FUNCTIONAL AND CLINICAL CONSEQUENCES OF IMMUNE-DRIVEN SEQUENCE VARIATION OF GAG-PROTEASE IN HIV-1 SUBTYPES A, C, D AND RECOMBINANTS

SUBMITTED BY: MARION WANGUI KIGUOYA

SUPERVISED BY: PROFESSOR THUMBI NDUNG'U
DR JACLYN MANN

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Virology in the Nelson R. Mandela School of Medicine, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa.

2016

The experimental work described in this thesis was carried out in the HIV Pathogenesis Programme, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, from April 2012 to December 2015 under the supervision of Professor Thumbi Ndung'u and co-supervised by Dr. Jaclyn Mann.

This project represents original work done by the author and has not otherwise been submitted in any form for any degree or diploma to any other university. Where use has made of the work others, it has been duly acknowledged in the text.

Signed …….. Date…14th April 2016……..

Marion Wangui Kiguoya

Signed……Date……14 April 2016……

Prof. Thumbi Ndung'u (Supervisor)

Signed……Date 14th April 2016……

Dr. Jaclyn Mann (Co- supervisor)

DECLARATION

I, Marion Wangui Kiguoya, declare that

(i) The research reported in this thesis, except where otherwise indicated, is my original work this thesis has not been submitted for any degree or examination at any other university.

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

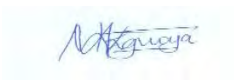
(iv) This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

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

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

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

(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 sections.



Signed …….. Date……..

ETHICAL APPROVAL

Ethical Approval was given from the Biomedical Research Ethics Committee of the Nelson R. Mandela School of Medicine, University of KwaZulu-Natal (ref: BE 181.12) and the Ethics Review Committee of the Kenyatta National Hospital Ethics Review Board (KNH/ERC/IR/121, protocol P211/09/2006). All patients gave written informed consent for this study.

CONFERENCE PRESENTATIONS AND MANUSCRIPTS

Conference presentations

1. Kiguoya, M.W., Mann, J.K., Omarjee, S., Kiogora, F.M., Ball, B.T., Kimani, J., T Ndung'u. Functional and Clinical Consequences of Immune-driven Sequence Variation in Gag-protease in HIV-1 subtypes A and D. University of Nairobi HIV/AIDS Collaborative Centre Annual Meeting, 26th-30th January 2015, Nairobi, Kenya.

2. Kiguoya, M.W., Mann, J.K., Omarjee, S., Kiogora, F.M., Ball, B.T., Kimani, J., T Ndung'u. HIV subtypes A, C, D and inter-subtype recombinants are associated with Gag-protease driven fitness differences. College of Health Sciences Research Symposium, 10th -11th September 2015, University of KwaZulu-Natal, Durban. South Africa.

3. Kiguoya, M.W., Mann, J.K., Omarjee, S., Kiogora, F.M., Gounder, K., McKinnon, L., Ball, B.T., Kimani, J., T Ndung'u. Gag-protease-mediated Replication Capacity Differs According to HIV-1 Subtype. Conference on Retroviruses and opportunistic infections (CROI), 22nd-25th February 2016, Boston, Massachusetts, Abstract number 16-145.

Manuscripts (in preparation)

Marion Kiguoya, Jaclyn K. Mann, Denis Chopera, Zabrina L. Brumme, Mark A. Brockman, Thumbi Ndung'u. Subtype-Specific Differences in Gag-Protease-Driven Replication Capacity are Consistent with Inter-Subtype Differences in HIV-1 Disease Progression.

DEDICATION

I dedicate this thesis to my late father, Mr. John Kiguoya Karaka, whom together with my mother, Mrs. Margaret Waringa Karaka, believed in me and sacrificed a lot to give me the best education. I hope this achievement will fulfil the dream you had for me from my childhood. I appreciate the priceless support and continuous care and thank God for you always.

ACKNOWLEDGEMENTS

First and above all, I praise God, the almighty for providing me this opportunity and granting me the capability to proceed successfully, without his grace and blessings I would not have made it.

I would like to thank my supervisor Professor Thumbi Ndung'u and co-supervisor Dr. Jaclyn Mann, for their patience, excellent guidance and support throughout my research project. My heartfelt gratitude to my mentors Professors Blake Ball and Joshua Kimani for their immeasurable support, patience and encouragement.

My sincere appreciation to the Majengo cohort women for the samples, staff and fellow students, my officemates both at Kenya Aids Control Project and HIV Pathogenesis Programme, friends Tarryn, Uche, Keshni, Lara, Sitembile, Miriam, Saleha, Sharlot, Irungu, Nyakio and Ethel.

CANSSA grant, HHMI grant, KACP and HPP for the financial assistance.

Sincere gratitude to my special friends; Nancy, Salma, Peninah, Margaret, Juliana, Sophy, Goreth, William, Daniel, Akeem and Jane, who have been my pillars throughout the study, thank you for the friendships above all, the support.

My family for their love support and encouragement, my parents Mr. and Mrs. Njau, my brothers Edward, Jeff, Eric, Thuo and Theuri, my sisters Silvia, Lucy, Eunice, Diana and Rose. My deepest gratitude to my mother and friend Mrs. Margaret Karaka, for taking care of our child throughout my study period, forever indebted for the support and love.

My spiritual mentors and friends Reverend Emmanuel Nkoitiko, Mrs. Nancy Gikemi, Mrs Munsanje Muleya, Kinloch family, Mrs Jane Mutua and Mrs Alice Kihonge for their immense prayers and words of encouragement. Lastly but not least my loving husband Nicholas Njau Ngomi and daughter Isabelle Waringa Njau; for the sacrifice, immeasurable love and support, visits and patience. I must express my profound gratitude you two have been my greatest inspiration. Thank you all for believing in me.

ABSTRACT

Introduction

HIV-1 has become one of the world's most serious health and development challenges. HIV/AIDS is a pandemic and represents a major development crisis for the African continent in particular, which is the worst affected region in the world. Despite highly active antiretroviral therapy (HAART) intervention which has dramatically reduced HIV-1-associated morbidity and mortality, HIV incidence globally remains unacceptably high, especially in low income countries. There is therefore an urgent need to develop an effective vaccine that will halt the spread of HIV. A major obstacle to HIV vaccine development is the genetic diversity of the virus world-wide, with various subtypes prevalent in different regions of the world. A successful HIV vaccine will need to protect against these diverse subtypes, especially the non-B strains that predominate in high-burdened countries. In addition to the implications of this diversity for vaccine development, subtype-specific differences in disease progression rate have been reported; however the mechanisms underlying this heterogeneity are not fully understood. An interplay of viral and host factors contribute to HIV control and may thus be responsible for differences in disease progression. In particular, HIV-1 infection induces a strong host cellular immune response, mediated by the highly polymorphic human leukocyte antigen (HLA) class I molecules. Cytotoxic T lymphocyte responses play a critical role in controlling viral replication and may drive viral escape, which in turn may impact on viral replicative fitness and ultimately influence the clinical outcome of infection. In particular, fitness costly escape mutations in the Gag protein have been associated with clinical benefit and there is evidence that Gag-driven viral fitness is a significant determinant of the rate of disease progression. The overall aim of this study was to investigate whether there were subtype-specific replication capacity differences in recombinant viruses possessing patient-derived *gag-protease* genes from subtypes A, C, D and inter-subtype recombinants which predominate in the HIV-1 epidemic in East Africa. We investigated whether

viral replication capacity was associated with markers of disease progression and whether viral replication capacities differed according to host HLA class I molecules. We also sought to identify amino acids associated with differences in replication capacity within different subtypes.

Methods

Antiretroviral therapy-naïve (ARV) patients were recruited (n=103) from the Majengo sex worker cohort based in Nairobi, Kenya. Recombinant NL4-3 viruses bearing patient-derived *gag-protease* genes were generated by homologous recombination and their replication capacities, normalized to the growth of the wild type NL4-3 virus, were assayed in an HIV-1-inducible green fluorescent protein reporter T cell line. Replication capacities of NL4-3 recombinant viruses bearing *gag-protease* genes from subtypes A, C, D and inter-subtype recombinants were compared. To further assess subtype differences in Gag-protease-driven replication capacity, 10 recombinant viruses each from patients infected with subtypes A, D and A/D recombinants were constructed from the UARTO cohort in Uganda. Associations between viral replication capacities and markers of disease progression (viral load and CD4 T cell counts) and HLA class I molecules were investigated. An exploratory codon-by-codon analysis was performed using the Kruskal-Wallis test to identify amino acid residues associated with differences in replication capacity for the subtype A recombinant viruses.

Results

There were 57 (56%) subtype A, 16 (15.5%) C, 13 (12.6%) D and 17 (16%) inter-subtype recombinant viruses from the Majengo cohort studied. Replication capacities differed significantly by subtype. Gag-protease inter-subtype recombinants had the highest mean replication capacity of 1.13 ($p<0.001$ compared to subtypes A, C and D) followed by subtype D with a mean replication capacity of 0.94 ($p<0.001$ compared to subtype A, $p<0.01$ compared to subtype C) then subtype C and A with means of 0.78 and 0.69, respectively. Overall, there was no correlation between replication capacities and CD4+ T cell counts or viral loads (Spearman's

correlation; r=0.16 and p=0.10 and r=0.08 and p=0.40 respectively). There were also no significant differences in CD4 T cell counts or viral loads according to subtypes (ANOVA; p=0.53 and p=0.91 respectively), indicating that this did not drive inter-subtype differences in Gag-protease-driven replication capacity. Our finding of subtype-specific differences in Gag-protease-driven replication capacity were reproduced in the UARTO cohort. Gag-protease AD recombinants had the greatest mean replication capacity of 1.42 ($p<0.01$ compared to subtype A) followed by subtype D with mean of 1.07 ($p<0.05$ compared to subtype A) and lastly subtype A with a mean 0.76. Overall, our replication capacity differences between subtypes were reproducible across different cohorts and geographical regions. We found no significant differences in replication capacity of patients expressing protective HLA alleles versus those with non-protective HLA alleles in HIV-1 subtype A, subtype D and inter-subtype recombinants. However, for subtype C, protective alleles were associated with lower replication capacity, suggesting that the protective alleles for HIV- 1 subtypes A and D may not be the same as for the well-defined protective alleles for subtypes B and C. We additionally showed that HLA alleles A*74, A*68 and A*03 were associated with lower replication capacities, however the associations did not remain significant after correction for multiple comparisons. We identified six amino acids that were significantly ($p<0.05$ and $q<0.2$) associated with reduced replication capacity in subtype A (L75I, I107L, S125S, S126S, N315N and S499S). The polymorphism 107L and consensus amino acid 315N were significantly more frequent in subtype A (Chi square test; $p=10^{-8}$ and p=0.01, respectively) and were associated with reduced replication capacity, suggesting that these amino acid variants may contribute to the lower replication capacity of subtype A. We additionally found that 247X in the TW10 epitope was associated with reduced replication capacity in subtype A.

Discussion

Our data show a hierarchy of Gag-protease driven replicative fitness where subtypes A/C are less fit than D, which is also less fit than inter-subtype recombinants. The data are consistent with reported subtype-specific differences in disease progression in East Africa, suggesting that Gag-protease-driven replication capacity is a determinant of differences in disease progression between subtypes. It is likely that the lower functionality of subtypes A and C Gag-proteases slows disease progression in individuals infected with these subtypes, leading to greater opportunity for transmission and consequently, increased prevalence of these subtypes. We found that none of the typical protective HLA alleles defined for subtypes B and C were associated with lower replication capacity in subtypes A and D, suggesting that protective HLA alleles may differ according to subtype and geographic region. We did however identify weak associations between specific HLA alleles and reduced replication capacity, including HLA-A*74, which has previously been associated with improved clinical outcome, suggesting a plausible mechanism of viral control by this allele. Furthermore, we identified amino acids associated with altered replication capacity in subtype A viruses and thus we have extended previous studies that described fitness associations in Gag-protease from subtypes B and C. Of interest, the subtype A-specific mutation 247X in the HLA-B*57 restricted TW10 epitope was associated with lower replication capacity, indicating that this mutation may contribute to the protective effect of HLA-B*57 in subtype A infection. Further studies to better understand the mechanisms underlying subtype-specific differences in replication capacity are warranted. Furthermore, these data may have implications for the differences observed in the rate of disease progression and for the uneven spread and expansion of HIV-1 subtypes in the global epidemic.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ANOVA, analysis of variance

ART, antiretroviral therapy

CD, cluster of differentiation

CO2, carbon dioxide

DEPC, diethyl pyrocarbonate

DNA, deoxyribonucleic acid

dNTP, deoxyribonucleotide triphosphate

Env, envelope glycoprotein

Gag, group specific antigen

GFP, green fluorescent protein

GXR, CEM-GXR25

HIV, human immunodeficiency virus

HIV-1, human immunodeficiency virus type 1

HLA, human leukocyte antigen

IQR, interquartile range

kb, kilobase

kDa, kilodalton

LTR, long terminal repeat

MHC, Major Histocompatibility Complex

ml, millilitre

mm, millimetre

MOI, multiplicity of infection

PCR, polymerase chain reaction

pNL4-3, NL4-3 plasmid

Pol, polymerase

RNA, ribonucleic acid

rpm, revolutions per minute

RT, reverse transcriptase

PCR, polymerase chain reaction

SD, standard deviation

SIV, simian immunodeficiency virus

°C, degrees Celsius

µl, microliter

µM, micromolar

CHAPTER 1

INTRODUCTION

CHAPTER I -INTRODUCTION

1.1 Human Immunodeficiency Virus (HIV)

1.1.1 HIV/AIDS history

Human immunodeficiency virus type 1 (HIV-1) is a lentivirus, belonging to the *Retroviridae* family, that causes acquired immunodeficiency syndrome (AIDS), a complex disease that severely affects the immune system leading to death [1]. There exist a number of theories proposed about the origin of AIDS; nonetheless, the unifying feature in all the theories is the fact that the AIDS virus originated from non-human primate sources which is strongly suggestive of cross-species transmission [2, 3]. One explanation is that cross-transmission occurred through the exposure of humans to the blood of chimpanzees butchered locally for bush meat [4]. AIDS was initially discovered in June 1981 in homosexual men presenting with *Pneumocystis carinii* pneumonia and mucosal candidiasis [5, 6]. Similar symptoms were later clinically described in non-homosexuals, intravenous drug users, hemophiliacs, blood transfusion patients, as well as the sexual partners of these individuals [7]. Initially, the first AIDS viruses isolated were branded lymphadenopathy-associated virus (LAV) by Dr. Luc Montagnier's laboratory, and human T-cell lymphotropic virus type III B (HTLV-IIIB) by Dr. Robert Gallo's laboratory at the Pasteur Institute and National Cancer Institute respectively [8, 9]. HIV was identified as the causative agent of AIDS in 1984 [10-12]. Since the isolation of HIV, the rapid development and application of various molecular tools have substantially improved understanding of the origins and evolution of HIV [13].

1.1.2 HIV/AIDS epidemiology

HIV infection has become one of the world's most serious health and development challenges. Effective control of HIV infection remains elusive. By the end of 2014, it was estimated that 37 million people were living with HIV/AIDS globally and 70 % of people living with HIV resided in sub-Saharan Africa [14]. In Kenya, where our study cohort is based, HIV prevalence was 5.6%

(95% CI: 4.9 to 6.3) and HIV incidence was 0.5% (95% CI: 0.2 to 0.9) at the end of 2013 [15, 16]. The epidemic in Kenya is deeply rooted in the general population, as well as concentrated in high risk populations, such as sex workers and their clients, intravenous drug users, and men who have sex with men. The HIV/AIDS pandemic represents a major crisis for the progress of the African continent, which is the most affected region in the world. Despite the intervention of highly active antiretroviral therapy (HAART), which has dramatically reduced HIV-associated morbidity and mortality, HIV incidence globally remains unacceptably high, especially in the low income countries in sub-Saharan Africa where nearly 1 in every 20 adults lives with the disease [17, 18]. Treatment regimens are based on WHO guidelines and generally consist of a combination of at least three antiretroviral (ARV) drugs to maximally suppress the HIV virus and stop the progression of disease [19, 20]. The drugs are divided into first line and second line treatment regimens. There is also third line treatment which is the last option available to an HIV-infected individual should second line treatment fail [21]. The antiretroviral drug classes include entry inhibitors, fusion inhibitors, reverse transcriptase inhibitors [(nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase Inhibitors (NNRTIs)], integrase inhibitors and protease inhibitors (PI) [22].

1.1.3 HIV classification and origin

There are two major types of HIV, namely HIV-1 and HIV-2. The two types are genetically distinct and greatly differ in terms of disease outcome and in geographic patterns of distribution. HIV-2 was first isolated from West African patients with AIDS in 1986 and is closely related to SIV that infects the sooty mangabeys (*Cercocebus atys atys*) from West Africa [4, 5, 23]. The HIV-2 virus has continued to be geographically restricted to West African countries, such as Cape Verde, Cote d'Ivoire, Gambia, Guinea-Bissau, Mali, Senegal, Sierra Leone and Nigeria [24, 25] where sooty mangabeys are hunted for food and are kept as household pets [3, 26-28]. HIV-2, which is less transmissible and less pathogenic than HIV-1, is generally characterized by a longer

asymptomatic stage [29-33]. HIV-2 is divided into eight groups A-H based on phylogenetic criteria. Groups A and B are the most prevalent subtypes [20] originating from common ancestors in 1940 and 1945 respectively [5]; however, groups C-H have only been identified in individuals [34]. On the other hand, HIV-1 viruses are closely related to SIVs from chimpanzees (*Pan troglodytes troglodytes*) [35-38]. The first case of an HIV-1 sequence from a human being was characterized from a frozen 1959 plasma sample of an apparently healthy man from Leopoldville (Kinshasa) in Central Africa, which turned out to be HIV infected when the sample was assessed in the mid-1980s and it was found to cluster around ancestral nodes of subtypes B and D in the phylogenetic tree [39]. HIV-1 accounts for most of the HIV infections globally and is divided into four major groups: M (major), O (outlier), N (non-M, non-O) and P [40, 41]. Groups O and N are mainly restricted to the Cameroon and the Democratic Republic of Congo [42]. Group M, which is assumed to have originated from the Democratic Republic of Congo [25, 43], constitutes approximately 95% of HIV-1 infections and is responsible for most of the global epidemic [44-49]. It was inferred that group M subtypes originated from a 1931 common ancestor based on sequence analyses [37, 50]. Group M is further divided into nine subtypes (A–D, F-H, J, K) [51-53]. HIV-1 subtypes A and F are further sub-divided into subtypes A1-A4 and F1-F2 [52]. There is a possible tenth group L from which two full length sequences have been identified [54]. The subtypes are able to combine their genetic material to form hybrid viruses called circulating recombinant forms (CRFs). To date, 72 CRFs and multiple unique recombinant forms (URFs) have been identified [55] which are unevenly distributed globally [48].

1.1.4 HIV-1 diversity and geographical distribution

HIV-1 is characterized by a high level of genetic heterogeneity [44, 46, 56, 57] which may have a dire impact on diagnostic methodologies, disease progression, virus transmission, treatment outcomes, and eventually, vaccine development [58-61]. HIV-1 subtypes have diversified to a great extent through a number of mechanisms. Firstly, several zoonotic cross-species

transmission of the simian lentivirus has led to the different major subtypes of HIV [3, 62, 63]. Secondly, the high error rate of its reverse transcriptase (RT) enzyme results in 0.2–2 mutations per genome per replication cycle [64-66], and this, combined with a high turnover rate of about 10^{10} virions per day, has contributed immensely to increased genetic diversity [66-68]. Thirdly, the diversity between HIV-1 subtypes allows for the virus to recombine effectively, creating virulent forms in the occurrence of co-infection where two individual subtypes of the virus are multiplying the same cell. Lastly, HIV-1 mutates to avoid host selective immune pressures such as antibodies and cytotoxic T-lymphocytes (CTL) leading to the creation of a highly diverse "quasi-species" [69-72]. Also therapeutic pressure could lead to mutations contributing to diversity [73-78]. Molecular epidemiological studies show that HIV-1 subtypes have distinct geographic distribution patterns [44, 79], and this is illustrated for the African continent in Figure 1.1[80]. HIV-1 subtype B is predominant in Europe and America, and is the well-characterized subtype, however, it only accounts for 11% of all HIV-1 infections. It is also present in some Asian countries like Korea and Singapore [44, 81]. Furthermore, subtypes A, C, D and inter-subtype recombinants prevail in sub-Saharan Africa, which bears the largest global burden of HIV-1 disease [82]. Subtype A accounts for 80% of the HIV infections in Eastern Africa (Kenya, Uganda, Tanzania, and Rwanda), and about 15-30% in West Africa [44, 83, 84]. Subtype C accounts for approximately 50 % of infections globally [44, 85] and is particularly predominant in southern, Central and Eastern Africa and has expanded faster than other HIV-1 subtypes [86]. Subtype D constitutes up to 40% of HIV infections in countries of East and Central Africa where it co-circulates with subtype A [87-92]. Subtype E conventionally defined as CRF_01 AE predominates in South East Asia [93, 94], whereas subtypes G and A/G recombinants have been found in Western and Eastern Africa [95, 96], and subtypes H and K are mostly found in Central Africa [95, 97, 98]. CRFs and URFs are estimated to account for more than 20% of HIV-1 global infections [99-104]. Kenya is one of the countries in sub-Saharan Africa where HIV-1 subtypes are heterogeneous with subtypes A, C, D and inter-subtype recombinants co-existing [92, 101, 105-109].

Figure 1.1 Distribution of most prevalent HIV subtypes

A map showing an overview of the geographical dispersal of the most prevalent HIV-1 subtypes in Africa from 2005-2015. It shows a distinct distribution pattern highlighting the predominance of different subtypes in specific regions, as is shown by the various colours.

Source: http://www.vaccineenterprise.org/sites/default/files/150316_S3_Modjarrad.Kayvon.pdf (accessed January 2016)


B, and other recombinants(recs.)
CRF02_AG, A, and C
CRF02_AG, G, and high proportion of their recombinants
A, C, D, F, G, H, J, K, CRF01_AE, CRF02_AG, and high proportion of their recombinants
A, D, and high proportion of their recombinants
A, D, C, and high proportion of their recombinants
A, C, and high proportion of their recombinants
C, and low proportion of recombinants
Insufficient data

These subtypes also dominate in Uganda, Rwanda and Tanzania [110-115]. A study done prior to 2000 revealed regional heterogeneity in Kenya based on the *env* C2-V3 classification, where subtype A predominated (71%-87%), with significant components of subtype D (7%-29%) and subtype C (7%-17%) [116]. Studies done after 2000 in Kenya show that subtype A continues to dominate in most of the regions [92, 101, 109, 117, 118].

1.1.5 HIV genome organization

The HIV genome comprises approximately 9,749 nucleotides situated in the viral capsid as two non-covalently linked positive stranded RNAs. The integrated provirus of the HIV genome is flanked by 5' and 3' long terminal repeat (LTR) regions. HIV encodes for 9 genes in its genomic structure, with three major genes - group specific antigen (*gag*), polymerase (*pol*) and envelope glycoprotein (*env*) - which code for structural proteins as well as enzymes. These genes are expressed as polyproteins. Additionally, HIV encodes for proteins which have regulatory and auxiliary functions. These include trans-activator of transcription (*tat*), regulator of ***v***irion protein expression (*rev*), negative replication factor (*nef*), virion infectivity factor (*vif*), viral protein R (*vpr*), viral protein U (*vpu*) (present in HIV-1) and virion-associated protein X (*vpx*) (present in HIV-2). The genome arrangement is illustrated in Figure 1.2 (i) [119]. The structure of HIV, which is composed of a nucleoprotein core enclosed by the proteolipid envelope, has been illustrated as a schematic in Figure 1.2 (ii) [120]. The structure comprises of an inner and outer core. The inner core comprises of matrix, capsid and nucleocapsid proteins, whereas as the outer core consists of a lipid bilayer which is synthesized from the host cell and consists of exposed surface glycoproteins (gp120) that are attached to trans-membrane proteins (gp41). The HIV genome can generally be divided into four portions - the LTR, core structural proteins, viral enzymes, and coat proteins. The LTR consists of the R (repeat) sequence segment, the U3 region and the untranslated 5' segment (U5). The structural proteins include matrix, capsid, nucleocapsid, the p6 protein and two spacer proteins (p2 and p1) which are all encoded by the

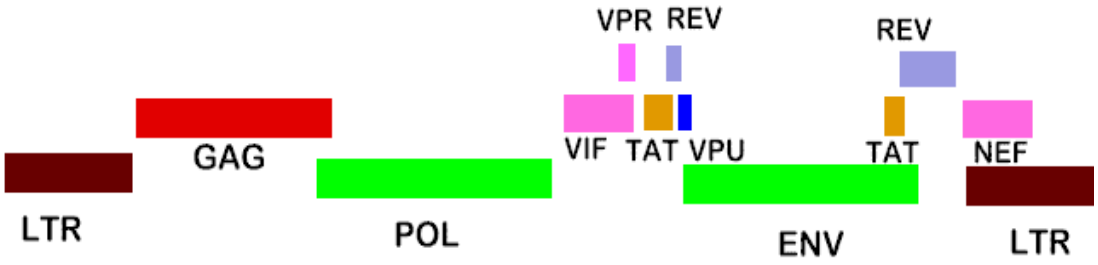
Figure 1.2 Schematic illustration of HIV-1 genome arrangement and virion and structure

(i) A diagram representing the map of the HIV genome illustrating the various structural and accessory genes. Long terminal repeats flank the 9 genes at 5' and 3' ends. Source: http://www.microbiologybook.org/lecture/Image138.gif (accessed December 2015).

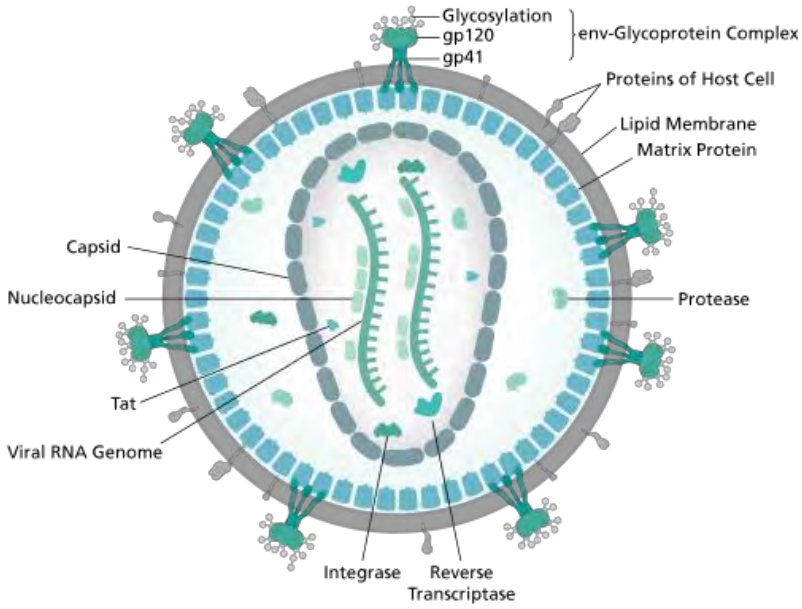
(ii) A drawing showing the structure of HIV with the approximate locations of Gag proteins, the Env glycoproteins, and the *pol*-encoded enzymes.

Source: https://en.wikipedia.org/wiki/Structure_and_genome_of_HIV#/media/File:HI-virion-structure_en.svg (accessed December 2015).

i


VPR
REV
REV
GAG
VIF
TAT
VPU
TAT
NEF
LTR
POL
ENV
LTR

ii


Glycosylation
gp120
gp41
env-Glycoprotein Complex
Proteins of Host Cell
Lipid Membrane
Matrix Protein
Capsid
Nucleocapsid
Protease
Tat
Viral RNA Genome
Integrase
Reverse
Transcriptase

gag gene [121]. The viral enzymes include protease (PR), reverse transcriptase (RT), and integrase (IN) which are encoded by the *pol* gene [122]. The coat proteins consist of the structural unit (gp120) and the trans-membrane unit (gp41) which are encoded by the *env* gene [123].

1.1.5.2 Accessory genes

The HIV genome has six additional genes known as accessory genes which have supporting functions whereas two of these genes, *tat* and *rev* code for regulatory proteins. Tat is a 101 kDa protein encoded by two different exons from spliced mRNAs that activates viral transcription through binding to the trans-activation response element (TAR) (TAR forms the necessary attachment sites for RNA pol II and cellular proteins). Tat is also used to induce apoptosis of predominantly uninfected cells [124]. Rev is a 13 kDa protein that facilitates the synthesis of viral structural proteins and enzymes by altering the splicing of genomic viral RNA [ordinarily cellular messenger RNA (mRNAs) are spliced before leaving the nucleus] via an arginine-rich RNA-binding motif during replication [125, 126]. The Nef protein, on the other hand, is 27 kDa in molecular weight. Its role is to modulate the expression of cluster of differentiation 4 (CD4) antigen by the infected cell hence facilitating viral budding and egress from infected cells. It manipulates the host's cellular machinery by interfering with signal transduction pathways, thereby enhancing viral infectivity and production [127]. The 23 kDa Vif protein increases the infectivity of the HIV particle by preventing the inhibitory effect of cellular host cell factor apoliprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G). Vif sequesters APOBEC3G and related proteins from the budding virion and promotes their ubiquitin-mediated proteosomal degradation [128]. The Vpr protein is 14 kDa and it transports the pre-integration complex (the complex consists of viral RNA, reverse transcriptase and integrase proteins) into the nucleus where viral integration is concluded [129, 130]. Additionally, the Vpr protein induces cell cycle arrest at the G2 phase (pre-mitotic phase) and apoptosis, which allows the virus to replicate more efficiently [131]. Lastly, the Vpu protein, which is 16 kDa,

facilitates assembly of new virus particles and promotes CD4 protein degradation within the infected cell thereby assisting in budding [132, 133]. It also interacts with and down-regulates a host cell factor called tetherin, aiding in virion release from the cell membrane [134, 135]. The role of the Vpx protein in HIV-2 is not clearly defined, but it is associated with enhancing viral replication by counteracting the host antiviral factor SAM domain and HD domain-containing protein 1 (SAMHD1) which blocks HIV replication [136]. The Vpx protein is approximately 16 kDa.

