UNIVERSITY OF KWAZULU-NATAL

GAG-PROTEASE DRIVEN VIRAL REPLICATION CAPACITY AMONG HIV-1 SUBTYPES: IMPLICATIONS FOR DISEASE PROGRESSION, EPIDEMIC SPREAD, AND VACCINE DESIGN

Omotayo Oluwaremilekun Farinre: 216073965

Supervisor: Prof. Thumbi Ndung'u

Co-Supervisor: Dr. Jaclyn K. Mann

A thesis submitted in fulfilment of the academic requirements for the degree of Doctor of Philosophy in Virology in the School of Laboratory Medicine and Medical Sciences, HIV Pathogenesis Programme, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal.

2020

This project represents original work done by the author and where others have made contributions it has been acknowledged in the text. The experimental work described in this thesis was performed at the HIV Pathogenesis Programme, in the Hasso-Plattner Research Laboratory, from June 2016 to March 2020 in the Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa, under the supervision of Prof. Thumbi Ndung'u and Dr. Jaclyn Mann.

Omotayo O. Farinre: _____ Date: _____

As the candidate's supervisors, we agree to the submission of this thesis.

Prof. Thumbi Ndung'u: _____ Date: _____

Dr. Jaclyn Mann: _____ Date: _____

Declaration

I, Omotayo Oluwaremilekun Farinre, declare that

i. The research reported in this thesis, is my original work, except where otherwise indicated.

ii. This thesis has not been submitted for any degree or examination at any other university or tertiary institution.

iii. This thesis does not contain other person's data, pictures, graphs, or other information, unless specifically stated as being sourced from other people.

iv. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other scientists. Where other written sources have been quoted, then:

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

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

v. Where I have reproduced a publication of which I am author, I have indicated in detail which part of the publication was written by myself alone and have fully acknowledged that publication.

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

Signed: _____ Date: _____

Publication(s)

Omotayo Farinre, Kamini Gounder, Tarylee Reddy, Marcel Tongo, Jonathan Hare, Beth Chaplin, Jill Gilmour, Phyllis Kanki, Jaclyn K. Mann, Thumbi Ndung'u. *Subtype-specific differences in Gag-protease replication capacity of HIV-1 isolates from East and West Africa. (RETV-D-21-00003R1). Journal of Retrovirolgy.*

Statement

This manuscript has been reproduced in part non-continuously throughout the thesis. The candidate performed the experiments described in this paper, and where others made contributions it has been duly acknowledged in the text. The candidate drafted this publication in full and it was reviewed by co-authors.

Omotayo O. Farinre: _____ Date: _____

Conference contributions

1. **Omotayo Farinre**, Kamini Gounder, Marcel Tongo, Jonathan Hare, Jill Gilmour, Jaclyn Mann, & Thumbi Ndung'u. HIV-1 Isolates from West Africa Show Higher Fitness than to those from East Africa. The Frontiers in Retrovirology Conference, 11-13 September 2018, Leuven, Belgium, Abstract 157,
   Retrovirology 2018, **15 (Suppl 1):** O17. Presented by Jaclyn K. Mann

2. **Omotayo Farinre**, Kamini Gounder, Marcel Tongo, Jonathan Hare, Jill Gilmour, Jaclyn Mann, & Thumbi Ndung'u. HIV-1 Recombinant Viruses from West Africa Display Higher Replication Capacity Compared to Recombinants from East Africa. HIV research for prevention (HIVR4P): From research to impact Conference on the 21-25 October 2018, Madrid, Spain. Abstract Poster: P11.05. Presented by Omotayo Farinre.

Dedication

I dedicate this work to my personal Lord and Saviour, Jesus Christ for giving me the strength push to the end even when my own wisdom and efforts failed me.

I also dedicate this work to my wife and children; *Funmi, Tobi, Timi and Tomisin*. Words are insufficient to express my gratitude for your patience, forbearance, and support throughout this journey.

Thank you.

Acknowledgements

This work was supported by the Poliomyelitis Research Foundation (PRF), the South African Department of Science and Innovation through the National Research Foundation (South African Research Chairs Initiative), the Victor Daitz Foundation and the International AIDS Initiative ((UKZNRSA1001). Additional funding was provided by the Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE), a DELTAS Africa Initiative [grant # DEL-15-006]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant # 107752/Z/15/Z] and the UK government.

I would also like to thank Professor Thumbi Ndung'u (Prof.) for granting me the opportunity to pursue my PhD at the HIV Pathogenesis Programme. Your mature, gentle, patience, kind nature and approach towards managing me as student and giving me some room to make mistakes and learn from them have been exceptional and exemplary. I am particularly inspired by your keen desire to develop critical mass of scientists on the African continent, and this shows in how you mentor your students. I am grateful to God for your support throughout this journey without which I may not have had the courage to finish.

I also want to thank God for my Co-supervisor in the person of Dr. Jaclyn K. Mann (Jaclyn). If it were not for you taking my request to pursue a PhD with HPP to Prof. Thumbi, I would not have been in HPP in the first place. I cannot forget in a hurry your contribution to my growth, your laser precision approach towards research, critical thought, strict principles and very importantly time-management and planning qualities I still hope to learn from you.

I am grateful to our collaborators who graciously donated plasma samples for this project such as Dr. Marcel Tongo who donated the Cameroonian samples, Dr. Hannah O. Ajoge who donated the Nigerian samples, and to Prof. Phyllis Kanki who donated the Senegalese samples and Beth Chaplin who assisted in amplifying the Senegalese samples. I will also thank Dr. Jill Gilmour, Dr. Jonathan Hare and the entire IAVI Protocol-C team for believing in this work enough to graciously donate the East African samples used for this study.

I would like to thank our Laboratory Manager at the Hasso-Plattner in HPP, in the person of Keshni Hiramen. She and her team work very hard to support postgraduate students who use the laboratory and to ensure that everything I needed for my experiments was made available to me and were in good working order. I would also like to thank my fellow students and colleagues whom we weathered the difficult storm of long and failed experiments together. I would like to thank my sister from Kenya Dr. Doty Ojwach for all her encouragement and putting me through most of the laboratory techniques I am now familiar with, Noni whom we are bracing this PhD storm together, Dr. Funsho Ogunshola, Dr. Omolara Baiyeghuni, Dr. Kewreshni Naidoo. I would like to appreciate the administrative team of Mrs. Tarryn Leslie and Mr. Silvester Olupot for their support. I would also like to say a big thank you to all my office colleagues Sithembile, Qiniso, Miriam, Anele, Nombali, Trevor and Delon. You all made my stay memorable.

To my mother Mrs. Tolulope Oluwatosin Farinre who is my number one fan, always praying, cheering, rooting, and believing in me even when I doubt myself. Thank you so much mum. To my sisters Ayes and Temiloobs who have been calling me Dr. Tayo for the last two years, and I have been telling them it is premature, well your wishes are coming true, thank you both for your love and support.

Abstract

Introduction:

The HIV-1 epidemic in sub-Saharan Africa is heterogeneous with diverse unevenly distributed subtypes and regional differences in prevalence. Subtype-specific differences in disease progression rate and transmission efficiency have been reported, but the underlying biological mechanisms have not been fully characterized. In this study, I tested the hypothesis that the subtypes prevalent in the East African epidemic, where adult prevalence rate is higher, have lower viral replication capacity (VRC) than their West African counterparts where adult prevalence rates are lower.

Materials and methods:

Gag-protease sequencing was performed on plasma samples from 213 and 160 antiretroviral-naïve participants from West and East Africa, respectively. Online bioinformatic tools were used to infer HIV-1 subtypes and recombination patterns. Replication capacities of patient-derived *gag-protease* chimeric viruses from West (n=178) and East (n=114) Africa were determined using a green fluorescent protein reporter-based cell assay. Subtype and regional differences in viral replication capacity and amino acid variants impacting replication capacity were identified using appropriate statistical methods.

Results:

Subtypes identified in West Africa were CRF02_AG (65%, n=139), G (7%, n=15), A3 (5%, n=10), other CRFs (12%, n=26), various pure subtypes (9%, n=19) and A1G recombinants (2%, n=4). Subtypes A1 (64%, n=103), D (22%, n=35), AD (11%, n=17) and AC (3%, n=5) were identified in East Africa. Chimeric viruses from West Africa had significantly higher VRC compared to those from East Africa ($p < 0.0001$), with subtype-specific differences found among strains within West and East Africa ($p < 0.0001$). Recombination patterns showed a

preference for subtypes D, G or J rather than subtype A in the p6 region of *gag*, with evidence that subtype-specific differences in this region impact viral replication capacity. Furthermore, the Gag A83V polymorphism was associated with reduced viral replication capacity in CRF02_AG (median < 0.86). HLA-A*23:01 ($p = 0.0014$) and HLA-C*07:01 ($p = 0.002$) were associated with significantly lower viral replication capacity in subtype A infected individuals from East Africa.

Conclusion:

Overall, the data showed that viruses from West Africa displayed higher replication capacity than those from East Africa, which is consistent with the hypothesis that lower viral replication capacity is associated with higher population prevalence.

Table of contents

List of Figures

List of Tables

Abbreviations

ADDC	Antibody dependent cellular cytotoxicity
AIDS	Acquired Immuno-deficiency syndrome
APC	Antigen presenting cells
APOBEC3G	Apolipoprotein-B mRNA editing enzyme catalytic polypeptide-like 3G
ART	Anti-retroviral therapy
ARV	AIDS Related Virus
BSA	Bovine serum albumin
CA	Capsid
CCR5	Chemokine receptor 5
CD4+	Cluster of differentiation 4 positive
CD8+	Cluster of differentiation 8 positive
CDC	Centres for disease control and prevention
cpx	Complex
cpz	Chimpanzee
CRF	Circulating recombinant form
CTL	Cytotoxic T lymphocytes
CXCR4	Chemokine receptor 4
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DRC	Democratic republic of Congo
dsDNA	Double stranded Deoxyribonucleic acid
EBF	Early B cell factor
env	Envelope
ER	Endoplasmic reticulum
FDC	Follicular dendritic cells
Gag	Group specific antigen
GM-CSF	Granulocyte-macrophage colony stimulating factor
GWAS	Genome wide association studies
gor	Gorilla
gp	Glycoprotein
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
HLA	Human leucocyte antigen
HMA	Heteroduplex mobility assay
HTLV	Human T-lymphotropic virus
IDU	Intravenous drug user
IFN-γ	Interferon gamma
Ig	Immunoglobulin
IL	Interleukin
IN	Integrase
LAS	Lymphadenopathy syndrome
LAV	Lymphadenopathy associated virus

LTR	Long Terminal Repeat
MA	Matrix
MACS	Multicentre AIDS Cohort Study
MHC	Major histocompatibility complex
MIP	Macrophage inflammatory protein
mRNA	Messenger ribonucleic acid
MSMs	Men that have sex with men
MTCT	Mother to child transmission
mtDNA	Mitochondrial deoxyribonucleic acid
nef	Negative factor
nt	Nucleotide
Nκ	Natural killer
ORF	Open reading frame
PAMPS	Pathogen associated molecular patterns
PAX	Paired box gene
PBS	Phosphate buffered saline
PFA	Paraformaldehyde
PIC	Pre-integration complex
Pol	Polymerase
PR	Protease
Prep	Pre exposure prophylaxis
PRRs	Pathogen recognition receptors
Ptt	*Pan troglodytes troglodytes*
rev	Regulator of expression of virion
RNA	Ribonucleic Acid
RNAse H	Ribonuclease H
RT	Reverse transcriptase
SIV	Simian immunodeficiency virus
smm	Sooty mangabey monkey
SU	Surface protein
tat	Transactivation protein
TCR	T cell receptor
TLR	Toll-like receptors
TM	Transmembrane protein
tMRCA	Time to most recent common ancestor
TNF-α	Tumour necrosis factor alpha
TRIM5α	Tripartite motif 5α
URF	Unique recombination form
VCT	Voluntary counselling and testing
vif	Viral infectivity factor
vpr	Viral protein R
vpu	Viral protein U
VRC	Viral replicative capacity
WHO	World Health Organisation.

1 INTRODUCTION

1.1 Background

Since the identification of the human immunodeficiency virus (HIV) as the etiologic agent of the acquired immunodeficiency syndrome (AIDS) almost four decades ago (1-3), over 60 million people have been infected globally, with an excess of 35 million currently living with the virus (3-6). AIDS is a set of symptoms and illnesses that develop because of advanced HIV infection, whereby the host immune system becomes wholly compromised. In 1981, the United States Centers for Disease Control and Prevention (CDC) reported a characteristic set of symptoms which included *Pneumocystis carinii* pneumonia and Kaposi's sarcoma that appeared to be prevalent within a community of homosexual men in the United States (7-9). These would later turn out to be the first set of AIDS-related cases reported (7). Since then, the epidemic has grown widely, and extended beyond homosexual men into the general population. Sub-Saharan Africa currently accounts for about 70% of the global burden of HIV infections (2-5). Over the years, the development of highly active antiretroviral therapy (HAART) and increased public awareness programs facilitated a significant reduction of the epidemic globally. However, the financial burden of life-long use of medication, patient non-adherence, social stigma among others are some of the challenges impeding the effectiveness of HAART globally. To complement the efficacy of HAART, the development of a safe, effective, and affordable vaccine against HIV infection remains a priority to reduce viral transmission. Unfortunately, this has remained elusive due to the characteristic genetic diversity of HIV (2, 5, 10). I will give a brief history of HIV from its origins and discovery, its biology, pathogenesis, disease progression, host immune responses and conclude with the rationale for the current project in this chapter.

1.2 Initial isolation of HIV

HIV is a lentivirus, a genus within the *retroviridae* family (1, 3). In 1983, Francoise Barre Sinuossi and colleagues reported the isolation of a virus from the lymph node of a homosexual man suffering from lymphadenopathy syndrome (LAS) which they named the Lymphadenopathy Associated Virus (LAV) (1, 9). Thereafter, the Gallo and Levy groups also independently reported the isolation of a human T-lymphotropic virus (HTLV III) (11), and an AIDS-related virus (ARV) (12) from patients within their respective studies. However, further characterization of these isolates showed that they shared similar biological properties with the LAV, indicating that they were the same (1, 13). To resolve the problem of having different names for the same isolate, a sub-committee was inaugurated and empowered by the International Committee on Taxonomy of Viruses to propose a unifying name. The proposed and adopted name for the isolates became the human immunodeficiency virus (HIV) (1, 14).

1.2.1 Origins of HIV

Following the discovery and isolation of HIV, it became imperative to understand the route of its introduction into the human population and factors that aided its efficient host-adaptation (3). HIV represents multiple distinct cross-species transmissions; lentiviruses were known to be more prevalent in non-human mammals prior to the discovery of HIV (3, 15-18) as summarised in Table 1.1. Retrospective sequence and phylogenetic analyses revealed distinct cross-species transmission events that led to the introduction of HIV into the human population (16). The search began for animal lentiviruses that could have been the source of HIV into the human population. Phylogenetic analysis showed that HIV was most closely related to lentiviruses found in non-human primates called simian immunodeficiency viruses (SIVs) (3, 18, 19).

Table 1.1: Lentivirus Family

Virus	Host	Primary cell type infected	Clinical disorder
Equine infectious anemia virus (EIAV)	Horse	Macrophages	Cyclical infection in the first year: hemolytic anemia and sometimes encephalopathy
Visna virus	Sheep	Macrophages	Encephalopathy
Caprine arthritis-encephalitis virus (CAEV)	Goat	Macrophages	Immune deficiency, encephalopathy
Bovine immune deficiency virus (BIV)	Cow	Macrophages	Lymphadenopathy, lymphocytosis, CNS disease
Feline immunodeficiency virus (FIV)	Cat	T lymphocytes	Immune deficiency
Simian immunodeficiency virus (SIV)	Primate	T lymphocytes	Immune deficiency, encephalopathy
Human immunodeficiency virus (HIV)	Human	Macrophages and T lymphocytes	Immune deficiency, encephalopathy

*Table 1 adapted from (12)

1.2.2 Primate reservoir theory

Forty different species of primates had been shown to harbour SIVs (6, 18, 20), indicating primates as reservoirs of the ancestral viruses that were transmitted into the human population (3, 15). Evidence to support this theory includes, similarities in the viral genome organization between SIVs and HIV, phylogenetic relatedness, the prevalence of SIVs in their natural hosts, a geographic coincidence of natural hosts of SIVs and the epicentre of the epidemic as well as plausible routes of transmission (15) as shown in Figures 1.1 and 1.2.

1.2.3 Primate sources of HIV-1 & HIV-2

The origin of HIV-1 was resolved in 1989 when a closely related virus was isolated from captive chimpanzees *Pan troglodytes troglodytes* (*Ptt*) in Gabon; the virus was designated SIV*cpz* (20). Corbet and colleagues were able to isolate additional samples of SIV*cpz* from wild chimpanzees (*Ptt*) (21, 22). Phylogenetic analyses showed that this virus gave rise to the Group M viruses responsible for the global pandemic and the N group that is largely restricted in West-Central Africa, while group O viruses were introduced into the human population by gorillas -*Gorilla gorilla* (SIV*gor*). The source of HIV-2 was also confirmed in the same year to be from sooty mangabey monkeys (SIV*smm*) (3, 16). Phylogenetic analyses of mitochondrial DNA (mtDNA) obtained from African primates by Gagneux et al., (1999) showed four distinct sub-species of chimpanzees namely *Pan troglodytes verus*, *Pan troglodytes ellioti*, *Pan troglodytes troglodytes* and *Pan troglodytes schweinfurthii* and two main sub-species of gorilla- *Gorilla gorilla* and *Gorilla beringei* all spread across sub-Saharan Africa (10, 20, 23). My discussion hereafter will focus on HIV-1 since it is the subject of this study.

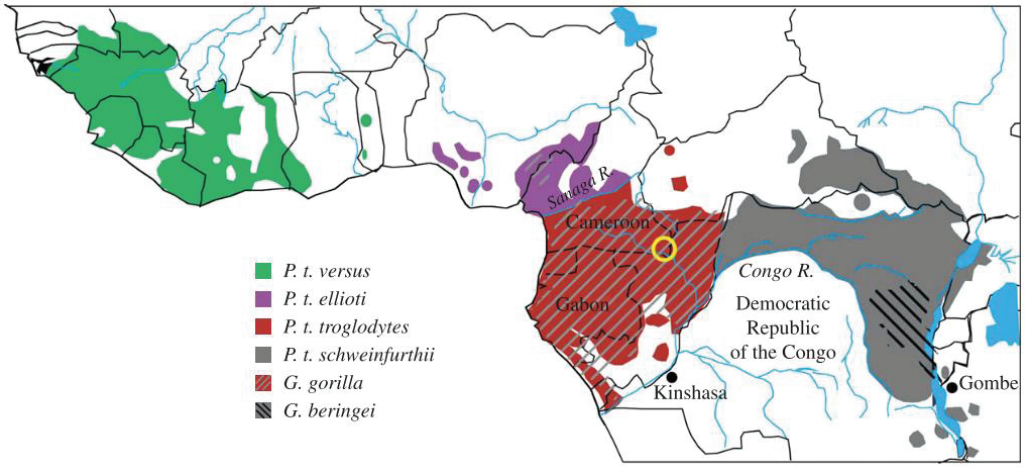


Figure 1.1: Distribution of non-human primates in West-Central Africa

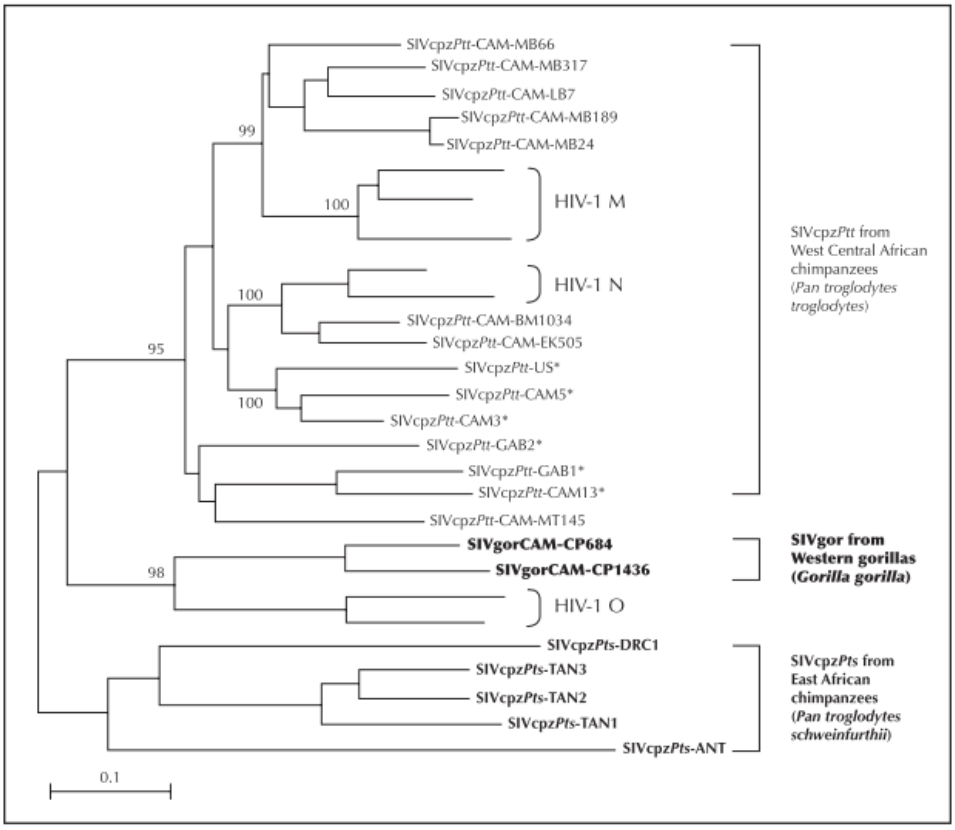


Figure 1.2: Phylogenetic analysis showing SIV and HIV relatedness

Figure 1.1: Map of West-Central Africa, showing the ranges of chimpanzee subspecies (colour coded). The gold circle denotes the region in southeast Cameroon where SIVcpz strains closely related to HIV-1 group M are found. (20). and **Figure 1.2** Evolutionary relationship of SIV*cpz*, SIV*gor*, and HIV-1 strains based on neighbour-joining phylogenetic analysis of partial *env*/*nef* sequences. Horizontal branch lengths are drawn to scale (3).

1.2.4 HIV-1 host adaptation

Fitness is a parameter that measures the sum of all characteristics of a pathogen that enables it to successfully transmit its genetic material from one generation to another within its immediate environment and has been associated with the rate of transmission in HIV-1 infections (24, 25). HIV-1 effectively adapted to its human hosts through several mechanisms including the transmission of mutation-prone viral particles that exist as a population of genetic variants or quasi-species within a single infection leading to positive selection of the fittest genetic variants (22, 26). HIV-1 has a short generation time, high mutation rate, high copy numbers and recombination of segments of its genome between circulating viruses of different clades (27-29) facilitates its fitness. Variation in its envelope (*env*) sequence length and glycosylation patterns facilitate the utilization of receptors and co-receptors for viral entry into target cells. High levels of intra-host amino acid variation within its hyper-variable (V1V2) region facilitates viral escape from immune effector functions such as antibodies and cytotoxic T-lymphocytes (CTLs) (26). Other examples of viral adaptation mechanisms include the inhibition of host antiviral immunity proteins such as the apolipoprotein-B mRNA editing enzyme catalytic polypeptide-like 3G (APOBEC3G) which inhibits the process of reverse transcription in HIV-1 (30), tripartite motif 5α (TRIM5α) which inhibits viral uncoating step (31) and tetherin which inhibits viral budding and the release of virions from infected cells (32). The entire sum of the viral characteristics plus the host cell environment, which includes the collection of $CD4^+$, $CD8^+$ cells and antibodies at the disposal of the host immune system ultimately determines the balance of viral adaptation whether positively or negatively (27).

1.2.5 Treatment and prevention strategies

Antiretroviral therapy (ART) lowers patient viral load to undetectable levels, thereby reducing the risks of further transmission (33); a concept called treatment as prevention (34). Following positive diagnosis of HIV-1 infected individuals, the World Health Organization (WHO) recommends the test and treat initiative, which simply encourages global public health systems to treat all infected individuals immediately, and not wait for depletion of $CD4^+$ cells to $200/mm^3$ (35, 36). ART, however is unable to completely eradicate virus infected cells from the body (33), but successfully targets different steps in the viral replication cycle. There are currently, six classes of ARTs in clinical use namely: the nucleoside and non-nucleoside reverse transcription inhibitors, protease inhibitors, integrase inhibitors, the fusion inhibitor (enfuvirtide (T-20)), and the chemokine receptor antagonist (maraviroc) (37). ART regimens vary from country to country (33). Prevention strategies recommended by the WHO (35, 38) include advocacy for sexual health education, access to voluntary counselling and testing (VCT) services, advocating use of condoms between sexual partners and high-risk groups like sex workers and men who have sex with men (MSMs). Biomedical interventions have also not been left out in the prevention strategy against HIV-1 infection and transmission. Pre-exposure prophylaxis (PrEP) is given to high-risk groups like health workers, sex workers and MSMs (39-42). Also, the administration of short course of nevirapine during labour to prevent mother to child transmissions (MTCT) (43). Vaginal microbicides developed for women have also been shown to be moderately effective against HIV-1 acquisition, although not efficacious enough for licensure (39). Unfortunately, vaccine development efforts have failed to overcome the genetic diversity of HIV-1, although some concepts are showing promise and undergoing further testing (5, 44-46).

1.3 HIV epidemiology

The HIV-1 epidemic remains a global public health concern with an estimated 37.9 million people currently infected (47). The pattern of transmission varies across regions based on adult prevalence and viral genetic composition. Some regions like Europe, Americas, and Australia experience a concentrated pattern of transmission where infection is high among key populations such as men that have sex with men (MSM) and intravenous drug users (IDUs), but generally low within the general population. Even in sub-Saharan Africa where the population prevalence rate of HIV-1 is high, key population groups are disproportionately affected (48-53). Global adult prevalence is reported to have stabilized at 0.8% between 2001 and 2007 (49), with a decline in new infections largely due to the successful expansion of ART coverage in poor developing countries resulting in the aversion of about 5.2 million deaths between 1995 and 2010 (54, 55). Adult HIV prevalence is highest within sub-Saharan Africa (5%), followed by the Caribbean (1%), Eastern Europe and Central Asia (0.8%), Central and South America (0.5%), South and Southeast Asia (0.3%), and East Asia (0.1%) (50). Heterosexual transmission is largely responsible for the spread of the virus globally (48, 49), although other modes have been reported in key populations such as IDUs and MSMs, mother-to-child transmission (MTCT) and sharing of contaminated equipment in clinical settings (56). Sub-Saharan Africa bears about 70% of the global burden of HIV-1 infection with an estimated 25.6 million currently living with the virus (47). However, prevalence rates in sub-Saharan Africa are heterogenous. Specifically, West Africa is reported to have an adult prevalence rate below 2%, except for Nigeria and Côte d'Ivoire with a prevalence of 5% and 11% respectively, while East Africa on the other hand has a higher prevalence rate above 5%, with Southern Africa having the highest prevalence rate of about 20% (48, 57-59).

1.3.1 HIV global subtype distribution

A major characteristic of HIV-1 is its genetic diversity. The virus evolves spontaneously in the absence of selection pressure and is found in infected persons in the form of a swarm of genetically diverse population termed quasi-species (60, 61). Viral evolution and thus diversity are driven by the multiple cross-species transmission events from non- human primates, the error prone RT (10^{-4}/nt) enzyme, high viral turnover (10^{10} particles/day), high mutation rate (3 × 10^{-5} mutations per base pair per cycle) and host immune-selection (24, 61, 62). HIV-1 subtype distribution varies from region to region. The most recent estimates show subtype C as the most prevalent subtype globally, accounting for about 46.6% of all infections, while B (12.1%), A (10.3%), CRF02_AG (7.7%), CRF01_AE (5.3%), G (4.6%), and D (2.7%) are less common. Subtypes F, H, J and K combined account for 0.9%. Other circulating recombinant forms CRFs account for 3.7%, while unique recombinant forms (URFs) account for 6.1% of total infections (51). These subtypes are unevenly distributed globally with subtype A being predominant in Central and East Africa, and Eastern Europe, subtype B in Western and Central Europe, Americas, Australia, North Africa, and Middle East. Subtype C is predominant in Southern Africa, India, and Ethiopia. Subtype D on the other hand is locally predominant in East Africa especially Uganda, while G is predominantly found in West Africa, particularly Nigeria. CRF02_AG is the most predominant subtype in West Africa (6, 57, 60-64). The epidemic in sub-Saharan Africa is notably heterogeneous, with West-central Africa having the most genetically diverse epidemic globally. This genetic variation in HIV may have an impact on accurate diagnosis, disease progression, treatment, and development of a safe and effective vaccine (52, 63, 65-67).

1.3.2 HIV-1 classification system

Characteristics of viruses like transmissibility, pathogenesis and immunogenicity are determined by viral evolution and host genetic composition. Viral genetic factors play a critical role in virulence, accurate diagnosis, vaccine design and epidemiological outlook required to predict and possibly prevent future outbreaks (63, 68). Prior to 1992, HIV-1 strains were classified based on their geographical origin; namely North American and African variants. However, as more isolates began to emerge, a robust classification system based on phylogenetic analyses of *env* coding sequences was implemented (63, 68). HIV-1 is classified into 4 groups namely M (Major) (69), O (Outlier) (70), N (Non-major) (71), and a rare P (72) group. Group M variants are the most common and thus responsible for global infection, while other groups are restricted to West-Central Africa (3, 73, 74). Group M viruses are classified into 10 discrete clades or subtypes namely (A-D, F-H, J, K) as shown in Figure 1.3 and very recently, subtype 'L' was identified from archived samples collected from the Democratic Republic of Congo (DRC) (75). The first set of subtypes classified were A, B, C, D & E. These isolates varied in their *env* sequences by about 30% and by 14% in the *gag* sequences. Sub-subtypes are a category of viral isolates closely related to a specific subtype lineage, but not distant enough to be identified as a new subtype. For example, subtype A has sub-subtypes A1-A7 (76, 77) and recently subtype D has been shown to have sub-subtypes D1-D3 (77) (Figure 1.4). An alternative classification system is based on viral co-receptor usage (78), however, classification based on viral genetic variation is more commonly used. Observed increase in complexity of newly identified HIV-1 strains called for the re-evaluation of HIV-1 nomenclature and classification. A 1999 meeting to resolve this challenge was convened to properly articulate a new classification system that would maintain historical consistencies with already existing literature; details of the outcomes of this meeting are fully described in (68), which surmised that phylogenetic and distance analysis must be used to identify new subtypes

and a branching index of 0.66 has been used as an acceptable cut-off to determine genetic distance (76, 78).

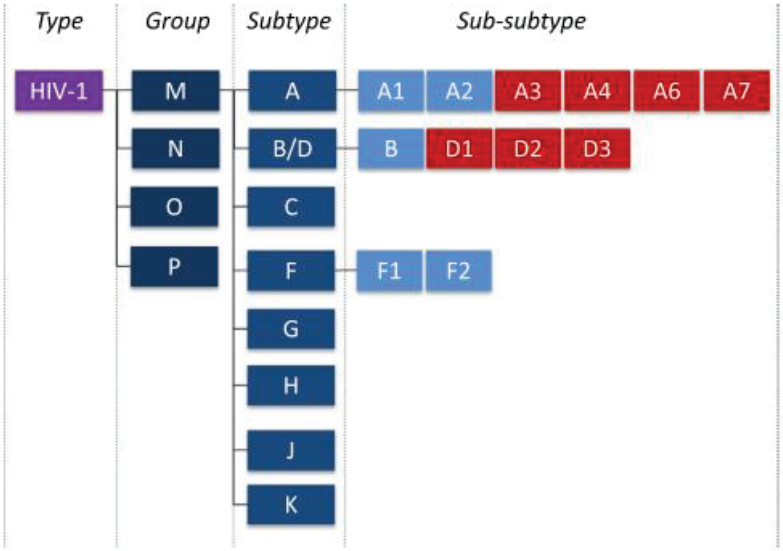


Figure 1.3: Updated classification of HIV-1 subtypes

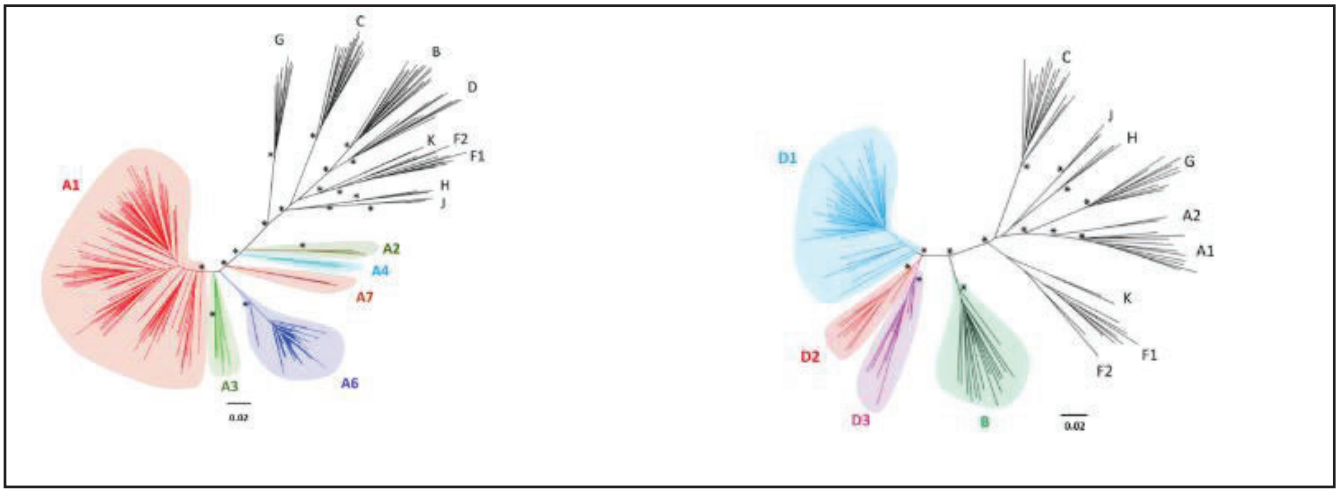


Figure 1.4: New sub-sub types for A and D

Figure 1.3 Updated classification of HIV-1 group M lineages. **Figure 1.4: Maximum** likelihood phylogenetic analysis of near full-length genome sequences of subtypes A and subtype D (77).

1.3.2.1 Circulating recombinant forms

Circulating recombinant forms (CRFs) of HIV-1 group M represent a recombinant lineage that plays a critical role in regional pandemics. A CRF must share an identical genome structure that descends from the same recombination event (68). Generally, when people think of CRFs, the tendency is to presume that the recombination occurs because of fragment breakage and re-joining as is the case for double stranded deoxy-ribonucleic acid (dsDNA); however, this is inaccurate. CRFs are formed due to phenomenon known as template switching which occurs during the reverse transcription process in a dually infected cell. This is what leads to the recombination event (57, 60, 76). The naming system for CRFs include the acronym 'CRF', followed by a number to reflect the order in which it was identified in relation to previously identified CRFs and two letters to identify the subtype composition; for example, CRF02_AG is the second CRF identified and has components of subtype A and G in its genome. In the event that there are more than 2 subtype components within the genome it is referred to as a complex and designated 'cpx' for example CRF11_cpx (68). Currently, over a 100 CRFs have been identified according to the Los Alamos HIV database (https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html). CRF01_AE and CRF02_AG are the most prominent. Subtype 'E' was initially identified in the *env*, but was not identified in *gag*, hence it did not 'meet' the criteria for becoming an independent subtype. In *gag*, it appeared to be more closely related to subtype A. At the 2000 HIV-1 classification meeting, it was proposed that subtype 'E' be designated CRF01_AE because it represented a putative recombinant (68), and to keep with historical consistencies of previous literature referring to it as subtype E. It is proposed that CRF01_AE formed early during the HIV-1 epidemic, and it is critical in regional epidemics in Asia and south east Asia, though it emerged from Central Africa (6, 68). CRF02_AG also plays a critical role in regional epidemic in Western and Central Africa. It is the most prevalent CRF, appearing to also have been formed

by an early recombination event at the start of the epidemic. It was initially thought to be a 'pure' subtype, while subtype 'G' was thought to be a recombinant. However, time to most recent common ancestor (tMRCA) analysis has shown that the ancestral sequences of pure subtypes A and G predate that of CRF02_AG. This confirms that CRF02_AG is indeed a recombinant of pure subtype A and G (6, 71, 79). Figure 1.5 shows the genome map of CRF01_AE and CRF02_AG.

