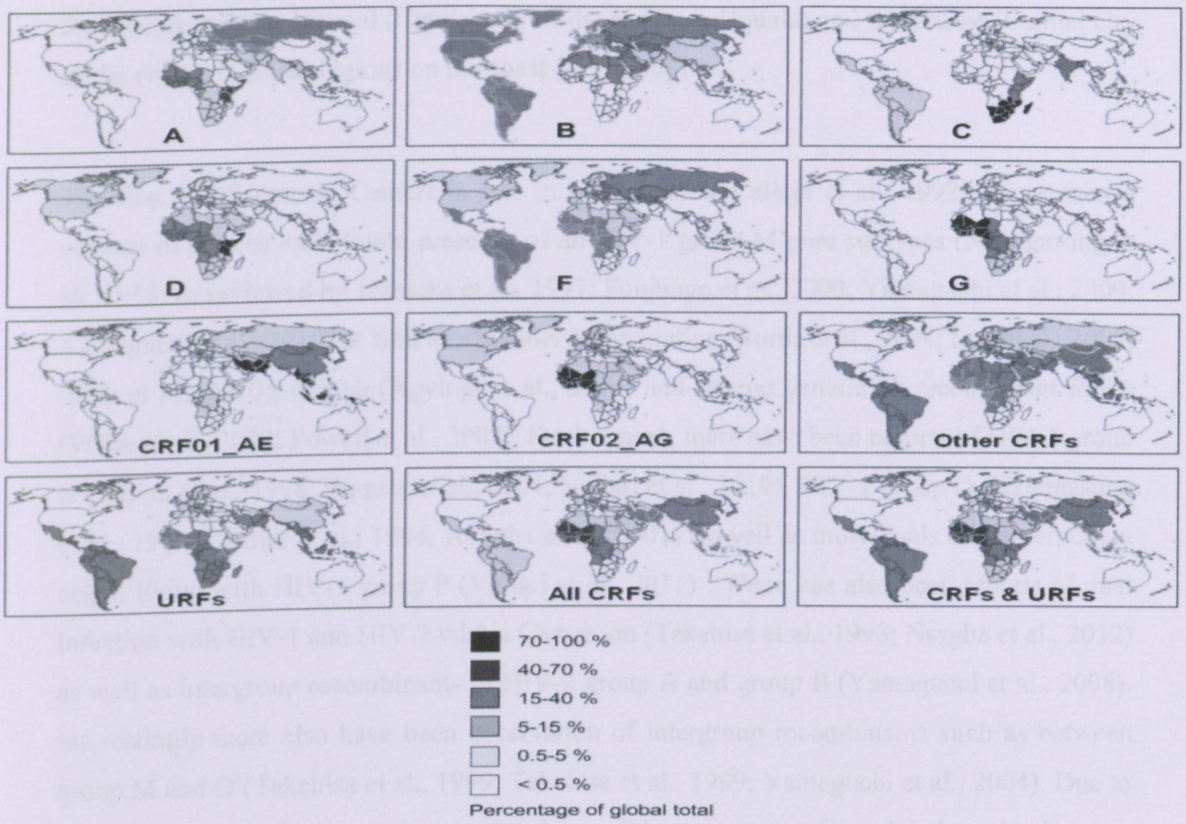


Figure 1.4: Global distribution of HIV-1 subtypes and recombinants

The figure shows nine grids with each grid showing particular subtype or CRF distribution globally in percentages. The left upper grid shows HIV-1 group M subtype A distribution globally. It shows that the highest percentage (70 to 100%) distribution of subtype A is within East Africa with a 40 to 70% distribution within China, Russia and Eastern Europe. The third grid from the top on the middle column shows that CRF02_AG is the dominant recombinant (70 to 100%) within west central Africa. West central Africa seems to have a high percentage (70 to 100%) of all CRFs and URFs as shown on the bottom grids on the middle and far right columns (as reviewed by Hemelaar, 2013)

1.2.9 HIV diversity in Cameroon

HIV diversity in Cameroon is the main focus of this study. Cameroon is a country located at the west coast of Africa, bordered by Nigeria to the west; Chad to the northeast; the Central African Republic to the east; Equatorial Guinea, Gabon, and the Republic of the Congo to the south. Its coastline lies on the Bight of Bonny, part of the Gulf of Guinea and the Atlantic Ocean. Cameroon is divided into 10 administrative regions with Yaoundé in Centre region as

the capital city and Littoral (Capital city of the region is Douala) and Southwest (Capital city of the region is Buea) regions on the coast (West, 2011).

The first AIDS case in Cameroon was in 1985 (Garcia-Calleja et al., 1992). Since then a number of studies have found presence of all HIV-1 group M pure subtypes (Nkengasong et al., 1994; as reviewed by Janssens et al., 1997; Fonjungo et al., 2000; Yamaguchi et al., 2009; Yamaguchi et al., 2010) as well as a number of circulating (Burda et al., 2004; Luk et al., 2007; Zhao et al., 2010), unique (Agyingi et al., 2014) and second generation recombinant forms (Wilbe et al., 2002; Powell et al., 2007). Furthermore, there have been reports of HIV-1 group N (Simon et al., 1998; Roques et al., 2004; Vallari et al., 2010), HIV-1 group O (Nkengasong et al., 1993; Gürtler et al., 1994; Ayouba et al., 2001) as well as individuals of Cameroonian origin living with HIV-1 group P (Vallari et al., 2011) . There has also been reports of dual infection with HIV-1 and HIV-2 within Cameroon (Takehisa et al., 1998; Nsagha et al., 2012) as well as intergroup recombinants of HIV-2 group A and group B (Yamaguchi et al., 2008). Interestingly there also have been observation of intergroup recombinants such as between group M and O (Takehisa et al., 1999; Takehisa et al., 1999; Yamaguchi et al., 2004). Due to the high diversity, Cameroon is considered one of the epicenters of the global pandemic.

The above studies show extensive HIV diversity within Cameroon (Heyndrickx et al., 2000) that has low prevalence compared to Southern African countries. It had the second highest HIV prevalence (4,8%) in 2015 within the West-Central African region after Democratic Republic of Congo (DRC) (6,2%) (UNAIDS, 2015). From sequence data analyses, it has been shown that the zoonotic transmission events (ZTE) that lead to the emergence of HIV-1 group M was a transmission of SIV infecting chimpanzees specie *Pan troglodytes troglodytes* (SIVcpzPtt), found in chimpanzees within southeastern Cameroon, to the human population (Keele et al., 2006). Furthermore HIV-1 group N have been linked to a separate ZTE of SIVcpzPtt from chimpanzee commonly found in south central Cameroon, to the human population (Keele et al., 2006). Whereas the ZTE that lead to emergence of HIV-1 Group O and group P was a transmission of SIV infecting gorillas found in southern Cameroon (D'arc et al., 2015). As a result, Cameroon is considered "ground zero" for HIV-1 group M because the zoonotic transmission to the human population of all the SIV that resulted to the epidemics of HIV-1 groups began in Cameroon. The beginning of HIV-1 group M and group O epidemics have been molecularly dated at the beginning of the 20th century and group N and P are dated much

later (Lemey et al., 2004; Worobey et al., 2008; as reviewed by Sharp and Hahn, 2011). This could partly explain why strains of HIV-1 group M and O are more diverse than those found of group N and P. It is likely that the availability of very few groups N and P sequences poses a challenge in dating these groups.

Based on numerous studies of rural and urban areas, CRF02_AG has been found to be the predominant strain in Cameroon (Carr et al., 2001; Nyambi et al., 2002; Burda et al., 2004). A study by Carr et al., (2010) studying HIV diversity at 14 remote locations in Cameroon found that a majority (66%) were CRF02_AG, 10,4% were URFs, 5.5% were CRF22_01A1 and the rest were a collection of 4 different HIV-1 group M subtypes (A2, D, F2, G) and 6 different CRFs (CRF-01, -11, -13, -18, -25, -37). Despite the predominance of CRF02_AG in Cameroon, it is believed to have originated from Democratic Republic of Congo and migrated to Cameroon and the rest of the region (Faria et al., 2012).

1.2.10 HIV-1 genetic diversity and recombination

There is a high sequence variability among HIV-1 genomes. Studies based on analysis of nucleotide sequences show that the mean percentage differences are: 48.3 % between HIV-1 and HIV-2, 37.5% difference among HIV-1 groups, 14,7% between HIV-1 group M subtypes (inter-subtype variation), 8.2% within each subtype of HIV-1 group M (intra-subtype variation) and 0.6% within a single individual's viral population. This correlated to a mean amino acid difference of 53.8% between types, 41.1% between HIV-1 groups, 18% between HIV-1 group M subtypes, 12.0% intra-subtype variation and 1.1% within an individual (Li et al., 2015). It is important to note that intra-subtype and inter-subtype variation can differ widely among specific subtypes and increases over time (as reviewed by Lau and Wong, 2013). Furthermore, there are amino acid differences for HIV proteins from individual strains of the same subtype as shown in Table 1. 3. The greatest difference is in *env* (gp120) and the least difference in *gag* (p24).

HIV genetic variability is due to a rapid viral turnover of up to 10^{10} viral particles per day in a newly infected drug-naïve individual. It has been observed that during the first year of infection, there is a linear increase in envelope (C2 –V5) viral diversity from the founder strain

(Ho et al., 1995; Shankarappa et al., 1999). This is due to the fact that each of the viral particle undergoes mutation because of the low fidelity of reverse transcriptase (RT), which results in nucleotide mis-incorporation errors during reverse transcription in an order of 10^{-4} per nucleotide per replication cycle (Preston et al., 1988; Santos and Soares, 2010). Before seroconversion, the infected individual produces a large mix of slightly different genotypic and phenotypic viral variants, that results in the viral population having some variants possessing unique phenotypes that can evade the host's defenses. This enables one or some variants to escape immune pressure upon seroconversion. In addition, RT can jump from one template to another during reverse transcription to make a single viral DNA from both of its positive (+) genomic RNA in a process known as recombination. Hence, if two genetically different HIV-1 strains are present, recombination can produce large genomic changes or even antigenic shifts. Hence, recombination is an integral part of HIV evolution.

Recombinant viruses arise when an individual is infected with two (dual infection) or more (super-infection) different HIV-1 clades. The subtypes can co-infect a target cell and each subtype undergoes the early replication stages, integration and expression of their proviral DNA within the cell independently of the other (Coffin, 1996). However, during viral assembly, one (+) genomic RNA strand from each viral subtypes can be co-packaged into a single virion to produce a heterodiploid or chimeric viral progeny. This chimeric or hybrid viral progeny buds out, matures and infect a new cell. In the newly infected cell, the chimeric virus can reverse transcribe its genomic RNA obtained from two different parental subtypes to produce a new recombined provirus (as reviewed by Temin, 1993; as reviewed by Quiñones-Mateu and Arts, 1999).

As discussed previously in section 1.2.3.1 during reverse transcription, RT tends to jump from one RNA template to the other during first and second strand transfer thereby switching templates. Furthermore, these template jumps do not occur only during strand transfers but can also occur when RT pauses during reverse transcription. Usually RT makes these template jumps at a minimum rate of 2.8 jumps per replication cycle. The more template jumps, the more complex the resulting chimeric virus (Hu and Temin, 1990; Zhuang et al., 2002).

Table 1. 3: Average amino acid diversity of viral proteins within individual HIV clades (%)

(Li et al., 2015)

HIV-1 Group M subtypes	MA	CA	NC	p6	PR	RT	IN	Vif	Vpr	Tat	Rev	Vpu/Vpx*	GP120	GP41	Nef
A1	13.07	7.3	7	19.89	5.59	6.44	4.76	13.63	9.86	18.14	16.83	23.04	23.34	13.38	14.64
B	12.92	4.95	10.83	14.99	8.21	6.04	4.91	14.71	11.29	19.32	18.34	20.04	23.93	15.59	17.79
C	16.11	5.77	9.39	15.52	6.15	5.84	4.45	10.61	10.64	15.37	16.10	20.21	22.89	14.19	14.96
D	14.78	5.08	10.85	13.16	7.68	6.98	4.80	11.57	11.74	17.89	15.87	19.40	23.58	13.61	14.42
F1	13.69	5.71	10.07	16.72	8.84	6.42	4.76	11.04	9.26	17.01	15.70	18.27	20.24	13.39	14.02
G	17.33	5.54	7.40	17.21	6.60	7.30	4.17	14.36	10.33	15.51	18.39	19.25	23.87	12.98	14.15
H	18.81	3.61	8.48	17.65	4.55	5.95	5.71	15.46	9.11	19.28	19.80	18.31	21.51	12.33	16.34
CRF01_AE	10.06	3.00	4.92	13.37	4.34	3.75	2.39	9.32	7.97	13.57	13.34	12.80	17.40	9.62	11.78
CRF02_AG	13.57	5.69	3.78	13.63	5.56	5.67	4.06	13.44	7.70	13.58	15.49	16.42	23.37	12.09	15.45
Group N	10.52	6.40	7.06	10.01	5.93	6.78	3.09	11.02	4.87	12.15	12.77	35.01	21.43	9.15	12.64
Group O	14.52	7.45	9.89	22.17	8.20	6.51	5.77	16.17	12.16	21.61	21.25	27.12	28.65	19.35	17.96
HIV-2															
Group A	10.15	4.67	11.82	11.02	7.14	6.89	6.78	12.74	22.55	22.59	22.75	10.40	18.55	16.47	21.98
Group B	12.14	3.54	13.15	15.4	5.76	6.4	6.19	12.71	12.66	20.44	15.82	11.45	21.20	15.29	20.59
Average difference	12.61	5.42	10.08	17.65	5.68	6.42	5.48	13.17	11.26	19.29	16.33	17.25	22.27	14.34	17.62

#: Only HIV groups, subtypes or CRFs with more than 2 genomic sequences are listed.
*: *Vpu* in HIV-1 and Vpx in HIV-2.

An earlier model to explain how retroviral recombination occurs is called forced copy-choice; which suggests that breakages in the RNA template causes RT to change templates during reverse transcription of the viral (–) DNA (Coffin, 1979). A more favoured and modified forced copy choice model known as dynamic copy choice, suggests that as RNAse H degrades the RNA template from the DNA+RNA hybrid during (–) DNA strand synthesis, the exposed DNA binds to the nearby free genomic RNA pulling the transcribing RNA+RT complex closer to the free genomic RNA. As RT has low template fidelity, it can switch from the transcribed RT template to the neighbouring RT template if the two templates are pulled close enough. It is possible too that the pull from the nascent (–) DNA bound to the neighbouring genomic RNA can cause the two template RNA to overlap with each other at a downstream region that is yet to be transcribed making it energy effficent for RT to switch templates (Hwang et al., 2001). Another model called strand displacement assimilation model proposes that during (+) DNA strand synthesis, RT synthesises short DNA fragments. These fragments bind to the (–) DNA strand and blocks the (–) DNA template from RT. Consequently, RT does not have access to the (–) DNA strand template and hence makes a template jumps to a nearby RNA and uses it as template (Boone and Skalka, 1981).

Recombination seems to require that there is homology within regions where RT switches template from the transcribed RNA template to the neighbouring template. Therefore, the more diverse two co-packaged RNA are the less likely they are to recombine. However, homology is not the only factor that determines recombination. It is likely that recombination *in vivo* may be driven by a combination of factors including presence of secondary structures like hairpins and stem-loops on the template RNA (Delviks-Frankenberry et al., 2011).

Recombinants (CRFs and URFs) are responsible for 20% of the global infections. The estimates of recombinants could be higher as most subtyping is usually done by using single (or two) regions of gene and not the whole genome (Santoro and Perno, 2013). Figure 1.5 shows the common CRFs globally. The most common CRFs are recombinant forms of subtype A and E (CRF_01 AE); recombinant forms of subtype A and G (CRF_02 AG); recombinant forms of subtype B and C, like CRF_07 BC and CRF_08 BC and recombinant forms of subtype B and F like CRF_12 BF and CRF_17 BF (Tebit et al., 2007).

UNIVERSITY OF VENDA
LIBRARY

1.2.11 Potential impact of recombination and viral diversity

1.2.11.1 Transmission and disease progression

A high number of CCR5 variants are observed during initial and early infection with HIV and a high number of CXCR4 variants are observed during late stages of infection (Schuitemaker et al., 2011). This suggests that HIV has evolved to rely on CCR5 co-receptor during initial infection or transmission. Hence CCR5 variants are easily transmitted than CXCR4 variants. Furthermore, the rate of disease progression is usually associated with the emergence of CXCR4 tropic strains which results in increased rates of CD4+ T cell depletion. A study done in Canada from 1996 to 2007 found that individuals with subtype B progressed to AIDS stage rapidly compared to African natives with non-B subtypes (that is subtypes A, C, F-K, AC recombinants, AE recombinants, AG recombinants, BF recombinants and DF recombinants) (Keller et al., 2009). Another study found that subtype D had significantly faster rates of CD4+ T cell depletion in comparison to subtype A, resulting in comparatively rapid disease progression (Kiwanuka et al., 2010). This could be due to the dual tropic nature of subtype D which consequently results in a lower rate of transmissibility in comparison to subtype A (Huang et al., 2007; Kaleebu et al., 2007). Dual infections have also been associated with elevated viral set point and rapid disease progression (Gottlieb et al., 2004). Cameroon has a higher frequency of dual-infections due to multiple subtypes co-circulating, resulting in numerous inter-subtype recombinants (Powell et al., 2009). As a result it is expected that some of these recombinants will show elevated viral set-point in those infected with them and show rapid progress to the AIDS stage.

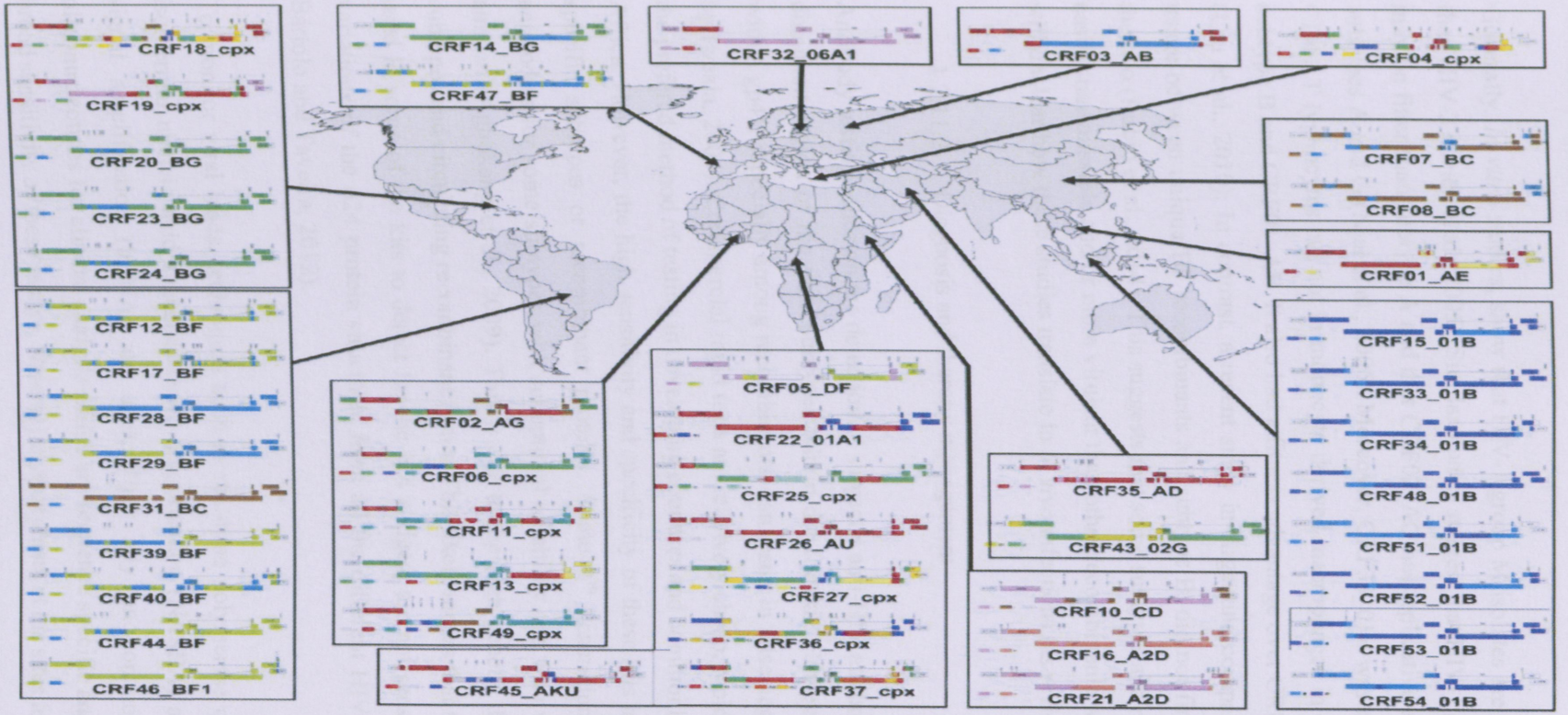


Figure 1.5: Global geographic distribution of circulating recombinant forms (CRFs)

The figure shows a wide distribution of recombinants globally with CRF31_BC and BF recombinants circulating mostly within Brazil. CRF18_cpx, CRF19_cpx and BG recombinants circulating within Cuba. CRF02_AG and CRF06/11/13/49_cpx within Ghana. CRF05_DF, CRF22_01A1, CRF26_AU, CRF25/27/36/37_cpx and CRF45_AKU within Democratic Republic of Congo. CRF10_CD and A2D recombinants within Ethiopia. CRF35_AD and CRF43_02G within Iran. CRF01_AE within Thailand and 01B recombinants within Indonesia. BC recombinant within China; CRF04_cpx in Greece; CRF03_AB in Russia; CRF32_06A1 in Lithuania; CRF14_BG and CRF47_BF in Spain (as reviewed by Hemelaar, 2013)

Generally, *In vitro* studies, show that HIV-1 group M isolates are more than 100 times fitter than HIV-2 (Ariën et al., 2005). Similar studies suggest that HIV-1 group M subtypes B and D may be fitter than subtype A and that CRF02_AG has replicative advantage over its parental subtypes A and G (Njai et al., 2006). Moreover, CRF33_01B was shown to replicate better in CD4+ T lyphocytes but not in monocyte derived macrophages in comparison to its parental subtype B and CRF01_AE. It also had a fitness advantage over CRF01_AE but not subtype B (Lau et al., 2010). In contrast, a recent study in Argentina comparing fitness in co-receptor usage between unique BF recombinants and subtype B, did not find any difference between the two (Rubio et al., 2014). This suggests that some subtypes and recombinants may be more easily transmissible and/or more virulent than other recombinants whereas others are not. It is still unclear how these studies translate in *in vivo* infections (as reviewed by Hemelaar, 2013).

1.2.11.2 Diagnosis and therapy management

Antibody testing relies on the detection of subject's anti-HIV antibodies to the gp41 immuno-dominant region of the viral subtypes within the test kit. However, polymorphisms found within gp41, especially among recombinants, can result in false negative results (as reviewed by Iweala, 2004). Commercial rapid tests are relatively cheap, rapid and instrument-free. It is the prefered method of testing in developing countries and commonly used within sub-saharan Africa. However, the high sensitivity and specificity of these kits have been optimised using specific subtypes or recombinant proteins. Some 4th generation assays detect anti-HIV antibodies for some subtypes with suboptimal reliability and with variable detection for p24 antigen (Aghokeng et al., 2009). This is due to the high genetic divergence among groups, subtypes and circulating recombinant forms. This likely means that the monoclonal antibodies used in some of the kits to detect for the p24 antigen, are not sensitive enough to detect all variations of the p24 protein structures from all the different HIV strains (as reviewed by Bártolo and Taveira, 2012).

To monitor viral loads, techniques such as real-time polymerase chain reaction (RT-PCR), isothermal nucleic acid sequence-based amplification (NASBA) and branched-chain DNA signal amplification (bDNA) are used. They rely on commercially made probes or oligonucleotides for clinical analysis which are sequence specific and recombination can limit broad sensitivity of these assays. Natural polymorphism at the sites for primer or probe binding

can result in reduced or unreliable quantification. The complexity observed in recombinant viruses means that there needs to be regular evaluation of effectiveness of these primers or probes to ensure they continue to function optimally (as reviewed by Bártolo and Taveira, 2012).

Similar challenges can be observed in genotypic assays used to monitor drug resistance. These assays rely on sequencing of the *pol* genes in order to check the sequences for known mutations for drug resistance. Genotypic assays are quicker than and as accurate as phenotypic methods. They have become a reliable method used by clinicians to monitor drug resistance and inform on options for therapy. Aghokeng et al., (2011) showed that commercial primers used in the ViroSeq assay failed to prime and sequence due to the high degree of polymorphism in non-B isolates commonly found within Cameroon. As recombination drives the virus to become more complex, improvements need to be made to test kits in order to maintain optimal antigen detection capability as well as anti-HIV antibody specificity and sensitivity.

1.2.11.3 Drug resistance

Combined therapy has been effective in reducing viral loads and stopping viral growth in HIV-infected people on treatment across subtypes. However, nucleotide variation within HIV strains means that each strain carries genetic signatures and polymorphisms that could alter the structure of viral proteins which are targeted by drugs, thereby impairing or reducing ARV drug binding and efficacy. Some subtypes have a low genetic barrier for drug resistance to a certain class of antiretroviral drug, since these subtypes require a single nucleotide substitution to acquire resistance (Santoro and Perno, 2013).

Despite the genetic variations, there has been no strain observed that has acquired multiple drug resistance mutations for all known classes of drugs (Santoro and Perno, 2013). The reason for this is that to acquire all the mutations required for multiple drug resistance to the different classes of antiretrovirals is a daunting evolutionary leap for the virus . However, recombination is one of the main tools that can make the virus achieve such evolutionary feats (Kellam and Larder, 1995). A possible scenario where this feat can be achieved is whereby, increases in co-circulating drug resitant viruses may lead to a possible recombinant virus that is resistant to multiple classes of drugs (Korber et al., 2001). However in such a case, it is also likely that as

drug-resistant recombinant for one class of drug co-circulate, it can recombine with other drug-susceptible subtypes and lose the drug resistance mutation through recombination, before acquiring a resistance mutation for another class of drugs. However if this was possible for a multi-drug resistant recombinant virus to arise, the virus would persist among the population and pose a challenge to treatment management (as reviewed by Korber et al., 2001; as reviewed by Fraser, 2005).

1.2.11.4 Vaccine development

An ideal vaccine has to provoke innate immunity, elicit neutralizing and non-neutralizing antibodies in large volumes as well as stimulate cytotoxic T lymphocyte (CTL) response to provide a durable, potent, and comprehensive immune response (as reviewed by Stephenson and Barouch, 2013). To this end, a live attenuated HIV vaccine could produce such effects; however, it has shown to cause persistent infection that develops to AIDS in animal tests. Inactive HIV vaccines and viral protein subunits stimulate only neutralizing antibodies with a poor stimulation of CTL response (McMichael et al., 2002).

An effective vaccine is one that can induce an immune response that protects the host during contact with the virus. For this to happen the vaccine needs to elicit a strong mucosal and innate immune response as well as neutralising antibodies that can clear the virus during early stages of infection prior to it reaching the target cells (as reviewed by Douek et al., 2006; as reviewed by Burton et al., 2012; as reviewed by van Gils and Sanders, 2013). Another strategy for an effective vaccine is to provoke responses by long living CD8+ T cells that can clear the virus from the host before latency can be established. A vaccine that elicits CTL response holds a lot of promise as CD8+ T cells can control viral loads by eliminating infected CD4+ cells. To elicit CD8+ T cell response in large volumes without using live virus, a recombinant viral vector has to be used (as reviewed by Tongo and Burgers, 2014). A recommended target for inclusion for such a vaccine would be viral capsid (24) protein as it is relatively conserved across types, groups, subtypes (Li et al., 2015). Furthermore, studies have shown that immune responses targetting p24 is associated with viral control (Zuñiga et al., 2006). Moreover, CD8+ T cell escape mutations within p24 epitopes, have been associated with a fitness costs to the virus (Martinez-Picado et al., 2006; Ammaranond et al., 2011). However, HIV can stay latent in a minority of long-living inactive CD4+ cells for a very long time. Since latency is

established within days or weeks of infection, CTL response alone cannot eliminate HIV infection (as reviewed by Johnston and Fauci, 2011). Therefore for the vaccine to be effective in preventing infection as well as controlling the virus upon intial infection, the vaccine would have to induce humoral as well as cellular immune responses.