1.1.5.3 Core structural proteins

Gag is synthesized as a 55 kDa polyprotein ($Pr55^{Gag}$) which is cleaved into the five domains essential in viral assembly and release namely: matrix (MA; p17), capsid (CA; p24), nucleocapsid (NC; p7), p6 and spacer peptides 1 and 2. The MA, at the N terminus of the Gag polyprotein, forms the inner layer of the virus and it is believed to be involved in the active transport of proviral DNA into the nucleus and in virion assembly upon myristoylation (addition of specific fatty acids) [137]. In addition, MA also contributes to infection in non-dividing cells, such as macrophages, because it is identified by cellular nuclear import machinery and is involved in the specific incorporation of viral glycoproteins into the virion [138]. The CA forms the main portion of the virion core shell. It contains precursor forms of the enzymes needed for replication and it regulates uncoating of the virus by interacting with cyclophilin A, a host factor necessary for HIV replication [139]. Basically, the CA is the portion of the virus that gets inserted into the host cell upon infection [140]. The NC coats the viral RNA, forming a protective complex by binding to the packaging signal of RNA. The NC transports the viral RNA during assembly and also shields it from nucleases [141]. The p6 comes from the C-terminal portion of the Gag polyprotein and it facilitates particle release of virions from infected cells [142]. Lastly, the spacer peptide 1, which separates NC and p6 regions, is derived from proteolytic cleavages. It coordinates binding of the membrane and mediates Gag-Gag lattice formation during Gag processing [143]. On the other

hand, space peptide 2, which separates CA and NC, facilitates the p6 protein cleavage efficiency during virion maturation [144, 145].

1.1.5.4 Viral enzymes

The Pol polyprotein encoded by the *pol* gene is consecutively cleaved into three enzymes - PR, RT, and IN - following the 5' to 3'end direction [146]. All 3 proteins are found within the capsid of free HIV-1 virions. The enzymes are not active in their monomeric forms; they must interact into their catalytic state to be active [147]. HIV PR is a retroviral aspartyl protease that acts as a dimer. It mainly mediates the proteolytic cleavage of specific amino acid sequences. During or shortly after the virus buds off from the host cell, PR splits $Pr55^{Gag}$ into mature Gag proteins MA, CA, NC and p6. This cleaving is an important step in the life cycle because it activates the viral particles to functional enzymes and proteins without which the virion would not be infectious [148, 149]. On the other hand, RT (a heterodimer of p65 and p50) synthesises DNA from the viral RNA and RNase H (cleaved from RT) degrades viral RNA during DNA synthesis [150, 151]. Lastly, IN, which is a tetramer, mediates the incorporation of the viral nucleic acid into that of the host. The enzyme consists of three domains namely: the N-terminal domain (the zinc binding portion that is involved in DNA binding), the core domain (the catalytic portion speeding up the incorporation process), and the C-terminal domain (where non-specific DNA binding occurs) [152-154]. The enzyme works in a three step direction: the exonuclease activity step, the endonuclease activity step, and lastly, the ligase activity step. These three enzymes have been used as targets for antiretroviral drugs [155, 156].

1.1.5.5 Envelope structural proteins

The Env protein is paramount in binding the target cell and is exposed to the outside environment during HIV-1 replication, hence it becomes an important target of host immune responses and is an important determinant of viral pathogenesis [157]. The 160 kDa Env (gp160) heavily

glycosylated polyprotein is encoded by the *env* gene and is cleaved by cellular furin in the endoplasmic reticulum into two main units namely, the external surface protein (gp120) and the transmembrane protein (gp41). The Env proteins, which are covalently bound to each other and are responsible for viral binding, are found in the lipid bilayer (derived from host membrane) that surrounds the viral capsid. The gp120 glycoprotein exists as trimers anchored on the viral membrane [158] and is responsible for the tropism of a virus because it is the unit that is concerned with the receptor binding function. It covers most of the exposed surface of the viral envelope. The crystal structure of core gp120 reveals an outer domain, an inner domain with respect to its termini, and a bridging sheet [159]. The inner domain sequences are not homologous to any other organisms, while the outer domains have sporadic homology [159]. It has five conserved core domains (C1-C5) scattered between five hypervariable glycosalated loops (V1-V5) and C4 region which mediates the binding to CD4 molecule leading to conformational changes in the gp120. On the other hand, the fusion gp41 subunit consists of a six helical bundle in its core and several peptide sites within its membrane domains [160].

1.1.6 HIV lifecycle

The life cycle of HIV-1 comprises a series of steps necessary for the successful infection of a human target cell, as illustrated in Figure 1.3 [161] and summarized in sections 1.1.6.1 - 1.1.6.4.

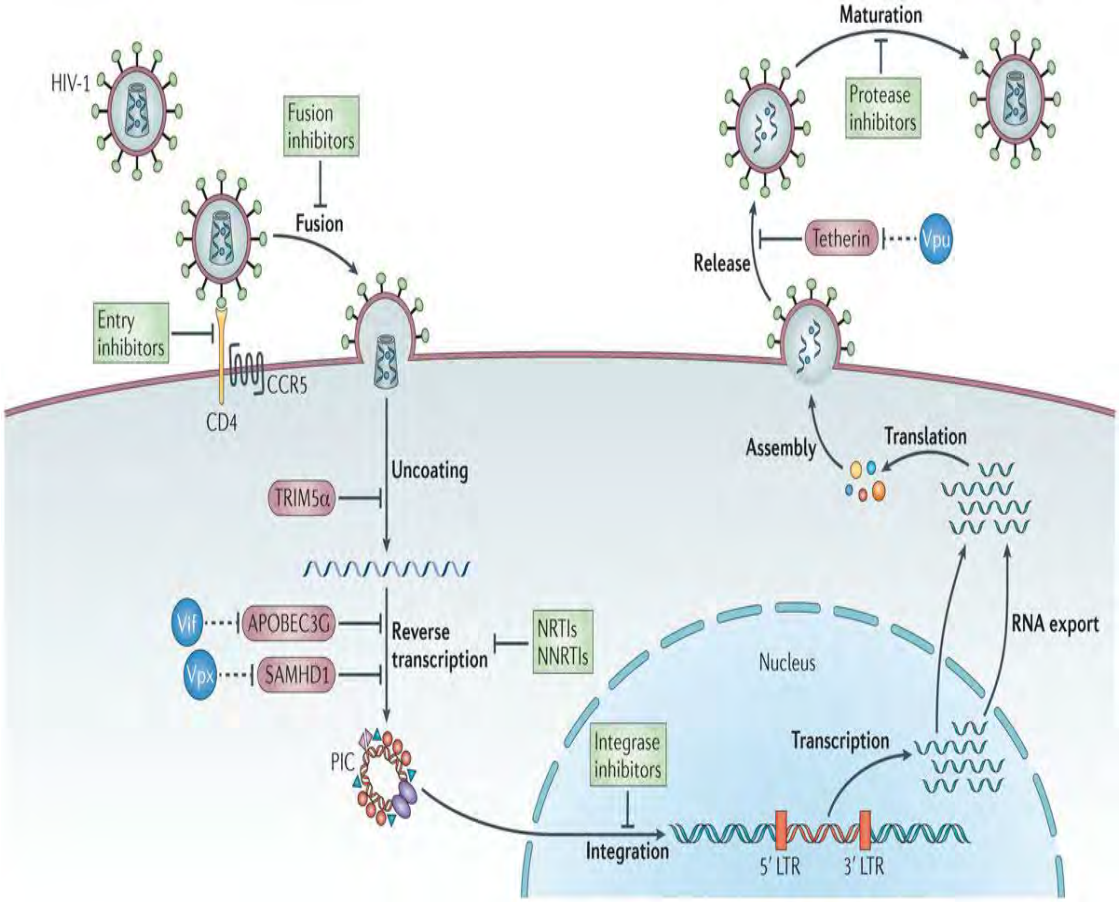
1.1.6.1 HIV-1 entry

The initial step of the life cycle is attachment, which occurs due to the high affinity interaction of the gp120 envelope glycoprotein and CD4 ligand on the host cell surface. The gp120 undergoes conformational changes [162], which facilitates its binding to the chemokine receptor molecules -C-C chemokine receptor 5 (CCR5) found on macrophages and dendritic cells or the CXC chemokine receptor 4 (CXCR4) found on T cells [163, 164]. The interaction with the chemokine receptor (HIV coreceptor) initiates the membrane fusion process which exposes gp41 fusion

Figure 1.3 Steps in the HIV-1 replication cycle

A figure showing the main steps of the HIV-1 replication cycle. Also highlighted are the major classes of antiretroviral drugs (in green) at every major step of the life cycle and the key restriction factors (in red) coupled with their corresponding antagonists (in blue).

Source: http://www.nature.com/nrmicro/journal/v11/n12/fig_tab/nrmicro3132_F2.html (accessed January 2015).


HIV-1
Fusion inhibitors
Fusion
Entry inhibitors
CCR5
CD4
Uncoating
TRIM5α
Vif
APOBEC3G
Vpx
SAMHD1
Reverse transcription
NRTIs
NNRTIs
PIC
Integrase inhibitors
Integration
5′ LTR
3′ LTR
Nucleus
Transcription
RNA export
Assembly
Translation
Maturation
Protease inhibitors
Release
Tetherin
Vpu

peptides facilitating entry into the cell [165, 166]. Depending on the coreceptor usage, viruses are known as M-tropic or T-tropic. Entry and fusion inhibitors work by interfering with binding or fusion by binding to either viral gp41 or host CD4+ cell or chemokine receptors [167-169].

1.1.6.2 Reverse transcription and integration

Upon entry into the host cell, uncoating of the capsid and reverse transcription of the viral genome occur. RT has two enzymatic units - DNA polymerase and RNAse H - that it uses to synthesize a double stranded complementary DNA from the single stranded genetic RNA template and degrade the RNA template, respectively [170]. The newly synthesised DNA is then translocated to the nucleus via the nuclear pore where it is inserted into the host cellular DNA genome. A key unique feature of HIV is to integrate its genome into that of the host cell mediated by IN. IN recognizes the LTRs at the ends of the newly synthesized viral DNA duplex and cleaves 2-3 bases from the 3' end. This is then followed by a trans-esterification reaction that takes place joining proviral and cellular DNA ends where 4-6 base pair gaps of mismatched ends are trimmed, filled and ligated [171]. NRTIs are nucleic acid analogues that mimic the normal building blocks of DNA, preventing transcription of viral RNA to DNA. NNRTIs block the genomic HIV binding site of RT [172], whereas IN inhibitors prevent the transfer of proviral DNA strands into the host chromosomal DNA [173].

1.1.6.3 Transcription and translation

After integration, transcription of the integrated DNA is initiated (initiated by Tat and using the host's cellular machinery), starting with the short spliced mRNAs encoding for Tat, Rev and Nef [174]. The RNAs generated could be spliced mRNA (used for translation for viral proteins) or unspliced genomic RNA [175]. Rev is involved in exporting viral RNAs into the cytoplasm [176]. The virus then uses the host's own cellular machinery to generate viral proteins. The mRNA is

transcribed to form polypeptide chains (long chains of amino-acids), which fold to form the protein and enzyme components of new virus particles [177].

1.1.6.4 Assembly and maturation

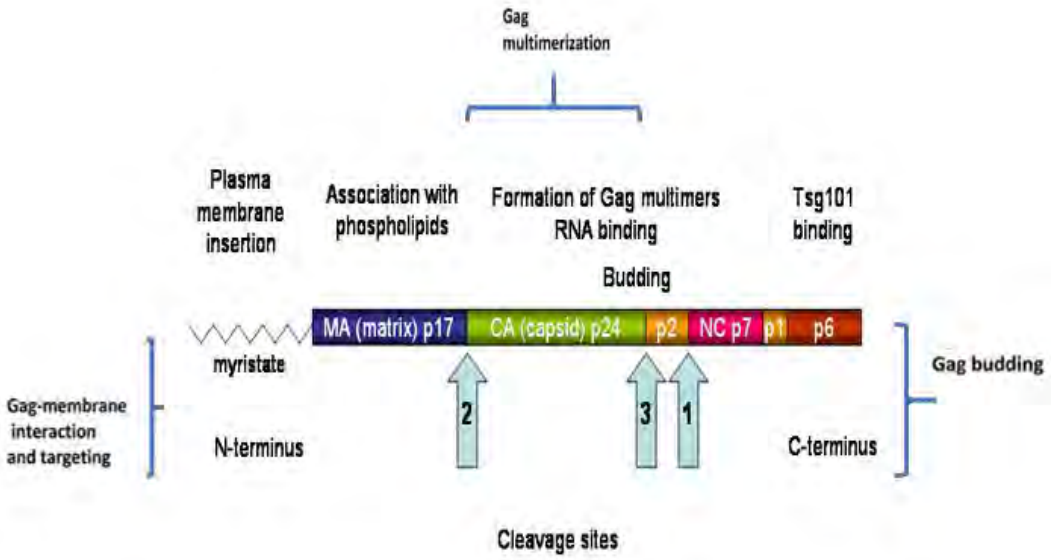
Once the viral subunits have been generated and processed, they move to the inner surface cellular membrane where they are assembled and packaged, and then separated for the final assembly of a new virus. The structural subunits network with the cell's membrane deforming a section of the membrane, which permits the nucleocapsid to take shape and the viral RNA is coiled tightly to fit inside the nucleocapsid [178, 179]. Once the new viral particles are assembled, they bud off the host cell, and create a new virus. The virus enters the maturation step which is the final step where HIV-1 PR cleaves Gag and Gag-Pol polyproteins precursors in a sequential manner into mature functional proteins in well-defined sites, hence making the viral particles infectious [179]. Maturation is dynamic and involves sequential conformational changes and structural protein rearrangement. Some of these include activation of the fusogenic activity of the viral Env protein, stabilization of genomic RNA, and CA assembly, among others [180, 181]. PIs block the final maturation stages of HIV replication, resulting in the formation of non-infective viral particles [182].

1.1.6.5 Role of Gag

The Gag protein is essential during the assembly of viral particles during the replication cycle. It recruits all the building blocks required for the formation of fully infectious mature viruses. Gag cleavage occurs sequentially at specific sites [183-185] as depicted in Figure 1.4 [185]. The MA remains associated with lipid envelope of the mature viral membrane. The MA takes part in the early stages of virus replication as well as in RNA targeting to the plasma membrane, incorporation of Env proteins into virions and particle assembly [137].

Figure 1.4 Structure of Gag protein

The diagram shows the arrangement of Gag proteins. The cleavage sites are highlighted in order of cleavage. Source: http://www.microbiologybook.org/lecture/hiv-gag.gif (accessed December 2015).


Gag
multimerization
Plasma
membrane
insertion
Association with
phospholipids
Formation of Gag multimers
RNA binding
Tsg101
binding
Budding
MA (matrix) p17
CA (capsid) p24
p2
NC p7
p1
p6
myristate
N-terminus
C-terminus
2
3
1
Gag-membrane
interaction
and targeting
Gag budding
Cleavage sites

The CA protein is paramount in the particle assembly and forms the shell of the core in a mature virion. The CA consists of protein hexamer subunits to which host restriction factors, such as TRIM5-alpha, bind to, causing premature capsid disassembly [139]. The NC encapsulates and protects viral dimeric unspliced genomic RNA (gRNA). The p6 region adheres to the tumor susceptibility protein 101 (Tsg101) (a human cellular protein required for HIV replication) during the viral budding process [186] and the C terminal part of p6 binds to endosomal sorting complex required for transport (ESCRT) machinery required for retroviral budding [187]. The peptide spacers are also important in HIV assembly and budding [188].

1.1.7 HIV-1 pathogenesis

HIV-1 infection leads to destruction of a person's immune system, particularly CD4+ T lymphocytes which are the preferred target cells [189]. The effect of HIV on the immune system is monitored by measuring the CD4+ T (T helper) lymphocyte count and, together with viral loads (amount of HIV in a body fluid), and markers of immune activation the three parameters are used as markers of HIV-1 disease progression. The course of progression is dependent on a number of factors which contribute to the pathogenesis of the infection including viral, immunological and other host factors, such as age, gender, mode of transmission, body mass index, haemoglobin and physiological factors [190, 191]. Individuals progress at different rates to AIDS, which is defined as a CD4+ T cell count below 200 cells per μL [192], and may be classified as rapid progressors (individuals who develop AIDS within 3-5 years post infection in the absence of ARVs), typical progressors (individuals who develop AIDS within approximately 10 years after seroconversion) or long term non progressors (LTNP; individuals who remain asymptomatic for over 10 years post infection in the absence of ARVs) [193, 194]. Based on the viral loads, individuals who are slow progressors may be categorized as either elite controllers (HIV RNA <50 copies/mL) or viremic controllers (HIV RNA <2,000 copies/mL) [195-197]. HIV-1 infection can be categorized into three distinct phases: primary infection, chronic infection and clinical

AIDS [198]. The first phase is the primary or acute or seroconversion phase. It develops within 2-4 weeks of infection, with many people developing flu-like symptoms referred to as acute retroviral syndrome (ARS). Severity varies considerably between individuals but the phase is characterized by high viral loads as the amount of virus released is massive. During this phase there is a rapid decrease in CD4 count as the virus replicates in activated CD4 T cells. This is the most contagious phase of infection due to the extremely high viral loads [199-201]. Next is the chronic or clinical latency phase or asymptomatic phase. Subjects in this phase usually recover from the symptoms associated with the previous phase as the viral load reaches a set point where it stabilizes. Despite relatively stable viral loads during the clinical latency phase, HIV replication is substantial and there is rapid turnover of CD4+ T cells. The clinically asymptomatic phase can last for years and during this phase there is gradual decrease in CD4 T cells [202, 203]. Lastly, clinical AIDS develops and this is the last phase of infection characterized by a damaged immune system which is vulnerable to opportunistic infections and cancers. The end point is usually death [121, 204, 205].

1.1.8 HIV-1 subtype differences

HIV-1 subtypes exhibit differences in their epidemiological and clinical disease outcomes as well as their biological properties, such as their genetic variation and cellular tropism [206-213].

1.1.8.1 Epidemiological and clinical outcome

1.1.8.1.1 Distribution and uneven expansion

The hallmark of HIV-1 is its extreme genetic diversity exhibited by the uneven global distribution and expansion of groups, subtypes, inter-subtype recombinants, CRFs and URFs, as discussed in sections 1.1.2 to 1.1.4. The reason for the uneven expansion and diverse distribution of HIV-1 subtypes is not fully understood; however, several theories have been assumed. In the first place,

the most common assumption is the founder effect influence - i.e., HIV-1 subtypes predominate according to the strains that first established infection in a particular area giving the subtype precedence for its dominance over other subtypes [214]. Then there is the prevalent route of transmission theory, which may favour one subtype over another, assuming that there are subtype specific differences in the preferred route of transmission rendering the particular subtype with a transmission advantage [85, 214]. Lastly, behavioral patterns/cultural practices and environmental limitations could also, to some extent, influence the geographical distribution of HIV-1 subtypes [215]. This global uneven expansion and distribution of HIV-1 subtypes has led to a divergent HIV evolution which could contribute to the differences in virulence between the subtypes [48], and may have a dire impact on public health, clinical consequences (diagnosis, treatment and prevention strategies) and, ultimately, vaccine development.

1.1.8.1.2 Transmission efficiency

HIV-1 subtypes differ in their ability to establish initial infection [216, 217]. There are recognized differences in transmission efficiency of the different HIV-1 subtypes [218-222], however, it is uncertain whether the geographical distribution or timing of initial entry of subtypes in the population determines the transmission dynamics. Primarily, high viral load burden during acute HIV-1 infection has been identified as a major determinant of HIV-1 transmission [200, 223]. Besides that, certain HIV-1 subtypes have been associated with particular modes of transmission. For instance, subtype B appears to spread better through homosexual contact and intravenous drug use, while subtype E and C may thrive better through heterosexual transmission [95, 224-227]. It was also reported that subtype E had a slightly higher potential of vertical transmission than subtype B in a study done in Japan [228]. Moreover, higher heterosexual transmission of subtype A compared to subtype D [88] has been observed, which may have contributed to the expansion of subtype A in Uganda [229]. Reports differ on whether subtype C is associated with higher transmission efficiency compared to other subtypes. Some studies suggest that subtype C

has equal transmission efficiency with other M group subtypes [230-232], whereas other studies reported that subtype C was associated with increased transmissibility compared to subtypes A and D [218, 222, 233, 234]. Interestingly, subtype A, which has been shown to have a slower disease progression to AIDS or death, has been associated with significantly higher transmission rate than subtype D [88]. Also, a study in Thailand suggested that subtype E had a higher potential of heterosexual transmission as compared to subtype B [221]. This was attributed to the higher tropism of subtype E for Langerhans cells which line the vaginal mucosa and the penile foreskin [235-237], however, these observations have been fiercely contested [221, 238].

1.1.8.1.3 Disease progression

There have been varying reports on HIV-1 subtype differences in the rate of disease progression [105, 239-246]. Several studies in East Africa, where subtypes A and D co-circulate, have established that subtype D is associated with a faster disease progression than subtype A [105, 206, 240-242, 245, 247-249]. In addition, there is evidence showing that recombinant forms have a faster rate of progression to AIDS when compared with subtype A [240, 250]; however, this was not observed in another cohort in Tanzania [242]. There are also inconsistent reports of HIV-1 subtype C disease progression relative to other subtypes [242, 246, 248, 251]. In Kenya, for instance, women infected with subtype C had a faster rate of disease progression and immunosuppression than those infected with subtypes A or D [116, 220], whereas a study in Tanzania did not find any difference in disease progression between subtypes C or A and recombinants in mothers infected with these subtypes [242]. In West Africa, subtype A has been associated with slower disease progression compared to other prevalent subtypes [252]. There are also a number of studies that have documented differences in viral loads between the different subtypes [213, 241]. These authors suggest that the viral load differences resulted from inter-subtype biological differences. A few studies that compared the clinical symptoms in acute infection between subtypes have shown that subtype D patients experienced more symptoms and

suggested that this was due to increased viral fitness of subtype D and therefore more inflammatory response in the subtype D infected patients [241, 244, 246, 253, 254]. However, the mechanisms underlying subtype differences in terms of HIV-1 disease progression have not been fully understood.

1.1.8.2 Biological differences

The high diversity of HIV has impacted the development of antiretroviral therapy and vaccines since these interventions target the viral proteins [255]. Diversity occurs as a result of high rates of mutations facilitated by the error-prone reverse transcriptase, recombination events, viral replication differences and pressures exerted by the immune system [256]. There is large genetic variation between subtypes. For example, 25-35% variation has been reported in the *env* gene between subtypes based on amino acid sequences [97, 241, 257, 258], and LTR sequences vary considerably from subtype to subtype in the copy numbers of enhancers and promoter structures used in the replication cycle [234, 259-262]. Furthermore, differences in efficiency of protein function between subtypes have been reported for several HIV proteins, namely Env [263], RT [264], PR [265], Vif [266], Pol [206], and Nef [267]. Furthermore, inter-subtype differences in LTR activity have been reported [259]. Most of these studies analysed small numbers of samples or lacked clinical data, therefore the implications of these inter-subtype differences for disease progression and the epidemic in general remain largely undetermined.

HIV-1 interacts with the CD4 molecule (primary receptor) and beta-chemokine receptors CCR5 and or CXCR4 (secondary receptors) and depending on coreceptor utilization the strain is classified as either M- tropic (R5 viruses) or T-tropic (R4 viruses). Coreceptor usage has been recognized to have an impact on the rate of disease progression [185, 247-252] and differences in coreceptor usage between subtypes have been reported. Studies have shown that HIV-1 subtype D exhibits a greater degree of dual-tropism than other subtypes [268, 269]. Interestingly, it has

been reported that subtype C is associated with predominant usage of the CCR5 coreceptor throughout infection, whereas a coreceptor switch to CXCR4 or an alternative coreceptor is quite common for other subtypes [207, 270]. Overall, although genetic differences between subtypes are uncontested, it remains unclear whether biological differences may explain epidemiological differences in geographic spread and reported differences in viral loads and rates of disease progression.

1.2 CTL immune control of HIV-1 infection

The design of potentially efficacious HIV-1 vaccine will be dependent on the identification of correlates of protection of the adaptive immune response such as virus-specific CTL responses. The cluster of differentiation 8 (CD8) is a membrane glycoprotein which acts as a coreceptor for the T cell receptor (TCR) and it is expressed on the surface of cytotoxic T cells. CD8 forms a dimer comprising of a pair of chains (CD8-α and CD8-β) both with an immunoglobulin variable (IgV)- like extracellular domain. CD8 has a high affinity for HLA class I molecules. HLA class I molecules present epitopes to the TCR during antigen-specific interaction leading to the formation of epitope-HLA complex [271]. CD8 assists with the binding of the TCR to the epitope-HLA complex. This activates CTL enabling them to recognize and lyse the infected cell by secretion of perforin and granzymes [272]. HIV-1 infection induces a strong host cellular immune response that is critical in controlling viral replication [273, 274]. CTL responses against HIV-1 exert a immune selection pressure on the virus, causing selection of HIV-1 variants that readily adopt to the selective pressure exerted by the host immune system through developing escape mutations [275-278]. Escape mutations alter the formation of the epitope-HLA complex by disturbing the processing [75, 277, 279, 280], presentation [281-283] and/or recognition of the target epitope [284]. Viral escape from CTL in certain epitopes and flanking regions have been associated with HIV-1 control [277, 280, 285], however other studies have associated CTL escape with loss of virus suppression [286-288]. Preferential Gag T-cell immune responses have been

associated with HIV-1 control [274, 289-292], although another study did not observe any such association with viral control [293]. Overall, CTL responses play an important role in HIV-1 control and targeting of specific Gag epitopes may be beneficial.

1.2.1 Impact of HLA on immune control

HLA proteins play a significant role in immune responses particularly to intracellular pathogens such as viruses. HLA class I molecules are well adapted to present epitopes of HIV proteins on the surfaces of infected cells dictating immune recognition, which viral epitopes are to be targeted and CTL repertoire [294]. HLA class I profile is an important genetic determinant of clinical outcome, and immune control of HIV-1 largely depends on the HLA genes expressed by each individual [295-298]. There are six HLA class I alleles expressed by every individual, namely two HLA-A, two HLA-B and two HLA-C alleles, and HLA frequencies vary significantly between different populations (which are infected with the different subtypes) [299-301]. HIV-1 viral control has been associated with specific protective HLA class I alleles, such as HLA-B*27, HLA-B*57:01, HLA-B*58:01 and HLA-B*81, mostly in subtype B and/or subtype C infected populations [296, 302-305]. These protective HLA restricted CTL responses target several epitopes within the HIV-1 Gag CA protein and this may mediate their association with better clinical outcome [276, 282, 303, 306-310]. Other HLA alleles, however, such as HLA-B*18:01, HLA-B*35 and HLA-B*58:02, have been associated with rapid disease progression [311, 312].

1.2.2 HLA-driven determinant of HIV-1 fitness

Replication fitness is the efficiency with which a virus replicates owing to the selective pressure of its environment [313]. Replication capacity is defined as the ability to infect target cells under ideal conditions, for instance in *in vitro* assays, and is usually used as a marker of viral fitness (the ability of the virus to reproduce). HIV replication fitness has been shown to influence the predominating variants in patient quasispecies [314]. In addition, it was shown that there were

viral fitness differences in ARV-naïve patients' isolates indicative of natural genetic polymorphisms existing in each viral strain [315]. A number of studies have associated some HLA class I alleles with a Gag-mediated fitness differences, attributed by ability of these molecules to exert immune pressure and thereby alter viral sequences as the virus attempts to evades immune recognition [306, 316-320]. Specifically HLA-B*57/HLA-B*58:01-restricted CTL responses to key epitopes in the Gag CA results in fitness costly mutations such as A146P [277], A163G [321] and T242N [281]. However, most studies investigating relationships between HLA, escape and viral fitness have been conducted on subtypes B and C. Identification of regions of viral vulnerability that lower viral fitness as a consequence of immune pressure may have implications for vaccine design, as these regions can be targeted in immunogen design strategies. Therefore, there is an urgent need to identify such regions in epidemiologically important subtypes such as A, D and intersubtype recombinants that have not been investigated before, and in regions of the world where HLA class I alleles that may impact on disease outcome have not been fully characterized.

1.3 Project aims and strategy

The HIV-1 epidemic is heterogeneous with various subtypes being prevalent. Subtype-specific differences in the rate of disease progression have been reported; however, the mechanisms underlying this heterogeneity are still unknown (Section 1.1.8). The Gag protein is a relatively conserved region of the HIV proteome, and it plays an essential role in the virus life cycle and virus tropism (Section 1.1.6.5). The Gag protein is also a known preferred target of HIV specific CTL. Some studies have shown that immune responses focused on Gag are associated with viral control (Section 1.2) and it has been shown that certain mutations in the *gag* gene which are driven by the host CTL responses may reduce viral fitness in subtypes B and C (Section 1.1.2). Additionally, previous studies of HIV-1 subtype B and C showed that Gag-protease-driven viral fitness correlated with markers of disease progression, suggesting that it is a significant

determinant of disease progression (Section 1.2.2). Therefore, our major aims were to determine whether subtype-specific sequence variability in Gag is associated with differences in Gag-mediated viral replication capacity and to investigate whether this may contribute inter-subtype differences in disease progression and prevalence. We sought to address these questions using the ARV naïve patients of the Majengo sex worker cohort based in Nairobi, Kenya. To do this, we generated patient derived *gag-protease* NL4-3 recombinant viruses and assayed their replication capacities in an HIV-1-inducible green fluorescent protein reporter T cell line, using previously published techniques [317, 322]. Protease was included in the patients' chimeric constructs because it cleaves the Gag protein and protease sequence may co-evolve the Gag protein sequences within an individual to allow for efficient Gag processing. Thus Gag-protease amplicons from patients preserved the naturally occurring interaction and between the protein and allows for efficient processing of the Gag polyproteins. In addition to our main aim of investigating whether there were subtype-specific replication capacity differences in recombinant viruses possessing patient-derived *gag-protease* genes, we aimed to determine whether individuals who express protective HLA class I alleles would have lower viral replication capacity. Furthermore, we sought to identify specific amino acid polymorphisms in HIV-1 subtype A (which represented the majority of sequences samples) that were associated with differences in replication capacity. An effective vaccine intervention would be that which would cover a wide range of subtypes and inter-subtype recombinants to curb further the spread of HIV-1, and our cohort allowed us to extend previous studies of subtypes B and C to identify regions of Gag-protease vulnerability in subtype A, D and intersubtype recombinants which, together with subtype C, characterize the East African epidemic. The objectives of our study were as follows:

- Generate Gag-protease NL4-3 recombinant viruses using plasma samples from ARV naïve HIV-1 infected patients and measure their replication capacities;

- Associate viral replication capacities with markers of disease progression, namely viral load and CD4 cell counts, and determine whether there are differences in Gag-protease-driven viral replication capacities according to HIV-1 subtype;
- Determine whether there are host HLA associations with viral replication capacities in the study population; and
- Identify amino acid sequences associated with decreased or increased viral replication capacity.