1.3.2.2 Unique recombinant forms

Unique recombinant forms (URFs) refer to unique mosaic sequences that currently lack evidence of onward transmission and thus do not meet the well-defined criteria to be classified as CRFs. They account for 6.1% of global HIV infections and are also critical to regional infections. For example, recombinants of AC, AD, CD and ACD are estimated to constitute about 30% of East Africa's infections in countries like Kenya, Rwanda, Tanzania, and Uganda, while BF is predominant in South America, particularly Brazil and Argentina (51, 56, 80, 81).

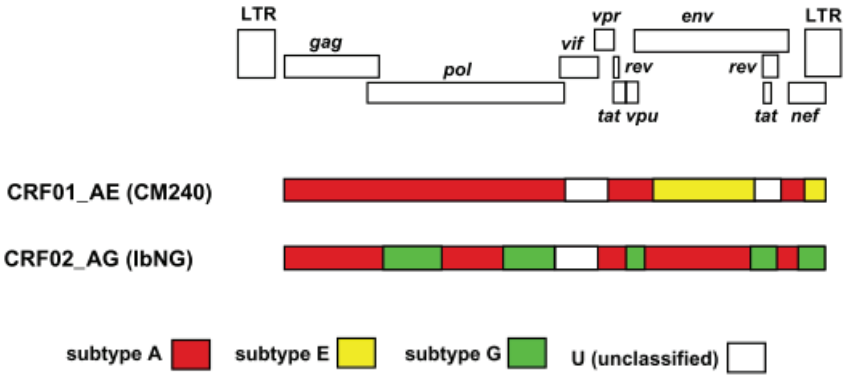


Figure 1.5: Mosaic genome structures of CRFs

Figure 1.5: Mosaic genome structures of CRF01_AE and CRF02_AG. An alignment of representative near full-length strains was used and gap-stripped prior to all analysis. The HIV-1 genome map shows the position of each open reading frame in the gap-stripped multiple alignment. Below the genome map, each bar represents the mosaic pattern each CRF, adapted from (82)

1.4 Morphology of HIV-1

HIV-1 particles are circular in nature and comprise of 3 major components namely an envelope lipid bilayer, a viral matrix, and the core. The particle is surrounded by a lipoprotein membrane that is about 100-150 nm in diameter (83) containing 72 glycoprotein complexes integrated into this membrane. The lipoprotein membrane comprises of external glycoprotein trimers gp120 surface protein (SU) that are anchored by the transmembrane protein gp41 (TM) (84). The matrix p17 (MA) is anchored inside of the viral lipoprotein membrane. The viral capsid p24 (CA) encases two copies of HIV-1 RNA genome (83-85). The HIV-1 RNA is part of a protein-nucleic acid complex composed of the nucleoprotein p7 and the reverse transcriptase p66 (RT) (85). The viral core contains the enzymatic machinery for replication, which are the reverse transcriptase (RT), integrase p32 (IN) and the protease p11 (PR) (86-88). A graphical representation of HIV-1 particle can be seen in Figure 1.6.

1.4.1 HIV-1 genome organisation

The HIV-1 genome is organized into nine open reading frames (ORF) just short of 10 kilo bases in length (84, 89). The composition of the genome can be broadly classified into structural and accessory genes. The group-specific antigen (*gag)* gene encodes the matrix, capsid, nucleocapsid and p6 proteins; while the envelope (*env*) gene encodes the gp120 and gp41 (84, 89). The polymerase gene (*pol*) codes the enzymatic machinery of the virus that facilitates effective replication which include the enzymes. All three genes make up the structural genes (89, 90). The virus also has 6 accessory/regulatory genes which include virion infectivity factor (*vif*), viral protein R (*vpr*), viral protein U (*vpu*), trans-activation of transcription (*tat*), regulator of expression of virion (*rev*) and the negative factor (*nef*). The proteins coded for by the genes all play important roles in the replication cycle of the virus (90, 91) and are well described in (86, 92). Figure 1.7

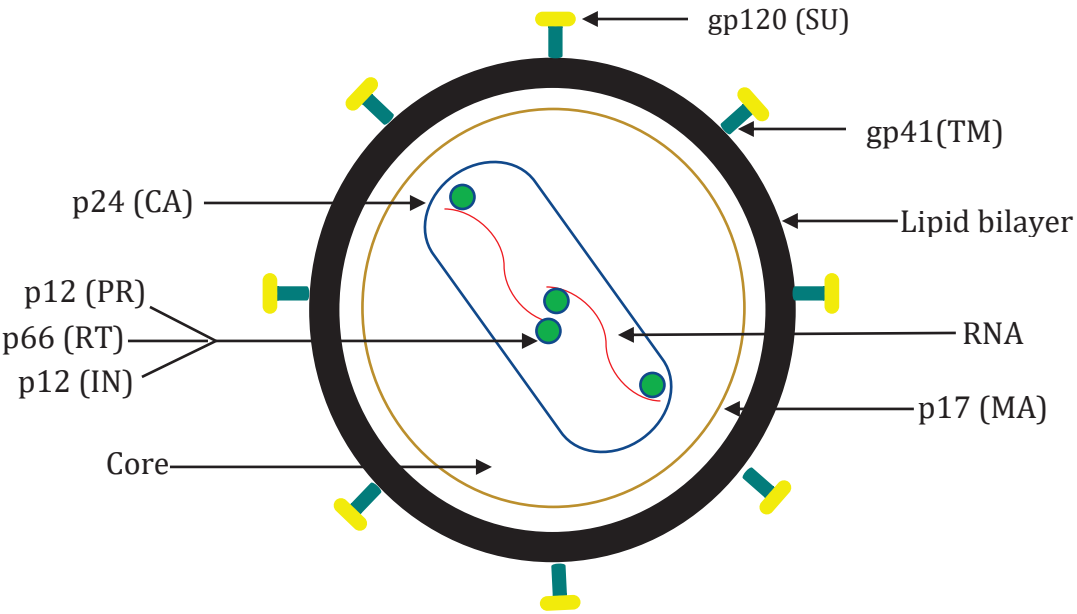


Figure 1.6: Structure of HIV-1 virion

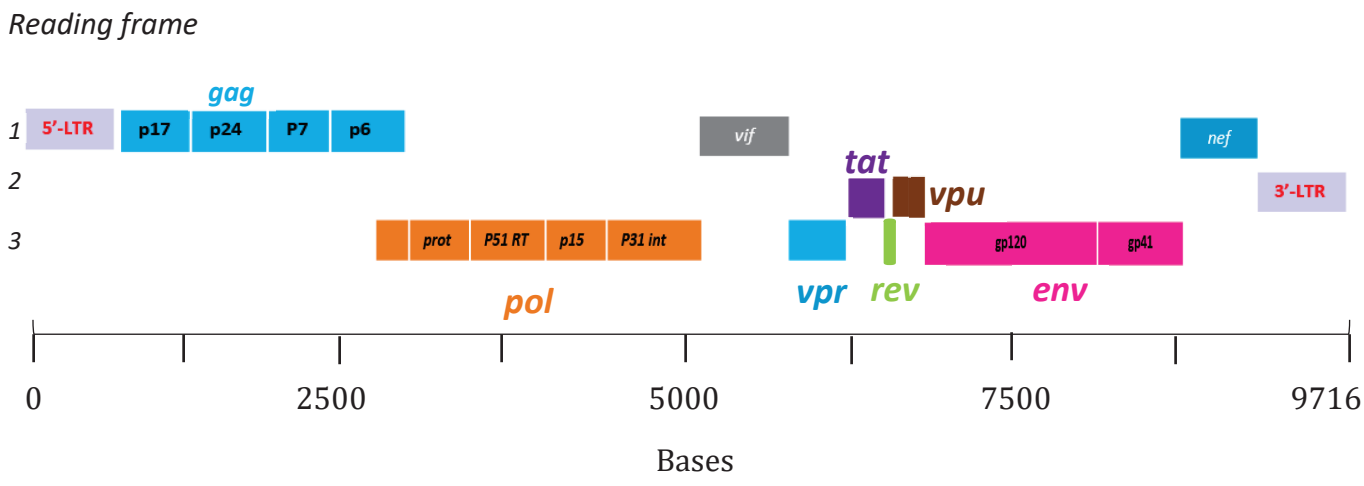


Figure 1.7: HIV-1 genome map

Figure 1.6: Graphical representation of the morphology of HIV-1 particle and **Figure 1.7:** HIV-1 genome map showing all three reading frames and nucleotide base positions.

1.4.2 HIV-1 replication cycle

Host cell activation, which involves the induction of cellular division is a crucial requirement for effective replication of HIV-1 (92). The HIV-1 replication cycle can be broadly divided into the following stages: a) cell binding and entry, b) viral uncoating, c) reverse transcription, d) viral DNA integration e) virus protein processing and assembly and f) virion budding (89, 92, 93). Cell binding and entry is facilitated through the affinity of viral gp120 protein for $CD4^+$ receptors found on a subset of T-lymphocytes and monocytes-macrophages. The binding between both molecules causes a conformational change allowing further interaction of gp120 with a co-receptor on target cells such as the chemokine receptor 4 (CXCR4) or chemokine receptor 5 (CCR5) resulting in the fusion of the viral envelope with the target cell membrane (92). Following successful entry, uncoating of viral core occurs releasing viral RNA into the host cell cytoplasm. This is immediately followed by the conversion of single stranded viral RNA into a double stranded DNA by the enzymatic action of RT. Through its ribonuclease H active site, RT degrades the viral RNA upon completion of the reverse transcription process. The entire process occurs within a pre-integration complex (PIC) that comprises of MA (p17) protein, Vpr and IN. This PIC is then trafficked into the cell nucleus where the viral double stranded DNA is integrated into the host genome through the enzymatic action of IN (93). IN creates sticky ends at the 3' ends of the host DNA allowing for the integration of the viral DNA into the host cell genome to form a pro-viral DNA. This process and the expression of the pro-viral DNA readily occurs when the host cell is in an activated state (93). This step is followed by the expression of the pro-viral DNA which encodes for the viral proteins and their pre-cursors. Expression of viral transcripts starts from the 5' long terminal repeat (LTR) region of the pro-viral genome with Tat enhancing the initiation of the transcription process. Following successful transcription, full length and spliced viral RNAs are transported from the nucleus to the cytoplasm for translation and packaging of proteins. This process is facilitated by the Rev

protein (89). Translation of viral messenger RNA (mRNA) occurs in the cytoplasm. Gag and Gag-Pol products are localized at the cell membrane, while translation of the *env* mRNA occurs in the endoplasmic reticulum (ER). Translation and post-translational processing of the mRNA result in the production of viral core proteins, MA, CA, NC, p6, PR, RT, IN and the accessory proteins such as Vif, Vpr, Nef, and genome RNA which leads to the gradual formation of immature virion and the initial step towards budding from the cell surface. The final stage of the process is the budding stage which also involves downregulation of surface $CD4^+$ receptors by the Vpu and Nef proteins; this is to make way for the provision of SU and TM for the outer membrane coat during viral budding. Viral maturation occurs by proteolytic cleavage of Gag and Gag-Pol polyproteins by PR enzyme. At this stage, the mature virus is ready to begin infection of another target cell, continuing the cycle in the absence of ART (89, 92, 93). As mentioned in an earlier section, every step of the HIV-1 replication cycle as well as every viral gene product is a potential target for therapeutic interventions. Following the discovery of HIV/AIDS, several strategies to inhibit viral replication cycle are put into consideration during drug development; which include RNA based strategies such as antisense RNA, RNA decoys, ribozymes and protein-based strategies such as monoclonal antibodies, chimeric proteins, intracellular single-chain antibodies (94) and ARTs which target viral enzymes and block viral entry (37). Continuous investigation of HIV-1 replication cycle is critical for determining potential new targets for the development of novel ARTs (92) as shown in Figure 1.8.

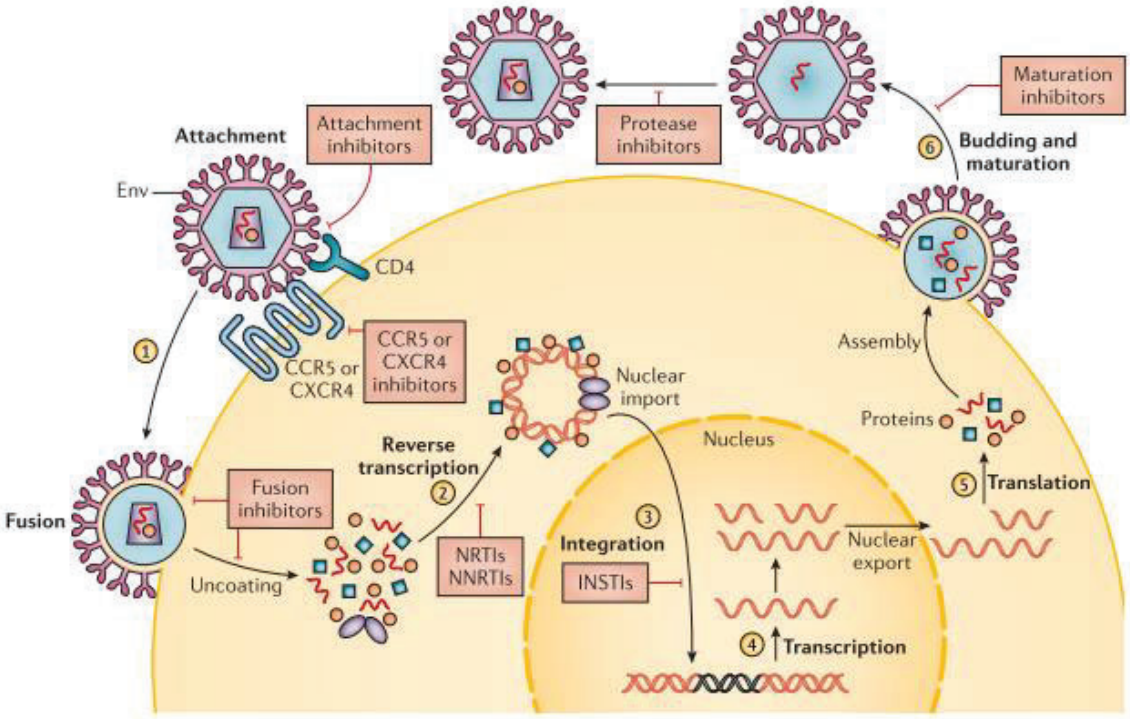


Figure 1.8: HIV-1 replication cycle with ART targets

Figure 1.8: HIV-1 replication cycle showing all known ART targets. Taken from (95).

1.4.3 Group specific antigen (Gag)

The group specific antigen (*gag*) is one of the 3 structural genes in the HIV-1 genome; it codes for a Gag polyprotein precursor, which when acted upon by PR leads to smaller products namely MA, CA, NC and p6. HIV-1 particle assembly is primarily coordinated by products of the *gag*, a machinery that recruits building blocks required for formation of a fully infectious viral particle, a full description of *gag* products, their functions and intra-cellular trafficking can be found in (96, 97). Its examination is of import to this thesis, and in subsequent sections I will elaborate on its role in viral replication.

1.5 Pathogenesis of HIV-1

Viral pathogenesis is the series of steps that occur when a viral isolate infects a susceptible host resulting in disease induction (98). Pathogenesis of HIV-1 infection and subsequent progression to AIDS are a consequence of the characteristics of the infecting viral isolate and the host immune response; the balance of effectiveness of both components determines the outcome of the infection (93).

1.5.1 Primary (acute) infection

Acute infection occurs when HIV-1 infects a susceptible host cell (98). $CD4^+$ and follicular dendritic cells (FDC) near the epithelial layer represent the initial target of HIV-1 infection following breakthrough into the mucosal barrier (93, 99, 100). Acute HIV-1 infection is generally characterized by three major events namely; an initial increase in plasma viral load, drastic depletion of naïve resting $CD4^+$ cells, and establishment of latent viral reservoir (101). The course of disease progression in an infected patient can be predicted within the first 6-12 months of infection. This finding is based on a 1996 Multicenter AIDS Cohort Study (MACS) data set, that showed rapid increase in plasma viral load following primary infection up to a

median of 10^6 -10^7 RNA copies/mL, and depletion of $CD4^+$ count can independently predict whether the patient will progress quickly or slowly to AIDS (101). Within the first 10-12 days post infection, viral RNA can be detected in the blood using RT-PCR amplification methods. Following peak plasma viremia levels, HIV-1 antibodies can be detected in the patient blood due to humoral immune responses to the infection. This point is referred to as seroconversion and it is about 20-25 days on average post primary infection (93, 99). During this time the patient experiences flu-like symptoms such as fever, lymphadenopathy, mononucleosis, weight loss etc. which takes place for about 7-10 days post infection (93, 99).

1.5.2 Chronic infection and AIDS

Following the resolution of acute infection and the establishment of a virologic quasi-steady state, a prolonged period of asymptomatic chronic infection ensues (93, 101); though patients appear to be asymptomatic, ongoing viral replication and $CD4^+$ depletion continues. During this phase, HIV-specific CTLs control viremia levels by recognizing viral antigens presented on the cell surface of infected cells (93). In addition, antibody dependent cellular cytotoxicity (ADCC) of infected cells is mediated by the natural killer (NK) cells. However, in the absence of therapeutic interventions, CD4+ T lymphocytes decline gradually and at a certain threshold (< 200 cells/μL), patients begin to suffer from certain symptoms, which constitute AIDS, such as prolonged fatigue, lymphadenopathy, oral and vaginal candidiasis, dermatological conditions, neurological conditions, atrophy, tuberculosis, and *Pneumocystis carinii* infection (93, 101).

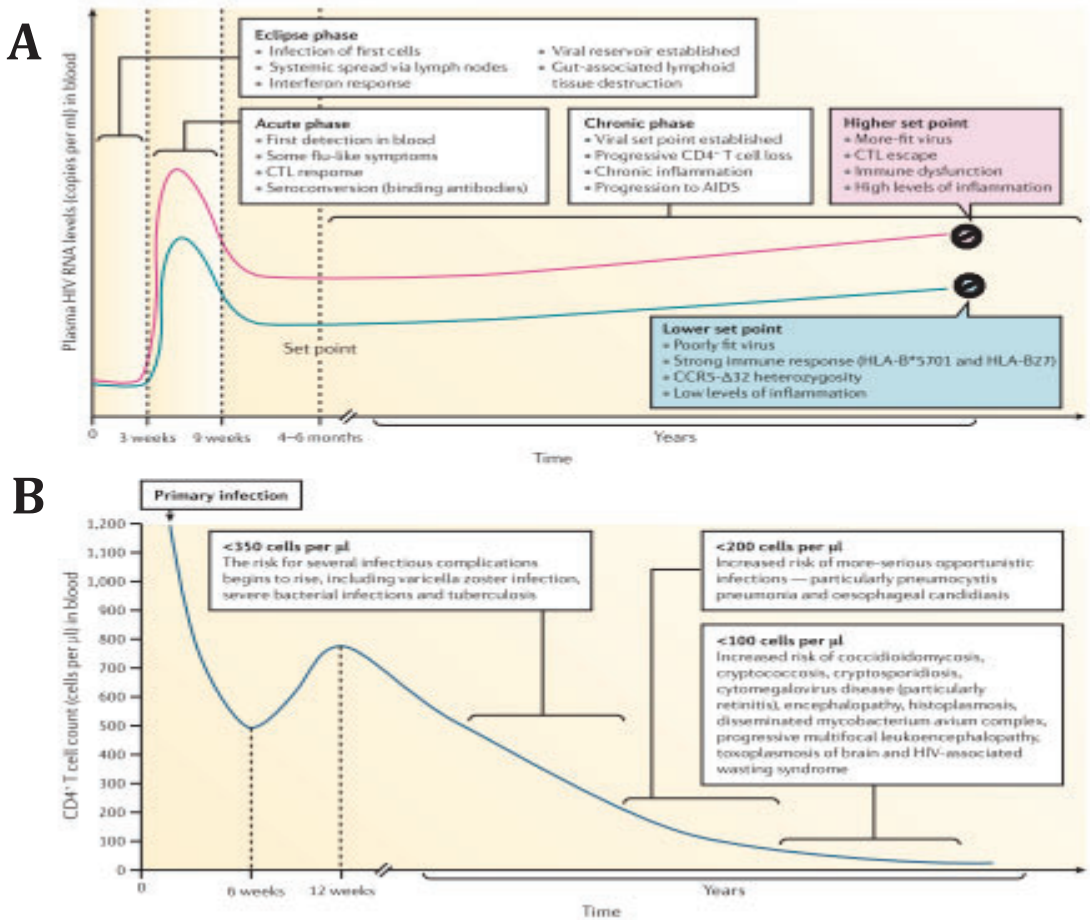


Figure 1.9: Phases of natural HIV-1 infection

Figure 1.9: (A) Showing phases of HIV-1 natural infection (Acute and Chronic), and (B) showing rate of CD4+ decline. Taken from (95).

1.6 Host immunity

The human immune system comprises of a complex array of protective mechanisms including cells, chemicals, and processes to overcome unfortunate exposure to pathogens and their products that have gained entry through mucosal and epithelial surfaces of the body. An important feature of a functional immune system is its ability to properly discriminate between host cells (self) and antigens, allergens, or toxins of pathogenic origin (non-self) to avoid causing damage to physiological functions of the body (102). The immune system comprises of the innate and adaptive mechanisms which work in complement to one another to mount an appropriate immune response to pathogenic exposure (103-105).

1.6.1 Innate immunity

The innate immune system is organized into a network of cells (macrophages, dendritic cells, natural killer cells, neutrophils, mast cells, eosinophils, and basophils) and signals that act as a first line of defence against pathogens. The onset of HIV-1 acute infection occurs following breaching of the host epithelial and mucosal barriers - the host innate immune response is triggered by the presence of pathogen associated molecular patterns (PAMPs) which are recognized by pathogen recognition receptors (PRRs) found on the surface of cells of the innate immune system (106, 107). PRRs can be broadly divided into toll-like receptors (TLRs) and intracellular endosomes (108, 109). Dendritic cells are large cells encountered by HIV at mucosal surfaces. They transport the virus from the site of entry to the lymphoid tissue where follicular dendritic cells (FDCs) provide signals for activation of B-cells to attack the virus (109). Natural killer (NK) cells are also important for control of HIV replication (110). When activated they release cytokines such as interferon-γ (IFN-γ), tumour necrosis factor-α (TNF-α) and chemokines that facilitate T-cell proliferation and have been reported to show improved function in the presence of KIR3DS1 HLA subtypes (110, 111).

Unfortunately, as the infection progresses it damages components of the innate immune system, especially macrophages which are linked to the adaptive immune system by inhibiting their antigen presenting function (112). As the first line of defence, the innate immune system is critical for sensing the presence of HIV particles and in turn setting in motion a series of immunological events to activate the adaptive immune response (108, 113).

1.6.2 Adaptive immunity

The adaptive immune system has evolved to provide a broad and fine-tuned collection of component assets for the recognition of both self and non-self thereby facilitating pathogen specific immune responses that result in the generation of long-term immune memory and regulation of immune homeostasis. The evolution of the adaptive immune system is driven by the vast variable antigenic structures of pathogens and their tendency to evade immune detection. Following initial pathogen encounters, cells expressing specific immune receptors for pathogen recognition persist in the host for life providing immunological memory and capacity for rapid response in the event of re-infection. The major components of adaptive immunity are the T-lymphocytes which are developed in the thymus and the antibody producing B-cells produced in the bone marrow (114). The adaptive immune system depends on customized receptors selected through somatic recombination events of specific germ-line segments which arose by gene duplication to form intact T cell receptor (TCR) and immunoglobulin (B cell antigen receptor; Ig) genes early in the evolution of vertebrates for the purpose of generating a highly flexible and specific immune response. The germline encoding genes elements facilitates the development of millions of different antigen receptors conferring each with the potential to be uniquely specific for the vast amount of antigen it may encounter during an infection (103).

1.6.2.1 T-cell immunity and HIV infection

T-lymphocytes mature in the thymus following the migration of hematopoietic stem cells originating in bone marrow (104). They express a unique series of antigen-binding receptors on their cell membranes commonly referred to as T-cell receptors (TCR). Each T-cell expresses a single form of TCR and has the capacity to rapidly proliferate and differentiate based on reception of appropriate induction signals. Following HIV-1 infection, host experiences persistent immune activation, followed by decline in CD4+ cells and subsequent disease progression to AIDS (115) and have been shown to correlate directly with viral load and inversely with $CD4^+$ count (116, 117). The major histocompatibility complex (MHC) class I on the cell surfaces present HIV-1 epitopes processed within the cells for recognition by TCRs on $CD8^+$ T cells. $CD8^+$ T cells then lyse HIV-1 infected cells and secrete cytokines such as IFN-γ, TNF-α, and chemokines such as macrophage inflammatory protein MIP-1 α and MIP β, that inhibit virus replication and block viral entry into $CD4^+$ T cells. $CD8^+$ T cells have been shown to be critical for control of HIV-1 replication (118-121). During acute infection, $CD4^+$ T cells lose their ability to expand rapidly and therefore their contribution to viral control is minor, however, during chronic infection this is reversed with $CD4^+$ T cells present and secreting interleukin-2 (IL-2) or cytokines, such as IFN-γ, to control viremia (117, 122). CTL escape is the variation in viral sequence that results in the loss of recognition by CTLs which is critical to the HIV pathogenesis (119). However, this in-turn impacts viral function positively or negatively on viral function and subsequent disease progression; and though it appears to be a mechanism of adaptation to its host, the range of variation of CTL epitopes is limited based on its impact on structure and function of the gene products (118), which further illustrates the effectiveness of CTL function on viral control albeit short-lived (123).

1.6.2.2 B-cell immunity and HIV infection

The adaptive humoral immune system is driven by antibody production from B cells under the signalling influence of T cells and dendritic cells. B cells themselves, like T cells, they also originate from the hematopoietic stem cells in the bone marrow (114). The humoral response occurs later in HIV-1 infection due to the time it takes to generate appropriate antibodies. Non-neutralizing antibodies to structural proteins such as p17 and p24 appear first but do not last. However, neutralizing antibodies specific to viral proteins involved in the entry of the virus into the cells, are subsequently produced (124). These antibodies are specific to the variable region of gp120 (V3), $CD4^+$ binding sites and chemokine receptors (CXCR4 and CCR5), and the transmembrane protein gp41 (102).

1.7 HIV-1 disease progression

A complex interplay between the subtype of the infecting virus and host immunogenetics ultimately determine the course of disease progression in HIV-1 infected patients (125). Subtype-specific variation in disease progression has been widely reported among HIV-1 infected individuals (126), with varying biological characteristics of different subtypes influencing rate or ease of transmission, pathogenesis, drug discovery and vaccine design (127) making it critical to understand the impact of these marked differences.

1.7.1 Subtype specific differences in disease progression

Following the discovery of HIV-1 and HIV-2, it was quickly observed that while both viruses descended from a common ancestor, however, a significant variance in their individual virulence was recognized (126). HIV-1, which is responsible for the global pandemic is more virulent and easily transmitted. This raised the question whether there could be differences in the biological characteristics of different subtypes and if this impacted disease progression. The best way to evaluate this question was the use of long-term natural history cohorts in the

absence of ART (126, 128). A prospective study conducted by Kanki and colleagues investigated disease progression in registered female sex workers in Dakar, Senegal. The results of their study showed that sex workers infected with HIV-1 non-A subtypes (which was mostly subtype D in the cohort) progressed to AIDS or death 8 times faster than patients infected with subtype A (126). Subsequent prospective studies conducted in Africa by other groups showed similar findings; for example, a 4-year prospective study by of 1045 infected adults conducted in Uganda by Kaleebu et al, showed that adults infected with subtype D had a faster progression of disease with a relative risk of 1.29 and 95% CI and a significantly lower CD4 cell count compared to those infected with subtype A (127). Another 2-year study of pregnant women by Vasal and colleagues (2006) conducted in Tanzania showed that relative to subtype A, patients with subtype D experienced the most rapid progression to death or to the WHO stage 4 of illness (129). A 12-year study by Baeten and colleagues of 145 infected adults in Kenya revealed a higher mortality ratio in women infected with subtype D over those infected with subtype A despite having no significant difference in their viral load counts (130). Other studies such conducted by Kiwanuka et al (2008), and Ssemwanga et al. (2013), showed similar results with patients infected with subtype D having a faster progression to disease than those infected with subtype A in Uganda (131, 132), while a multi-site study conducted in Uganda and Zimbabwe by Venner et al (2016) showed similar trends in tier results, however this time with patients infected with subtype C having the least disease progression relative to A and then to subtype D (128). All these studies therefore underlie the clear observation that there are subtype-specific differences in disease progression in HIV-1 infection.

1.7.2 HIV-1 subtype-specific differences in transmission rates

It has been particularly challenging to establish an association between HIV-1 subtypes and their respective risk of transmission (66, 133), however, several attempts have been made to demonstrate that such association exists. For example, a matched case-control study by Blackard and colleagues in 2001 showed that subtype A and inter-subtype LTRs were 3.2 and 4.8 times more likely to be transmitted from mother-to-child than subtype D perinatally (134); another such study by Odaibo et al., 2006 highlighted a marked variance in MTCT showing a 36.4%, 66.7% and 100% transmission in subtypes CRF02_AG, G and B in naive infected pregnant Nigerian women. Though the statistical power of the study was low, it did give an indication of the differences in the rates of MTCT based on infecting subtype (135). On the other hand, several risky behaviours such as intravenous drug usage, as well as hetero or homo -sexual relationships lead to the sustenance of HIV-1 transmission, which results in the segregation of HIV-1 subtypes into different groups within the population (136). For example, subtype B has a high transmission rate among MSMs and IDUs in Chile, Thailand, South Africa, Argentina, Europe, and most western countries, while non-B subtypes are more commonly transmitted amongst heterosexuals in those same counties (136, 137). A retrospective-observational study that controlled for confounders such as age, viral load, stage of disease, genital ulcers etc., showed that subtype A had a higher transmission potential than subtype AD and subtype D in Rakai, Uganda (138). Taken together, these studies highlight the significant effect of HIV-1 subtype diversity on transmission efficiency.

1.7.3 HIV-1 subtype-specific functional differences

HIV-1 subtype B is the most well characterized subtype and serves as the benchmark for development of therapeutic interventions, however the prospective studies conducted in Africa have shown that there are differences in biological characteristics of HIV-1 subtypes. With 70% of HIV-1 infected individuals living in sub-Saharan Africa where non-B subtypes are most prevalent, studying disease progression in non-B subtypes is necessary in order to develop robust interventions that can cater for all subtypes (58). HIV-1 subtypes vary in all protein regions, for instance by an average of 40% and 20% within the *env* and *gag* region respectively, which bears a marked impact on their corresponding protein structure and function (139). For example, the LTR, which encodes the transcriptional promoter, has a unique configuration for each subtype thereby affecting rates of replication in pairwise competitive experiments (139-141). Similarly such functional variation had been observed in the Envelope (142) and Pol (143) proteins. We have also previously demonstrated the impact of subtype-specific sequence variation in Gag on replication capacity, where a subtype-specific hierarchy in viral replication capacity was observed using patient-derived sequences from East Africa (144). In addition, our group observed variation in Nef's ability to down regulate CD4+ and HLA class I allele expression based on subtype (145). These studies underscore the impact of HIV-1 genetic diversity on viral protein structure and function, it is therefore imperative to monitor their distribution in sub-Saharan Africa to see if these differences impact the epidemic variance observed in each sub-region and to account for these differences when developing therapeutic interventions to slow down or eradicate viral transmission.

1.8 Viral replication capacity

Viral fitness is simply the sum of all characteristics (mostly genetic) that contribute to the successful transmission of its genetic material from one generation to another under specific conditions such as host immune response, drug treatment (24, 25, 119), and thus accounts for inter subtype variation in disease progression. The inherent characteristics of HIV-1 to generate mutation-prone viral particles that exist as a population of genetic variants or quasi-species within a single infection is an evolutionary mechanism that leads to the positive selection of the fittest variants (22, 26, 119) that go on to establish infection in the host (24). Group M viruses have been shown to be more fit within the human population than their close genetic relatives which are the N, O, P and HIV-2 (146). However, the complex interplay between the viral and host-immune factors makes fitness particularly difficult to measure in vivo, hence viral replication capacity is used in its place within clinical and research settings as an indicator of the viral fitness (147). Viral replicative capacity (VRC) is an absolute measure of the ability of HIV-1 to replicate under standard conditions. This can be measured using several ex-vivo and in-vitro assays which are well described in the literature (24, 25, 148, 149). VRC impacts viral load, drug resistance, rate of transmission and disease progression within the host (143, 144, 150-154)

1.8.1 Gag function and viral replication capacity

Gag sequences vary between HIV-1 subtypes by an average of about 14% (63, 79), revealing the highly conserved nature of the region and the important role it plays for effective viral replication. Several studies have shown how the complex interaction between viral and host-immune factors ultimately determine disease progression in infected patients (154-157), with host human leukocyte antigen (HLA) composition as an individual factor correlating most strongly with disease progression rate (158-160). While this is logical *in-vivo*, it does little to delineate and identify viral factors that may enhance disease progression in patients who lack

protective HLAs (154). Gag, being highly immunogenic (155) and conserved, is a prime target of HLA-restricted CTL responses leading to selection of viral variants with escape mutations that may compromise the replicative capacity of the virus and thus disease progression (154), especially in cases of positively linked HLA-associated mutations such in patients infected with HLA B*57 and B*27 which have been strongly associated with HIV-1 controllers (115). Previous work from our group has shown that Gag function has a significant impact on viral replication capacity *in vitro*, and strong association with markers of disease progression such as viral load and $CD4^+$ count (156), suggesting that it is a suitable candidate for vaccine design strategies. Gag-driven variation in replication capacity within different subtypes has also been documented (144).

1.8.1.1 Impact of Gag escape mutations on vaccine design

The anti-viral function of CTLs is well documented. However, CTL effectiveness is limited by escape mutations within the dominant epitopes of HIV-1 (161-166), which are driven by the net balance between the immune selection pressure and fitness cost to the virus making it challenging to be studied *in vivo* (162). While there is variation in the timing of occurrence of escape mutations, be it at the beginning of the infection or later stages (166, 167), there is evidence that they are transmissible. For example, a study on perinatal transmission showed that escape mutations in the KK10 epitope restricted by HLA B*27 were passed on from infected mothers to their respective infants upon delivery (161). However, due to the conserved nature of *gag* and the prominent role of Gag protein products in viral replicative function, escape mutations within particular regions of *gag* are limited by structural considerations impacting its protein products (162). *Gag* escape mutations with significant costs to viral replication ability (as measured *in vitro*) were shown to be strongly associated with reduced viral loads when transmitted to recipients, suggesting that CTL responses focussed on

vulnerable regions of Gag where escape is fitness costs could be of benefit to both the donor and recipient (168). A study by Chopera *et al.* 2017 showed that high frequency of escape mutations within different regions of *gag* had differential impact on disease progression with escape mutations within the variable regions of p17 p6p7 region leading to faster disease progression (167). Therefore, an in depth understanding of the replicative fitness and disease progression consequences of immune escape mutations bear significant implications for CTL based vaccine design (163, 164). It would be worthwhile to develop a prophylactic or therapeutic intervention that would target the specific *gag* regions where the generation of escape mutations would come at a significantly high fitness cost, weakening the capacity of the virus to effectively propagate itself (163, 168). HLA class one driven CTL escape mutations can be exploited for HIV-1 vaccine design capable of stimulating CTL responses against the highly conservative regions within the HIV-1 genome such as gag thereby forcing viral evolution towards a less-fit state (169, 170).