The 2003 to 2005 RV144 vaccine trial in Thailand was an attempt at such a vaccine. The vaccine administered was a DNA vaccine made of a recombinant canarypox viral vector with subtype B and CRF01_AE *gag*, *pol* and *env* sequences (Rerks-Ngarm et al., 2009). The vaccine was boosted by two succesive injections of recombinant glycoprotein 120 subunit vaccine from subtype B and CRF01_AE. The trial looked at the protective effect the vaccine had towards inhibiting infection and in lowering viral loads upon infection. The vaccine showed a moderate efficacy of 31%, inducing T-cell–line adapted neutralizing antibody, cell-mediated cytotoxicity, CD4+ lymphoproliferation and CD8+ T cells. However, it did not completely inhibit infection or lower viral loads after intital infection (Rerks-Ngarm et al., 2009). Other trials that have been attempted include the STEP (conducted in the Americas and Australia) and Phambili (conducted in South Africa) trials, HVTN505 trial in America, VAX003 and VAX004 trial in Thailand and Americas. These trials revealed no protective efficacy for the vaccines used (as reviewed by Barouch and Michael, 2014).

The design and development of a vaccine relies on understanding the complex dynamics between host immune response and viral adaptation to selective pressure exerted by the host. These dynamics are further complicated by the high diversity and complexity of the virus (McMichael et al., 2002). To come up with an adequate vaccine, there has to be an awareness of the introduction of recombinant forms and their effect on vaccine design (as reviewed by Stephenson and Barouch, 2013; as reviewed by Tongo and Burgers, 2014).

1.3 Rationale for the study

The high HIV-1 diversity is due to a high viral turnover rate as well as the mutational events caused by HIV-1 heterodiploid RNA genome and poor RNA template fidelity, poor proof reading ability and recombinogenic properties of the reverse transcriptase enzyme (Tebit and Arts, 2011). A combination of these factors foster genetic variants within an infected individual's viral population that facilitate evasion of the host's immune responses. Consequently, this may result in improved viral fitness, which has an impact on diagnosis, disease progression, therapy management as well as vaccine design.

The number of recombinant viral forms observed keeps increasing. For example, in 2004, there were only 19 CRFs (Hemelaar et al., 2006) but by April 2015, there were 72 CRFs (HIV-LANL, 2016). This is partly due to better subtyping resulting from the advancements in sequencing technologies that aid in the extraction of enormous details of the genetic structure and complexity. Moreover, there is an increase in the global spread of CRFs from 5% in 2004 to 20% in 2013 (Hemelaar et al., 2006; as reviewed by Hemelaar, 2013). What has been observed, is the virus' enormous ability to recombine into very complex forms and spread globally to form regional epidemics.

HIV-1 group M, which is responsible for most of the global infection, has its possible origins in Cameroon. Cameroon as well as it's neighbouring countries, are known to be "hotspots" for a high diversity of the virus and are considered to be "ground zero" for all known types, groups and subtypes (Tebit et al., 2007; Courtney et al., 2015). Despite the predominance of recombinant CRF_02 AG within Cameroon, a study on possible unclassified recombinant viral strains that may be discovered within this region, could reveal information on the possible newer variations of the virus that may become the next regional epidemic or pandemic (Nyambi et al., 2002; Tebit et al., 2002). This study, which focuses on samples obtained from HIV-infected Cameroonian patients, therefore, aimed at finding and genetically describing these unclassified recombinants and hence provide some insight into the complexity of the identified unclassified recombinant viruses as well as predict what impact they may have on the course of managing the virus.

LIBRARY

1.4 Study hypothesis

There are multiple infections and recombination events going on within Cameroon where all known types and subtypes are co-circulating.

1.5 Objectives

Main Objective

The main objective of this study was to describe the genetic diversity and recombination events in HIV-1 in a cross-sectional population in Cameroon.

Specific Objectives

The specific objectives are:

- To screen for genetic diversity and recombinants by genotyping the *gag* gene and complementing them with *pol* and *env* gene generated from a previous study
- To characterize the full-length genomic sequence of an observed "novel" recombinant
- To predict Cytotoxic T lymphocytes epitopes

CHAPTER 2: MATERIALS AND METHODS

2.1 Ethical considerations

The study protocol was approved by the Ethics Committee of the Cameroon Baptist Health Board (IRB2012-01) and the Research Ethics Committee of the University of Venda, South Africa. Signed informed consents were obtained from all participants and from legal guardians of participants below the age of consent.

2.2 Study setting, population and sample collection

Mutengene Baptist Hospital is located 200 meters from the main road leading from Mutengene to Limbe in the Fako Division of the SW Region of Cameroon. The hospital has an HIV & AIDS Prevention and Care Unit that provides care and treatment to outpatients from the Southwest and Littoral Regions of Cameroon. Some of the cities from which participating patients came from are highlighted in Figure 2.1. Blood samples (5 ml) were collected into EDTA tubes from 83 drug naïve and 24 drug experienced individuals. Plasma was separated by centrifugation from whole blood and both plasma and total cells were shipped on dry ice to the University of Venda, South Africa.

Figure 2.1: Geographical location of the study population's provenance. This figure represents towns (blue circle) where the study participants originated (Google maps, 28 November 2014 17:52 pm).

2.3 Experimental protocols and data analysis

2.3.1 Investigating objective 1: To screen for genetic diversity and recombinants by genotyping the *gag* gene and complementing them with *pol* and *env* gene generated from a previous study

2.3.1.1 DNA extraction

DNA was extracted from total cells using the QIAamp DNA blood midi kit (Qiagen) according to manufacturer's instructions from blood. Eluted DNA was stored in aliquots at -20^{0}C.

2.3.1.2 Polymerase chain reaction (nested PCR)

The primers used for PCR and sequencing of partial *gag* were as defined previously (Tebit et al., 2002). Amplification targeted continuous partial p17 and partial p24 sub genomic regions starting from nucleotide position 838 to 1608 (relative to HXB2) as shown in table 2.1.

Table 2.1: Primer sets used for pre-nested and nested PCR for partial HIV *gag* amplification

Primer Name	Sequence (5' → 3')	Primer Location relative to HXB2	Expected amplicon Length
Gag 1F (Forward primer)	TAT GGG TGC GAG AGC GTC AGT ATT AAG	792 → 815	863 bp
Gag 2R (Reverse primer)	TTT GGC CCT TGT CTT ATG TCC AAA ATG C	1 655 → 1 631	
Gag 3F (Forward primer)	TGG GAG AAA ATT CGG TTA AGG CC	838 → 857	770 bp
Gag 4R (Reverse primer)	TAC TAT TTT ATT TAA TCC CAG GAT TAT	1 608 → 1 585	

Gag 1F and 2R were pre-nested PCR primers and Gag 3F and 4R were nested and sequencing primers. Primer name and sequence obtained from Tebit et al., 2002.

Table 2.2 shows the details of the PCR reagents and concentrations used for pre-nested and nested PCR. Table 2.3 shows the cycling conditions used for both the pre-nested and nested PCR. The reagent concentration and cycling conditions are as described in Tebit et al., 2002.

Table 2.2: Pre-nested and nested PCR reaction concentrations for partial *gag* amplification

Gag PCR	**Pre-nested PCR**	**Nested PCR**
Reagents	Concentration	Concentration
10X PCR buffer (Invitrogen)	1X	1X
50 mM $Mgcl_2$ (Invitrogen)	1.5 mM	1.5 mM
10 mM dNTP mix (Invitrogen)	0.2 mM	0.2 mM
5U/µl Platinum Taq DNA polymerase (Invitrogen)	0.02U/µl	0.02U/µl
Gag 1F/Gag 2R	0.2 µM	0.2 µM
Gag 3F/Gag 4R	0.2 µM	0.2 µM
PCR H_20 (Gibco by Life Technologies distilled water DNase/RNase free)	-	-
Total Reaction Volume	25 µl	50 µl

The extracted DNA was used as a template for pre-nested PCR and amplicons from pre-nested PCR were used as template for nested PCR.

Table 2.3: Cycling conditions for pre-nested and nested PCR

Cycling Step	**Number of Cycles**	**Temperature (^{0}C)**	**Duration**
Initial Denaturation	1	94	3 minutes
Denaturation	30X	94	30 seconds
Annealing		52	30 seconds
Extension		72	2 minutes
Final Extension	1	72	7 minutes
Hold		4	Infinity

The resulting nested PCR amplicons were either stored overnight at 4^0C within the thermocycler then at -20 ^{0}C the next day or were stored immediately after the PCR run at – 20 ^{0}C.

2.3.1.3 Quality control measures for limiting PCR contamination

PCR provides a rapid and highly sensitive method to amplify DNA that can result in 10^9 amplicons per PCR cycle (as reviewed by Aslanzadeh, 2004). To restrict PCR contamination,

good laboratory practices were followed; for instance, gloves and lab coats were always worn when working. The PCR mastermixes were prepared in a separate room from where the template DNA was added and from where the PCR was run. The PCR mastermixes were prepared in a biosafety class 2 cabinet using DNase, RNase and Pyrogen free aerosol resistant pipetting tips which were disposed of after each use (Mifflin, 1997). The biosafety cabinet had UV light which was turned on for at least 10 minutes prior to use of the cabinet with pipettes, tips and other PCR equipment minus reagents exposed within the cabinet to the UV light. Regularly, the PCR biosafety cabinet surface was disinfected with 5% sodium hypochlorite solution. Prior to each mastermix preparation, the cabinet surface and pipettes were wiped down with DNA AWAY (Molecular BioProducts), then RNase AWAY (Molecular BioProducts) before wiping down with 70 % ethanol. The pipettes sets were dedicated to each room or area and were wiped with 70 % ethanol prior and after use. The PCR reagents were opened only within the cabinet after disinfection and the samples only in the area dedicated for adding samples. PCR mastermix was prepared for multiple PCR reactions to avoid repeated pipetting and possible contamination. A negative control that contained the required volume of PCR mastermix and same volume of PCR water as used for the template DNA within the sample PCR reactions, was used for every PCR ran as a means to detect contamination. The PCR amplicons were analyzed in a separate area from where the PCR was prepared or run (as reviewed by Aslanzadeh, 2004; as reviewed by Lo and Chan, 2006). All waste was disposed of in biosafety bins. To confirm precise amplification of desired product, the sequenced amplicons were analyzed on NCBI BLAST to confirm homology with HIV-1 subtypes and not other viral, non-viral, plasmid, bacterial or human DNA. Quality of the received sequences was checked using the Los Alamos National Laboratory (LANL) HIV Database Quality Control tool (http://www.hiv.lanl.gov/content/sequence/QC/index.html).

2.3.1.4 Gel electrophoresis of PCR products

A 2% agarose gel was prepared by completely dissolving 2 g of molecular grade agarose (Invitrogen) in 100 ml of 1X Tris, acetic acid and ethylenediaminetetraacetic acid (EDTA) (TAE) buffer (Qiagen). Complete dissolution was ensured by intermittent heating in a microwave oven. Molten agarose was allowed to cool and stained with 6 µl of ethidium bromide (10 mg/µl), followed by casting in a gel tray with gel combs positioned to create wells as the agarose solidifies. The nested PCR amplicons (3 µl) were loaded, followed by

3 µl 100 bp DNA molecular weight marker and controls in designated wells. DNA was resolved by electrophoresis for 45 minutes at 80 volts. The resolved PCR products were viewed under a UV transilluminator (Syngene, Germany) and documented.

2.3.1.5 Purification of PCR products

PCR amplicons that showed single bands from the resolved gels, were purified using QIAquick PCR Purification Kit (Qiagen) according to manufacturer's instructions. The eluted products (purified PCR amplicons) were confirmed by resolving two µl on a 2% agarose gel.

The PCR products which showed multiple bands were gel extracted and purified using a QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions. Two µl of the eluted DNA was run on a 2% agarose gel.

2.3.1.6 Sanger sequencing

Direct sequencing was done on the forward and reverse strands of the generated nested purified PCR amplicons with the ABI V3.1 using a BigDye Direct cycle sequencing kit and run on the ABI 3500XL instrument (Applied Biosystems). This sequencing was outsourced to Inqaba Biotec., Pretoria, South Africa.

2.3.1.7 Sequencing and phylogenetic Analyses

Nucleotide sequences were assembled and manually edited using SeqMan software (SeqMan Pro Version 8.1.2 DNASTAR). The sequences were aligned with a representative set of sequences of known HIV-1 group M subtypes (A-D, F-H,J,K), as well as known circulating recombinants and one representatives sequences of HIV-1 group O obtained from the LANL database (http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html) using Clustal W implemented within the Molecular Evolutionary Genetic Analysis (MEGA) v. 6.0 software (Tamura et al., 2013). Rough/draft trees were constructed and edited by eliminating recombinants and subtypes that did not cluster with any of the patient samples sequences until a final tree was constructed under the Kimura 2-parameter model with a bootstrap of 1000 replicates. Two online tools namely, Recombination Identification Program (RIP) from the Los Alamos National Laboratory HIV database

http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html (Siepel et al., 1995) and Jumping Profile Hidden Markov Model (jpHMM) http://jphmm.gobics.de (Zhang et al., 2006) were used to analyze for recombination patterns.

2.3.1.8 Plot identity for partial *gag* sequences

The partial *gag* nucleotide sequences were translated and the p17 and p24 portions of the *gag* alignment was clipped and codon aligned using the Gene Cutter tool from HIV database (http://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html). The sequences were then grouped based on determined subtypes and aligned with a subtype consensus retrieved from HIV sequence database (http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html). These groupings were subtype A1, CRF01_AE, subtype G and subtype F2. A plot identity was constructed for p17 and p24 for each of the subtype groups using BioEdit Sequence Alignment Editor (BioEdit v. 7.2.5) software (Hall, 1999).

2.3.2 Investigating objective 2: Characterize the full-length genomic sequence of an observed "novel" recombinant

Genotyping results for partial *gag* was matched with genotyping results for partial *pol* (complete protease and partial reverse transcriptase) and partial *env* (partial gp120 and partial gp41) for samples that were available from an earlier study done within the laboratory. Based on the matching of the results, samples were selected for full-length amplification and sequencing based on their perceived unique recombination patterns.

To generate near full-length amplicon, a nested PCR was done to amplify two overlapping fragments of approximately 4.5 kb each that span the length of the HIV-1 genome. The 5 prime (5') fragment spanned from 5' LTR U5 to *vif* start (552 to 5068 relative to HXB2) covering a 4 516 bp length and the 3 prime (3') fragment spanned from near end of integrase to 3' LTR U5 (4 956 to 9636 relative to HXB2) covering a 4 680 bp length. This summed up to an amplification of a near full-length genome of 9 084 bp.

UNIVERSITY OF VENDA
LIBRARY

2.3.2.1 Primer design

Primers used for near full-genome amplification were as described (Salazar-Gonzalez et al., 2009). Table 2.4 shows the primer sets used for pre-nested and nested PCR for amplifying the 5' fragment (from 5' LTR U5 to *vif* start) and table 2.5 shows the primer sets for amplifying the 3' fragment (from near end of integrase to 3' LTR U5).

Table 2.4: Primer sets for pre-nested and nested PCR for amplifying the 5' fragment

Primer Name	Sequence (5' → 3')	Primer Location relative to HXB2	Melting Temp (^{0}C)	Expected amplicon Length
PRE-NESTED PRIMERS				
1.U5.B1F (Forward Primer)	CCT TGA GTG CTT CAA GTA GTG TGT GCC CGT CTG T	538→571	71.9	4 539 bp
B5R1 (Reverse Primer)	CCT GCC ACA CAA TCA TCA CCT GCC AT	5 077→5 052	67.8	
NESTED PRIMERS				
2.U5.B4F (Forward Primer)	AGT AGT GTG TGC CCG TCT GTT GTG TGA CTC	552→581	70.1	4 516 bp
B5R2 (Reverse Primer)	CAA TCA TCA CCT GCC ATC TGT TTT CCA TA	5 068→5 040	64.6	

Primer name and sequence obtained from Salazar-Gonzalez et al., 2009.

Table 2.5: Primer sets for pre-nested and nested PCR for amplifying the 3' fragment

Primer Name	Sequence (5' → 3')	Primer Location relative to HXB2	Melting Temp (^{0}C)	Expected amplicon Length
PRE-NESTED PRIMERS				
B3F1 (Forward Primer)	ACA GCA GTG CAA ATG GCA GTA TT	4 749→4 771	60.9	4 893 bp
1.R3.B3R (Reverse Primer)	ACT ACT TGA AGC ACT CAA GGC AAG CTT TAT TG	9 642→9 611	65.9	
NESTED PRIMERS				
B3F3 (Forward Primer)	TGG AAA GGT GAA GGG GCA GTA GTC ATA C	4 956→4 983	67.5	4 680 bp
2.R3.B6R (Reverse Primer)	TAA AGC ACT CAA GGC AAG CTT TAT TGA GGC	9 636→9 607	66.0	

Primer name and sequence obtained from Salazar-Gonzalez et al., 2009.

2.3.2.2 Nested PCR, gel electrophoresis and gel purification

Tables 2.5 shows the PCR reagents and concentrations used for pre-nested and nested PCR for both 5' and 3' fragments and table 2.6 shows the cycling conditions used. The reagent concentration and cycling conditions are as described in Salazar-Gonzalez et al., 2009.

Table 2.6: Pre-nested and nested PCR reaction concentrations for 5' and 3' fragment amplification

Near-full length PCR	**Pre-nested PCR**	**Nested PCR**
Reagents	Concentration	Concentration
10 × Buffer with 17,5 mM $MgCl_2$ (Buffer 1)	1X	(1X)
10 µM 1.U5.B1F / B3F1	0.3 µM	0.3 µM
10 µM B5R1 / 1.R3.B3R	0.3 µM	0.3 µM
10mM dNTP mix	0.35 mM	0.35 mM
5 U/µL Expanded Long Template Enzyme (Roche)	3.75 U/µL	3.75 U/µL
ddH_20 (Gibco by Life Technologies distilled water DNase/RNase free)		
Reaction Volume	25 µl	50 µl

Table 2.7: Cycling conditions used for both the pre-nested and nested PCR for both fragments

Cycling Step	**Number of Cycles**	**Temperature (^{0}C)**	**Duration**
Initial Denaturation	1	94	2 minutes
Denaturation	10X	94	15 seconds
Annealing		60	30 seconds
Extension		68	5 minutes
Denaturation	25X	94	15 seconds
Annealing		60	30 seconds
Extension		68	5 minutes with 20 seconds increment per cycle
Final Extension	1	68	15 minutes
Hold		4	∞

The resulting nested PCR amplicons were stored at – 20 ^{0}C.

A 0.8 % agarose gel was prepared and 2 µL of each amplicon was loaded into each respective well. Two µL of 1 kb plus DNA ladder (ThermoFisher Scientific GeneRuler) was loaded into a single empty well. DNA was resolved by electrophoresis for 60 minutes at 140 volts. The resolved PCR products were viewed under a UV transilluminator (Syngene) and documented. The amplicons were gel extracted and purified using the QIAquick gel extraction kit (Qiagen) as mentioned in section 2.3.1.5.

2.3.2.3 Quality control

The same measures to minimize PCR contamination described under section 2.3.1.3 were observed. Furthermore, to ensure that the desired template was amplified, partial *pol* fragment was sequenced from the 5' (5' LTR U5 to *vif* start) fragment and partial *env* fragment was sequenced from the 3' (near end of integrase to 3' LTR U5) fragment using sanger sequencing (ABI 3500XL). The sequencing primers for partial *pol* were Gag P1 and Pol 2R; for partial *env* were Env 2F and Env 2R as listed in table 2.8.

Table 2.8: Sequencing primers for partial *pol* and partial *env* from the 5' and 3' fragments

Primer Name	Primer Sequence (5' → 3')	Primer Location relative to HXB2	Expected sequence length
GAG P1 (Forward)	CAA GGG GAG GCC AGG GAA TTT	2112 → 2132	917 nucleotides
POL 2R (Reverse)	TAA TAT TGC TGG TGA TCC TTT CC	3029 → 3007	
ENV 2F (Forward)	CAA CTC AAC TGC TGT TAA ATG GCA G	6991 → 7015	687 nucleotides
ENV 2R (Reverse)	TTA TAT AAT TCA CTT CTC CAA TT	7678 → 7656	

The resulting sequences were checked for quality using the quality control online tool from Los Alamos National Laboratory (LANL) HIV Database Quality Control tool (http://www.hiv.lanl.gov/content/sequence/QC/index.html). After manual editing, the same recombinant tools mentioned in section 2.3.2.1 were used to confirm that the same subtypes and recombination patterns observed in the earlier analysis were present within the same samples.

2.3.2.4 Illumina sequencing

Near-full length genome sequencing was done for sample BM189 by sequencing two overlapping fragments of about 4,5 kb each in length using Illumina Miseq sequencer. Deep sequencing was outsourced to Inqaba Biotec., Pretoria, South Africa where the purified PCR products were couriered to.

Miseq is a next-generation sequencer that uses massive parallel sequencing to generate gigabytes of sequence data. It does this through three main steps namely: Library preparation, cluster generation and sequencing by synthesis (as reviewed by Grada and Weinbrecht, 2013; as reviewed by Bahassi and Stambrook, 2014). During library preparation, the sample DNA is randomly chopped to give fragments of 150 to 250 base-pairs long. The sheared ends are trimmed and polyadenylated. Adapter oligos are then bound to each ends to tag the DNA. The adapter oligos have unique sequences that complimentary bind to oligos within the flow cell and also have sequences complimentary to sequencing primers (as reviewed by Dijk et al., 2014). The fragments are then size selected to remove unbound oligos and the library is purified.

During cluster generation, the library DNA is introduced into a flow cell (as reviewed by Casals et al., 2012). The flow cell is made up of a sealed glass slide with 8 micro-fluid channels with each channel having nanopores whereby many oligos complementary to the adapter oligos are covalently bound at their 3' end (as reviewed by Mardis, 2013). Each library fragment introduced to the flow cell is denatured and allowed to hybridize to a complementary oligo on the flow cell based on the unique sequence it has on its adapter. The cluster generation cycle begins with the library fragments bound to the flow cell being amplified and afterwards the original strand is denatured and washed off. In the next cycle, the DNA strand left bound to the flow cell hybridizes its free 5' end to a neighboring oligo on the flow cell and then amplified (Morey et al., 2013).

The amplified double stranded DNA is untethered at one end to give single strands when denatured. In the next cycle, each single strand can bind its free ends to a neighboring oligo on the flow cell and then amplified. This process of clonal amplification through bridge amplification generates many copies of the original template that eventually form hundreds of millions of unique spatially separated clusters on the flow cell (as reviewed by Zhang et al.,

2011). After cluster generation, the reverse strand is removed and washed off while the forward strand 3' ends are blocked to avoid unwanted priming with sequencing primers.

Reversible terminator sequencing by synthesis begins with the sequencing primer being hybridized to the forward strands at the adapter oligo. Polymerase adds a fluorescently labelled nucleotide that has its carbon-3 blocked to inhibit further polymerization. The nucleotide is added in a complementary fashion to the base on the template (as reviewed by Metzker, 2010; as reviewed by Morey et al., 2013). The bound nucleotide is then exited by a laser and the fluorescence given off that is unique for that nucleotide is captured by a camera and recorded. The florescence label and the blocking agent are then removed from the bound nucleotide and washed away and another cycle of nucleotide addition followed by excitation and recording follows (as reviewed by Mardis, 2013). The cycles are continued until the length DNA is read. A reverse strand can be amplified from the forward strand and read in a similar fashion as the forward strand using a reverse sequencing primer to give paired end reads (as reviewed by Dijk et al., 2014). This sequencing process takes place on all the strands within all the clusters on the flow cell in a parallel fashion. As a result, it can give up to an average of 1 giga-base-pairs (ranging from 440 mega-base-pairs to 7 giga-base-pairs) reads per run with a raw base accuracy of greater than 99.5% (as reviewed by Zhang et al., 2011; as reviewed by Liu et al., 2012; as reviewed by Rizzo and Buck, 2012).

2.3.2.5 Genetic analysis of BM189

The sequencing run resulted in 1 519 451 pair-ended sequence reads that were assembled to give a consensus sequence of the dominant strain of the sample. The genome nucleotide sequence as well as the viral gene sequences were aligned using MEGA v. 6.0 software (Tamura et al., 2013) and neighbor-joining trees were constructed. The subtype was inferred from the clustering pattern observed. The genome was analyzed for recombination using Simplot version 3.5.1 (Lole et al., 1999). The positions of recombination breakpoints was obtained from the Simplot results and used to map the breakpoints on a HIV-1 gene map using HIV-1 Recombinant Drawing Tool (http://www.hiv.lanl.gov/cgi-bin/DRAW_CRF/crfmap.cgi). Simplot was also used to analyze BM189 viral genes. Online

tools jpHMM (Zhang et al., 2006) and RIP (Siepel et al., 1995) were also used to determine recombination patterns within the genome and viral gene nucleotide sequences.

A drug resistance report (Advanced Biological Laboratories) was generated as part of the post-sequencing data analysis. It provided the drug resistance profile of the sample based on the analysis of drug resistant mutations in protease, reverse transcriptase, integrase and glycoprotein 120 and glycoprotein 41.

The V3 loop amino acid sequence was extracted from the envelope sequence using Gene Cutter tool for sequence alignment and protein extraction from LANL HIV sequence database (http://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html). It was aligned using Clustal W from BioEdit Sequence Alignment Editor (BioEdit v. 7.2.5) software (Hall, 1999) to a consensus A1, and CRF01_AE as well as A1 subtype sequence that gave the highest score on NCBI BLAST (KF716475) and a plot identity was constructed.

Co-receptor usage was predicted using two different web based algorithms namely the geno2pheno (g2p) co-receptor 2.5 tool (Lengauer et al., 2007) from Max Plank Institute for Informatics website and the position specific scoring matrices (web PSSM) tool from the University of Washington, Seattle. For g2p, a false positive rate (FPR) of 10% was used according to recommendations from the European Consensus Group on clinical management of HIV-1 tropism testing (Vandekerckhove et al., 2011). All FPR below 10 were considered CXCR4. The PSSM method gave a prediction score of 1 for X4 tropic virus or viral sequence and a score of 0 for predictions of R5 tropic virus. PSSM has been reported to reliably predict CXCR4 usage (Jensen et al., 2003). The prediction score was calculated from position scoring of amino acid residues at position 11 and position 25 of the V3 sequence, the net charge of residues within the query V3 sequence and the total number of positively charged residues within the sequence. PSSM calculated the likelihood of the sequence to be a V3 loop sequence and gave percentiles score. Percentiles higher than 0.95 were considered not to be V3 loop sequences.

CTL epitope prediction for *gag* and *env* genes was done using MHC-1 tool from IEDB and NETCTLpan as described in section 2.3.3.

2.3.3 Investigating objective 3: Predicting CTL epitopes

The p17 and p24 amino acid sequences were examined for potential CTL epitopes using immune epitope database (IEDB) tool called MHC-1 binding (Moutaftsi et al., 2006). The MHC-I binding predictions were made using the IEDB analysis resource consensus tool (Kim et al., 2012) which combines predictions from artificial neural network (ANN) (Nielsen et al., 2003; Lundegaard et al., 2008), stabilized matrix method (SMM) (Peters and Sette, 2005) and scoring matrices derived from combinatorial Peptide Libraries (Comblib) (Sidney et al., 2008). The amino acid sequences were queried for epitopes that are restricted by HLA alleles with high frequencies (10% to 23,1% allele frequency) among populations from the coastal region of Cameroon (Southwest and Littoral region) (Torimiro et al., 2006). These alleles include: A*3004, A*3303, B*3501, B*5301, B*5801, C*0202, C*0401, C*0602, C*0701, C*0702, C*0706 and C*0718. Default parameters were used and epitopes that gave a percentile score less than one were considered to have a higher than average binding affinity. Those epitopes that gave an inhibitory concentration 50 (IC_{50}) of less than 50 nanomolar (nM) concentration were predicted to have the best binding affinities. NetCTLpan was also used to predict CTL epitopes (Stranzl et al., 2010). The program predicted three processes vital to MHC-1 epitope presentation and calculated a combined score. These three processes are namely: proteasomal cleavage of viral proteins into peptides, transporter associated with antigen processing (TAP) and MHC-1 binding. MHC class I binding and proteasomal cleavage was performed using artificial neural networks method. TAP transport efficiency was predicted using weight matrix (Peters et al., 2003). The prediction values were calculated as a weighted average of the MHC-1 binding, TAP efficiency and C terminal cleavage scores were given as a percentage rank to a set of 200 000 random natural peptides. ANN, SMM, Comblib and weight matrix methods are machine learning methods that rely on training using experimentally known epitopes in order to predict epitopes from query protein sequences. The methods work by randomly fragmenting the query protein into peptides and mixing them with thousands of fragments of known peptides retrieved from databases as well as *in-silico* simulated peptides and scoring each peptides by the parameters that predict a good binder/epitope.