CHAPTER 2

MATERIALS AND METHODS

CHAPTER 2- MATERIALS AND METHODS

2.1 Ethics and study populations

2.1.1 Majengo cohort

The Pumwani Majengo female sex worker cohort was established in 1985 and to date has a total of over 4,000 women enrolled [323]. It is one of the largest, most comprehensively studied sex worker cohorts and has been used to conduct HIV immunobiology and sexually transmitted infection (STI) research. 103 patients were recruited in this study with the following inclusion criteria; sufficient plasma sample available for analysis, CD4 counts with follow up for at least two years post sample collection, HLA profile available and ARV naïve. The study subjects gave informed consent at recruitment and enrollment for immunological, genetic and virological studies for which ethical approval was given by the Kenyatta National Hospital Ethics Review Board (KNH/ERC/IR/121, protocol P211/09/2006). Our study protocol was also approved by Biomedical Research Ethics Committee of the University of KwaZulu-Natal (BREC no: BE181/12). Plasma samples (stored at -80 ◦C) that were collected between 2006 and 2010 were the study material used to generate Gag-protease recombinant viruses, for which replication capacity was subsequently measured. CD4 counts and HLA typing were previously determined using standard assays for all participants, as previously described [324]. Viral loads were measured at Global labs using Nuclisens Easy Q platform with a detection limit of 50 RNA copies/ml, according to the manufacturer's instructions. Cohort characteristics and demographics are summarized in Table 2.1.

2.1.2 Uganda AIDS Rural Treatment Outcomes (UARTO) cohort

The UARTO cohort study was approved by the University of British Columbia (UBC) Providence Health Care Research Institute (UBC-PHC REB NO: H11-01642). In order to validate our results from the Majengo cohort, we used *gag-protease* reverse transcription polymerase chain reaction (RT-PCR) products from 30 HIV-1 infected patients (10 each from subtypes A, D and AD)

Table 2.1 Clinical and demographic characteristics of the Majengo cohort

Characteristics	Median (interquartile ranges)
Age	30 [a] (30-40) years
CD4 T cell count	359 (342-520) cells/mm^3
Viral load	5.23 (3.82-5.27) log_{10} copies/ml

[a] The mean is reported.

from the UARTO cohort (kindly donated by Dr Zabrina Brumme from Simon Fraser University) to generate Gag-protease recombinant viruses and measure their replication capacities. The clinical characteristics of the cohort are summarized in Table 2.2.

2.2 Generation of Gag-protease NL4-3 recombinant virus stocks

HIV-1 RNA was extracted from the patients' plasma and the, *gag–protease* region was amplified to generate amplicons for recombination. Gag-protease NL4-3 recombinant viruses were generated by co-transfection of a CEM-derived T cell line with patient-derived *gag-protease* amplicons and NL4-3 *gag-protease*-deleted plasmid (pNL4-3Δ*gag-protease*) based on methods previously described [325]. The procedures followed are summarized in Figure 2.1.

2.2.1 Plasma RNA extraction

The plasma samples were thawed at room temperature. HIV-1 RNA was extracted from 150 µl plasma using the Nucleospin RNA virus kit (Machery-Nagel, Versmold, Germany) as per manufacturer's instructions. In brief, the kit employs the silica membrane technology using mini spin columns with a binding capacity of 40 µg for nucleic acids. All plasma samples were concentrated by centrifugation (Jouan MR23i, Thermo Scientific, Delaware, USA) at 14,000 rpm (radius of rotor 100 mm) for 2 hours at 4 °C prior to RNA extraction. The extracted RNA was stored at -80 ˚C until use. The extraction process is summarized in Figure 2.2.

2.2.2 Amplification of HIV-1 *gag-protease*

The *gag–protease* region was amplified by RT-PCR using the Superscript III One-Step RT-PCR system with platinum Taq DNA polymerase (Invitrogen, Carlsbad, USA) and the following forward and reverse primers, respectively, were used: 5' GAG GAG ATC TCT CGA CGC AGG AC3' (HXB2 nucleotides 675- 697) and 5'GAG TAT TGT ATG GAT TTT CAG GCC CAA T 3'(2,696-2,724). The master mix for the RT-PCR reaction consisted of 14.4 µl

Table 2.2 Clinical characteristics of the UARTO cohort

Characteristics	Median (interquartile ranges)
CD4 T cell count	164 (112-216) cells/mm^3
Viral load	4.98 (4.48-5.71) log_{10} copies/ml

Figure 2.1 Generation of Gag-protease NL4-3 recombinant virus stocks

Diagram showing the recombination of patient-derived HIV-1 *gag-protease* PCR product and *gag-protease*-deleted viral NL4-3 backbone (pNL4-3Δ*gag-protease*) to produce recombinant viruses.

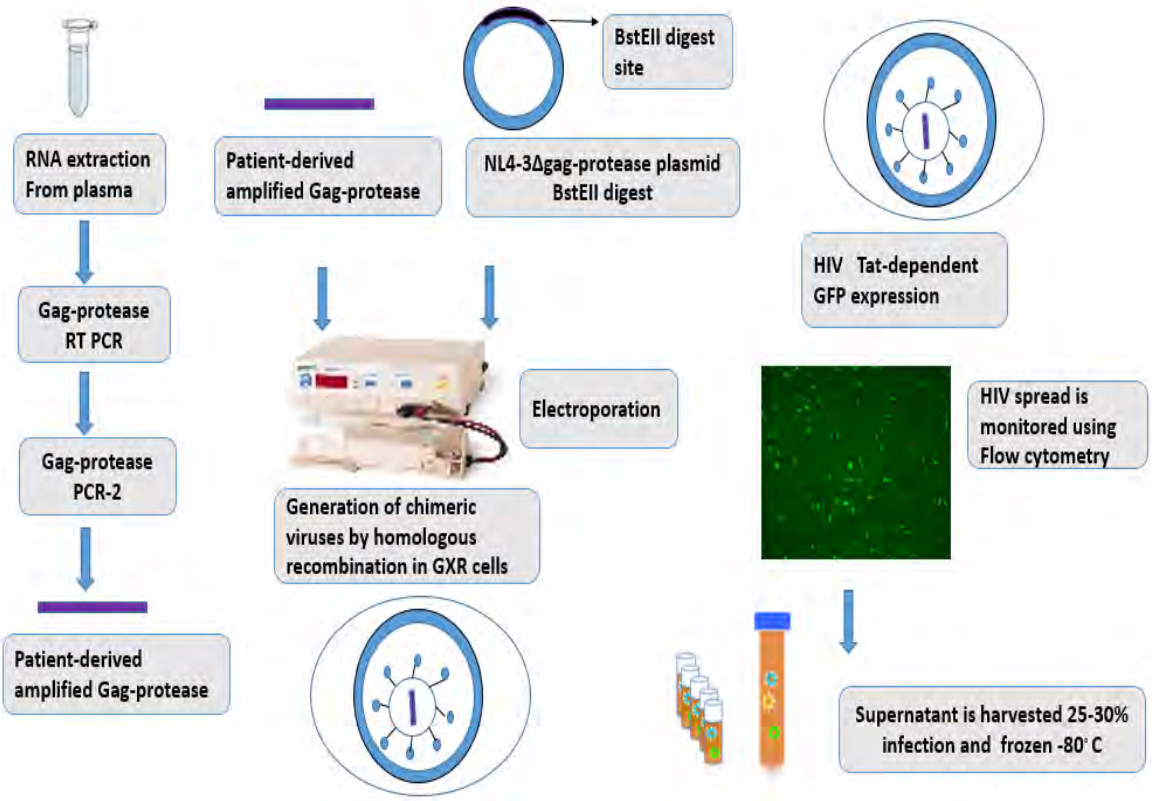

RNA extraction From plasma
Gag-protease RT PCR
Gag-protease PCR-2
Patient-derived amplified Gag-protease
Patient-derived amplified Gag-protease
BstEII digest site
NL4-3Δgag-protease plasmid BstEII digest
Electroporation
Generation of chimeric viruses by homologous recombination in GXR cells
HIV Tat-dependent GFP expression
HIV spread is monitored using Flow cytometry
Supernatant is harvested 25-30% infection and frozen -80° C

Figure 2.2 Summary of the extraction process

The flow chart summarizes the RNA extraction process using the Nucleospin RNA virus kit from Machery-Nagel (taken from the kit manual).


plasma
serum
+ carrier RNA
lysis
with GITC + ethanol
binding
washing (2 buffers)
elution
with low-salt buffer
or RNase-free water
ready-to-use viral RNA

diethylpyrocarbonate (DEPC) treated water (Invitrogen), 20 µl 2X buffer, 0.8 µl of each primer (10 µM), 0.8 µl RT/Taq enzyme mix, and 4 µl RNA extract. The reaction mixture was then run in a thermocycler (Gene Amp PCR system 9700, Applied Bio systems, Foster City, USA) at 30 °C for 55 minutes (cDNA synthesis) and 94 °C for 2 minutes (initial denaturation), followed by 35 cycles of 94 °C for 15 seconds (denaturation), 55 °C for 30 seconds (primer annealing), and 72 °C for 2 minutes (extension), and ended with a 7 minute incubation at 72 °C (final extension). Once the RT-PCR program was completed, the samples were held in the thermocycler at 8 °C for at least 10 minutes (on hold). Nested PCR was performed using a TaKaRa Ex Taq HS enzyme kit (Takara, Shiga, Japan) and 100-mer forward (5' 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 3'; HXB2 nucleotides 695–794) and reverse (5' GGC CCA ATT TTT GAA ATT TTT CCT TCC TTT TCC ATT TCT GTA CAA ATT TCT ACT AAT GCT TTTATT TTT TCT TCT GTC AAT GGC CAT TGT TTA ACT TTT G 3'; 2,706-2,805) primers that were exactly complementary to NL4-3 on either side of *gag-protease*, resulting in an overlap that was necessary for recombination of the *gag-protease* PCR product and pNL4-3Δ*gag-protease* following cotransfection. Two 50 µl PCR reactions were prepared per sample. Each reaction consisted of 37 µl DEPC water, 5 µl 10X Ex Taq buffer, 4 µl deoxyribonucleotide triphosphate (dNTP), 0.8 µl forward primer (10 µM), 0.8 µl reverse primer (10 µM), 0.25 µl Ex Taq and 2 µl RT-PCR product. The mixture was then incubated in a thermocycler at 94 °C for 2 minutes (initial denaturation), 35 cycles of 94 °C for 15 seconds (denaturation), 55 °C for 30 seconds (primer annealing), and 72 °C for 2 minutes (extension), followed by 72 °C for 7 minutes (final extension). Once the nested PCR was completed, the amplicons from the two 50 µl PCR reactions were combined and stored at -20 °C. Gel electrophoresis was used to confirm *gag-protease* amplification. The 1% polyacrylamide gel was prepared by adding one 0.5 g agarose tablet (Bioline, Taunton, USA) to 50 ml of 1X TBE buffer (Sigma-Aldrich, St. Louis, USA; composition: 89 mM tribase, 89 mM boric acid, and 2 mM ethylenediaminetetraacetic acid [EDTA]) in a flask. A low DNA mass ladder (Roche, Penzberg,

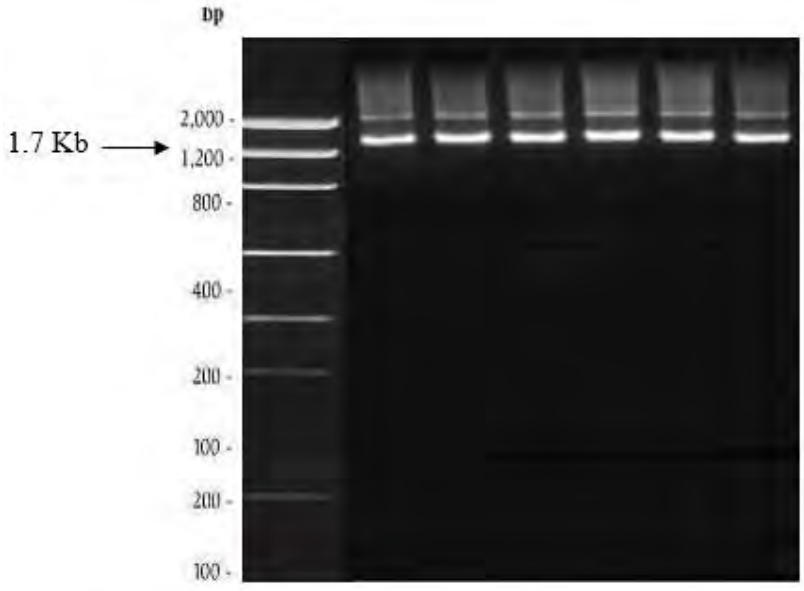
Germany) and 5 µl PCR product and 2 µl of gel loading buffer mix which comprised of Gelred Nucleic Acid Stain (Biotium, Hayward, USA) and loading dye (Sigma-Aldrich) at a ratio of 1:4) were run on the gel at 120 V, 500 mA for 30 minutes using an Electrophoresis Power Supply - EPS 301 (Amersham Biosciences, Uppsala Sweden). The gel was viewed under ultraviolet (UV) light using the GelVue UV Trans illuminator (SynGene, London, United Kingdom). The *gag-protease* PCR product corresponded to a size of approximately 1.7 kb as shown in Figure 2.3. Approximately 5 µl of the product was set aside for bulk sequencing and the remaining 80-90 µl of PCR product was used in the generation of recombinant viruses.

2.2.3 Sequencing of patient-derived gag-protease and sequence analysis

PCR products were diluted 1:15 in DEPC water for bulk sequencing using the Big Dye ready reaction termination mix V3 (Applied Biosystems) and the following sequencing primers: 5' GAC GCA GGA CTC GGC TTG CTG A 3' (688-710), 5' TTA TCT AAA GCT TCC TTG GTG TCT 3' (1,073-1,097), 5' CAG CAT TAT CAG AAG GAG CCA C 3'(1,307-1,329), 5' GGT TCT CTC ATC TGG CCT GGT 3' (1,461-1,482), 5' TGA CAT GCT GTC ATC ATT TCT TCT A 3' (1,816-1,841), 5' GAA GGG CAC ACA GCC AGA AAT TGC 3'(1,981-1,953), 5' GAT AAA ACC TCC AAT TCC3' (2,396-2,414), 5' TCT TCT GTC AAT GGC CAT TGT TTA AC 3'(2,609-2,635). The sequencing reaction was performed in a Micro Amp optical 96-well plate (Applied Biosystems). Each well consisted of 0.4 µl big dye mix, 2.6 µl sequencing primer, 2 µl sequencing buffer, 4 µl DEPC water, and 1 µl diluted template. The reaction was incubated at 96 °C for 1 minute (initial denaturation), followed by 25 cycles of 96 °C for 10 seconds (denaturation), 50 °C for 5 seconds (primer annealing), and 60 °C for 4 minutes (extension). Once the sequencing reaction was complete, the products were immediately purified. Briefly, the sequencing product was firstly diluted with 1 µl of 125 mM EDTA solution at pH 8.0 (Sigma-Aldrich). Thereafter, 26 µl of a mixture of 1 µl 3M sodium acetate (NaOAc) at pH 5.2 (Sigma-Aldrich) and 25 µl 99% chilled ethanol solution was added to each well, followed by

Figure 2.3 Image showing an agarose gel loaded with low DNA mass ladder and gag-protease PCR products

A representative image of six *gag-protease* amplicons corresponding to approximately 1.7 kb.


bp
1.7 Kb
2,000
1,200
800
400
200
100
200
100

centrifugation (Eppendorf centrifuge 5810R, Merck, Germany) at 3,000 g for 20 minutes. The products were dried by inverting the plate on paper towel and then centrifuging at 150 g for 1 minute. Subsequently, 35 µl of 70% ethanol was added to the products to resuspend the pellet followed by centrifugation at 3,000 g for 5 minutes. Following inversion and centrifugation at 150 g for 1 minute, products were dried in a thermocycler at 50 ℃ for 5 minutes and then stored at -20 ℃ until analysis. The sequencing products were suspended in HiDi formamide and denatured in a thermocycler at 95 ℃ for 3 minutes followed by cooling at 4 ℃ for 3 minutes before being run on an ABI 3130 XL Genetic analyzer (Applied Biosystems). Sequences were edited in Sequencher version ®5.1 (Gene Codes, Corporation, Ann Arbor, MI USA; (http://www.genecodes.com) and aligned to the HIV-1 subtype B reference strain HXB2 (Genbank accession number K03455) using the Gene Cutter tool from Los Alamos HIV sequence database (http://www.hiv.lanl.gov/content-sequence/GENECUTTER/cutter.html). Insertions with respect to HXB2 were stripped out manually in Bioedit Version 7.2.5 [5]. Classifications of the subtypes were based on the *gag* gene using the Rega HIV1 subtyping tool from the Stanford database (http://dbpartners.stanford.edu:8080/RegaSubtyping/stanfordhiv/typingtool).

A maximum likelihood phylogenetic tree was drawn using the Phyml tool from the Los Alamos HIV sequence database (http://www.hiv.lanl.gov/content/sequence/PHYML/interface) [326].

The tree was rooted using a subtype B reference sequence (GenBank accession number DQ383746). All reference subtypes and circulating recombinant forms were obtained from the Los Alamos HIV Sequence Database and were incorporated to construct the tree. The recombination breakpoints were determined by Simplot version 3.5.1 software of the Phylip package program (http://sray.med.som.jhmi.edu/SCRoftware/simplot).

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

The pNL4-3Δ*gag-protease* plasmid in STBL3 cells was donated by Dr. Toshiyuki Miura (University of Tokyo, Japan). pNL4-3Δ*gag-protease* was previously prepared by the introduction of the unique restriction enzyme site BstEII at the 5' end of the *gag* gene and then 45 bases downstream from the 3'end of the *protease* gene of pNL4-3, followed by BstEII digestion to remove *gag-protease* and self-ligation of the plasmid [317]. To propagate the plasmid, 100 ml of Luria-Bertani (LB) broth (Sigma-Aldrich) with 100 µg/ml ampicillin was inoculated with 25 µl STBL3 stock and incubated for 12-16 hours at 37 °C at 230 rpm in a shaking incubator (Infors HT, Bottmingen, Switzerland). The plasmid was then purified using the Plasmid Maxi kit (Qiagen, Valencia, USA) as per manufacturer's instructions. The purified DNA was later quantified using a Nano drop spectrophotometer (Thermo Scientific) and stored in a -80 °C ultra-freezer (Snijders Scientific, Holland) until use.

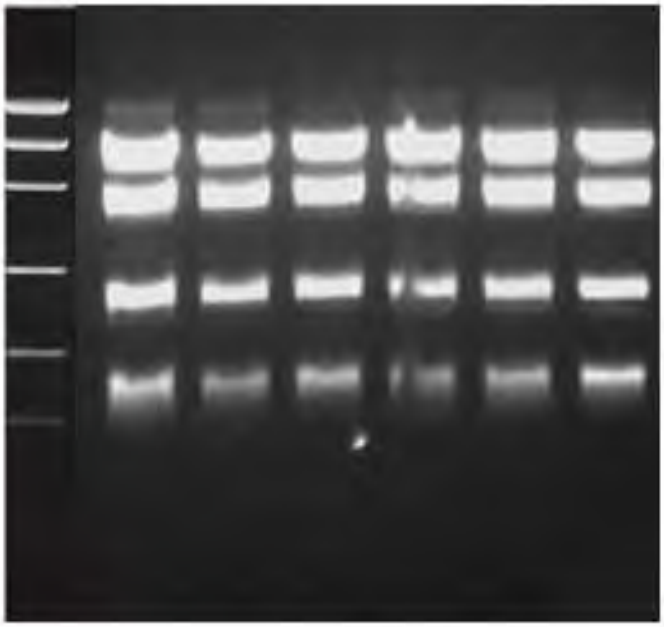
2.2.4.1 Confirming specificity of the purified DNA product

The specificity of the purified plasmid was then confirmed by digestion with the restriction endonuclease Hind III (Thermo scientific). Restriction endonucleases cleave the double–stranded (dsDNA) at specific sites within their recognition sequences. Briefly, the reaction consisted of 15 µl of nuclease free water, 2 µl of 10x fast digest buffer, 1 µl of fast digest enzyme and 2 µl of diluted (up to 1 µg) purified plasmid product. The mixture was gently mixed, spun down and then incubated at 37 °C in a thermocycler for 30 minutes. Following gel electrophoresis of the digested plasmid 4 bands were observed as shown in Figure 2.4.

-

Figure 2.4 An agarose gel loaded with low DNA mass ladder and HindIII digested pNL4-3Δgag-protease

A gel confirming the specificity of the maxiprepped pNL4-3Δ*gag-protease*. Four bands are observed following HindIII digest**.**



2.2.5 Thawing, culturing, and storing of CEM-GXR25 cells

The reporter T-cell line CEM–GXR25 (GXR cells) donated by Dr Mark Brockman (Simon Fraser University, Canada) was used [6]. This cell line encodes for green fluorescent protein (GFP) which is driven by HIV-1 long terminal repeat (LTR). The LTR plays a major role in viral gene expression. Upon HIV infection, GFP expression is induced in a Tat-dependent manner and HIV1 spread is monitored by flow cytometry. A frozen 1 ml aliquot of GXR cells stock was transported from the liquid nitrogen freezer (Custom Biogenics Systems, Romeo, USA) on dry ice and incubated in a 37 °C water bath by swirling until thawed. Afterwards, the thawed cells were moved to 4 ml of pre-warmed R10 medium (RPMI-1640 supplemented with 2 mM L-glutamine [both from Sigma-Aldrich], 50 U/ml penicillin-streptomycin, 10% foetal calf serum, and 10 mM HEPES [all from Gibco, New York, USA]) in a T25 flask (Corning-Costar, New York, USA**)** and incubated in a humidified Heraeus incubator (Thermo Scientific) at 37 °C and 5% CO_2. The next day, the GXR cells were spun at 1,500 rpm (radius of rotor 100 mm) for 10 minutes (Heraeus multifuge 3SR+, Thermo Scientific) and then suspended in 10 ml fresh R10 medium to remove dimethylsulfoxide (DMSO; Sigma-Aldrich) used in the storing of cells. Cell counts were performed by mixing 10 µl cell suspension with 10 µl of trypan blue dye, loading this onto a slide (Bio-rad, South Africa) and reading the count using an automated TC20 cell counter (Bio-rad). On growing to approximately 90% viability, the cells were transferred to a T75 flask (Corning-Costar) and a stock culture was maintained in 25 ml R10 medium for maximum of three months. Every two days approximately 20% of the culture medium was discarded and replaced with fresh medium. GXR cells were stored in 1 ml R10 medium at a concentration of approximately 5 million per ml in 10% DMSO. Following approximately 20 minutes of equilibration in DMSO, cells were placed in a Mr Frosty container (Nalgene, Thermo Scientific), which insulates the cells making them to cool at a rate of approximately 1 °C per minute, in the -80 °C freezer (Snijders Scientific) overnight and thereafter in liquid nitrogen until use.

2.2.6 Co-transfection of GXR cells with *gag-protease* and pNL4-3Δ*gag-protease*

The purified plasmid (10 µg per sample) was digested with BstEII (Promega, Madison, USA) to obtain a linear DNA fragment. Digestion was performed on the morning of the co-transfection experiment to reduce re-ligation of the restriction site [327]. The digestion reaction consisted of the plasmid, 1/10 reaction volume of the 10X buffer, and 1/100 reaction volume of bovine serum albumin (BSA) and was made up to the final reaction volume with sterile deionized water. The reaction was incubated in a water bath at 60 °C for 2 hours. The *gag-protease* PCR amplified fragment (80-90 µl) was then co-transfected with the linearized pNL4-3Δ*gag-protease* (10 µg digested plasmid) into GXR cells (4 million cells per sample) by electroporation (Gene Pulser II, Bio-rad, Hercules, USA) in 4 mm cuvettes (Bio-rad, South Africa) at 250 V and 950 µF with resistance at ∞ (infinity). The electric field induces pore formation in the cell wall and enhances the permeability of the host cells, which allows for the uptake of DNA [328, 329]. The cells were then rested for 5 minutes before being moved to a T25 flask with 10 ml of prewarmed R10 medium (supplemented with 4 µg/ml of polybrene [EMD Millipore, Darmstadt, Germany], which enhances virus infection of cells) and 1 million non-electroporated GXR cells. The cells were then incubated at 37 °C and 5% CO_2 and 5 days later 5 ml of R10 media was added to the flask followed by further incubation.

2.2.7 Monitoring of virus infection by flow cytometry

The percentage infected cells was monitored quantitatively using a FACS Calibur flow cytometer (Becton Dickinson, San Jose, USA). Flow cytometry measures multiple characteristics of individual cells flowing in single file in a stream of fluid. Light scattering at different angles can distinguish differences in size and internal complexity, and light is also emitted from fluorescent molecules [330]. On the 10^{th} day post cotransfection, and every 2 days thereafter, HIV-induced expression of GFP was quantified as described in published protocols [322]. Briefly, the culture suspension was well mixed, after which 2 ml was removed and replaced with fresh R10 medium.

1 ml was discarded and the remaining 1 ml was transferred to a matrix cluster tube (Corning-Costar) followed by centrifugation at 1,500 rpm for 10 minutes at 20 °C. The excess supernatant was aspirated leaving approximately 50 µl with the pellet, which was then fixed with 200 µl of 2% paraformaldehyde (PFA) for 15 minutes at room temperature before it was read on the FACS calibur. Instrument settings used were: forward scatter = E-1, side scatter = 411 FL1 = 2806, FL2 = 493, and FL3 = 721. The gating for GFP (using the uninfected negative control sample) was set at 0.05, above which cells were considered GFP positive. 15,000 events were acquired. Data analysis to accurately determine the amount of GFP positive cells was conducted using FlowJo version 8.1 (TreeStar, USA) (http://www.flowjo.com/v8/html/reference.html). When cells reached ~30% infection, the culture supernatant was harvested.

2.2.8 Harvesting of the recombinant virus

At ~30% GFP positive (infected) cells, the culture was centrifuged at 1,700 rpm for 5 minutes at 4 °C. The supernatant was aliquotted and stored at -80 °C prior to use in assays.

2.3 Viral replication assay

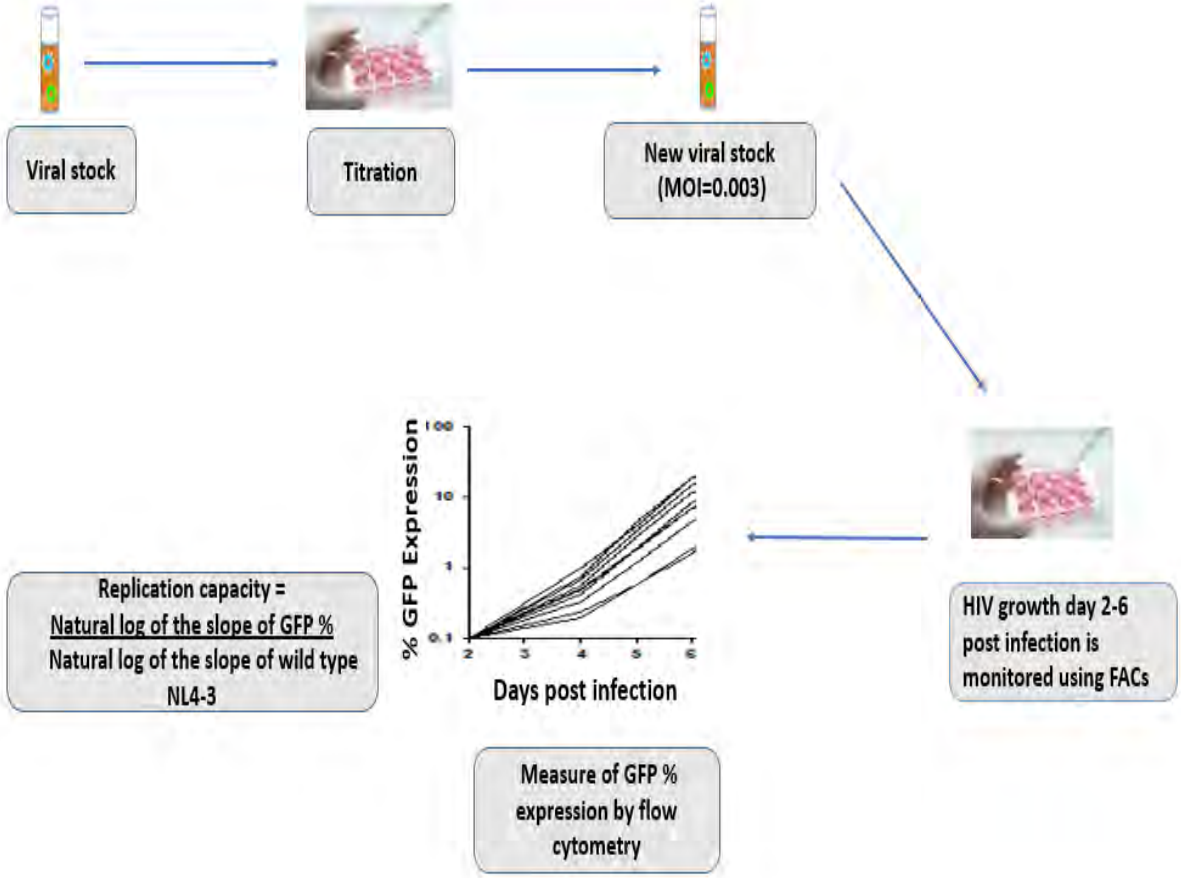
The replication assay was performed in duplicate to confirm reproducibility. Prior to this assay all the recombinant viruses were titered. These methods were previously established in our laboratory [325] as described briefly below. The process is summarized in Figure 2.5.