1.9 Study rationale

The adult prevalence rate of HIV-1 infection and its genetic diversity vary widely across sub-Saharan Africa where an estimated 70% of global population of people living with HIV infection reside. This diversity bears significant impact on accurate diagnosis, rate of disease progression, transmission, epidemic spread, as well as the design and development of therapeutic interventions for controlling the epidemic. While several prospective studies have shown marked differences in subtype-specific rate of disease progression and transmissibility, there is very limited understanding of the underlying biological mechanisms which may account in part for the reported subtype-specific differences in disease progression or explain the uneven distribution and spread of subtypes in sub-Saharan Africa. The Gag protein is highly immunogenic and is a candidate target in HIV-1 vaccine design. Previous work from our group

has shown that Gag-protease is an important genetic determinant of viral replication capacity *in vitro*, and that differences in Gag protease-driven replication capacity of circulating subtypes in South Africa and Kenya correspond with differences in disease progression rate (144, 154). Furthermore, studies link Gag-protease-driven replication capacity to transmission. Therefore, in this study, we sought to better examine the diversity of circulating HIV-1 strains in East and West Africa, two regions of sub-Sharan African with distinct patterns of high versus low prevalence of HIV and to investigate whether there are differences in *Gag-protease* mediated replication capacity that may explain the reported differences in prevalence and subtype-specific rates of disease progression. Moreover, we wanted to identify some of the viral genetic and host determinants of differences in virus replication capacity, considering that this information may have important implications for biomedical prevention and treatment strategies against HIV.

1.9.1 Hypothesis

I therefore plan to test the following hypotheses

a) Gag-protease driven viral replicative capacity differs among HIV-1 subtypes.
b) Viral replicative capacity is associated with indicators of disease progression such as viral load and $CD4^+$ count.
c) Host HLA Class I alleles expressed is associated with viral replication capacity in subtype A from East Africa.
d) That these subtype specific differences could account for or contribute to the epidemiological picture of HIV-1 infection in West and East Africa.

1.9.2 Project aims and objectives

Specific aims:

1. Determine if there are differences in Gag-protease-driven viral replication capacity according to HIV-1 subtypes.

2. Associate viral replication capacities with indicators of disease progression namely viral load, CD4 cell counts and the rate of CD4 cell decline.

3. Determine if there are subtype-specific HLA class I associations with viral replication capacities in the study population.

4. Identify amino acid sequences associated with decreased or increased viral replication capacity within each subtype using all available patients.

To achieve these aims the following specific objectives were fulfilled:

- ✓ Amplification and sequencing of *gag-protease* of patient samples from West (n=213) an East (n=160) Africa.
- ✓ Subtyping of patient derived sequences through online bioinformatic tools.
- ✓ Recombination analysis of patient derived sequences through online bioinformatic tools.
- ✓ Construction of recombinant viruses encoding *gag-protease* sequences derived from various subtypes identified from chronically infected patients West (n=178) and East (n=114) Africa.
- ✓ Analysis of *in vitro* replication capacity for *gag-protease* chimeric viruses using a flow cytometry-based GFP-reporter T cell assay.
- ✓ Statistical comparative analysis of *in-vitro* viral replication capacities of chimeric viruses from West and East Africa.

- ✓ Association of *gag-protease* replication capacity values to HLA class I profiles of subtype A patients from East Africa.
- ✓ Qualitative and quantitative analysis to determine genetic determinants of viral replication capacity in multiple subtypes.

.

2 MATERIALS AND METHODS

2.1 Ethical Approval

Ethical clearance was received from The Biomedical Research Ethics Committee (BREC) at the University of KwaZulu-Natal for this study (**BE313/16).** Stored plasma was used as study material and written informed consent was provided by all study participants at the time of recruitment. All investigations were undertaken in accordance with the principles in the Declaration of Helsinki.

2.2 Study samples

This was a retrospective cross-sectional study, where ART-naive plasma samples collected between the years 2000 and 2010 from chronically infected individuals in West and East Africa were analysed. In West Africa, study participants included ineligible blood donors, female commercial sex workers and members of the general population from the Cameroon (n=169) (171), Nigeria (n=31) (172) and Senegal (n=96) (173). In East Africa, study participants were from the well characterized IAVI Protocol-C cohort which included discordant couples, heterosexuals, and MSMs from the general population in Kenya (n=73), Rwanda (n=61) and Uganda (n=107). However, samples from East Africa were preselected for subtypes A, AD and D based on previously sequenced pol genes (174). Viral load and $CD4^+$ T cell count data was available for all participants from East Africa. HLA class I data for participants infected with subtype A1 from East Africa was also available for analysis. Clinical and demographic characteristics for both samples from West and East African groups are summarized in Table 3.1.

2.3 Plasma RNA extraction

Plasma samples were obtained from storage in -80°C freezer and thawed at room temperature. HIV-1 RNA was extracted from individual patient plasma sample using the QIAamp® Viral RNA Mini extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's prescribed protocol. Viruses from plasma samples with lower viral loads (<5,000 copies/mL) were concentrated by spinning 500µL of plasma sample at 14,000 rpm (rotor radius of 100mm) for 2 hours at 4^{o}C. Approximately 350µL of supernatant was removed and discarded. The pellet was then re-suspended in the remaining volume of plasma and used for RNA extraction.

2.3.1 Amplification and viral sequencing of *Gag-protease*

Extracted viral RNA from section 2.3 was converted into cDNA and then double-stranded DNA by a one-step RT-PCR using the Superscript III One-Step RT-PCR, Platinum Taq High Fidelity kit (Invitrogen, San Diego, USA). Each reaction was made up of 14.4µL UltraPure^{TM} Diethyl pyrocarbonate (DEPC) treated water (Invitrogen), 20µL of 2X reaction mix, 0.8 µL of each of the10µM forward Gag+1 (5' GAG GAG ATC TCT CGA CGC AGG AC 3'; **675-697**) and reverse 3'RVP (5' GGA GTG TTA TAT GGA TTT TCA GGC CCA ATT 3'; **2725-2696**) primers, 0.8 µL of Superscript III RT/Platinum Taq High Fidelity Enzyme Mix and 4µL of extracted RNA. The resulting reaction mixture was incubated in a thermocycler (Applied Biosystems, Foster City, USA) under the following cycling conditions; 55^{o} C for 30 minutes and 94^{o} C for 2 minutes followed by 35 cycles of 94^{o} C for 15 seconds, 55^{o} C for 30 seconds and 68^{o} C for 2 minutes. The reaction ended after a final extension at 68^{o} C for 5 minutes. A second round of PCR was performed to amplify the 1760bp *gag-protease*, HXB2 **790-2550** (175) using the Takara Ex Taq® Hot Start kit (Takara, Shiga, Japan). The 50 µL PCR reaction mixture included 37µL nuclease free water, 5 µL of 10Ex Taq Buffer, 4 µL 2.5mM dNTP, 0.8 µL each of 10µM forward (GAC TCG GCT TGC TGA AGC GCG CAC GGC AAG AGG CGA GGG GCG GCG ACT GGT GAG TAC GCC AAA AAT TTT GAC TAG CGG AGG

CTA GAA GGA GAG AGA TGG G; **695-794**) and reverse (GGC CCA ATT TTT GAA ATT TTT CCT TCC TTT TCC ATT TCT GTA CAA ATT TCT ACT AAT GCT TTT ATT TTT TCT TCT GTC AAT GGC CAT TGT TTA ACT TTT G; **2704-2605**) primers that are the NL4-3 sequence immediately flanking the *gag-protease* to facilitate homologous recombination, 0.25 µL of ExTaq enzyme and 2 µL of the RT-PCR product. 5 µL of the 2° *Gag-protease* amplicons of each patient was mixed with 2 µL gel loading buffer (SigmaAldrich, Missouri, USA) containing Gel-red nucleic acid gel stain (Biotium, California, USA) at a ratio of 1:50, and were resolved on 1% agarose gel (prepared using agarose tablets, Bioline, Atlanta, USA). The loaded gel was run at 95 volts for 1 hour to verify products of approximately 1760 bp (Figure 3.1) using the 1kb high DNA mass ladder (Invitrogen™, life Technologies Carlsbad, USA). The gel was viewed on a Chemidoc™ imaging system using the Image Lab™ software version 5.0 (BioRad, CA, USA). On confirmation of the positive bands, the 2° amplicons were diluted in nuclease free water at a ratio of 1:15 and set aside for sequencing. The resulting solution was population sequenced using the ABI PRISM Big Dye Terminator Ready reaction mix V3 (Applied Biosystems, Waltham, USA) and run on an ABI 3130xl Genetic Analyzer (Applied Biosystems, Waltham). The sequencing reaction comprised 4 µL of nuclease free water, 2 µL of 5X sequencing buffer, 2.6 µL of individual primers at 2µM concentration, 0.4 µL of Big Dye Terminator mix and 1 µL template of diluted 2° PCR. These 10 µL reactions were amplified in a MicroAmp® optical 96 well reaction plate (Applied Biosystems, Foster City, USA) at the following cycling conditions; 96^{o} C for 1 minute, followed by 25 cycles of 96^{o} C for 10 seconds, 50^{o} C for 5 seconds and 60^{o} C for 4 minutes. The sequencing products were then cleaned by adding 1 µL ethylenediaminetetraacetic acid (EDTA; 125 mM Sigma-Aldrich, Johannesburg, South Africa) in each well followed by 26 µL of the volatile mixture of sodium acetate (NaOAc; 3 M Sigma-Aldrich) and absolute cold ethanol (100%) at a ratio of 1:25. This was followed by centrifugation (Eppendorf centrifuge

5810R, ThermoFisher scientific, Carlsbad, USA) of the 96 well plate at 3,000*g* for 20 minutes. Plates were then inverted on paper towel and centrifuged at 150*g* for 5 minutes to remove the liquid from the sequencing products. The pellets were immediately washed in cold 70% ethanol followed by centrifugation at 3,000*g* for 5 minutes. The liquid was removed from the plate again by inversion on paper towel followed by centrifugation at 150*g* for 1 minute. The products were dried at 50°C for 5 minutes in a thermocycler and stored at -20 °C (Samsung RZ28H6150SS, Seoul, South Korea) until further analysis on the ABI 3130xl Genetic Analyser (Applied Biosystems, Foster City, USA). Sequence chromatograms generated from genetic analysis were assembled and edited using Sequencher v5.1 (Gene Codes Corporation, California, USA). Nucleotide mixtures were defined as double peaks in the sequence chromatogram where the secondary peak was a minimum of 25% of the height of the dominant peak. All plasma sequences were aligned to the HIV-1 subtype B reference strain HXB2 (GenBank accession no. K03455) using Gene Cutter. A summary of viral extraction to sequence generation procedure is shown in Figure 2.1.

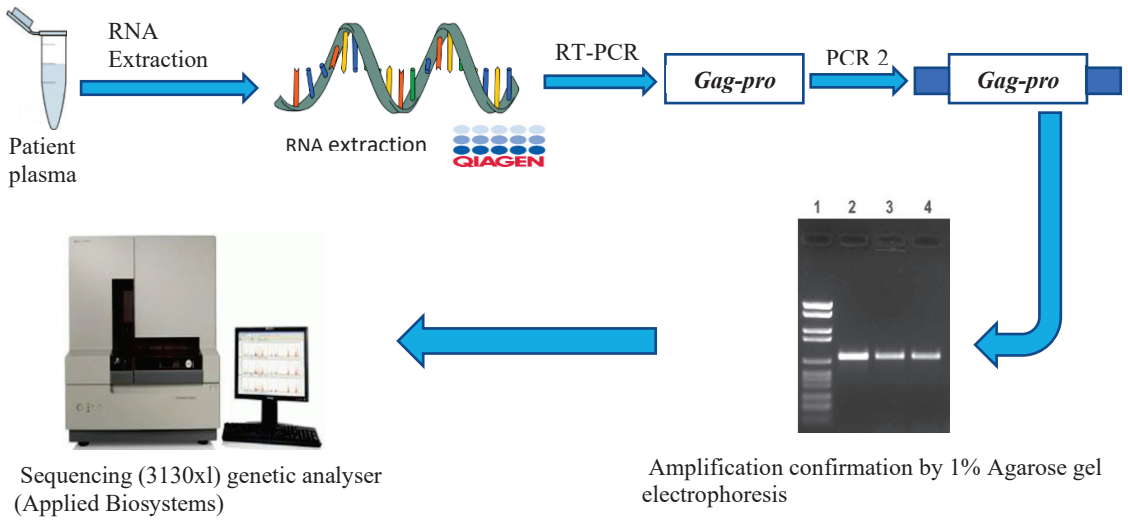


Figure 2.1: Amplification and sequencing summary

Figure 2.1: Summary of amplification and sequencing of patient plasma showing experimental procedures from RNA extraction from patient plasma all the way to sequencing of patient samples.

2.3.2 HIV-1 subtyping

Edited sequences were exported as a FASTA file, and were then uploaded to the REGA (http://www.bioafrica.net/subtypetool/html/subtypinghiv.html) HIV-1 subtyping tool v3.46 (176) and COMET (https://comet.lih.lu/) subtyping tool (177) for independent subtype classification. Sequence alignment and translation was performed using the Gene-Cutter tool from the Los Alamos National Laboratory (LANL) database (https://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html) for subsequent phylogenetic analyses.

2.3.3 Phylogenetic analysis

Maximum-likelihood phylogenetic trees were constructed from the multiple sequence alignments using PhyML (178) to determine evolutionary relationships between patient sequences and subtype reference strains obtained from LANL. Instances where the subtyping methods were not in agreement, the phylogenetic tree was used for the final subtype assignment. Figtree v1.4.4 was used to infer the branching topology.

2.3.4 HIV-1 subtype recombination analysis

Prediction of inter-subtype recombination breakpoints was performed using the Jumping Profile Hidden Markov Model (jpHMM) online tool (179) developed by the Institute of Microbiology and Genetics at the Department of Bioinformatics at the University of Gottingen in Germany in collaboration with the Los Alamos HIV-1 sequence database (http://jphmm.gobics.de/jphmm.html) and the Simplot (180) program. Default settings were used for jpHMM, while Simplot was optimized to a window size of 350, base pair step size of 30 and a stringent consensus value of 50%. Each identified region between breakpoints was extracted, trimmed, and realigned together with subtype reference sequences and phylogenetic

trees were constructed to verify the subtype for each sequence fragment between predicted breakpoints. A summary of methods for HIV-1 subtyping, and inter-subtype recombination prediction is shown in Figure 2.2.

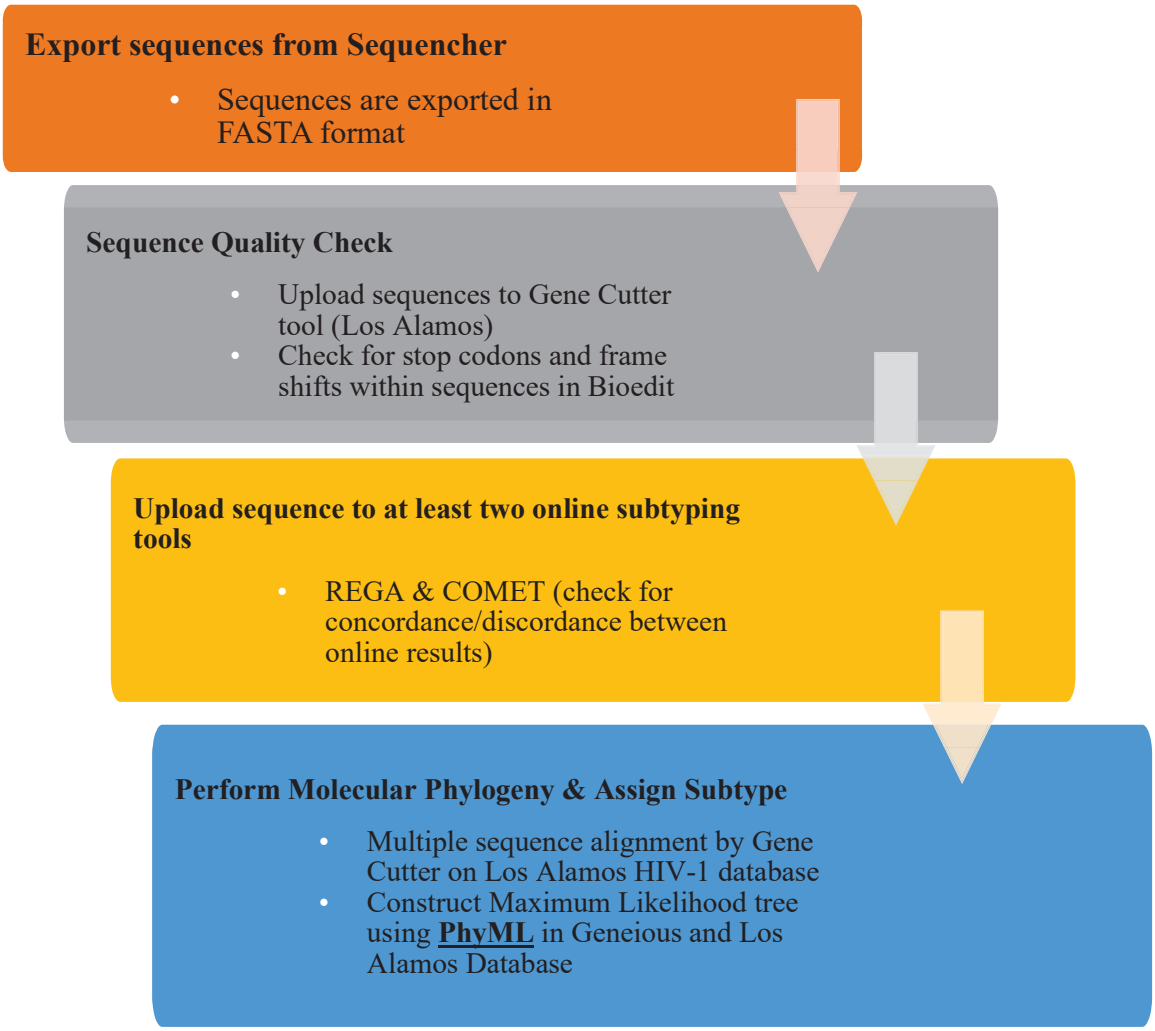


Figure 2.2: Summary of HIV-1 subtyping protocol

Figure 2.2: Summary of HIV-1 subtyping protocol: Flow chart summarising bioinformatic techniques used to analyse sequence data from sequence editing to subtyping, and phylogenetic analysis using PhyML to draw trees and infer phylogenetic relatedness of sequences.

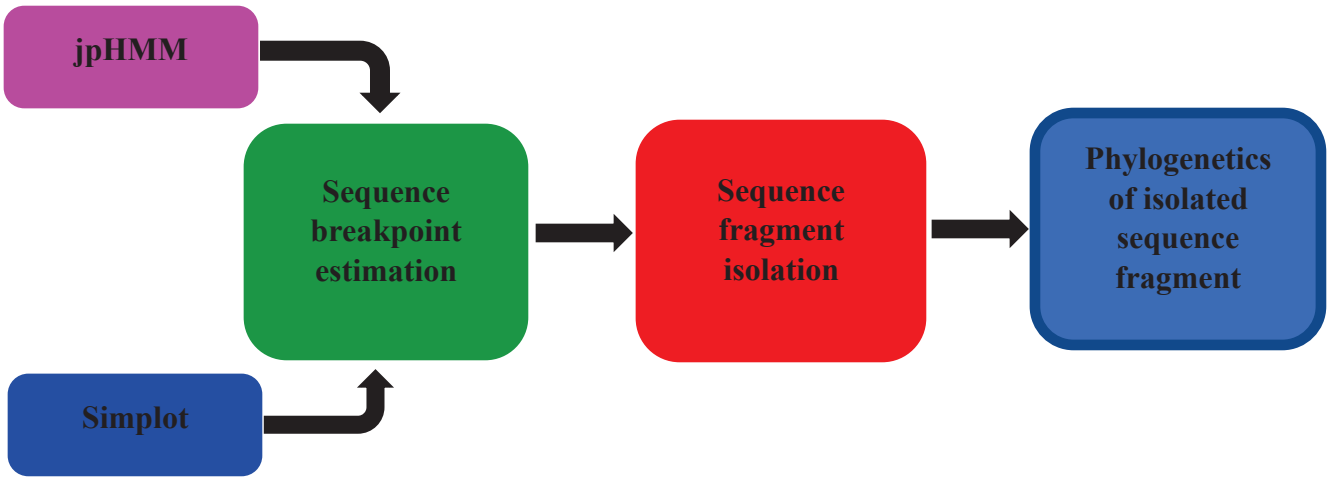


Figure 2.3: Inter-subtype recombination analysis

Figure 2.3: Summary of HIV-1 inter-subtype recombination analysis: Flow chart showing two bioinformatic tools employed to predict inter-subtype recombination points, extract sequence fragment, and confirm subtype by phylogeny.

2.4 Preparation of the pNL4-3Δ*gag-protease* backbone

The pNL4-3Δ*gag-protease* plasmid is available in the HIV Pathogenesis Programme laboratory and was prepared as originally described in (181). The plasmid contains a restriction enzyme BstEII site on either side of the deleted *gag-protease* coding region. pNL4-3Δ*gag-protease* plasmid stock was prepared by mixing a grown culture transformed with a glycerol-containing cocktail of 1.6 mL LuriaBertani (LB) media (Sigma, Johannesburg, SA), 400 µL glycerol (Sigma) and 2 µL ampicillin in a 1:1 ratio. Large concentrations of the plasmid were obtained by inoculating 100 mL of LB media containing 100 µL ampicillin with 25 µL of thawed glycerol sock and incubated at 37°C and 230 rpm for 16-17 hours in a shaking incubator (Infors HT, Switzerland). The culture was then cooled to 4°C for 30-60 minutes and pelleted by centrifugation (Eppendorf centrifuge 5810R ThermoFisher™ scientific, Carlsbad, USA) at 5850g for 20 minutes. The pellet was purified using plasmid maxi kit (Qiagen, Shiga, Japan) according to the manufacturer's instructions and quantified using a Nanodrop 2000 spectrophotometer (ThermoFisher™ Scientific). The purified plasmid was then frozen at -80°C until ready for use. The specificity of the purified plasmid was confirmed by Hind III

digest (ThermoFisher™ scientific). The 21 µL reaction included 15 µL nuclease free water, 2 µL Buffer 2, 1 µL Hind III enzyme and 3 µL purified plasmid. This was incubated for 1 hour at 37ºC. The digested products were resolved on a 2% agarose gel.

2.4.1 Maintenance of the cell line

Cell culture experiments were conducted using CEM-GXR cell lines donated by the Brockman group (182). They are T-cell lines that have been engineered to express the green, fluorescent reporter protein (GFP) to monitor HIV-1 infection levels within the cells; the cells are susceptible to both X4 and R5 viruses. The CEM-derived GXR GFP-reporter cells contain a Tat-dependent GFP expression system in which the GFP gene was inserted downstream of the TAR sequence in a retroviral vector. The cells were thawed from liquid nitrogen by placing them into a water bath at 37°C after which the cells are mixed with pre-warmed R10 RPMI medium with 10% foetal calf serum, 2 mM L-Glutamine, 10 nM HEPES and 50 U/ml penicillin –streptomycin (ThermoFisher™ scientific, Carlsbad, USA), and incubated in a humidified incubator at 37°C and 5% CO_2 overnight. The cells were then pelleted by spinning at 1,500 rpm for 10 minutes and resuspended in fresh R10 removing the Dimethyl sulfoxide (DMSO) (Sigma-Aldrich, South Africa) a cryo-protective agent used during storage. The stock culture was maintained in R10 medium up to for two months. Growth medium was replaced every two days by discarding 50-90% of culture medium and replacing it with fresh medium. The R10 medium contains a phenol red indicator, which changes colour from red to yellow signalling cell growth and consequently a need for medium replacement. The cells were used for experiments as required and maintained in culture for no more than 3 months before a new vial of GXR cells was thawed. CEM-GXRs were stored in liquid nitrogen (Antech Incorporated, Los Angeles, USA) within two weeks of maintenance of newly thawed GXR cells, to maintain consistency in all the culture experiments. Proper storage involved counting the cultured GXRs

using TC10™ automated cell counter (BioRad, Hercules, California, USA) with the BioRad dual chamber counting slides. Cells were mixed with 0.4% trypan blue dye (BioRad, Hercules, California, USA) in a 1:1 ratio and loaded onto the counting slides. The average reading was used as the cell count. The ideal count was 1 million cells/ml in 40 ml flask at >90% cell viability.

2.4.2 Co-transfection Experiments

Patient-derived viral particles were generated by co-transfecting secondary (2°) *Gag-protease* amplicons from each patient and pNL4-3Δgag-protease backbone into CEM–GXR cells. Firstly, the reaction per *gag-protease* amplicon included 10 µg plasmid, 10X buffer (a tenth in total reaction volume), 100X BSA (one hundredth in total reaction volume), 2U BstEII enzyme (Promega, Madison, USA) per µg plasmid (i.e. 20 U) and nuclease free water. This reaction was incubated for 2 hours at 60 °C in a water bath, facilitating the linearization of the plasmid at the site of the *gag-protease* gene allowing for insertion of amplified *gag-protease*. The HIV-1 *gag-protease* amplicons from the second round nested PCR were co-transfected with the pNL4-3Δ*gag-protease* by electroporation into CEM–GXR cells using the GenePulser Xcell™ electroporation system (BioRad, New York, USA) to generate recombinant viruses by homologous recombination as described by (183). A negative control CEM–GXR cells was electroporated with the pNL4-3Δ*gag-protease* only. The electroporated CEM–GXR cells was incubated for 5 days then R10 media is added. To allow time for recombination of *gag-protease* amplicons and pNL4-3Δ*gag-protease* plasmid, the electroporated CEM–GXR cells will then be incubated at 37°C and 5% CO_2. On day 3 and day 5 post-transfection, 3mL and 2mL R10 medium was added to the cell-culture flasks, respectively. Thereafter, the cultures were monitored every second day for infection by removing 2 mL of culture for flow cytometry analysis and replacing with 2 mL fresh R10 medium.

2.4.3 Harvesting the viral stocks

Homologous recombination of insert and vector produces infectious recombinant HIV-1; upon infection of new GXR cells in culture, GFP is expressed thereby allowing monitoring of infection by flow cytometry. On day 9 post-transfection, 2 mL of culture was removed and 1mL of this was added to individual cluster tubes (Corning® costor, Sigma, Saint Louis, USA) for centrifugation at 1500 rpm for 10 minutes. The supernatant was carefully discarded, and the cells re-suspended in 200 µL of 2% paraformaldehyde (PFA) fixative and incubated for 10 minutes prior to flow cytometry analysis. The fixative was prepared by dissolving 20g PFA (Merck, Durban, South Africa) in 500 mL phosphate buffered saline (PBS) (ThermoFisher™ scientific, Carlsbad, USA) at 60°C for 30 minutes on heat stir (Stuart CB162, Johannesburg, SA) using a magnetic stirrer. The pH of the prepared PFA was adjusted to 7.0 and stored at -20 °C in aliquots. The thawed PFA aliquot was stored at 4°C after each use and discarded within 7 days of thawing. The cluster tubes containing the fixed samples were placed in re-usable FACS tubes (BD Biosciences, New Jersey, USA) to enable easy access to the flow cell chamber of the FACSCalibur (BD Biosciences). Fluorescence information was captured for 15000 intact cells/events per sample using optimised instrument settings. The negative control (uninfected GXR cells) was included in each co-transfection experiment and was used to distinguish GFP positive cells from GFP negative cells as well as intact cells from debris. Gating for GFP (using the uninfected negative control sample) was set at 0.05, above which cells were considered GFP positive. The florescence data was analysed using FlowJo software (FloJo LLC, Ashland, Oregon). Upon reaching a threshold infection rate of about 25-30%, the culture was pelleted by centrifugation at 1700 rpm for 5 minutes at 4°C. The supernatant was collected, then aliquoted and stored at -80°C for subsequent cell culture experiments.

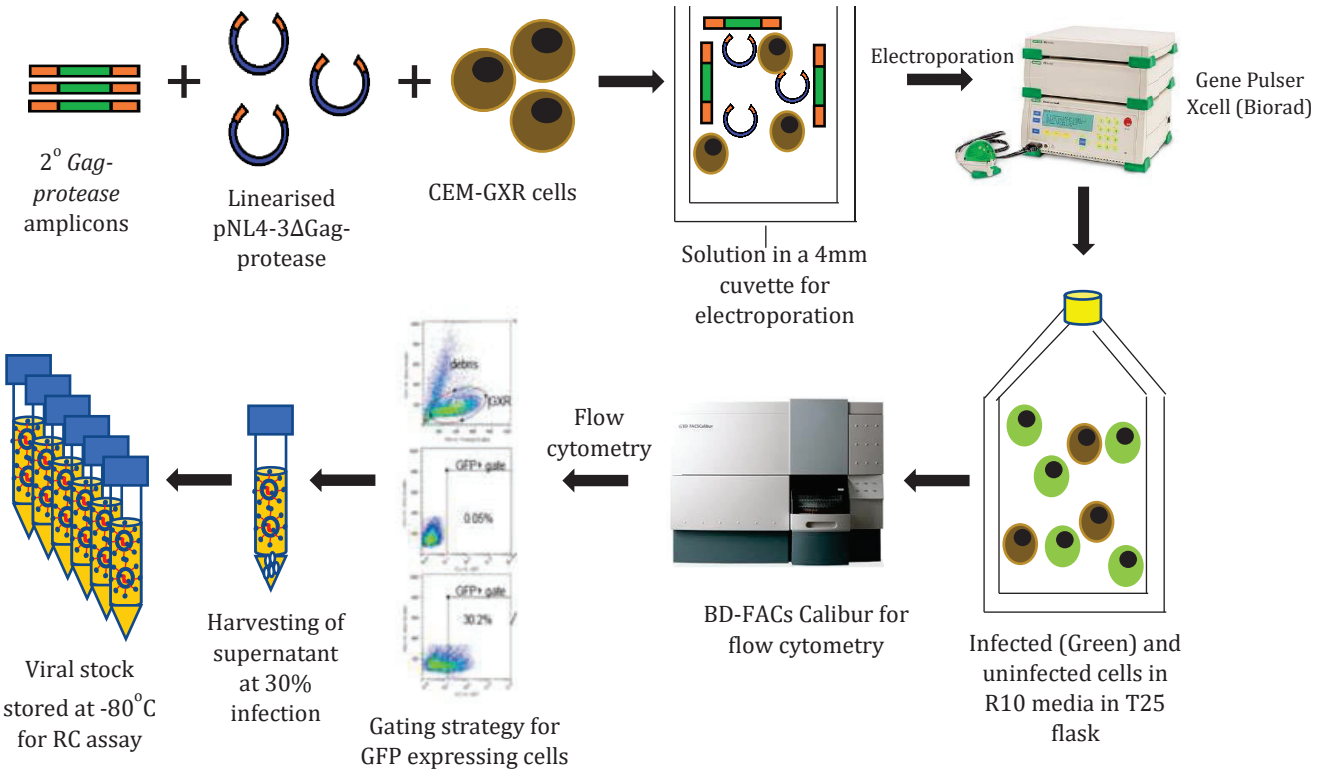


Figure 2.4: Summary of co-transfection and viral harvest protocol

2.4.4 Validation of generated recombinant pNL43 RT-integrase viruses

Validation of patient derived chimeric viruses was done as a quality control measure to ascertain that chimeric viruses had identical genetic composition as the original plasma viruses. RNA was extracted from 140µL of viral culture supernatant for 10% of randomly selected viral stocks using the QIAamp® Viral RNA Mini extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's prescribed protocol described in section 2.3. *Gag-protease* was amplified and sequenced as previously described (section 2.3.1).

2.4.5 Titration assay

Viral titre of stored supernatants was determined by infecting CEM–GXR cells with harvested virus to determine the volume of viral stock required for replication capacity experiments. 0.4mL of virus is incubated with 0.1mL containing 1×10^6 CEM–GXR cells to make a final volume of 0.5 mL and left overnight at 37°C in a 24 well flat bottom tissue culture plate with a low evaporation lid (Multiwell™ 24 well Falcon BD). Cells were then be resuspended in 1mL R10 and incubated for a further 24 hours. The percentage of infected cells was measured using fluorescence-activated cell sorter (FACS Calibur, BD Sciences, New Jersey, USA) analysis to determine green fluorescent protein (GFP) expression at 48 h post infection. The titre results were used to calculate the volume of virus stock to infect GXR cells at a multiplicity of infection (MOI) of 0.003 for the replication capacity assay. Values were then normalized to those for the wild type NL4-3.

Volume of virus stock = (0.3/proportion GFP-expressing cells) X 400 μL.

2.4.6 Replication capacity assay

The viral replication assay was performed in duplicate as previously described (154). Briefly, CEM-GXR cells were infected, using the same method as described in the titration assay, with the volume of virus stock calculated by the titration assay made up to a final volume of 0.4 mL using R10 medium. A negative (R10 medium) and positive (wild-type NL4-3 virus generated by electroporation of GXR cells with 10 μg NL4-3 wild-type plasmid) control was included in each assay. After a 24 hour-incubation (day one), 1 mL of R10 medium was added and the percentage of infected cells was monitored by flow cytometry each day for 5 days by pipetting out 500 μL from each well and replacing with 500 μL of fresh prewarmed R10 medium. This experiment was performed in duplicate independently for each sample. The replication capacity of the chimeric viruses determined by Gag-protease function was calculated by the

mean slope of exponential growth from 3-6 days post infection using the semi-log method in Microsoft Excel and normalised to the slope of growth of the wild-type NL4-3 control included in each assay to control for assay-to-assay variation. The 'LOGEST' function in Microsoft Excel calculates a $\log_{10}$ exponential curve that best fits the generated data and returns the slope of the curve. LOGEST values were converted to natural log for interpretation of exponential growth.