CHAPTER 3: RESULTS

3.1 Characteristics of study population

The characteristics of patients enrolled for the study are summarized in Table 3.1. There were 83 drug naïve and 24 drug experienced patients. Ten subjects were less than 15 years old, while females comprised about 73% (78/107). There were 60 patients with CD4 cell count below 350 cells/mm^3 and 47 above 350 cells/mm^3. There was no information available on viral load measurements.

Table 3.1: Demographic and immunologic profile of the study population

Parameters	**Adults >15 years**	**Pediatrics < 15 years**
Overall patients	97	10
Sex *Female* (%)	70 (72.2%)	8 (80%)
Age range (Median)	19-59 (39)	2-14 (8)
CD4 cells/mm^3 range (Median)	12-1357 (684.5)	64-1498 (781)
ARV experience *Drug naïve* (%) *Drug experienced* (%)	 74 (76.3%) 23 (23.7%)	 9 (90%) 1 (10%)

3.2 Amplification of HIV-1 partial *gag* gene

Nested PCR was performed on DNA from 83 drug naïve and 24 drug experienced patients. The PCR was done to target the region from nucleotide position 838 to 1608 relative to HXB2 which was expected to amplify a *gag* fragment of 770 bp. The band covered the junction between the matrix protein (p17) and capsid protein (p24) within the *gag* sub-genomic region covering 348 nucleotides from p17 end and 422 nucleotides from p24 start. Out of the 107 samples that PCR was attempted on, 52 samples were successfully amplified (48.60 %) with 49 samples from drug naïve patients and three from drug experienced under 15 years old patients. The 55 samples that failed amplification were from 35 drug naïve patients and 20 drug experienced patients. Figure 3.1 is a representative gel of PCR amplification of the *gag* fragments.

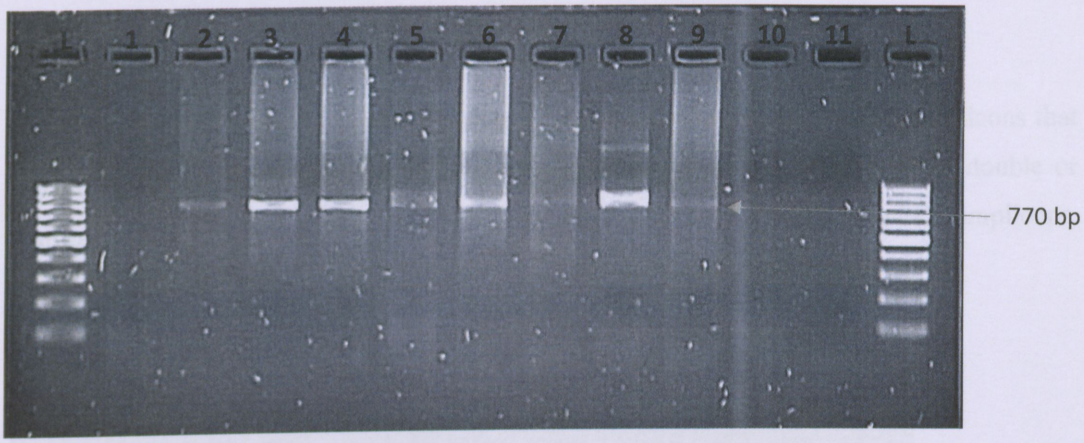


Figure 3.1: Amplification of 770 bp partial *gag*.

2% ethidium bromide stained agarose gel electrophoresis of proviral DNA after nested PCR. Lane L - 100 bp Ladder (ThermoFisher Scientific O'GeneRuler), Lane 1 – 9 are nested PCR amplicons of patient DNA, Lane 10 – Negative control for nested PCR and Lane 11 is duplicate negative control for nested PCR.

As noted, some of the amplicons had multiple bands (for example Lane 8 had a distinct undesired band of over 1 000 base-pairs). Attempts were made to optimize the PCR conditions to produce a higher yield of desired product and improve specificity of primers to amplify only desired target DNA. These attempts included changing annealing temperatures by one or two degrees for each nested PCR run for temperature ranging from 53 ^{0}C to 64 ^{0}C. Moreover attempts were made to use different annealing temperatures for pre-nested and nested PCR of each nested PCR using the same temperature range (53 ^{0}C to 64 ^{0}C). The temperature range was guided by the melting temperature of the primers (± 5 ^{0}C) however higher and lower temperatures were considered. Other attempts included using higher volume of sample DNA for pre-nested PCR (from 2 µL to 3 µL then 5 µL in a 25 µL PCR reaction) for samples that showed low yield or no amplification on initial run. Despite getting better PCR products for some samples, some samples did not amplify and seven samples yielded multiple bands and had to be gel extracted and purified.

3.3 PCR purification, gel extraction and sequencing

The PCR amplicons were purified either using PCR purification kit for those amplicons that yielded single bands or gel extraction and purification kit for those that yielded double or multiple bands. Figure 3.2 is a representative gel of purified HIV-1 partial *gag* PCR amplicons.

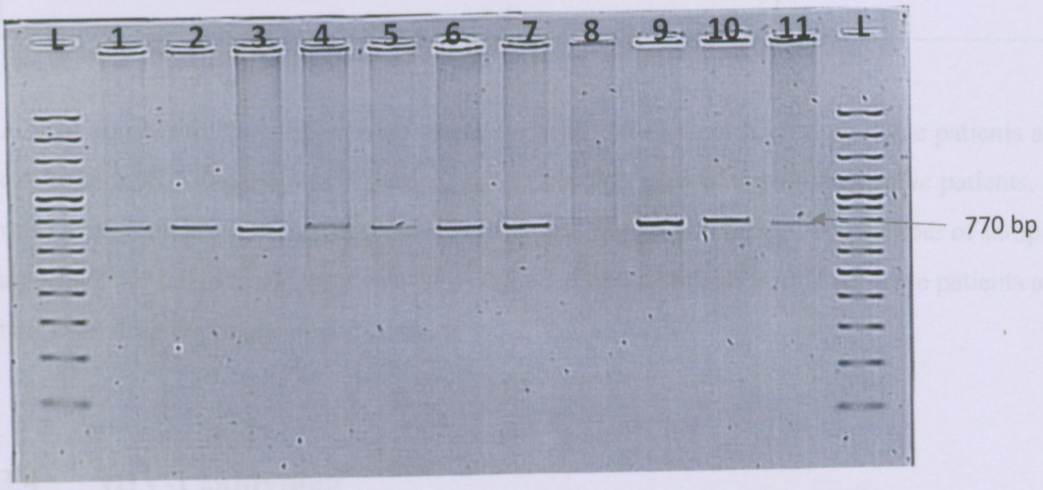


Figure 3.2: PCR purification of 770 bp partial *gag*.

2% ethidium bromide stained agarose gel electrophoresis of PCR purified amplicons. Lane L - 100 bp Ladder (ThermoFisher Scientific O'GeneRuler), Lane 1 – 11 are PCR purified amplicons.

As shown on Figure 3.2 some PCR amplicons showed very weak desired band after purification there were barely visible. These samples were not considered for sequencing.

Some amplicons failed to pass sequencing quality control and were not sequenced. Some of the sequenced amplicons had poor sequence signals. Hence these amplicons/samples were not considered for further analysis. Table 3.2 gives a summary of PCR amplification and sequencing outcomes.

UNIVERSITY OF VENDA
LIBRARY

Table 3.2: Outcome of partial *gag* gene amplification and sequencing

	Number of samples run	**Percentage success per method done (%)**
Samples PCR was attempted	107	
Samples successfully amplified	52	48.60
Samples successfully purified (had the right intensity bands for sequencing)	48	92.3
Samples successfully sequenced	39	81.25

A total number of 39 samples were sequenced with 38 of them from drug naïve patients and one from a drug experienced 7 year old male (BM402). Out of the 38 drug naïve patients, 24 were adult females, 12 were adult males and 2 were female pediatrics. The number of samples amplified yet failed to obtain a sequence were 13 with 11 of them from drug naïve patients and two from drug experienced pediatrics.

3.4 HIV-1 subtyping

The sequences obtained were assembled and manually edited using SeqMan. They ranged in length from 702 to 758 nucleotides spanning from position 838 to 1602 relative to HXB2 with the exception of BM236, which was 524 nucleotides long spanning 1072 to 1596 relative to HXB2. The sequences were aligned using Clustal W in MEGA 6.0 and trimmed to ensure that the same nucleotide positions for each sequence was considered for further analyses. For 38 partial *gag* sequences, 658 nucleotides each were considered for further analysis and this spanned the region from 868 to 1526 relative to HXB2. For one sequence out of the 39, that is BM236 all 524 nucleotides were considered. Quality control (QC) was done on each of the sequences using HIV LANL database Quality control software (http://www.hiv.lanl.gov/content/sequence/QC/index.html). All the sequences passed QC showing no hypermutations or stop codons. The sequences were analyzed using RIP 3.0 from the HIV Los Alamos National Laboratory database with a window size of 600 and a confidence threshold of > 0.9. It showed that 33 sequences were subtype A1 or/and CRF01_AE with two (BM250 and BM548) being assigned as subtype G and two (BM189 and BM488) being assigned subtype F2. One sample (BM526) was a recombinant for subtype A and G (Figure

3.3) whereas BM236 RIP was unclassified. Figure 3.4 is a RIP analysis of the *gag* gene of BM189 showing it to be a F2 subtype with a confidence threshold of 0.96.

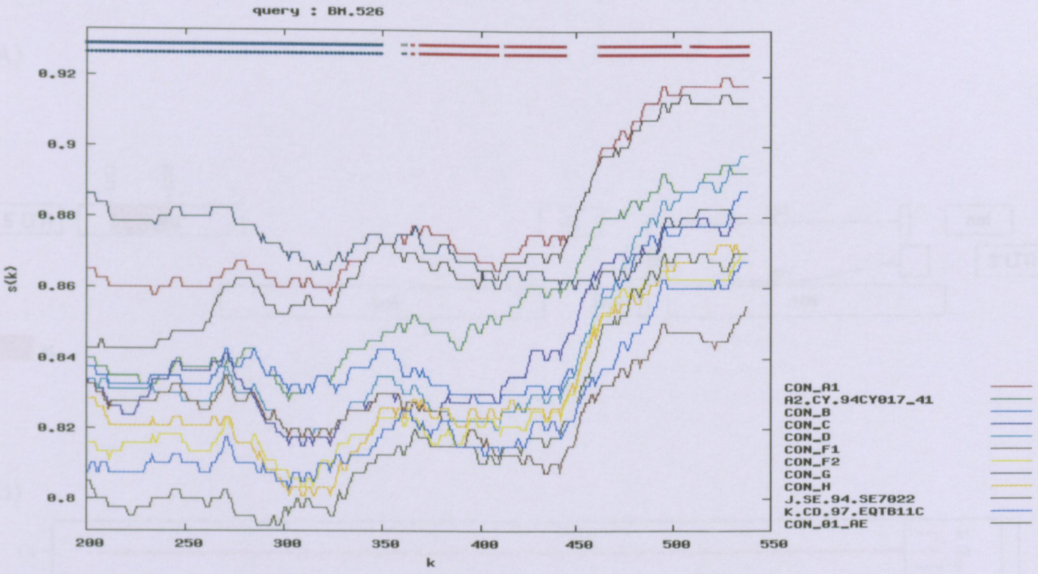


Figure 3.3: RIP analysis of BM526 partial *gag* gene.

Recombination analysis of BM526 *gag* sequence which was assigned subtypes A1 and G recombinant with a high threshold support > 0.9 and window size of 600.

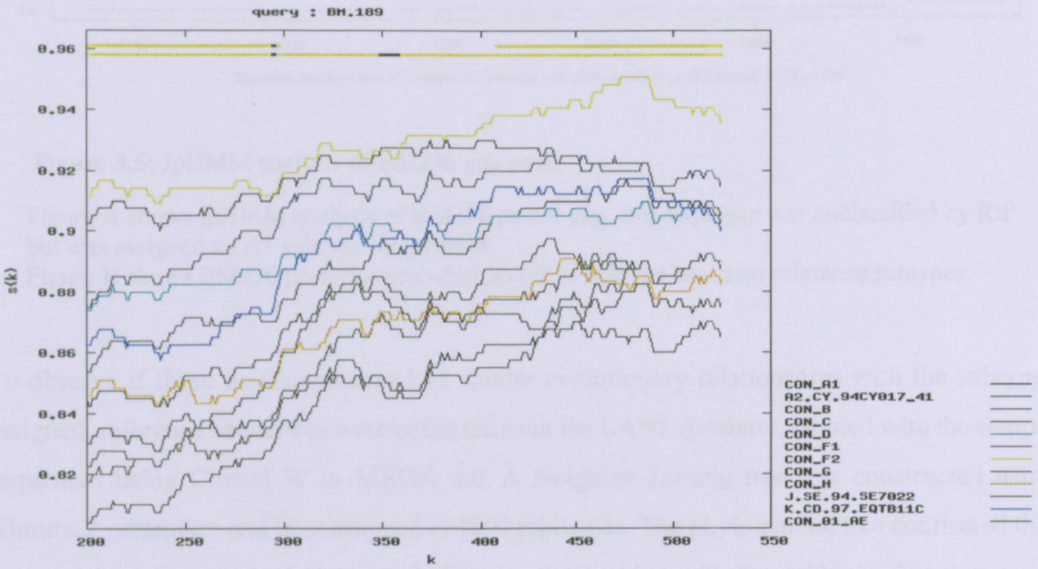
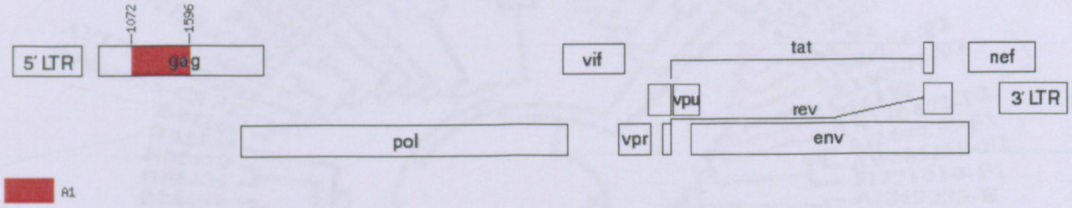


Figure 3.4: RIP analysis of BM189 partial *gag* gene.

Recombination analysis of BM189 *gag* sequence which was assigned subtype F2. Threshold support > 0.9 with a window size of 600.

These results were compared to those obtained from analyzing the same sequences with jpHMM. The results confirmed the subtypes observed with RIP. Figure 3.5 shows Subtype A1 was assigned for BM236 by jpHMM with a posterior probability approaching 1.

(A)



(B)

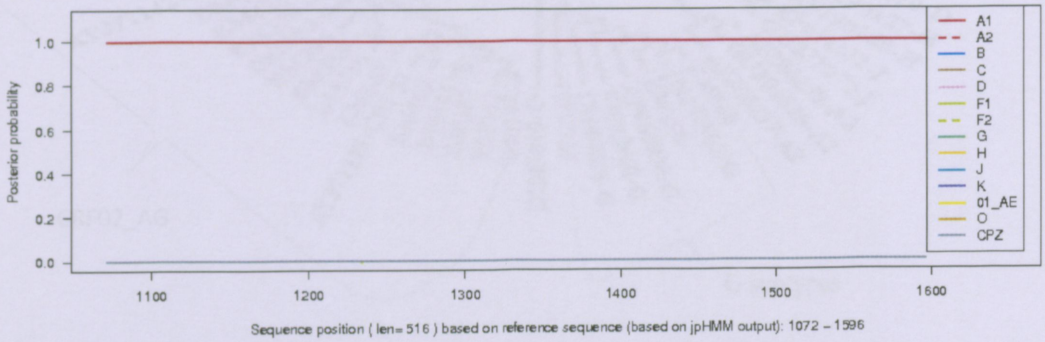


Figure 3.5: JpHMM analysis of BM236 *gag* gene.

Figure A shows jpHMM analysis of BM236 partial *gag*, this sequence was unclassified by RIP but was assigned an A1 subtype by jpHMM.
Figure B shows BM236 posterior probabilities of the subtype based on reference subtypes.

To observe if these new sequences had similar evolutionary relationships with the subtypes assigned, reference sequences were collected from the LANL database, aligned with the sample sequences using Clustal W in MEGA 6.0. A Neighbor-Joining tree was constructed using Kimura-2 parameter and bootstrapped at 1000 replicates. The phylogenetic tree confirmed that the sample sequences had close evolutionary relationships with the subtypes that they were assigned from RIP and jpHMM. A separate alignment and phylogenetic tree was constructed for BM236 due to its short length and it showed that it was closely related to A1.

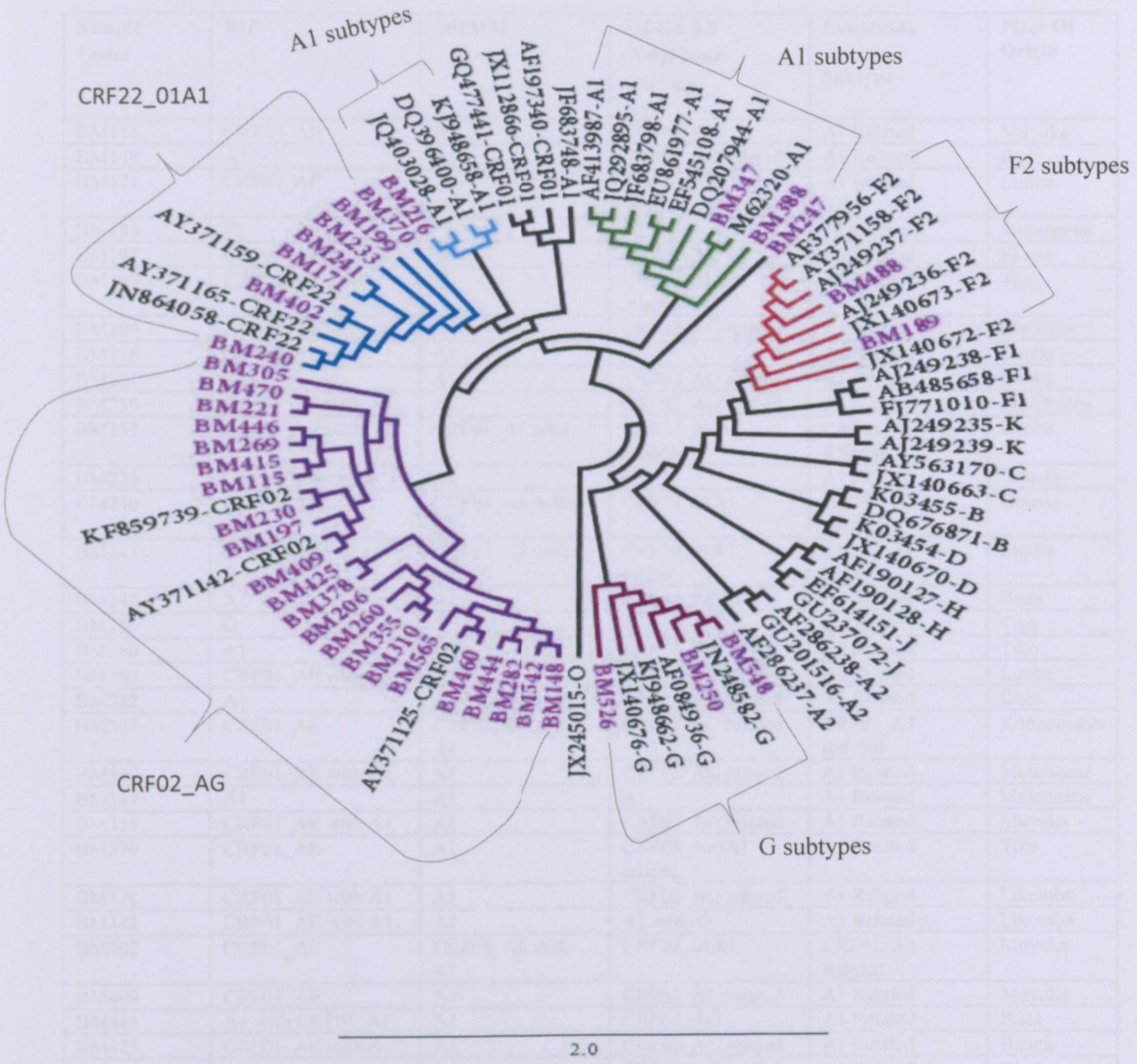


Figure 3.6: Circular layout of neighbour-joining phylogenetic tree of HIV-1 *gag* sequences derived from the study cohort.

The Neighbour Joining tree was generated using the Kimura 2-parameter method, bootstrapped to 1000 replicates and rooted with a group O reference sequence (JX245015). The novel *gag* sequences identified as "BM with a number" are presented in pink distributed among reference CRF02_AG, CRF22_01A1, subtypes G, and A1 and F2 sub-subtypes with published sequences retrieved from LANL HIV database in black. The tree was constructed using MEGA 6.0 and visually edited using Fig Tree v1.4.2

Table 3.3: Assignment of genotypes based on different tools using on partial *gag* sequences

Sample Codes	RIP	jpHMM	MEGA 6.0 (Neighbour-Joining)	Consensus Subtype	Place Of Origin
BM115	CRF01_AE	A1	CRF02_AG	A1 Related	Muyuka
BM148	A1	A1	CRF02_AG related	A1 Related	Buea
BM171	CRF01_AE	A1	CRF22_01A1 related	A1 Related	Limbe
BM189	F2	F2	F2	F2 Related	Mutengene
BM197	CRF01_AE with A1	A1	CRF02_AG	A1 Related	Ekona
BM199	CRF01_AE	A1	CRF22_01A1 related	A1 Related	Tiko
BM206	A1 with CRF01_AE	A1	CRF02_AG related	A1 Related	Likomba
BM216	CRF01_AE, A1	A1	A1	A1 Related	Limbe
BM221	A1	A1	CRF02_AG related	A1 Related	Limbe
BM230	CRF01_AE with A1	A1	CRF02_AG related	A1 Related	Mutengene
BM233	CRF01_AE with A1	CRF01_AE with A1	CRF22_01A1 related	CRF01_AE Related	Limbe
BM236	Cannot Determine	A1	A1	A1 Related	Douala
BM240	CRF01_AE	CRF01_AE with A1	CRF22_01A1	CRF01_AE Related	Douala
BM241	CRF01_AE	CRF01_AE with A1	CRF22_01A1 related	CRF01_AE Related	Limbe
BM247	A1	A1	CRF02_AG/G	A1 Related	Buea
BM250	G	G	G	G Related	Tiko
BM260	A1	A1	CRF02_AG related	A1 Related	Tiko
BM269	CRF01_AE with A1	A1	CRF02_AG related	A1 Related	Limbe
BM282	A1	A1	CRF02_AG related	A1 Related	Buea
BM305	CRF01_AE	CRF01_AE with A1	CRF02_AG related	CRF01_A1 Related	Kongpaynya
BM310	CRF01_AE with A1	A1	CRF02_AG related	A1 Related	Mutengene
BM347	A1	A1	A1	A1 Related	Mutengene
BM355	CRF01_AE with A1	A1	CRF02_AG related	A1 Related	Muyuka
BM370	CRF01_AE	A1	CRF01_AE/A1 related	A1 Related	Tiko
BM378	CRF01_AE with A1	A1	CRF02_AG related	A1 Related	Likomba
BM388	CRF01_AE with A1	A1	A1 related	A1 Related	Likomba
BM402	CRF01_AE	CRF01_AE with A1	CRF22_01A1	CRF01_A1 Related	Muyuka
BM409	CRF01_AE	A1	CRF02_AG related	A1 Related	Muyuka
BM415	A1 with CRF01_AE	A1	CRF02_AG	A1 Related	Buea
BM425	CRF01_AE with A1	A1	CRF02_AG related	A1 Related	Banga
BM444	A1	A1	CRF02_AG related	A1 Related	Buea
BM446	A1	A1	CRF02_AG	A1 Related	Pendaboko
BM460	CRF01_AE	A1	CRF02_AG	A1 Related	Buea
BM470	A1 with CRF01_AE	A1	CRF02_AG related	A1 Related	Likomba
BM488	F2	F2	F2	F2 Related	Buea
BM526	A1 and G	A1 and G	G related	G Related	Edenau
BM542	A1	A1	CRF02_AG related	A1 Related	Limbe
BM548	G	G	G related	G Related	Buea
BM565	A1	A1	CRF02_AG related	A1 Related	Buea

CRF = Circulating Recombinant Form

Based on analysis of *gag* gene from 39 patients, subtype A1-related strains were more commonly observed (n=29; 74.4%), followed by CRF01_AE (n=5; 12.8%), then subtype G-related (n=3; 7.7%) and subtype F2-related (n=2; 5.1%). Subtyping of genetically diverse viruses like HIV is challenging when subtyping is based on genetic analysis of one sub-

genome; for example partial *gag* in this case. Therefore, it is recommended that for accurate subtyping, at least three different sub-genomic regions be considered. Consequently, the subtype distribution did not meet the 3 gene regions criterion used to classify a sequence as a subtype or CRF (Robertson et al., 1999) hence subtyping only inferred a particular subtype or CRF but did not identify the sequences to be that subtype. The subtypes observed were distributed widely among the populations with no specific clustering according to geography.

Partial *pol* (complete protease and partial RT) and partial *env* (gp120 C2- C3) sequences for some of the samples that were subtyped for *gag* were available in the lab and used to check for discordance. Table 3.4 shows this additional data for 20 samples.

Table 3.4: Subtyping using protease, partial RT, partial *pol* and partial *gag* of HIV-1.

Sample ID	Consensus Protease	Consensus RT	Consensus Env	Consensus Gag	Possible recombinant
BM115	G	G	A1/A2	A1	CRF02_AG
BM189	F2	F2	A1/A2	F2	URF
BM199	A1	A1	A1	A1	CRF22_01A1
BM216	C	CRF02_AG	A1	A1	URF
BM221	A1/G	G	A1	A1	CRF02_AG
BM233	CRF01_AE	A1	A1	CRF01_AE	CRF22_01A1
BM236	G/A1	G	A1	A1	CRF02_AG
BM241	A1	A1	D (A1/H)	CRF01_AE	URF
BM260	C	C	A1	A1	URF (Similar to BM282 AND BM355)
BM282	C	C	A1	A1	URF (Similar to BM260 AND BM355)
BM310	G	G	A1	A1	CRF02_AG
BM355	C	C	A1	A1	URF (Similar to BM260 AND BM282)
BM370	G	G/A1	A1	A1	CRF02_AG
BM409	G/A1	G	A2	A1	CRF02_AG
BM415	G/CRF02_AG	G/A1	A2	A1	CRF02_AG
BM444	K	G/A1	A1	A1	URF
BM446	G	G, CRF02	A1	A1	CRF02_AG
BM460	G	G	A1	A1	CRF02_AG
BM470	G	G/A1	A1	A1	CRF02_AG
BM488	F2	F2	F2	F2	F2

URF is Unique Recombinant Forms.

Based on subtyping analysis of partial *gag*, *pol* and *env* genes from 20 patients, the dominant strain was CRF02_AG (n=10; 50.0%) followed by sequences that did not have known recombination patterns and are highly likely to be URFs (n=7; 35.0%), then CRF22_01A1 (n=2; 10.0%) and finally subtype F2 (n=1; 5.0%).

3.5 HIV-1 *gag* variation

A plot identity was constructed for *gag* p17 and p24 as shown in Figure 3.7 to Figure 3.14. The amino acid mismatches for each sequence was counted and a percentage difference was calculated out of the total number of amino acids within the sequence. *Gag* p17 showed an average of 16.48% amino acid difference in comparison to p24 that showed an average of 7.51%. The highest variation observed was for subtype G-like *gag* p17 sequences which was 18.95% amino acid difference and the lowest was for subtype F2-like *gag* p24 showing 3.23 % difference. CRF01_AE-like *gag* p17 displayed 17.05% difference, subtype A1-like *gag* p17 showed 16.46% difference and subtype F2-like *gag* p17 showed 11.58% difference. For *gag* p24, subtype A1-like displayed 7.77% amino acid difference, subtype G-like showed 10.22% difference whereas CRF01_AE-like showed 6.29% amino acid difference. Subtype G-like sequences showed the highest variation on p17 and p24 in comparison to the other subtypes, whereas subtype F2 showed the least variability on p17 and p24.