2.3.1 Titration of the viruses

The objective of the replication assay is to measure the ability of recombinant viruses generated to reproduce by assessing multiple rounds of infection. It is therefore paramount to ensure that all viruses start at a similar infection level, which we set at a multiplicity of infection (MOI) of 0.003 i.e. 0.3% infection after 2 days. To determine the volume of virus stock that would yield 0.3% infection after 2 days, we infected 1 million GXR cells (corresponding to 0.1 ml) with 0.4 ml

Figure 2.5 Replication assay summary

The diagram shows a summary of the titration and replication assay.


Viral stock
Titration
New viral stock (MOI=0.003)
HIV growth day 2-6 post infection is monitored using FACs
Measure of GFP % expression by flow cytometry
% GFP Expression
Days post infection
Replication capacity =
Natural log of the slope of GFP %
Natural log of the slope of wild type NL4-3

virus stock in a 24 well plate (Corning-Costar) followed by incubation at 37 °C and 5% CO_2. 24 hours later, 1 ml of pre-warmed R10 medium was added to each well and incubation was continued under same conditions. At 48 hours post-infection, the percentage of infected cells was determined by flow cytometry as described in Section 2.2.7 (with the exception that 0.5 ml instead of 1 ml was prepared for flow cytometry). The volume of virus stock to use in the replication assay was calculated as follows:

Volume of virus stock (ml) = (0.3%/% cells infected on day 2 of the titration)*0.4 ml.

2.3.2 Replication capacity assay

To perform the replication capacity assay, a new viral stock was thawed and the calculated volume of virus stock (from the titration experiment, Section 2.3.1) was diluted to a final volume of 0.4 ml with R10 medium. As described in Section 2.3.1, 1 million GXR cells in 0.1 ml of R10 medium were infected with the 0.4 ml diluted virus stock. The assay was set up in a 24 well culture plate (Corning-Costar) and was incubated at 37 °C and 5% CO_2 for 6 days. At 24 hours post-infection, 1 ml of fresh R10 medium was added to each well. On days 2-6 post-infection, 0.5 ml of the culture was prepared for flow cytometry (as described in Section 2.2.7) and was replaced with 0.5 ml fresh R10 medium. Wild-type NL4-3 virus was included as a positive control in each assay (the average of three measurements was taken in each assay). The negative control was 1 million uninfected cells. The replication capacity was defined as the natural log of the slope of exponential increase in GFP percentage monitored on day 3 to 6 post-infection, expressed relative to the wild type NL4-3 growth, by employing the semi-log method in MS Excel. To ensure calculation of the slope only during the period of exponential viral spread, if the percentage infected cells exceeded a cut off of 11% on day 5 post-infection, the slope was calculated from day 3-5 post-infection instead.

2.4 Data analysis

Statistics were performed and graphs drawn using GraphPad Prism version 5.01 (GraphPad Software, Inc., San Diego, CA, USA) unless otherwise stated. For all analyses, $p<0.05$ was considered significant. The clinical parameters and demographics were compared between the patient groups infected with different subtypes using the Kruskal-Wallis test. The replication capacity frequency distribution of the entire cohort, as well as grouped according to subtype, was analysed using histograms on the IBM SPSS statistics program 23. To confirm reproducibility of duplicate replication measurements and concordance between our replication measurements with those of previously assayed samples, we used Pearson's correlation tests. Analysis of variance (ANOVA) was used to assess whether there were significant differences in replication capacities between the different subtypes. Tukey post- hoc tests were used to compare individual subtypes, with $p<0.05$ (*), $p<0.01$ (**) and $p<0.001$ (***) considered for the level of significance. Students T tests or Mann-Whitney U tests (when the data failed assumptions of the Students T test) were used to compare the replication capacities of viruses when divided into 2 groups based on different parameters (for example geographical location or the presence/absence of a mutation). Spearman's correlation tests were used to correlate viral replication capacities with several parameters including the percentage similarity of sequences to the consensus B sequence, markers of disease progression (namely CD4 counts and viral loads) and the number of HLA-associated polymorphisms present in sequences. To assess whether there were significant differences in replication capacities within expressed HLA-A, -B and –C allele groups, ANOVA was used. Student's t-tests were employed for each individual allele to compare replication capacities of viruses from hosts expressing the particular allele and those not expressing that allele. Finally, to assess relationship between individual amino acids and replication capacity in our cohort, an exploratory codon-by-codon tool that uses the Kruskal-Wallis test was employed using a cut off of $q<0.2$ (http://brockman-srv.mbb.sfu.ca/~B_Team_iMac/Codon_by_codon) [331]. Briefly, for each coordinate, the algorithm identifies an amino acid that satisfies the minimum count. (The

minimum count is used to determine whether or not to test the cases in the test group). A minimum count of 3 is provided by the tool by default in order to achieve a p value of less than 0.05. This tool independently looks at each coordinate and determines whether the presence or absence of a particular amino acid is significantly associated with change in replication capacity functionality. The Chi square test was used [332] to test for significant differences in the frequencies of amino acid variants associated with altered replication capacity between subtypes A and D.

CHAPTER 3

RESULTS

CHAPTER 3 -RESULTS

Introduction

The purpose of this study was to generate Gag-protease NL4-3 recombinant viruses using plasma samples from ARV naïve patients in the Majengo sex worker cohort from Nairobi, Kenya; where A, C, and D subtypes as well as inter-subtype recombinants co-exist. The replication capacity of each virus was measured using a standardized GFP based reporter T cell line assay. We investigated whether there were differences in Gag-protease driven viral replication capacities according to HIV-1 subtype. We then associated the replication capacities of the recombinant viruses with markers of disease progression and host HLA class I types. We also identified specific amino acids associated with decreased or increased viral replication capacity.

3.1 Subtype classification

Viral RNA was extracted from patient plasma samples and the *gag-protease* region of the HIV-1 genome was amplified by nested RT-PCR. The patient-derived *gag-protease* amplicons were then sequenced using Sanger bulk sequencing for HIV-1 subtyping. Classifications of the subtypes were based on the *gag* gene using the Rega HIV-1 subtyping tool from the Stanford database (http://dbpartners.stanford.edu/RegaSubtyping). A maximum likelihood phylogenetic tree drawn using Phyml tool on Los Alamos Database illustrated that the patient-derived sequences clustered with the respective reference sequences for each subtype as shown in Figure 3.1 (http://www.hiv.lanl.gov/content/sequence/PHYML/interface). The tree was rooted using a subtype B reference sequence from Los Alamos. The inter-subtype recombinants were confirmed by boot scanning analysis using Simplot version 3.5.1 [200 base pairs window, 20 base pairs step increment and 50-1000 bootstrap replicates at every window (JC, Kimura (2-parameter) model)] and further assessed for recombination patterns, as illustrated in Figure 3.2. Interestingly, common patterns of recombination breakpoints were observed: specifically, breakpoints at amino acids 150, 410 and 435 occurred 5 times each.

Figure 3.1 Phylogenetic tree showing the subtypes classification

A maximum likelihood phylogenic tree showing clustering of *gag* from all patient-derived sequences into distinct subtypes. Subtype references, A, C, D, and inter-subtype recombinant sequences are represented by black, red, purple, blue and green colours respectively. The scale bar indicates 2% nucleotide sequence divergence.

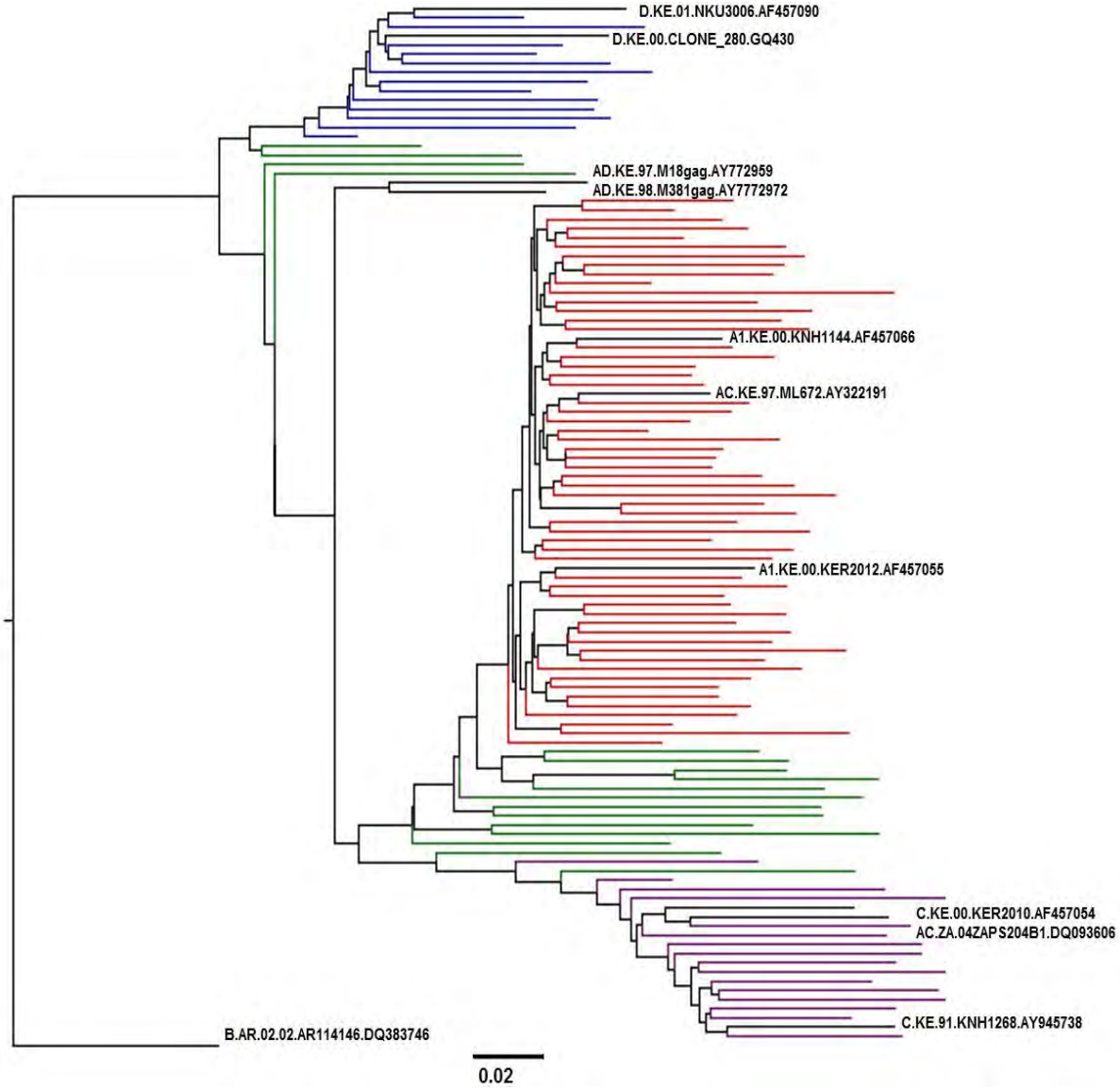
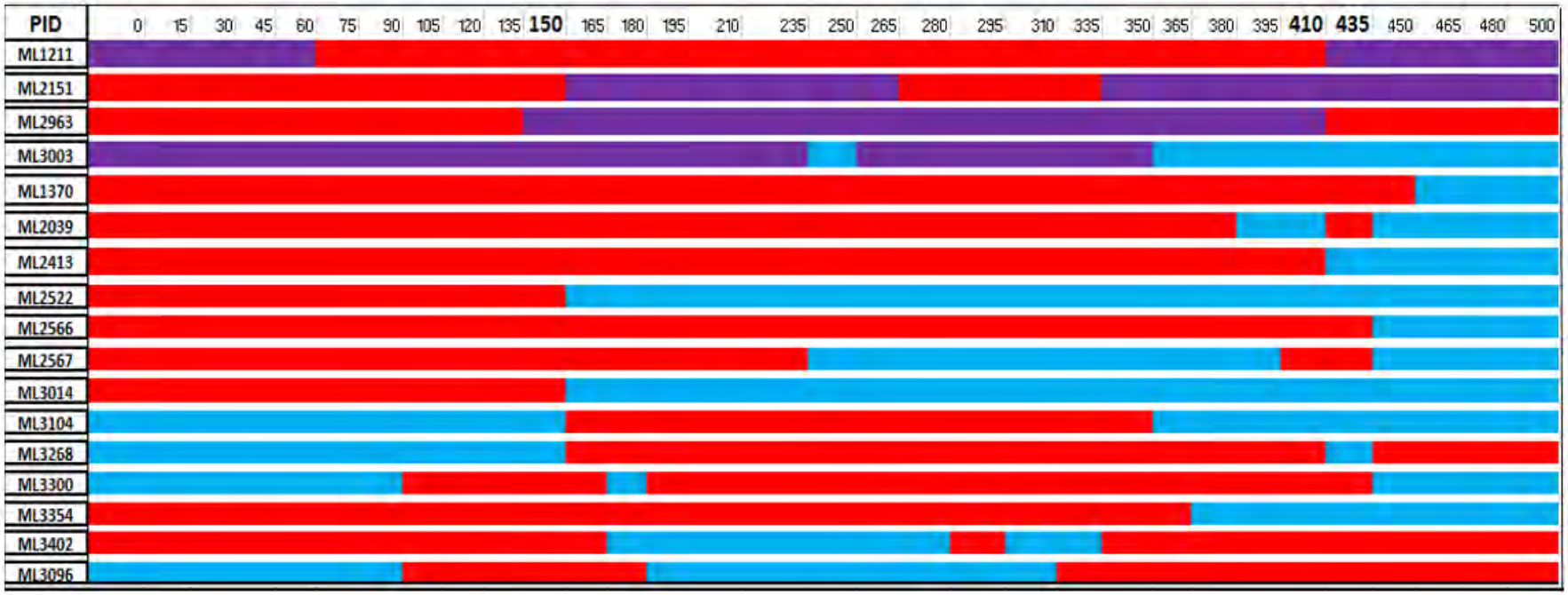

D.KE.01.NKU3006.AF457090
D.KE.00.CLONE_280.GQ430
AD.KE.97.M18gag.AY772959
AD.KE.98.M381gag.AY7772972
A1.KE.00.KNH1144.AF457066
AC.KE.97.ML672.AY322191
A1.KE.00.KER2012.AF457055
C.KE.00.KER2010.AF457054
AC.ZA.04ZAPS204B1.DQ093606
C.KE.91.KNH1268.AY945738
B.AR.02.02.AR114146.DQ383746
0.02

Figure 3.2 Recombination patterns of inter-subtype recombinants

An illustration map showing a distinct pattern of recombination breakpoints of the inter-subtype recombinants. The subtypes A, C and D are represented by red, purple and blue colours respectively. The amino acids codons are positioned according to *gag* HXB2 reference strain numbering. Residues at which breakpoints were common are highlighted in bold.


PID
0 15 30 45 60 75 90 105 120 135 150 165 180 195 210 235 250 265 280 295 310 335 350 365 380 395 410 435 450 465 480 500
ML1211
ML2151
ML2963
ML3003
ML1370
ML2039
ML2413
ML2522
ML2566
ML2567
ML3014
ML3104
ML3268
ML3300
ML3354
ML3402
ML3096

3.2 Demographic and clinical data

The patients were categorized according to infecting subtype (percentage composition of infecting subtypes) as displayed in Figure 3.3 and the demographic and clinical data was compared between subtype groups (Table 3.1) since differences in these parameters may potentially confound analyses comparing replication capacities between subtypes. However, there were no significant differences in age, CD4 counts or viral loads between the different subtype groups (Table 3.1).

3.3 Replication capacity measurements

The amplified patient-derived *gag-protease* amplicons and the *gag-protease* deleted plasmid (pNL4-3Δ*gag-protease*) were co-transfected into CEM-GXR T cell line via electroporation to generate Gag-protease NL4-3 recombinant viruses. The replication capacities of the Gag-protease recombinant viruses were assayed in duplicate independent experiments by infecting the cell line at a multiplicity of infection (MOI) of 0.003 and then calculating the mean slope of exponential growth from day 3-6 post infection normalized to wild-type NL4-3. Duplicate measurements were in strong agreement indicating reproducibility of the assay as shown in Figure 3.4. Furthermore, supporting reproducibility of the assay, we used 20 *gag-protease* amplicons for which Gag-protease replication capacity was previously measured [325], constructed Gag-protease recombinant viruses and found that the replication capacities measured correlated strongly with previous measurements (Pearson's correlation r=0.92, p=0.0001) as shown in Figure 3.5. The overall replication capacities of recombinants viruses encoding patient-derived Gag-protease were gathered to draw a frequency distribution histogram for the 103 patients (Figure 3.6). Histograms of the replication capacities of viruses within subtype A, subtype C, subtype D and inter-subtype recombinant groups were also drawn and showed a normal distribution of replication capacities within each subtype group (Figure 3.7)

Figure 3.3 Distribution of subtypes

A pie chart illustrating the percentage composition of the different subtypes in the 103 study patients. The subtypes A, C, and D and inter-subtype recombinants are indicated by red, purple, blue and green colours respectively.

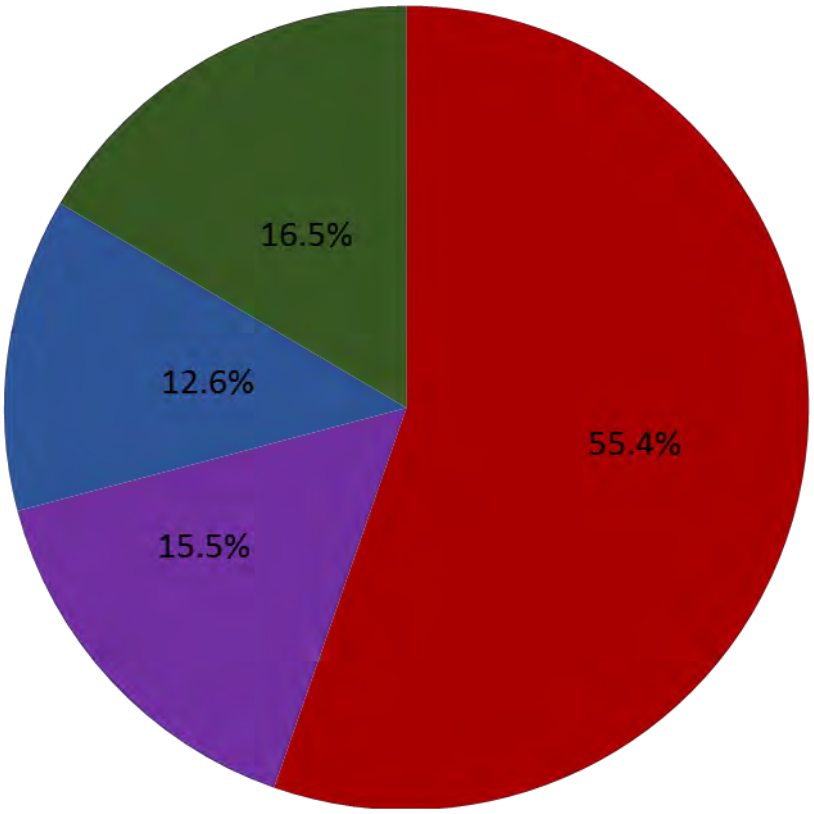

16.5%
12.6%
55.4%
15.5%

Table 3.1 Cohort characteristics and demographic data

HIV-1 Subtype **n=103**	**A** **n=57**	**C** **n=16**	**D** **n=13**	**Recombinants** **n=17**	**p-values**[b]
Age (years; mean)	35	37	31	35	0.39
CD4 count (cells/mm^3) [a]	421 [342-590]	376 [307-471]	427 [356-509]	358 [351- 400]	0.14
Viral load (log_{10} copies/ml)[a]	4.20 [3.71-5.18]	4.04 [3.81-4.95]	5.00 [4.21-5.73]	5.23 [4.65-5.32]	0.16

[a] Medians with inter-quartile ranges in brackets are shown.

[b] p-values calculated using the Kruskal-Wallis test

HIV – human immunodeficiency virus; CD4 – cluster of differentiation 4

Figure 3.4 Quality control: concordance of duplicate measurements

A graph showing good concordance of the duplicate measurements of the replication capacity assay (Pearson's correlation; r=0.98, p=0.0001).

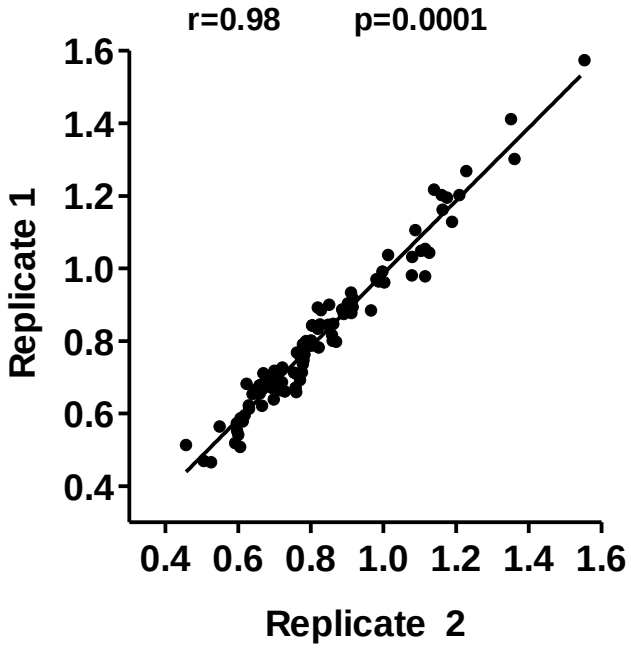

r=0.98
p=0.0001
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Replicate 1
0.4 0.6 0.8 1.0 1.2 1.4 1.6
Replicate 2

Figure 3.5 Quality control: concordance with previous measurements

A graph showing good concordance of replication capacity of previously assayed measurements and replication capacity from same amplicons (Pearson's correlation; r=0.92, p=0.0001). On the X-axis (JM SK) are measurements done previously and on the Y-axis (MWK SK) are repeat measurements from the same samples.

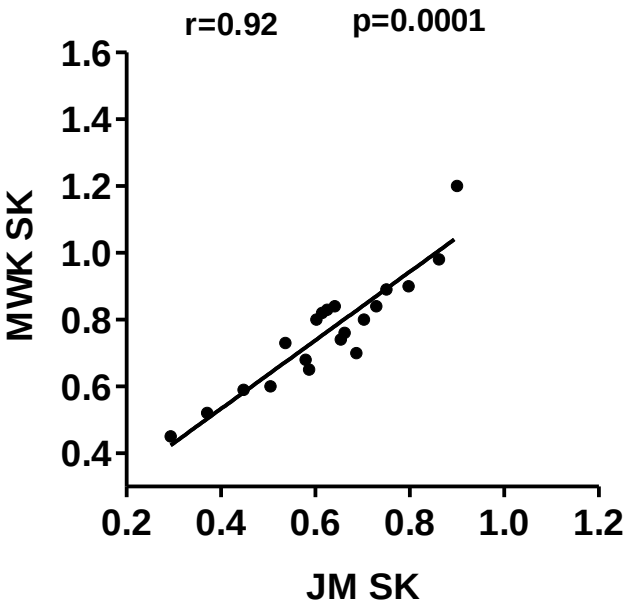

r=0.92
p=0.0001
MWK SK
JM SK
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.4
0.6
0.8
1.0
1.2

Figure 3.6 The frequency distribution of the replication capacities of the entire cohort

A histogram showing a wide range distribution of the viral replication capacities from the 103 patients. The mean replication capacity was 0.81 with a standard deviation of 0.2.

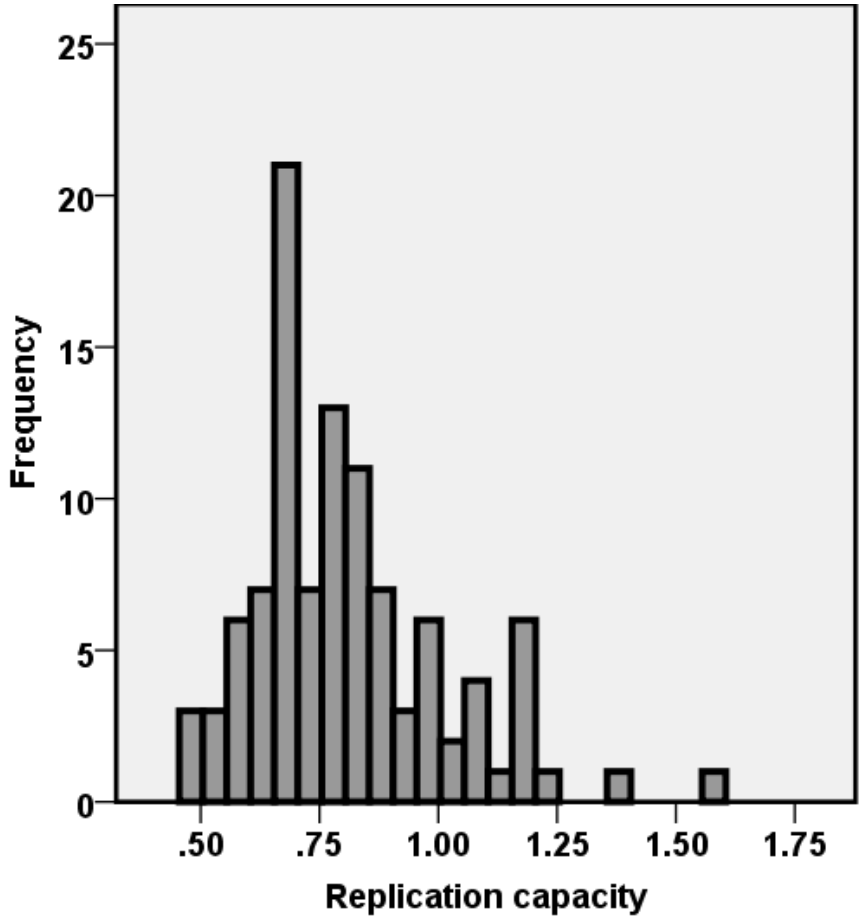
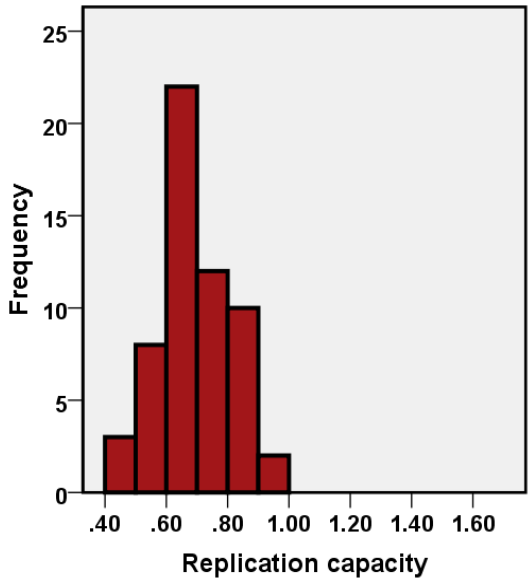

25
20
15
10
5
0
Frequency
.50
.75
1.00
1.25
1.50
1.75
Replication capacity

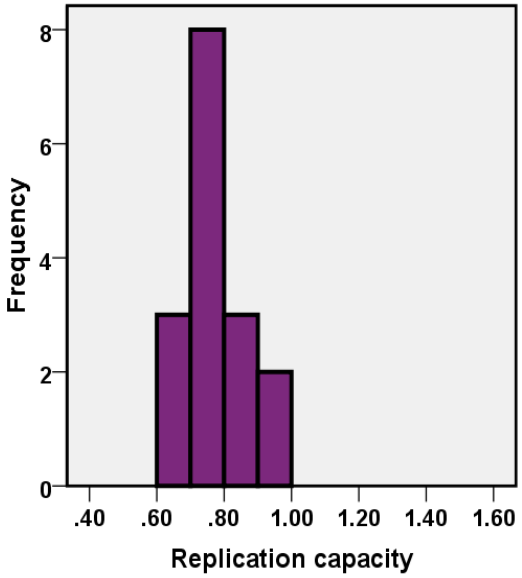
Figure 3.7 Replication capacity frequency distribution among specific subtypes

Histograms (i-iv) showing the distribution of viral replication capacities within subtypes A, C and D as well as inter-subtype recombinants, indicated by red, purple, blue and green colours respectively. Replication capacities approximated a normal distribution with mean replication capacities of 0.69 (standard deviation) (SD=0.1), 0.78 (SD=0.08), 0.94 (SD=0.1) and 1.13 (SD=0.2) respectively.