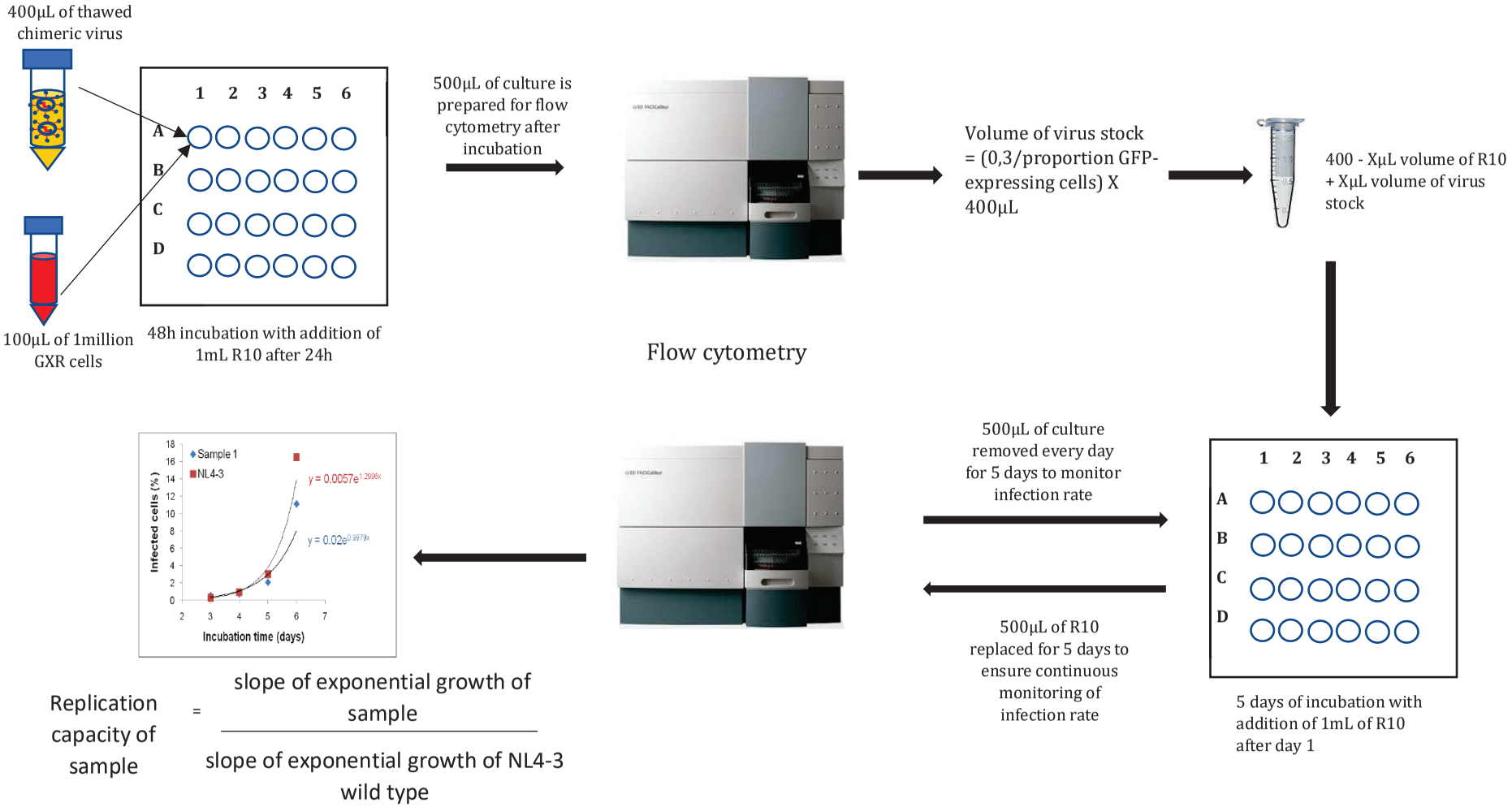


Figure 2.5: Summary of viral titration and replication capacity assay

2.5 Data analysis

Viral replication capacities of patient-derived chimeric viruses were compared based on HIV-1 subtype classification and geographical region using either the Students T test or Mann-Whitney U test if two groups were compared, or ANOVA with Tukey post-hoc tests where more than two groups were compared. ANOVA was used to test for differences in replication capacity across each group of HLA alleles (A, B and C) for participants infected with subtype A1 from East Africa, while the Students T test was used to compare differences in RC values between participants that expressed certain class I HLA alleles and those that did not express such alleles. HLA class I alleles expressed in a minimum of 5 individuals were included. These analyses were performed using Graphpad Prism v.8.4.3 (Graphpad Software, Sandiego, California, USA). A multivariable linear regression model performed using Stata 15 (StataCorp LLC, Texas, USA), was used to assess the relationship between subtype and VRC after adjusting for potential confounders. A variable was included in the final model if its inclusion resulted in a 10% or greater change in the coefficients of the subtype category variable. Multiple imputation using chained equations, was used to assess the sensitivity of results to missing data. Codon-by-codon Mann-Whitney U tests with q values (available at https://bblab-hivresearchtools.ca/django/tools/codon_by_codon/) were used to identify specific amino acid variants associated with increased or decreased *gag-protease*-driven VRC.

3 RESULTS

3.1 Results

HIV-1 sequence and VRC data generated from patient samples were correlated with markers of disease progression such as viral load and CD4+ count. A comparative analysis of subtype-specific differences between *gag-protease* driven VRC of isolates from West and East Africa with the view to see if it contributes to the variances observed between both sub-regions. Furthermore, a multivariable regression analyses was performed to see what factors were driving VRC among HIV-1 subtypes between the different regions, followed by an evaluation of the association between HLA allele expression and viral replication capacity on East African subtype A1, this could unfortunately not be done for CRF02_AG in East Africa due to lack of HLA data. Inter-subtype recombination within the p6 region of *gag* was prevalent and investigated to see if it bore any impact on low versus high VRC. An exploratory codon-by-codon analysis was performed to identify amino acid variants that could impact VRC in subtypes CRF02_AG and A1 isolates from West and East Africa, respectively.

3.2 Study Participants

Samples collected from HIV-infected participants from East and West Africa were analysed in this study. Of the 296 samples collected from West Africa, 213 (73%) were successfully amplified. 21% of samples from West Africa had extremely low viral load (<log 2 copies/mL), while 6% of samples were of low sample quality making them difficult to amplify. In East Africa, only 160 (66%) of the 241 samples from East Africa were successfully amplified. 20.6% of samples had extremely low viral loads (<log 2 copies/mL) while another 13% appeared to be of low sample quality. A total of 178 (84%) and 114 (71%) of amplified samples were successfully co-transfected and assayed for replication capacity for West and East Africa, respectively. A disproportionately large number of subtype A1 samples failed to replicate in

culture. Table 3.1 shows a summary of the study population demographic and clinical characteristics. Low sample quality is defined as samples having extremely low viral loads (≤Log 2copies/ml) which indicates degradation possibly due to poor storage.

Table 3.1: Demographic and clinical characteristics of participants

	West Africa	East Africa	*p*-value
No. of participants (% male)	213 (34%)	160 (64%)	<0.0001[a]
Age, years (IQR)[b]	32 (27 – 40)	28 (24 – 34)	0.0006[c]
CD4+ cells, cells/mm^3 (IQR)[b]	478 (305 – 619)[d]	487 (347 – 625)	0.274[c]
Viral load, log$_{10}$ copies/ml (IQR)[b]	4.86 (4.35 – 5.37)[e]	4.44 (3.95 – 4.93)	<0.0001[c]

[a]p-values calculated using Fisher's exact test
[b]medians with inter-quartile ranges shown in brackets
[c]p-values calculated using the Mann-Whitney test
[d]only 66% of participants had CD4+ T cell count data
[e]only 49% of participants had viral load data

3.3 Sequence amplification

Secondary PCR products of *gag-protease* amplification was confirmed using 1% Agarose gel electrophoresis prior to commencing the sequence procedure described in section 2.3.1. A representative gel of all samples successfully amplified is shown in figure 3.1 below.

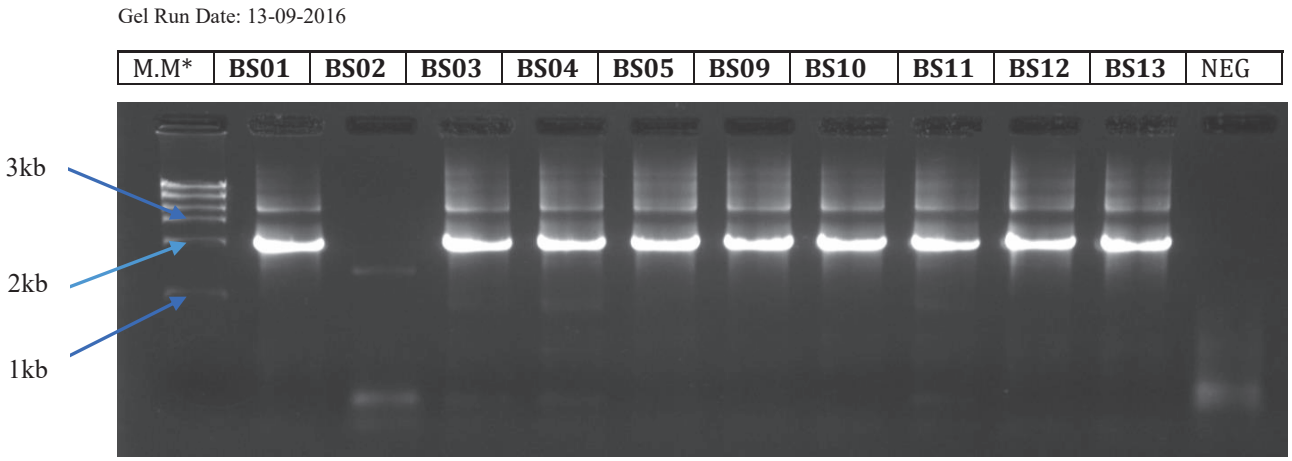


Figure 3.1: Image of secondary amplicons resolved by 1% agarose gel electrophoresis.

3.4 HIV-1 sequence analysis

3.4.1 HIV-1 subtyping

HIV-1 *Gag-protease* sequencing and subtyping was performed on samples from patients from West (n=213) and East Africa (n=160) (Table 3.1). A high level of subtype diversity was observed in West Africa, while this characteristic feature was less prevalent in the East African isolates. Subtype assignment was performed using REGA and COMET online tools independently so as to arrive at a consensus using bootstrap values of >70% to demonstrate high confidence in subtyping results. There was a marked variance in the performance of both tools at subtyping sequences (Tables 3.2 and 3.3), however, this variance was resolved by drawing maximum likelihood phylogenetic trees to make a final decision on the sequence subtype assignment.

Table 3.2: Subtyping data for West Africa

SAMPLE ID	REGA	COMET	Mphy
BS 001	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 002	A (A1)	A1	A1-like
BS 003	G	G	G
BS 004	G	G	G
BS 005	A (02_AG)	02_AG	CRF02_AG
BS 006	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 009	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 010	A (A1)	A1	A1
BS 011	CRF02_AG	?	CRF02_AG
BS 012	G	G	G
BS 013	A (A1)	02_AG	CRF02_AG
BS 014	A (02_AG)	02_AG	CRF02_AG
BS 016	A (02_AG)	02_AG	CRF02_AG
BS 019	A (02_AG)	02_AG	CRF02_AG
BS 020	A (02_AG)	02_AG	CRF02_AG

BS 021	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 022	A (02_AG)	02_AG	CRF02_AG
BS 023	A (02_AG)	02_AG	CRF02_AG
BS 024	A (37_cpx)	37_CPX	CRF37_CPX
BS 025	A (02_AG)	02_AG	CRF02_AG
BS 026	A (A1)	01_AE	CRF01_AE
BS 027	A (37_cpx)	02_AG	CRF37_CPX
BS 029	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 030	D	D	D
BS 032	A (02_AG)	02_AG	CRF02_AG
BS 035	11_CPX, A1	CRF11_CPX	CRF11_CPX
BS 038	A (02_AG)	02_AG	CRF02_AG
BS 039	A (A1)	02_AG	CRF02_AG
BS 040	A (A1)	CRF36_CPX	CRF36_CPX
BS 042	A (02_AG)	A (02_AG)	CRF02_AG
BS 043	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 044	A (02_AG)	02_AG	CRF02_AG
BS 045	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 046	G	G	G
BS 047	F1,K	?	F2
BS 048	G	G	G
BS 049	F2, K	F2	F2
BS 050	A (02_AG)	02_AG	CRF02_AG
BS 051	G	G	G
BS 053	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 054	D	D	D
BS 055	A (02_AG)	02_AG	CRF02_AG
BS 056	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 057	CRF11_CPX-like	11_cpx	CRF11_CPX
BS 064	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 065	A (A1)	?	CRF22_01A1
BS 066	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 071	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 072	A1, F1, K	?	F2
BS 073	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG

BS 074	A (02_AG)	02_AG	CRF02_AG
BS 075	A (02_AG)	02_AG	CRF02_AG
BS 076	A (A1)	A1 (check for 09_cpx)*	CRF09_cpx
BS 077	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 078	A (01_AE)	?	CRF22_01A1
BS 081	A (02_AG)	02_AG	CRF02_AG
BS 083	A (02_AG)	02_AG	CRF02_AG
BS 085	A (02_AG)	02_AG	CRF02_AG
BS 086	A (02_AG)	02_AG	CRF02_AG
BS 088	A (01_AE)	01_AE	CRF01_AE
BS 090	A (A1)	01_AE	CRF01_AE
BS 092	CRF11_cpx	?	CRF11_CPX
BS 094	A (02_AG)	02_AG	CRF02_AG
BS 096	A (02_AG)	02_AG	CRF02_AG
BS 097	A (A1)	A1	A1
BS 099	A (02_AG)	02_AG	CRF02_AG
BS 101	A (A1)	A1	CRF45_cpx
BS 102	A (02_AG)	02_AG	CRF02_AG
BS 105	A (02_AG)	02_AG	CRF02_AG
BS 107	A (A1)	01_AE	CRF22_01A1
BS 114	A (A1)	01_AE	CRF22_01A1
BS 115	A (A1)	?	CRF22_01A1
BS 117	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 118	CRF11_CPX	11_cpx	CRF11_CPX
BS 119	A (02_AG)	02_AG	CRF02_AG
BS 120	CRF02_AG	02_AG	CRF02_AG
BS 123	A (02_AG)	02_AG	CRF02_AG
BS 125	A (02_AG)	02_AG	CRF02_AG
BS 126	A2	A2	A2
BS 129	CRF02_AG	02_AG	CRF02_AG
BS 130	CRF02_AG	02_AG	CRF02_AG
BS 133	H	H	H
BS 135	H	H	H
BS 136	D	D	D
BS 137	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
BS 139	D	D	D

BS 140	A (02_AG)	02_AG	CRF02_AG
BS 141	G	G	G
BS 142	A (02_AG)	02_AG	CRF02_AG
BS 143	A (02_AG)	02_AG	CRF02_AG
BS 144	A(01_AE)	A1	CRF01_AE
BS 145	A (01_AE)	01_AE	CRF22_01A1
BS 147	A (02_AG)	02_AG	CRF02_AG
BS 159	A (A1)	A1	CRF45_cpx
BS 160	CRF11_CPX	11_cpx	CRF11_CPX
BS 163	A (02_AG)	A1 (check for 02_AG)*	A1, G
BS 166	G, A1	?	A1, G
BS 168	CRF02_AG	02_AG	CRF02_AG
BS 170	G, A1	?	A1, G
BS 175	A (01_AE)	01_AE	CRF22_01A1
BS 178	A (02_AG)	02_AG	CRF02_AG
BS 179	D	D	D
BS 180	A (02_AG)	A1 (check for 02_AG)*	CRF02_AG
NG 001	G, A1*	?	CRF43_02G
NG_002	CRF02_AG-like	?	CRF02_AG
NG 003	G	G	G
NG_006	CRF02_AG	02_AG	CRF02_AG
NG 007	G	?	G
NG 012	G	G	G
NG 032	CRF02_AG	A1(check for 02_AG)*	CRF02_AG
NG 033	CRF02_AG	CRF02_AG	CRF02_AG
NG 035	CRF02_AG	02_AG	CRF02_AG
NG 036	G	G	G
NG 038	G	G	G
NG 039	CRF02_AG	02_AG	CRF02_AG
NG 046	CRF02_AG	CRF02_AG	CRF02_AG
NG 050	G	G	G
NG 053	CRF02_AG	CRF02_AG	CRF02_AG
NG 069	02_AG, G	?	CRF02_AG
SN 001	CRF02_AG	02_AG	CRF02_AG
SN 002	CRF02_AG	02_AG	CRF02_AG
SN 003	C	C	C

SN 004	A (02_AG)	A1/02_AG*	CRF02_AG
SN 005	A (02_AG)	02_AG	CRF02_AG
SN 006	CRF02_AG	02_AG	CRF02_AG
SN 007	A (A1)	A1	A3
SN 008	CRF02_AG	A1/02_AG*	CRF02_AG
SN 009	09_cpx, A1, G*	?	CRF02_AG
SN 010	A (A1)	A1/02_AG*	CRF02_AG
SN 011	CRF02_AG	02_AG	CRF02_AG
SN 012	CRF02_AG	A1/02_AG*	CRF02_AG
SN 013	A (02_AG)	A1/02_AG*	CRF02_AG
SN 014	CRF02_AG	02_AG	CRF02_AG
SN 015	A (02_AG)	02_AG	CRF02_AG
SN 016	CRF02_AG	02_AG	CRF02_AG
SN 017	CRF02_AG	02_AG	CRF02_AG
SN 018	A (02_AG)	02_AG	CRF02_AG
SN 019	CRF02_AG	02_AG	CRF02_AG
SN 020	A (02_AG)	02_AG	CRF02_AG
SN 021	A (02_AG)	02_AG	CRF02_AG
SN 022	A (02_AG)	A1/02_AG*	CRF02_AG
SN 023	CRF02_AG	?	CRF02_AG
SN 024	CRF02_AG	?	CRF02_AG
SN 025	A (02_AG)	02_AG	CRF02_AG
SN 026	A (02_AG)	A1/02_AG*	CRF02_AG
SN 027	CRF02_AG	A1/02_AG*	CRF02_AG
SN 028	A (02_AG)	02_AG	CRF02_AG
SN 029	A (02_AG)	02_AG	CRF02_AG
SN 030	CRF02_AG	02_AG	CRF02_AG
SN 031	A (A1)	A1	A3
SN 032	CRF02_AG	02_AG	CRF02_AG
SN 033	CRF02_AG	02_AG	CRF02_AG
SN 034	CRF02_AG	02_AG	CRF02_AG
SN 035	CRF02_AG	02_AG	CRF02_AG
SN 036	A (02_AG)	02_AG	CRF02_AG
SN 037	CRF02_AG	02_AG	CRF02_AG
SN 038	A (02_AG)	02_AG	CRF02_AG
SN 039	CRF02_AG	A1/02_AG*	CRF02_AG

SN 040	CRF02_AG	02_AG	CRF02_AG
SN 041	G*	06_cpx	CRF06_cpx
SN 042	CRF02_AG	02_AG	CRF02_AG
SN 043	A (A1)	01_AE	CRF22_01A1
SN 044	G, A1*	?	CRF06_cpx
SN 045	CRF02_AG	02_AG	CRF02_AG
SN 046	A (A1)	A1	A3
SN 047	CRF02_AG	02_AG	CRF02_AG
SN 048	A (A1)	A1	A3
SN 049	CRF02_AG	02_AG	CRF02_AG
SN 050	CRF02_AG	02_AG	CRF02_AG
SN 051	CRF02_AG	02_AG	CRF02_AG
SN 052	CRF02_AG	02_AG	CRF02_AG
SN 053	A (A1)	A1	A3
SN 054	CRF02_AG	A1 (02_AG)	CRF02_AG
SN 055	A (02_AG)	A1 (02_AG)	CRF02_AG
SN 056	G	G	G
SN 057	CRF02_AG	A1 (02_AG)	CRF02_AG
SN 058	A (02_AG)	02_AG	CRF02_AG
SN 059	CRF02_AG	A1 (02_AG)	CRF02_AG
SN 060	D	D	D
SN 061	A (02_AG)	A1 (02_AG)	CRF02_AG
SN 062	A (02_AG)	02_AG	CRF02_AG
SN 063	CRF02_AG	?	CRF02_AG
SN 064	A (02_AG)	02_AG	CRF02_AG
SN 065	A (A1)	?	A3
SN 066	CRF02_AG	?	CRF02_AG
SN 067	CRF02_AG	?	CRF02_AG
SN 068	G	G	G
SN 069	02_AG, A1	?	CRF02_AG
SN 070	CRF02_AG	?	CRF02_AG
SN 071	CRF02_AG	?	CRF02_AG
SN 072	D	D	D
SN 073	CRF02_AG	?	CRF02_AG
SN 074	A1	?	CRF02_AG
SN 075	D	D	D

SN 076	A (A1)	A1	A3
SN 077	CRF02_AG	02_AG	CRF02_AG
SN 078	CRF02_AG	A1 (02_AG)	CRF02_AG
SN 079	A (A1)	A1	A3
SN 080	CRF02_AG	A1 (02_AG)	CRF02_AG
SN 081	CRF02_AG	A1 (02_AG)	CRF02_AG
SN 082	A (02_AG)	A1 (02_AG)	CRF02_AG
SN 084	CRF02_AG	?	CRF02_AG
SN 085	CRF02_AG	?	CRF02_AG
SN 086	A (02_AG)	A1 (02_AG)	CRF02_AG
SN 087	CRF02_AG	A1 (02_AG)	CRF02_AG
SN 088	A (02_AG)	02_AG	A3
SN 089	A (A1)	A1 (02_AG)	CRF02_AG
SN 090	A (02_AG)	A1 (02_AG)	CRF02_AG
SN 091	B	B	B
SN 092	A1, G	A1, G	A1, G
SN 093	A1, G	?	CRF02_AG
SN 094	A (02_AG)	02_AG	CRF02_AG
SN 096	A1	?	A3

*Low Confidence or bootstrap (<70%); ?Sequence could not be subtyped by online tool; Mphy: Molecular phylogeny

Table 3.3: Subtyping data for East Africa

SAMPLE ID	REGA	COMET	Mphy
PC 002	A (A1)	A1	A1
PC 003	A1, C	A1C*	A1C
PC 004	A (A1)	A1	A1
PC 005	A (A1)	A1	A1
PC 006	A (A1)	A1	A1
PC 008	A (A1)	A1	A1
PC 009	A (A1)	A1	A1
PC 010	A (A1)	A1	A1
PC 011	A1	A1	A1
PC 012	C, A1	?	A1C
PC 013	A1, C	A1C*	A1C
PC 014	A (A1)	A1	A1
PC 015	D, A1	A1D*	A1D
PC 016	A (A1)	A1	A1
PC 017	A (A1)	?	A1
PC 018	A (A1)	A1	A1
PC 020	A (A1)	A1	A1
PC 022	D	D	D
PC 023	D	D	D
PC 025	D	D	D
PC 026	A1, D	?	A1D
PC 027	A (A1)	A1	A1
PC 028	A (A1)	A1	A1
PC 029	A (A1)	A1	A1
PC 030	A (A1)	A1	A1
PC 031	A (A1)	A1	A1
PC 032	A (A1)	A1	A1
PC 033	A (A1)	A1	A1
PC 034	A (A1)	A1	A1
PC 036	C, A1	A1C*	A1C
PC 039	A1, D	?	A1D
PC 040	A (A1)	A1	A1

PC 041	A (A1)	A1	A1
PC 042	A (A1)	A1	A1
PC 043	A (A1)	A1	A1
PC 046	C, A1	A1C*	A1C
PC 047	A (A1)	A1	A1
PC 049	D	D	D
PC 050	A (A1)	A1	A1
PC 051	A1, D	?	A1D
PC 053	A1, D	?	A1D
PC 054	D	D	D
PC 055	A (A1)	A1	A1
PC 056	A1D	A1D	A1D
PC 057	A (A1)	A1	A1
PC 058	D	D	D
PC 060	D	D	D
PC 061	A (A1),	A1	A1D
PC 062	A (A1)-like	A1D*	A1D
PC 063	A1	A1	A1
PC 066	D	D	D
PC 067	A (A1)	A1C*	A1
PC 068	D	D	D
PC 070	A1	A1	A1
PC 071	D, A1	A1D*	D
PC 076	A (A1)	A1	A1
PC 077	A (A1)	A1	A1
PC 078	A1	A1	A1
PC 079	A1, H	?	A1D
PC 080	A1	A1	A1
PC 081	A (A1)	A1	A1
PC 083	A (A1)	A1	A1
PC 087	A (A1)	A1	A1
PC 088	A (A1)	A1	A1
PC 089	A (A1)	A1	A1
PC 092	A (A1)	A1	A1
PC 093	A (A1)	A1	A1
PC 095	D	D	D

PC 096	A1, D	A1D*	A1D
PC 097	A (A1)	A1	A1
PC 098	A (A1)	A1	A1
PC 099	D	A1	D
PC 100	A1	A1	A1
PC 101	A (A1)	A1	A1
PC 104	A (A1)	A1	A1
PC 105	A (A1)	A1	A1
PC 106	A (A1)	A1	A1
PC 107	A (A1)	A1	A1
PC 109	A (A1)	A1	A1
PC 112	A (A1)	A1	A1
PC 113	A (A1)	A1	A1
PC 114	D	A1D*	A1D
PC 116	A (A1)	A1	A1
PC 117	A (A1)	A1	A1
PC 118	A (A1)	A1	A1
PC 121	A1	A1	A1
PC 122	A (A1)	A1	A1
PC 123	D	D	D
PC 125	D	D	D
PC 128	A (A1)	A1	A1
PC 129	A (A1)	A1	A1
PC 132	A (A1)	A1	A1
PC 137	A (A1)	A1	A1
PC 141	A (A1)	A1	A1
PC 142	A1, D	?	A1D
PC 143	A (A1)	A1	A1
PC 144	A (A1)	A1	A1
PC 148	A1	A1	A1
PC 149	D	D	D
PC 150	A1, D	A1D*	A1D
PC 151	A (A1)	A1	A1
PC 154	A (A1)	A1	A1
PC 155	A1	A1	A1
PC 158	A (A1)	A1	A1

PC 159	A (A1)	A1	A1
PC 161	A (A1)	A1	A1
PC 162	A (A1)	A1	A1
PC 163	D	D1	D
PC 164	A (A1)	A1	A1
PC 165	D, A1	A1D*	D
PC 166	D	D	D
PC 167	D, A1	A1D*	A1D
PC 168	D	D	D
PC 170	D	D	D
PC 171	D	D	D
PC 172	D	D	D
PC 173	A (A1)	A1	A1
PC 174	A (A1)	A1	A1
PC 175	A1, D	A1D*	A1D
PC 176	A (A1)	A1	A1
PC 177	A (A1)	A1	A1
PC 178	D	D	D
PC 179	A (A1)	A1	A1
PC 181	D	D	D
PC 182	D	D	D
PC 183	D	D	D
PC 185	D	D	D
PC 186	D	D	D
PC 187	D	D	D
PC 188	A (A1)	A1	A1
PC 190	D	D	D
PC 191	D	D	D
PC 192	D	D	D
PC 193	D	D	D
PC 194	A (A1)	A1	A1
PC 195	A (A1)	A1	A1
PC 196	D	D	D
PC 197	A (A1)	A1	A1
PC 198	A (A1)	A1	A1
PC 199	A (A1)	A1	A1

PC 200	A (A1)	A1	A1
PC 201	A (A1)	A1	A1
PC 202	A (A1)	A1	A1
PC 203	A (A1)	A1	A1
PC 204	A (A1)	A1	A1
PC 205	A (A1)	A1	A1
PC 206	A (A1)	A1	A1
PC 213	D	D	D
PC 217	A (A1)	A1	A1
PC 222	A (A1)	A1	A1
PC 223	D, A1	?	A1D
PC 224	A (A1)	A1	A1
PC 225	A (A1)	A1	A1
PC 227	A1D	A1D	A1D
PC 229	A (A1)	A1	A1
PC 230	A (A1)	A1	A1
PC 232	A (A1)	A1	A1
PC 233	A (A1)	A1	A1
PC 235	A (A1)	A1	A1
PC 239	A1, D	?	A1

*Low Confidence or bootstrap (<70%); ?Sequence could not be subtyped by online tool; Mphy: Molecular phylogeny

3.4.2 Sequence Phylogenetic Analysis

Phylogenetic analysis was conducted by drawing maximum likelihood trees using the online PhyML tool on the LANL database. The trees showed clustering of patient derived query sequences with their corresponding reference sequences downloaded from the LANL. CRF02_AG which was the most predominant subtype formed two sub-clusters within its major cluster which were largely country specific for Cameroonian sequences and a separate cluster for Senegalese/Nigerian seqeuences. CRF06_cpx and CRF11_cpx formed their independent clusters, while CRF09_cpx, CRF45_cpx, A1, A3, and A2 all formed different lineages from within the A1 cluster which showed how closely related these subtypes were. Furthermore, recombination analyses of these CRFs showed that they were pure subtype A1 within the gag region, hence their positioning on the phylogram. Due to the closeness of CRF01_AE and CRF22_01A1, the online tools found it challenging to distinguish between both and in most cases identified them as pure A1. On the other hand, phylogenetic analysis was able to resolve both into separate clades despite their close genetic relatedness. Other CRFs such as CRF36_cpx, CRF37_cpx and CRF43_02G formed their own separate lineages, though CRF43_02G was very close to the pure subtype G cluster. Pure subtypes such as B, C, D, F2, G, and H were also identified in West Africa making it highly diverse as shown in Figure 3.2A, while Figure 3.2B showed the branching topology within the West African tree to highlight how the different subtypes related to one another. In East Africa, the phylogenetic analysis showed that there is less diversity within the region.The major clusters within the maximum likelihood tree were subtype A1, A1D and D, there was to a much less extent subtype A1C (Figure 3.3). Having identified subtype A1 and D as common denominators within each region (though they were more predominant in East Africa), another phylogenetic analysis was performed comparing these subtypes which showed the presence of region-specific evolutionary adaptation within subtypes (Figure 3.4).

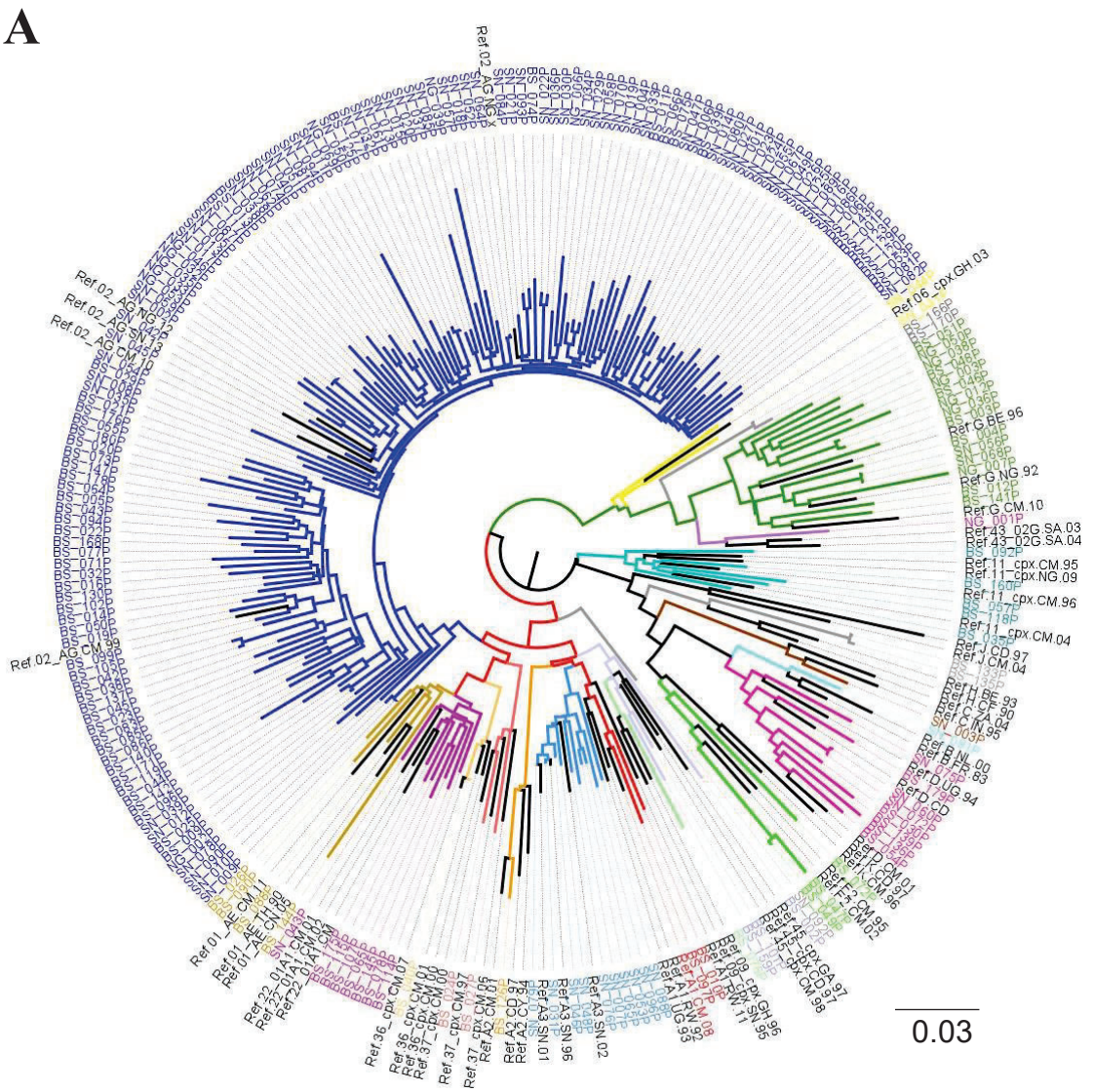


Figure 3.2: Phylogenetic analysis of gag-protease sequences for West Africa

Subtype Key

A1	A2	A3	B	C
CRF01_AE	CRF02_AG	CRF06_cpx	CRF09_cpx	CRF11_cpx
CRF22_01A1	CRF36_cpx	CRF37_cpx	CRF43_02G	CRF45_cpx
D	F2	G	H	A1, G

Figure 3.2: Phylogenetic analysis of *gag-protease* sequences for West Africa (A) Maximum-likelihood phylogenetic tree showing all *gag-protease* sequences in a polar tree layout; and **(B)** phylogram layout to highlight relationships between each cluster using the PhyML tool on the LANL HIV database. Subtypes and inter-subtype recombinants were confirmed using reference sequences downloaded from the LANL HIV database. Subtype classification was colour coded as shown in the key. Branches representing subtype reference sequences are labelled in black.