3.6 Cytotoxic T-lymphocyte epitopes

Table 3. 5 and Table 3. 6 shows the predicted epitope results for p17 and p24 respectively from IEDB MHC-1 binding tool and NetCTLpan tool. The predicted epitopes were mapped onto the plot identity of the subtypes or CRF they were predicted for, as shown on Figure 3.7 to Figure 3.14. The epitopes that gave a percentage rank less than 1 were discarded. There were some epitope motifs that were similar across subtypes and some motifs that were exactly the same. For p17, LVWASRELER which was 10 amino acids (10 aa) was predicted for sub-subtype A1 and F2 and RFALNPGLL (9 aa) was predicted for sub-subtype F2. The exact sequence for p24 epitope prediction were: MVHQPLSPR (9 aa) for sub-subtype F2 and CRF01_AE, KAFSPEVIPMF (11 aa) for sub-subtype A1, F2 and subtype G, FSPEVIPMF (9 aa) for sub-subtype A1, F2 and subtype G and FSPEVIPM for sub-subtype F2 and subtype G. There were

UNIVERSITY OF VENDA
LIBRARY

some predicted epitopes with similar motifs but differed in one or two amino acids within their sequences. For instance in p17 epitope prediction, there were 6 epitopes of varying lengths with a motif of VL-VHQR with the gap (-) in the sequence filled with C or Y. All these p17 motifs were restricted by HLA-A*33:03. Whereas for p24 epitope prediction, there were 8 epitopes of varying lengths with the T-QEQI-W motif whereby the first gap was filled with a L or P and the second gap with either a A, R or G. All these p24 motifs were predicted to be restricted by HLA-B*58:01 except epitope TPQEQIGW which was restricted by HLA-B*53:01.

The epitopes were examined against the HIV molecular immunology database http://www.hiv.lanl.gov/content/immunology/tables/optimal ctl summary.html to determine if the epitopes predicted had been experimentally determined to be optimal CTL epitopes. For p17 epitope TLYCVHQR which was predicted to be restricted by HLA-A*33:03, had been shown to be optimally restricted by HLA-A*11:01 within the database. RSLYNTVATLY was predicted to be restricted by HLA-B*58:01 and shown to be optimally restricted by HLA-B58:01. LVWASRELER was predicted to be restricted by HLA-A*33:03 and shown to be restricted by HLA A*30. For p24 epitope TSTLQEQIGW was predicted to be restricted by HLA-B*58:01 whereas it was described to be optimally restricted by HLA-B*57:01 and HLA-B*58:01. KAFSPEVIPMF was predicted to be restricted by HLA-B*58:01 whereas it was described to be optimally restricted by HLA-B*57:01 and HLA-B*57:03. QAISPRTLNAW was predicted to be restricted by HLA-B*53:01 whereas it was described to be optimally restricted by HLA-A*25:01. FSPEVIPMF was predicted to be restricted by HLA-C*02:02 whereas it was described in literature to be optimally restricted by HLA-B*57.

Figure 3. 7: P17 epitope map and amino acid plot identity for CRF01_AE like *gag* sequences.

The predicted epitopes for CRF01_AE p17 gag sequences mapped onto CRF01_AE p17 plot identity with the allele predicted and the HLA allele frequency within the coastal population (in percentage) shown in brackets next to the allele.

*Denotes positions of disagreement among consensus BM and consensus CRF01_AE and . denotes positions of agreeement.

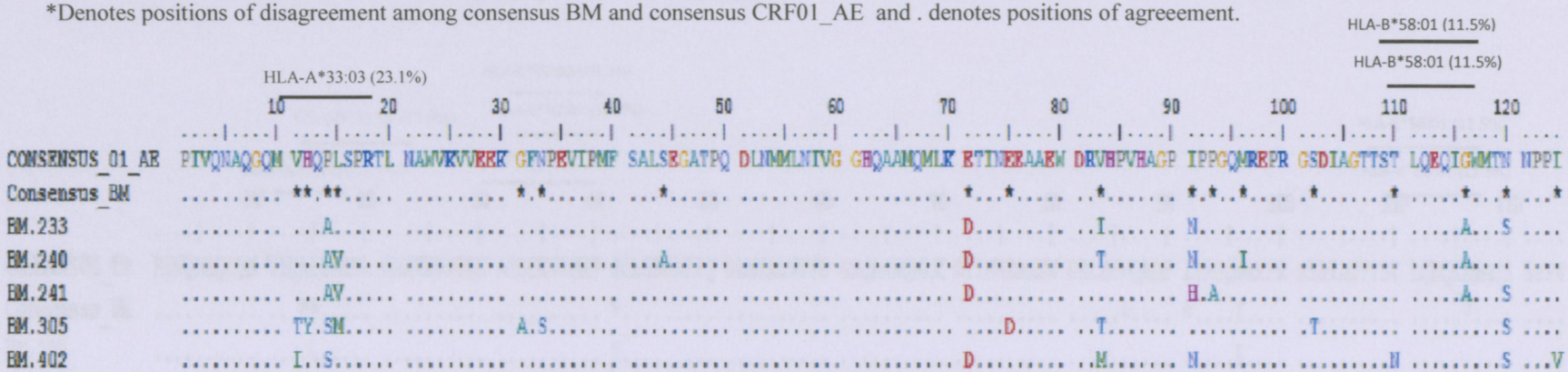


Figure 3. 8: P24 epitope map and amino acid plot identity for CRF01_AE like *gag* sequences.

The predicted epitopes for CRF01_AE p24 gag sequences mapped onto CRF01_AE p24 plot identity with the allele predicted and the HLA

allele frequency within the coastal population (in percentage) shown in brackets next to the allele.

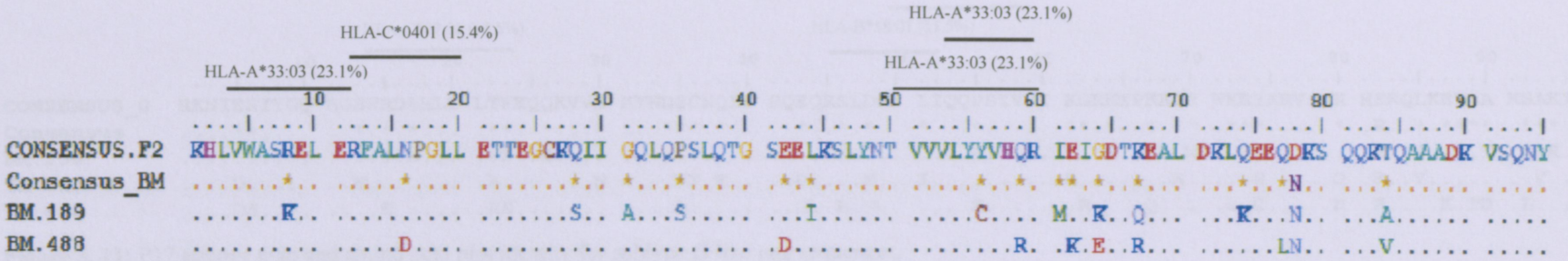


Figure 3. 9: P17 epitope map and amino acid plot identity for subtype F2 like *gag* sequences.

The predicted epitopes for subtype F2 p17 gag sequences mapped onto subtype F2 p17 plot identity with the allele predicted and the HLA allele frequency within the coastal population (in percentage) shown in brackets next to the allele.

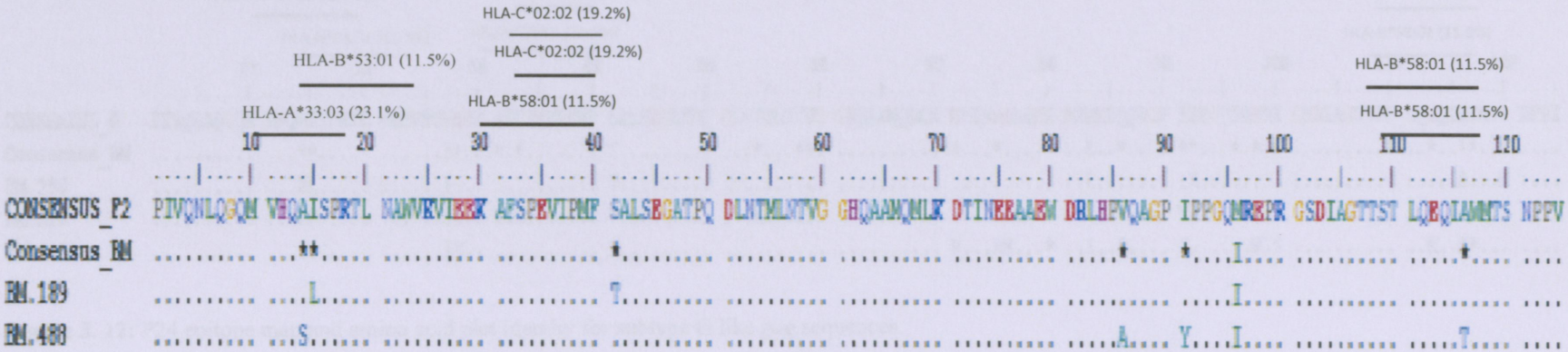


Figure 3. 10: P24 epitope map and amino acid plot identity for subtype F2 like *gag* sequences.

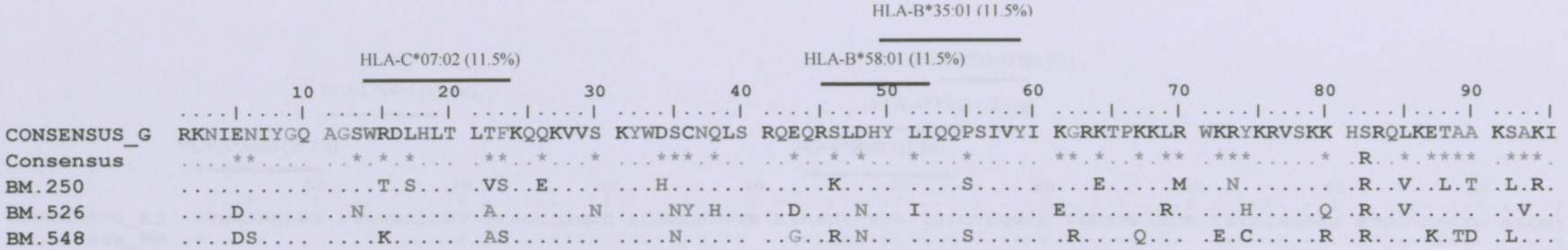


Figure 3. 11: P17 epitope map and amino acid plot identity for subtype G like *gag* sequences.

The predicted epitopes for subtype G p17 gag sequences mapped onto subtype G p17 plot identity with the allele predicted and the HLA allele frequency within the coastal population (in percentage) shown in brackets next to the allele.

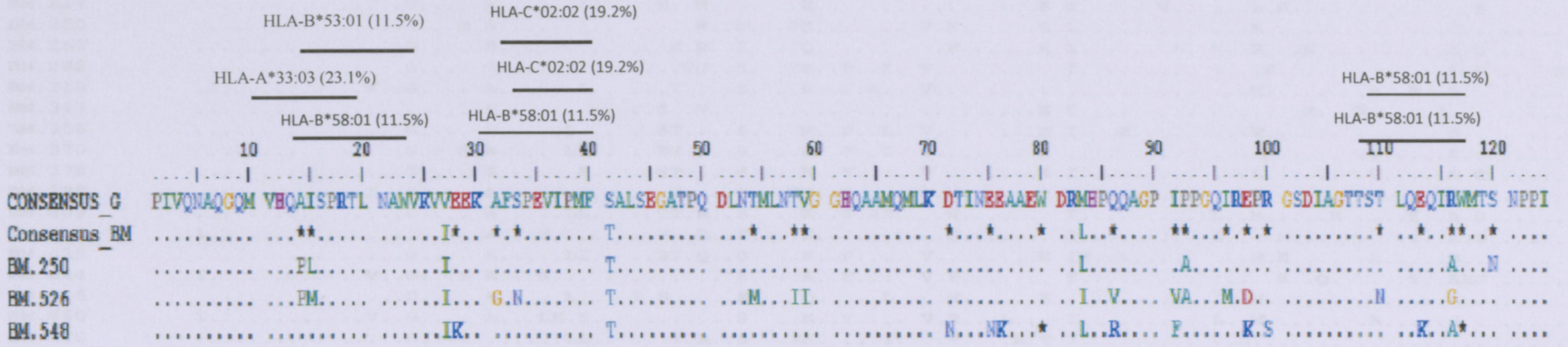


Figure 3. 12: P24 epitope map and amino acid plot identity for subtype G like *gag* sequences.

The predicted epitopes for subtype G p24 gag sequences mapped onto subtype G p24 plot identity with the allele predicted and the HLA allele frequency within the coastal population (in percentage) shown in brackets next to the allele.

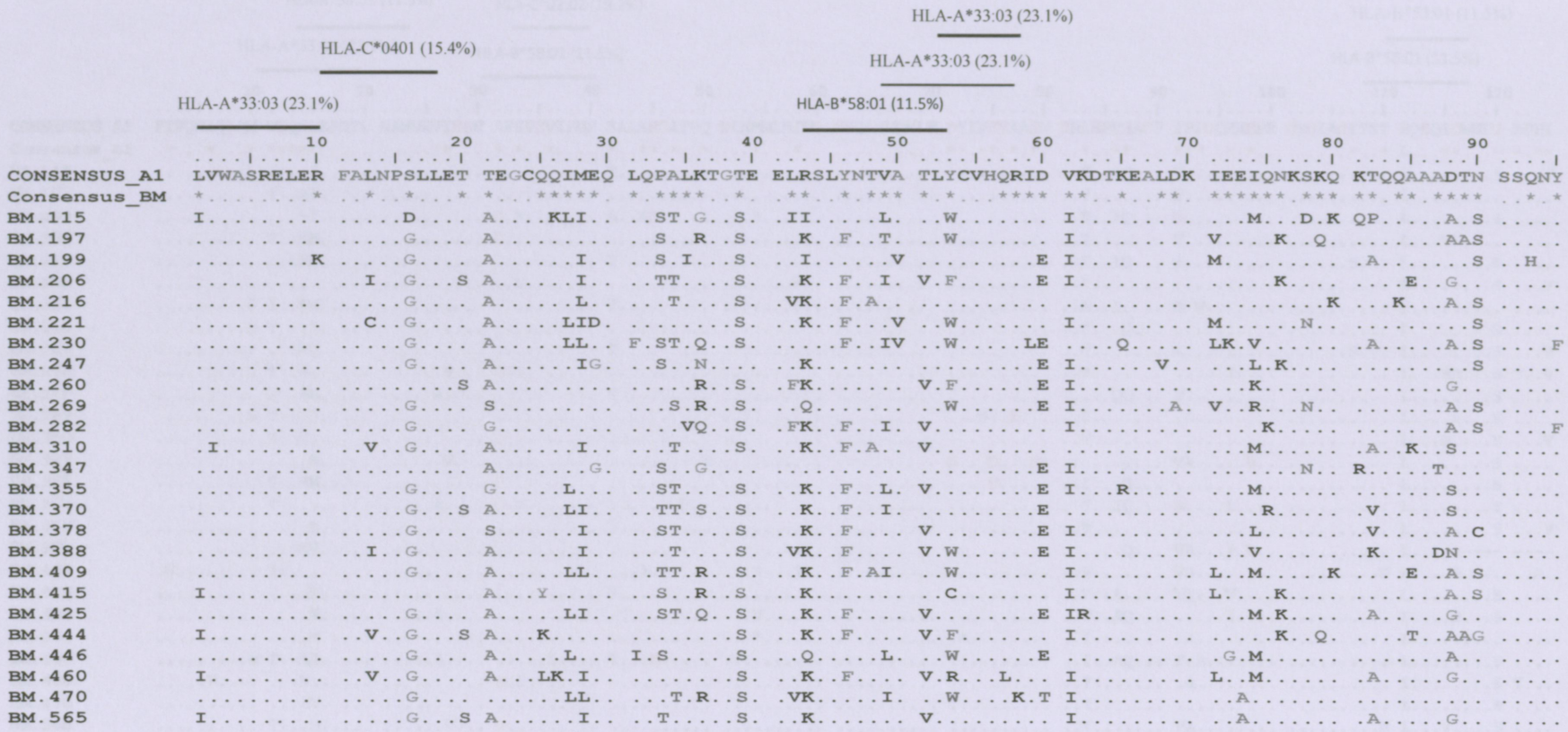


Figure 3. 13: P17 epitope map and amino acid plot identity for subtype A1 like *gag* sequences.

The predicted epitopes for subtype A1 p17 gag sequences mapped onto subtype A1 p17 plot identity with the allele predicted and the HLA allele frequency within the coastal population (in percentage) shown in brackets next to the allele.

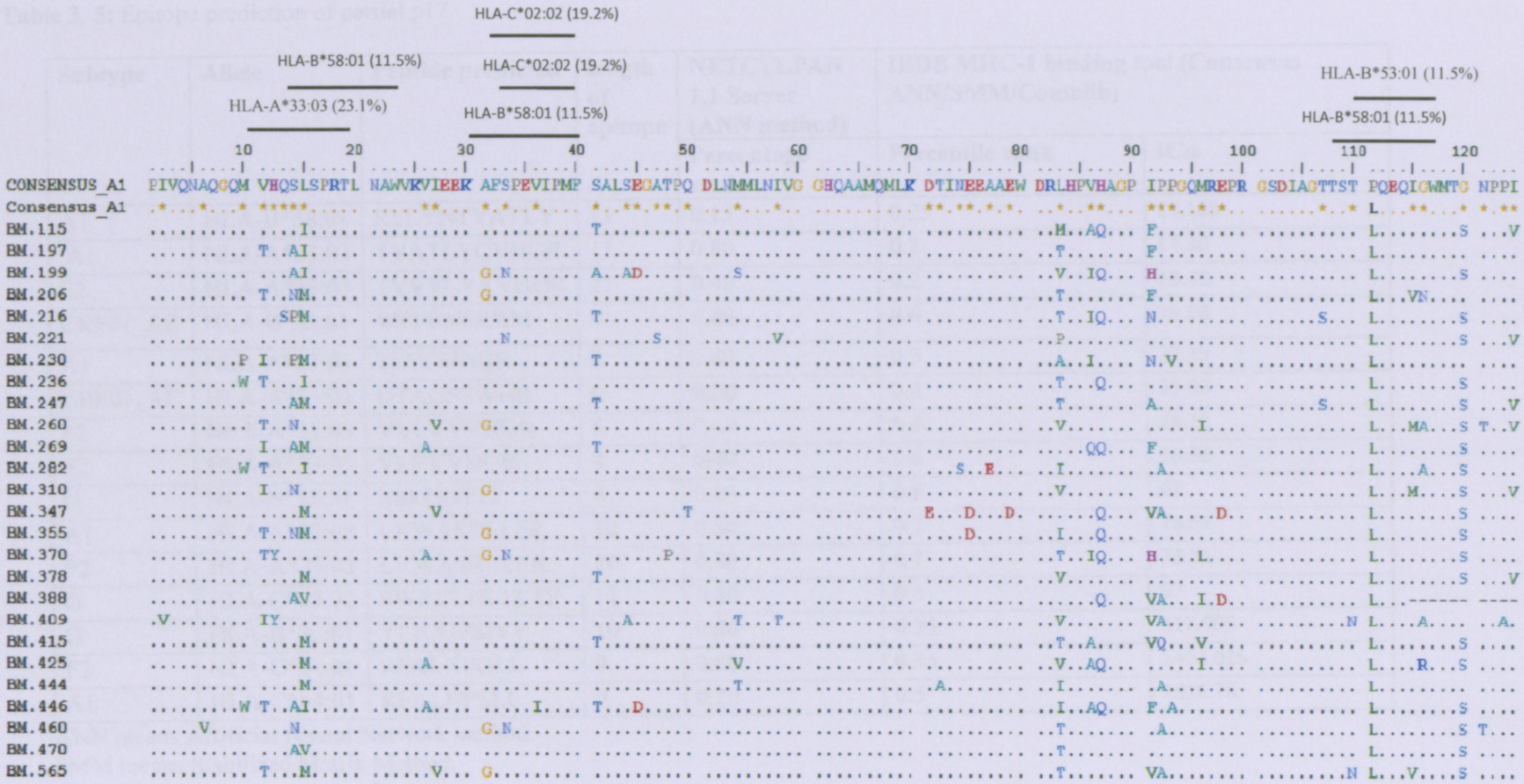


Figure 3. 14: P24 epitope map and amino acid plot identity for subtype A1 like *gag* sequences.

The predicted epitopes for subtype A1 p24 gag sequences mapped onto subtype A1 p24 plot identity with the allele predicted and the HLA allele frequency within the coastal population (in percentage) shown in brackets next to the allele.

Table 3. 5: Epitope prediction of partial p17

Subtype	Allele	Peptide predicted	length of epitope	NETCTLPAN 1.1 Server (ANN method)	IEDB MHC-1 binding tool (Consensus ANN/SMM/Comblib)	
				Percentage rank	Percentile rank	IC_{50}
A1	HLA-B*58:01	RSLYNTVATLY	11	0.15	0.25	13.125
A1	HLA-A*33:03	TVATLYCVHQR	11	0.80	0.2	14.30
F2	HLA-A*33:03	TVVVLYYVHQR	11	0.40	0.2	18.45
CRF01_AE	HLA-B*35:01	WPFRNSRRM	9	0.80	0.6	19.68
A1	HLA-A*33:03	TLYCVHQR	8	0.40	0.3	22.19
CRF01_AE	HLA-A*33:03	DTAGSSWHR	9	0.20	0.4	25.67
F2	HLA-A*33:03	VVLYYVHQR	9	0.40	0.4	28.12
F2	HLA-A*33:03	VLYYVHQR	8	0.40	0.6	56.56
G	HLA-B*58:01	RSLDHYLI	8	0.80	0.8	67
A1	HLA-A*33:03	LVWASRELER	10	0.80	0.7	73.51
F2	HLA-A*33:03	LVWASRELER	10	0.80	0.7	73.51
G	HLA-C*07:02	SWRDLHLTLTF	11	0.80	0.3	97
G	HLA-B*35:01	YLIQQPSIVY	10	0.80	0.35	117.005
F2	HLA-C*04:01	RFALNPGLL	9	0.80	0.35	1494.015
A1	HLA-C*04:01	RFALNPSLL	9	0.20	0.5	1816.78

ANN means Artificial Neural Network method.
SMM means Stabilized Matrix Method.
IC_{50} = inhibitory concentration 50.

Table 3. 6: Epitope prediction of partial p24

Subtype	Allele	Peptide predicted	length of epitope	NETCTLPAN 1.1 Server (ANN method)	IEDB MHC-1 binding tool (Consensus ANN/SMM/ Comblib)	
				Percentage rank	Percentile rank	IC_{50}
CRF 01_AE	HLA-A*33:03	MVHQPLSPR	9	0.15	0.2	7.57
F2	HLA-A*33:03	MVHQAISPR	9	0.15	0.2	7.99
A1	HLA-A*33:03	MVHQSLSPR	9	0.20	0.2	8.48
A1	HLA-B*58:01	QSLSPRTLNAW	11	0.80	0.35	31.27
G	HLA-B*58:01	QAISPRTLNAW	11	0.15	0.55	35.93
G	HLA-B*58:01	TSTLQEQIRW	10	0.30	0.25	52.03
CRF01_AE	HLA-B*58:01	STLQEQIGW	9	0.15	0.6	55.52
A1	HLA-B*58:01	KAFSPEVIPMF	11	0.10	0.5	57.255
F2	HLA-B*58:01	KAFSPEVIPMF	11	0.10	0.5	57.255
G	HLA-B*53:01	QAISPRTLNAW	11	0.10	0.95	57.255
F2	HLA-B*53:01	QAISPRTLNAW	11	0.10	0.95	62.895
F2	HLA-B*58:01	STLQEQIAW	9	0.80	0.7	65.715
G	HLA-B*58:01	KAFSPEVIPMF	11	0.80	0.5	65.715
F2	HLA-B*58:01	TSTLQEQIAW	10	0.30	0.25	74.765
G	HLA-B*58:01	STLQEQIRW	9	0.30	0.4	76.145
CRF01_AE	HLA-B*58:01	TSTLQEQIGW	10	0.30	0.25	78.675
A1	HLA-C*02:02	FSPEVIPM	8	0.40	0.2	144.26
F2	HLA-C*02:02	FSPEVIPMF	9	0.40	0.4	144.26
G	HLA-C*02:02	FSPEVIPMF	9	0.40	0.4	144.26
A1	HLA-B*58:01	TSTPQEQIGW	10	0.30	0.4	190.965
A1	HLA-C*02:02	FSPEVIPMF	9	0.30	0.4	213.04

G	HLA-A*33:03	MVHQAISPR	9	0.30	0.2	213.04
F2	HLA-C*02:02	FSPEVIPM	8	0.80	0.2	1073.765
G	HLA-C*02:02	FSPEVIPM	8	0.80	0.2	1073.765
A1	HLA-B*53:01	TPQEQIGW	8	0.30	0.5	3898.79

ANN means Artificial Neural Network method.
SMM means Stabilized Matrix Method.
IC50 = inhibitory concentration 50.

3.7 Amplification of near full-length sequence

Near full-length amplification was attempted on five samples. Samples BM488 was selected because it seemed to be a pure F2 subtype based on the assignment of F2 subtype for the partial *gag*, protease, partial RT and partial *env* for the sample. It was of interest because there are few near full-length F2 subtypes that have been so far described in GenBank. BM189 was selected as it seemed to have mostly F2 within the HIV-1 structural genes studied with A subtype assigned to its partial *env* analysis. BM260, BM282 and BM355 were selected for further study due to their unique recombinant nature that had not been observed among known recombinants observed within the HIV Los Alamos National Laboratory Circulating Recombinant Forms database. Furthermore, BM260, BM282 and BM355 were assigned subtype C within the *pol* gene which is rarely observed in Cameroon and it could have suggested a possible introduction of subtype C into Cameroon as a recombinant of this subtype within the study site.

PCR was successful for 4 out of the five samples (BM189, BM260, BM282 and BM355) with each having amplicons for 5' fragment and 3' fragment. For the fifth sample (BM488), only the 3' fragment was amplified. Figure 3.15 is a representative gel of nested products of both 5' and 3' fragment amplification of the 5 samples.

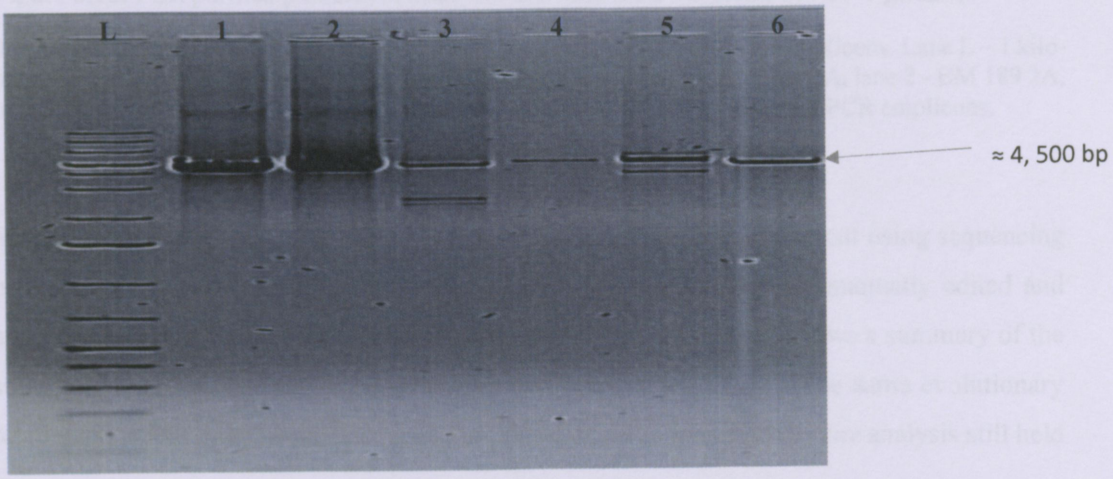


Figure 3.15: Amplification of about 4.5 kilo-base pairs 5' and 3' fragments of HIV-1 genome.

0.8% ethidium bromide stained agarose gel electrophoresis of drug-naïve HIV-1 positive patients' DNA after nested PCR. Lane L - I kilo-base pair plus Ladder (ThermoFisher Scientific GeneRuler), lane 1 - BM 189 2A, lane 2 - BM 260 2A, lane 3 - BM 282 2A, lane 4 - BM 282 2B, lane 5 - BM 260 2B and lane 6 - BM 189 2B nested PCR amplicons of patient DNA.

Most of the PCR products had secondary undesired bands upon viewing the gel that had to be gel extracted and purified. Figure 3.16 is a representative gel of the purified PCR amplicons.