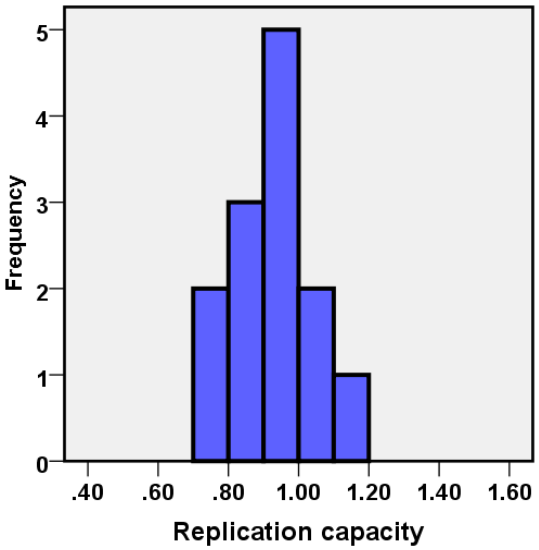
i


Frequency
25
20
15
10
5
0
.40
.60
.80
1.00
1.20
1.40
1.60
Replication capacity

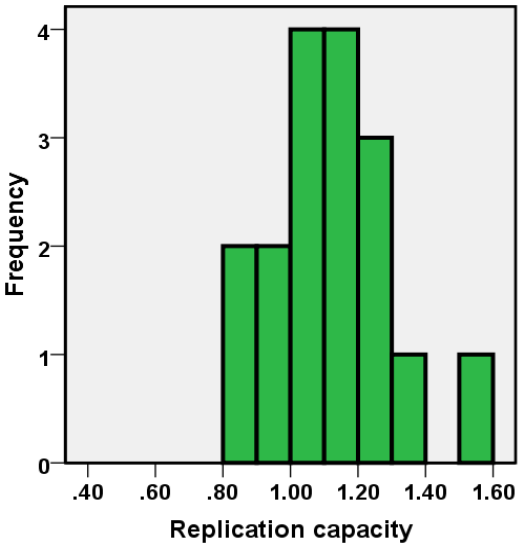
ii


Frequency
8
6
4
2
0
.40
.60
.80
1.00
1.20
1.40
1.60
Replication capacity

iii


Frequency
5
4
3
2
1
0
.40
.60
.80
1.00
1.20
1.40
1.60
Replication capacity

iv


Frequency
4
3
2
1
0
.40
.60
.80
1.00
1.20
1.40
1.60
Replication capacity

3.3.1 Replication capacity differences among subtypes

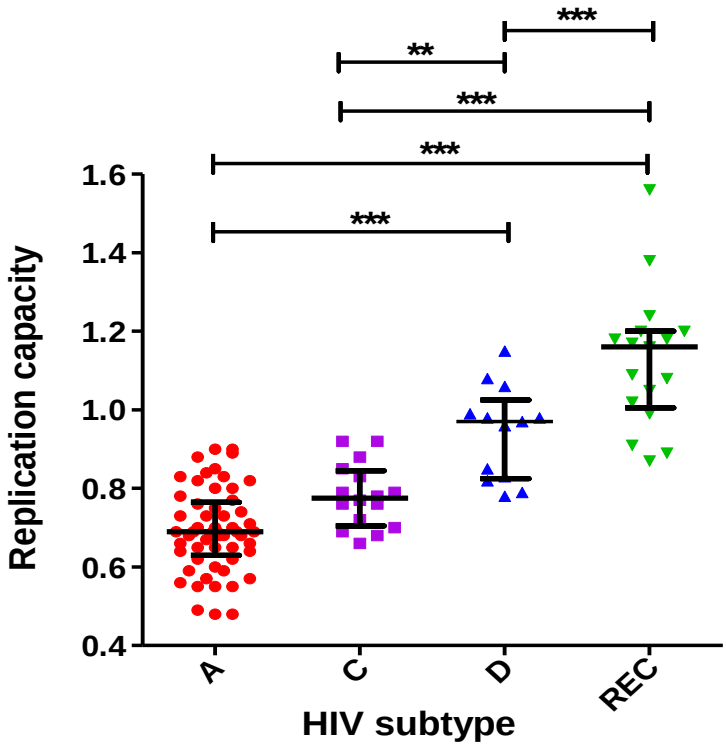
Previous studies have shown differences in disease progression, immunogenicity and transmission among HIV-1 subtypes [105, 239, 242, 243, 245, 333] however the underlying biological mechanisms remains unknown. We investigated whether there were differences in Gag-protease-driven viral replication capacities according to HIV-1 subtype as viral replication capacity is a factor influencing disease progression [325, 334, 335]. Marked differences in replication capacities were observed among subtypes (Figure 3.8). Gag-protease inter-subtype recombinants had the highest mean replication capacity of 1.13 ($p<0.001$ compared to subtypes A, C and D) followed by subtype D with a mean replication capacity of 0.94 ($p<0.001$ compared to subtype A, $p<0.01$ compared to subtype C) then subtype C and A with means of 0.78 and 0.69 respectively. There were no significant differences in CD4 count or viral load between the different subtype groups (Table 3.1), indicating that these factors were not driving the replication differences between subtypes.

3.3.2 Reproducibility of subtype differences

We subsequently wanted to investigate whether similar inter-subtype differences in viral replication capacity would be observed in a different cohort from another region with similar infecting subtypes. Therefore, we generated and assessed the replication capacity of 30 Gag-protease recombinant viruses, 10 samples each from subtypes A, D and AD recombinants, from the UARTO cohort in Uganda. In order to control for CD4 count and viral load as potential confounding factors, the samples from the different subtype groups were matched for CD4 counts and viral loads. There was no significant difference in the CD4 counts (ANOVA; $p=0.53$) or viral

Figure 3.8 Replication capacity differences in subtypes

A graph showing differences in Gag-protease replication capacities among the subtypes and inter-subtype recombinants (ANOVA; p=0.0001). Subtypes A, C, D and inter-subtype recombinants are represented in red, purple, blue and green respectively. The bars and whiskers represent the means and interquartile ranges respectively. The number of asterisks denotes the level of significance, namely, $p<0.05$ (*), $p<0.01$ (**) and $p<0.001$ (***).


***
**
***
***
***
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Replication capacity
A
C
D
REC
HIV subtype

loads (ANOVA; p=0.91) between the different subtype groups. The medians of CD4 T cell counts and viral loads respectively for each group were as follows: 171 cells/mm^3 and 5.20 $\log_{10}$ copies/ml for subtype A, 154 cells /mm^3 and 4.98 $\log_{10}$ copies/ml for subtype D and 147 cells /mm^3 and 4.98 $\log_{10}$ copies/ml for AD recombinants. Consistent with our observations from the Majengo cohort, there were marked differences in replication capacity observed between subtypes shown in Figure 3.9. AD recombinants had the greatest replication capacity of 1.42 ($p<0.01$) compared to subtype A) followed by subtype D with mean of 1.07 ($p<0.05$) compared to subtype A) and lastly subtype A with a mean 0.76. Overall, our replication capacity results for the different subtypes are reproducible across different cohorts and geographical regions. To further assess comparability between cohorts of different geographical regions, we measured the Gag-protease driven replication capacities of 20 subtype C viruses derived from the Sinikithemba cohort in Durban and compared these to the replication capacities of the subtype C viruses generated from the Majengo cohort in Nairobi. Replication capacities of the subtype C viruses from the two geographically diverse regions were similar (Mann-Whitney U test; p=0.75) (Figure 3.10).

3.4 Impact of NL4-3 backbone on replication capacity

One of the limitations of our study is the use of a subtype B backbone (NL4-3) to generate recombinant viruses encoding Gag-protease from different subtypes, which could potentially influence our analysis of inter-subtype replication capacity differences. For instance, a number of studies have previously reported that subtype D is nearer to subtype B than subtypes A or C are to subtype B in terms of sequence phylogeny analysis [319, 320]. Therefore, it would be expected that the subtype D Gag-protease recombinants may be more compatible with the subtype B backbone resulting in fitter viruses compared to subtype A or C Gag-protease recombinants. We sought to exclude any concerns that this factor could have influenced the differences we observed in replication capacity between the subtypes in our cohort. Firstly, we compared the

Figure 3.9 Reproducibility of subtype differences in the UARTO cohort

A graph showing significant differences in Gag-protease viral replication capacities among the subtypes A, D and inter-subtype recombinants (ANOVA; $p=0.0029$) represented in red, purple and green respectively. The bars and whiskers represent the means and interquartile ranges respectively. The number of asterisks denotes the level of significance, namely, $p<0.05$ (*), $p<0.01$ (**).

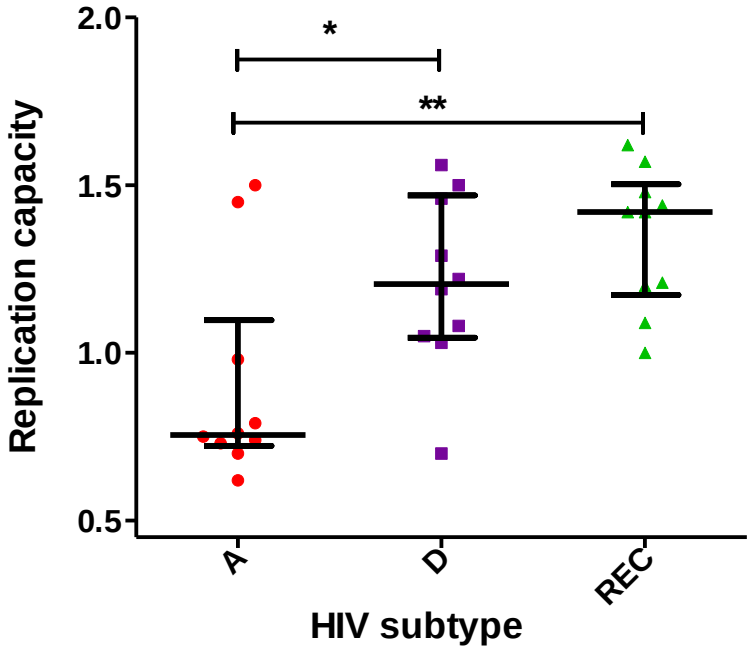
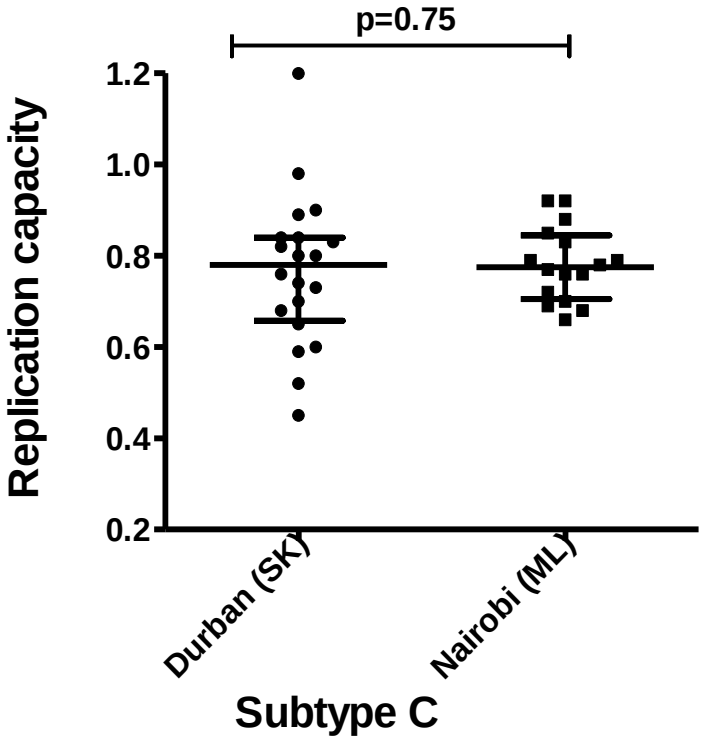

2.0
1.5
1.0
0.5
Replication capacity
*
**
A
D
REC
HIV subtype

Figure 3.10 Replication capacity comparison of subtype C in different cohorts

Graph showing no significant differences between Gag-protease driven replication capacities of subtype C viruses from Sinikithemba cohort (SK) in Durban and Majengo cohort (ML) in Nairobi (Mann-Whitney U test; p=0.75). The bars and whiskers represent the means and interquartile ranges respectively.


p=0.75
Replication capacity
1.2
1.0
0.8
0.6
0.4
0.2
Durban (SK)
Nairobi (ML)
Subtype C

replication capacities of subtype D recombinants versus non-subtype D recombinants to investigate whether the subtype B NL4-3 backbone could bias subtype D recombinants in favour of higher replication capacity. However, we established that there were no significant differences in replication capacity of the subtype D recombinants and non-subtype D recombinants (Mann-Whitney U test; p=0.73) (Figure 3.11). Secondly, we expected that if the subtype B NL4-3 backbone influenced replication capacity results that the viruses most similar to the consensus B sequence would be fitter [319, 320]. We therefore calculated the similarity of Gag from all patient-derived sequences to the consensus B Gag sequence and correlated the percentage similarity with the replication capacities of the corresponding viruses (Figure 3.12). No significant correlation was observed between these two parameters (Spearman's correlation; r=0.13, p=0.34 for subtype A; r=0.44, p=0.09 for subtype C; r=-0.32, p=0.29 for subtype D and r=0.14, p=0.63 for inter-subtype recombinants), supporting that the subtype B backbone was not driving the differences in Gag-protease driven replication capacity observed between viruses encoding different subtypes in our cohort.

3.5 Replication capacity versus markers of disease progression

Previous studies of HIV-1 subtypes B and C [318, 327, 337, 338] have demonstrated an association between Gag-protease replication capacity and markers of disease progression. We therefore assessed the relationship between Gag-protease driven replication capacity and markers of disease progression, namely CD4 counts and viral loads, in the different subtypes. As shown in Figure 3.13, overall there were no significant correlations observed between replication capacity and CD4 T cell counts (Spearman's correlation; r=0.16 and p=0.10 or viral loads (Spearman's correlation; r=0.08 and p=0.40). However, since we had shown significant differences in Gag-protease replication capacity among the subtypes we further assessed the correlation of replication capacity and markers of disease progression by subtype.

Figure 3.11 Replication capacities of subtype D and non-subtype D recombinants

Graph showing no significant differences in replication capacities between subtype D and non-subtype D inter-subtype recombinants (Mann-Whitney U test; p=0.73). The bars and whiskers represent the means and interquartile ranges respectively.

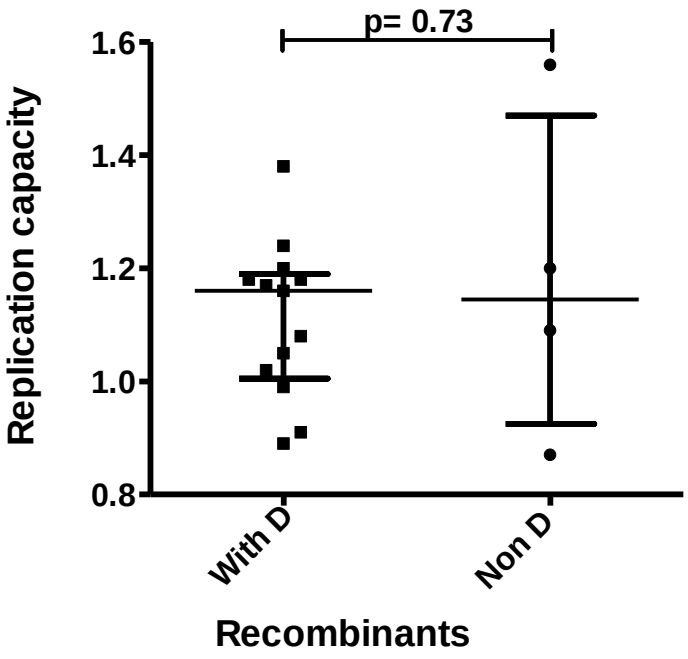
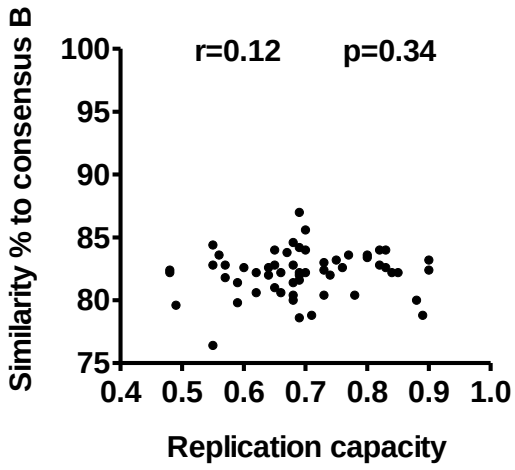

p= 0.73
1.6
1.4
1.2
1.0
0.8
Replication capacity
With D
Non D
Recombinants

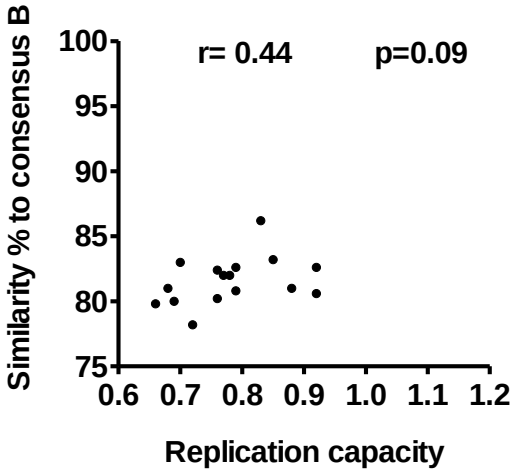
Figure 3.12 Similarity to consensus B versus replication capacity

Graphs showing no correlation between percentage similarities of Gag sequences to the consensus B Gag sequence and replication capacity (Spearman's correlation). Subtype A, subtype C, subtype D and inter-subtype recombinants are represented in panels (i-iv) respectively.

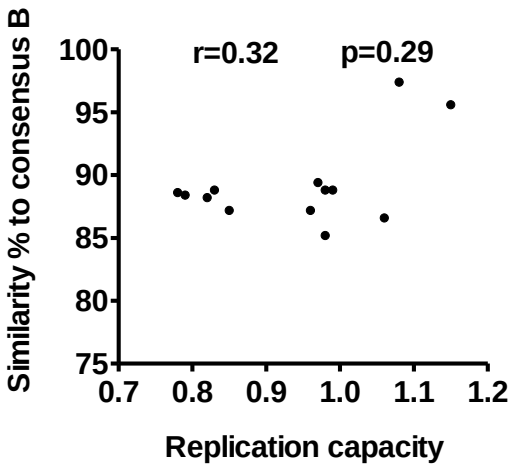
i


r=0.12
p=0.34
Similarity % to consensus B
100
95
90
85
80
75
0.4 0.5 0.6 0.7 0.8 0.9 1.0
Replication capacity

ii


r= 0.44
p=0.09
Similarity % to consensus B
100
95
90
85
80
75
0.6 0.7 0.8 0.9 1.0 1.1 1.2
Replication capacity

iii


r=0.32
p=0.29
Similarity % to consensus B
100
95
90
85
80
75
0.7 0.8 0.9 1.0 1.1 1.2
Replication capacity

iv

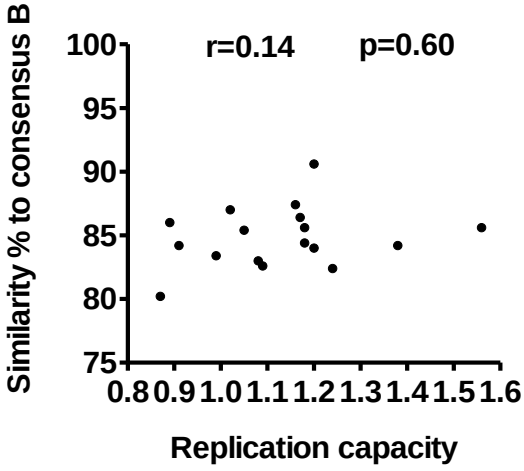
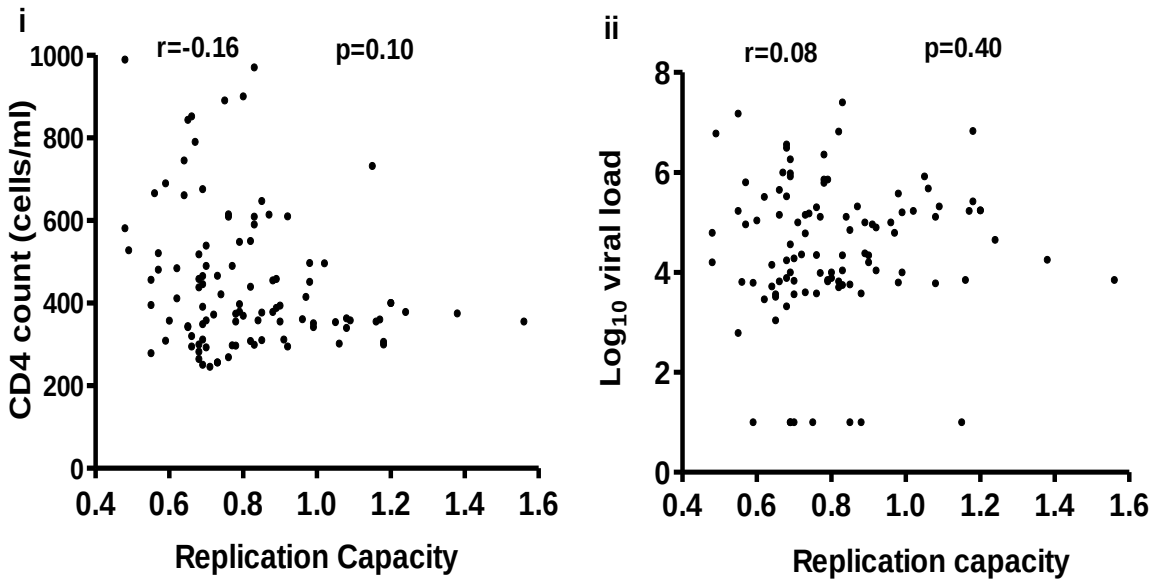

r=0.14
p=0.60
Similarity % to consensus B
100
95
90
85
80
75
0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6
Replication capacity

Figure 3.13 Replication capacity versus markers of disease progression

Graphs showing no correlation between Gag-protease driven replication capacity and markers of disease progression, namely CD4 counts and viral loads (Spearman's correlation).


i
r=-0.16
p=0.10
CD4 count (cells/ml)
1000
800
600
400
200
0
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Replication Capacity
ii
r=0.08
p=0.40
Log10 viral load
8
6
4
2
0
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Replication capacity

Similarly, no significant correlations were observed between replication capacities and CD4 counts (Spearman's correlation; r=-.0.13 and p=0.32 for subtype A, r=-0.09 and p=0.71 for subtype C, r=-0.03 and p=0.92 for subtype D, and r=-0.18 and p=0.44 for inter-subtype recombinants) or viral loads (Spearman's correlation; r=0.09 and p=0.48 for subtype A, r=0.48 and p=0.08 for subtype C, r=-0.15 and p=0.62 for subtype D, and r=-0.19 and p=0.43 for inter-subtype recombinants) as shown in Figure 3.14.

3.6 Impact of HLA class I on replication capacity

The vastly polymorphic human leukocyte antigen (HLA) class I locus has been shown to play a major role in immune control of HIV-1 infection [296, 336, 337]. In particular, host HLA-B alleles exert a considerable influence on HIV-1 disease progression [309, 338]. Recent studies on the impact of HLA-mediated immune pressure on HIV-1 Gag-protease fitness in HIV-1 subtype B and C have demonstrated lower replication capacities in patients who have protective alleles [325, 339, 340]. Consequently, we compared viral replication capacity from patients expressing known protective alleles (HLA-B*57:01, -B*57:02, -B*57:03, -B*58:01 -B*27 and –B*81:01) [338, 341-343] versus those who did not express protective alleles (Figure 3.15). However, for subtype A, HLA-B*58:01 was excluded from the protective HLA allele group as this HLA allele was not associated with lower CD4 counts or viral loads in subtype A infection [344, 345]. Generally we show that there were no significant differences in replication capacity of patients expressing protective alleles versus those with non-protective HLA alleles within subtype A, subtype D and inter-subtype recombinant groups (Mann-Whitney U test; p=0.92 for subtype A, p=0.87 for subtype D and p=0.53 for inter-subtype recombinants) (Figure 3.15). However, for subtype C, protective alleles were associated with lower replication capacity (Mann-Whitney U test; p=0.004). Interestingly, three out of four of the subtype C sequences from the patients expressing the protective HLA alleles harboured mutations, namely 186S and 242N, which were previously reported to have a fitness cost [346-348].

Figure 3.14 Replication capacity versus markers of disease progression in subtypes

Graphs showing no correlation between Gag-protease driven replication capacity and markers of disease progression, namely CD4 counts and viral loads, in the different subtypes (Spearman's correlation). Subtype A is shown in panels (i) and (ii), subtype C in (iii) and (iv), subtype D in (v) and (vi) and finally inter-subtype recombinants in (vii) and (viii) respectively.

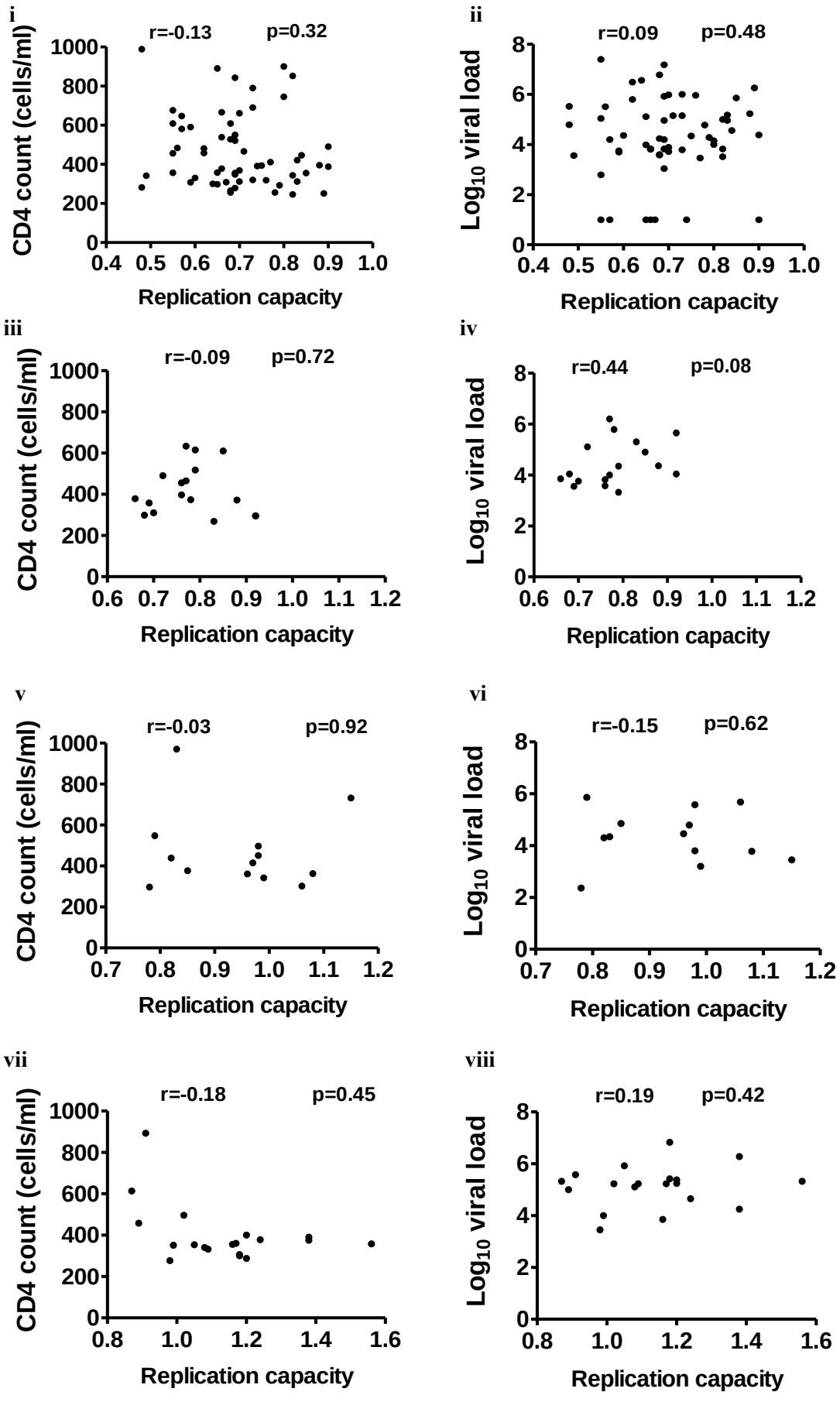
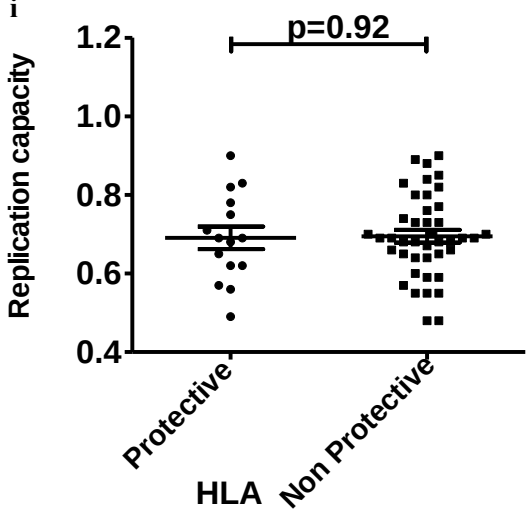

i
r=-0.13
p=0.32
CD4 count (cells/ml)
Replication capacity
ii
r=0.09
p=0.48
Log10 viral load
Replication capacity
iii
r=-0.09
p=0.72
CD4 count (cells/ml)
Replication capacity
iv
r=0.44
p=0.08
Log10 viral load
Replication capacity
v
r=-0.03
p=0.92
CD4 count (cells/ml)
Replication capacity
vi
r=-0.15
p=0.62
Log10 viral load
Replication capacity
vii
r=-0.18
p=0.45
CD4 count (cells/ml)
Replication capacity
viii
r=0.19
p=0.42
Log10 viral load
Replication capacity

Figure 3.15 Impact of protective HLA class I alleles on replication capacity in individual subtypes

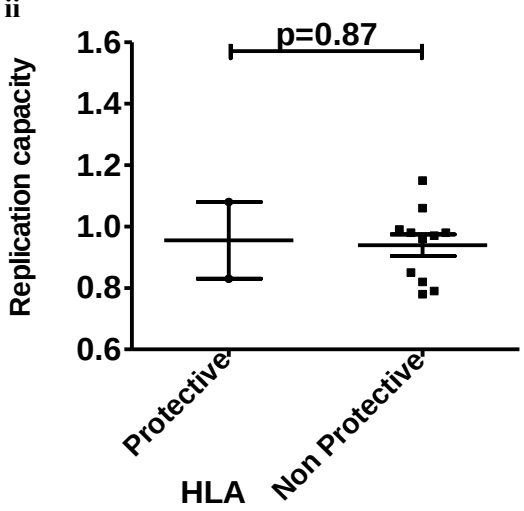
Graphs (i-iii) showing no significant differences in replication capacity of patients expressing protective HLA alleles versus non-protective HLA alleles in subtype A, subtype D and inter-subtype recombinants.