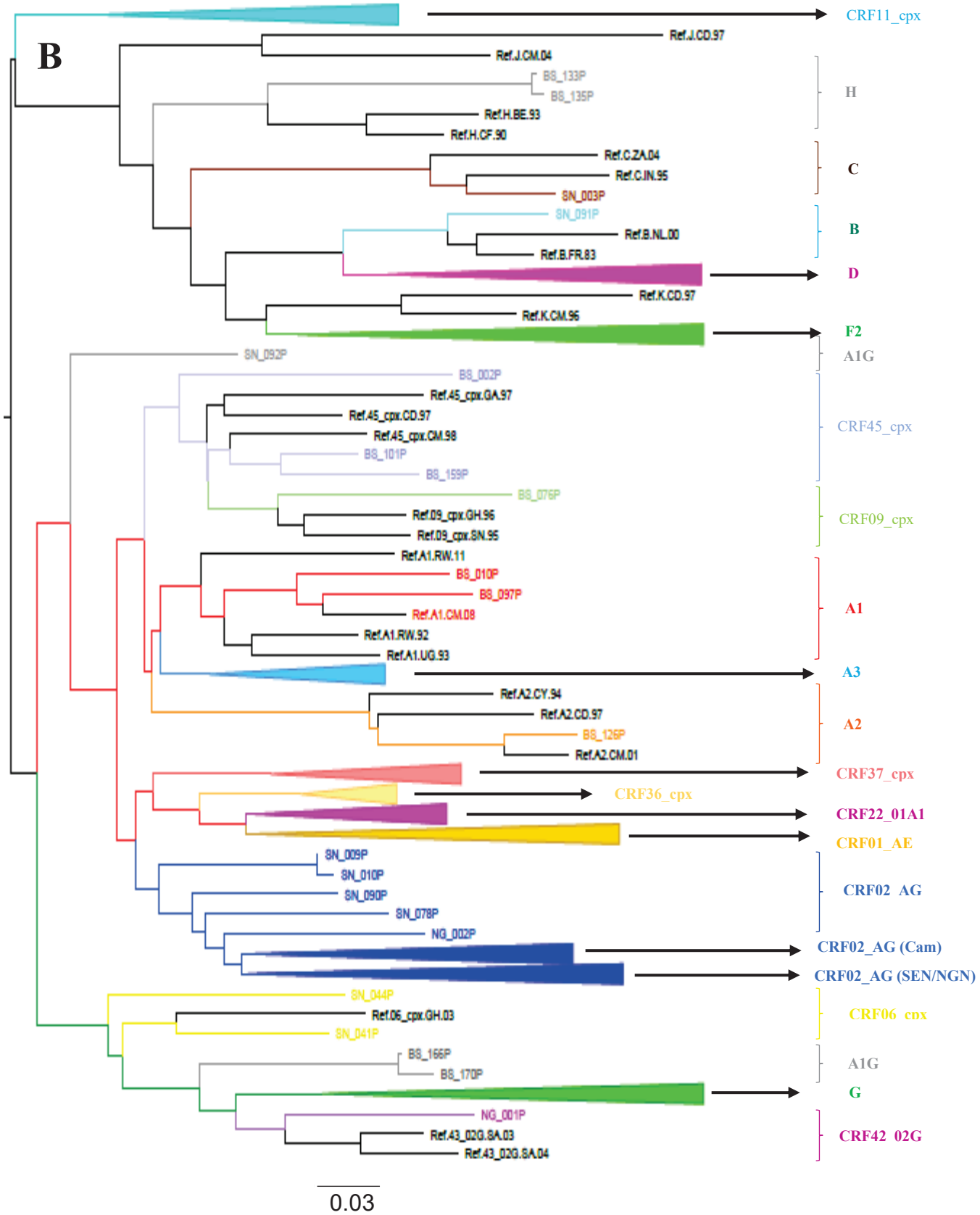

B
CRF11_cpx
Ref.J.CD.97
Ref.J.CM.04
BS_133P
BS_135P
H
Ref.H.BE.93
Ref.H.CF.90
Ref.C.ZA.04
Ref.C.IN.95
SN_003P
C
SN_091P
Ref.B.NL.00
Ref.B.FR.83
B
D
Ref.K.CD.97
Ref.K.CM.96
F2
SN_092P
A1G
BS_002P
Ref.45_cpx.GA.97
Ref.45_cpx.CD.97
Ref.45_cpx.CM.98
BS_101P
BS_159P
CRF45_cpx
BS_076P
Ref.09_cpx.GH.96
Ref.09_cpx.SN.95
CRF09_cpx
Ref.A1.RW.11
BS_010P
BS_097P
Ref.A1.CM.08
Ref.A1.RW.92
Ref.A1.UG.93
A1
A3
Ref.A2.CY.94
Ref.A2.CD.97
BS_126P
Ref.A2.CM.01
A2
CRF37_cpx
CRF36_cpx
CRF22_01A1
CRF01_AE
SN_009P
SN_010P
SN_090P
SN_078P
NG_002P
CRF02_AG
CRF02_AG (Cam)
CRF02_AG (SEN/NGN)
SN_044P
Ref.06_cpx.GH.03
SN_041P
CRF06_cpx
BS_166P
BS_170P
A1G
G
NG_001P
Ref.43_02G.SA.03
Ref.43_02G.SA.04
CRF42_02G
0.03

Figure 3.3:Phylogenetic analysis of gag-protease sequences for East Africa

Subtype Key

A1	A1C	A1D	D

Figure 3.3: Phylogenetic analysis of *gag-protease* sequences for East Africa. Maximum likelihood trees were constructed using the PhyML tool on the LANL HIV database. Subtypes and inter-subtype recombinants were confirmed using reference sequences downloaded from the LANL HIV database. Subtype classification was colour coded as shown in the key. Branches representing subtype reference sequences are labelled in black.

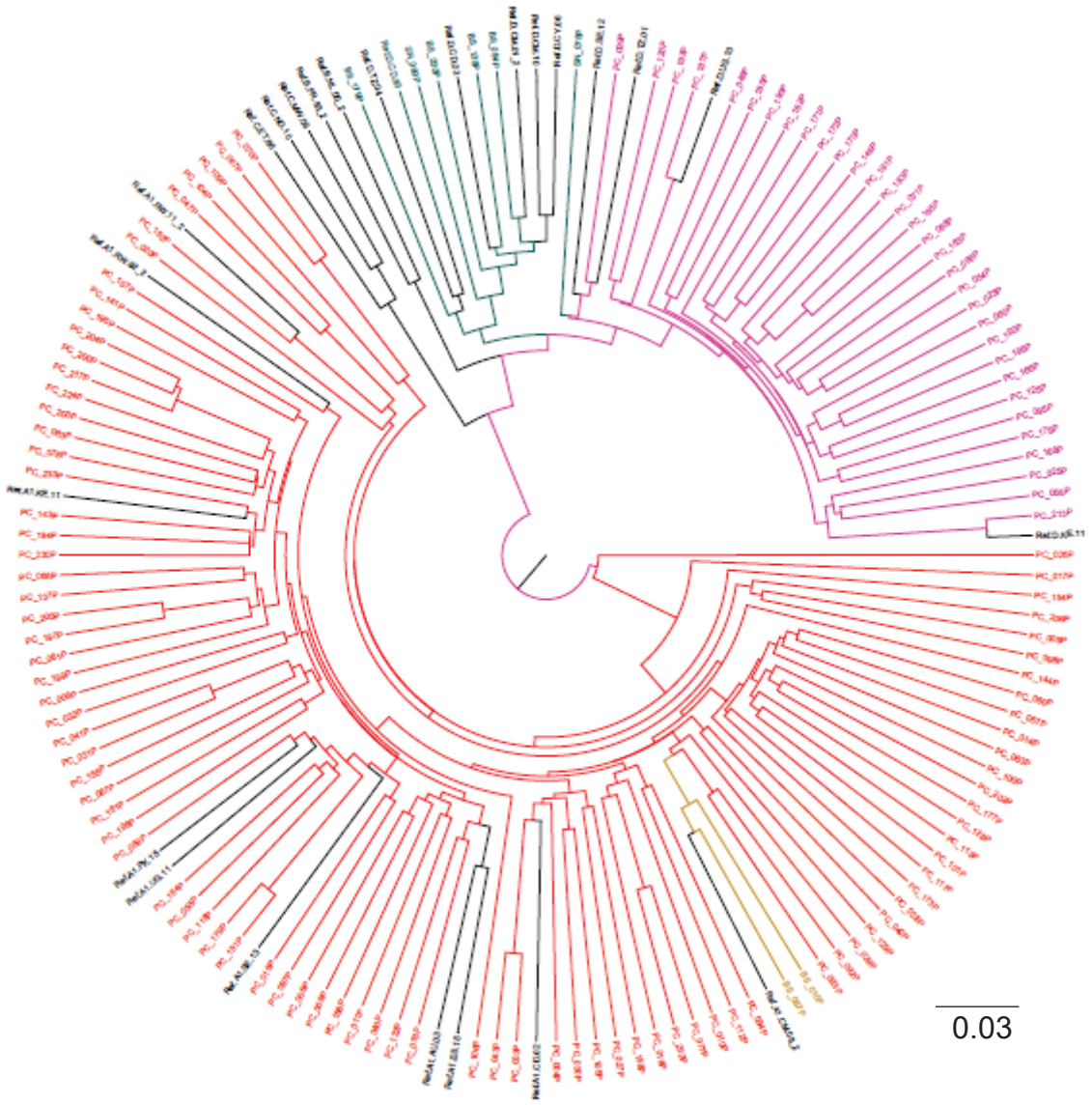


Figure 3.4: Phylogenetic tree showing common subtypes in West and East Africa

Subtype key

A1 (East Africa)	A1 (West Africa)	D (East Africa)	D (West Africa)

Figure 3.4: Maximum likelihood tree showing common subtypes A1 and D in West and East Africa with region-specific subtype evolution.

3.4.3 HIV-1 subtype geographic distribution

CRF02_AG was the most prevalent circulating subtype in West Africa at 65%, followed by subtype G and A3 at 7% and 5% respectively, while the remaining pure subtypes and CRFs were present at <5% prevalence Figure 3.5A. It was also the most predominant subtype across the 3 different countries. It was interesting to note an apparent decline in subtype A1 across the region. Cameroon and Senegal seemed to have the most genetic diversity of the three countries within the region, though more sampling from Nigeria would give more information about the diversity of its epidemic. Senegal was the only country with subtype A3 (Figure 3.5 D). Cameroon had the most complex expidemic within the region (Figure 3.5B).

To extend previous analysis of *Gag-protease* subtype in East Africa (144) by our group, we increased the sample size for the most prevalent subtypes in the region. Preselected subtypes A, D and AD recombinants based on previous Pol subtyping data (174) were used. The data showed that subtype A1 was most prevalent in the population sampled followed by subtype D, A1D and A1C (Figures 3.6 A-D). Subtype A1 remained the most predominant in Kenya and Rwanda (Figures 3.6 B and C). In Uganda there was almost an equal split between subtype A1 (40%) and D (44%), which was follwed by URF A1D at 16% (Figure 3.6D). 80% of URF A1D came from Uganda.

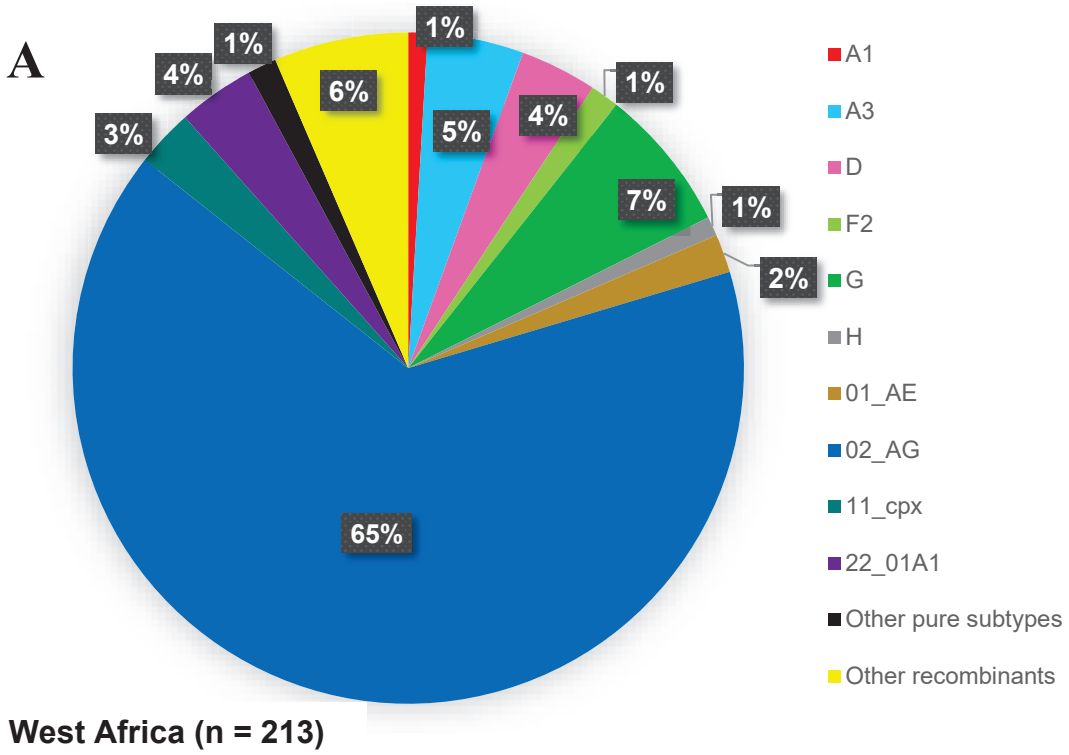
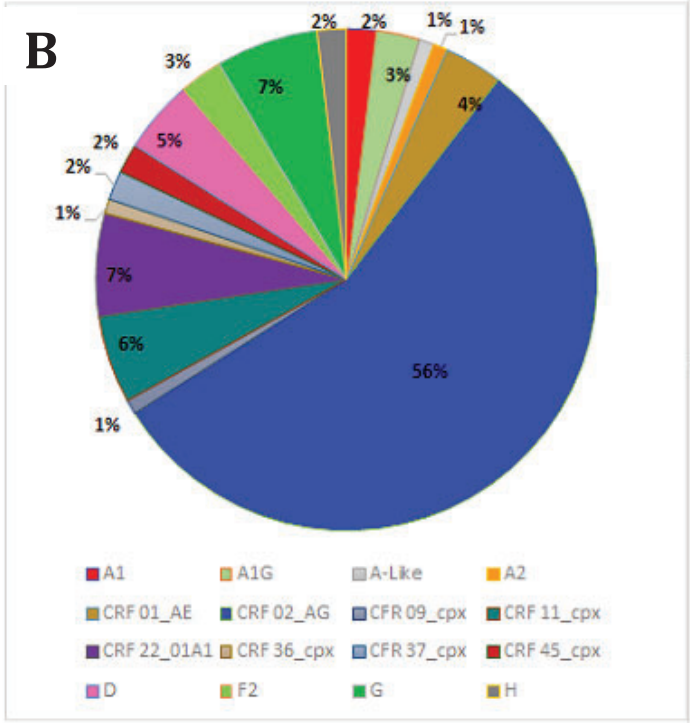
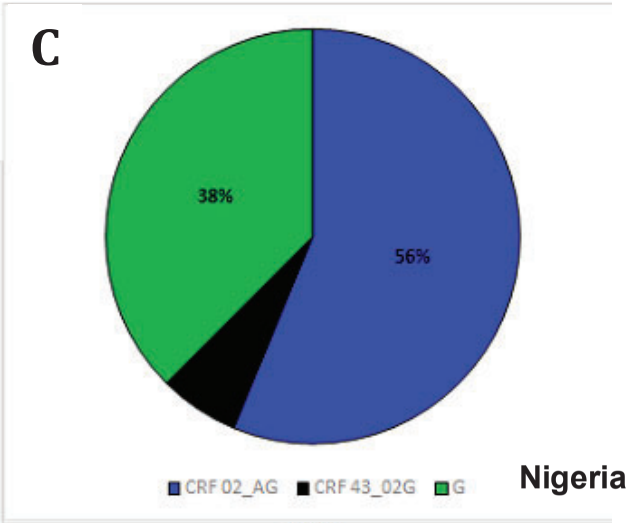


Figure 3.5: Subtype distribution in West Africa

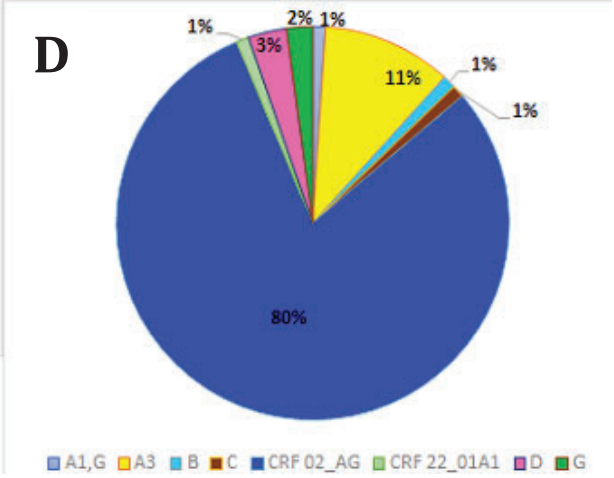
Figure 3.5: Subtype geographical distribution. Genetic subtype distribution in West Africa **(A)** and by countries Cameroon **(B)**, Nigeria **(C)** and Senegal **(D)**.



Cameroon (n = 106)



Nigeria (n = 16)



Senegal (n = 92)

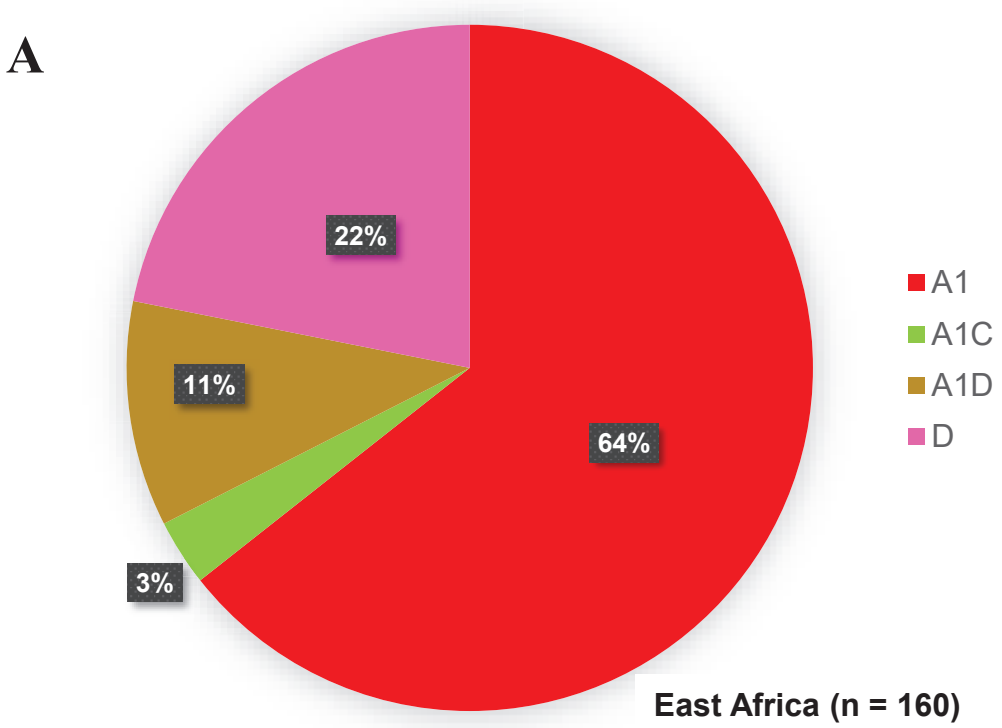
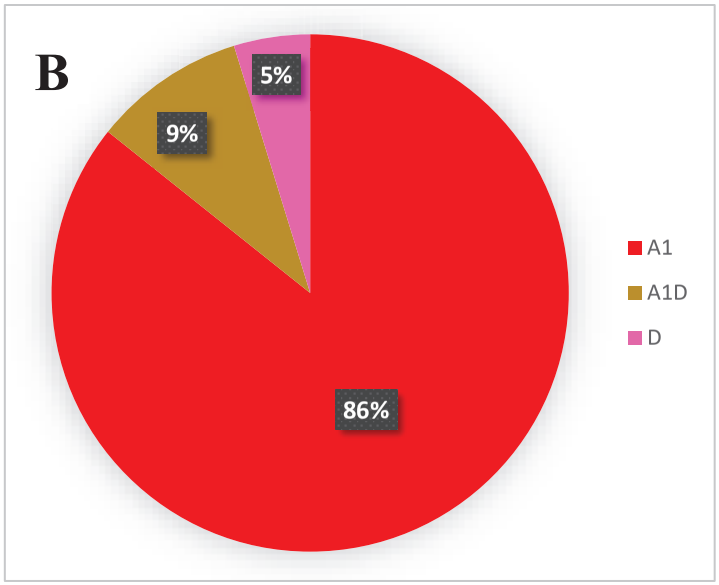
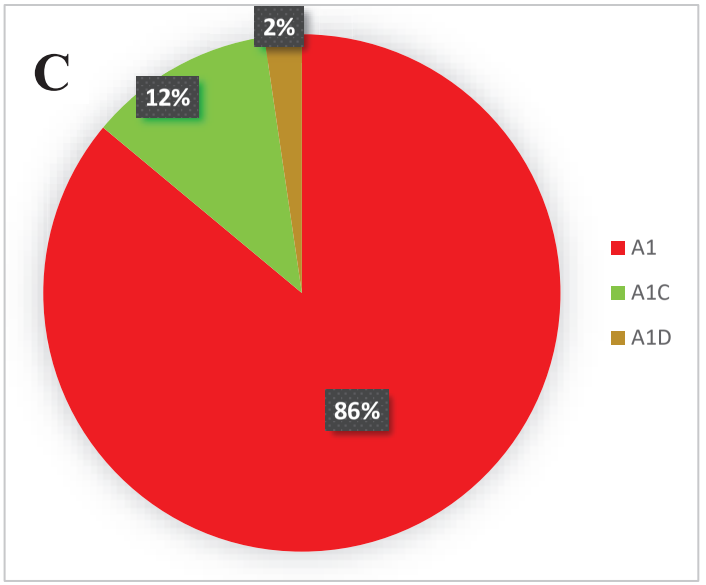


Figure 3.6: Subtype distribution in East Africa

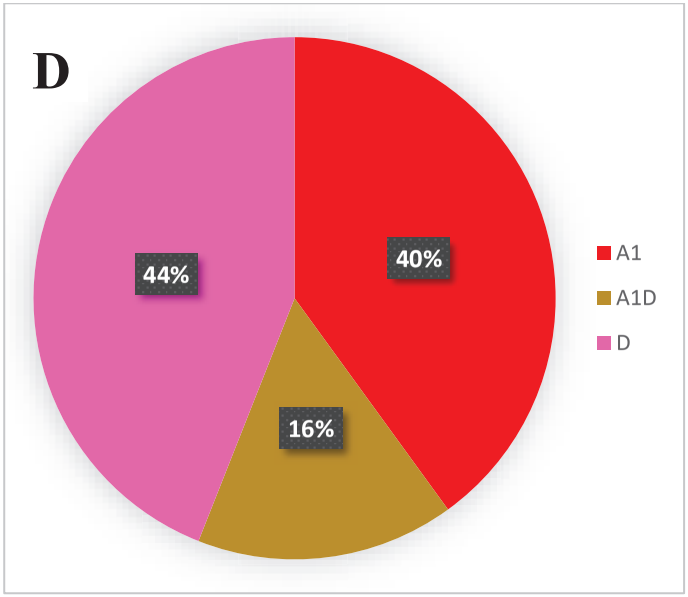
Figure 3.6: Subtype geographical distribution in East Africa. Genetic subtype distribution in East Africa **(A)** and by countries Kenya **(B)**, Rwanda **(C)** and Uganda **(D)**.



Kenya (n = 42)



Rwanda (n = 43)



Uganda (n = 75)

3.4.4 Gag-protease inter-subtype recombination analysis

There are a myriad of bioinformatics methods and tools that are employed to infer HIV-1 subtypes and inter-subtype recombination patterns; however, following trials with different tools, the REGA, jpHMM and Simplot (184) tools were chosen for my analysis. These three methods were used to complement and validate one another due to inherent limitations in each tool, thereby increasing confidence in the accuracy of the results. Wherever there was conflict between tools, 2 out of 3 that were in concordance with one another would be chosen (Figure 3.7), and the results still confirmed through phylogenetic analysis of the fragments of the sequences (Figure 3.8). Inter-subtype recombination analysis within the *gag-protease* region indicates that all HIV-1 CRFs and other recombinant forms in West and East Africa are recombinants of A1 (Figure 3.9). In West Africa, the analysis shows that the most prevalent inter-subtype recombinants were those of subtypes A1 and G, and less common inter-subtype recombinants were combinations of A1 and J, or A1, G and J, with CRF02_AG being the most common followed by CRF11_cpx. In the *gag* region, the remaining West African CRFs were close genetic relatives of subtype A1 (CRF01_AE, CRF09_cpx, and CRF22_01A1, CRF36_cpx, CRF37_cpx and CRF45_cpx) or A1, G (CRF06_cpx and CRF43_02G) recombinants. Majority of the inter-subtype recombinants have a breakpoint within the p6 region in *gag-protease*. In East Africa, inter-subtype recombination patterns between subtype A1 and C or A1 and D were identified. All 5 A1C recombinants were identified from Rwanda and had subtype A component towards the -3'end of the sequences, while 80% (n=12) of all A1D identified originated from Uganda where A1 and D co-circulate in approximately equal proportions, with subtype D (80%) being the predominant component at the -3'end of the sequences. This recombinant pattern is consistent with previous work done in our group (144) indicating an evolutionary preference in A1D inter-subtype recombinant Gag sequences for subtype D towards the 3' end of the sequence.

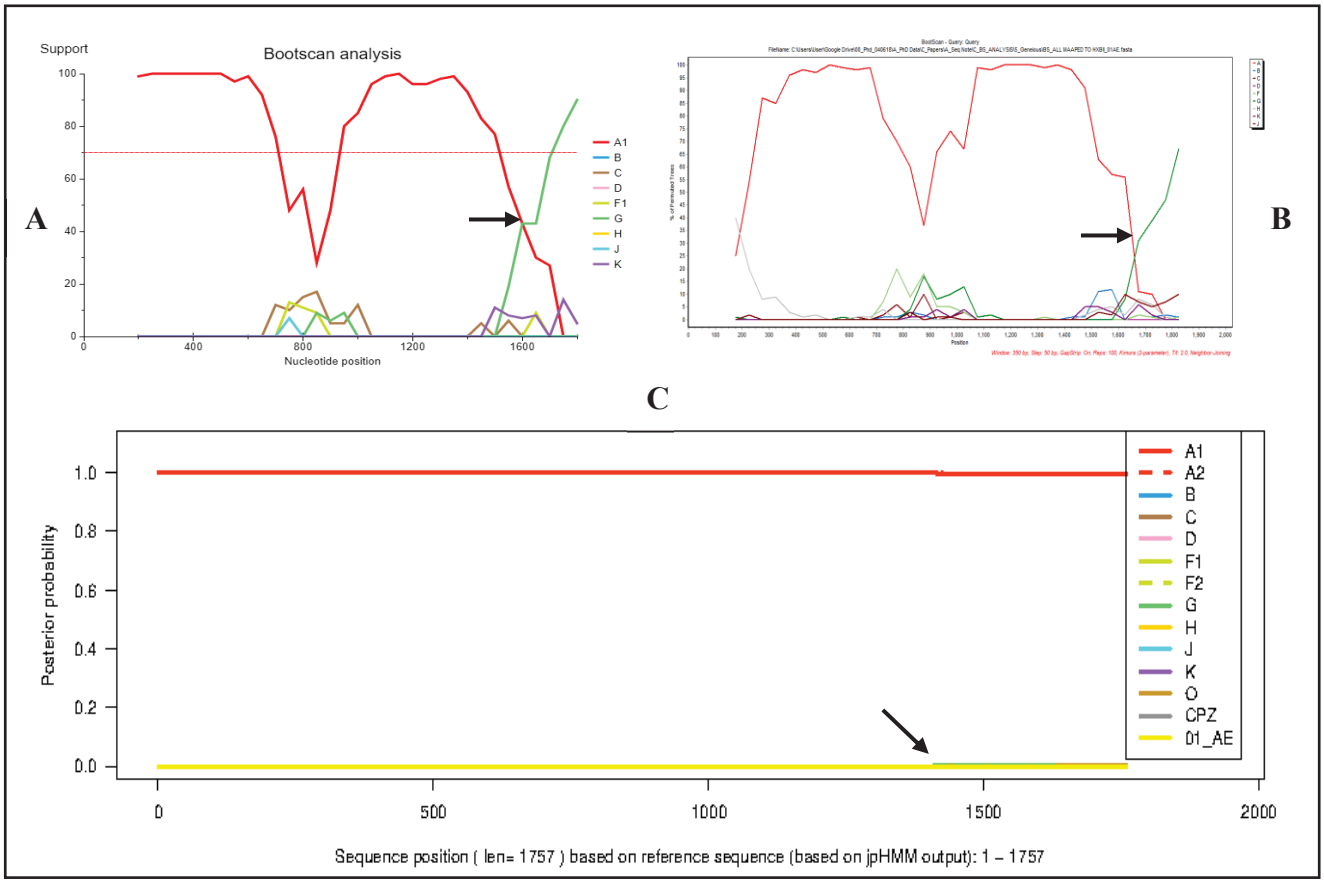


Figure 3.7: Output of inter-subtype recombination analysis tools

Figure 3.7: A representative output showing different bioinformatic tools used to infer inter-subtype recombination analysis in HIV-1 sequences. REGA **(A)** and Simplot **(B)** both used bootstrapping method to compare query sequences to reference sequences while the jpHMM **(C)** uses posterior probability to infer recombination points within the sequences. Tools were not always in agreement, however a concordance of 2 out of 3 was acceptable to infer recombination coordinates. Black arrows point at predicted recombination points along the sequence within each tool. This report was generated for BS_001 a CRF02_AG recombinant.

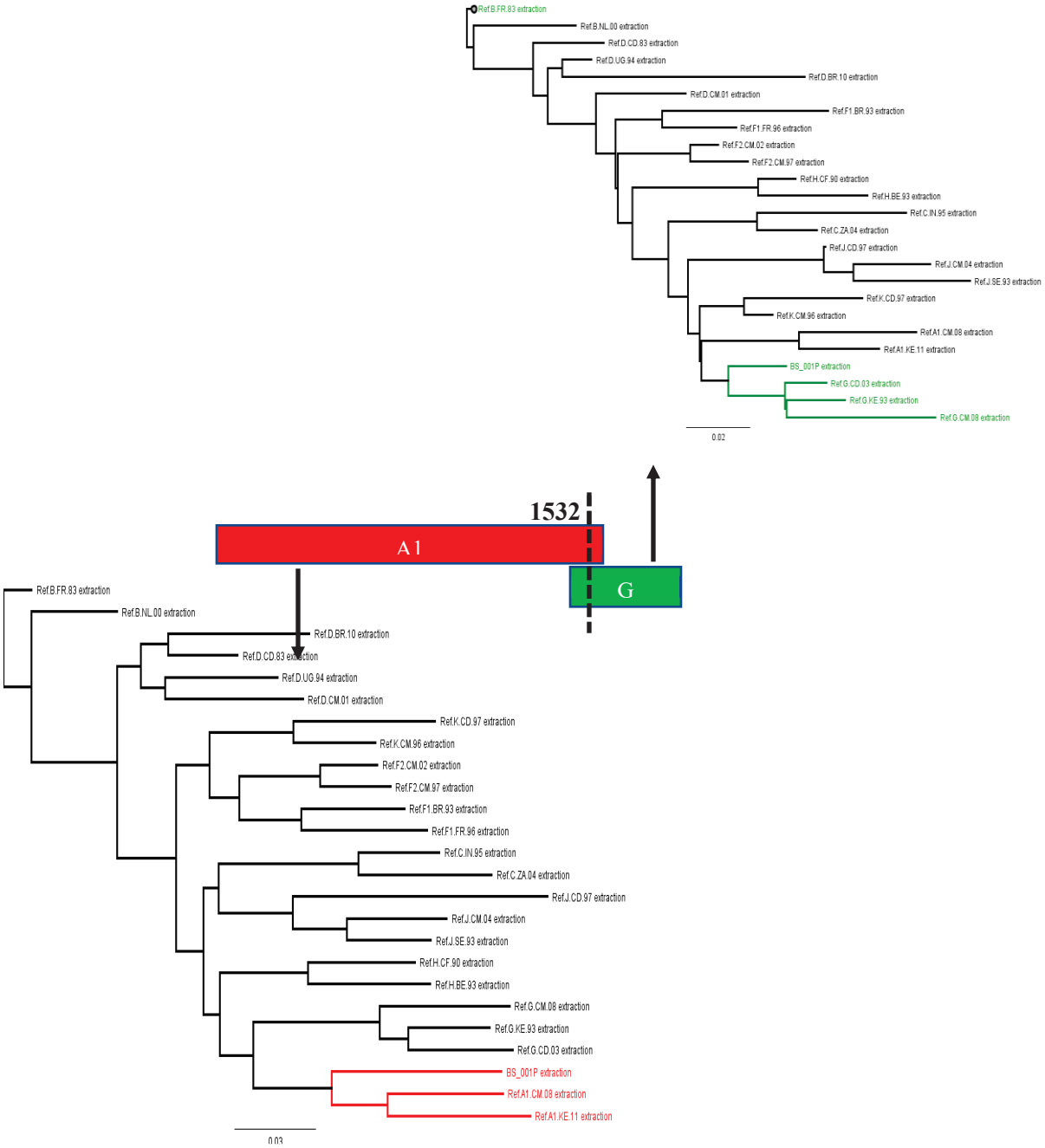


Figure 3.8: Phylogenetic confirmation of sequence fragments

Figure 3.8: Phylogenetic confirmation of sequence fragments. Predicted sequence fragments by REGA, Simplot and jpHMM tools were further confirmed using phylogenetic analysis by extracting sequence fragment and drawing maximum likelihood trees. Sequence fragments clustered with their respective reference sequences. Number indicated in the figure is the actual nucleotide position within the sequence where the inter-subtype recombination took place. Numbering on figure represents actual nucleotide position in the sequence and not HXB2 numbering.

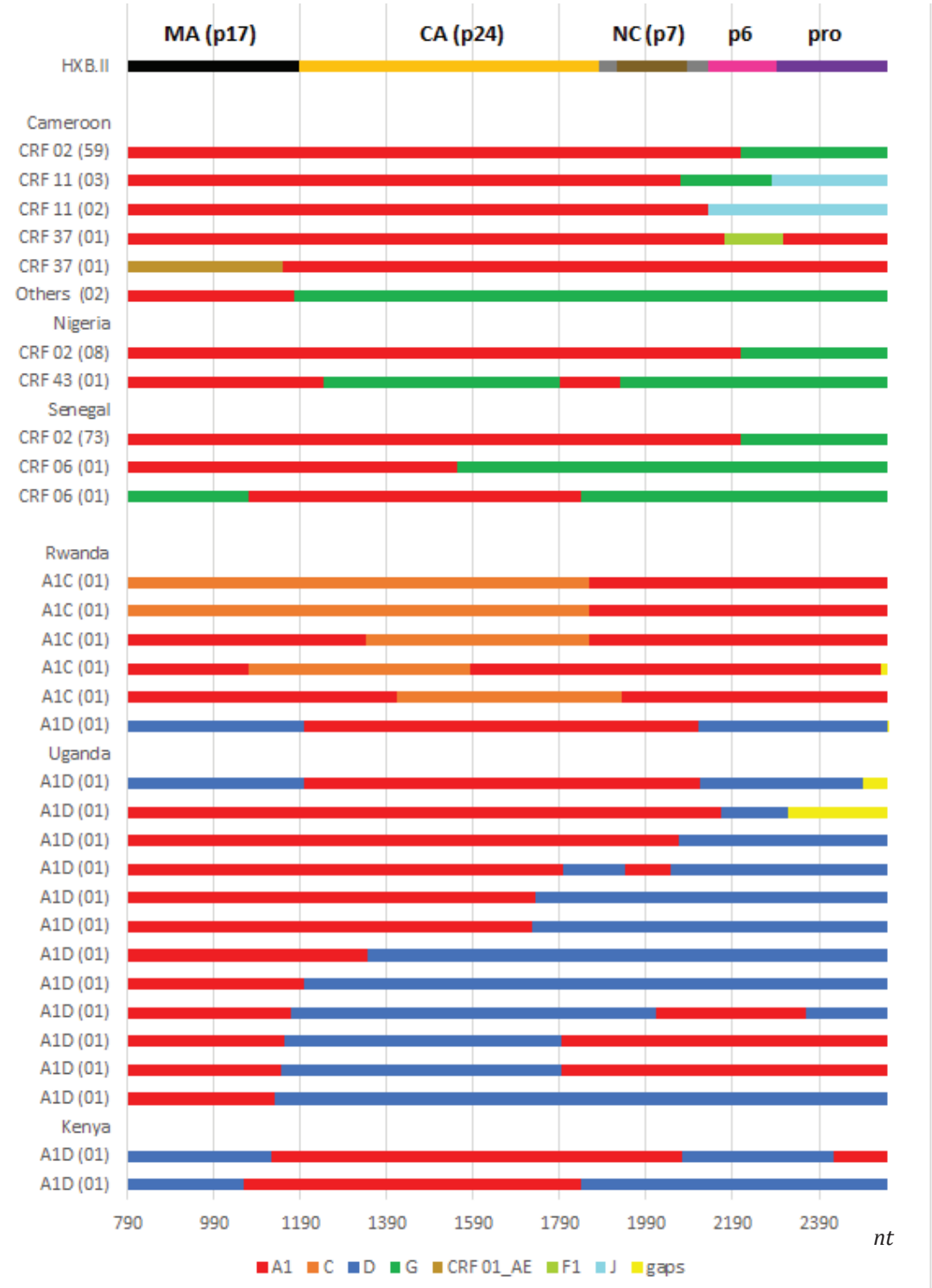


Figure 3.9: Inter-subtype recombination analysis

Figure 3.9: Graphic summary of inter-subtype recombination patterns within HIV-1 *gag-protease* from West and East Africa. Consensus sequence patterns of recombinants circulating within different countries in West Africa. Individual sequences are represented in East Africa. The number of sequences identified are indicated in brackets. Missing sequence data is shown in yellow. Gene lengths and breakpoints were drawn according the HXB2 numbering.