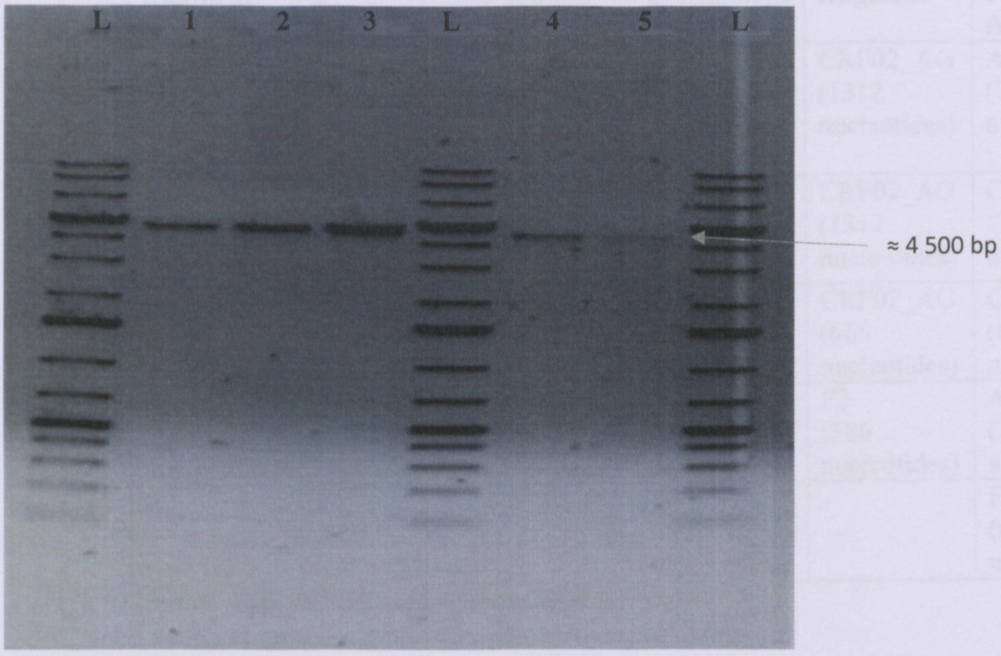


Figure 3.16: PCR purified products of about 4.5 kilo-base pairs fragments of HIV-1 genome.

0.8% ethidium bromide stained agarose gel electrophoresis of PCR purified amplicons. Lane L – 1 kilo-base pairs plus Ladder (ThermoFisher Scientific GeneRuler), lane 1 - BM 282 2A, lane 2 - BM 189 2A, lane 3 - BM 260 2A, lane 4 - BM 189 2B and lane 5 - BM 282 2B are purified PCR amplicons.

Sanger sequencing was done for the four samples and BM488 3' fragment using sequencing primers that targeted partial *pol* and partial *env*. The sequences were manually edited and analyzed with the same tools as were used for partial *gag*. Table 3.7 shows a summary of the subtyping analysis. A neighbor-joining tree was constructed to see if the same evolutionary relationships that were observed during the initial partial *pol* and partial *env* analysis still held true.

Table 3.7: Summary of subtyping from 5' and 3' fragments.

Sample ID	GENOTYPING (RIP)		GENOTYPING (JpHMM)		GENOTYPING (MEGA 6.0 Neighbour-joining)	
	***Pol* from 5' fragment**	***Env* from 3' fragment**	***Pol* from 5' fragment**	***Env* from 3' fragment**	***Pol* from 5' fragment**	***Env* from 3' fragment**
BM260	G and C (1312 nucleotides)	Cannot be determine (769 nucleotides)	A1 and G (1312 nucleotides)	N/A (769 nucleotides)	CRF02_AG (1312 nucleotides)	A1 (769 nucleotides)
BM282	A1 and G (1312 nucleotides)	A1 (769 nucleotides)	A1 and G (1312 nucleotides)	A1 AND J (769 nucleotides)	CRF02_AG (1312 nucleotides)	CRF02_AG (769 nucleotides)
BM355	G and C (665 nucleotides)	A1 (642 nucleotides)	G and A1 (665 nucleotides)	A1 (642 nucleotides)	CRF02_AG (665 nucleotides)	CRF02_AG (642 nucleotides)
BM189	F2 (386 nucleotides)	A1 and B (278 nucleotides)	K and F1 (386 nucleotides)	N/A (278 nucleotides)	F2 (386 nucleotides)	A1 (278 nucleotides)
BM488	-	F2 (610 nucleotides)	-	F2 (610 nucleotides)	-	F2 (610 nucleotides)

Cannot determine: Sample was untypeable on RIP.
N/A (Not Available) means sample was untypeable by jpHMM.
- Means the 5' fragment could not be amplified.

3.8 Genome subtyping

A near full-length sequence of BM189 was obtained. The sequence was 9077 nucleotides (589 to 9666 relative to HXB2) which covered the middle region of the 5' long terminal repeat unique 5 (5' LTR U5) to its end (that is positions 589 to 634 relative to HXB2), complete *gag*, complete *pol*, complete *vif*, complete *vpr*, complete *tat*, complete *rev*, complete *vpu*, complete *env*, complete *nef* and beginning of 3' long terminal repeat unique 5 (3' LTR U5) to the middle region (that is positions 9086 to 9077 relative to HXB2). The sequence passed QC showing no hypermutations or stop codons (http://www.hiv.lanl.gov/content/sequence/QC/index.html).

A sequence search on National Center of Biotechnology Information Basic Local Alignment Search Tool (NCBI BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) revealed 88% identity with a A/F2/G recombinant isolate (Accession number: AF377958.1) obtained from a 37 year old male from Yaoundé, Cameroon collected in 1997. It also revealed a 87% similarity with subtype F molecular clone (AB480298.1) obtained from Brazil in 2004, 87% similarity with

subtype F1 isolate (DQ979025.1) from Angola and 87% similarity with a BF recombinant isolate (AF332867.1) from Argentina. HIV BLAST results from the HIV database (http://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html) revealed a 89% identity with a sub-subtype F2 isolate (Accession number: JX140672) from Cameroon collected in 2010, a 89% identity with a sub-subtype F2 isolate (AF377956) collected from a 40 year old male from Douala, Cameroon in 1997. 88% identity with a sub-subtype F2 isolate (JX140673) from Cameroon collected in 2010 and 88% identity with a sub-subtype F1 isolate (AY173957 and AB485656) and a sub-subtype F1 clone from Brazil collected in 1989 and 1990 respectively. All the mentioned blast result alignments gave a 0.0 Expected (E) value and covered regions beginning from *gag* to end of *nef* that is nucleotide positions 790 to 9417 relative to HXB2.

Clustal W was used to align the genome sequence of BM189 to a representative subtype set retrieved from the LANL website (http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html). Figure 3. 17 shows a neighbor-joining tree that shows BM189 clustering with subtype F. Recombination analysis of BM189 genome was done using jpHMM. It was observed that the sequence was a recombinant virus made up of F2, CRF01_AE and A1. The sequence had 3 recombination breakpoints alternating between the parental sub-subtype F2 and CRF01_AE and A1. The first breakpoint was observed in the *env* gene at position ±8 nucleotides from 6324 relative to HXB2 and marked the beginning of CRF01_AE. The second breakpoint was at position ±52 nucleotides from position 6679 relative to HXB2 and marked the end of CRF01_AE sequence and beginning of the A1 sequence. The third breakpoint was ±47 nucleotides from position 8335 and it marked the end of the A1 sequence within the *env* gene. Posterior probability analysis confirmed the subtypes were observed with 99% confidence. There were some regions of uncertainty within *gag* (positions ±101 nucleotides from 2061 relative to HXB2), *nef* (positions ±218 nucleotides from 9034 relative to HXB2) and 3' LTR (nucleotide positions 9407 to 9666 relative to HXB2) that could not be typed and were labelled as N/A as shown in Figure 3. 18.

Simplot analysis was done using genome alignment of BM189 with HIV-1 pure subtype reference sequences. Bootscan analysis showed 8 recombination breakpoints. The first breakpoint at position 4438, second at position 4500, third at 6125, fourth at 6625, fifth at 8125, sixth at 8250, seventh at 9000 and the eighth breakpoint at 9125. The subtypes identified were

sub-subtype F2 at positions 250 to 4438, subtype G at positions 4489 to 4500, sub-subtype F2 at positions 4501 to 6125, sub-subtype C and A1 at positions 6126 to 6625, sub-subtype A1 at positions 6626 to 8125, subtype G at positions 8126 to 8250, sub-subtype A1 at positions 8251 to 9000 and sub-subtype F1 at positons 9001 to 9125 and subtype C at 9126 to 9250. The subtypes that had more than 75% of permutated trees were sub-subtypes F2 and A1. Bootscan values were exported as comma separated values (.csv) file and the highest score at each position was selected and used to annotate the sequence. The annotation of the sequence was used to construct a gene map of the BM189 sequence as shown in Figure 3. 20 relative to HXB2 using HIV-1 Recombinant Drawing Tool from the HIV database (http://www.hiv.lanl.gov/cgi-bin/DRAW_CRF/crfmap.cgi)

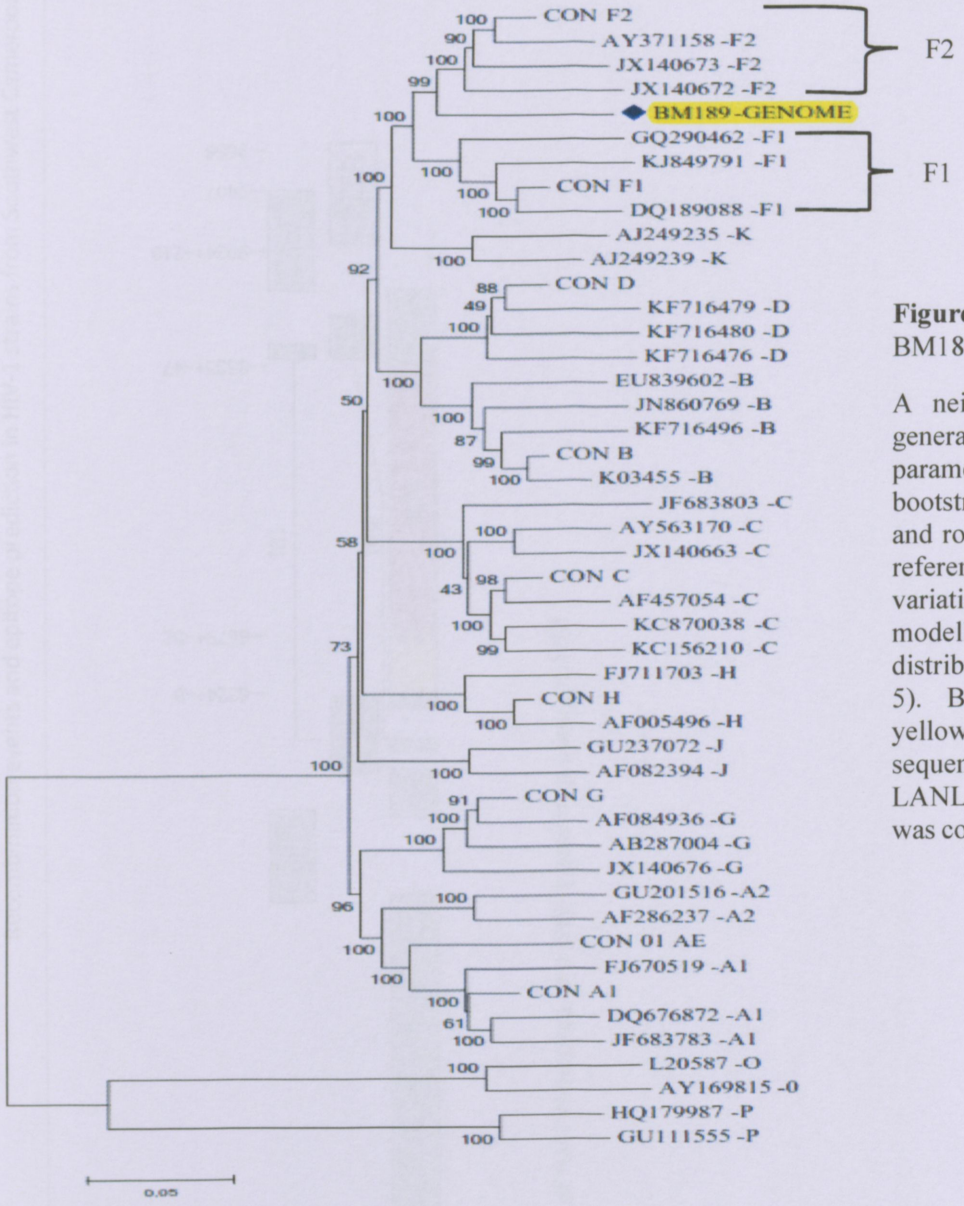


Figure 3. 17: Phylogenetic tree of BM189 genome sequence.

A neighbour Joining tree was generated using the Kimura 2-parameter method and bootstrapped to 1000 replicates and rooted with a group O and P reference sequence. The rate variation among sites was modelled with a gamma distribution (shape parameter = 5). BM189 is highlighted in yellow with published reference sequences retrieved from HIV LANL database in black. The tree was constructed using MEGA 6.0.

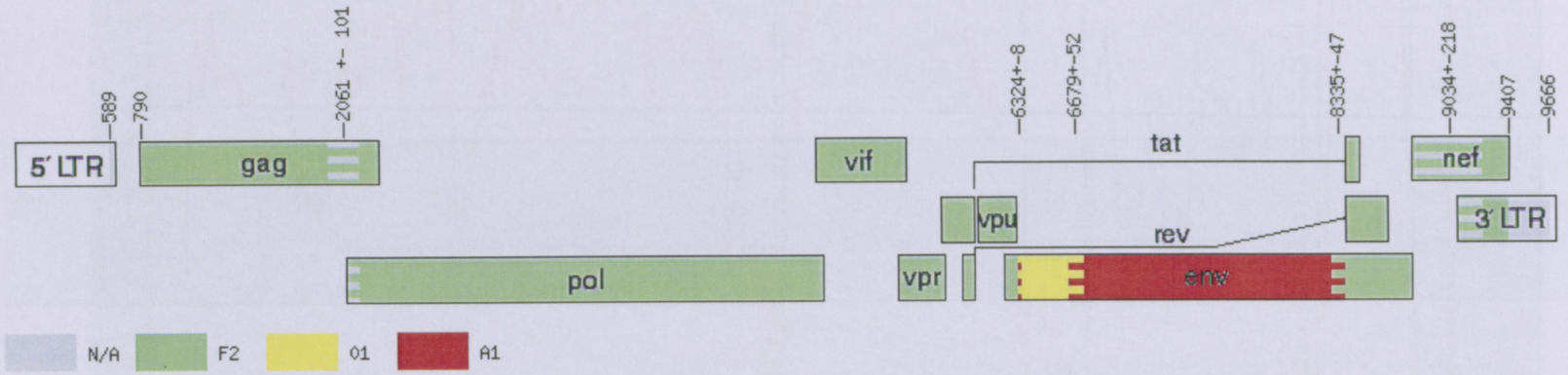


Figure 3. 18: Recombination breakpoints of BM189 mapped on a HIV-1 gene map using jpHMM.

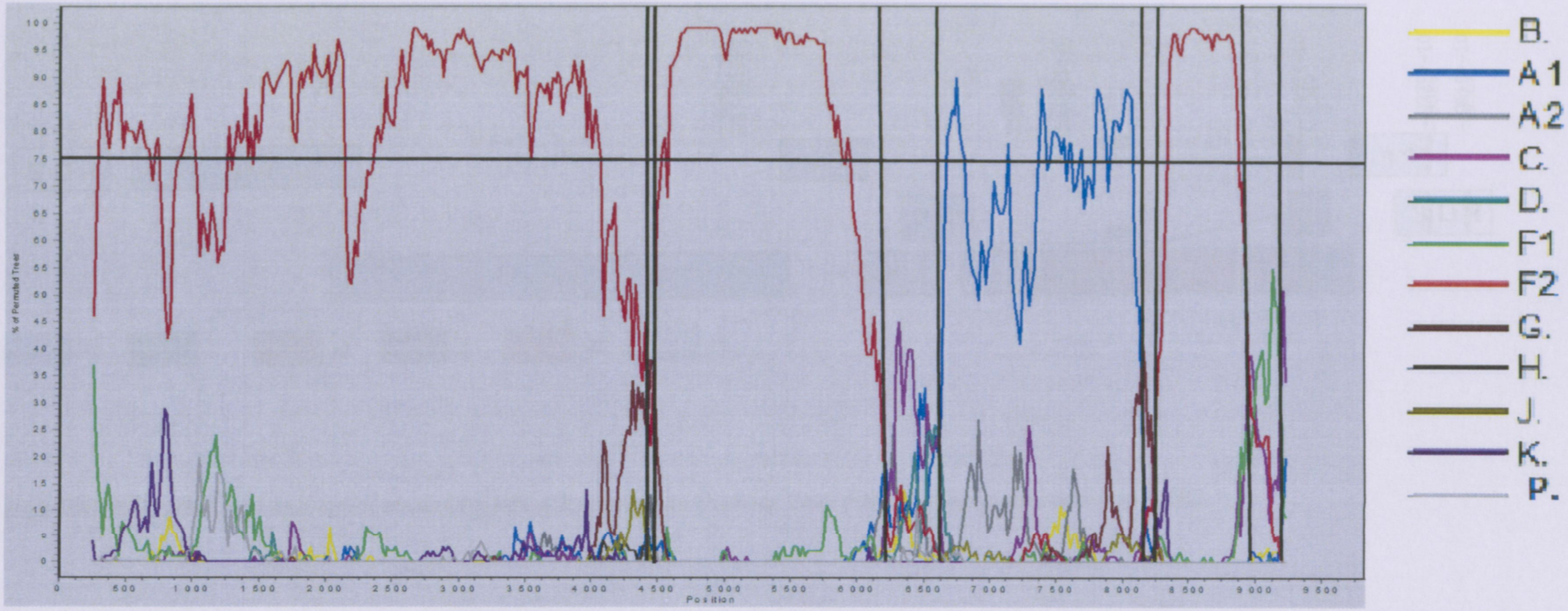


Figure 3. 19: Bootscan analysis of BM189 genome using Simplot.

The bootscan analysis was performed using a strict consensus with a window size of 500 bp and a step size of 20 bp using Kimura-2 parameter method with bootstrap of 1000 replicates. The black vertical lines across the graph marks the positions at which recombinant breakpoints occurred. The black horizontal line splits the graph at the 75% mark and the region above the line indicates the sub-subtypes that appeared in more than 75% of permutated trees. The legend for the colored graph is on the far right.

UNIVERSITY OF VENDA
LIBRARY

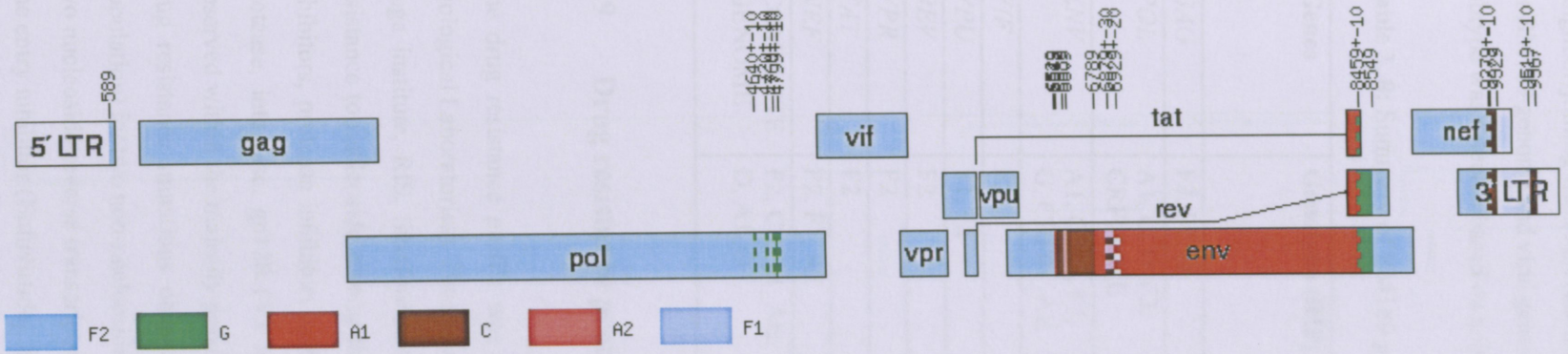


Figure 3. 20: BM189 gene map constructed using HIV-1 Recombinant Drawing Tool. Positions shown are relative to HXB2.

Neighbor-joining trees and recombination analysis using RIP, jpHMM and Simplot was done for BM189 genome and viral genes nucleotide sequences as shown in Table 3. 8. The consensus subtype was selected based on subtypes that were identified by more than 3 tools.

Table 3. 8: Summary of BM189 genome and viral genes analysis

Genes	Genotype (RIP)	Genotype (JpHMM)	Genotype (MEGA 6) Neighbour Joining	Simplot	Consensus subtype
GAG	F2, F1, D	F2	F2	F2, F1	F2
POL	A1, C, D, F2, CRF01_AE	F2	F2	F2	F2
ENV	A1, C, F1, F2, G, CRF01_AE	A1, F2, CRF01_AE	CRF01_AE	A2, A1, F2	A1, F2, CRF01_AE
VIF	F2	F2	F1, F2	F2	F2
VPU	F1, F2	F1	F2	F2	F2
REV	F2	F2	F2	F2	F2
VPR	F2	F2	F2	F2	F2
TAT	F2	F2	F2	F2	F2
NEF	F2, F1, A1	F1, F2, N/A	F2, F1	F2, F1	F2, F1
COMPLETE GENOME	F2, CRF01_AE, D, A1, G	F2, CRF01_AE AND A1	F1/F2	A1, F1, F2, G, C	F2, A1

3.9 Drug resistance profile

The drug resistance profile was generated using DeepChek-HIV software 1.4 (Advanced Biological Laboratories). The software uses Centre Hospitalier de Luxembourg, ANRS, Grade, Rega institute, RIS, Stanford and RenaGeno algorithms independently to determine drug resistance to nucleoside reverse transcriptase inhibitors, non-nucleoside reverse transcriptase inhibitors, protease inhibitors and entry inhibitors from the translated sequence data of RT protease, integrase, gp120 (V3 loop) and gp41. There was no drug resistance mutations observed within the majority population (> 20% viral population). However, there were some drug resistance mutations observed for the minority population (< 5% and < 1% viral population) for two non-nucleoside reverse transcriptase inhibitors (Efavirenz and Nevirapine), two nucleoside reverse transcriptase inhibitors (Emtricitabine and Lamivudine) as well as for one entry inhibitor (Enfuvirtide). See Table 3. 9 to Table 3. 11.

For viral populations less than 5% and 1 %, Centre Hospitalier de Luxembourg, ANRS, Grade, Rega institute, RIS, Stanford and RenaGeno were concordant for Lamivudine (a Nucleoside Reverse Transcriptase inhibitors) resistance, Efavirenz and Nevirapine (both non-Nucleoside Reverse Transcriptase Inhibitors) resistance. All algorithms apart from RenaGeno agreed on Emtricitabine (a Nucleoside Reverse Transcriptase inhibitors) resistance. ANRS, Grade, Rega institute and RIS agreed on Enfuvirtide (an Entry Inhibitor) resistance for less than 5% and 1% viral populations whereas Rega institute only agreed to Enfuvirtide resistance within less than 1% population and Centre Hospitalier de Luxembourg, Stanford and RenaGeno showed no resistance at the 5% or 1% population. There were no mutations associated with resistance for protease inhibitors within all viral populations.

Table 3. 9: Viral populations at which drug resistance was present for nucleoside reverse transcriptase inhibitors.

Nucleoside Reverse Transcriptase inhibitors	**Algorithm**	**>5.00% viral population**	**>1.00% viral population**	**Mutations within sequence associated with resistance**
Emtricitabine (FTC)	Centre Hospitalier de Luxembourg	R	R	M184V
	ANRS	R	R	
	Grade	R	R	
	Rega institute	R	R	
	RIS	R	R	
	Stanford	R	R	
	RenaGeno	NA	NA	
Lamivudine (3TC)	Centre Hospitalier de Luxembourg	R	R	M184V
	ANRS	R	R	
	Grade	R	R	
	Rega institute	R	R	
	RIS	R	R	
	Stanford	R	R	
	RenaGeno	R	R	

R means Resistance for Centre Hospitalier de Luxembourg, ANRS, Grade, RIS, RenaGeno, Rega institute and Stanford. NA means that algorithm could not determine susceptibility or resistance.

Table 3. 10: Viral populations at which drug resistance was present for non-nucleoside reverse transcriptase inhibitors.

Non-Nucleoside Reverse Transcriptase Inhibitors	Algorithm	>5.00% viral population	>1.00% viral population	Mutations within sequence associated with resistance
Efavirenz (EFV)	Centre Hospitalier de Luxembourg	R	R	K103N
	ANRS	R	R	
	Grade	R	R	
	Rega institute	R	R	
	RIS	R	R	
	Stanford	R	R	
	RenaGeno	R	R	
Nevirapine (NVP)	Centre Hospitalier de Luxembourg	R	R	K103N
	ANRS	R	R	
	Grade	R	R	
	Rega institute	R	R	
	RIS	R	R	
	Stanford	R	R	
	RenaGeno	R	R	

R means Resistance for Centre Hospitalier de Luxembourg, ANRS, Grade, RIS, RenaGeno, Rega institute and Stanford.

Table 3. 11: Viral populations at which drug resistance was present for entry inhibitors.

HIV Entry Inhibitor	Algorithm	>5.00% viral population	>1.00% viral population	Mutations within sequence associated with resistance
Enfuvirtide (INN)	Centre Hospitalier de Luxembourg	NA	NA	G36S
	ANRS	R	R	
	Grade	R	R	
	Rega institute	I	R	
	RIS	R	R	
	Stanford	NA	NA	
	RenaGeno	NA	NA	

R means Resistance for Centre Hospitalier de Luxembourg, ANRS, Grade, RIS, RenaGeno, Rega institute and Stanford. NA means that algorithm could not determine susceptibility or resistance.

3.10 Co-receptor and CTL epitope prediction for BM189

BM189 V3 loop sequence was queried for predictions on co-receptor usage using web PSSM and gene2pheno tools. Web PSSM predicted a 0.85 percentile score which was below 0.95. This indicated that the sequence queried was indeed a V3 loop sequence. It further predicted a 0 predictive score indicating that BM189 was a R5-tropic or a non-syncytium inducing virus. The sequence had Glycine (G) and Aspartic Acid (D) as amino acid residues at V3 sites 11 and 25 as shown with the black box on Figure 3. 21. Usually basic amino acid residues (that is Histidine, Lysine or Arginine) at either or both of these sites is predictive of CXCR4 use. The presence of G and D at these positions indicated an R5-tropic virus. The net charge of BM189 V3 loop sequence was +4. A net charge of 5 or higher is indicative of an X4-tropic virus. Gene2pheno predicted with a 65% confidence that the sequence had a low risk of X4 emergence. It further suggested that CCR5 antagonists like Maraviroc (Celsentri/Selzentry) may be effective.

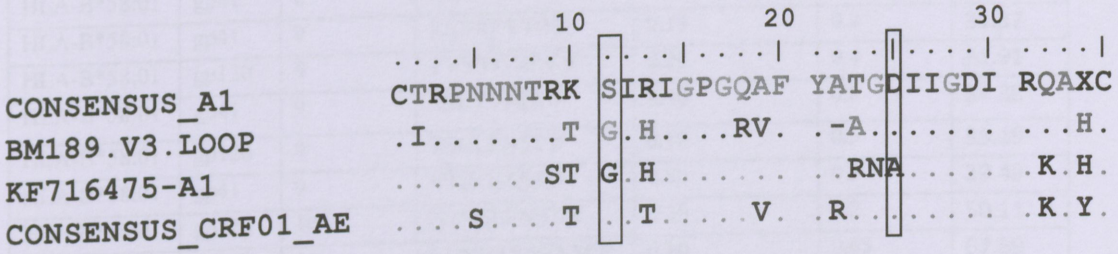


Figure 3. 21: Plot identity for V3 loop.

KF716475-A1 = Sequence that gave highest similarity on a NCBI BLAST search of BM189 V3 loop nucleotide sequence.

The *gag* and *env* genes were queried for epitopes restricted by frequently occurring HLA alleles within the study region and those that gave percentile scores below 1 were considered as the best binders as shown in tables Table 3. 12 and Table 3. 13. There were some similar but not identical epitopes with differing epitope lengths. The epitopes were WSSKTYNEIW and SSKTYNEIW; TVYYGVPVW and VTVYYGVPVW; TSGLFNSTW and NTSGLFNSTW; CTTNVPWNHSW and TTNVPWNHSW all restricted by HLA-B*58:01. The epitopes predicted were examined for presence of experimentally determined, optimally defined epitopes on the list of defined epitopes at HIV molecular immunology database http://www.hiv.lanl.gov/content/immunology/tables/optimal_ctl_summary.html. Predicted

Epitopes RSLCLFSY and LVNVTENFNMW were found on the database however, there was no HLA allele that had been experimentally determined to restrict them.