Graph (iv) showing significant differences in replication capacity of patients expressing protective HLA alleles versus those with non-protective HLA alleles in subtype C.

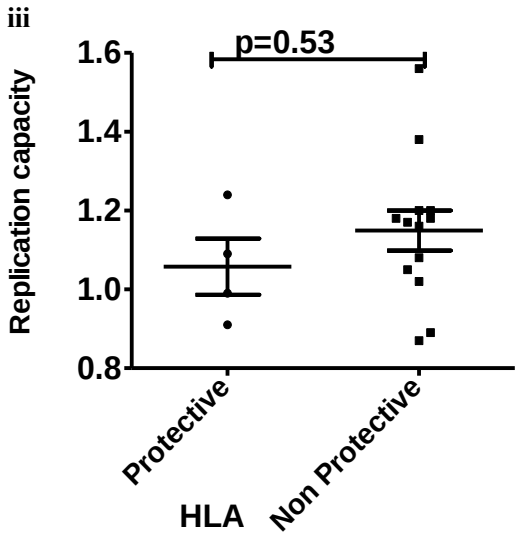
i


p=0.92
Replication capacity
1.2
1.0
0.8
0.6
0.4
Protective
Non Protective
HLA

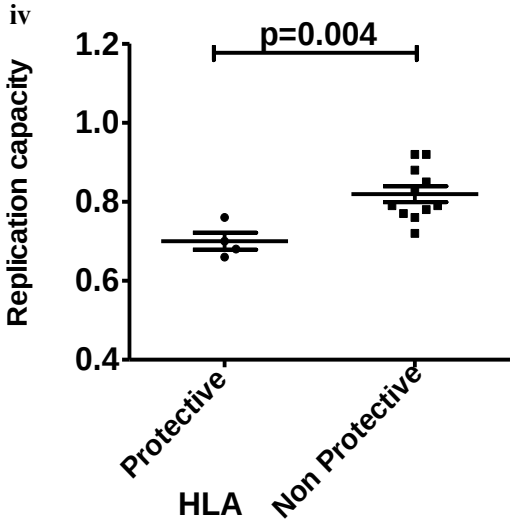
ii


p=0.87
Replication capacity
1.6
1.4
1.2
1.0
0.8
0.6
Protective
Non Protective
HLA

iii


p=0.53
Replication capacity
1.6
1.4
1.2
1.0
0.8
Protective
Non Protective
HLA

iv


p=0.004
Replication capacity
1.2
1.0
0.8
0.6
0.4
Protective
Non Protective
HLA

3.6.1 HLA class I effect on replication capacity within subtype A

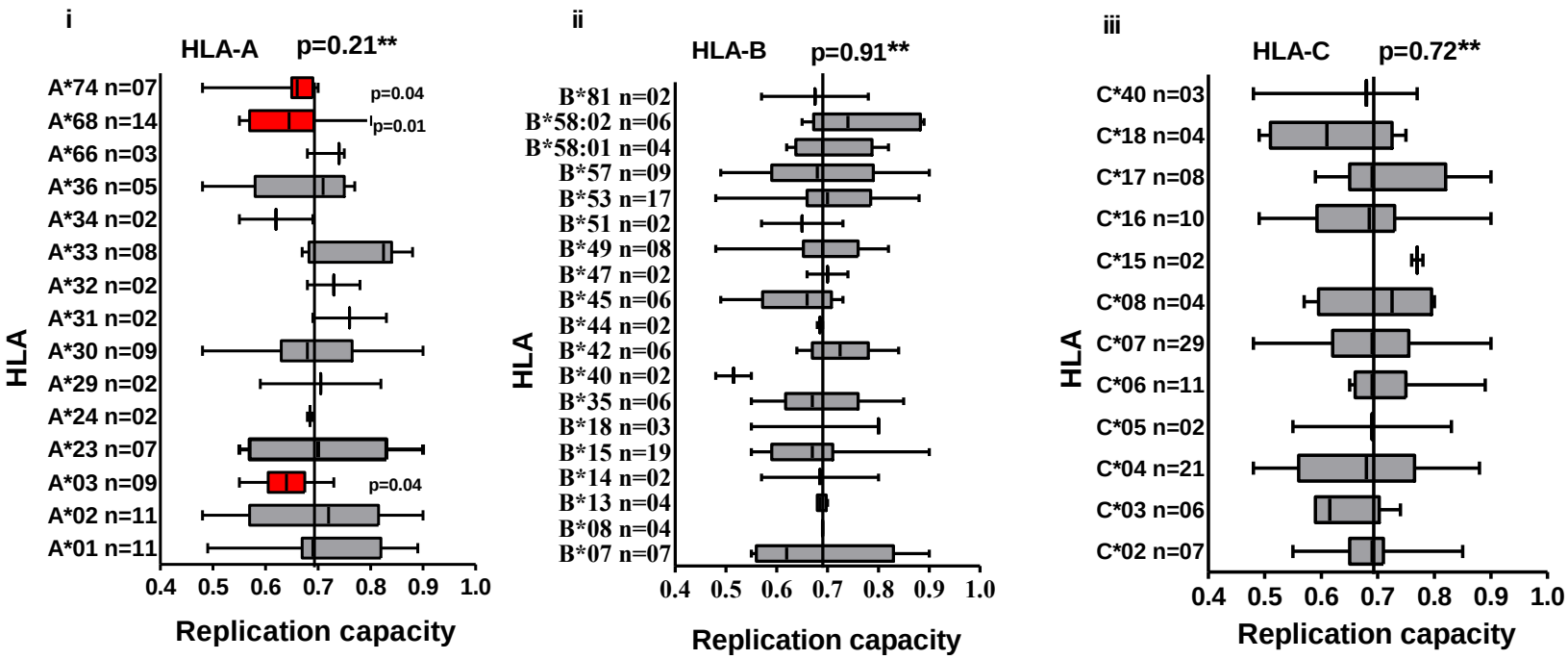
A number of studies have shown that different individual HLA class I alleles are associated with loss of viral fitness in subtype B and C due to immune selection pressure and sequence changes in the *gag* gene [306, 316-318]. We sought to assess the impact of specific HLA class I alleles on Gag-protease driven replication capacity in the different subtypes in our cohort, however we were limited in numbers for subtypes C and D and inter-subtype recombinants ($n<20$ for all), therefore we focussed this analysis on subtype A for which we had the most samples ($n=57$). Viral replication capacities were grouped according to the HLA class I alleles expressed by the host for subtype A infected patients. There were no significant differences in replication capacity overall between different HLA-A, -B, or –C alleles as assessed by ANOVA (Figure 3.16). For each individual allele, the replication capacities of patients expressing that particular HLA allele versus those who did not express the allele were compared (only for HLA alleles present at $n\geq5$), and it was observed that A*74, A*68 and A*03 were associated with lower replication capacities (Student's T test; $p=0.04$, $p=0.01$ and $p=0.04$, respectively). However, when Bonferroni correction was applied to account for multiple comparisons, these alleles were no longer significantly associated with lower replication capacity. Therefore, we did not find strong evidence of HLA alleles driving viral attenuation in subtype A.

3.7 Sequence determinants of replication capacity

We subsequently wanted to identify sequence determinants of Gag-protease driven replication capacity. Specifically we investigated whether there was any association between replication capacity and different recombination breakpoints as well as the number of HLA-associated polymorphisms present in a given sequence. In addition, we performed analyses to identify specific amino acids associated with altered replication capacity.

Figure 3.16 HLA allele effect on Gag-protease driven replication capacity within subtype A

Graphs showing Gag-protease replication capacities grouped according to HLA-A (i), -B (ii) and –C (iii) alleles expressed by patients infected with subtype A. Double asterisks (**) indicate ANOVA p values for HLA-A, -B and –C groups. Individual HLA alleles associated with reduced replication capacity as determined by the Student's T test ($p<0.05$) are highlighted in red. The vertical line shows the mean replication capacity, the edges of the boxes show interquartile ranges, and the edges of the whiskers show the most extreme value.


i
HLA-A
p=0.21**
A*74 n=07
A*68 n=14
A*66 n=03
A*36 n=05
A*34 n=02
A*33 n=08
A*32 n=02
A*31 n=02
A*30 n=09
A*29 n=02
A*24 n=02
A*23 n=07
A*03 n=09
A*02 n=11
A*01 n=11
p=0.04
p=0.01
p=0.04
HLA
0.4 0.5 0.6 0.7 0.8 0.9 1.0
Replication capacity
ii
HLA-B
p=0.91**
B*81 n=02
B*58:02 n=06
B*58:01 n=04
B*57 n=09
B*53 n=17
B*51 n=02
B*49 n=08
B*47 n=02
B*45 n=06
B*44 n=02
B*42 n=06
B*40 n=02
B*35 n=06
B*18 n=03
B*15 n=19
B*14 n=02
B*13 n=04
B*08 n=04
B*07 n=07
HLA
0.4 0.5 0.6 0.7 0.8 0.9 1.0
Replication capacity
iii
HLA-C
p=0.72**
C*40 n=03
C*18 n=04
C*17 n=08
C*16 n=10
C*15 n=02
C*08 n=04
C*07 n=29
C*06 n=11
C*05 n=02
C*04 n=21
C*03 n=06
C*02 n=07
HLA
0.4 0.5 0.6 0.7 0.8 0.9 1.0
Replication capacity

3.7.1 Recombination breakpoints

We hypothesised that the pattern of recombination might affect the replication capacities of the inter-subtype recombinants. Therefore, we compared the replication capacities of inter-subtype recombinants with and without the common breakpoints 150, 410 and 435. However, there were no significant differences in replication capacities in the presence or absence of these breakpoints (Mann-Whitney U test; p=0.79, p=0.79, p=0.53, respectively). Furthermore, there was no significant difference in replication capacity between recombinants with a single breakpoint and those with multiple breakpoints (Mann-Whitney U test; p=0.92) as shown in Figure 3.17.

3.7.2 HLA-associated polymorphisms

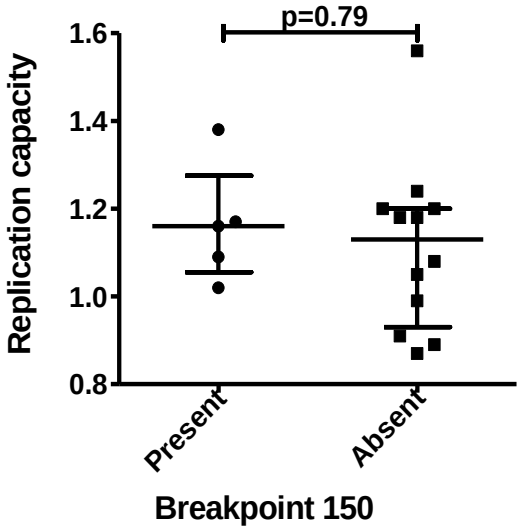
Studies done previously in subtypes B and C identified negative relationships between HLA-associated polymorphisms and replication capacity [316, 319, 325]. Therefore we explored the relationship between the number of HLA-associated polymorphisms and replication capacity in our cohort. For each subtype, we used a list of HLA-associated polymorphisms that were based on sequences from that subtype and were previously defined using methods that account for the phylogenetic relatedness of sequences, amino acid co-variation, and HLA linkage disequilibrium effects [349]. The list of HLA-associated polymorphisms for subtype C (Table A2) was previously published [325], and the lists for subtypes A (Table A1) and D (Table A3) were kindly provided by Dr Zabrina Brumme from Simon Fraser University [350]. We considered only those polymorphisms that were non-consensus and positively associated with the presence of a particular HLA allele (i.e. adapted associations) at $p<0.05$ and $q<0.2$. We found that the quantity of HLA-associated polymorphisms specific to the host HLA alleles did not correlate with the replication capacity (Spearman's correlation; r=-0.22 and p=0.10 for subtype A, r=-0.17 and p=0.52 for subtype C, r=-0.57 and p=0.85 for subtype D, and r=0.06 and p=0.86 for inter-subtype recombinants (Figure 3.18). Since HLA-B-associated polymorphisms in particular, especially those in or immediately adjacent to epitopes, were previously shown to correlate

Figure 3.17 Recombination patterns and replication capacity

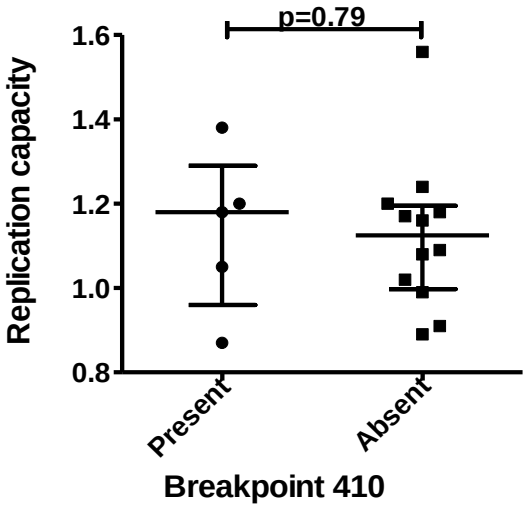
Graphs panels (i-iii) showing no significant differences in replication capacities in the presence or absences of breakpoints at positions 150, 410 and 435, respectively.

Graph (iv) showing no significant differences in replication capacities of recombinants with single breakpoints or multiple breakpoints.

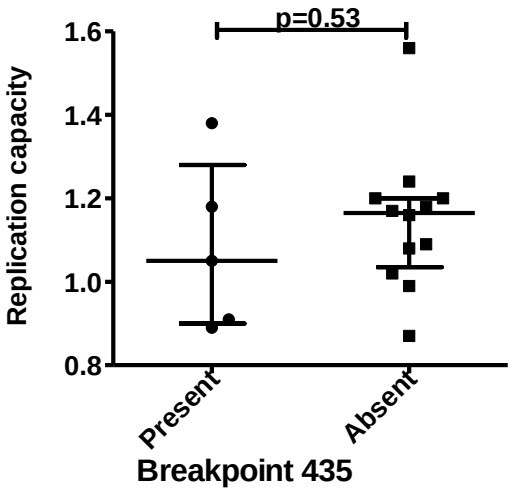
i


p=0.79
1.6
1.4
1.2
1.0
0.8
Replication capacity
Present
Absent
Breakpoint 150

ii


p=0.79
1.6
1.4
1.2
1.0
0.8
Replication capacity
Present
Absent
Breakpoint 410

iii


p=0.53
1.6
1.4
1.2
1.0
0.8
Replication capacity
Present
Absent
Breakpoint 435

iv

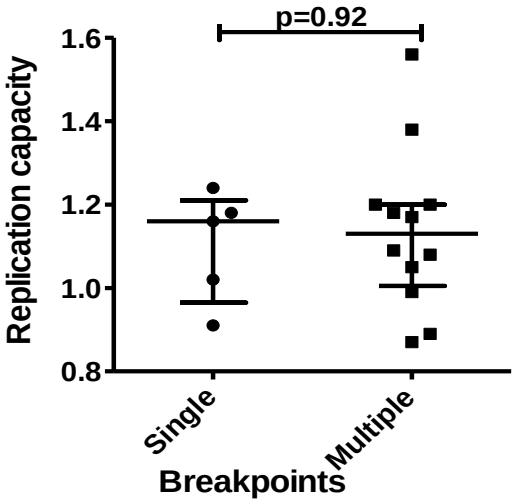
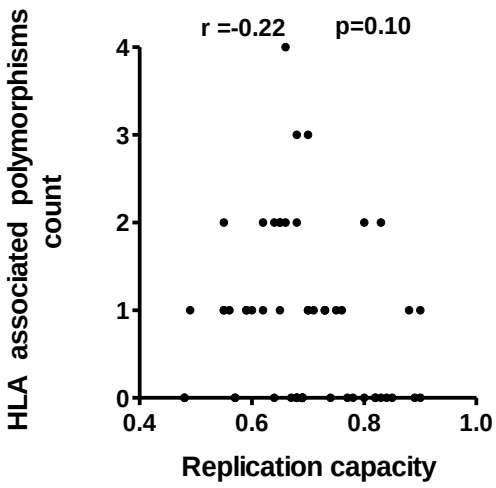

p=0.92
1.6
1.4
1.2
1.0
0.8
Replication capacity
Single
Multiple
Breakpoints

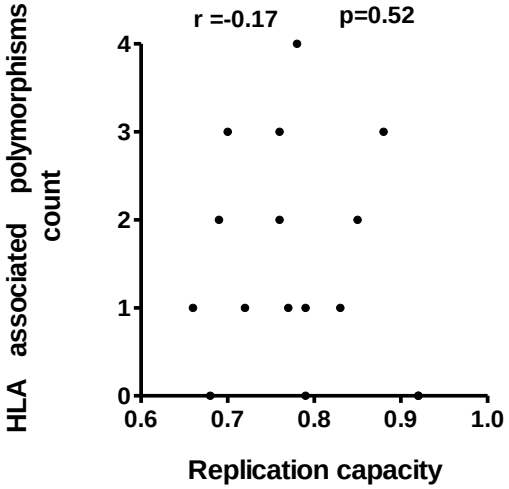
Figure 3.18 Correlation of host HLA-associated polymorphism counts with replication capacity within the individual subtypes

Graphs showing no correlation between the number of HLA-associated polymorphisms that the host could have selected based on their HLA alleles and replication capacity (Spearman's correlation). Subtype A, subtype C, subtype D and inter-subtype recombinants are represented in panels i-iv, respectively.

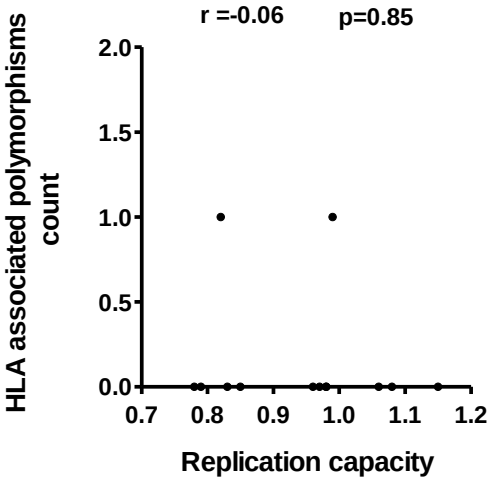
i


r =-0.22
p=0.10
HLA associated polymorphisms count
Replication capacity
0
1
2
3
4
0.4
0.6
0.8
1.0

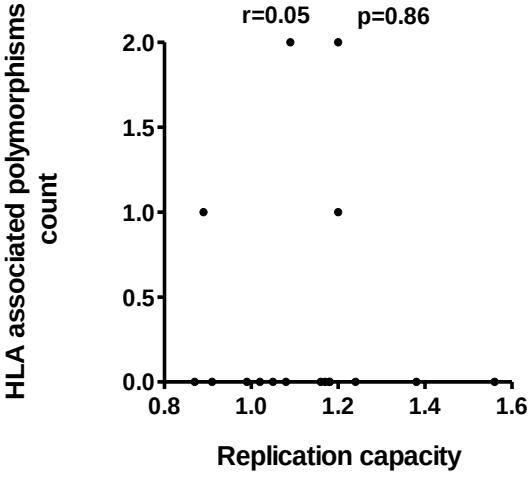
ii


r =-0.17
p=0.52
HLA associated polymorphisms count
Replication capacity
0
1
2
3
4
0.6
0.7
0.8
0.9
1.0

iii


r =-0.06
p=0.85
HLA associated polymorphisms count
Replication capacity
0.0
0.5
1.0
1.5
2.0
0.7
0.8
0.9
1.0
1.1
1.2

iv


r=0.05
p=0.86
HLA associated polymorphisms count
Replication capacity
0.0
0.5
1.0
1.5
2.0
0.8
1.0
1.2
1.4
1.6

significantly with reduced replication capacity in previous studies on subtypes B and C [1, 10, 12], we repeated this analysis considering only HLA-B-associated polymorphisms in or within 5 amino acids of optimal epitopes. We used the optimal list also referred to "A list" (Table A4) from the Los Alamos HIV database (http://www.hiv.lanl.gov/content/immunology/variants/variantsearch.html). These are more likely to represent CTL escape mutations. With these specific criteria we observed no correlation between the number of the HLA-associated polymorphisms restricted by HLA-B alleles and replication capacity (Spearman's correlation; r=-0.12 and p=0.37 for subtype A, r=-0.06 and p=0.85 for subtype D, and analysis for subtype C and inter-subtype recombinants was obsolete as there were no counts associated with any HLA alleles) (data not shown). We additionally counted the polymorphisms regardless of the HLA alleles expressed by the host. With the exception of a weak positive relationship between the number of polymorphisms and replication capacity in subtype D (p=0.04; data not shown), there were no significant relationships observed (data not shown). When only HLA-B-associated polymorphisms in or next to optimal epitopes were included, we similarly found that there were no significant associations between HLA-associated polymorphisms and replication capacity (data not shown). In summary, we did not observe an attenuating effect of HLA-associated polymorphisms on Gag-protease driven replication capacity.

3.7.3 Specific amino acid variants

A number of studies have shown an impact of specific amino acids in subtype B and C Gag–protease on viral fitness [319, 351-353]. We performed an exploratory codon-by-codon analysis (http://brockman-srv.mbb.sfu.ca/~B_Team_iMac/Codon_by_codon) to assess whether we could identify specific amino acid variants that were significantly associated with altered Gag-protease driven replication capacity in our cohort. Due to limitation in numbers for the other subtypes, only subtype A was considered for this analysis. Considering only amino acid variants occurring at a frequency of $n \geq 5$, we identified six amino acids which were significantly ($p<0.05$ and $q<0.2$)

Table 3.2 Amino acid variants significantly associated with altered replication capacity

Codon	Amino acid	Consensus	With amino acid	Without amino acid	n=with amino acid	n=without amino acid	p-value	q-value
75	I	L	0.65	0.70	13	40	0.07	0.04
107	L	I	0.64	0.69	9	43	0.007	0.03
125	S	S	0.70	0.65	37	15	0.02	0.03
126	S	S	0.69	0.62	47	8	0.01	0.03
315	N	N	0.67	0.71	18	36	0.05	0.04
499	S	S	0.70	0.62	48	6	0.03	0.04

associated with reduced or increased replication capacity (Table 3.2). In order to explore potential sequence determinants of the inter-subtype differences in replication capacity we compared the frequency of these amino acid variants between subtypes A and D, for which 400 sequences each were available from the Los Alamos HIV sequence database, (http://www.hiv. lanl.gov/ components/sequence/HIV/search/search.html). Four out of the six amino acid variants associated with altered replication capacity differed significantly in frequency between subtypes A and D. Interestingly, the mutation 107L and consensus amino acid 315N were significantly more frequent in subtype A (Chi square test; $p=10^{-8}$ and p=0.01 respectively) and were associated with reduced replication capacity (Table 3.3). Therefore it is possible that these differences contribute to the lower replication capacity of subtype A sequences.

3.7.4 TW10 epitope mutations in subtype A

The HLA-B*57-restricted TW10 epitope is one of the most extensively studied epitopes in light of immune escape and viral control in HIV-1 infection. HLA-B*57-mediated selection pressure has been shown to result in a predictable escape pathway in TW10 epitope [283, 323, 348, 349, 356]. Specifically, the 242N mutation is common in subtypes B, C and D, and has been shown to result in a significant fitness cost [323, 348]. However, it was shown recently that there is clade-specific evolution mediated by HLA-B*57 in HIV-1 clade A1 p24: T242N is rare in subtype A, while escape mutations at residues 243 and 247 are more common [346]. These authors also observed that mutations at codon 247 are associated with a longer time for CD4 counts to drop below 500 (L McKinnon 2014, personal communication, 13 August 2014) suggesting that these mutations may carry a fitness cost. Therefore, we compared the replication capacity of viruses with and without mutations at this codon, and found that viruses encoding 247X had a lower replication capacity than those expressing the consensus residue at codon 247 (Student's T test; p=0.01) (Figure 3.19). A comparison of the replication capacities of viruses with and without mutations at codon 243 showed no significant difference (Student's T test; p=0.27) (Figure 3.20).

Table 3.3 Frequencies of amino acids associated with altered replication capacity in subtypes A and D

Codon	Subtype A n=400	Subtype D n=400	p- value
75I	264	262	0.88
107L	32	0	10^{-8}
125S	346	320	0.01
126S	274	280	0.64
315N	264	229	0.01
499S	302	163	0.77

Figure 3.19 The effect of mutations at codon 247 in the TW10 epitope on viral replication capacity

Graph showing a significant difference in replication capacity between viruses encoding a mutation at position 247 (I247X) versus those that did not a have a mutation at that position (247I)

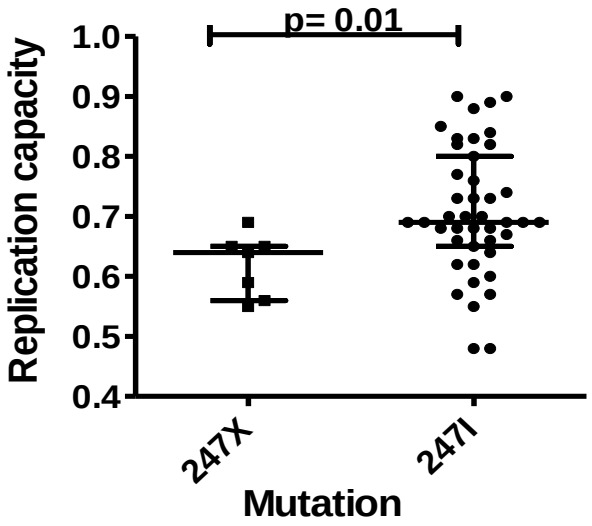
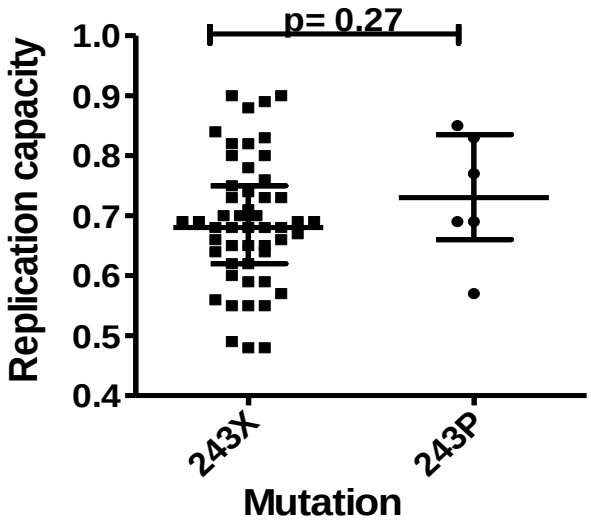

p= 0.01
Replication capacity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
247X
247I
Mutation

Figure 3.20 The effect of mutations at codon 243 in the TW10 epitope on viral replication capacity

Graph showing no significant difference in replication capacity between viruses encoding a mutation at position 243 (P243X) versus those that did not a have a mutation at that position (243P).


p= 0.27
1.0
0.9
0.8
0.7
0.6
0.5
0.4
Replication capacity
243X
243P
Mutation

CHAPTER 4

DISCUSSION

CHAPTER 4- DISCUSSION

4.1 Background

The HIV-1 epidemic is heterogeneous with various subtypes in circulation. The subtypes are unevenly distributed globally, unevenly expanding and have been associated with differences in rates of disease progression [48, 49, 354, 355]. The reasons for these subtype-dependent differences in prevalence, rates of expansion in the epidemic and clinical progression are unknown [249]. The Gag protein plays an essential role in the replication cycle of HIV [356-361] and is a preferred target of the immune system [290, 291, 362]. Since there is evidence supporting Gag-protease-driven replication capacity as a determinant of disease progression [363-365], we hypothesized that differences in Gag-protease driven replication capacity between subtypes may be responsible for HIV-1 subtype specific differences in disease progression. We sought to address subtype differences within a relatively genetically homogenous population, since HLA class I (the most polymorphic human genetic loci) has been shown to significantly impact Gag-protease-driven virus replication capacity [316, 317, 366]. We therefore undertook this study in the well characterized Majengo sex worker cohort based in Nairobi, Kenya, where multiple subtypes and inter-subtype recombinants coexist [367, 368].

4.2 Subtype classification and distribution

In East Africa, heterogeneity has been reported with HIV-1 subtypes A, C, D and their recombinants being prevalent [25-28]. In this study of samples collected between the years 2000 and 2010, HIV-1 subtype distribution based on *gag* gene sequencing was as follows: A (n=57, 55.4%), C (n=16, 15.5%), D (n=13, 12.6%) and inter-subtype recombinants (n=17, 16.5%). The HIV-1 subtype distribution pattern in our study is comparable to that of other studies done in East Africa between the years 1995 - 2007 [92, 105, 240, 242, 369]. Subtype A is the most prevalent subtype and this could possibly be attributed to founder effects [48, 229, 370, 371]. We also found that there was a relatively high percentage of inter-subtype recombinants in our cohort as

compared to studies done prior to the year 2000 [49, 87] (for example, a study in Kenya done in the year 1999 reported that *env* inter-subtype recombinants had a prevalence of 2-5% [116]), indicating that inter-subtype recombinants have increased in prevalence [111]. The increase in inter-subtype recombinants has been thought to arise from the increase in HIV-1 dual infections [372-377]. In addition, the increase and spread of inter-subtype recombinants has also been linked to human migration and movements from different regions/countries in search of better job opportunities [376-378]. It should be noted that it is possible that we are underestimating the prevalence of inter-subtype recombinants since our classification is based on one genomic region and therefore excludes recombination that could occur in other genomic regions [379]. The knowledge of HIV-1 subtype distribution and continuous changes in the genetic profile of the HIV epidemic is essential in understanding HIV-1 pathogenesis and these insights will have an impact on diagnostics, virulence, treatment strategies and, ultimately, contribute to vaccine developments.