.

3.5 Gag-protease replication capacity analysis

3.5.1 Distribution and variability of replication capacity assay

The VRC of patient-derived *gag-protease* chimeric viral stock was reported as the slope of viral spread from day 3-6 post-infection in comparison to that of the control NL4-3 strain. The assays were performed in independent duplicate experiments. The frequency distribution was driven by the prevalent subtype in each region which were subytpes CRF02_AG and A1 in West and East Africa respectively. In West Africa VRC distribution was skewed to the right (Figure 3.10A) showing higher mean VRC among subtypes within the region, while a normal distribution was observed on the East African histogram (Figure 3.10B) showing a more balanced distribution in the VRC within the region. Duplicate independent measurements performed to validate experiments and results were highly concordant. They were reported as the average of the two measurements. Spearman and Pearson correlation analysis were performed for West (Figure 3.10C) and East (Figure 3.10D) Africa respectively based on their frequency distribution patterns.

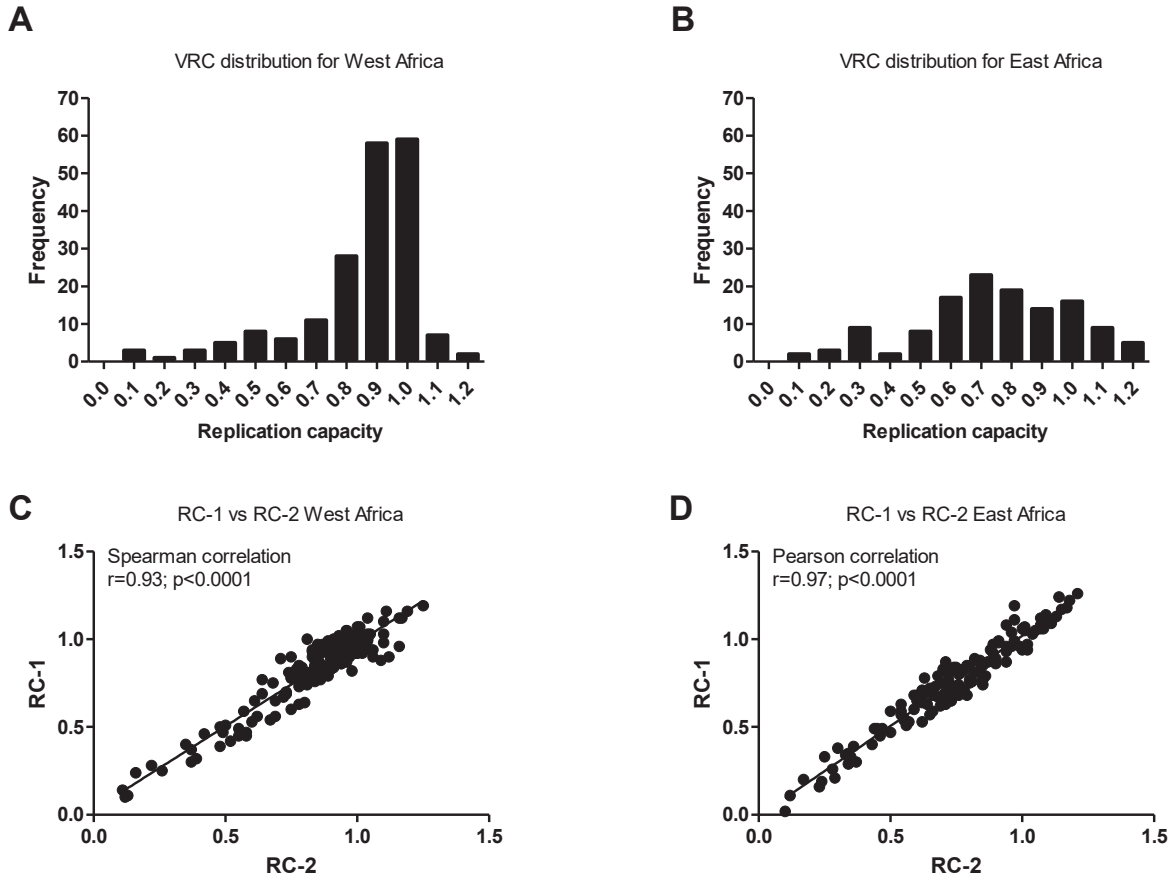


Figure 3.10: VRC distribution analysis for West and East Africa.

Figure 3.10: VRC distribution and variation in West and East African Isolates (A and B); distribution and reproducibility of *gag-protease* driven replication capacity assay in West and East Africa (C and D) by performing duplicate measurements for all samples. RC-1 is 1st replicate for all RC samples, while RC-2 is the second replicate. This experiment was done to demonstrate how reproducible the assay is.

3.5.2 Correlation of VRC with markers of disease progression

Previous work from our group have shown strong association of VRC with markers of disease progression in subtype C from South Africa and various subtypes from East Africa (143, 144, 154); here, VRC of all isolates regardless of their region were correlated with viral load and $CD4^+$ counts separately. The results showed a weak positive correlation between viral load measured in $\log_{10}$ copies/mL and VRC (Spearman correlation r = 0.12; p = 0.071), while it

negatively correlated with $CD4^+$ count measured in cells/μL (Spearman correlation r = -0.06; $p = 0.322$).

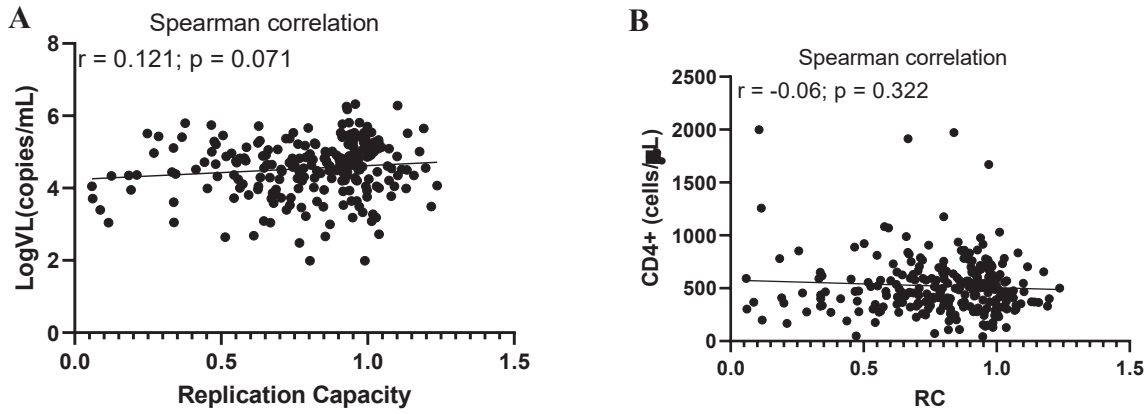


Figure 3.101: VRC and markers of disease progression

Figure 3.11: Gag-protease driven viral replication capacity and markers of disease progression such as viral load **(A)** and $CD4^+$ count **(B).**

3.5.3 *Gag-protease* driven replication capacity in West and East African subtypes

Previous work from our group showed that there was subtype hierachy in gag-driven VRC (144) and the literature had also shown that individuals infected with certain subtypes progressed faster over patients infected with other subtypes (126, 128, 131, 185). We hypothesised that these subtype-specific differences in VRC could contribute to the mosaic nature of the epidemic in sub-Saharan Africa. To evaluate these observations, a comparative analyses was performed using the student t tests and ANOVA to identify possible subtype specific differences in *gag-protease*-driven VRC. Overall, viruses from West Africa had significantly higher VRC compared to those from East Africa (Figure 3.12A). A subsequent comparison of the most prevalent circulating subtype in each region showed that CRF02_AG, which was most predominant in West Africa had a higher viral replication capacity than subtype A1 (Figure 3.12B) which was the most predominant subtype from the samples collected in East Africa, indicating that the prevalent subtypes were driving the overall difference in VRC observed between both regions. The same subtypes – A1 and D - compared between regions showed no difference in viral replication capacity according to region (Figure 3.12C and D). In West Africa pure subtype A1 had the lowest *gag-protease* driven VRC when compared to subtypes D, G, CRF02_AG, and CRF11_cpx; and though A1 was used as the reference subtype for this analysis, other genetically close A1 relatives like A3/CRF22_01A1 also had significantly lower *gag-protease* driven VRCs compared to the other subtypes (Figure 3.13A). In East Africa pure subtype A1 also had the lowest *gag-protease* driven replication capacity when compared to other subtypes, and a hierarchy of A<A1C<A1D<D was observed, which is consistent with previous reports by our group (144) (Figure 3.13B). Comparison between the most prevalent CRF and URF in West and East African respectively showed no significant difference (Figure 3.13C).

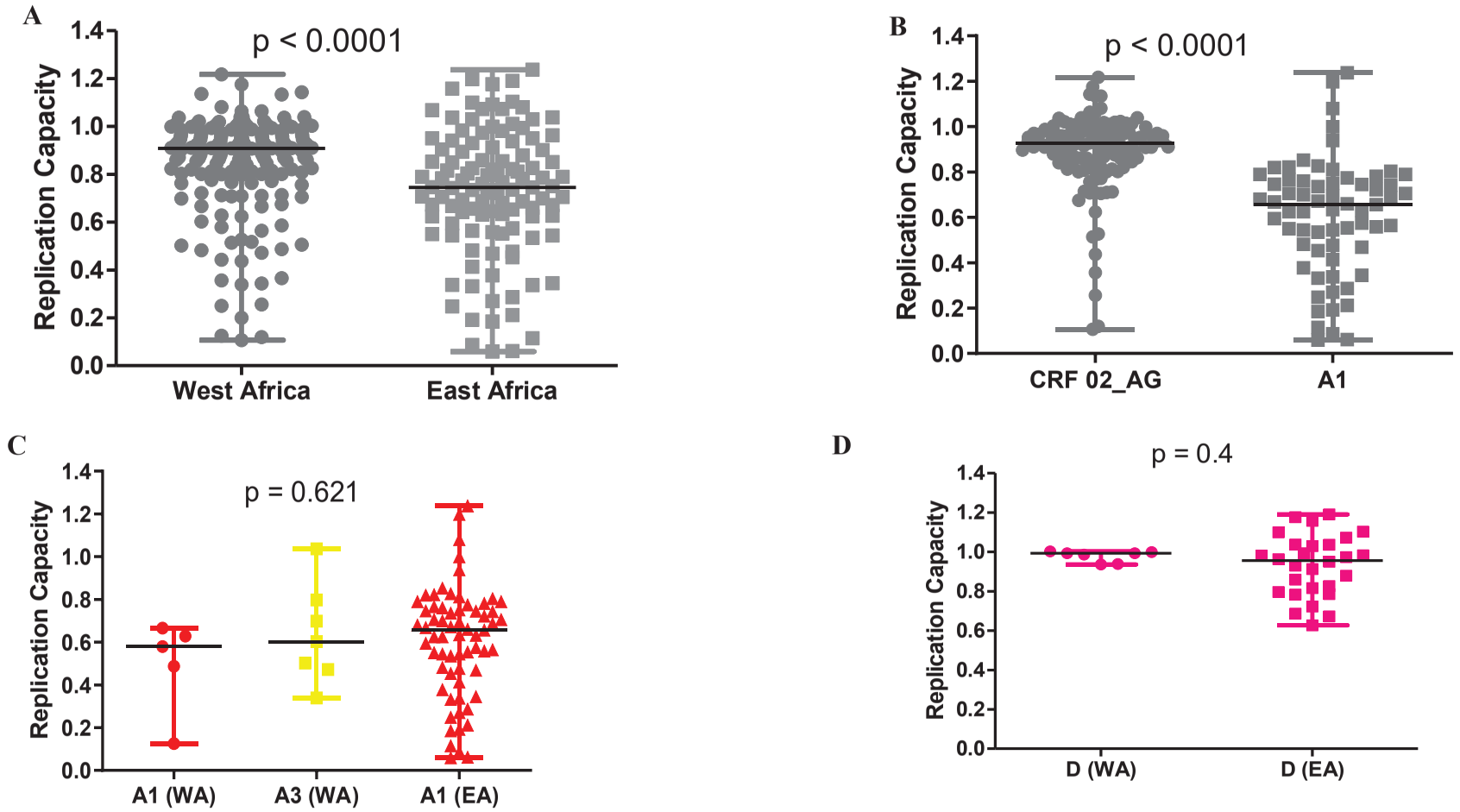


Figure 3.112: Gag-protease VRC within West and East Africa

Figure 3.12: Gag-protease-driven replication capacity within East and West African subtypes A) Patient-derived *gag-protease* recombinant viruses from West Africa have significantly higher replication capacity compared to those from East Africa irrespective of subtype. **B)** Recombinant viruses of predominant strains in West Africa (CRF02_AG) displayed higher replication capacity compared to recombinants of the predominant strains (A1) in East Africa. **C)** There was no significant difference in replication capacity between subtype A1 and A3 viruses from West Africa and subtype A1 from East Africa as well as the subtype D variants from West versus East Africa **D)** The 1-way ANOVA and Mann-Whitney tests were used, accordingly.

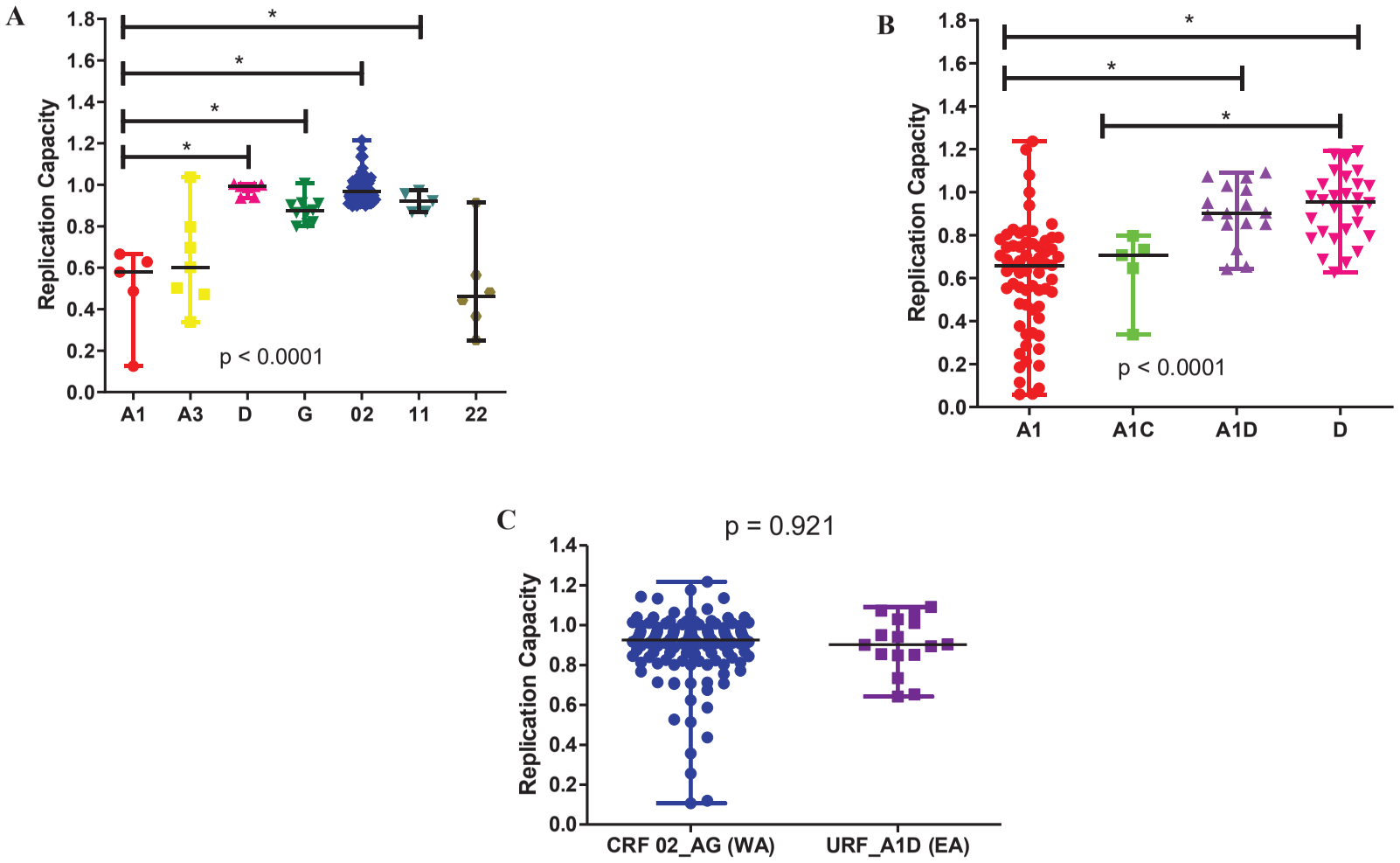


Figure 3.123: Gag-protease VRC within East and West Africa

Figure 3.13: Gag-protease-driven replication capacity within East and West African subtypes A) Significant differences in patient derived *gag-protease* replication capacities between subtypes A1 and subtype D, G, CRF02_AG and CRF11_cpx. Similar differences were observed in A3/CRF22_01A1 and these other subtypes. **B)** East Africa also display significant differences in replication capacity between circulating subtypes with *p*-values less than 0.0001 and 0.005. A 1-way ANOVA showed a significant difference in viral replication capacity of $p < 0.0001$ for West and East African viruses; followed by a Tukey multiple comparison test which showed A1 and its variants as having lower replication capacity compared to other subtypes between both regions. $*p < 0.05$. CRFs are represented by their respective numbers (CRF02_AG, CRF11_cpx, CRF22_01A1) **C)** No significant difference between the most prevalent CRF and URF in West and East African epidemic.

3.5.4 Multvariable regression analysis (West and East Africa)

Multvariable linear regression was used to assess the relationship between subtype and VRC. However, potential confounders such as participant's sex, age and viral load counts which were significantly higher in West African participants than in East African participants (Table 1) had to be controlled for. Historically, men have been shown to have higher viral loads than women (186), and this higher viral load and sex bears an impact on the region's VRC (150, 187) making it was necessary to decouple all them. There were 196 participants with sequence-VRC pairs from West Africa, 91 (46.4%), 14 (7.1%) and 91 (46.4%) from Cameroon, Nigeria, and Senegal, respectively. Viral Load measurements were only available for participants from Cameroon, and $CD4^+$ count measurements were only available for 48 participants from Cameroon and 84 participants from Senegal. Information in age was only available for 84 participants from Cameroon. In West Africa, the most frequently observed subtype was CRF02_AG which was observed in 139 (~71%) of participants with sequence-VRC pairs. In the univariate analysis, subtypes A1, A1G recombinants, A3 and CRF22_01A1 were all associated with significantly lower VRC when compared to CRF02_AG. Under 25% of participants had complete information on all the following parameters: $CD4^+$ count, viral load, and VRC, rendering the construction of a multivariable or imputation model challenging. In view of country being a significant predictor of VRC, and country-specific missing data patterns, we constructed separate multivariable models for Cameroon and Senegal (but not Nigeria due to the small sample size) to gain insight into the effect of subtype on VRC while adjusting for confounders (Table 3.4). The country-specific multivariate models collectively supported the significant findings in the univariate analysis: A1, A1G recombinants, A3 and CRF22_01A1 had 0.49 ($p < 0.0001$), 0.35 ($p = 0.01$), 0.2 ($p = 0.03$) and 0.4 ($p < 0.0001$) significantly lower mean VRC compared to CRF02_AG.

In East Africa, the most frequently observed subtype, A1, was found in 103 (64%) of participants. 5 (3.1%), 17 (10.6%) and 35 (21.9%) of participants were subtypes A1C, A1D and D, respectively; however, only 114 participants had seqeunce-VRC pairs which were included in the regression models. In the univariate analysis, A1D and D subtypes were associated with significantly higher VRC values, after adjusting country, CD4 count, viral load, age and gender. Using multiple imputation with multivariate regression the dataset was imputed 50 times and estimates were combined on the imputed dataset, yielding revised regression estimates. After multiple imputation A1D had a 0.24 significantly higher mean VRC compared to A1 ($p < 0.001$), while D had a 0.29 significantly higher mean VRC compared to A ($p < 0.001$) after adjusting for country, $CD4^{+}$, viral load, age, and gender (Table 3.5).

Table 3.4:Multivatiable model associating subtype and VRC for West Africa.

	Univariate model		**Multivariate model**			
	West Africa		Cameroon		Senegal	
Observations	166		77		66	
Prob > f	<0.0001		<0.0001		0.022	
r^2	0.290		0.653		0.193	
[a]Subtypes	Co-efficient	*p*-value	Co-efficient	*p*-value	Co-efficient	*p*-value
A1	-0.388	<0.0001	-0.489	<0.0001	-	-
A1G[b]	-0.190	0.013	-0.012	0.854	-0.352	0.010
A3	-0.250	<0.0001	-	-	-0.199	0.032
CRF11_cpx	0.033	0.691	-0.006	0.907	-	-
CRF22_01A1	-0.038	<0.0001	-0.398	<0.0001	-	-
D	0.093	0.187	0.024	0.681	0.137	0.308
G	-0.017	0.775	-0.098	0.223	-0.067	0.768
Meta data						
Age	-	-	-0.001	0.629	-	-
Log VL	-	-	0.033	0.054	-	-
$CD4^+$ count	-	-	-	-	0.000	0.137

[a]Subtype A1 is used as reference, *A1G recombinant is not a CRF.

Table 3.5: Multivariable model associating subtype and VRC for East Africa.

	Univariate model		Multivariate model	
Observations	114		111	
Prob > f	<0.0001		<0.0001	
r^2	0.30		0.34	
[a]Subtypes	Co-efficient	*p*-value	Co-efficient	*p*-value
A1C	0.07	0.5	0.113	0.298
A1D	0.25	<0.0001	0.239	<0.0001
D	0.31	<0.0001	0.292	<0.0001
[b]Country				
Rwanda	-	-	-0.075	0.272
Uganda	-	-	-0.032	0.622
meta data				
Age	-	-	-0.003	0.175
Log VL	-	-	0.012	0.727
$CD4^+$ count	-	-	<0.0001	0.858
Sex	-	-	-0.034	0.481

[a]Subtype A1 is used as reference, [b]Kenya is used as reference

3.5.5 HLA alleles associated with *Gag-Protease* driven replication capacity

Genome wide association studies (GWAS) have demonstrated that the Human Leucocyte Antigen (HLA) class I alleles are the most significant genetic determinant of clinical outcome in HIV-1 infection (188). To investigate the impact of HLA allele expressed on the Gag-protease driven VRC on subtype A1 patient-derived isolates from East Africa, VRC data were grouped according to HLA alleles expressed by the hosts for each loci (Figure 3.11). VRC did not differ significantly overall between the different HLA-A (Figure 3.11A), HLA-B (Figure 3.11B) and HLA-C (Figure 3.11C) alleles. However, an analysis of individual HLA alleles showed that HLA-A*23:01 and HLA-C*07:01 were strongly associated with lower VRC (Student's T test; $p = 0.0014$ and $p = 0.002$) respectively, while HLA-B*07:02 was strongly associated with higher VRC (Student's T test; $p < 0.05$). This analysis could not be done on CRF02_AG isolates from West Africa due to the lack of HLA typing data on the acquired samples.

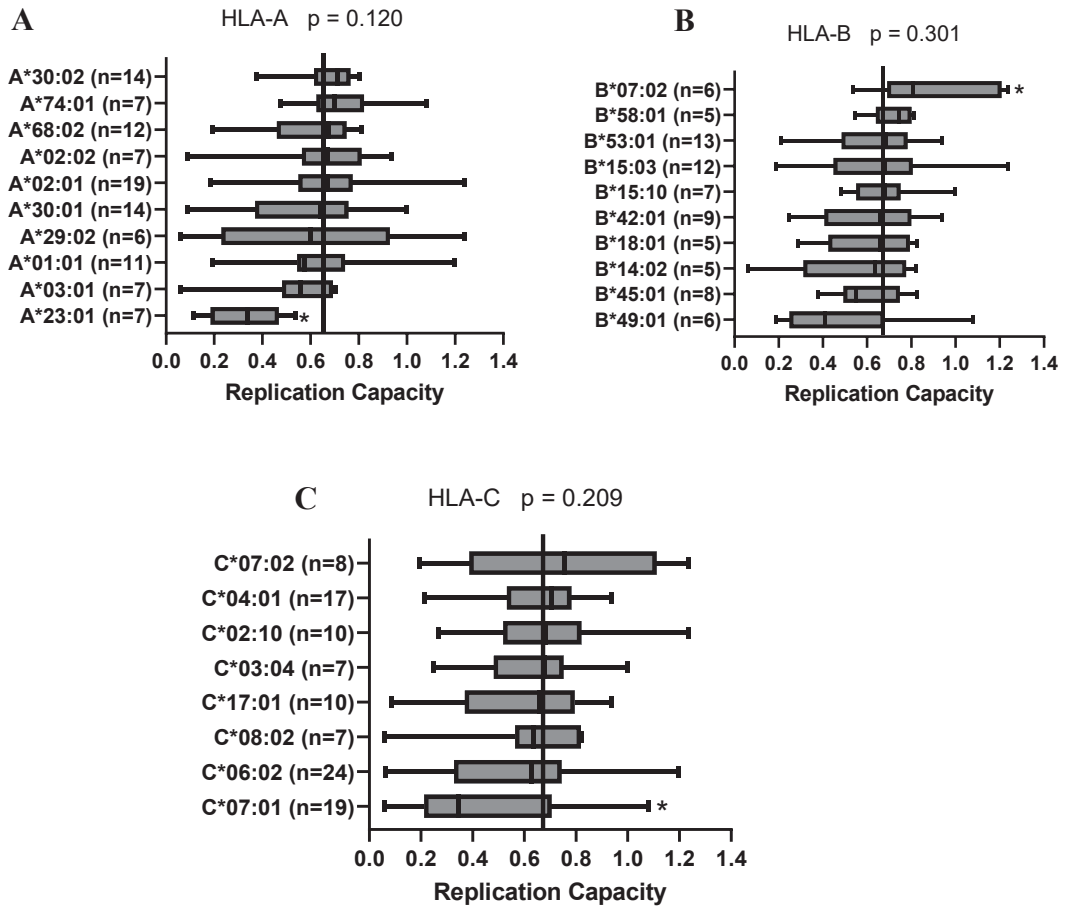


Figure 3.134: VRC association with HLA class I expression

Figure 3.14: Associations between HLA class I allele expression and *Gag-protease* driven replication capacity. *Gag-protease* driven viral replication capacities (VRC) of chronically infected participants from East Africa with subtype A1 infection were grouped according to HLA class I alleles expressed. The box plots display VRC results arranged from lowest median VRC at the bottom and the highest median VRC at the top. Boundaries of the boxes indicate the interquartile ranges, while the whiskers display the maximum and minimum RC values. The continuous vertical line on each graph indicates the median VRC (0.6560) for all subtype A1 viruses. HLA-I alleles with a minimum of n = 5 are shown. Asterisks indicate HLA alleles that are significantly ($p < 0.05$) associated with either higher or lower VRC (Student's t test).

3.5.6 Gag p6 subtype predicts VRC

Depending on the subtype, the Gag p6 is a 50-52 amino-acid protein unique to lentiviruses such as HIV-1. It contains the PTAPP and LYPLASL motifs which interact with the host cell TSG101 and Alix factors respectively to facilitate efficient virion maturation and budding from host cells (189, 190). Recombination patterns in West and East Africa showed a preference for subtypes D, G or J rather than subtype A in the p6 region of Gag, highlighting an important evolutionary trait that may impact VRC. In support of this, a comparative analysis of East African recombinants with A1 and D components within the p6 region showed a significant difference between the groups, where sequences with a subtype D p6 had a higher VRC ($p = 0.0021$; Figure 3.15A). Previous reports (191) and work from our own group (144) support that variation in the LYPLASL motif in *gag* affects VRC. Therefore, to investigate subtype-specific amino acid sequence in the PTAP and LYPLASL motifs of the p6 region of *gag*, consensus sequences of individual subtypes irrespective of region was generated A1 (n=104), A1D (n=16), A3 (n=10), CRF02_AG (n=139), CRF11_cpx (n=5), CRF22_01A1 (n=8), D (n=43) and G (n=15) and aligned with HXB2 subtype B sequence as a reference. While amino acid sequence in the PTAP domain remained conserved irrespective of subtype, there were marked variations in the LYPLASL domain. Highlighted in Figure 3.15B are the sequence variations in Alix motif of p6 by subtype and how these correspond to subtypes of low versus high replication capacity. Further work to interrogate the effect of these subtype specific LYPLASL motifs on VRC is warranted.

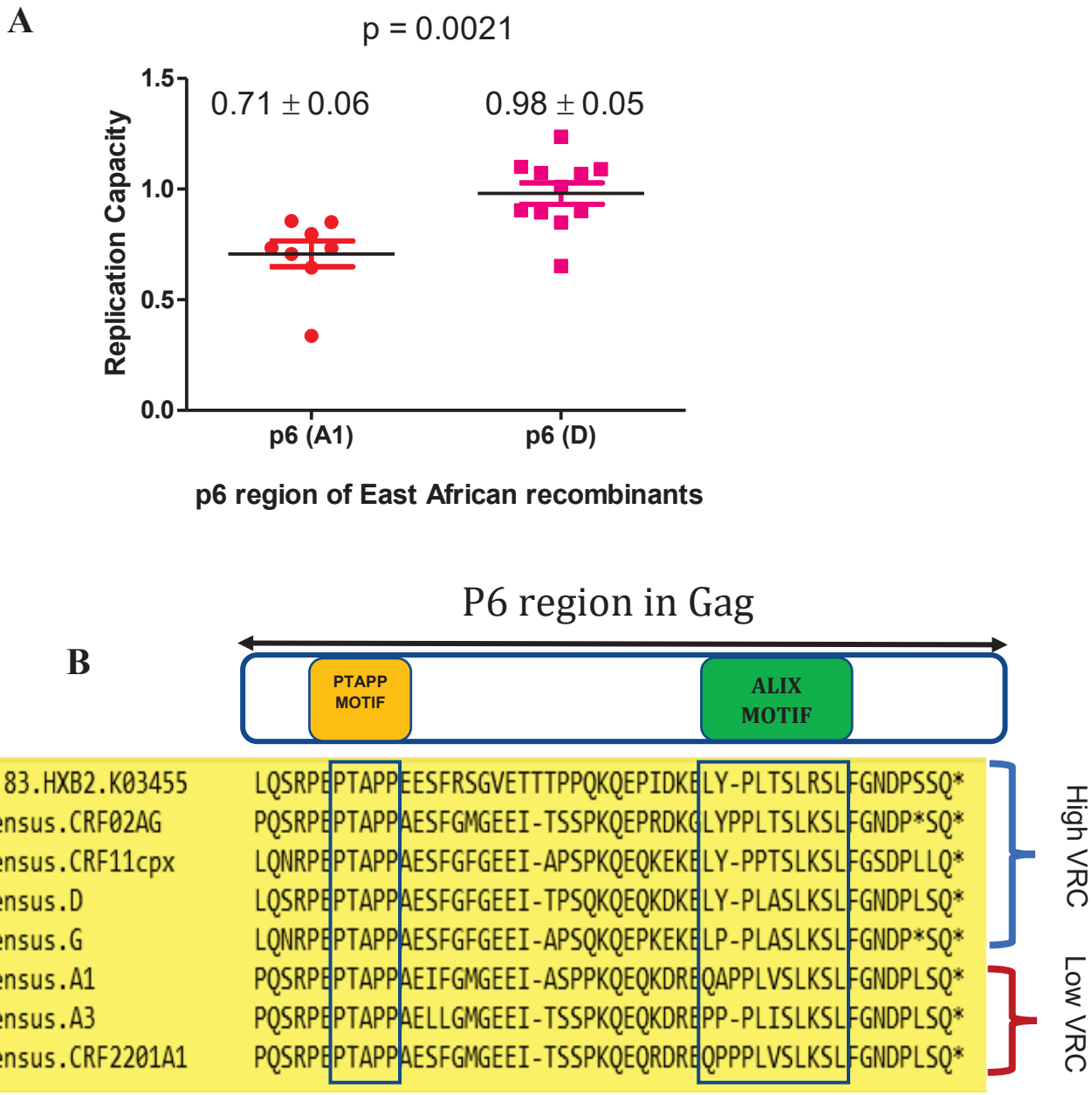


Figure 3.14: Amino Acid variation in Gag p6 region

Figure 3.15: Analysis of the amino acid sequence of the p6 region of Gag. Analysis of the amino acid sequence of the p6 region of Gag. (A) Comparative analysis of VRC of A1D recombinants with A1 and D components respectively within the p6 region of *gag*. **(B)** Consensus amino acid sequence of subtypes A1, A3, CRF02_AG, CRF11_cpx, CRF22_01A1, D, and G show mutations within the Alix motif showing which sequences correspond to subtypes with high VRC or low VRC.

3.5.7 Sequence codon-by-codon analysis

An exploratory codon-by-codon analysis to identify amino acid polymorphic variants associated with differences in VRC was performed for the predominant subtypes in West (CRF02_AG) and East (A1) Africa using methods previously described (143). The analysis identified several amino acid polymorphisms that associated with altered VRC at $p < 0.05$, however, upon correction for multiple codon comparisons none of the associations met the significant threshold of $q < 0.2$ (Tables 3.6 and 3.7). In CRF02_AG sequences, A83V, a previously inferred escape variant (165) in known CTL epitopes in the p17 region of *gag* was the only non-consensus amino acid variant associated with a lower VRC (Table 3.6). This result was consistent with previous reports linking A83V to reduced viral fitness in subtypes B and CRF01_AE (165, 192). In subtype A1 sequences, the only non-consensus polymorphism associated with decreased VRC was P497L, while non-consensus amino acid variants K69Q, S125N, L147I, A158V and R380K, were associated with significantly higher VRC (Table 3.7). However, the consensus amino acids at positions 12 and 28 were associated with increased VRC, which suggests that mutations at these codons, within known CTL epitopes, confer a fitness cost. Consistent with this, polymorphisms at codon 28 have been reported to alter VRC in other subtypes (155, 193). Further work is required to validate these findings since the relatively small number of sequences available for analysis resulted in limited statistical power.