Table 3. 12: Epitope predictions of BM189 *env* gene using NetCTLpan and MHC-1 binding tool.

Allele	*Env*	Length	Peptide	NETCTLPAN v1.1 Server (ANN method)	IEDB MHC-1 binding tool (Consensus ANN/SMM/Comblib)	
				Percentage rank	Percentile rank	IC_{50}
HLA-B*35:01	gp120	11	HSFTCGGEFFY	0.80	0.65	99.61
HLA-B*35:01	gp120	11	FAILKCNDAEF	0.30	0.9	292.39
HLA-B*53:01	gp41	9	SPLSLQTLI	0.80	0.6	101.27
HLA-B*53:01	gp120	9	NPSGGDLEL	0.80	0.8	167.07
HLA-B*53:01	gp41	10	LALDKWSNLW	0.30	0.5	244.48
HLA-B*58:01	gp41	9	VTGFLALFW	0.80	0.2	14.19
HLA-B*58:01	gp41	10	WSSKTYNEIW	0.30	0.15	27.9
HLA-B*58:01	gp41	10	LGWEALKYLW	0.80	0.2	30.58
HLA-B*58:01	gp41	8	RSLCLFSY	0.15	0.6	31
HLA-B*58:01	gp41	8	ISNWLWYI	0.80	0.6	31
HLA-B*58:01	gp41	9	RSVRFVTGF	0.15	0.2	31.37
HLA-B*58:01	gp120	9	TVYYGVPVW	0.20	0.4	31.91
HLA-B*58:01	gp41	9	SSKTYNEIW	0.80	0.6	34.38
HLA-B*58:01	gp120	9	TSGLFNSTW	0.10	0.5	36.19
HLA-B*58:01	gp41	9	VTRGLQLGW	0.80	0.5	39.43
HLA-B*58:01	gp120	10	NTSGLFNSTW	0.20	0.2	50.11
HLA-B*58:01	gp120	11	LVNVTENFNMW	0.30	0.65	67.89
HLA-B*58:01	gp41	11	CTTNVPWNHSW	0.05	0.6	88.81
HLA-B*58:01	gp41	9	YLWNLAQYW	0.30	0.9	93.87
HLA-B*58:01	gp41	10	TTNVPWNHSW	0.10	0.75	277.54
HLA-B*58:01	gp120	10	VTVYYGVPVW	0.05	0.95	341.73
HLA-B*58:01	gp120	10	QMQEDIISLW	0.80	0.8	360.95
HLA-C*07:02	gp120	10	FFYCNTSGLF	0.20	0.2	75

IC_{50} = inhibitory concentration 50.

For the *gag* gene, there were similar epitopes that varied in length namely IAKNCRAPR and HIAKNCRAPR; TVVVLYCVHQR, VVLYCVHQR and VLYCVHQR; DYVDRFFKTLR and DYVDRFFK all restricted by HLA-A*33:03. KAFSPEVIPMF restricted by HLA-B*58:01, FSPEVIPM and FSPEVIPMF both restricted by HLA-C*02:02; FALNPGLL restricted by HLA-C*07:02 and RFALNPGLL restricted by HLA-C*04:01; STLQEQIAW and TSTLQEQIAW restricted by HLA-B*58:01. Epitopes KAFSPEVIPMF, STLQEQIAW, FSPEVIPM, FSPEVIPMF, FALNPGLL and RFALNPGLL were observed in partial *gag*

analysis (on Table 3. 5 and Table 3. 6). They all were sub-subtype F2 p17 and p24 epitopes and were restricted by the same alleles. BM189 *gag* epitope KAFSPEVIPMF was predicted to be restricted by HLA-B*58:01 whereas it was described in the database as restricted by HLA-B*57:01 and HLA-B*57:03. FSPEVIPMF was predicted to be restricted by HLA-C*02:02 whereas it was described in the HIV Molecular Immunology Database (http://www.hiv.lanl.gov/content/immunology/index) to be restricted by HLA-B*57.

Table 3. 13: Epitope predictions of BM189 *gag* gene using NetCTLpan and MHC-1 binding tool.

Allele	*Gag*	Length	Peptide	NETCTLPAN v1.1 Server (ANN method)	IEDB MHC-1 binding tool (Consensus of ANN/SMM/Comblib)	
				Percentage Rank	Percentile rank	IC_{50}
HLA-A*33:03	P24	9	MVHQALSPR	0.30	0.2	9.66
HLA-A*33:03	P7	9	IAKNCRAPR	0.40	0.2	14.25
HLA-A*33:03	P7	10	HIAKNCRAPR	0.80	0.3	23.58
HLA-A*33:03	P7	11	MMQKSNFKGQR	0.40	0.2	23.82
HLA-A*33:03	P17	11	TVVVLYCVHQR	0.80	0.3	27.62
HLA-A*33:03	P24	11	DYVDRFFKTLR	0.30	0.3	27.90
HLA-A*33:03	P6	10	ELLPPLASLR	0.10	0.3	37.86
HLA-A*33:03	P17	9	VVLYCVHQR	0.80	0.6	47.46
HLA-A*33:03	P24	8	DYVDRFFK	0.80	0.6	57.64
HLA-A*33:03	P7	8	QMKDCTER	0.80	0.6	63.53
HLA-A*33:03	P17	9	DAWEKIRLR	0.40	0.8	80.90
HLA-A*33:03	P17	8	VLYCVHQR	0.80	0.8	86.27
HLA-B*35:01	P24	10	EPFRDYVDRF	0.30	0.85	380.24
HLA-B*53:01	P24	10	VPVGEIYKRW	0.80	0.25	81.36
HLA-B*53:01	P6	10	EPTAPPAENF	0.30	0.7	339.93
HLA-B*58:01	P17	9	LSGGKLDAW	0.80	0.8	53.67
HLA-B*58:01	P24	11	KAFSPEVIPMF	0.10	0.5	57.26
HLA-B*58:01	P24	9	STLQEQIAW	0.10	0.7	62.90
HLA-B*58:01	P24	10	TSTLQEQIAW	0.30	0.25	74.77
HLA-B*58:01	P24	11	QALSPRTLNAW	0.80	0.7	95.04
HLA-C*02:02	P24	8	FSPEVIPM	0.40	0.2	144.26
HLA-C*02:02	P24	9	FSPEVIPMF	0.30	0.4	213.04
HLA-C*04:01	P24	8	MYSPVSIL	0.15	0.2	64
HLA-C*04:01	P17	9	RFALNPGLL	0.80	0.35	1494.02
HLA-C*07:02	P17	8	FALNPGLL	0.80	0.4	351.56

IC_{50} = inhibitory concentration 50.

CHAPTER 4: DISCUSSION, RECOMMENDATIONS AND CONCLUSION

HIV-1 diversity has been recognized as a challenge not only in classifying the different variants (Robertson et al., 1999) but also in drug resistance and vaccine design. There has been numerous reports of recombinant viruses based on sub-genome analysis, however the first reported full-genome inter-subtype recombinant was CRF01_AE in Thailand (Carr et al., 1996; Gao et al., 1996). Since then there has been numerous studies documenting a variety of inter-subtype recombinants, unique recombinant forms as well as some inter-group recombinants (Yamaguchi et al., 2008). Expectedly, the trend in the last two decades, has been an increase in recombinant forms and complexity of the virus that has led to second generation recombinants like CRF11_cpx (Tscherning-Casper et al., 2000; Montavon et al., 2002). Travel and migration contribute to the diversity and spread of the virus from one region where a subtype is endemic to another region where it is not; resulting in new subtypes being introduced and recombining with endemic subtypes (as reviewed by Perrin et al., 2003). All this has an impact on the diagnosis, vaccine development, disease progression and management (as reviewed by Bártolo and Taveira, 2012). Cameroon has had a low HIV-1 prevalence ranging from 5.5% in 2000 to 4.8% in 2014 (UNAIDS, 2015). However, there is a high genetic diversity with observations of all types, groups, subtypes, CRFs and URFs (Heyndrickx et al., 2000; Ndongmo et al., 2006; Powell et al., 2010; Tongo et al., 2013). Furthermore there have been reports of inter-group recombinants (Peeters et al., 1999; Yamaguchi et al., 2004). This diversity is partly due to the fact that Cameroon is located within the epicenter where the first zoonotic transmissions of the virus are believed to have occurred and the virus likely circulated and evolved for some time before spreading globally (Vidal et al., 2000; Sharp and Hahn, 2008).

This was a retrospective study using samples that were obtained for a Malaria study done at the study site in 2013. The samples satisfied the criteria for this study's objectives and research questions. There was a low amplification success (that is 52 samples were amplified from 107 samples that amplification was attempted on) despite numerous attempts to optimize the PCR protocol for the samples that did not amplify. The PCR primers were obtained from (Tebit et al., 2002) study that amplified HIV-1 *gag* gene from similar study site. Although there was

no clear explanation for the challenge in amplifying the 55 samples, it is within the realm of possibility that the long period between sample collection and shipment to University of Venda could have resulted in lowered quality of proviral DNA. It is also possible that the shipment conditions (at room temperature) may have led to the low amplification success as PCR success relies on good quality of DNA to produce good amplification yields.

The aim of this study was to screen for recombinants and characterize a unique recombinant form found within the study region. This would inform on the nature of the recombinant and provide some insight into the complexity of the recombinants found within the region. To this end, 39 partial *gag* sequences obtained from blood samples from study participants living within the Southwest and Littoral regions of Cameroon were analyzed. The analysis showed subtypes A related, G related, CRF01_AE and sub-subtype F2. Partial *gag*, partial *pol* and partial *env* analysis of 20 sequences, revealed that CRF02_AG, CRF22_01A1, subtype F2 and unique recombinant forms were present within the study population. There was no appreciable subtype variability across age groups or gender. The rate of amplification success was lower (52 amplifications out of 107 samples) than in most average cases which is 8 amplification per 10 samples). This could be due to the use of plasma as the source for viral DNA which is usually low in HIV-infected individuals who have seroconverted.

Subtyping of 20 partial *gag*, partial *pol* and partial *env* sequences revealed that the dominant strain was CRF02_AG (n=10/20; 50.0%) followed by URFs (n=7/20; 35.0%), then CRF22_01A1 (n=2/20; 10.0%) and finally subtype F2 (n=1/20; 5.0%). This correlates to what has been recently observed in Limbe town within Southwest region of Cameroon by Agyingi et al., (2014), whereby they observed CRF02_AG as the dominant strain (n=54/111; 48.6%) of their study population. URFs were the second dominant strain (n=40/111; 36%), CRF22_01A1 third (n=8/111; 7.2%) and subtype F2 (n=4/111; 3.6%). An earlier study conducted in Douala (Littoral region) and Yaoundé (Centre region) towns of Cameroon from 1996 to 2004 found that CRF02_AG was the dominant strain (n=334/574; 58.2%), followed by URFs (n= 85/574; 14.8%), CRF22_01A1 (n=35/574; 6.1%) and subtype F2 (n=27/574; 4.7%) (Brennan et al., 2008). The predominance of CRF02_AG was also confirmed by another study done in Northwestern regions of Cameroon (Yamaguchi et al., 2004). Despite the presence of all subtypes, groups and HIV types within Cameroon, CRF02_AG seems to predominate and this points towards a favorable adaptation this recombinant over the numerous other strains within

the country. With those samples that had *gag, pol* and *env* sequences, there was a high number of URFs, suggesting that subtyping based on 4 short genomic fragments may not be adequate to determine subtype for URFs. Hence, longer fragments (1 kb and above) or complete genome sequences would be advisable for future studies in order to correctly subtype these URFs. The presence of a high number of URFs in this study as well as previous studies suggest that recombination is driving HIV evolution towards more complex forms particularly in regions with high diversity. This study attempted to genetically describe one of these URFs as discussed below inorder to get a clearer picture on the complexity of the URFs within the region.

Genetic variability occurs within individual subtypes as well as within genes of individual subtypes. Generally, the *env* gene has the highest amino acid variability across different HIV-1 subtypes (18.0%) and *pol* has the least variability (5.8%), however this varies across individual subtypes and CRFs (Li et al., 2015). On average, HIV-1 *env* Gp120 protein has the highest amino acid variability (22.3%) across different subtypes and the most conserved protein is *gag* capsid (CA) protein (5.3%). Plot identities of the various subtypes revealed that p17 had on average, higher amino acid difference (16.48%) than p24 (7.51%) across the subtypes analyzed. This indicated that variability is higher within p17 than in p24 for the samples analyzed. This correlates with a recent study that shows that p17 had a greater variability (13.54%) than p24 (5.39%) for the same subtypes that were studied (Li et al., 2015). Subtype G on average showed the highest amino acid variability on p17 (18.95%) and on p24 (10.22%) which indicates that it is more diverse in comparison to the other subtypes studied. Previous study show subtype G has an amino acid variation of 17.33% on p17 and 5.54% on p24 (Li et al., 2015). Hence the higher amino acid variation for subtype G-like p24 sequences than the expected 5.5% variation, indicates that subtype G-like sequences may have been recombinants and not pure subtype G. These were BM250 (6.45% variation) BM526 (13.71%) and BM548 (10.48%) which may be an indication that the sequences identified as subtype G may be recombinants. BM548 showed a higher variation for p17 (21.05% variation) than the expected 17.33%. BM250 and BM526 showed a 17.89% variation which was close to the expected 17.33% which indicated that the p17 protein was likely subtype G. BM250 and BM526 may be a recombinant with subtype G for p17 and a recombinant subtype for the p24 protein.

All the CRF01_AE like sequences (BM 233, BM 240, BM 241, BM 305 and BM 402) showed higher variation for p17 and p24 than what was expected indicating a likelihood that these sequences may be second generation recombinants. Subtype F2-like sequences (BM189 and BM488) showed the expected amino acid variation for p17 and p24 indicating that these sequences were likely to be subtype F2. For the subtype A1-like sequences, BM 216, BM 247, BM 260, BM 347, BM 415, BM 446 and BM 565 showed p17 amino acid variations close to what was expected (13.07%) indicating that these were subtype A1 sequences. The rest of the subtype A1-like sequences BM 115, BM 197, BM 199, BM 206, BM 221, BM 230, BM 269, BM 282, BM 310, BM 355, BM 370, BM 378, BM 388, BM 409, BM 425, BM 444, BM 460, BM 470 and BM 565 had higher than expected variation for p17 indicating they were likely subtype A1 recombinants. For the A1-like sequences, BM 216, BM 260, BM 347, BM 446, BM 199, BM 230, BM 269, BM 282, BM 370, BM 409, BM 425 and BM 565 showed a higher p24 variation than the expected 7.3% and indicating these sequences as possible A1 recombinants. Whereas subtype A1-related sequences, BM 115, BM 197, BM 206, BM 221, BM 310, BM 355, BM 378, BM 388, BM 444, BM 460 and BM 470 showed close p24 variation to what was expected. BM 247 and BM 415 showed similar amino acid variations for p17 and p24 indicating that these sequences were subtype A1. In summary, BM 189 and BM 488 were subtype F2 sequences and BM 247 and BM 415 were subtype A1 sequences for partial *gag*. The rest of the sequences were likely recombinants.

The low variability of p24 across the subtypes studied indicates that it may be a good target for vaccine design and anti-retroviral drugs besides the current use as a diagnostic tool in fourth generation kits. Some of the amino acid sequences analyzed showed higher variability than what was expected, indicating that they were recombinants. The length of the sequence analyzed could only identify the dominant subtype from the sequence provided. Therefore, a complete *gag* gene could help clearly identify these recombinants and future studies should attempt to obtain longer fragments or sequences of *gag* as well as *pol* and *env* sequences so that they can be accurately identified.

BM189 was a sample from a 28 year old single female who was a Mutengene resident with no history of treatment. She weighed 54 kilograms at the time she participated in the study and had a white blood cell count of 4400 cells per µl with a CD4+ cell count of 104 cells/mm^3. She was clinically diagnosed as AIDS stage 3. A near full-length sequence (9077 nucleotides) of

BM189 was obtained that revealed, according to NCBI BLAST search, a 89% similarity with two sub-subtype F2 isolates from Cameroon, 88% similarity with A/F2/G recombinant isolate from Cameroon; a sub-subtype F2 isolate from Cameroon; with a sub-subtype F1 isolate and a sub-subtype F1 clone from Brazil. It also revealed 87% similarity with a subtype F molecular clone from Brazil; subtype F1 isolate from Angola and BF recombinant from Argentina. This indicated that BM189 was possibly related to sub-subtype F2 and F1 and could be a recombinant of subtypes A and G.

This was also observed from neighbor-joining tree that showed BM189 genome clustering with subtype F sequences however it clustered between sub-subtype F1 and F2 sequences. A look at jpHMM results revealed the genome to be a CRF01_AE/A1/F2 recombinant. The portion within *env* gene that was determined as CRF01_AE within *env* gene (nucleotide positions 6321 to 6728) by jpHMM was cross- referenced with a CRF01_AE reference sequence. CRF01_AE sequence showed that most of the region that was cross-referenced within *env* was a sub-subtype A1 sequence with a small portion of uncertainty (Gao et al., 1996). The difference between CRF01_AE and subtype A1 for the cross-referenced region could be due to subtype A1 intra-genetic variability between CRF01_AE sequences and pure subtype A1 sequences.

Genome analysis using Simplot determined that the sequence was A1/F2 recombinant based on 75% of permutated trees. However, at less than 75% of permutated trees, subtype G was observed within integrase, subtypes C within gp120, subtypes G and F1 within gp41 and subtypes F1 and C within *nef*. Viral protein subtyping using jpHMM, RIP, MEGA and Simplot revealed that the consensus subtypes agreed with some of the subtypes observed during BM189 genome analysis with Simplot. All the tools agreed that *gag*, *pol*, *vif*, *vpr*, *vpu*, *tat* and *rev* were sub-subtype F2. This was in agreement with BM189 genome analysis using Simplot and jpHMM that showed these genes to be sub-subtype F2. Simplot and jPHMM showed agreement within the recombination patterns observed namely a sub-subtype F2 backbone with sub-subtype A1 intermingled within *env* however the breakpoints did not show an exact match. This could be because jpHMM analysis labels breakpoints relative to HXB2 whereas Simplot does not. However it is worth noting that some of the breakpoints were within close range such as the first jpHMM breakpoint at 6324 ± 8 nucleotides was close to Simplot breakpoint at 6625. Also jpHMM third breakpoint at 8335 ±47 was in close range to breakpoint at 8250 observed from Simplot analysis. There was similarity between the two tools with exception of *pol* that

showed subtype G within integrase during Simplot genome analysis. As subtype G was observed for small subset of permutated trees, it was not worth considering. RIP, jpHMM and Simplot agreed that *env* was made of CRF01_AE and sub-subtype F2 and A1. This was in agreement with jpHMM analysis of BM189 genome. Simplot analysis of BM189 genome showed that *env* gene was a recombinant made up of subtype G, C, CRF01_AE and sub-subtype F2 and A1. CRF01_AE is likely a variation of sub-subtype A1 sequence. Subtypes G and C were only observed within a subset of permutated trees that were less than 75% of the total trees constructed during Simplot analysis. Hence the observation of these subtypes suggest a sequence similarity between the subtypes and BM189 however, not great enough to indicate that the regions are these subtypes.

BM189 genome analysis revealed the sequence to be a second generation, unique recombinant form made up of a dominant sub-subtype F2 with a minor sub-subtype A1 in most of *env* and CRF01_AE at the beginning portions of *env* (nucleotide positions 6321 to 6728 relative to HXB2) and sub-subtype F1 in most of *nef*. A search on the HIV-database (http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html) showed no CRFs with this subtype makeup or its unique recombination pattern, however there were B/F1 circulating recombinants mostly from South America. A search on NCBI GenBank revealed no full length sequence with a similar recombination pattern, however there were six near-full length sub-subtype F2 sequences. To our knowledge, a sequence with this unique recombination pattern has not been described previously in literature. It would be of interest to find out how transmissible or virulent this virus is in future studies.

Drug resistance profile for BM189 showed that the dominant strain (>20% viral population) in the quasispecies possessed no mutations associated with drug resistance. This indicated that the strain was susceptible to known drugs. However, the minority strains (<5% and <1% viral population) showed a major drug mutations (M184V) associated with high levels of resistance to two nucleoside reverse transcriptase inhibitors (NRTI) namely: Emtricitabine (FTC) and Lamivudine (3TC). The same populations showed a major drug mutation (K103N) associated with high levels of resistance to two non-nucleoside reverse transcriptase inhibitors (NNRTI) namely Efavirenz (EFV) and Nevirapine (NVP). According to the Stanford algorithm, K103N is a non-polymorphic mutation that can cause approximately 50-fold reduced susceptibility for NVP and about 20-fold reduced susceptibility to EFV. Sequencing of BM189 was done from

proviral DNA indicating that the virus population studied was a latent population. It is interesting that within the minority viral population of the latent virus, drug resistant mutations are present. This indicates the highly polymorphic nature of the virus which is caused by high mutation rates and viral turnover. The polymorphic nature of the virus results in numerous variants that have differing drug susceptibilities to known anti-retroviral drugs. Due to the high viral turnover, the virus can produce variants faster than our ability to make new drugs. This could explain why drug resistance mutations were observed among the minority viral populations. The majority viral population may have a replicative advantage over the minority population, resulting in it being the dominant strain. However, if the subject initiates therapy and drug pressure begins to act on the viral population, replication of the majority viral population would be suppressed as it is drug susceptible. Consequently, the strains within the minority viral population with the appropriate drug resistance mutation to the treatment regimen used will replicate and in time become the dominant strain.

A mutation (G36S) associated with resistance for the entry inhibitor Enfuvirtide (INN) was observed within < 5% and < 1% populations, despite no widespread use of entry inhibitors within Cameroon. Therefore it is remotely likely that the observed mutation could be due to a transmitted resistant virus that was spread by Cameroonians nationals visiting home from their travels abroad. However, it is more likely the mutation could have arisen due to natural polymorphism observed within non-B subtypes (Carmona et al., 2005).

Co-receptor prediction showed that the dominant strain was a R5 tropic virus that was susceptible to entry inhibitors like maraviroc. This was supported by the drug resistance profile that showed the dominant strain was susceptible to entry inhibitors. The use of an entry inhibitor may work to suppress the dominant strain in combination with other classes of drugs. However, the presence of G36S mutation associated with resistance for the entry inhibitor Enfuvirtide within the minority strains indicate that the drug may fail to suppress viral growth once the minority strains gain a selective advantage due to drug pressure.

Cameroon like many other resource limited countries apply the WHO Public Health approach for management of HIV infection (W.H.O., 2013). NRTIs and NNRTIs constitute the first line of therapy which involves a combination cocktail of Tenofovir (TDF) with 3TC and EFV. TDF is a nucleoside reverse transcriptase inhibitors. Despite presence of 3TC and EFV resistant viruses within the minority population, therapy with TDF, 3TC and EFV can manage to

suppress growth of the dominant strain which is susceptible to the drugs. Moreover, the presence of M184V increases susceptibility of TDF and is associated with reduced viral replication and delays the emergence of TDF resistant viruses (Wolf et al., 2003; Ross et al., 2004). However, the subject BM189 would have to be monitored through resistance testing or/and viral load measurements regularly to ensure she switches to second line therapy once first line begins to fail. Other first line options involve using NVP in place of EFV or FTC instead of 3TC (W.H.O., 2013). However as there is resistance for NVP and FTC within the minority strains, a similar outcome is to be expected. Second line therapy involves a cocktail of one protease inhibitors (lopinavir/ritonavir) with TDF and 3TC. Due to having viral populations resistant to 3TC it would be advisable to switch to abacavir as an alternative NRTI.

Through next-generation sequencing (NGS), the minority populations i.e. <5% and <1% viral populations could be screened for drug resistance and showed mutations associated with resistance for two NRTIs, two NNRTIs and one entry inhibitor. The majority population (>20% viral population) showed susceptibility to known drugs. If Sanger sequencing was used for screening, these drug resistance mutations observed within the minority population would not have been detected. Sanger sequencing lacks the sensitivity to sequence the minority populations <20% (Samuel et al., 2014). The implication of this is that if clinical genotypic drug resistance profiling or monitoring is done using Sanger sequencing, this could lead to rapid failure if not closely monitored. Hence the use of NGS can inform on the minority population and aid clinicians in stream-lining patient's drug regimen to give the patient the best possible therapeutic outcome. As costs for per sequencing run for next-generation sequencing drops, it is recommended that next-generation sequencing be used as the standard method for clinical genotypic drug resistance profiling (as reviewed by Quiñones-Mateu et al., 2014).

Due to the large number of epitopes present within viral protein sequences across strains, *in vitro* experimentation to determine optimal epitopes can be laborious, expensive and time consuming. Genotypic prediction methods help by narrowing down the possible epitopes for phenotypic testing (as reviewed by Yang and Yu, 2009; as reviewed by Soria-Guerra et al., 2015). HLA B*57, B*58:01, B*27, B*51 and B*81:01 have been shown to be associated with controlling viral growth whereas HLA B*58:02, B*35 and C*04 have been shown to be associated with rapid disease progression (Carrington et al., 1999; Hendel et al., 1999; Migueles et al., 2000). Subjects from the study region display a high frequency (11.5%) of

HLA-B*58:01 which is associated with disease control (Torimiro et al., 2006). The p17 epitope RSLYNTVATLY and p24 epitope TSTLQEQIGW from both subtypes B and C have been shown to be restricted by HLA-B*58:01 (Llano et al., 2013; as reviewed by Chakraborty et al., 2014). P17 epitope RSLYNTVATLY was predicted for subtype A1 related sequences, however the epitope had a frequency (n=1/25; 4.00%) among subtype A1 related sequences, with BM347 having the epitope. The remaining subtype A1-related sequences had different amino acids at different positions within the RSLYNTVATLY epitope. The epitope was predicted from the consensus "BM" sequence of all subtype A1 related sequences. Generally, the amino acids at the N and the C-terminus of the epitope plays an important role in anchoring the epitope on the MHC-1 molecule. Hence a different amino acids at the N and C terminus can lead to unstable binding and/or lowered binding affinity (Janeway et al., 2001).

For RSLYNTVATLY epitope, arginine (R) and tyrosine (Y) at the N and C terminus respectively, anchors the epitope ensuring stable binding. However, among the subtype A1 related sequences, 23 out of 25 "BM" sequences had amino acids that were not R at their N-terminus and 14 out of 25 of them had amino acids that were not Y at their C terminus. BM197, BM206, BM216, BM221, BM247, BM260, BM282, BM310, BM355, BM370, BM378, BM388, BM409, BM415, BM425, BM444, BM460, BM470 and BM565 had lysine (K) at the N terminus instead on R. Whereas BM115 and BM199 had isoleucine, BM446 had glutamine (Q) instead of R at the N-terminus. Only BM230 and BM347 had R at the N-terminus. BM115, BM197, BM221, BM230, BM269, BM388, BM409, BM446, BM470 and BM565 had tryptophan (W) instead of tyrosine (Y) at the C-terminus. BM206, BM260 and BM444 had phenylalanine (F) at the C-terminus whereas BM415 had cysteine (C) and BM460 had arginine (R). This indicates that RSLYNTVATLY epitope may have lower binding affinity as well as lower stability for the "BM" sequences that did not have R and Y at their N and C terminus.

The p24 epitope TSTLQEQIGW was predicted for CRF01_AE related sequences which were BM233, BM240, BM241, BM305 and BM402 with (n=1/5; 20%) frequency within CRF01_AE related sequences. All the sequences had threonine (T) at the N-terminus and tryptophan (W) at the C terminus, indicating that they would have similar binding affinities as TSTLQEQIGW despite some sequences having single amino acid differences. BM233, BM240 and BM241 had alanine (A) instead of glycine (G) at amino acid position 9 of the epitope, whereas BM402 had asparagine (N) instead of threonine (T) at position 3. BM305 had

the exact epitope sequence. These single amino acid changes may have a minimal effect on binding affinity as these changes are not at the anchor regions that is N and C terminus. The presence of the p24 epitope TSTLQEQIGW suggests that study subjects expressing HLA-B*58:01 can to some degree restrict growth of these strains *in vivo*. This could also be true of the epitopes that had a similar motif as p24 epitope TSTLQEQIGW. However, phenotypic tests need to be carried out to determine if antigen presentation, CTL binding and potent immune response will occur. Moreover, further studies need to be done to determine if restriction of this epitope will result in slower disease progression.