4.3 Recombination patterns

Interestingly, the inter-subtype recombinants had distinct breakpoint patterns, with breakpoints commonly occurring at amino acids positions 150, 410 and 435 (according to HXB2 numbering). This suggests that recombination in the *gag* gene is not random. The common breakpoints may represent a protein structural advantage favouring recombination at these points. This finding is consistent with other studies showing that recombination has discrete breakpoints along the HIV-1 genome with potential localized "hot spots" [369, 380-385]. Similarly, another study done in the same cohort showed a pattern of recombination breakpoints within the *gag* region of the inter-subtype recombinants [369]. Other studies in the *env* gene have illustrated that recombination breakpoint patterns are not random [382, 386-388]. The recombination breakdown from our study (3 AD patients, 12 AC patients and 1 CD patient) demonstrated that for most recombinants subtype A is a component, likely reflecting the predominance of subtype A.

4.4 Replication capacity differences among subtypes

In this study, we hypothesized that Gag-protease NL4-3 recombinant viruses derived from subjects infected with different HIV-1 subtypes would show differences in viral replicative fitness. We observed a hierarchy of Gag-protease driven replicative fitness where subtypes A/C were less fit than D, which was also less fit than inter-subtype recombinants, and found that this hierarchy was reproducible across different cohorts and geographical regions. These inter-subtype differences in Gag-protease driven replication capacity are largely consistent with those observed for the fitness of whole virus isolates in studies where competition assays were performed in PBMCs for limited numbers of isolates of different subtypes [127, 230, 231]. Consistent with our findings, in these previous studies subtype C isolates from Africa were shown to have lower fitness than other M group isolates and subtype A isolates had a lower fitness than those of subtype D [231]. However, these previous studies found that subtype C isolates had a significantly lower fitness than those of subtype A; while in our study we observed that subtype C Gag-protease-driven replication capacity was higher than that of subtype A, but the two means were not statistically different. This may suggest that genetic determinants other than Gag-protease may distinguish the fitness of subtype A and C isolates. Our finding that inter-subtype recombinants have the highest Gag-protease-driven replication capacity is consistent with studies showing that the circulating recombinant form CRF02_AG has a higher *ex vivo* replicative fitness than its parental subtypes [250]. In addition, a study by Tarosso *et al*., demonstrated that subtype recombinant BF was associated with a faster $CD4^+$ T cell loss than parent subtype B in Brazil, suggesting that the recombinant form had a greater fitness than the parent subtypes [389]. Our data is consistent with recombination conferring replicative advantage of viable progeny compared to parental strains [390]. In line with this, a previous study found that recombination was associated with fitness recovery in viral quasispecies [391]. Our findings that inter-subtype differences in Gag-protease driven-replication capacity are largely consistent with fitness differences between whole isolates of different subtypes indicates that Gag-protease is a

significant determinant of overall viral fitness, in addition to other genes, namely *env* [230], *pol* [206] and *protease-reverse transcriptase* [392], which have been shown to correlate with whole isolate fitness.

The hierarchy of Gag-protease-driven replication capacity that we observed in East African cohorts is also consistent with reported inter-subtype differences in disease progression in these populations; where subtype A has exhibited a slower disease progression than subtype D in cohorts from Uganda [88, 241], Kenya [105, 245] and Tanzania [242]. Inter-subtype recombinants in Uganda have resulted in faster disease progression to AIDS/death than subtype A and had a slightly higher hazard ratio for death compared with subtype D [240]; however, this increased virulence of inter-subtype recombinants was not observed in Tanzania [242]. In addition, in Guinea-Bissau, infection with recombinant A3/02 was associated with increased risk of AIDS or death compared with subtype A3 [250]. Together with previous studies showing an association between Gag-protease or Gag driven replication capacity and markers of disease progression [316, 325, 363, 393], the consistency between inter-subtype differences in Gag-protease driven replication capacity and disease progression indicates an important influence of Gag-protease on disease progression. We speculate that the lower Gag-protease driven replication capacity of subtypes A and C slows disease progression in individuals infected with these subtypes, leading to greater opportunity for transmission [394] and consequently increased prevalence of these subtypes [48]. Indeed, subtype C is the most prevalent subtype world-wide and subtype A is the predominant subtype in East Africa [48, 49, 104, 395, 396]. Furthermore, the increase in prevalence and spread of subtype C has been reported in Brazil and in Western Europe, the Americas, and Australia [85, 397, 398]. Additionally, a study by Conroy *et al*., showed the expansion of subtype A at the expense of subtype D in a cohort in Uganda [229]. This expansion has been observed in Kenyan cohorts as well [371]. Increasing expansion of subtypes A and C may also be partly attributed to a transmission advantage of these subtypes[88, 399]. High transmission efficiency of subtype C may be at least partly due to its low frequency of

classical switch from CCR5-tropic virus to CXCR4-tropic virus relative to other subtypes and maintenance of predominant CCR5 tropism throughout infection [400].

4.5 Replication capacity of the same subtype across different geographical regions

In this study we also investigated Gag-protease driven replication capacities of 20 subtype C viruses derived from the Sinikithemba cohort in Durban and compared these to the replication capacities of the subtype C viruses generated from the Majengo cohort in Nairobi. We found that the replication capacities of the subtype C viruses from the two geographically diverse regions were similar. Therefore, we can conclude that, at least within the African continent, the same subtype portrays similar Gag-protease driven replication capacity. However, it is possible that the Gag-protease-driven replication capacity for subtype C from India may differ from subtype C from Africa, since it has been shown subtype C isolates from India replicated faster in PBMCs than subtype A isolates [233].

4.6 Replication capacity versus markers of disease progression

In previous studies, it has been established that Gag-protease-driven or Gag-driven replication capacity correlates with markers of disease progression for subtypes B and C [316, 319, 325, 335, 363]. We therefore hypothesized that there would be a correlation of Gag-protease replication capacity with markers of disease progression in our data set; however, this was not observed. The lack of correlation with markers of disease progression maybe be due to compensatory mutations which occur in chronic infection, considering these particular patients were chronically infected and it has been shown that during chronic infection compensatory mutations restore fitness [273, 319, 321, 393, 401]. More recently, a study in Uganda showed restored fitness in subtypes A and D due to compensatory mutations in chronic infection [402]. In addition it could be attributed to the limitation in sample size in our cohort. In previous studies showing a significant correlation

between Gag-protease driven replication capacity and markers of disease progression, the correlation co-efficient is weak, suggesting that a large population is needed to observe a significant correlation. Similarly, a study in Japan did not find any significant impact of Gag-Protease-mediated HIV-1 replication capacity on clinical parameters in treatment-naïve patients [403]. Furthermore, this correlation tested only one gene (*gag-protease*) of the viral genome and thus may not fully capture viral replication capacity of the entire HIV-1 genome from the patients.

4.7 Impact of HLA class I on replication capacity

4.7.1 Protective HLA alleles

Previous studies of HIV-1 subtypes B and C have demonstrated that Gag-protease viral replication capacities differ across HLA-B alleles [319, 325] and have hypothesized that some protective HLA class I alleles mediate their effects by driving the virus to a less fit state. We thus hypothesized that viruses from individuals who express protective HLA class I alleles would have lower viral replication capacity irrespective of subtype. We only considered already defined protective alleles: HLA-B*57:01, -B*57:02, -B*57:03, -B*58:01 -B*27 and –B*81:01 [309, 338, 404-406]. Contrary to what was expected, there were no significant differences in replication capacity of patients expressing protective alleles versus those with non-protective HLA alleles in HIV-1 subtypes A, subtype D and inter-subtype recombinants. However, for subtype C, protective alleles were associated with lower replication capacity. Interestingly, three out of four of the subtype C sequences from the patients expressing the protective HLA alleles, harboured mutations namely 186S and 242N, which were previously reported to have a fitness cost [346-348]. These results could perhaps suggest that the protective alleles for HIV- 1 subtypes A and D are not the same as the well-defined protective alleles for subtype B and C. We speculate that different HIV-1 subtypes have distinct HLA class I associations with clinical outcome and that mechanisms or patterns of Gag-protease-mediated differences in replication capacity are not overlapping. However, it should also be noted that there are other non-HLA and non-CTL factors

that may impact on viral fitness and these are largely undetermined. Nevertheless, our data suggests that there are differences in HLA-associated pressure, with protective HLA alleles possibly different across the various HIV-1 subtypes. A combination of HLA-mediated pressure and intrinsic differences in viral fitness may therefore contribute to the differences seen in pathogenesis and disease progression.

4.7.2 HLA class I effect on replication capacity within subtype A

We further assessed the impact of each specific HLA class I allele on Gag-protease driven replication capacity for subtype A since we had the most samples (n=57) for this subtype. We found that A*74, A*68 and A*03 were associated with lower replication capacities ($p<0.05$), although these associations did not remain significant after Bonferroni correction for multiple comparisons was employed which may be attributed to a limitation in numbers. A study done in the Majengo cohort recently (2014) showed that the frequencies of A*74, A*68 and A*03 were 9.12%, 6.76% and 4.72%, respectively [407]. Interestingly, A*74:01 has been associated with protection from HIV-1 acquisition and disease progression in Tanzania [408], and also in children in a Kenyan cohort [409]. In addition, HLA-A*74:01 has also been shown to mediate control of viremia in a subtype C population [410]. Moreover, HLA-A*74:01 been associated with a favourable CD4:CD8 ratio in a Zambian cohort [411]. Thus, our data suggest that lowered viral fitness could be a mechanism contributing to the beneficial effect of A*74:01. Even though HLA-A*68 has been associated with strong Gag CTL responses in subtype C (HLA-A*68:02 is one of the most prevalent alleles in African populations) [412], it has not been directly linked with altered disease progression or protection in subtype A. HLA-A*03 has not been linked with altered disease progression, however it has been negatively associated with HIV-1 infection in China [413]. In addition, HLA-A*03 has been reported to play an important role in inducing immune responses to a variety of CTL epitopes [414].

4.8 Sequence determinants of replication capacity

4.8.1 Recombination patterns

Amino acid sequence variations could have a substantial impact on the secondary and tertiary structure of a protein, thereby affecting its function. We initially investigated whether recombination breakpoint patterns could be associated with differences in replication capacity, since we identified certain breakpoints that occurred more frequently than others. However, we found no significant differences in Gag-protease mediated replication capacity in the presence or absence of the common breakpoints in the inter-subtype recombinants and there was no significant difference in replication capacity between recombinants with a single breakpoint and those with multiple breakpoints. This could imply that the recombination patterns observed in nature generally represent an advantage for the virus. It has been speculated that distribution of recombinant breakpoints across the HIV-1 genome is dependent on sequence identity and mechanism of recombination, and the relative functionality of the recombinant genes [380, 383].

4.8.2. HLA-associated polymorphisms

Studies have previously identified negative relationships between HLA-associated polymorphisms and Gag-protease or Gag driven replication capacity in HIV subtype B and C [316, 319, 325]. Therefore, we hypothesized that the number of HLA-associated polymorphisms would affect replication capacity. However, we found no evidence that HLA-associated mutations in Gag-protease attenuate HIV for any of the subtypes. One potential explanation may be the chronic infection status of the patients. In chronic infections, viral escape mutations from CTL are usually accompanied by compensatory mutations that restore viral fitness [273, 319, 321, 401]. A study of patients chronically infected with HIV-1 subtype B similarly found no relationship between the number of HLA-associated polymorphisms and Gag-protease driven replication capacity, while this relationship was evident in acute infection [339], suggesting that fitness costs of escape mutations were largely compensated in late infection. Another study of

patients chronically infected with HIV-1 subtype C detected a significant relationship between the number of HLA-associated polymorphisms and Gag-protease driven replication capacity; however the correlation was relatively weak [325]. It should be appreciated that our sample size may have limited our ability to detect this relationship. It is worth noting that a recent study in Uganda showed that HLA-associated Gag mutations associated with protection, namely A163X in KF11 epitope and I147X in the ISW9 epitope, were more frequent in subtype A than subtype D [402]. These mutations have previously been demonstrated to carry a fitness cost [321, 346]. Thus these data are consistent with our finding that subtype A has a lower Gag-protease mediated fitness than subtype D.

4.8.3 TW10 epitope mutations in subtype A

Since it was previously found that there are unique TW10 epitope escape variants that are associated with the protective allele HLA-B*57 for subtype A as compared to the other subtypes [344] and the fitness consequences of these are currently unknown, we explored the relationship between these variants and replication capacity in our data set. The 242N mutation is selected by the majority of individuals expressing B*57 or B*58:01 in subtypes B, C and D; however it is rare in subtype A-infected HLA-B*57/B*58:01 positive individuals [344]. Instead, escape variants at codons 243 and 247 are observed in subtype A-infected individuals who possess HLA-B*57 but not HLA-B*58:01, and interestingly HLA-B*57 but not HLA-B*58:01 is associated with a protective effect in subtype A infection, suggesting that these unique escape variants may be responsible for the protective effect of B*57 in subtype A infection [344]. Since the same authors observed that mutations at codon 247 are associated with a longer time for CD4 counts to drop below 500 cells/µl (L McKinnon 2014, personal communication, 13 August 2014), we hypothesized that these mutations would be associated with a fitness cost. Accordingly, we found viruses encoding 247X had a lower replication capacity than those expressing the consensus residue at codon 247. Furthermore, we did not find a fitness cost associated with mutations at

243 which is consistent with the observation that mutations at this codon were stable (i.e. did not revert) in the absence of HLA-B*57 (L McKinnon 2014, personal communication, 13 August 2014). Thus our data suggests that a fitness cost associated with 247X may partly explain the protective effect of HLA-B*57 in subtype A infected individuals.

4.8.4 Codon-by-codon analysis

Previous studies have shown an impact of specific amino acids in subtype B and C Gag–protease on viral fitness [319, 348, 415]. Although limited by sample size, we performed an exploratory codon-by-codon analysis on our subtype A sequences (the subtype for which we had the largest sample size) to identify associations between single amino acid variants and Gag-protease driven replication capacity. We identified six amino acids, namely 75I, 107L, 125S, 126S, 315N and 499S (of which amino acids at codons 125, 126, 315 and 499 were the consensus amino acids), that were significantly associated with altered replication capacity in subtype A. The amino acids identified as associated with differences in replication capacity for subtype A in this study do not overlap with those previously identified for subtype C [325]. The residue 75I is located on an alpha helix in the folded MA protein. It is hypothesized that mutations occurring in the alpha helix permit greater flexibility in the secondary structure of Gag, which in turn enhances the MA/CA cleavage site availability to the protease protein [416]; however 75I was associated with reduced replication. A previous study provides an indirect insight that 75I could be an escape variant: it was positively associated with HLA-A*02, and prevented both peptide processing and recognition in the flanking regions of SL9 [417, 418]. The residues 107L, 125S and 126S are located in the p17 region, while residues 315N and 499S are located in p24 and p6 regions, respectively, and none of these residues have been reported as escape variants to the best of our knowledge. Interestingly, the polymorphism 107L and consensus amino acid 315N were significantly more frequent in subtype A than subtype D and were associated with reduced replication capacity. This could possibly mean that the occurrence of these mutations confers a

fitness cost and hence may contribute to the lower replication capacity of subtype A compared to subtype D.

4.9 Future directions and recommendations

A significant limitation of our study is that Gag-protease-driven viral fitness was measured in the subtype B backbone. Although this approach allows for only the influence of Gag-protease on viral fitness to be measured, it may bias the data on Gag-protease-driven viral fitness if particular Gag-protease variants preferentially interact with NL4-3 proteins outside of Gag. To address this potential NL4-3 backbone bias, the Gag-protease variants analysed in our study may have to be tested in other subtype backbones- such as the HIV-1 subtype C MJ4 to comprehensively investigate the influence of backbone subtype on subtype-specific differences we observed in our study. It should be noted that the current study was restricted to the HIV-1 subtypes and intersubtype recombinants that are circulating in East Africa. However, the HIV-1 pandemic is composed of numerous other subtypes that are also unevenly distributed globally. Further studies will be needed to better understand subtype-specific Gag-driven fitness differences more comprehensively.

Although we found subtype-specific differences in this study based on Gag-protease-driven viral fitness, it remains unclear whether there is correlation between Gag-protease-driven viral replication capacity and whole virus isolate fitness. Future studies will be needed to address this question. For such future studies, whole virus isolates will need to be generated from study participants, with subsequent measurement of the fitness of the viruses by assays such as the one utilized in our study or by ex-vivo competition assays as has been previously attempted with a limited number of isolates [419]. Another approach would be to generate infectious molecular clones that represent different subtypes. Although technically challenging, this approach has the advantage that specific differences between subtypes can further be interrogated by molecular approaches such as site-directed mutagenesis studies that are not possible with virus isolates.

Furthermore, in this study, we found that a significant percentage of Gag-protease sequences analysed were intersubtype recombinants. It is unclear what subtype these viruses are in other genetic loci. Full-length sequencing will be needed to address this question and to provide additional data that may be informative for better understanding of HIV pathogenesis and vaccine design.

It should be noted that although we found subtype-specific differences in Gag-protease-driven virus replication capacity, the underlying biological mechanisms remain poorly understood. Recent studies have shown that Gag-driven replication capacity is associated with viral load set point and with markers of disease progression in recently infected individual [316, 363, 420]. Additionally, in recently infected persons, high replicative fitness virus was associated with higher levels of immune activation and proviral load [420]. Future studies will therefore need to investigate the specific host pathways impacted by highly fit virus to cause detrimental clinical effects and how this can be reversed through vaccine or therapeutic interventions.

Our data also suggests that future studies are required to define protective HLA alleles in different geographic locations for different subtypes. In this study, we analysed for differences in viral replicative fitness according to HLA class I alleles for HIV-1 subtype A. However, considering the enormous genetic heterogeneity in the HLA class I alleles, this question was not fully resolved and a larger cohort with diverse subtype is needed to fully interrogate whether there are differences in viral fitness according to HLA class I alleles for different subtypes. It is also possible that other genetic loci may influence viral fitness and future studies should address this issue. Furthermore, we recommend that site-directed mutagenesis studies are undertaken to confirm the fitness consequences of the amino acid variants that we found to be statistically associated with altered replication capacity and to further elucidate the fitness effect of subtype A-specific escape variants in the TW10 epitope. We propose that mutations should be introduced into a patient-derived subtype A gag-protease sequence that is similar to the consensus A sequence. In addition to testing the most common escape variants at codons 243 (243T) and 247

(247L), we recommend that mutagenesis is performed to test for the ability of 248Q and 248A to compensate for the hypothesised fitness cost of 247L, since it was observed that the 247L mutation was stable in the presence of the 248Q/248A mutations and that when both mutations were present the time for CD4 count to drop below 500 was less than when 247X was present alone (L McKinnon 2014, personal communication, 13 August 2014).

4.10 Conclusions/summary

In summary, we provide evidence that Gag-protease is an important determinant of viral fitness and show that it differs substantially in functionality between HIV-1 subtypes. We demonstrate a hierarchy of Gag-protease driven replicative fitness in East African populations where subtypes A/C are less fit than D, which is also less fit than inter-subtype recombinants. Since this hierarchy is consistent with reported subtype differences in disease progression in East Africa, our data supports the finding that Gag-protease-driven replication capacity is a determinant of differences in disease progression between subtypes. It is likely that the lower functionality of subtype A and C Gag-proteases slows disease progression in individuals infected with these subtypes, leading to greater opportunity for transmission and consequently, increased prevalence of these subtypes. Our study thus sheds light on mechanisms underlying the differential spread and expansion of HIV-1 subtypes in the global epidemic. It is therefore evident that an interplay of viral factors as well as host genetics factors, correlates of immune protection should all be incorporated in future vaccine strategies.

.

APPENDICES

APPENDIX

Table A1. HLA-associated amino acids in HIV-1 subtype A Gag-protease

Protein	Association	HLA allele	Codon	Target amino acid	Consensus	p-value	q-value
Gag	Adapted	A*7401	20	Q	R	1.99E-04	0.1555131
Gag	Adapted	C*0602	26	R	K	1.34E-04	0.1518822
Gag	Adapted	A*0301	28	Q	K	2.33E-08	1.40E-04
Gag	Adapted	A*3001	28	Q	K	1.77E-09	1.55E-05
Gag	Adapted	C*0802	49	G	G	1.60E-05	0.0295338
Gag	Adapted	A*0202	75	I	L	5.64E-05	7.08E-02
Gag	Adapted	B*5101	75	I	L	8.94E-05	0.1207346
Gag	Adapted	A*7401	91	K	R	1.81E-05	0.0303098
Gag	Adapted	A*0301	93	A	D	1.54E-04	0.1575459
Gag	Adapted	B*5703	93	E	D	1.56E-04	0.1575459
Gag	Adapted	C*0304	114	Q	K	1.66E-04	0.161668
Gag	Adapted	C*0602	146	N	A	1.11E-04	0.134405
Gag	Adapted	A*6802	223	I	I	9.41E-05	0.1224212
Gag	Adapted	B*5702	242	N	T	1.93E-06	0.0061577
Gag	Adapted	B*3501	260	E	D	9.22E-06	0.0193479
Gag	Adapted	B*4415	310	S	S	5.34E-10	6.25E-06
Gag	Adapted	B*4901	310	S	S	4.71E-06	0.0110699
Gag	Adapted	B*4415	312	D	D	2.09E-05	2.93E-02
Gag	Adapted	B*4901	312	D	D	3.28E-07	0.001048
Gag	Adapted	C*0210	312	E	D	1.95E-04	0.1555131
Gag	Adapted	C*0401	322	M	L	1.59E-04	0.1434228
Gag	Adapted	B*0702	357	G	S	4.36E-05	5.67E-02
Gag	Adapted	C*0702	357	G	S	1.62E-06	0.0061577
Gag	Adapted	C*0407	403	K	R	2.68E-04	0.1918226
Gag	Adapted	A*3001	484	G	Y	1.50E-04	0.1575459

Table A2. HLA-associated amino acids in HIV-1 subtype C Gag-protease

Protein	Association	HLA allele	Codon	Target amino acid	p-value	q-value
Gag	Adapted	B*1503	7	V	0.000452	0.105734
Gag	Adapted	B*1503	11	G	0.000954	0.176149
Gag	Adapted	A*74	12	N	3.04E-10	4.32E-07
Gag	Adapted	B*5802	14	E	0.000942	0.175546
Gag	Adapted	C*06	20	K	0.000796	0.165084
Gag	Adapted	A*6801	20	R	0.000462	0.105894
Gag	Adapted	A*74	20	K	0.000352	0.094699
Gag	Adapted	B*13	28	K	7.37E-05	0.025748
Gag	Adapted	B*42	28	Q	7.02E-08	6.82E-05
Gag	Adapted	C*17	28	S	0.000195	0.0564
Gag	Adapted	B*42	30	K	5.92E-05	0.021819
Gag	Adapted	A*3001	54	A	0.000972	0.186114
Gag	Adapted	A*6802	54	T	0.000517	0.115996
Gag	Adapted	B*1402	62	N	0.000456	0.105734
Gag	Adapted	C*16	67	S	0.000747	0.159962
Gag	Adapted	C*1601	67	S	0.000304	0.082405
Gag	Adapted	A*2902	79	Y	6.58E-06	0.003275
Gag	Adapted	A*74	93	G	0.000526	0.116515
Gag	Adapted	B*08	93	K	0.000446	0.109795
Gag	Adapted	A*3001	103	R	0.000785	0.164502
Gag	Adapted	B*5801	106	K	0.000108	0.035523
Gag	Adapted	C*04	128	A	0.000486	0.115298
Gag	Adapted	B*57	146	P	3.81E-10	5.05E-07
Gag	Adapted	A*2911	147	L	8.94E-06	0.003967
Gag	Adapted	B*57	147	L	4.62E-06	0.00236
Gag	Adapted	C*15	147	L	0.000683	0.147808
Gag	Adapted	B*1503	147	I	8.96E-06	0.003967
Gag	Adapted	B*5703	163	G	2.29E-05	0.009293
Gag	Adapted	B*5703	163	N	0.00039	0.101329
Gag	Adapted	A*0202	165	N	0.000884	0.1743
Gag	Adapted	B*81	182	S	6.08E-05	0.022027
Gag	Adapted	B*81	186	S	2.04E-09	2.54E-06
Gag	Adapted	C*18	190	A	7.6E-06	0.003603
Gag	Adapted	A*43	215	M	0.000935	0.175546
Gag	Adapted	C*07	215	I	0.000204	0.058118
Gag	Adapted	B*5801	219	P	0.000306	0.082405
Gag	Adapted	A*2911	223	I	0.00055	0.119207
Gag	Adapted	B*57	242	N	3.2E-20	2.12E-16
Gag	Adapted	B*5801	242	N	3.76E-24	7.48E-20
Gag	Adapted	B*4201	252	A	0.000646	0.135608

Gag	Adapted	B*35	260	E	1.36E-05	0.005754
Gag	Adapted	B*14	302	R	1.27E-07	0.000105
Gag	Adapted	C*0304	303	A	6.99E-06	0.003397
Gag	Adapted	B*44	312	E	1.83E-11	3.64E-08
Gag	Adapted	B*15	339	A	7.63E-05	0.028181
Gag	Adapted	C*0304	340	A	2.26E-06	0.00125
Gag	Adapted	B*07	357	G	2.91E-13	8.27E-10
Gag	Adapted	B*1401	370	A	0.00024	0.067542
Gag	Adapted	A*3002	371	N	0.00018	0.054374
Gag	Adapted	B*5802	377	M	0.000317	0.083085
Gag	Adapted	A*01	382	K	3.3E-06	0.001731
Gag	Adapted	B*4201	386	S	1.23E-06	0.000718
Gag	Adapted	A*74	403	K	7.47E-11	1.24E-07
Gag	Adapted	A*03	403	R	0.001058	0.197033
Gag	Adapted	A*3001	403	R	2.2E-07	0.000169
Gag	Adapted	A*0301	411	R	0.000908	0.175546
Gag	Adapted	B*13	437	L	0.000823	0.16829
Gag	Adapted	A*74	441	N	4.94E-07	0.000352
Gag	Adapted	C*0304	467	G	0.000836	0.16829
Gag	Adapted	C*18	474	P	0.000835	0.16829
Gag	Adapted	A*6802	477	D	0.000668	0.138729
Gag	Adapted	C*04	478	M	0.001036	0.187919
Gag	Adapted	C*17	487	I	4.92E-05	0.019227
Gag	Adapted	B*4201	488	A	0.000392	0.101329
Protease	Adapted	B*44	35	D	3.52E-06	0.004092
Protease	Adapted	B*45	63	H	6.3E-05	0.048816

Table A3. HLA-associated amino acids in HIV-1 subtype D Gag-protease

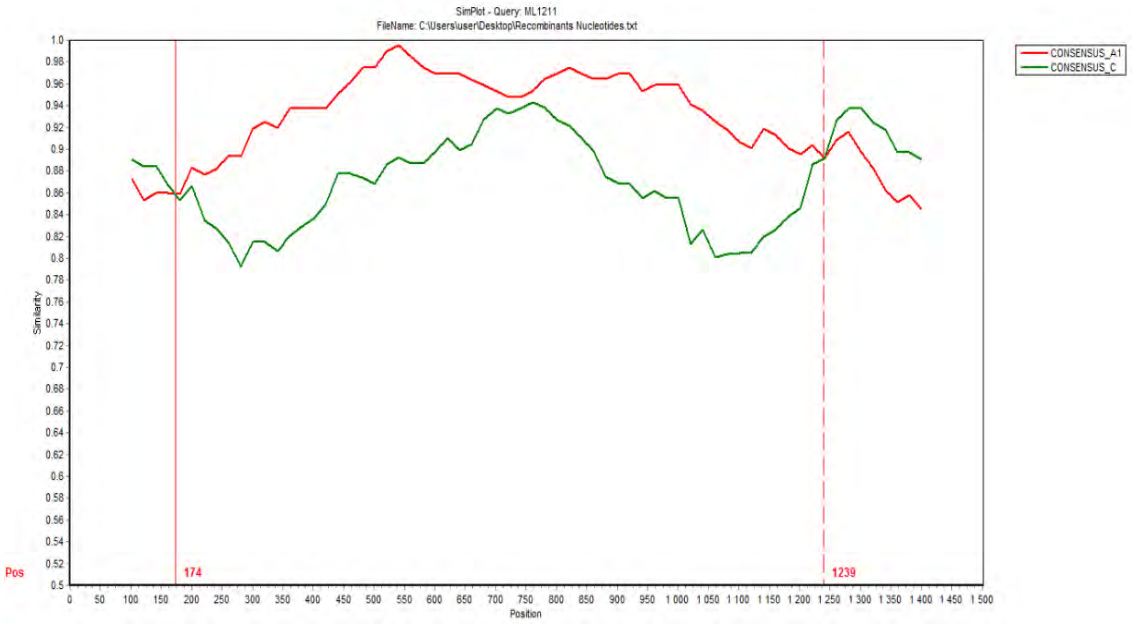
Protein	Association	HLA allele	Codon	Target amino acid	Consensus	p-value	q--value
Gag	Adapted	B*4101	9	R	S	3.47E-05	0.036354
Gag	Adapted	A*7401	20	Q	R	8.45E-08	0.000354
Gag	Adapted	B*4415	74	G	E	0.00028	0.177824
Gag	Adapted	C*0802	119	A	A	0.000294	0.180753
Gag	Adapted	B*5703	146	P	A	2.71E-05	0.031549
Gag	Adapted	B*1402	147	I	L	0.000187	0.126287
Gag	Adapted	B*5703	161	D	E	1.07E-06	0.00258
Gag	Adapted	B*8101	186	S	M	1.36E-05	0.020358
Gag	Adapted	A*0101	219	p	H	0.000139	0.100053
Gag	Adapted	B*5703	242	N	T	1.77E-10	1.24E-06
Gag	Adapted	B*5101	326	S	A	1.02E-04	0.076158
Gag	Adapted	B*5703	342	S	T	1.11E-05	0.017874
Gag	Adapted	B*0702	357	G	S	3.44E-12	7.02E-08
Gag	Adapted	A*0201	389	T	I	7.77E-06	0.013545
Gag	Adapted	A*3402	490	R	K	2.92E-06	0.005549