Table 3.6: Codon-by-codon analysis for subtype CRF02_AG

[a]Consensus AA/Codon	[b]AA Signal	Median Replication with aa	Median Replication without aa	No. of patients with aa	No. of patients without aa	*p*-value	*q*-value	A-list epitope
L34	L	0.91	0.95	83	27	0.022	0.525	KW9 (A*2402), HL9, LF11 (A*30)
L34	I	0.95	0.91	27	83	0.022	0.525	
A83	A	0.93	0.73	109	6	0.015	0.525	RY11 (B*0801), SL9 (A*0201, A*0202, A*0205), SY10 (A*0201), LY9 (A*2902, B*4403)
A83	V	0.73	0.93	6	109	0.015	0.525	
S125	S	0.91	0.95	69	40	0.029	0.525	NY9 (B*3501)
A146	N	1.00	0.91	13	101	0.006	0.514	GI8 (B*1302), HL9 (B*1510), QW11 (A*2501), IW9 (B*5701, B*63)
E260	D	0.97	0.91	19	96	0.011	0.525	NY10, PY9 (B*3501), EI8 (B*0801)
E260	E	0.91	0.97	94	21	0.017	0.525	
V267	V	0.91	0.94	58	56	0.032	0.525	EI10(B*0801), RK10 (B*2703), KK10 (B*2705)
V267	I	0.94	0.91	56	58	0.032	0.525	
S332	A	1.04	0.92	5	110	0.005	0.514	DL9 (B*0801)
R335	R	0.91	0.95	73	40	0.025	0.525	DL9 (B*0801)
R335	K	0.95	0.91	40	73	0.025	0.525	

[a]This column represents the consensus amino acid at specified at that locus; [b]This column represents the amino acid variant being reported by the codon analysis; [c]A-list epitope as available from http://www.lanl.com

Table 3.7: Codon-by-codon analysis for subtype A1

Consensus AA/Codon	AA Signal	Median Replication with aa	Median Replication without aa	No. of patients with aa	No. of patients without aa	*p*-value	*q*-value	A-list epitope
K12	K	0.68	0.23	52	8	0.045	0.862	GI9 (B*4002)
K28	K	0.69	0.47	45	13	0.008	0.862	RK9 and RY10 (A*0301)
A67	A	0.63	0.82	57	5	0.023	0.862	-
K69	Q	0.77	0.63	8	52	0.048	0.862	-
S125	N	0.76	0.63	8	52	0.042	0.862	NY9 (B*3501)
L147	I	0.79	0.63	5	55	0.039	0.862	GI8 (B*1302), HL9 (B*1510) QW11 (A*2501), IW9 (B*5701, B*63)
V158	V	0.63	0.75	54	9	0.038	0.862	VF (B*1503)
V158	A	0.75	0.63	9	54	0.038	0.862	
I223	I	0.63	0.73	49	14	0.028	0.862	HA9 (B*3501, B7)
R380	K	0.75	0.63	9	53	0.036	0.862	-
P497	L	0.54	0.68	16	42	0.021	0.862	-

[a]This column represents the consensus amino acid at specified at that locus; [b]This column represents the amino acid variant being reported by the codon analysis; [c]A-list epitope as available from http://www.lanl.com

4 DISCUSSION

A hallmark of the HIV-1 epidemic in sub-Saharan Africa is its heterogeneity in terms of adult prevalence rates and viral subtype diversity. Marked variance in prevalence rates across sub-regions has been widely reported such that southern African countries tend to have the highest adult prevalence rates followed by East African countries, while West and Central African countries having the lowest prevalence rates (50, 51, 194). It is also noteworthy that HIV-1 subtypes are unevenly distributed across the continent, with subtype C being predominant in southern Africa, subtypes A, D, C and their recombinants are common in East Africa and almost all known subtypes have been identified to be present in West-Central Africa (51, 195). The possible contribution of viral genetic and functional characteristics to the uneven distribution of subtypes is unresolved. Differences in transmission efficiency and rate of disease progression have been confirmed in epidemiological and clinical studies, suggesting that viral factors may partially explain the heterogeneity in prevalence and uneven spread of HIV-1 subtypes within the continent (126, 127, 134, 135, 138, 144, 196). In this study, samples from HIV-1 infected drug naïve individuals from 3 West African countries (Cameroon, Nigeria, and Senegal) and 3 East African countries (Kenya, Rwanda, and Uganda) were analyzed in order to perform a comparative analysis of the epidemic in both regions using patient-derived *gag-protease* sequences which have been reported previously to be a strong determinant of VRC. I then analyzed the data according to geographic regions and viral subtypes with a view to gain further insights into how viral characteristics are associated with and possibly contribute to the heterogeneity of the epidemic between and within the regions.

4.1 HIV-1 diversity and geographical distribution

The HIV-1 epidemic in West Africa displayed a higher level of subtype diversity than that of East Africa, which was consistent with previous surveillance data from both regions (61, 63, 171, 174, 197). West Africa has been theorized to be where the cross-species transmission events occurred which gave rise to the group M HIV-1 lineage that is responsible for the global pandemic (3, 10, 15, 16, 20). This suggests that all subtypes expanded out of the region into the different parts of the world through migration, trade and other possible early human interactions (10), however, it is interesting to note that while subtypes A1 and D were common to both West and East Africa, though in significantly higher proportions in East Africa, the expansion of other subtypes such as B, F, G, H, J and K have rarely been reported in East Africa. This suggests that subtype evolution and expansion into different regions could be driven by a combination of viral factors and host genetic factors such as HLA which select for the fittest variants and allow for their subsequent expansion within a specific region (119, 163). It is therefore critical to consider HLA class I expression and other host genetic factors within each region's population to identify any differences that may account for the subtype differences observed within both regions. Furthermore, research efforts into the development of a universal vaccine that can address HIV-1 subtype diversity should be intensified in West Africa where all known subtypes have been previously identified.

4.2 In-vitro VRC was higher in West African Isolates

This data showed that VRC varied widely among subtypes both inter and intra regionally. Overall, isolates from West African displayed a higher *gag-protease* driven VRC than isolates from East Africa. CRF02_AG and A1 were the most prevalent subtypes in West and East Africa respectively, with both subtypes being the main drivers of VRC within their respective regions. CRF02_AG had a significantly higher VRC than A1 (Figure 3.12B). The VRC in both West and East African isolates differed significantly by HIV-1 subtype with a clear hierarchy where A1/A3 had generally lower VRC when compared to pure subtypes D and G or A1 recombinants except for A/C recombinants with a median = 0.71 and CRF22_01A1 (median = 0.46), the latter being a very close genetic relative of A1. My data is thus consistent with previous studies of Gag-protease functional differences among HIV-1 subtypes and recombinants that may also be extended to other HIV-1 proteins (140, 142, 145, 198-201). Enhanced *gag-protease*-driven VRC has been associated with faster rate of $CD4^+$ T cell decline and disease progression (156), and my data is consistent with reported subtype-specific differences in disease progression (126-128). Descriptive statistical analysis of the combined *gag-protease* driven VRC data for West and East African participants gave a median of 0.86. This allowed for the categorization of the VRC values into low or high, -values less than the median were considered low while those higher were categorized as high. HIV-1 surveillance studies have shown that adult prevalence rate is much lower in West Africa than in East Africa. However, the current study shows that overall, West African isolates have a higher *gag-protease* driven VRC than the East African isolates even after controlling for potential confounders such as sex, age of participants, plasma viral load and CD4+ count. This is consistent with the hypothesis that higher replication capacity is associated with lower transmission potential.

There are a couple of plausible mechanisms described in literature by which these observations could be rationalized; these include the theory of evolution of virulence in pathogens, and functional differences driven by *gag-protease* VRC amongst different HIV-1 subtypes. The evolution of virulence theory proposed that intermediary virulence is likely to maximize the fitness of a pathogen as it facilitates a much-required balance between its ongoing replication and its onward transmission within susceptible members of the population which results in a trade-off between virulence and transmission (202). In RNA viruses such as HIV-1, several factors such as transmission bottlenecks, accumulation of deleterious mutations and within host adaptation affect evolution of virulence by driving down its severity (202-205). In time, these impact virulence and in-turn reduce transmission potential of the virus. The data from the current study points towards functional differences in *gag-protease* driven VRC among regional HIV-1 subtypes as another reason for the observed inverse relationship between adult prevalence rate and prevailing subtype transmission potential. Previous work from our group showed a strong correlation between *gag-protease*-mediated VRC and the VRC of whole isolates (144). As highlighted in the introductory chapter, prospective studies have clearly shown that time of infection to AIDS/death varies amongst HIV-1 subtypes (126-132); and others have shown that *in vitro* VRC to be a predictor of *in vivo* VRC or viral fitness, which in turn impacts disease progression in infected patients (24, 119, 148, 149, 206). Putting all these together, it is proposed that patients infected with subtypes with higher VRC progress to disease/death faster than those infected with subtypes lower VRC making it easier for the latter to transmit the virus to other members of the population according to the evolution of virulence theory. This may explain the observation that in East Africa, subtype A1 which was shown to have a lower VRC is more transmissible compared to subtype D and yet the latter has been associated with a faster rate of disease progression (130-132, 202, 207).

On the other hand, in West Africa a recombinant form of A1 has become predominant suggesting that recombination may be a mechanism that both compensates for replication capacity defects while allowing adaptations that facilitate transmission, another factor that may influence the expansion of CRF02_AG in West Africa could be prevalent HLA-I expression that selects for the subtype. A 2005 comparative study of VRC of group M, O and HIV-2 isolates which showed that VRCs of group O and HIV-2 isolates were significantly lower than their group M counterparts and by implication, this limited viral function impaired their transmission potential causing them to be restricted to west-central Africa (146); hence if our hypothesis is accurate these subtypes should be the most prevalent globally. However, these contrasting positions could possibly be explained by significant difference in respective viral-host adaptations between group M, N, O, P and HIV-2 isolates. In summary, group M, N and HIV-2 isolates represent multiple cross-species transmissions into the human population (3, 10, 15, 16, 20, 21), and a consequence of the evolutionary pathway of group M viruses has made them the most adaptable to the human population (146); thereby making subtypes within this clade able to thrive above and beyond group N, O, P and even HIV-2 isolates. Nevertheless, further studies of other genetic determinants of transmission enhancement are required.

It is also noted from the data, that studies of inter-subtype recombinants may be informative. Specifically, that non-A subtype was preferred for the p6 region of Gag. Recombination patterns suggested that the 3' region of the Gag protein is a recombination hotspot particularly for A1 recombinants. The p6 contains two well studied segments, the PTAP and LYPLASL domains which interact with the host cell TSG101 and Alix factors respectively to facilitate efficient virion maturation and budding from host cells (189, 190). Whereas the PTAP motif showed no amino acid variation across subtypes, the LYPLASL motif showed variation across subtypes which corresponded to either low or high VRC. Furthermore, a comparative analysis

of East African recombinants with A1 and D components within the p6 region showed a lower VRC for those with subtype A1 ($p < 0.05$). We and others have shown previously that variation in the Alix budding motif in Gag affects viral replication capacity (144, 191). Further work to interrogate the effect of these subtype-specific amino acid variations within the LYPLASL motifs on VRC is warranted.

4.3 Impact of HLA expression on subtype A1

Isolates from the East African cohort had complete HLA data, and since subtype A1 was the most predominant within the region we decide to evaluate the impact of HLA allele expression on VRC of patients infected with subtype A1 from the region, this was done as a follow-up to similar work done by our group on subtype C from Southern Africa. We identified HLA class I alleles associated with differences in VRC in this study, VRC did not differ significantly across HLA-A, HLA-B and HLA-C alleles. However, HLA-A*23:01 and HLA-C*07:01 were associated with lower replication capacity, suggesting an underappreciated role in immune selection pressure on HIV-1 subtype A1 by certain HLA-A and HLA-C alleles. This finding appears consistent with previous findings of subtype-specific differences in HLA-driven viral evolution that may have consequences for natural and vaccine-mediated immunity (208). Overall, the data highlights the need for further studies to identify mechanisms of immune control or regions of viral vulnerability by HLA alleles common in regions of the world with non-B and C subtypes as this knowledge may be useful for universal virus attenuation-based vaccine strategies.

4.4 Recombination as a mechanism for increased fitness in subtype A1

Recombination analyses of the West African isolates identified several CRFs, consistent with the multiple pure subtypes that have been identified within the region (16). In West Africa, most recombinants comprised of A1 and G genetic fragments and breakpoint analysis suggested that recombination patterns were not random, with a strong preference for subtype G or J over A1 at the 3'-end of most recombinant or mosaic sequences, with a recombination hotspot around the Gag p6 region. A similar trend was noted for the East African isolates where there was also a preference for subtype D over A within the same region in recombinant sequences. Overall, the data suggested that subtype A1 in both West and East Africa has a high propensity for inter-subtype recombination with non-A subtypes at its 3' end particularly from around the Gag p6 region; [96% (146/152) of West and 80% (12/15) of East African recombinants had a break point within the p6 region of *gag*]. It is therefore hypothesized that this could be a mechanism to facilitate the adaptation of HIV-1 subtype A1 into a fitter virus. Indeed, it has been reported that throughout sub-Saharan Africa, pure forms of subtype A1 are in decline, with a concomitant increase in its recombinants (51). The notion that recombination may be a non-random mechanism to enhance viral fitness is backed up by our replication capacity data. Except for A1C, all other A1 recombinants had higher VRC than the median VRC of the total population studied, with A1C having slightly lower VRC than the median. Furthermore, direct comparison of the most prevalent A1 recombinant in West (CRF02_AG) and East (A1D) Africa showed no significant difference in their VRC as opposed to when both are compared with pure subtype A1 which showed a significant difference and therefore attribute this difference to the recombination event within the p6 region of *gag*. We found that inter-subtype recombination within *gag-protease* was non-existent within less common West African CRFs such as CRF01_AE, CRF22_01A1 CRF09_cpx CRF45_cpx due to their very close genetic relatedness to subtype A1 within the *gag-protease* region.

4.5 HIV-1 subtyping

Accurate regional HIV-1 genetic surveillance is critical in directing public health response to accurate diagnosis and confirmation of prevalent subtype(s), its impact on viral transmissibility, virulence, disease progression, drug resistance and very importantly, vaccine development efforts to complement the efficacy of ARTs in reducing viral transmission (57, 63, 197). In the past, the heteroduplex mobility assay (HMA) was commonly used to identify HIV-1 genetic subtypes (63, 209, 210); however, in more recent times, it has been replaced with more sophisticated multiple online bioinformatic tools developed for the same purpose since they are more user friendly. However, each tool comes with its own strengths and limitations, hence the need to explore different approaches to increase confidence that the right subtype is being identified. In this study the REGA and COMET tools were used independently to confirm subtype of each isolate. Both tools accept FASTA format which can be uploaded as a file or pasted into the space created on each tool for sequences. REGA seemed to give the more comprehensive report of HIV-1 subtyping by mapping the nucleotide positions of the query sequence to HXB2 reference sequence, it also provides a pie chart summary of subtypes identified and highlights possible drug resistance mutations in its report output; due to the multiple analyses it conducts, it takes moderately longer time to process each sequence, as opposed to COMET which has a much faster output, but with very limited information other than HIV-1 subtype identified. Both algorithms are limited by the training/reference sequence data set used to identify subtypes, so an expansion of the training/reference sets used is critical for accurate subtype identification, especially with the expansion of new pure subtypes, CRFs and URFs. In many cases both tools could not distinguish between pure subtype A1, CRF01_AE and CRF22_01A1 due to how closely related these lineages are. Other CRFs such as CRF09_cpx and CRF45_cpx were assigned subtype A1 because there was no recombination event within the *gag* region of these subtypes, and so by implication, the initial subtyping

results overestimated the incidence of A1 in West and even East Africa. Phylogenetic analysis which is the generally accepted gold standard for HIV-1 subtyping (211) was used to resolve discrepancies between both tools and final subtype assignment. Phylogenetic analysis has more control with the number of reference sequences that can be used to train the algorithm in subtype assignment process, it is therefore highly recommended that phylogenetic analysis be used to confirm any subtyping exercise. The methods used to arrive at subtype interpretation are consistent with current methods currently being employed (73, 212) giving confidence in the subtyping results obtained which bears significant impact on the subsequent downstream analysis of the sequences generated, and equally relevant within clinical settings where it is imperative to consider HIV-1 subtype when designing and implementing new preventative and therapeutic strategies such as vaccines (174, 213).

4.6 Study limitations

Limitations of the current study include missing clinical data such as CD4+ T cell counts and HLA-I genotypes, particularly for the West African samples, thus limiting the robustness of statistical analysis. East African samples were preselected for subtypes A, D and their recombinants based on prior pol genotyping and therefore not allowing for comprehensive inter-subtype comparison of the strains that constitute the East Africa epidemic. It should also be noted that all replication assays were performed in the subtype B pNL4-3Δ*gag-protease* backbone, however, we have previously demonstrated that viral replication data generated from this assay is generally reflective of whole virus isolates in HIV-1 subtype C. The study compared VRC of HIV-1 subtype C isolated from patients and the corresponding NL4-3 recombinant viruses encoding Gag-protease derived from the same patients. The results showed a significant positive correlation between both groups (144, 154).

A higher sample size would also have benefited the robustness of the codon-by-codon analysis to identify amino acids associated with altered VRC. Another limitation of the study was the singular use of GXR cells to perform the replication capacity analysis. The study would have benefitted from the use of multiple cell lines to see if differences in cell lines could impact the viral replication capacity. It is likely that the cell line or cell type used for the replication assay could influence result outcomes. Nevertheless, the advantage of using the GXR cell line for the assay is that it is optimised for high throughput, and it has an indicator system that allows analysis of replication capacity using a flow cytometry-based assay. Moreover, it shows good correlation of replication capacity of viruses when compared with assays performed in peripheral blood mononuclear cells (144).

4.7 Future Work and recommendations

My project focused on *gag-protease* as a virological factor that drives VRC as well as HIV-1 viral diversity in sub-Saharan Africa, with a minor emphasis on immunological factors that may drive escape mutations. Immunological factors in combination with viral genetic factors highlighted within this thesis ultimately determine the variance in adult prevalence and HIV-1 subtype diversity observed in sub-Saharan Africa.

Expansion of this study will strongly benefit from investigating immunological factors that also contribute to the epidemiological picture of both regions studied. An exploratory transcriptome wide association study to profile upregulated and downregulated host genes that determine immune responses to infection by different HIV-1 subtypes could be investigated. Data from this work could be critical in answering the questions as to why certain subtypes irrespective of their VRC are more prevalent in certain locations over others, why there is a strong compartmentalization of HIV-1 subtypes between different regions. It is plausible that infection with viruses with high VRC could cause aberrant gene expression profiles that leads

to elevated immune activation leading to faster disease progression independent of viral load. The identification of host molecules that are differentially expressed in infected cells may lead to novel interventions to counteract the effects of viruses with high VRC. The study will therefore be performing an exploratory transcriptome-wide analyses to identify host factors and genetic pathways that may distinguish between uninfected cells, and cells infected with viruses with high VRC and those with low VRC. This study could provide important insights on intrinsic differences between subtypes to explain epidemiological spread and disease progression rates and may help to advance rational vaccine design approaches for these subtypes. The approach could be by way of ligation mediated RNA sequencing (LM-RNA-Seq) which is a modified, but less costly RNA-Seq protocol as described by Hou et al., 2015 (214). cDNA libraries will be constructed, mRNA is isolated from purified 100 ng total RNA using oligo-dT beads (NEB). Based on data generated from VRC assays, selection of the 20 high VRC viruses and 20 low VRC viruses will be made and put through the transcriptome gene expression experiments in cells infected with these as well as uninfected cells. Isolated mRNA is fragmented in reverse transcription buffer with heat and then reverse-transcribed with SmartScribe reverse transcriptase (Clontech) using a random hexamer oligo. Following reverse transcription, RNA is removed by RNaseA and RNAaseH treatment. A partial illumine 59 adaptors is then ligated to the single stranded cDNA using T4 RNA ligase 1 (NEB) and incubated overnight at 22°C. Following purification, ligated cDNA is amplified by 18 cycles of PCR using oligos that contain full illumine adaptors. The cDNA library is thereafter sequenced and mapped to its source genes.

Due to the possible limitation associated with using only pNL4-3Δ*gag-protease* backbone to generate chimeric viruses for my study, it will be useful to consider generating full length HIV-1 genomes for a smaller number of patient derived sequences. Full-length genome primers could be designed and optimized for subtypes with high VRC from West Africa such as CRF02_AG and CRF11_cpx, and those with low VRC such as CRF01_AE and CRF22_01A1, while full length sequences can be generated for subtypes A1 with low VRC and subtypes A1D and D from East Africa. VRC assay will be conducted using the protocol outlined in Chapter 3, and the data generated can be compared to the VRC data generated from my study to see if there are similar trends, or if other genetic loci are bearing significant impact on overall VRC. This approach is particularly important to address the disproportionate poor growth of subtype A1 isolates in culture which could be because of poor compatibility with the NL4-3 backbone.

Further investigation is also warranted to critically review the hypothesis that viruses with lower VRC are more successfully transmitted within the population studied than those with a higher VRC as an evolutionary mechanism to ensure sustained transmission within the population. It is possible that viral variants with lower replication capacity are preferentially transmitted because they induce a less prominent innate immune response (such as type I interferons) upon exposure to cells in the genital tract, thus facilitating the establishment of infection, as opposed to a more robust innate immune response that may lead to viral clearance. Our group has also previously proposed that viruses with lower replication capacity may be associated with a longer infected cell half-life, which could potentially then favour cell to cell transmission of these viruses compared to viruses with higher replicative capacity that may quickly kill carrier cells and thus lessen the probability of transmission (144). Studies to better understand the effect and kinetics of transmission of viruses with high versus low replication capacity particularly in cells within the genital tract that are the portal of entry in most

heterosexual transmission of HIV in sub-Saharan Africa are warranted. Supernatants from infected cell cultures of different subtypes can be collected and measured for IFNα to via flow-cytometry to subtype-specific induction of immune responses which could address the hypothesis raised in my study.

Variance in viral characteristics such as host adaptation and viral function amongst the different HIV-1 groups can be evaluated by methods such as growth competition assay which have the advantage of directly comparing the growth and virulence characteristics of isolates exposed to the same cells in culture to see which viruses expand over one another or have other cytopathic effects (24, 148). Viral adaptation and transmission efficiency could further be established by engineering cell-lines with specific receptor on cells to see which isolates will still be able to replicate despite the limited receptors available on the cell lines. For example, T-cells lines with R5 and X4 receptors, cells with either R5 of X4 receptors and cells with neither of these receptors as a negative control, as a positive control cells with both receptors but receptor blocking agents introduced into the culture. Furthermore, enzyme kinetic studies of orthologs within different isolates could also be performed to evaluate their respective functional capacity (215); these are some proposed methods of evaluating viral fitness and host adaptability that could further validate results from the present study and test the hypothesis generated as well. Another means of measuring transmission efficiency would be to collect culture supernatants of infected cells and quantify IFNα expression via flow-cytometry, the immune response data collected can then be matched to the HIV-1 strain that establishes infection which will give an overview of how well it is able to establish infection in the presence or absence of adequate immune responses. This can be further matched with genetic regulatory pathway data generated from the exploratory transcriptome wide association study. It is possible that viral variants with lower replication capacity are preferentially transmitted

because they induce a less prominent innate immune response (such as type I interferons) upon exposure to cells in the genital tract, thus facilitating the establishment of infection, as opposed to a more robust innate immune response that may lead to viral clearance. Our group has also previously proposed that viruses with lower replication capacity may be associated with a longer infected cell half-life, which could potentially then favours cell to cell transmission of these viruses compared to viruses with higher replicative capacity that may quickly kill carrier cells and thus lessen the probability of transmission (144). Studies to better understand the effect and kinetics of transmission of viruses with high versus low replication capacity particularly in cells within the genital tract that are the portal of entry in most heterosexual transmission of HIV in sub-Saharan Africa are warranted.

Finally, though the adult prevalence rate in West Africa is lower relative to other parts of sub-Saharan Africa, the population affected is still significantly high and ART coverage programs are relatively low due to poor execution unlike in eastern and southern Africa. A renewed focus on HIV-1 research work in terms of VRC and HLA class-I expression within the region is urgently needed, being ground zero for the cross-species transmissions and global spread of the global HIV-1 epidemic, concentrated efforts within the region could yield to faster development of a vaccine that can account for the genetic diversity of the HIV-1.

4.8 Conclusion

Overall, in this study we show that HIV-1 viruses from West Africa displayed higher VRC than those from East Africa, consistent with the hypothesis that higher prevalence is associated with lower replication capacity of circulating strains. Subtype-specific differences in replication capacity in agreement with previous studies were noted, and consistent with reported differences in the rate of clinical disease progression. Our study identified inter-subtype recombination pattern as a driver of VRC differences and identified specific amino acid that may alter VRC. Validation of these amino acids in additional studies is required. Finally, our study identified HLA-A class I class to have a differential impact on HIV-1 replication capacity for HIV-1 subtype A1, the predominant subtype in East Africa which suggest the need for further studies to identify regions of viral vulnerability for HIV-1 vaccine design strategies.

REFERENCES

1. Levy JA. Pathogenesis of human immunodeficiency virus infection. Microbiol Rev. 1993;57(1):183-289.
2. Chin'ombe N, Ruhanya V. HIV/AIDS vaccines for Africa: scientific opportunities, challenges and strategies. Pan Afr Med J. 2015;20:386.
3. Van Heuverswyn Fran & Peeters M. The origins of HIV and implications for the global epidemic. Curr Infect Dis Rep. 2007;9(4):9.
4. King SR. Hiv - Virology and Mechanisms of Disease. Ann Emerg Med. 1994;24(3):443-9.
5. Barouch DH. Challenges in the development of an HIV-1 vaccine. Nature. 2008;455(7213):613-9.
6. Hemelaar J. The origin and diversity of the HIV-1 pandemic. Trends Mol Med. 2012;18(3):182-92.
7. Morison L. The global epidemiology of HIV/AIDS. Br Med Bull. 2001;58:7-18.
8. Al-Jabri AA. Mechanisms of Host Resistance Against HIV Infection and Progression to AIDS. Sultan Qaboos University Medical Jorunal. 2007;7(2):14.
9. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science. 1983;220(4599):868.
10. Peeters M, Jung M, Ayouba A. The origin and molecular epidemiology of HIV. Expert Rev Anti Infect Ther. 2013;11(9):885-96.
11. Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, et al. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science. 1983;220(4599):865.
12. Levy JA, Hoffman AD, Kramer SM, Landis JA, Shimabukuro JM, Oshiro LS. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. Science. 1984;225(4664):840.
13. Vahlne A. A historical reflection on the discovery of human retroviruses. Retrovirology. 2009;6:40.
14. Coffin J HA, Levy JA, Montagnier L, Oroszlan S, Teich N, Temin H, Toyoshima K, Varmus H, Vogt P, et al. What to call the AIDS virus? Nature. 1986;321(6065):10-.

15. Feng Gao EB, David L. Robertson, Yalu Chen, Cynthia M. Rodenburg, Scott F. Michael, Larry B. Cummins, Larry O. Arthure, Martine Peeters, George M. Shaw, Paul M. Sharp & Beatrice H. Hahn. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. Nature. 1999;397.
16. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med. 2011;1(1):a006841.
17. Fehervari Z. Origin Sotry. Nat Immunol. 2018.
18. Paraskevis D, Lemey P, Salemi M, Suchard M, Van De Peer Y, Vandamme AM. Analysis of the evolutionary relationships of HIV-1 and SIVcpz sequences using bayesian inference: implications for the origin of HIV-1. Mol Biol Evol. 2003;20(12):1986-96.
19. Esbjornsson J, Mild M, Mansson F, Norrgren H, Medstrand P. HIV-1 molecular epidemiology in Guinea-Bissau, West Africa: origin, demography and migrations. PLoS One. 2011;6(2):e17025.
20. Sharp PM, Hahn BH. The evolution of HIV-1 and the origin of AIDS. Philos Trans R Soc Lond B Biol Sci. 2010;365(1552):2487-94.
21. Corbet S, Müller-Trutwin MC, Versmisse P, Delarue S, Ayouba A, Lewis J, et al. env sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. Journal of virology. 2000;74(1):529-34.
22. Heeney JL, Dalgleish AG, Weiss RA. Origins of HIV and the evolution of resistance to AIDS. Science. 2006;313(5786):462-6.
23. Gagneux P, Wills C, Gerloff U, Tautz D, Morin PA, Boesch C, et al. Mitochondrial sequences show diverse evolutionary histories of African hominoids. Proc Natl Acad Sci U S A. 1999;96(9):5077-82.
24. Quiñones-Mateu ME, Arts EJ. HIV-1 Fitness: Implications for Drug Resistance, Disease Progression, and Global Epidemic Evolution. 2001.
25. Maldarelli F. HIV-1 Fitness and Replication Capacity: What Are They and Can They Help in Patient Management? Curr Infect Dis Rep. 2003;5(1):77-84.
26. Demma LJ, Logsdon JM, Jr., Vanderford TH, Feinberg MB, Staprans SI. SIVsm quasispecies adaptation to a new simian host. PLoS Pathog. 2005;1(1):e3.
27. van Opijnen T, de Ronde A, Boerlijst MC, Berkhout B. Adaptation of HIV-1 depends on the host-cell environment. PLoS One. 2007;2(3):e271.
28. Vanderford TH, Demma LJ, Feinberg MB, Staprans SI, Logsdon JM, Jr. Adaptation of a diverse simian immunodeficiency virus population to a new host is revealed through a

systematic approach to identify amino acid sites under selection. Mol Biol Evol. 2007;24(3):660-9.

29. Carlson JM, Le AQ, Shahid A, Brumme ZL. HIV-1 adaptation to HLA: a window into virus-host immune interactions. Trends Microbiol. 2015;23(4):212-24.
30. Sheehy AM, Gaddis, N. C., Choi, J. D., & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature. 2002;418(6898):5.
31. Matthew Stremlau CMO, Michel J. Perron, Michael Kiessling, Patrick Autissier & Joseph Sodroski. The cytoplasmic body component TRIM5-alpha restricts HIV-1 infection in Old World monkeys. Nature 2004;427 (848):6.
32. Neil SJ, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature. 2008;451(7177):425-30.
33. Simon V, Ho DD, Abdool Karim Q. HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. The Lancet. 2006;368(9534):489-504.
34. Smith K, Powers KA, Kashuba AD, Cohen MS. HIV-1 treatment as prevention: the good, the bad, and the challenges. Curr Opin HIV AIDS. 2011;6(4):315-25.
35. WHO. PROGRESS REPORT 2016 PREVENT HIV, TEST AND TREAT ALL. WHO. 2016.
36. Rufu A, Chitimbire VTS, Nzou C, Timire C, Owiti P, Harries AD, et al. Implementation of the 'Test and Treat' policy for newly diagnosed people living with HIV in Zimbabwe in 2017. Public Health Action. 2018;8(3):145-50.
37. Sierra-Aragon S, Walter H. Targets for inhibition of HIV replication: entry, enzyme action, release and maturation. Intervirology. 2012;55(2):84-97.
38. ORGANISATION WH. HIV/AIDS TREATMENT AND CARE CLINICAL PROTOCOLS FOR THE WHO EUROPEAN REGION. WHO 2007.
39. Mujugira A, Baeten JM, Donnell D, Ndase P, Mugo NR, Barnes L, et al. Characteristics of HIV-1 serodiscordant couples enrolled in a clinical trial of antiretroviral pre-exposure prophylaxis for HIV-1 prevention. PLoS One. 2011;6(10):e25828.
40. Heffron R, Ngure K, Mugo N, Celum C, Kurth A, Curran K, et al. Willingness of Kenyan HIV-1 serodiscordant couples to use antiretroviral-based HIV-1 prevention strategies. J Acquir Immune Defic Syndr. 2012;61(1):116-9.
41. Murnane PM, Celum C, Mugo N, Campbell JD, Donnell D, Bukusi E, et al. Efficacy of preexposure prophylaxis for HIV-1 prevention among high-risk heterosexuals: subgroup analyses from a randomized trial. AIDS. 2013;27(13):2155-60.

42. Heffron R, Mugo N, Were E, Kiarie J, Bukusi EA, Mujugira A, et al. Preexposure prophylaxis is efficacious for HIV-1 prevention among women using depot medroxyprogesterone acetate for contraception. AIDS. 2014;28(18):2771-6.
43. Stefano Bertozzi NSP, Jeny Wegbreit, Lisa M. DeMaria, Becca Feldman, Helene Gayle, Julian Gold, Robert Grant, and Michael T. Isbell. HIV/AIDS Prevention and Treatment. . Disease Control Priorities in Developing Countries 2nd edition Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2006 Chapter 18. 2006.
44. Sekaly RP. The failed HIV Merck vaccine study: a step back or a launching point for future vaccine development? J Exp Med. 2008;205(1):7-12.
45. Rubens M, Ramamoorthy V, Saxena A, Shehadeh N, Appunni S. HIV Vaccine: Recent Advances, Current Roadblocks, and Future Directions. J Immunol Res. 2015;2015:560347.
46. Choi E, Michalski CJ, Choo SH, Kim GN, Banasikowska E, Lee S, et al. First Phase I human clinical trial of a killed whole-HIV-1 vaccine: demonstration of its safety and enhancement of anti-HIV antibody responses. Retrovirology. 2016;13(1):82.
47. UNAIDS. UNAIDS DATA 2019. 2019.
48. Morison L. The global epidemiology of HIV/AIDS. Br Med Bull. 2001;58:7-18.
49. Kilmarx PH. Global epidemiology of HIV. Current opinion in HIV and AIDS. 2009;4(4):240-6.
50. Shao Y, Williamson C. The HIV-1 epidemic: low- to middle-income countries. Cold Spring Harb Perspect Med. 2012;2(3):a007187.
51. Hemelaar J, Elangovan R, Yun J, Dickson-Tetteh L, Fleminger I, Kirtley S, et al. Global and regional molecular epidemiology of HIV-1, 1990-2015: a systematic review, global survey, and trend analysis. Lancet Infect Dis. 2019;19(2):143-55.
52. Ippolito G, Rezza G, Girardi E, Malkovsky M, Cairns JS. HIV infections: the global epidemiology and goals for vaccine research. Mol Med. 2000;6(2):69-85.
53. Kharsany AB, Karim QA. HIV Infection and AIDS in Sub-Saharan Africa: Current Status, Challenges and Opportunities. Open AIDS J. 2016;10:34-48.
54. Fettig J, Swaminathan M, Murrill CS, Kaplan JE. Global epidemiology of HIV. Infect Dis Clin North Am. 2014;28(3):323-37.

55. Vandormael A, Akullian A, Siedner M, de Oliveira T, Barnighausen T, Tanser F. Declines in HIV incidence among men and women in a South African population-based cohort. Nat Commun. 2019;10(1):5482.
56. McCutchan FE. Global epidemiology of HIV. J Med Virol. 2006;78:S7-S12.
57. Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. The challenge of HIV-1 subtype diversity. N Engl J Med. 2008;358(15):1590-602.
58. Pai NP, Shivkumar S, Cajas JM. Does Genetic Diversity of HIV-1 Non-B Subtypes Differentially Impact Disease Progression in Treatment-Naive HIV-1-Infected Individuals? A Systematic Review of Evidence: 1996-2010. Jaids-Journal of Acquired Immune Deficiency Syndromes. 2012;59(4):382-8.
59. Abdool Karim SS, Abdool Karim Q, Gouws E, Baxter C. Global epidemiology of HIV-AIDS. Infect Dis Clin North Am. 2007;21(1):1-17, vii.
60. Sharp MPaPM. Genetic Diversity of HIV-1: the moving target. AIDS. 2000;14:12.
61. Thomson MM, Pérez-Álvarez L, Nájera R. Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. The Lancet Infectious Diseases. 2002;2(8):461-71.
62. Gupta RM, Sahni AK, Jena J, Nema SK. Genomic Diversity of Human Immunodeficiency Viruses. Medical Journal Armed Forces India. 2005;61(3):267-70.
63. Janssens W, Buve A, Nkengasong JN. The puzzle of HIV-1 subtypes in Africa. AIDS. 1997;11(6):705-12.
64. Carr JK. Viral diversity as a challenge to HIV-1 vaccine development. Curr Opin HIV AIDS. 2006;1(4):294-300.
65. Peeters M, Toure-Kane C, Nkengasong JN. Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials. Aids. 2003;17(18):2547-60.
66. Spira S. Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. Journal of Antimicrobial Chemotherapy. 2003;51(2):229-40.
67. Korber B, Hraber P, Wagh K, Hahn BH. Polyvalent vaccine approaches to combat HIV-1 diversity. Immunol Rev. 2017;275(1):230-44.
68. D.L. Robertson JPA, J.A. Bradac, J.K. Carr, B. Foley, R.K. Funkhouser, F. Gao, B.H. Hahn, C. Kuiken, G.H. Learn, T. Leitner, F. McCutchan, S. Osmanov, M. Peeters, D. Pieniazek, M.L. Kalish, M. Salminen, P. Sharp, S. Wolinsky, and B. Korber. HIV-1 Nomenclature Proposal. Science 2000;288(5463).