Prediction for p17 epitopes WPFRNSRRM and YLIQQPSIVY suggested that these epitopes were restricted by HLA-B*35:01. Since HLA B*35 serotype is associated with rapid disease progression (Carrington et al., 1999), the strong binding affinities predicted for these epitopes indicates a potential risk towards rapid disease progression for the study participants that express this allele and are infected with CRF01_AE related or subtype G related strains. CRF01_AE related sequences, BM233 and BM305, showed the exact WPFRNSRRM epitope sequence indicating that these subjects were at a potential risk of rapid disease progression if they expressed HLA-B*35:01 allele. CRF01_AE related sequences BM240 had serine (S) instead of phenylalanine (F) at amino acid position 3 of the epitope and BM241 had glycine (G) instead of arginine (R) at position 8. CRF01_AE related sequences BM402 had serine (S) instead of proline (P) at position 2 and isoleucine (I) instead of asparagine (N) at position 5. It is possible that BM240, BM241 and BM402 may have lower binding affinities due to the amino acid differences for their epitopes. The frequency of P17 epitopes WPFRNSRRM was 40% (n=2/5) among CRF01_AE related sequences. P17 epitopes YLIQQPSIVY was predicted for subtype G related sequences namely; BM250, BM526 and BM548. BM250 and BM548 had serine (S) instead of proline (P) at position 6 whereas BM526 had isoleucine (I) instead of leucine (L) at position 2. These changes may affect binding affinities of the epitopes for the allele. Prediction for p17 epitope RFALNPGLL showed that it could be restricted by HLA-C*04:01. However due to the high IC_{50} value (1494.02 nM for subtype F2 related sequences namely BM189 and BM488) it is unlikely that this epitope could have a strong binding affinity for the allele. Generally, epitopes that are predicted to have IC_{50} values greater than 500 are considered poor MHC-1 binders and hence unlikely to be restricted. This could imply that BM189 and BM488 will progress slowly even if they express HLA-C*04:01.

The p24 epitope TSTLQEQIGW would be worth considering for future vaccine designs targeting this population considering that they are restricted by HLA B*58:01, which has been associated with slower disease progression and also observed in a high frequency (11.5%) within the study area. Furthermore, the epitopes have been found within subtype B, C and G and has been predicted to occur at 20% frequency among CRF01_AE sequences, hence it could be used in a subtype specific vaccine. However due to the high diversity within Cameroon, it would be worthwhile to determine if these epitopes are present within other subtypes and CRF02_AG. If they are, then these epitopes could be used as part of immunogens for a universal vaccine.

BM189 *gag* epitope prediction showed that epitopes LSGGKLDAW, KAFSPEVIPMF, STLQEQIAW, TSTLQEQIAW and QALSPRTLNAW were strong binders for HLA-B*58:01. Furthermore, *env* epitope prediction also showed VTGFLALFW, WSSKTYNEIW, LGWEALKYLW, RSLCLFSY, ISNWLWYI, RSVRFVTGF, TVYYGVPVW, SSKTYNEIW, TSGLFNSTW, VTRGLQLGW, NTSGLFNSTW, LVNVTENFNMW, CTTNVPWNHSW, YLWNLAQYW, TTNVPWNHSW, VTVYYGVPVW and QMQEDIISLW were strong binders for HLA-B*58:01. This allele has been associated with slow disease progression. Furthermore, *env* epitopes (HSFTCGGEFFY and FAILKCNDAEF) were predicted to be strong binders of HLA-B*35:01 and *gag* epitope (MYSPVSIL) was predicted to be strong binders for HLA-C*04:01. HLA typing was not done to determine whether the subject expressed any of these allele so there was no way to speculate if the subject was able to suppress the virus or susceptible to fast disease progression. It is worth noting that association of HLA-B*58:01 to slow disease progression and association of HLA-B*35:01 and HLA-C*04:01 to fast disease progression depends on genetic as well as environmental factors (Carrington et al., 1999; Migueles et al., 2000).

Numerous epitopes were predicted to have good binding affinities for the HLA alleles tested. It has been shown that CD8+ T cell responses targeting *gag* gene are associated with lower viremia whereas responses targeting *env* or accessory genes were associated with increased viremia response (Kiepiela et al., 2007). Therefore, it is recommended that future studies using an assay like interferon-gamma ELISPOT be done to determine if the predicted *gag* epitopes have the suggested binding affinities (IC_{50}) for the respective alleles. This would indicate if the predicted epitopes would produce a strong enough CD8+ T cell response as predicted.

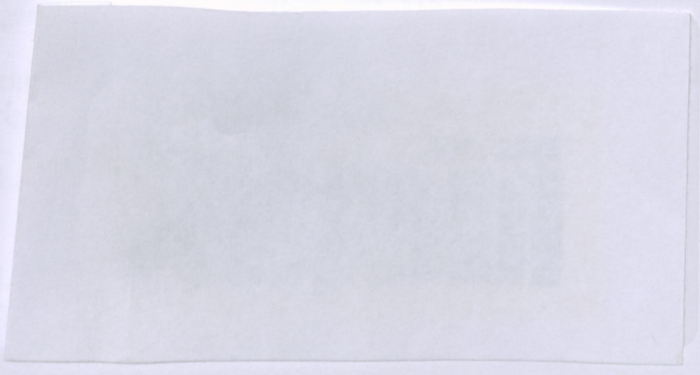
UNIVERSITY OF VENDA
LIBRARY

To conclude, in this study, *gag* gene from proviral DNA was amplified, sequenced and subtyped to detect possible recombination. From the 39 *gag* sequences, it was observed that they were subtypes A related, G related, CRF01_AE related and sub-subtype F2 related sequences, with most of subtypes A related, G related and CRF01_AE related sequences being possible recombinants. Analyses of 20 sequences that had partial *gag*, *pol* and *env* sequences showed that the predominant strains found within the within study region were CRF02_AG, URFs, CRF22_01A1 and sub-subtype F2. A unique recombinant form (BM189) was selected, amplified, deep sequenced and drug resistance was determined for >20%, <5% and <1% viral populations. The dominant population (>20%) was subtyped and co-receptor usage was predicted. The dominant strain was determined to be F2/CRF01_AE/A1/F1 recombinant. Its unique recombination pattern had not been observed circulating among the Cameroonian population based on a genbank search nor was it found among the samples that were subtyped using partial sub-genomes in this study. The recombinant or dominant strain was susceptible to all known drugs however, there were mutations associated with drug resistance within the minority population. It is likely that these mutations were a result of transmitted drug-resistant strains for NRTI and NNRTI inhibitors. For mutation associated with Enfuvirtide resistance it is likely that the mutation occurred due to the natural polymorphism commonly found among non-B subtypes. BM189 was predicted to be a R5 tropic virus. CTL epitopes prediction based on high frequency alleles within the study area, showed that there were some epitopes which were restricted by alleles associated with both slow disease progression and rapid disease progression. It also identified epitopes that had been described previously as optimal CTL epitopes. This study highlighted the importance of full-genome sequencing for determining genetic makeup of recombinants. It also illustrated the importance of next generation sequencing for genotypic drug resistance as a tool of selecting and monitoring therapy.

Study limitations

A larger sample size may elucidate greater details on all the subtypes, CRFs and URFs present within the study population. However, the limited sample size and lack of corresponding *pol* or *env* sequences for some of the samples meant that the findings of this work can only provide a marginal confirmation of the current genetic variants present within the study area. The samples with only *gag* sequences could not be confidently subtyped. Further studies can be done to look at a larger sample size and possibly look at *gag, pol* and *env* sequence to obtain a clearer picture of the viral diversity within the study region.

Possible CTL epitopes were predicted however, epitope prediction algorithms are trained mostly using 9 mer epitopes because experimental data show that most epitopes processed *in vivo* are 9 mer epitopes. Eleven mer and even longer epitopes are processed *in vivo* and usually have higher binding affinities in comparison to 9 mer epitopes although, they are rarely made. Hence a limitation with the predictive methods is that they give relatively good predictions for 9 mer epitopes however not as good prediction for 11 or longer predictions. This could mean that some 11 or greater mer epitopes with very strong binding affinity *in vivo* could have been neglected because it showed weak binding affinity *in silico*.



Reference

Aghokeng A.F., Mpoudi-Ngole E., Chia J.E., Edoul E.M., Delaporte E., and Peeters M. 2011. High failure rate of the ViroSeq HIV-1 genotyping system for drug resistance testing in Cameroon, a country with broad HIV-1 genetic diversity. *Journal of Clinical Microbiology*. **49(4):** 1635–1641.

Aghokeng A.F., Mpoudi-Ngole E., Dimodi H., Atem-Tambe A., Tongo M., Butel C., Delaporte E., and Peeters M. 2009. Inaccurate diagnosis of HIV-1 group M and O is a key challenge for ongoing universal access to antiretroviral treatment and HIV prevention in Cameroon. *PloS One*. **4(11):** e7702.

Agyingi L., Mayr L.M., Kinge T., Orock G.E., Ngai J., Asaah B., Mpoame M., Hewlett I., and Nyambi P. 2014. The evolution of HIV-1 group M genetic variability in Southern Cameroon is characterized by several emerging recombinant forms of CRF02_AG and viruses with drug resistance mutations. *Journal of Medical Virology*. **86(3):** 385–393.

Aldrich C. and Hemelaar J. 2012. Global HIV-1 diversity surveillance. *Trends in Molecular Medicine*. **18(12):** 691–694.

Ammaranond P., van Bockel D.J., Petoumenos K., McMurchie M., Finlayson R., Middleton M.G., Davenport M.P., Venturi V., Suzuki K., Gelgor L., Kaldor J.M., Cooper D.A., and Kelleher A.D. 2011. HIV Immune Escape at an Immunodominant Epitope in HLA-B* 27-Positive Individuals Predicts Viral Load Outcome. *The Journal of Immunology*. **186(1):** 479–488.

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):** 8979–8990.

Aslanzadeh J. 2004. Preventing PCR amplification carryover contamination in a clinical laboratory. *Annals of Clinical & Laboratory Science*. **34(4):** 389–396.

Ayouba A., Mauclère P., Martin P., Cunin P., Mfoupouendoun J., Njinku B., Souquières S., and Simon F. 2001. HIV-1 group O infection in Cameroon, 1986 to 1998. *Emerging Infectious Diseases*. **7(3):** 466.

Bahassi E.M. and Stambrook P.J. 2014. Next-generation sequencing technologies: breaking the sound barrier of human genetics. *Mutagenesis*. **29(5):** 303–310.

Bailes E., Gao F., Bibollet-Ruche F., Courgnaud V., Peeters M., Marx P.A., Hahn B.H., and Sharp P.M. 2003. Hybrid origin of SIV in chimpanzees. *Science*. **300(5626):** 1713.

Barouch D.H. and Michael N.L. 2014. Accelerating HIV-1 vaccine efficacy trials. *Cell*. **159(5):** 969–972.

Barré-Sinoussi F., Chermann J., and Rey F. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*.

220(4599): 868–871.

Bártolo I. and Taveira N. 2012. HIV-1 Diversity and Its Implications in Diagnosis, Transmission, Disease Progression, and Antiretroviral Therapy. INTECH Open Access Publisher.

Basu V.P., Song M., Gao L., Rigby S.T., Hanson M.N., and Bambara R.A. 2008. Strand transfer events during HIV-1 reverse transcription. *Virus Research*. **134(1):** 19–38.

Bieniasz P.D. 2006. Late budding domains and host proteins in enveloped virus release. *Virology*. **344(1):** 55–63.

Biswas P., Jiang X., Pacchia A.L., Dougherty J.P., and Peltz S.W. 2004. The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. *Journal of Virology*. **78(4):** 2082–2087.

Bolinger C. and Boris-Lawrie K. 2009. Mechanisms employed by retroviruses to exploit host factors for translational control of a complicated proteome. *Retrovirology*. **6(8):** 20.

Boone L.R. and Skalka A. 1981. Viral DNA synthesized in vitro by avian retrovirus particles permeabilized with melittin. II. Evidence for a strand displacement mechanism in plus-strand synthesis. *Journal of Virology*. **37(1):** 117–126.

Bour S., Geleziunas R., and Wainberg M.A. 1995. The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection. *Microbiological Reviews*. **59(1):** 63–93.

Brennan C.A., Bodelle P., Coffey R., Devare S.G., Golden A., Hackett Jr J., Harris B., Holzmayer V., Luk K.-C., Schochetman G., Swanson P., Yamaguchi J., Vallari A., Ndembi N., Ngansop C., Makamche F., Mbanya D., Gurtler L.G., Zekeng L., and Kaptue L. 2008. The prevalence of diverse HIV-1 strains was stable in Cameroonian blood donors from 1996 to 2004. *Journal of Acquired Immune Deficiency Syndromes*. **49(4):** 432–439.

Briggs J. and Kräusslich H.G. 2011. The molecular architecture of HIV. *Journal of Molecular Biology*. **410(4):** 491–500.

Bukrinskaya A.G. 2004. HIV-1 assembly and maturation. *Archives of Virology*. **149(6):** 1067–1082.

Burda S.T., Konings F.A., Williams C.A., Anyangwe C., and Nyambi P.N. 2004. HIV-1 CRF09_cpx circulates in the North West Province of Cameroon where CRF02_AG infections predominate and recombinant strains are common. *AIDS Research & Human Retroviruses*. **20(12):** 1358–1363.

Burton D.R., Ahmed R., Barouch D.H., Butera S.T., Crotty S., Godzik A., Kaufmann D.E., McElrath M.J., Nussenzweig M.C., Pulendran B., Scanlan C.N., Schief W.R., Silvestri G., Streeck H., Walker B.D., Walker L.M., Ward A.B., Wilson I.A., and Wyatt R. 2012. A Blueprint for HIV Vaccine Discovery. *Cell Host & Microbe*. **12(4):** 396–407.

Carmona R., Perez-Alvarez L., Munoz M., Casado G., Delgado E., Sierra M., Thomson M., Vega Y., de Parga E.V., Contreras G., and others 2005. Natural resistance-associated mutations to Enfuvirtide (T20) and polymorphisms in the gp41 region of different HIV-1 genetic forms from T20 naive patients. *Journal of Clinical Virology*. **32(3):** 248–253.

Carr J.K., Salminen M.O., Albert J., Sanders-Buell E., Gotte D., Birx D.L., and McCutchan F.E. 1998. Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. *Virology*. **247(1):** 22–31.

Carr J.K., Salminen M.O., Koch C., Gotte D., Artenstein A.W., Hegerich P.A., St Louis D., Burke D.S., and McCutchan F.E. 1996. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *Journal of Virology*. **70(9):** 5935–5943.

Carr J.K., Torimiro J.N., Wolfe N.D., Eitel M.N., Kim B., Sanders-Buell E., Jagodzinski L.L., Gotte D., Burke D.S., Birx D.L., and others 2001. The AG recombinant IbNG and novel strains of group M HIV-1 are common in Cameroon. *Virology*. **286(1):** 168–181.

Carr J.K., Wolfe N.D., Torimiro J.N., Tamoufe U., Mpoudi-Ngole E., Eyzaguirre L., Birx D.L., McCutchan F.E., and Burke D.S. 2010. HIV-1 recombinants with multiple parental strains in low-prevalence, remote regions of Cameroon: Evolutionary relics? *Retrovirology*. **7(1):** 1.

Carrington M., Nelson G.W., Martin M.P., Kissner T., Vlahov D., Goedert J.J., Kaslow R., Buchbinder S., Hoots K., and O'Brien S.J. 1999. HLA and HIV-1: heterozygote advantage and B* 35-Cw* 04 disadvantage. *Science*. **283(5408):** 1748–1752.

Casals F., Idaghdour Y., Hussin J., and Awadalla P. 2012. Next-generation sequencing approaches for genetic mapping of complex diseases. *Journal of Neuroimmunology*. **248(1-2):** 10–22.

CDC 1981. Follow-up on Kaposi's sarcoma and Pneumocystis pneumonia. *Morbidity and Mortality Weekly Report*. **30(33):** 409–410.

Chakraborty S., Rahman T., and Chakravorty R. 2014. Characterization of the protective HIV-1 CTL epitopes and the corresponding HLA class I alleles: A step towards designing CTL based HIV-1 vaccine. *Advances in Virology*. **2014**.

Chan D.C. and Kim P.S. 1998. HIV entry and its inhibition. *Cell*. **93(5):** 681–684.

Checkley M.A., Luttge B.G., and Freed E.O. 2011. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *Journal of Molecular Biology*. **410(4):** 582–608.

Clavel F., Guyader M., Guetard D., Salle M., Montagnier L., and Alizon M. 1986. Molecular cloning and polymorphism of the human immune deficiency virus type 2. *Nature*. **324** 691–695.

Clements J. and Zink M.C. 1996. Molecular biology and pathogenesis of animal lentivirus infections. *Clinical Microbiology Reviews*. **9(1):** 100–117.

Coffin J. 1996. Retroviridae: the viruses and their replication. In Fields, BN and Knipe, DM and Howley, PM, (Ed.) *Fundamental Virology. 3rd edition.* Raven Press, New York, pp. 763–844.

Coffin J. and Swanstrom R. 2013. HIV pathogenesis: dynamics and genetics of viral populations and infected cells. *Cold Spring Harbor Perspectives in Medicine*. **3(1):** a012526.

Coffin J.M. 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *Journal of General Virology*. **42(1):** 1–26.

Courtney C.R., Agyingi L., Fokou A., Christie S., Asaah B., Meli J., Ngai J., Hewlett I., and Nyambi P.N. 2015. Monitoring HIV-1 Group M Subtypes in Yaoundé, Cameroon Reveals Broad Genetic Diversity and a Novel CRF02_AG/F2 Infection. *AIDS Research & Human Retroviruses*. **31(00):** 1–5.

D'arc M., Ayouba A., Esteban A., Learn G.H., Boué V., Liegeois F., Etienne L., Tagg N., Leendertz F.H., Boesch C., Madinda N.F., Robbins M.M., Gray M., Cournil A., Ooms M., Letko M., Simon V.A., Sharp P.M., Hahn B.H., Delaporte E., Ngole E.M., and Peeters M. 2015. Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proceedings of the National Academy of Sciences*. **112(11):** E1343–E1352.

Delviks-Frankenberry K., Galli A., Nikolaitchik O., Mens H., Pathak V.K., and Hu W.-S. 2011. Mechanisms and factors that influence high frequency retroviral recombination. *Viruses*. **3(9):** 1650–80.

Dijk E.L., Auger H., Jaszczyszyn Y., and Thermes C. 2014. Ten years of next-generation sequencing technology. *Trends in Genetics*. **30(9):** 418–426.

Doms R.W. and Peiper S.C. 1997. Unwelcomed guests with master keys: how HIV uses chemokine receptors for cellular entry. *Virology*. **235(2):** 179–190.

Douek D.C., Kwong P.D., and Nabel G.J. 2006. The rational design of an AIDS vaccine. *Cell*. **124(4):** 677–681.

Engelman A. and Cherepanov P. 2008. The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication. *PLoS pathogens*. **4(3):** e1000046.

Engelman A., Mizuuchi K., and Craigie R. 1991. HIV-1 DNA integration: Mechanism of viral DNA cleavage and DNA strand transfer. *Cell*. **67(6):** 1211 – 1221.

Faria N.R., Suchard M.A., Abecasis A., Sousa J.D., Ndembi N., Bonfim I., Camacho R.J., Vandamme A.-M., and Lemey P. 2012. Phylodynamics of the HIV-1 CRF02_AG clade in Cameroon. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*. **12(2):** 453–60.

Fonjungo P.N., Mpoudi E.N., Torimiro J.N., Alemnji G.A., Eno L.T., Nkengasong J.N., Gao F., Rayfield M., Folks T.M., Pieniazek D., and Lal R.B. 2000. Presence of diverse human immunodeficiency virus type 1 viral variants in Cameroon. *AIDS Research & Human Retroviruses*. **16(13):** 1319–24.

Frankel A.D. and Young J.A. 1998. HIV-1: fifteen proteins and an RNA. *Annual Review of Biochemistry*. **67(1):** 1–25.

Fraser C. 2005. HIV recombination: what is the impact on antiretroviral therapy? *Journal of the Royal Society Interface*. **2(5):** 489–503.

Freed E.O. 2002. HIV-1 Replication. *Somatic Cell and Molecular Genetics*. **26(1):** 13–33.

Freed E.O. and Martin M.A. 2001. HIVs and their replication. In Fields, B.N. and Knipe, D.M. and Howley, P.M. and Griffin, D.E., (Ed.) *Fields' Virology*. Fields Virology. Lippincott Williams & Wilkins Philadelphia, pp. 1971–2041.

Gallo R., Sarin P., Gelmann E., Robert-Guroff M., Richardson E., Kalyanaraman V., Mann D., Sidhu G., Stahl R., Zolla-Pazner S., Leibowitch J., and Popovic M. 1983. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science*. **220(4599):** 865–867.

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

Gao F., Bailes E., Robertson D.L., Chen Y., Rodenburg C.M., Michael S.F., Cummins L.B., Arthur L.O., Peeters M., Shaw G.M., Sharp P.M., and Hahn B.H. 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature*. **397(6718):** 436–441.

Gao F., Robertson D.L., Morrison S.G., Hui H., Craig S., Decker J., Fultz P.N., Girard M., Shaw G.M., Hahn B.H., and Sharp P.M. 1996. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *Journal of Virology*. **70(10):** 7013–7029.

Gao F., Vidal N., Li Y., Trask S.A., Chen Y., Kostrikis L.G., Ho D.D., Kim J., Oh M.-D., Choe K., Salminen M., Robertson D.L., Shaw G.M., Hahn B.H., and Peeters M. 2001. Evidence of two distinct subsubtypes within the HIV-1 subtype A radiation. *AIDS Research & Human Retroviruses*. **17(8):** 675–688.

Garcia-Calleja J.M., Zekeng L., Louis J.P., Mvondo J.L., Trebocq A., Sokal D., Yanga D., Ndoumou A., Andela D., Salla R., Gardon J., Abbenyi S., and Kaptué L. 1992. HIV infection in Cameroon: 30 months' surveillance in Yaounde. *AIDS*. **6(8):**

Van Gils M.J. and Sanders R.W. 2013. Broadly neutralizing antibodies against HIV-1: templates for a vaccine. *Virology*. **435(1):** 46–56.

Goodrich D.W. and Duesberg P.H. 1990. Retroviral recombination during reverse transcription. *Proceedings of the National Academy of Sciences*. **87(6):** 2052–2056.

Gottlieb G.S., Nickle D.C., Jensen M.A., Wong K.G., Grobler J., Li F., Liu S.-L., Rademeyer C., Learn G.H., Karim S.S.A., Williamson C., Corey L., Margolick J.B., and Mullins J.I. 2004. Dual HIV-1 infection associated with rapid disease progression. *Lancet*. **363(9409):** 619–622.

Grada A. and Weinbrecht K. 2013. Next-generation sequencing: methodology and application. *The Journal of Investigative Dermatology*. **133(8):** e11.

Gürtler L.G., Hauser P.H., Eberle J., Von Brunn A., Knapp S., Zekeng L., Tsague J.M., and Kaptue L. 1994. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *Journal of Virology*. **68(3):** 1581–1585.

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

Hall T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic Acids Symposium Series*. pp. 95–98.

Hayward J.J. and Rodrigo A.G. 2010. Molecular epidemiology of feline immunodeficiency virus in the domestic cat (Felis catus). *Veterinary Immunology and Immunopathology*. **134(1-2):** 68–74.

Hemelaar J. 2013. Implications of HIV diversity for the HIV-1 pandemic. *Journal of Infection*. **66(5):** 391–400.

Hemelaar J. 2012. The origin and diversity of the HIV-1 pandemic. *Trends in molecular medicine*. **18(3):** 182–192.

Hemelaar J., Gouws E., Ghys P.D., and Osmanov S. 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS*. **20(16):** W13–W23.

Hemelaar J., Gouws E., Ghys P.D., and Osmanov S. 2011. Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS*. **25(5):** 679–689.

Hendel H., Caillat-Zucman S., Lebuanec H., Carrington M., O'Brien S., Andrieu J.-M., Schächter F., Zagury D., Rappaport J., Winkler C., and others 1999. New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. *The Journal of Immunology*. **162(11):** 6942–6946.

Hetzer C., Dormeyer W., Schnölzer M., and Ott M. 2005. Decoding Tat: the biology of HIV Tat posttranslational modifications. *Microbes and Infection*. **7(13):** 1364–1369.

Heyndrickx L., Janssens W., Ndumbe P.M., Vereecken K., Coppens S., De Houwer K., Fransen K., Van der Auwera G., and van der Groen G. 2000. HIV-1 genetic variability in Cameroon. *AIDS*. **14(12):** 1862–1864.

Hirsch V.M., Olmsted R.A., Murphey-Corb M., Purcell R.H., and Johnson P.R. 1989. An African primate lentivirus (SIVsm closely related to HIV-2). *Nature*. **339(6223):** 389–392.

HIV-LANL 2016. HIV Circulating Recombinant forms (CRFs). Available at: http://www.hiv.lanl.gov [Accessed January 25, 2016].

HIV-LANL 2008. HIV Sequence Compendium 2008 Kuiken C, Leitner T, Foley B, Hahn B, Marx P, McCutchan F, Wolinsky S, and Korber B, (Ed.). Theoretical Biology and Biophysics Group, Los Alamos National Laboratory.

Ho D.D., Neumann A.U., Perelson A.S., Chen W., Leonard J.M., and Markowitz M. 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*. **373(6510):** 123–126.

Hu W. and Temin H.M. 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proceedings of the National Academy of Sciences*. **87(4):** 1556–1560.

Huang W., Eshleman S.H., Toma J., Fransen S., Stawiski E., Paxinos E.E., Whitcomb J.M., Young A.M., Donnell D., Mmiro F., and others 2007. Coreceptor tropism in human immunodeficiency virus type 1 subtype D: high prevalence of CXCR4 tropism and heterogeneous composition of viral populations. *Journal of Virology*. **81(15):** 7885–7893.

Huet T., Cheynier R., Meyerhans A., Roelants G., and Wain-Hobson S. 1990. Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature*. **345(6273):** 356–359.

Hurwitz J. and Leis J.P. 1972. RNA-dependent DNA polymerase activity of RNA tumor viruses I. Directing influence of DNA in the reaction. *Journal of Virology*. **9(1):** 116–129.

Hutchinson J.F. 2001. The biology and evolution of HIV. *Annual Review of Anthropology*. **30(1):** 85–108.

Hwang C.K., Svarovskaia E.S., and Pathak V.K. 2001. Dynamic copy choice: steady state between murine leukemia virus polymerase and polymerase-dependent RNase H activity determines frequency of in vivo template switching. *Proceedings of the National Academy of Sciences*. **98(21):** 12209–12214.

Iweala O.I. 2004. HIV diagnostic tests: an overview. *Contraception*. **70(2):** 141–147.

Janeway C.J., Travers P., Walport M., and Capra J. 2001. Antigen recognition by T cells. In *Immunobiology: The Immune System in Health and Disease.* Garland Science Publishing New York.

Janssens W., Buvé A., and Nkengasong J.N. 1997. The puzzle of HIV-1 subtypes in Africa. *AIDS*. **11(6):** 705–712.

Jensen M.A., Li F.-S., van't Wout A.B., Nickle D.C., Shriner D., He H.-X., McLaughlin S., Shankarappa R., Margolick J.B., and Mullins J.I. 2003. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *Journal of Virology*. **77(24):** 13376–13388.

Johnston M.I. and Fauci A.S. 2011. HIV vaccine development-improving on natural immunity. *The New England Journal of Medicine*. **365(10):** 873–875.

Joshi S. and Joshi R.L. 1996. Molecular biology of human immunodeficiency virus type-1. *Transfusion Science*. **17(3):** 351–378.

Kaleebu P., Nankya I.L., Yirrell D.L., Shafer L.A., Kyosiimire-Lugemwa J., Lule D.B., Morgan D., Beddows S., Weber J., and Whitworth J.A. 2007. Relation between chemokine receptor use, disease stage, and HIV-1 subtypes A and D: results from a rural Ugandan cohort. *Journal of Acquired Immune Deficiency Syndromes*. **45(1):** 28–33.

Karn J. and Stoltzfus C.M. 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harbor Perspectives in Medicine*. **2(2):** a006916.

Keele B.F., Van Heuverswyn F., Li Y., Bailes E., Takehisa J., Santiago M.L., Bibollet-Ruche F., Chen Y., Wain L.V., Liegeois F., and others 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*. **313(5786):** 523–526.