69. Myers G, MacInnes K Fau - Korber, B., Korber, B. The emergence of simian/human immunodeficiency viruses. AIDS Res Hum Retroviruses. 1992;8(3):14.
70. Gurtler LG, Hauser Ph Fau - Eberle J, Eberle J Fau - von Brunn A, von Brunn A Fau - Knapp S, Knapp S Fau - Zekeng L, Zekeng L Fau - Tsague JM, et al. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. J Virol. 1994;68(3):6.
71. Simon F, Mauclere P Fau - Roques P, Roques P Fau - Loussert-Ajaka I, Loussert-Ajaka I Fau - Muller-Trutwin MC, Muller-Trutwin Mc Fau - Saragosti S, Saragosti S Fau - Georges-Courbot MC, et al. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat Med. 1998;4(9):6.
72. Vallari A, Holzmayer V, Harris B, Yamaguchi J, Ngansop C, Makamche F, et al. Confirmation of putative HIV-1 group P in Cameroon. J Virol. 2011;85(3):1403-7.
73. Vidal N, Koyalta D, Richard V, Lechiche C, Ndinaromtan T, Djimasngar A, et al. High genetic diversity of HIV-1 strains in Chad, West Central Africa. J Acquir Immune Defic Syndr. 2003;33(2):239-46.
74. Tongo M, Dorfman JR, Martin DP. High Degree of HIV-1 Group M (HIV-1M) Genetic Diversity within Circulating Recombinant Forms: Insight into the Early Events of HIV-1M Evolution. J Virol. 2015;90(5):2221-9.
75. Yamaguchi J, McArthur C, Vallari A, Sthreshley L, Cloherty GA, Berg MG, et al. Complete genome sequence of CG-0018a-01 establishes HIV-1 subtype L. J Acquir Immune Defic Syndr. 2019.
76. Foley BT, Leitner T, Paraskevis D, Peeters M. Primate immunodeficiency virus classification and nomenclature: Review. Infect Genet Evol. 2016;46:150-8.
77. Desire N, Cerutti L, Le Hingrat Q, Perrier M, Emler S, Calvez V, et al. Characterization update of HIV-1 M subtypes diversity and proposal for subtypes A and D sub-subtypes reclassification. Retrovirology. 2018;15(1):80.
78. E. A. Berger RWD, E.-M. Fenyö, B. T. M. Korber, D. R. Littman, J. P. Moore, Q. J. Sattentau, H. Schuitemaker, J. Sodroski, and R. A. Weiss. A new classification for HIV-1. Nature 1998;391:1.
79. Joris Hemelaar EG, Peter D. Ghys and Saladin Osmanov. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS. 2006;20:11.
80. Martine Peeters MJ, and Ahidjo Ayouba. The origin and molecular epidemiology of HIV. Expert Review of Anti Infect Therapy. 2013;11(9).

81. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. Aids. 2006;20(16):W13-23.
82. D.L. Robertson JPA, J.A. Bradac, J.K. Carr, B. Foley, R.K. Funkhouser, F. Gao, B.H. Hahn, C. Kuiken, G.H. Learn, T. Leitner, F. McCutchan, S. Osmanov, M. Peeters, D. Pieniazek, M.L. Kalish, M. Salminen, P. Sharp, S. Wolinsky, and B. Korber. HIV-1 Nomenclature Proposal. Science. 2000;288(5463).
83. Briggs JA, Riches JD, Glass B, Bartonova V, Zanetti G, Krausslich HG. Structure and assembly of immature HIV. Proc Natl Acad Sci U S A. 2009;106(27):11090-5.
84. Christian Hoffmann JKR, Hamburg, Bonn et al.,. HIV BOOK. 2015.
85. Sarafianos SG, Marchand B, Das K, Himmel DM, Parniak MA, Hughes SH, et al. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. J Mol Biol. 2009;385(3):693-713.
86. Andrea Rubbert GBaMO. Pathogenesis of HIV-1 Infection. HIVBookcom. 2011.
87. Engelman A, Cherepanov P. The structural biology of HIV-1: mechanistic and therapeutic insights. Nat Rev Microbiol. 2012;10(4):279-90.
88. Rainer Seitz P-E. Human Immunodeficiency Virus (HIV). Transfusion medicine and hemotherapy : offizielles Organ der Deutschen Gesellschaft fur Transfusionsmedizin und Immunhamatologie. 2016;43(3):203-22.
89. Young ADFaJAT. HIV-1: Fifteen Proteins and an RNA. Annu Rev Biochem. 1998;67:26.
90. Swanson CM, Malim MH. SnapShot: HIV-1 proteins. Cell. 2008;133(4):742, e1.
91. Rappaport MKaJ. HIV-1 Accessory Protein Vpr: Relevance in the pathogenesis of HIV and potential for therapeutic intervention. Retrovirology. 2011;8(25):20.
92. Monique R. Ferguson P, Daniel R. Rojo, PhD, Jana J. von Lindern, BS, William A. O'Brien, MD, MS. HIV-1 replication cycle. Clin Lab Med. 2002;22.
93. Fanales-Belasio E, Raimondo M, Suligoi B, Butto S. HIV virology and pathogenetic mechanisms of infection: a brief overview. Ann Ist Super Sanita. 2010;46(1):5-14.
94. Nielsen MH, Pedersen FS, Kjems J. Molecular strategies to inhibit HIV-1 replication. Retrovirology. 2005;2:10.
95. Steven G. Deeks, Julie Overbaugh , Andrew Phillips and Susan Buchbinder. HIV infection. Nature Reviews Disease Primers. 2015;1.
96. Göttlinger HG. HIV-1 Gag: a Molecular Machine Driving Viral Particle Assembly and Release. 2005:26.

97. Chen EC, Parent LJ. Chapter 8 - Role of Host Factors in the Subcellular Trafficking of Gag Proteins and Genomic RNA Leading to Virion Assembly. In: Parent LJ, editor. Retrovirus-Cell Interactions: Academic Press; 2018. p. 273-315.
98. Virgin MTHaHW. Pathogenesis of Viral Infection. Fields Virology 6th Edition: Chapter 10. 2013;1.
99. Nancy R. Calles DE, DeLouis Terlonge. Pathophysiology of the human immunodeficiency virus. Baylor International Pdiatric AIDS Initiative. 2010:8.
100. Swanstrom R, Coffin J. HIV-1 pathogenesis: the virus. Cold Spring Harb Perspect Med. 2012;2(12):a007443.
101. Koup DRKaR. HIV-1: Pathogenesis, Clinical Manifestations, and Treatment. Fields Virology 6th Edition: Chapter 50. 2013;2.
102. Ganapathy S, Vedam V, Rajeev V, Arunachalam R. Autoimmune Disorders – Immunopathogenesis and Potential Therapies. Journal of Young Pharmacists. 2017;9(1):14-22.
103. Chaplin DD. Overview of the immune response. J Allergy Clin Immunol. 2010;125(2 Suppl 2):S3-23.
104. Marshall JS, Warrington R, Watson W, Kim HL. An introduction to immunology and immunopathology. Allergy Asthma Clin Immunol. 2018;14(Suppl 2):49.
105. Smith NC, Rise ML, Christian SL. A Comparison of the Innate and Adaptive Immune Systems in Cartilaginous Fish, Ray-Finned Fish, and Lobe-Finned Fish. Front Immunol. 2019;10:2292.
106. Guha D, Ayyavoo V. Innate immune evasion strategies by human immunodeficiency virus type 1. ISRN AIDS. 2013;2013:954806.
107. Altfeld M, Gale M, Jr. Innate immunity against HIV-1 infection. Nat Immunol. 2015;16(6):554-62.
108. Jakobsen MR, Olagnier D, Hiscott J. Innate immune sensing of HIV-1 infection. Curr Opin HIV AIDS. 2015;10(2):96-102.
109. Alberts B JA, Lewis J, et al. Innate Immunity. Molecular Biology of the Cell 4th edition New York: Garland Science; 2002. 2002.
110. Shyam Kottilil T-WC, Susan Moir, Shuying Liu, Mary McLaughlin, Claire W. Hallahan, Frank Maldarelli, Lawrence Corey, and Anthony S. Fauci Innate Immunity in Human Immunodeficiency Virus Infection: Effect of Viremia on Natural Killer Cell Function. The Journal of Infectious Diseases. 2003;187.

111. Alter G, Martin MP, Teigen N, Carr WH, Suscovich TJ, Schneidewind A, et al. Differential natural killer cell-mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. J Exp Med. 2007;204(12):3027-36.
112. Sarah Howie RRaTH. Innate Immune System Damage in Human Immunodeficiency Virus Type 1 Infection Implications for Acquired Immunity and Vaccine Design. MERICAN JOURNAL OF RESPIRATORY AND CRITICAL CARE MEDICINE. 2000;162.
113. Joanna Mikulak CDVaDM. HIV-1-induced inflammation shapes innate immunity and induces adaptive traits in NK cells. Nature Immunology. 2020;21.
114. Bonilla FA, Oettgen HC. Adaptive immunity. J Allergy Clin Immunol. 2010;125(2 Suppl 2):S33-40.
115. Blankson JN, Siliciano RF. Elite suppression of HIV-1 replication. Immunity. 2008;29(6):845-7.
116. Eller MA, Blom KG, Gonzalez VD, Eller LA, Naluyima P, Laeyendecker O, et al. Innate and adaptive immune responses both contribute to pathological CD4 T cell activation in HIV-1 infected Ugandans. PLoS One. 2011;6(4):e18779.
117. Mohan T, Bhatnagar S, Gupta DL, Rao DN. Current understanding of HIV-1 and T-cell adaptive immunity: progress to date. Microb Pathog. 2014;73:60-9.
118. da Silva J, Hughes AL. Conservation of cytotoxic T lymphocyte (CTL) epitopes as a host strategy to constrain parasite adaptation: evidence from the nef gene of human immunodeficiency virus 1 (HIV-1). Mol Biol Evol. 1998;15(10):1259-68.
119. Goulder PJ, Watkins DI. HIV and SIV CTL escape: implications for vaccine design. Nat Rev Immunol. 2004;4(8):630-40.
120. Brumme ZL, John M, Carlson JM, Brumme CJ, Chan D, Brockman MA, et al. HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. PLoS One. 2009;4(8):e6687.
121. Ndhlovu ZM, Kamya P, Mewalal N, Kloverpris HN, Nkosi T, Pretorius K, et al. Magnitude and Kinetics of CD8+ T Cell Activation during Hyperacute HIV Infection Impact Viral Set Point. Immunity. 2015;43(3):591-604.
122. Makvandi-Nejad S. Human Immunodeficiency Virus (HIV). British Society for Immunology. 2020.
123. Mlotshwa M, Riou C, Chopera D, de Assis Rosa D, Ntale R, Treunicht F, et al. Fluidity of HIV-1-specific T-cell responses during acute and early subtype C HIV-1 infection and associations with early disease progression. Journal of virology. 2010;84(22):12018-29.

124. Muenchhoff M, Prendergast AJ, Goulder PJ. Immunity to HIV in Early Life. Front Immunol. 2014;5:391.
125. Alaeus A, Lidman K, Björkman A, Giesecke J, Albert J. Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A-D. AIDS. 1999;13(8):901-7.
126. Kanki PJ, Hamel DJ, Sankale JL, Hsieh C, Thior I, Barin F, et al. Human immunodeficiency virus type 1 subtypes differ in disease progression. J Infect Dis. 1999;179(1):68-73.
127. Kaleebu P, Ross A, Morgan D, Yirrell D, Oram J, Rutebemberwa A, et al. Relationship between HIV-1 Env subtypes A and D and disease progression in a rural Ugandan cohort. Aids. 2001;15(3):293-9.
128. Venner CM, Nankya I, Kyeyune F, Demers K, Kwok C, Chen PL, et al. Infecting HIV-1 Subtype Predicts Disease Progression in Women of Sub-Saharan Africa. EBioMedicine. 2016;13:305-14.
129. Ashwin Vasal BR, Ellen Hertzmark, Beth Chaplin, Gernard Msamanga, Max Essex, Wafaie Fawzi and David Hunter. Different Rates of Disease Progression of HIV Type 1 Infection in Tanzania Based on Infecting Subtype. Clinical Infectious Diseases. 2006;42.
130. Baeten JM, Chohan B, Lavreys L, Chohan V, McClelland RS, Certain L, et al. HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. J Infect Dis. 2007;195(8):1177-80.
131. Kiwanuka N, Laeyendecker O, Robb M, Kigozi G, Arroyo M, McCutchan F, et al. Effect of human immunodeficiency virus Type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. J Infect Dis. 2008;197(5):707-13.
132. Ssemwanga D, Nsubuga RN, Mayanja BN, Lyagoba F, Magambo B, Yirrell D, et al. Effect of HIV-1 subtypes on disease progression in rural Uganda: a prospective clinical cohort study. PLoS One. 2013;8(8):e71768.
133. Geretti AM. HIV-1 subtypes: epidemiology and significance for HIV management. Curr Opin Infect Dis. 2006;19:8.
134. Blackard JT, Renjifo B, Fawzi W, Hertzmark E, Msamanga G, Mwakagile D, et al. HIV-1 LTR subtype and perinatal transmission. Virology. 2001;287(2):261-5.
135. G.N. Odaibo DOO, L. Heyndrickx, K. Vereecken, K. Houwer and W. Jassens. Mother-To-Child Transmission of Different HIV-1 Subtypes Among ARV Naive Iinfected Pregnant Woment in Nigeria. Rev Inst Med trop S Paulo. 2006;48(2).

136. Buonaguro L, Tornesello ML, Buonaguro FM. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. J Virol. 2007;81(19):10209-19.
137. M. Rios JF, P. Jaramillo, V. Paredes, J.L. Sanchez, V.A. Laguna-Torres, J.K. Carr, and E. Ramirez. Molecular Epidemiology of HIV Type 1 in Chile: Differential Geographic and Transmission Route Distribution of B and F Subtypes. AIDS Research and Human Retroviruses. 2005;21(10):5.
138. Kiwanuka N, Laeyendecker O, Quinn TC, Wawer MJ, Shepherd J, Robb M, et al. HIV-1 subtypes and differences in heterosexual HIV transmission among HIV-discordant couples in Rakai, Uganda. AIDS. 2009;23(18):2479-84.
139. van Opijnen T, Jeeninga RE, Boerlijst MC, Pollakis GP, Zetterberg V, Salminen M, et al. Human immunodeficiency virus type 1 subtypes have a distinct long terminal repeat that determines the replication rate in a host-cell-specific manner. J Virol. 2004;78(7):3675-83.
140. Rienk E. Jeeninga MH, Mercedes Armand-Ugon, Michel De Baar, Koen Verhoef, and Ben Berkhout. Functional Differences between the Long Terminal Repeat Transcriptional Promoters of Human Immunodeficiency Virus Type 1 Subtypes A through G. JOURNAL OF VIROLOGY. 2000;74(8):3740–51.
141. Turk G, Carobene M, Monczor A, Rubio AE, Gomez-Carrillo M, Salomon H. Higher transactivation activity associated with LTR and Tat elements from HIV-1 BF intersubtype recombinant variants. Retrovirology. 2006;3:14.
142. Rebecca M. Lynch TS, S. Gnanakaran, and Cynthia A. Derdeyn. Appreciating HIV Type 1 Diversity: Subtype Differences in Env. AIDS RESEARCH AND HUMAN RETROVIRUSES. 2008;25(3):237-48.
143. Doty B. A. Ojwach DM, Tarylee Reddy, Vladimir Novitsky, Zabrina L. Brumme, Mark A. Brockman, Thumbi Ndung'u, Jaclyn K. Mann. Pol-Driven Replicative Capacity Impacts Disease Progression in HIV-1 Subtype C Infection. Journal of Virology. 2018;92(19):eoo811.
144. Kiguoya MW, Mann JK, Chopera D, Gounder K, Lee GQ, Hunt PW, et al. Subtype-Specific Differences in Gag-Protease-Driven Replication Capacity Are Consistent with Intersubtype Differences in HIV-1 Disease Progression. J Virol. 2017;91(13).
145. Jaclyn K Mann HB, Xiaomei T Kuang, Anh Q Le, Chanson J Brumme, Philip Mwimanzi, Saleha Omarjee, Eric Martin, Guinevere Q Lee, Bemuluyigza Baraki, Ryan Danroth,

Rosemary McCloskey, Conrad Muzoora, David R Bangsberg, Peter W Hunt, Philip JR Goulder, Bruce D Walker, P Richard Harrigan, Jeff N Martin, Thumbi Ndung'u, Mark A Brockman, and Zabrina L Brumme,. Ability of HIV-1 Nef to downregulate CD4 and HLA class I differs among viral subtypes. Retrovirology. 2013;10(100):1-16.

146. Arien KK, Abraha A, Quinones-Mateu ME, Kestens L, Vanham G, Arts EJ. The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. J Virol. 2005;79(14):8979-90.
147. Haubrich RH. Resistance and replication capacity assays: clinical utility and interpretation. Topics in HIV medicine : a publication of the International AIDS Society, USA. 2004;12(2):52-6.
148. Quinones-Mateu ME, Arts EJ. Fitness of drug resistant HIV-1: methodology and clinical implications. Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy. 2002;5(6):224-33.
149. Quinones-Mateu ME, Arts EJ. Virus fitness: concept, quantification, and application to HIV population dynamics. Curr Top Microbiol Immunol. 2006;299:83-140.
150. Goetz MB, Leduc R, Wyman N, Kostman JR, Labriola AM, Lie Y, et al. HIV replication capacity is an independent predictor of disease progression in persons with untreated chronic HIV infection. J Acquir Immune Defic Syndr. 2010;53(4):472-9.
151. Miura T, Brumme ZL, Brockman MA, Rosato P, Sela J, Brumme CJ, et al. Impaired replication capacity of acute/early viruses in persons who become HIV controllers. J Virol. 2010;84(15):7581-91.
152. Prado JG, Prendergast A, Thobakgale C, Molina C, Tudor-Williams G, Ndung'u T, et al. Replicative capacity of human immunodeficiency virus type 1 transmitted from mother to child is associated with pediatric disease progression rate. J Virol. 2010;84(1):492-502.
153. Philippe Selhorst SDI, Natasha Samsunder,Carina Combrinck,Salim S. Abdool Karim,Melissa-Rose Abrahams, Nigel Garrett,Nonkululeko Ndabambi,Miguel Lacerda,Quarraisha Abdool Karim,Carolyn Williamsona,. Replication Capacity of Viruses from Acute Infection Drives HIV-1 Disease Progression. Journal of Virology. 2017;91(8).
154. Wright JK, Brumme ZL, Carlson JM, Heckerman D, Kadie CM, Brumme CJ, et al. Gag-protease-mediated replication capacity in HIV-1 subtype C chronic infection: associations with HLA type and clinical parameters. J Virol. 2010;84(20):10820-31.

155. Brockman MA, Brumme ZL, Brumme CJ, Miura T, Sela J, Rosato PC, et al. Early selection in Gag by protective HLA alleles contributes to reduced HIV-1 replication capacity that may be largely compensated for in chronic infection. J Virol. 2010;84(22):11937-49.
156. Wright JK, Novitsky V, Brockman MA, Brumme ZL, Brumme CJ, Carlson JM, et al. Influence of Gag-protease-mediated replication capacity on disease progression in individuals recently infected with HIV-1 subtype C. Journal of virology. 2011;85(8):3996-4006.
157. Prince JL, Claiborne DT, Carlson JM, Schaefer M, Yu T, Lahki S, et al. Role of transmitted Gag CTL polymorphisms in defining replicative capacity and early HIV-1 pathogenesis. PLoS Pathog. 2012;8(11):e1003041.
158. Nomura S, Hosoya N, Brumme ZL, Brockman MA, Kikuchi T, Koga M, et al. Significant reductions in Gag-protease-mediated HIV-1 replication capacity during the course of the epidemic in Japan. J Virol. 2013;87(3):1465-76.
159. Saina MC, Bi X, Lihana R, Lwembe R, Ishizaki A, Panikulam A, et al. Comparison of HIV-1 nef and gag Variations and Host HLA Characteristics as Determinants of Disease Progression among HIV-1 Vertically Infected Kenyan Children. PLoS One. 2015;10(8):e0137140.
160. Sakai K, Chikata T, Brumme ZL, Brumme CJ, Gatanaga H, Oka S, et al. Lack of a significant impact of Gag-Protease-mediated HIV-1 replication capacity on clinical parameters in treatment-naive Japanese individuals. Retrovirology. 2015;12:98.
161. Philip J. R. Goulder CB, Yanhua Tang, Cecile Tremblay, Robert A. Colbert, Marylyn M. Addo, Eric S. Rosenberg, Thi Nguyen, Rachel Allen, Alicja Trocha, Marcus Altfeld, Suqin He, Michael Buncek, Robert Funkhouser, Stephen I. Pelton, Sandra K. Burchett, Kenneth McIntosh, Bette T. M. Korber & Bruce D. Walker. Evolution and transmission of stable CTL escapemutations in HIV infection. Nature. 2001;412:334-8.
162. Yang OO, Sarkis PT, Ali A, Harlow JD, Brander C, Kalams SA, et al. Determinant of HIV-1 mutational escape from cytotoxic T lymphocytes. J Exp Med. 2003;197(10):1365-75.
163. Smith SM. HIV CTL escape: at what cost? Retrovirology. 2004;1(8):8.
164. Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, Kadie C, et al. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. PLoS Comput Biol. 2008;4(11):e1000225.
165. Boutwell CL, Carlson JM, Lin TH, Seese A, Power KA, Peng J, et al. Frequent and variable cytotoxic-T-lymphocyte escape-associated fitness costs in the human immunodeficiency virus type 1 subtype B Gag proteins. J Virol. 2013;87(7):3952-65.

166. Roberts HE, Hurst J, Robinson N, Brown H, Flanagan P, Vass L, et al. Structured observations reveal slow HIV-1 CTL escape. PLoS Genet. 2015;11(2):e1004914.
167. Chopera DR, Ntale R, Ndabambi N, Garrett N, Gray CM, Matten D, et al. Early evolution of human leucocyte antigen-associated escape mutations in variable Gag proteins predicts CD4+ decline in HIV-1 subtype C-infected women. AIDS. 2017;31(2):191-7.
168. Goepfert PA, Lumm W, Farmer P, Matthews P, Prendergast A, Carlson JM, et al. Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. J Exp Med. 2008;205(5):1009-17.
169. Chopera DR, Wright JK, Brockman MA, Brumme ZL. Immune-mediated attenuation of HIV-1. Future Virol. 2011;6(8):917-28.
170. Payne R, Muenchhoff M, Mann J, Roberts HE, Matthews P, Adland E, et al. Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence. Proc Natl Acad Sci U S A. 2014;111(50):E5393-400.
171. Tongo M, Martin DP, Zembe L, Mpoudi-Ngole E, Williamson C, Burgers WA. Characterization of HIV-1 gag and nef in Cameroon: further evidence of extreme diversity at the origin of the HIV-1 group M epidemic. Virol J. 2013;10:29.
172. Ajoge HO, Gordon ML, de Oliveira T, Green TN, Ibrahim S, Shittu OS, et al. Genetic characteristics, coreceptor usage potential and evolution of Nigerian HIV-1 subtype G and CRF02_AG isolates. PLoS One. 2011;6(3):e17865.
173. Hamel DJ, Sankale JL, Eisen G, Meloni ST, Mullins C, Gueye-Ndiaye A, et al. Twenty years of prospective molecular epidemiology in Senegal: changes in HIV diversity. AIDS Res Hum Retroviruses. 2007;23(10):1189-96.
174. Price MA, Rida W, Kilembe W, Karita E, Inambao M, Ruzagira E, et al. Control of the HIV-1 Load Varies by Viral Subtype in a Large Cohort of African Adults With Incident HIV-1 Infection. J Infect Dis. 2019;220(3):432-41.
175. Tözsér J, Bláha I Fau - Copeland TD, Copeland Td Fau - Wondrak EM, Wondrak Em Fau - Oroszlan S, Oroszlan S. Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in Gag and Gag-Pol polyproteins. (0014-5793 (Print)).
176. Pineda-Pena AC, Faria NR, Imbrechts S, Libin P, Abecasis AB, Deforche K, et al. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. Infect Genet Evol. 2013;19:337-48.

177. Struck D, Lawyer G, Ternes AM, Schmit JC, Bercoff DP. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic Acids Res. 2014;42(18):e144.
178. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307-21.
179. Schultz AK, Zhang M, Bulla I, Leitner T, Korber B, Morgenstern B, et al. jpHMM: improving the reliability of recombination prediction in HIV-1. Nucleic Acids Res. 2009;37(Web Server issue):W647-51.
180. KAVITA S. LOLE RCB, RAMESH S. PARANJAPE, DEEPAK GADKARI, SMITA S. KULKARNI, NICOLE G. NOVAK, ROXANN INGERSOLL,HAYNES W. SHEPPARD, AND STUART C. RAY. Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination. Journal of Virology 1999;73(1):9.
181. Miura T, Brockman MA, Brumme ZL, Brumme CJ, Pereyra F, Trocha A, et al. HLA-associated alterations in replication capacity of chimeric NL4-3 viruses carrying gag-protease from elite controllers of human immunodeficiency virus type 1. J Virol. 2009;83(1):140-9.
182. Brockman MA, Tanzi GO, Walker BD, Allen TM. Use of a novel GFP reporter cell line to examine replication capacity of CXCR4- and CCR5-tropic HIV-1 by flow cytometry. J Virol Methods. 2006;131(2):134-42.
183. Brockman MA, Schneidewind A, Lahaie M, Schmidt A, Miura T, Desouza I, et al. Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A. Journal of virology. 2007;81(22):12608-18.
184. Lavinia Fabeni GB, Joseph Fokam, Ada Bertoli, Claudia Alteri, Caterina Gori, Federica Forbici, Desiré Takou, Alessandra Vergori, Mauro Zaccarelli, Gaetano Maffongelli, Vanni Borghi, Alessandra Latini, Alfredo Pennica, Claudio Maria Mastroianni, Francesco Montella, Cristina Mussini, Massimo Andreoni, Andrea Antinori, Carlo Federico Perno, Maria Mercedes Santoro, and The Resistance Study Group. Comparative Evaluation of Subtyping Tools for Surveillance of Newly Emerging HIV-1 Strains. Journal of Clinical Microbiology. 2017;55(9).

185. Mann JK, Chopera D, Omarjee S, Kuang XT, Le AQ, Anmole G, et al. Nef-mediated down-regulation of CD4 and HLA class I in HIV-1 subtype C infection: association with disease progression and influence of immune pressure. Virology. 2014;468-470:214-25.
186. Monica Gandhi PB, Paolo Miotti, Thomas C Quinn, Fulvia Veronese, Ruth M Greenblatt. Does Patient Sex Affect Human Immunodeficiency Virus Levels? Clinical Infectious Diseases. 2001;35:313-22.
187. Colleen F. Kelley M, Jason D. Barbour, PhD, and Frederick M. Hecht, MD. The Relation Between Symptoms, Viral Load, and Viral Load Set Point in Primary HIV Infection. Journal of Acquired Immune Deficiency Syndromes. 2007;45(4):445-8.
188. Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science. 2010;330(6010):1551-7.
189. Solbak SM, Reksten TR, Hahn F, Wray V, Henklein P, Henklein P, et al. HIV-1 p6 - a structured to flexible multifunctional membrane-interacting protein. Biochim Biophys Acta. 2013;1828(2):816-23.
190. Barbara Mu¨ller TP, and Hans-Georg Krausslich. The Late-Domain-Containing Protein p6 Is the Predominant Phosphoprotein of Human Immunodeficiency Virus Type 1 Particles. JOURNAL OF VIROLOGY. 2002;76(3):10.
191. Aralaguppe SG, Winner D, Singh K, Sarafianos SG, Quinones-Mateu ME, Sonnerborg A, et al. Increased replication capacity following evolution of PYxE insertion in Gag-p6 is associated with enhanced virulence in HIV-1 subtype C from East Africa. J Med Virol. 2017;89(1):106-11.
192. Takahashi N, Matsuoka S, Thi Minh TT, Ba HP, Naruse TK, Kimura A, et al. Human leukocyte antigen-associated gag and nef polymorphisms in HIV-1 subtype A/E-infected individuals in Vietnam. Microbes Infect. 2019;21(2):113-8.
193. Naidoo VL, Mann JK, Noble C, Adland E, Carlson JM, Thomas J, et al. Mother-to-Child HIV Transmission Bottleneck Selects for Consensus Virus with Lower Gag-Protease-Driven Replication Capacity. J Virol. 2017;91(17).
194. Abdool Karim Q. The global HIV epidemic: current status and challenges. Current HIV/AIDS reports. 2013;10(2):111-2.
195. Joris Hemelaar EG, Peter D. Ghys and Saladin Osmanov. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS. 2006;20:11.

196. Faria NR, Vidal N, Lourenco J, Raghwani J, Sigaloff KCE, Tatem AJ, et al. Distinct rates and patterns of spread of the major HIV-1 subtypes in Central and East Africa. PLoS Pathog. 2019;15(12):e1007976.
197. Aleaeus A. Significance of HIV-1 Genetic Subtypes. Scand J Infect. 2000;32:9.
198. Jin SW, Mwimanzi FM, Mann JK, Bwana MB, Lee GQ, Brumme CJ, et al. Variation in HIV-1 Nef function within and among viral subtypes reveals genetically separable antagonism of SERINC3 and SERINC5. PLoS Pathog. 2020;16(9):e1008813.
199. Sergey Iordanskiy MW, Yanjun Feng, Charles Wood. Subtype-associated differences in HIV-1 reverse transcription affect the viral replication. Retrovirology. 2010;7(85).
200. Binka M, Ooms M, Steward M, Simon V. The activity spectrum of Vif from multiple HIV-1 subtypes against APOBEC3G, APOBEC3F, and APOBEC3H. J Virol. 2012;86(1):49-59.
201. Nagata S, Imai J, Makino G, Tomita M, Kanai A. Evolutionary Analysis of HIV-1 Pol Proteins Reveals Representative Residues for Viral Subtype Differentiation. Front Microbiol. 2017;8:2151.
202. Blanquart F, Grabowski MK, Herbeck J, Nalugoda F, Serwadda D, Eller MA, et al. A transmission-virulence evolutionary trade-off explains attenuation of HIV-1 in Uganda. Elife. 2016;5.
203. Carl T. Bergstrom PM, and Leslie A. Real. Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. Proc Natl Acad Sci U S A. 1999;96:5095–100.
204. Ester La´ zaro CE, Juan Pe´ rez-Mercader, Susanna C. Manrubia, and Esteban Domingo. Resistance of virus to extinction on bottleneck passages: Study of a decaying and fluctuating pattern of fitness loss. Proc Natl Acad Sci U S A. 2003;100(9):10830–5.
205. Lythgoe KA, Gardner A, Pybus OG, Grove J. Short-Sighted Virus Evolution and a Germline Hypothesis for Chronic Viral Infections. Trends Microbiol. 2017;25(5):336-48.
206. Kouyos RD, von Wyl V, Hinkley T, Petropoulos CJ, Haddad M, Whitcomb JM, et al. Assessing predicted HIV-1 replicative capacity in a clinical setting. PLoS Pathog. 2011;7(11):e1002321.
207. Vasan A, Renjifo B, Hertzmark E, Chaplin B, Msamanga G, Essex M, et al. Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2006;42(6):843-52.

208. McKinnon LR, Capina R, Peters H, Mendoza M, Kimani J, Wachihi C, et al. Clade-specific evolution mediated by HLA-B*57/5801 in human immunodeficiency virus type 1 clade A1 p24. J Virol. 2009;83(23):12636-42.
209. Celine Montavon TT-K, Florian Liegeois, Eitel Mpoudi, Anke/Bourgeois, Laurence Vergne, Jean-Luc Perret, Annie Boumah, Eric Saman, Souleymane Mboup, Eric/Belaporte, and Martine Peeters Most env and gag Subtype A HIV-1 Viruses Circulating in West and West Central Africa Are Similar to the Prototype AG Recombinant Virus IBNG. Journal of Acquired Immune Deficiency Syndromes. 2000;23:12.
210. Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, Ilunga W, et al. 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. J Virol. 2000;74(22):10498-507.
211. Lavinia Fabeni CG, Giulia Berno, Mauro Zaccarelli, Alfredo Pennica, Cristina Mussini, Federica Forbici, Joseph Fokam, Desiré Takou, Gaetano Maffongelli, Maria Mercedes Santoro, Claudio Maria Mastroianni, Massimo Andreoni, Ada Bertoli, Vanni Borghi, Andrea Antinori, The Resistance Study Group, . Comparative Evaluation of Subtyping Tools for Surveillance of Newly Emerging HIV-1 Strains. Journal of Clinical Microbiology. 2017;55(9):11.
212. Mir D, Jung M, Delatorre E, Vidal N, Peeters M, Bello G. Phylodynamics of the major HIV-1 CRF02_AG African lineages and its global dissemination. Infect Genet Evol. 2016;46:190-9.
213. Wainberg MA. Editorial Commentary: Clinical Significance of HIV Subtype Variability in Regard to Treatment Outcome. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2015;60(10):1550-1.
214. Hou Z, Jiang P, Swanson SA, Elwell AL, Nguyen BK, Bolin JM, et al. A cost-effective RNA sequencing protocol for large-scale gene expression studies. Sci Rep. 2015;5:9570.
215. John J. Holland JCdLT, David K. Clarke, and Elizabeth Duarte. Quantitation of Relative Fitness and Great Adaptability of Clonal Populations of RNA Viruses. Journal of Virology. 1991;65(6):2960-7.

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

13 June 2016

Mr OO Farinre (216073965)
HIV Pathogenesis Programme
School of Laboratory Medicine and Medical Sciences
farinre@gmail.com

Dear Mr Farinre

Study Title: Gag-protease-driven viral fitness among HIV-1 subtypes: Implications for disease progression, epidemic spread and vaccine development.
Degree: PhD
BREC REF NO: BE313/16

The Biomedical Research Ethics Committee has considered and noted your application received on 19 May 2016.

The conditions have been met and the study is given **full ethics approval.**

This approval is valid for one year from **13 June 2016.** To ensure uninterrupted approval of this study beyond the approval expiry date, an application for recertification must be submitted to BREC on the appropriate BREC form 2-3 months before the expiry date.

Any amendments to this study, unless urgently required to ensure safety of participants, must be approved by BREC prior to implementation.

Your acceptance of this approval denotes your compliance with South African National Research Ethics Guidelines (2015), South African National Good Clinical Practice Guidelines (2006) (if applicable) and with UKZN BREC ethics requirements as contained in the UKZN BREC Terms of Reference and Standard Operating Procedures, all available at http://research.ukzn.ac.za/Research-Ethics/Biomedical-Research-Ethics.aspx.

BREC is registered with the South African National Health Research Ethics Council (REC-290408-009). BREC has US Office for Human Research Protections (OHRP) Federal-wide Assurance (FWA 678).

The sub-committee's decision will be **RATIFIED** by a full Committee at its meeting taking place on 12 **July 2016.**

We wish you well with this study. We would appreciate receiving copies of all publications arising out of this study.

Yours sincerely

[redacted]

Professor J Tsoka-Gwegweni
Chair: Biomedical Research Ethics Committee

cc supervisor: ndurse@ukzn.ac.za
cc postgraduate administrator: dudhrajhp@ukzn.ac.za
cc HPP Administrator: leslie@ukzn.ac.za