Kellam P. and Larder B.A. 1995. Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. *Journal of Virology*. **69(2):** 669–74.

Keller M., Lu Y., Lalonde R.G., and Klein M.B. 2009. Impact of HIV-1 viral subtype on CD4+ T-cell decline and clinical outcomes in antiretroviral naive patients receiving universal healthcare. *AIDS*. **23(6):** 731–737.

Ketema F., Zeh C., Edelman D.C., Saville R., and Constantine N.T. 2001. Assessment of the performance of a rapid, lateral flow assay for the detection of antibodies to HIV. *Journal of Acquired Immune Deficiency Syndromes*. **27(1):** 63–70.

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

Kim Y., Ponomarenko J., Zhu Z., Tamang D., Wang P., Greenbaum J., Lundegaard C., Sette A., Lund O., Bourne P.E., Nielsen M., and Peters B. 2012. Immune epitope database analysis resource. *Nucleic Acids Research*. **40(Web Server issue):** W525–W530.

Kiwanuka N., Robb M., Laeyendecker O., Kigozi G., Wabwire-Mangen F., Makumbi F.E., Nalugoda F., Kagaayi J., Eller M., Eller L.A., Serwadda D., Sewankambo N.K., Reynolds S.J., Quinn T.C., Gray R.H., Wawer M.J., and Whalen C.C. 2010. HIV-1 viral subtype differences in the rate of CD4+ T-cell decline among HIV seroincident antiretroviral naive persons in Rakai district, Uganda. *Journal of Acquired Immune Deficiency Syndromes*. **54(2):** 180–184.

Korber B., Gaschen B., Yusim K., Thakallapally R., Kesmir C., and Detours V. 2001. Evolutionary and immunological implications of contemporary HIV-1 variation. *British*

Medical Bulletin. **58(1):** 19–42.

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

Lackner A., Lederman M.M., and Rodriguez B. 2012. HIV pathogenesis: the host. *Cold Spring Harbor Perspectives in Medicine*. **2(9):** a007005.

Lau K.A., Wang B., Miranda-Saksena M., Boadle R., Kamarulzaman A., Ng K.-P., and Saksena N.K. 2010. Evidence for Possible Biological Advantages of the Newly Emerging HIV-1 Circulating Recombinant Form from Malaysia - CRF33_01B in Comparison to its Progenitors CRF01_AE and Subtype B. *Current HIV Research*. **8(3):** 259–271.

Lau K.A. and Wong J.J.L. 2013. Current trends of HIV recombination worldwide. *Infectious Disease Reports*. **5(SUPPL.1):** 15–20.

Lemey P., Pybus O.G., Rambaut A., Drummond A.J., Robertson D.L., Roques P., Worobey M., and Vandamme A.M. 2004. The molecular population genetics of HIV-1 group O. *Genetics*. **167(3):** 1059–1068.

Lengauer T., Sander O., Sierra S., Thielen A., and Kaiser R. 2007. Bioinformatics prediction of HIV coreceptor usage. *Nature Biotechnology*. **25(12):** 1407–1410.

Levy J.A. 1993. Pathogenesis of human immunodeficiency virus infection. *Microbiological reviews*. **57(1):** 183–289.

Leys R., Vanderborght B., Haesevelde M.V., Heyndrickx L., Van Geel A., Wauters C., Bernaerts R., Saman E., Nijs P., Willems B., Taelman H., Van Der Groen G., Piot P., Tersmette T., Huisman J.S., and Van Heuverswyn H. 1990. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *Journal of Virology*. **64(3):** 1207–1216.

Li G., Piampongsant S., Faria N.R., Voet A., Pineda-Peña A.-C., Khouri R., Lemey P., Vandamme A.-M., and Theys K. 2015. An integrated map of HIV genome-wide variation from a population perspective. *Retrovirology*. **12(18):** 1–16.

Lihana R.W., Ssemwanga D., Abimiku A., and Ndembi N. 2012. Update on HIV-1 diversity in Africa: a decade in review. *AIDS Reviews*. **14(2):** 83–100.

Liu L., Li Y., Li S., Hu N., He Y., Pong R., Lin D., Lu L., and Law M. 2012. Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*. **2012** 1–11.

Llano A., Williams A., Olvera A., Silva-Arrieta S., and Brander C. 2013. Best-characterized HIV-1 CTL epitopes: the 2013 update. *HIV Molecular Immunology*. **2013** 3–25.

Lo D.Y.M. and Chan A.K.C. 2006. Setting up a polymerase chain reaction laboratory. In Lo, Dennis Y M and Chiu, Rossa W K and Chan, Allen K C, (Ed.) *Clinical Applications of PCR*. Methods in Molecular Medicine. Humana Press 999 Riverview Drive, Suite 208

Totowa, New Jersey 07512, pp. 11–18.

Lodmell J.S., Ehresmann C., Ehresmann B., and Marquet R. 2001. Structure and dimerization of HIV-1 kissing loop aptamers. *Journal of Molecular Biology*. **311(3):** 475–490.

Lole K.S., Bollinger R.C., Paranjape R.S., Gadkari D., Kulkarni S.S., Novak N.G., Ingersoll R., Sheppard H.W., and Ray S.C. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology*. **73(1):** 152–160.

Luk K.-C., Holzmayer V., Yamaguchi J., Swanson P., Brennan C.A., Ngansop C., Mbanya D., Gayum H., Djuidje M.N., Ndembi N., Kamdem D., Kaptué L., Gürtler L., Devare S.G., and Hackett J. 2007. Near full-length genome characterization of three additional HIV type 1 CRF13_cpx strains from Cameroon. *AIDS Research & Human Retroviruses*. **23(2):** 297–302.

Lundegaard C., Lamberth K., Harndahl M., Buus S., Lund O., and Nielsen M. 2008. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Research*. **36(Web Server issue):** W509–W512.

Mardis E.R. 2013. Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*. **6** 287–303.

Martinez-Picado J., Prado J.G., Fry E.E., Pfafferott K., Leslie A., Chetty S., Thobakgale C., Honeyborne I., Crawford H., Matthews P., Pillay T., Rousseau C., Mullins J.I., Brander C., Walker B.D., Stuart D.I., Kiepiela P., and Goulder P. 2006. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. *Journal of Virology*. **80(7):** 3617–3623.

McMichael A., Mwau M., and Hanke T. 2002. Design and tests of an HIV vaccine. *British Medical Bulletin*. **62(1):** 87–98.

Metzker M.L. 2010. Sequencing technologies—the next generation. *Nature Reviews Genetics*. **11(1):** 31–46.

Mifflin T.E. 1997. Control of Contamination Associated with PCR and Other Amplification Reactions. Molecular BioProducts, Inc. 9389 Waples St, San Diego, CA 92121, United States of America.

Migueles S.A., Sabbaghian M.S., Shupert W.L., Bettinotti M.P., Marincola F.M., Martino L., Hallahan C.W., Selig S.M., Schwartz D., Sullivan J., and Connors M. 2000. HLA B* 5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proceedings of the National Academy of Sciences*. **97(6):** 2709–2714.

Montavon C., Vergne L., Bourgeois A., Mpoudi-Ngole E., Malonga-Mouellet G., Butel C., Toure-Kane C., Delaporte E., and Peeters M. 2002. Identification of a new circulating recombinant form of HIV type 1, CRF11-cpx, involving subtypes A, G, J, and CRF01-AE,

in Central Africa. *AIDS Research & Human Retroviruses*. **18(3):** 231–236.

Moore J.P., Trkola A., and Dragic T. 1997. Co-receptors for HIV-1 entry. *Current Opinion in Immunology*. **9(4):** 551–562.

Morey M., Fernández-Marmiesse A., Castiñeiras D., Fraga J.M., Couce M.L., and Cocho J.A. 2013. A glimpse into past, present, and future DNA sequencing. *Molecular Genetics and Metabolism*. **110(1-2):** 3–24.

Moutaftsi M., Peters B., Pasquetto V., Tscharke D.C., Sidney J., Bui H.-H., Grey H., and Sette A. 2006. A consensus epitope prediction approach identifies the breadth of murine TCD8+-cell responses to vaccinia virus. *Nature Biotechnology*. **24(7):** 817–819.

Murakami T. 2012. Retroviral env glycoprotein trafficking and incorporation into virions. *Molecular Biology International*. **2012** 1–11.

Ndongmo C.B., Pieniazek D., Holberg-Petersen M., Holm-Hansen C., Zekeng L., Jeansson S.L., Kaptue L., and Kalish M.L. 2006. HIV genetic diversity in Cameroon: possible public health importance. *AIDS Research & Human Retroviruses*. **22(8):** 812–816.

Nielsen M., Lundegaard C., Worning P., Lauemøller S.L., Lamberth K., Buus S., Brunak S., and Lund O. 2003. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Science*. **12(5):** 1007–1017.

Nikolaitchik O.A., Dilley K.A., Fu W., Gorelick R.J., Tai S.-H.S., Soheilian F., Ptak R.G., Nagashima K., Pathak V.K., and Hu W.-S. 2013. Dimeric RNA Recognition Regulates HIV-1 Genome Packaging. *PLoS Pathogens*. **9(3):** e1003249.

Nisole S. and Saïb A. 2004. Early steps of retrovirus replicative cycle. *Retrovirology*. **1(9):**

Njai H.F., Gali Y., Vanham G., Clybergh C., Jennes W., Vidal N., Butel C., Mpoudi-Ngolle E., Peeters M., and Ariën K.K. 2006. The predominance of Human Immunodeficiency Virus type 1 (HIV-1) circulating recombinant form 02 (CRF02_AG) in West Central Africa may be related to its replicative fitness. *Retrovirology*. **3(40):** 11.

Nkengasong J., Peeters M., Musi S., Willems B., Ndumbe P., Delaporte E., Perret J., Piot P., van den Groen G., and others 1993. Antigenic evidence of the presence of the aberrant HIV-1 ANT70 virus in Cameroon and Gabon [letter]. *AIDS*. **7(11):** 1536–8.

Nkengasong J.N., Janssens W., Heyndrickx L., Fransen K., Ndumbe P.M., Motte J., Leonaers A., Ngolle M., Ayuk J., and Piot P. 1994. Genotypic subtypes of HIV-1 in Cameroon. *AIDS (London, England)*. **8(10):** 1405–1412.

Nsagha D., Njunda A., Kamga H., Assob J., and Bongkem E. 2012. HIV-1/HIV-2 co-infection among voluntary counselling and testing subjects at a regional hospital in Cameroon. *African Health Sciences*. **12(3):** 276–281.

Nyambi P., Heyndrickx L., Vereecken K., Burda S., De Houwer K., Coppens S., Urbanski M., Williams C., Ndumbe P., and Janssens W. 2002. Predominance of infection with HIV-1 circulating recombinant form CRF02_AG in major Cameroonian cities and towns. *AIDS*.

16(2): 295–296.

Ono A., Demirov D., and Freed E. 2000. Relationship between human immunodeficiency virus type 1 Gag multimerization and membrane binding. *Journal of Virology*. **74(11):** 5142–5150.

Ott M., Geyer M., and Zhou Q. 2011. The control of HIV transcription: keeping RNA polymerase II on track. *Cell Host & Microbe*. **10(5):** 426–435.

Pathak V.K. and Hu W.-S. 1997. "Might as well jump!" Template switching by retroviral reverse transcriptase, defective genome formation, and recombination. In *Seminars in Virology*. Academic Press, pp. 141–150.

Peeters M., Courgnaud V., Abela B., Auzel P., Pourrut X., Bibollet-Ruche F., Loul S., Liegeois F., Butel C., Koulagna D., Mpoudi-Ngole E., Shaw G.M., Hahn B.H., and Delaporte E. 2002. Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. *Emerging Infectious Diseases*. **8(5):** 451–457.

Peeters M., Liegeois F., Torimiro N., Bourgeois A., Mpoudi E., Vergne L., Saman E., Delaporte E., and Saragosti S. 1999. Characterization of a highly replicative intergroup M/O human immunodeficiency virus type 1 recombinant isolated from a Cameroonian patient. *Journal of Virology*. **73(9):** 7368–7375.

Pereira L.A., Bentley K., Peeters A., Churchill M.J., and Deacon N.J. 2000. A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic Acids Research*. **28(3):** 663–668.

Perrin L., Kaiser L., and Yerly S. 2003. Travel and the spread of HIV-1 genetic variants. *The Lancet Infectious Diseases*. **3(1):** 22–27.

Peters B., Bulik S., Tampe R., Van Endert P.M., and Holzhütter H.-G. 2003. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *The Journal of Immunology*. **171(4):** 1741–1749.

Peters B. and Sette A. 2005. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics*. **6(132):** 9.

Pincetic A. and Leis J. 2009. The mechanism of budding of retroviruses from cell membranes. *Advances in Virology*. **2009** 1–9.

Plantier J., Leoz M., Dickerson J., De Oliveira F., Cordonnier F., Lemee V., Damond F., Robertson D., and Simon F. 2009. A new human immunodeficiency virus derived from gorillas. *Nature Medicine*. **15** 871–872.

Powell R., Barengolts D., Mayr L., and Nyambi P. 2010. The Evolution of HIV-1 Diversity in Rural Cameroon and its Implications in Vaccine Design and Trials. *Viruses*. **2(2):** 639–654.

Powell R.L., Urbanski M.M., Burda S., Kinge T., and Nyambi P.N. 2009. High frequency of HIV-1 dual infections among HIV-positive individuals in Cameroon, West Central Africa. *Journal of Acquired Immune Deficiency Syndromes*. **50(1):** 84–92.

Powell R.L., Zhao J., Konings F.A., Tang S., Nanfack A., Burda S., Urbanski M.M., Saa D., Hewlett I., and Nyambi P.N. 2007. Identification of a novel circulating recombinant form (CRF) 36_cpx in Cameroon that combines two CRFs (01_AE and 02_AG) with ancestral lineages of subtypes A and G. *AIDS Research & Human Retroviruses*. **23(8):** 1008–1019.

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

Quiñones-Mateu M. and Arts E. 1999. Recombination in HIV-1: update and implications. *AIDS Reviews*. **1(2):** 89–100.

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):** 9–19.

Rerks-Ngarm S., Pitisuttithum P., Nitayaphan S., Kaewkungwal J., Chiu J., Paris R., Premsri N., Namwat C., de Souza M., Adams E., Benenson M., Gurunathan S., Tartaglia J., McNeil J.G., Francis D.P., Stablein D., Birx D.L., Chunsuttiwat S., Khamboonruang C., Thongcharoen P., Robb M.L., Michael N.L., Kunasol P., and Kim J.H. 2009. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *The New England Journal of Medicine*. **361(23):** 2209–20.

Rice A.P. 2010. The HIV-1 Tat Team Gets Bigger. *Cell Host and Microbe*. **7(3):** 179–181.

Rizzo J.M. and Buck M.J. 2012. Key principles and clinical applications of "next-generation" DNA sequencing. *Cancer Prevention Research*. **5(7):** 887–900.

Robertson D., Anderson J., Bradac J., Carr J., Foley B., Funkhouser R., Gao F., Hahn B., Kalish M., Kuiken C., Learn G., Leitner T., McCutchan F., Osmanov S., Peeters M., Pieniazek D., Salminen M., Sharp P., Wolinsky S., and Korber B. 1999. HIV-1 Nomenclature Proposal: A Reference Guide to HIV-1 Classification. In Korber, TM and Brander, C and Haynes, BF and Koup, R and Kuiken, C and Moore, JP and Walker, BD and Watkins, D, (Ed.) *Human Retroviruses and AIDS*. Theoretical Biology and Biophysics Group, Los Alamos, New Mexico, pp. IV–55–65.

Robinson H.L. 2002. New hope for an AIDS vaccine. *Nature Reviews Immunology*. **2(4):** 239–250.

Roques P., Robertson D.L., Souquière S., Apetrei C., Nerrienet E., Barré-Sinoussi F., Müller-Trutwin M., and Simon F. 2004. Phylogenetic characteristics of three new HIV-1 N strains and implications for the origin of group N. *AIDS (London, England)*. **18(10):** 1371–81.

Ross L., Parkin N., Chappey C., Fisher R., Clair M.S., Bates M., Tisdale M., and Lanier E.R. 2004. Phenotypic impact of HIV reverse transcriptase M184I/V mutations in combination with single thymidine analog mutations on nucleoside reverse transcriptase inhibitor

resistance. *AIDS*. **18(12):** 1691–1696.

Rubio A.E., Abraha A., Carpenter C.A., Troyer R.M., Reyes-Rodríguez Á.L., Salomon H., Arts E.J., and Tebit D.M. 2014. Similar Replicative Fitness Is Shared by the Subtype B and Unique BF Recombinant HIV-1 Isolates that Dominate the Epidemic in Argentina. *PLoS One*. **9(4):** e92084.

Salazar-Gonzalez J.F., Salazar M.G., Keele B.F., Learn G.H., Giorgi E.E., Li H., Decker J.M., Wang S., Baalwa J., Kraus M.H., and others 2009. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *The Journal of Experimental Medicine*. **206(6):** 1273–1289.

Samuel R., Paredes R., Parboosing R., Moodley P., and Gordon M. 2014. Minority HIV-1 drug-resistant mutations and prevention of mother-to-child transmission: perspectives for resource-limited countries. *AIDS Reviews*. **16(4):** 187–198.

Santoro M.M. and Perno C.F. 2013. HIV-1 Genetic Variability and Clinical Implications. *ISRN microbiology*. **2013** 1–20.

Santos A.F. and Soares M.A. 2010. HIV genetic diversity and drug resistance. *Viruses*. **2(2):** 503–531.

Schuitemaker H., van 't Wout A.B., and Lusso P. 2011. Clinical significance of HIV-1 coreceptor usage. *Journal of Translational Medicine*. **9 Suppl 1** S5.

Serio D., Aguzzi A., Briddon R.W., Flores R., Ghabrial S.A., and others 2012. Family - Retroviridae. In King, Andrew M.Q. and Adams, Michael J. and Carstens, Eric B. and Lefkowitz, Elliot J., (Ed.) *Virus taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier San Diego, pp. 477 – 495.

Shankarappa R., Margolick J.B., Gange S.J., Rodrigo A.G., Upchurch D., Farzadegan H., Gupta P., Rinaldo C.R., Learn G.H., He X., and others 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *Journal of Virology*. **73(12):** 10489–10502.

Sharp P.M., Bailes E., Robertson D.L., Gao F., and Hahn B.H. 1999. Origins and Evolution of AIDS Viruses. *Biological Bulletin*. **196(3):** 338–342.

Sharp P.M. and Hahn B.H. 2008. AIDS: Prehistory of HIV-1. *Nature*. **455(7213):** 605–606.

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

Shaw G.M. and Hunter E. 2012. HIV transmission. *Cold Spring Harbor Perspectives in Medicine*. **2(11):** a006965.

Sidney J., Assarsson E., Moore C., Ngo S., Pinilla C., Sette A., and Peters B. 2008. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Research*.

4(2): 14.

Siepel A.C., Halpern A.L., Macken C., and Korber B.T. 1995. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Research & Human Retroviruses*. **11(11):** 1413–1416.

Simon F., Mauclere P., Roques P., Loussert-Ajaka I., Muller-Trutwin M.C., Saragosti S., Georges-Courbot M.C., Barre-Sinoussi F., and Brun-Vezinet F. 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nature Medicine*. **4(9):** 1032–1037.

Soria-Guerra R.E., Nieto-Gomez R., Govea-Alonso D.O., and Rosales-Mendoza S. 2015. An overview of bioinformatics tools for epitope prediction: Implications on vaccine development. *Journal of Biomedical Informatics*. **53** 405–414.

Stephenson K. and Barouch D. 2013. A global approach to HIV-1 vaccine development. *Immunological Reviews*. **254(1):** 295–304.

Stopak K., de Noronha C., Yonemoto W., and Greene W.C. 2003. HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Molecular Cell*. **12(3):** 591–601.

Stranzl T., Larsen M.V., Lundegaard C., and Nielsen M. 2010. NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics*. **62(6):** 357–368.

Takehisa J., Zekeng L., Ido E., Mboudjeka I., Moriyama H., Miura T., Yamashita M., Gürtler L.G., Hayami M., and Kaptué L. 1998. Various types of HIV mixed infections in Cameroon. *Virology*. **245(1):** 1–10.

Takehisa J., Zekeng L., Ido E., Yamaguchi-Kabata Y., Mboudjeka I., Harada Y., Miura T., Kaptu L., and Hayami M. 1999. Human immunodeficiency virus type 1 intergroup (M/O) recombination in cameroon. *Journal of Virology*. **73(8):** 6810–20.

Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*. **30(12):** 2725–2729.

Taylor B.S., Sobieszczyk M.E., McCutchan F.E., and Hammer S.M. 2008. The challenge of HIV-1 subtype diversity. *The New England Journal of Medicine*. **358(15):** 1590–602.

Tebit D.M. and Arts E.J. 2011. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *The Lancet Infectious Diseases*. **11(1):** 45–56.

Tebit D.M., Nankya I., Arts E.J., and Gao Y. 2007. HIV diversity, recombination and disease progression: how does fitness "fit" into the puzzle? *AIDS Reviews*. **9(2):** 75–87.

Tebit D.M., Zekeng L., Kaptué L., Salminen M., Kräusslich H.-G., and Herchenröder O. 2002. Genotypic and phenotypic analysis of HIV type 1 primary isolates from western

UNIVERSITY OF VENDA
LIBRARY

Cameroon. *AIDS Research & Human Retroviruses*. **18(1):** 39–48.

Temin H.M. 1993. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proceedings of the National Academy of Sciences*. **90(15):** 6900–6903.

Tongo M. and Burgers W.A. 2014. Challenges in the design of a T cell vaccine in the context of HIV-1 diversity. *Viruses*. **6(10):** 3968–3990.

Tongo M., Martin D.P., Zembe L., Mpoudi-Ngole E., Williamson C., and Burgers W.A. 2013. Characterization of HIV-1 gag and nef in Cameroon: further evidence of extreme diversity at the origin of the HIV-1 group M epidemic. *Virology Journal*. **10(29):** 1–7.

Torimiro J.N., Carr J.K., Wolfe N.D., Karacki P., Martin M.P., Gao X., Tamoufe U., Thomas A., Ngole E.M., Birx D.L., McCutchan F.E., Burke D.S., and Carrington M. 2006. HLA class I diversity among rural rainforest inhabitants in Cameroon: identification of A*2612-B*4407 haplotype. *Tissue Antigens*. **67(1):** 30–37.

Triques K., Bourgeois A., Saragosti S., Vidal N., Mpoudi-Ngole E., Nzilambi N., Apetrei C., Ekwalanga M., Delaporte E., and Peeters M. 1999. High diversity of HIV-1 subtype F strains in Central Africa. *Virology*. **259(1):** 99–109.

Tscherning-Casper C., Dolcini G., Mauclère P., Fenyö E.M., Barré-Sinoussi F., Albert J., and Menu E. 2000. Evidence of the existence of a new circulating recombinant form of HIV type 1 subtype A/J in Cameroon. *AIDS Research & Human Retroviruses*. **16(13):** 1313–1318.

Turner B.G. and Summers M.F. 1999. Structural biology of HIV. *Journal of Molecular Biology*. **285(1):** 1–32.

UNAIDS 2015. How AIDS changed everything. Available at: http://www.unaids.org/sites/default/files/media_asset/MDG6Report_en.pdf [Accessed December 20, 2015].

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 & Human Retroviruses*. **26(1):** 109–15.

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

Vandekerckhove L.P.R., Wensing A.M.J., Kaiser R., Brun-Vézinet F., Clotet B., De Luca A., Dressler S., Garcia F., Geretti A.M., Klimkait T., Korn K., Masquelier B., Perno C.F., Schapiro J.M., Soriano V., Sönnerborg A., Vandamme A.-M., Verhofstede C., Walter H., Zazzi M., and Boucher C.A.B. 2011. European guidelines on the clinical management of HIV-1 tropism testing. *The Lancet. Infectious Diseases*. **11(5):** 394–407.

Vidal N., Peeters M., Mulanga-Kabeya C., Nzilambi N., Robertson D., Ilunga W., Sema H., Tshimanga K., Bongo B., and Delaporte E. 2000. Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *Journal of Virology*. **74(22):** 10498–10507.

W.H.O. 2013. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: summary of key features and recommendations, June 2013. Available at: http://apps.who.int/iris/bitstream/10665/85321/1/9789241505727_eng.pdf [Accessed February 25, 2016].

W.H.O. 2005. Interim WHO clinical staging of HIV/AIDS and HIV/AIDS case definitions for surveillance. African Region. Available at: http://www.who.int/hiv/pub/guidelines/clinicalstaging.pdf [Accessed July 12, 2015].

West B. 2011. Cameroon. Bradt Travel Guides.

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

Wolf K., Walter H., Beerenwinkel N., Keulen W., Kaiser R., Hoffmann D., Lengauer T., Selbig J., Vandamme A.-M., Korn K., and Schmidt B. 2003. Tenofovir resistance and resensitization. *Antimicrobial Agents and Chemotherapy*. **47(11):** 3478–3484.

Wolfe N.D., Switzer W.M., Carr J.K., Bhullar V.B., Shanmugam V., Tamoufe U., Prosser A.T., Torimiro J.N., Wright A., Mpoudi-Ngole E., and others 2004. Naturally acquired simian retrovirus infections in central African hunters. *The Lancet*. **363(9413):** 932–937.

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., Gilbert M.T.P., and Wolinsky S.M. 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature*. **455(7213):** 661–4.

Wu Y. 2004. HIV-1 gene expression: lessons from provirus and non-integrated DNA. *Retrovirology*. **1(13):** 10.

Yamaguchi J., Bodelle P., Vallari A.S., Coffey R., McArthur C.P., Schochetman G., Devare S.G., and Brennan C.A. 2004. HIV infections in northwestern Cameroon: identification of HIV type 1 group O and dual HIV type 1 group M and group O infections. *AIDS Research & Human Retroviruses*. **20(9):** 944–957.

Yamaguchi J., Ndembi N., Ngansop C., Mbanya D., Kaptué L., Gürtler L.G., Devare S.G., and Brennan C.A. 2009. HIV type 1 group M subtype G in Cameroon: five genome sequences. *AIDS Research & Human Retroviruses*. **25(4):** 469–73.

Yamaguchi J., Vallari A., Ndembi N., Coffey R., Ngansop C., Mbanya D., Kaptué L., Gürtler L.G., Devare S.G., and Brennan C.A. 2008. HIV type 2 intergroup recombinant identified

in Cameroon. *AIDS Research & Human Retroviruses*. **24(1):** 86–91.

Yamaguchi J., Vallari A., Ngansop C., Makamche F., Ndembi N., Mbanya D., Kaptué L., Gürtler L.G., Devare S.G., and Brennan C.A. 2010. Near full-length sequence of HIV type 1 subtype J strain 04CMU11421 from Cameroon. *AIDS Research & Human Retroviruses*. **26(6):** 693–697.

Yang X. and Yu X. 2009. An introduction to epitope prediction methods and software. *Reviews in Medical Virology*. **19(2):** 77–96.

Yedavalli V.S. and Jeang K.-T. 2011. Rev-ing up post-transcriptional HIV-1 RNA expression. *RNA Biology*. **8(2):** 195–199.

Zhang J., Chiodini R., Badr A., and Zhang G. 2011. The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*. **38(3):** 95–109.

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(Web Server issue):** W463–W465.

Zhao J., Tang S., Ragupathy V., Carr J.K., Wolfe N.D., Awazi B., and Hewlett I. 2010. Identification and genetic characterization of a novel CRF22_01A1 recombinant form of HIV type 1 in Cameroon. *AIDS Research & Human Retroviruses*. **26(9):** 1033–1045.

Zhou D., Mei Q., Li J., and He H. 2012. Cyclophilin A and viral infections. *Biochemical and Biophysical Research Communications*. **424(4):** 647–650.

Zhu P., Chertova E., Bess J., Lifson J.D., Arthur L.O., Liu J., Taylor K.A., and Roux K.H. 2003. Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. *Proceedings of the National Academy of Sciences*. **100(26):** 15812–15817.

Zhuang J., Jetzt A.E., Sun G., Yu H., Klarmann G., Ron Y., Preston B.D., and Dougherty J.P. 2002. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *Journal of Virology*. **76(22):** 11273–11282.

Zuñiga R., Lucchetti A., Galvan P., Sanchez S., Sanchez C., Hernandez A., Sanchez H., Frahm N., Linde C.H., Hewitt H.S., Hildebrand W., Altfeld M., Allen T.M., Walker B.D., Korber B.T., Leitner T., Sanchez J., and Brander C. 2006. Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control. *Journal of Virology*. **80(6):** 3122–3125.