Table A4. Best-defined (A-list) optimal Gag epitopes from the Los Alamos HIV molecular immunology database

Epitope	Protein	HXB2 codon	Subtype	HLA restriction
GELDRWEKI	Gag	19-27		B*4002
KIRLRPGGK	Gag	18-26		A*0301
IRLRPGGKK	Gag	19-27	B	B*2705
RLRPGGKKK	Gag	20-28		A*0301
RLRPGGKKKY	Gag	20-29	B	A*0301
RPGGKKKYKL	Gag	22-31	B	B*5101
GGKKKYKLK	Gag	24-32	B	B*0801
KYKLKHIVW	Gag	28-36	B	A*2402
HLVWASREL	Gag	33-41		Cw*0804
LVWASRELERF	Gag	34-44		A30
WASRELERF	Gag	36-44	B	B*3501
ELRSLYNTV	Gag	74-82		B*0801
RSLYNTVATLY	Gag	76-86	B	A*3002, B58, B63
SLYNTVATL	Gag	77-85	B	A*0201, A*0202, A*0205
SLYNTVATLY	Gag	77-86	B	A*0201
LYNTVATL	Gag	78-85		Cw14
LYNTVATLY	Gag	78-86		A*2902, B*4403
TLYCVHQK	Gag	84-91		A*1101
IEIKDTKEAL	Gag	92-101		B*4001
NSSKVSQNY	Gag	124-132	B	B*3501
VQNLQGQMV	Gag	135-143		B13
HQAISPRTL	Gag	144-152		B*1510
QAISPRTLNAW	Gag	145-155	B	A*2501
ISPRTLNAW	Gag	147-155		B*5701, B63
SPRTLNAWV	Gag	148-156		B*0702
VKVIEEKAF	Gag	156-164		B*1503
EEKAFSPEV	Gag	160-168		B*4415
KAFSPEVI	Gag	162-169	B	B*5703
KAFSPEVIPMF	Gag	162-172	B	B*5701, B*5703, B63
FSPEVIPMF	Gag	164-172		B57
EVIPMFSAL	Gag	167-175	B	A*2601, A*2602, A*2603
VIPMFSAL	Gag	168-175	B	Cw*0102
SEGATPQDL	Gag	176-184		B*4001
TPQDLNTML	Gag	180-188	B	B*0702, B*3910,

				B*4201, B*8101, Cw*0802
TPQDLNMML	Gag	180-188	A	B53
TPYDINQML	Gag	180-188	HIV-2	B*5301
GHQAAMQML	Gag	193-201	B	B*1510, B*3901
KETINEEAA	Gag	202-210		B*4002
ETINEEAAEW	Gag	203-212		A*2501
AEWDRVHPV	Gag	210-218		B*4002
HPVHAGPIA	Gag	216-224		B*3501, B7
GQMREPRGSDI	Gag	226-236		B13
TSTLQEQIGW	Gag	240-249	B	B*5701, B*5801
NPPIPVGDIY	Gag	253-262		B*3501
PPIPVGDIY	Gag	254-262	B	B*3501
EIYKRWII	Gag	260-267	B	B*0801
RRWIQLGLQK	Gag	263-272		B*2703
KRWIILGLNK	Gag	263-272	B	B*2705
GLNKIVRMY	Gag	269-277	B	B*1501, B62
VRMYSPVSI	Gag	274-282		Cw18
RMYSPTSI	Gag	275-282		B*5201
YSPVSILDI	Gag	277-285	CRF01_AE	Cw*0102
FRDYVDRFF	Gag	293-301		Cw18
FRDYVDRFYK	Gag	293-302	B, D	B*1801
RDYVDRFFKTL	Gag	294-304	A	A*2402
RDYVDRFYKTL	Gag	294-304	B	B*4402
YVDRFYKTL	Gag	296-304		A*0207
YVDRFFKTL	Gag	296-304		B*1503, Cw*0303, Cw*0304
DRFYKTLRA	Gag	298-306	B	B*1402
AEQASQDVKNW	Gag	306-316	B	B*4402
AEQASQEVKNWM	Gag	306-317		Cw5
QASQEVKNW	Gag	308-316	B	B*5301, B*5701, B*5801
VKNWMTETL	Gag	313-321	B	B*4801
DCKTILKAL	Gag	329-337	B	B*0801
ACQGVGGPGHK	Gag	349-359		A*1101
GPGHKARVL	Gag	355-363	B	B*0702
AEAMSQVTNS	Gag	364-373		B*4501

Figure A1: This bootscan graph shows the patient ML 1211 sequence as a representative simplot figure of the recombinants. The exact position of recombination breakpoints are illustrated by the red dotted lines.


SimPlot - Query: ML1211
FileName: C:\Users\user\Desktop\Recombinants Nucleotides.txt
CONSENSUS_A1
CONSENSUS_C
Similarity
Position
Pos
174
1239

REFERENCES

1. Douek, D.C., M. Roederer, and R.A. Koup, *Emerging concepts in the immunopathogenesis of AIDS.* Annual review of medicine, 2009. **60**: p. 471.
2. Chitnis, A., D. Rawls, and J. Moore, *Origin of HIV type 1 in colonial French Equatorial Africa?* AIDS Res Hum Retroviruses, 2000. **16**(1): p. 5-8.
3. Hahn, B.H., et al., *AIDS as a zoonosis: scientific and public health implications.* Science, 2000. **287**(5453): p. 607-614.
4. Sharp, P.M. and B.H. Hahn, *AIDS: prehistory of HIV-1.* Nature, 2008. **455**(7213): p. 605-606.
5. Lemey, P., et al., *Tracing the origin and history of the HIV-2 epidemic.* Proceedings of the National Academy of Sciences, 2003. **100**(11): p. 6588-6592.
6. Henrard, D.R., et al., *Natural history of HIV-1 cell-free viremia.* Jama, 1995. **274**(7): p. 554-558.
7. Montagnier, L., *A history of HIV discovery.* Science, 2002. **298**(5599): p. 1727-1728.
8. Wain-Hobson, S., et al., *LAV revisited: origins of the early HIV-1 isolates from Institut Pasteur.* Science, 1991. **252**(5008): p. 961-965.
9. Chang, S.-Y.P., et al., *The origin of HIV-1 isolate HTLV-IIIB.* Nature, 1993. **363**(6428): p. 466-469.
10. Anderson, R., et al., *A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS.* Mathematical Medicine and Biology, 1986. **3**(4): p. 229-263.
11. Barré-Sinoussi, F., et al., *Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).* Science, 1983. **220**(4599): p. 868-871.
12. Popovic, M., et al., *Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS.* Science, 1984. **224**(4648): p. 497-500.
13. Wain-Hobson, S., et al., *Nucleotide sequence of the AIDS virus, LAV.* Cell, 1985. **40**(1): p. 9-17.
14. UNAIDS. *How AIDS changed everything fact sheet*. 2015 [cited 2015 November]; Available from: http://www.unaids.org/en/resources/campaigns/HowAIDSchangedeverything/factsheet.
15. Kimanga, D.O., et al., *Prevalence and incidence of HIV infection, trends, and risk factors among persons aged 15–64 years in Kenya: results from a nationally representative study.* JAIDS Journal of Acquired Immune Deficiency Syndromes, 2014. **66**: p. S13-S26.
16. Waruiru, W., et al., *The status of HIV testing and counseling in Kenya: results from a nationally representative population-based survey.* JAIDS Journal of Acquired Immune Deficiency Syndromes, 2014. **66**: p. S27-S36.
17. Mahy, M., et al., *What will it take to achieve virtual elimination of mother-to-child transmission of HIV? An assessment of current progress and future needs.* Sexually transmitted infections, 2010. **86**(Suppl 2): p. ii48-ii55.
18. May, M.T. and S.M. Ingle, *Life expectancy of HIV-positive adults: a review.* Sexual health, 2011. **8**(4): p. 526-533.
19. Organization, W.H., *March 2014 supplement to the 2013 consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach.* 2014.
20. Post, J. and M. Kelly, *New developments in antiretroviral therapy for HIV infection.* Australian Prescriber, 2005. **28**(6): p. 146.
21. Organization, W.H., *Antiretroviral therapy of HIV infection in infants and children: towards universal access: recommendations for a public health approach-2010 revision*2010: World Health Organization.

22. Barré-Sinoussi, F., A.L. Ross, and J.-F. Delfraissy, *Past, present and future: 30 years of HIV research.* Nature Reviews Microbiology, 2013. **11**(12): p. 877-883.
23. Hirsch, V.M., et al., *An African primate lentivirus (SIVsmclosely related to HIV-2.* 1989.
24. Campbell-Yesufu, O.T. and R.T. Gandhi, *Update on human immunodeficiency virus (HIV)-2 infection.* Clinical Infectious Diseases, 2011. **52**(6): p. 780-787.
25. Sharp, P., et al., *Origins and evolution of AIDS viruses: estimating the time-scale.* Biochemical Society Transactions, 2000. **28**(1): p. 24-24.
26. Apetrei, C., et al., *Detection and partial characterization of simian immunodeficiency virus SIVsm strains from bush meat samples from rural Sierra Leone.* Journal of virology, 2005. **79**(4): p. 2631-2636.
27. Santiago, M.L., et al., *Simian immunodeficiency virus infection in free-ranging sooty mangabeys (Cercocebus atys atys) from the Tai Forest, Cote d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2.* Journal of virology, 2005. **79**(19): p. 12515-12527.
28. Chen, Z., et al., *Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop.* Journal of virology, 1996. **70**(6): p. 3617-3627.
29. Kanki, P., et al., *Slower heterosexual spread of HIV-2 than HIV-1.* The Lancet, 1994. **343**(8903): p. 943-946.
30. Marlink, R., et al., *Reduced rate of disease development after HIV-2 infection as compared to HIV-1.* Science, 1994. **265**(5178): p. 1587-1590.
31. Matheron, S., et al., *Factors associated with clinical progression in HIV-2 infected-patients: the French ANRS cohort.* Aids, 2003. **17**(18): p. 2593-2601.
32. Popper, S.J., et al., *Lower human immunodeficiency virus (HIV) type 2 viral load reflects the difference in pathogenicity of HIV-1 and HIV-2.* Journal of Infectious Diseases, 1999. **180**(4): p. 1116-1121.
33. Conway, S. and J.G. Bartlett, *The 2003/2004 Southern African abbreviated guide to medical management of HIV infection*2003: Johns Hopkins University, Division of Infectious Diseases.
34. Damond, F., et al., *Identification of a highly divergent HIV type 2 and proposal for a change in HIV type 2 classification.* AIDS Res Hum Retroviruses, 2004. **20**(6): p. 666-672.
35. Huet, T., et al., *Genetic organization of a chimpanzee lentivirus related to HIV-1.* 1990.
36. Gao, F., et al., *Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes.* Nature, 1999. **397**(6718): p. 436-441.
37. Korber, B., et al., *Timing the ancestor of the HIV-1 pandemic strains.* Science, 2000. **288**(5472): p. 1789-1796.
38. Rambaut, A., et al., *Human immunodeficiency virus: phylogeny and the origin of HIV-1.* Nature, 2001. **410**(6832): p. 1047-1048.
39. Zhu, T., et al., *An African HIV-1 sequence from 1959 and implications for the origin of the epidemic.* Nature, 1998. **391**(6667): p. 594-597.
40. Plantier, J.-C., et al., *A new human immunodeficiency virus derived from gorillas.* Nature medicine, 2009. **15**(8): p. 871-872.
41. Simon, F., et al., *Identification of a new human immunodeficiency virus type 1 distinct from group M and group O.* Nature medicine, 1998. **4**(9): p. 1032-1037.
42. Niama, F.R., et al., *HIV-1 subtypes and recombinants in the Republic of Congo.* Infection, Genetics and Evolution, 2006. **6**(5): p. 337-343.
43. Vidal, N., et al., *Distribution of HIV-1 variants in the Democratic Republic of Congo suggests increase of subtype C in Kinshasa between 1997 and 2002.* JAIDS Journal of Acquired Immune Deficiency Syndromes, 2005. **40**(4): p. 456-462.
44. Hemelaar, J., et al., *Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004.* Aids, 2006. **20**(16): p. W13-W23.

45. Kuiken, C., et al., *HIV sequence compendium 2009.* Los Alamos, New Mexico: Los Alamos National Laboratory, Theoretical Biology and Biophysics, 2009.
46. McCutchan, F.E., *Global epidemiology of HIV.* J Med Virol, 2006. **78**(S1): p. S7-S12.
47. Perrin, L., L. Kaiser, and S. Yerly, *Travel and the spread of HIV-1 genetic variants.* The Lancet infectious diseases, 2003. **3**(1): p. 22-27.
48. Tebit, D.M. and E.J. Arts, *Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease.* The Lancet infectious diseases, 2011. **11**(1): p. 45-56.
49. Hemelaar, J., *The origin and diversity of the HIV-1 pandemic.* Trends in molecular medicine, 2012. **18**(3): p. 182-192.
50. Salemi, M., et al., *Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution.* The FASEB Journal, 2001. **15**(2): p. 276-278.
51. Meloni, S.T., et al., *Distinct human immunodeficiency virus type 1 subtype A virus circulating in West Africa: sub-subtype A3.* Journal of virology, 2004. **78**(22): p. 12438-12445.
52. Vidal, N., et al., *Identification and molecular characterization of subsubtype A4 in central Africa.* AIDS Research & Human Retroviruses, 2006. **22**(2): p. 182-187.
53. Vidal, N., et al., *Genetic characterization of eight full-length HIV type 1 genomes from the Democratic Republic of Congo (DRC) reveal a new subsubtype, A5, in the A radiation that predominates in the recombinant structure of CRF26_A5U.* AIDS Res Hum Retroviruses, 2009. **25**(8): p. 823-832.
54. Mokili, J.L., et al., *Identification of a novel clade of human immunodeficiency virus type 1 in Democratic Republic of Congo.* AIDS Res Hum Retroviruses, 2002. **18**(11): p. 817-823.
55. Los Alamos National Laboratory. *The HIV Sequence Database*. 2015 [cited 2015 November 5]; Available from: http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html.
56. Peeters, M. and P.M. Sharp, *Genetic diversity of HIV-1: the moving target.* Aids, 2000(Suppl. 3): p. S129-S140.
57. Wainberg, M.A., *HIV-1 subtype distribution and the problem of drug resistance.* Aids, 2004. **18**: p. S63-S68.
58. Phanuphak, N., et al., *HIV epidemic in Asia: implications for HIV vaccine and other prevention trials.* AIDS Res Hum Retroviruses, 2015.
59. Taylor, B.S., et al., *The challenge of HIV-1 subtype diversity.* New England Journal of Medicine, 2008. **358**(15): p. 1590-1602.
60. Rhee, S.-Y., et al., *Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors.* Retrovirology, 2008. **5**(1): p. 74.
61. Herbinger, K.-H., et al., *Frequency of HIV type 1 dual infection and HIV diversity: analysis of low-and high-risk populations in Mbeya Region, Tanzania.* AIDS Research & Human Retroviruses, 2006. **22**(7): p. 599-606.
62. Marx, P.A., P.G. Alcabes, and E. Drucker, *Serial human passage of simian immunodeficiency virus by unsterile injections and the emergence of epidemic human immunodeficiency virus in Africa.* Philosophical Transactions of the Royal Society of London B: Biological Sciences, 2001. **356**(1410): p. 911-920.
63. Apetrei, C., D.L. Robertson, and P.A. Marx, *The history of SIVS and AIDS: epidemiology, phylogeny and biology of isolates from naturally SIV infected non-human primates (NHP) in Africa.* Front Biosci, 2004. **9**: p. 225-254.
64. Preston, B.D., B.J. Poiesz, and L.A. Loeb, *Fidelity of HIV-1 reverse transcriptase.* Science, 1988. **242**(4882): p. 1168-1171.
65. Santoro, M.M. and C.F. Perno, *HIV-1 genetic variability and clinical implications.* ISRN microbiology, 2013. **2013**.

66. Salemi, M., *The intra-host evolutionary and population dynamics of human immunodeficiency virus type 1: a phylogenetic perspective.* Infectious disease reports, 2013. **5**(Suppl 1).
67. Domingo, E. and J. Holland, *RNA virus mutations and fitness for survival.* Annual Reviews in Microbiology, 1997. **51**(1): p. 151-178.
68. Nowak, M. and R.M. May, *Virus dynamics: mathematical principles of Immunology and virology: mathematical principles of Immunology and virology*2000: Oxford university press.
69. Spira, S., et al., *Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance.* Journal of Antimicrobial Chemotherapy, 2003. **51**(2): p. 229-240.
70. Tatt, I.D., et al., *The public health significance of HIV-1 subtypes.* Aids, 2001. **15**: p. S59-S71.
71. Peut, V., et al., *Balancing reversion of CTL and neutralizing antibody escape mutations within HIV-1 Env upon transmission.* Journal of virology, 2009.
72. Richman, D.D., et al., *Rapid evolution of the neutralizing antibody response to HIV type 1 infection.* Proceedings of the National Academy of Sciences, 2003. **100**(7): p. 4144-4149.
73. Altman, J.D. and M.B. Feinberg, *HIV escape: there and back again.* Nature medicine, 2004. **10**(3): p. 229-230.
74. McBurney, S.P. and T.M. Ross, *Viral sequence diversity: challenges for AIDS vaccine designs.* 2008.
75. Allen, T.M., et al., *Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution.* Journal of virology, 2005. **79**(21): p. 13239-13249.
76. Johnson, V.A., et al., *Update of the drug resistance mutations in HIV-1: December 2009.* Top HIV Med, 2009. **17**(5): p. 138-145.
77. Johnson, V.A., et al., *Update of the drug resistance mutations in HIV-1: Fall 2006.* Top HIV Med, 2006. **14**(3): p. 125-130.
78. Johnson, V.A., et al., *2011 update of the drug resistance mutations in HIV-1.* Topics in antiviral medicine, 2011. **19**(4): p. 156-164.
79. Kuiken, C., et al., *Genetic analysis reveals epidemiologic patterns in the spread of human immunodeficiency virus.* Am J Epidemiol, 2000. **152**(9): p. 814-822.
80. U.S. Military HIV Research Program. 2015 [cited 2016 JANUARY]; Available from: http://www.vaccineenterprise.org/sites/default/files/150316_S3_Modjarrad.Kayvon.pdf.
81. Hunt, R., *Virology Chapter Seven Human Immunodeficiency Virus and AIDS Part One Introduction.* 2010.
82. UNAIDS. *2015* HIV Stastics. Available from: http://www.unaids.org/en/resources/campaigns/HowAIDSchangedeverything/factsheet.
83. HIV/AIDS, J.U.N.P.o., *Global report: UNAIDS report on the global AIDS epidemic 2013. Geneva, Switzerland: UNAIDS, 2013*, 2014.
84. Esbjörnsson, J., et al., *HIV-1 molecular epidemiology in Guinea-Bissau, West Africa: origin, demography and migrations.* 2011.
85. Abecasis, A.B., et al., *HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics.* Retrovirology, 2013. **10**(1): p. 7.
86. Essex, M. and S. Mboup, *Regional variations in the African epidemics.* AIDS in Africa, 2002. **2**: p. 631-640.
87. Lihana, R.W., et al., *Update on HIV-1 diversity in Africa: a decade in review.* Aids Rev, 2012. **14**(2): p. 83-100.

88. Kiwanuka, N., et al., *HIV-1 subtypes and differences in heterosexual HIV transmission among HIV-discordant couples in Rakai, Uganda.* AIDS (London, England), 2009. **23**(18): p. 2479.
89. Masimba, P., et al., *Prevalence of drug resistance mutations and HIV type 1 subtypes in an HIV type 1-infected cohort in rural Tanzania.* AIDS Res Hum Retroviruses, 2013. **29**(9): p. 1229-1236.
90. Hu, D.J., et al., *Predominance of HIV-1 subtype A and D infections in Uganda.* Emerging infectious diseases, 2000. **6**(6): p. 609.
91. Nyombi, B.M., et al., *Diversity of human immunodeficiency virus type 1 subtypes in Kagera and Kilimanjaro regions, Tanzania.* AIDS Res Hum Retroviruses, 2008. **24**(6): p. 761-769.
92. Khoja, S., et al., *Genetic analysis of HIV-1 subtypes in Nairobi, Kenya.* PLoS One, 2008. **3**(9): p. e3191.
93. Beyrer, C., et al., *Overland heroin trafficking routes and HIV-1 spread in south and south-east Asia.* Aids, 2000. **14**(1): p. 75-83.
94. Buonaguro, L., M. Tornesello, and F. Buonaguro, *Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications.* Journal of virology, 2007. **81**(19): p. 10209-10219.
95. Piot, P. and M. Bartos, *The epidemiology of HIV and AIDS*, in *AIDS in Africa*2002, Springer. p. 200-217.
96. Tanaka, Y., et al., *African origin of GB virus C/hepatitis G virus.* FEBS letters, 1998. **423**(2): p. 143-148.
97. Triques, K., et al., *Near-full-length genome sequencing of divergent African HIV type 1 subtype F viruses leads to the identification of a new HIV type 1 subtype designated K.* AIDS Res Hum Retroviruses, 2000. **16**(2): p. 139-151.
98. Bikandou, B., et al., *Genetic subtypes of HIV type 1 in Republic of Congo.* AIDS Res Hum Retroviruses, 2000. **16**(7): p. 613-619.
99. Yang, R., et al., *On-going generation of multiple forms of HIV-1 intersubtype recombinants in the Yunnan Province of China.* Aids, 2002. **16**(10): p. 1401-1407.
100. Leitner, T., et al., *HIV-1 subtype and circulating recombinant form (CRF) reference sequences, 2005.* HIV sequence compendium, 2005. **2005**: p. 41-48.
101. Dowling, W.E., et al., *Forty-one near full-length HIV-1 sequences from Kenya reveal an epidemic of subtype A and A-containing recombinants.* Aids, 2002. **16**(13): p. 1809-1820.
102. Ragupathy, V., et al., *Identification of new, emerging HIV-1 unique recombinant forms and drug resistant viruses circulating in Cameroon.* Virol J, 2011. **8**(1): p. 185.
103. Jacobs, G.B., et al., *HIV-1 subtypes B and C unique recombinant forms (URFs) and transmitted drug resistance identified in the Western Cape Province, South Africa.* PLoS One, 2014. **9**(3): p. e90845.
104. Hemelaar, J., et al., *Global trends in molecular epidemiology of HIV-1 during 2000–2007.* AIDS (London, England), 2011. **25**(5): p. 679.
105. Baeten, J.M., et al., *HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads.* Journal of Infectious Diseases, 2007. **195**(8): p. 1177-1180.
106. Yang, C., et al., *Genetic diversity of HIV-1 in western Kenya: subtype-specific differences in mother-to-child transmission.* Aids, 2003. **17**(11): p. 1667-1674.
107. Chohan, B., et al., *Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels.* Journal of virology, 2005. **79**(10): p. 6528-6531.
108. Songok, E.M., et al., *Identification of env CRF-10 among HIV variants circulating in rural western Kenya.* AIDS Res Hum Retroviruses, 2003. **19**(2): p. 161-165.
109. Khamadi, S.A., et al., *HIV type 1 subtypes in circulation in northern Kenya.* AIDS Research & Human Retroviruses, 2005. **21**(9): p. 810-814.

110. Mosha, F., et al., *Prevalence of genotypic resistance to antiretroviral drugs in treatment-naive youths infected with diverse HIV type 1 subtypes and recombinant forms in Dar es Salaam, Tanzania.* AIDS Res Hum Retroviruses, 2011. **27**(4): p. 377-382.
111. Kiwelu, I.E., et al., *HIV-1 subtypes and recombinants in Northern Tanzania: distribution of viral quasispecies.* 2012.
112. Ssemwanga, D., et al., *Multiple HIV-1 infections with evidence of recombination in heterosexual partnerships in a low risk Rural Clinical Cohort in Uganda.* Virology, 2011. **411**(1): p. 113-131.
113. Shao, E.R., et al., *Vaccines & Vaccination.* 2014.
114. Harris, M.E., et al., *Among 46 near full length HIV type 1 genome sequences from Rakai District, Uganda, subtype D and AD recombinants predominate.* AIDS Res Hum Retroviruses, 2002. **18**(17): p. 1281-1290.
115. Kemal, K.S., et al., *Molecular Epidemiology of HIV Type 1 Subtypes in Rwanda.* AIDS Res Hum Retroviruses, 2013. **29**(6): p. 957-962.
116. Neilson, J.R., et al., *Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya.* Journal of virology, 1999. **73**(5): p. 4393-4403.
117. Emery, S., et al., *Evaluation of performance of the Gen-Probe human immunodeficiency virus type 1 viral load assay using primary subtype A, C, and D isolates from Kenya.* Journal of clinical microbiology, 2000. **38**(7): p. 2688-2695.
118. Lihana, R.W., et al., *HIV type 1 subtype diversity and drug resistance among HIV type 1-infected Kenyan patients initiating antiretroviral therapy.* AIDS Res Hum Retroviruses, 2009. **25**(12): p. 1211-1217.
119. hunt, D.R., *Microbiology and Immunology online book.*
120. splattstoesser, T.
121. Sierra, S., B. Kupfer, and R. Kaiser, *Basics of the virology of HIV-1 and its replication.* Journal of Clinical Virology, 2005. **34**(4): p. 233-244.
122. Watts, J.M., et al., *Architecture and secondary structure of an entire HIV-1 RNA genome.* Nature, 2009. **460**(7256): p. 711-716.
123. Gomez, C. and T.J. Hope, *The ins and outs of HIV replication.* Cellular microbiology, 2005. **7**(5): p. 621-626.
124. Chen, D., et al., *HIV-1 Tat targets microtubules to induce apoptosis, a process promoted by the pro-apoptotic Bcl-2 relative Bim.* The EMBO journal, 2002. **21**(24): p. 6801-6810.
125. Emerman, M. and M.H. Malim, *HIV-1 regulatory/accessory genes: keys to unraveling viral and host cell biology.* Science, 1998. **280**(5371): p. 1880-1884.
126. Abbas, A.K., A.H. Lichtman, and S. Pillai, *Basic immunology: functions and disorders of the immune system*2012: Elsevier Health Sciences.
127. Ariën, K.K., et al., *The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates.* Journal of virology, 2005. **79**(14): p. 8979-8990.
128. Rose, K.M., et al., *The viral infectivity factor (Vif) of HIV-1 unveiled.* Trends in molecular medicine, 2004. **10**(6): p. 291-297.
129. Zhao, R.Y. and M.I. Bukrinsky, *HIV-1 Accessory Proteins: VpR*, in *Human Retroviruses*2014, Springer. p. 125-134.
130. Morellet, N., et al., *NMR structure of the HIV-1 regulatory protein VPR.* J Mol Biol, 2003. **327**(1): p. 215-227.
131. Muthumani, K., et al., *HIV-1 Vpr induces apoptosis through caspase 9 in T cells and peripheral blood mononuclear cells.* Journal of Biological Chemistry, 2002. **277**(40): p. 37820-37831.
132. Park, S.H., et al., *Three-dimensional structure of the channel-forming trans-membrane domain of virus protein "u"(Vpu) from HIV-1.* J Mol Biol, 2003. **333**(2): p. 409-424.
133. Bour, S. and K. Strebel, *The HIV-1 Vpu protein: a multifunctional enhancer of viral particle release.* Microbes and Infection, 2003. **5**(11): p. 1029-1039.

134. Karetnikov, A. and M. Suomalainen, *Tethered virions are intermediates in the assembly and release of HIV-1 particles.* Virology, 2010. **407**(2): p. 289-295.
135. Neil, S.J., T. Zang, and P.D. Bieniasz, *Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu.* Nature, 2008. **451**(7177): p. 425-430.
136. Goujon, C., et al., *SIVSM/HIV-2 Vpx proteins promote retroviral escape from a proteasome-dependent restriction pathway present in human dendritic cells.* Retrovirology, 2007. **4**(1): p. 2.
137. Fiorentini, S., et al., *Functions of the HIV-1 matrix protein p17.* MICROBIOLOGICA-BOLOGNA-, 2006. **29**(1): p. 1.
138. Freed, E.O., G. Englund, and M.A. Martin, *Role of the basic domain of human immunodeficiency virus type 1 matrix in macrophage infection.* Journal of virology, 1995. **69**(6): p. 3949-3954.
139. von Schwedler, U.K., et al., *Functional surfaces of the human immunodeficiency virus type 1 capsid protein.* Journal of virology, 2003. **77**(9): p. 5439-5450.
140. Ganser-Pornillos, B.K., M. Yeager, and W.I. Sundquist, *The structural biology of HIV assembly.* Current opinion in structural biology, 2008. **18**(2): p. 203-217.
141. Krishnamoorthy, G., et al., *DNA condensation by the nucleocapsid protein of HIV-1: a mechanism ensuring DNA protection.* Nucleic Acids Res, 2003. **31**(18): p. 5425-5432.
142. Demirov, D.G., J.M. Orenstein, and E.O. Freed, *The late domain of human immunodeficiency virus type 1 p6 promotes virus release in a cell type-dependent manner.* Journal of virology, 2002. **76**(1): p. 105-117.
143. Wright, E.R., et al., *Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells.* The EMBO journal, 2007. **26**(8): p. 2218-2226.
144. Adamson, C.S., et al., *Research Polymorphisms in Gag spacer peptide 1 confer varying levels of resistance to the HIV-1maturation inhibitor bevirimat.* 2010.
145. Checkley, M.A., et al., *The capsid-spacer peptide 1 Gag processing intermediate is a dominant-negative inhibitor of HIV-1 maturation.* Virology, 2010. **400**(1): p. 137-144.
146. Chang, Y.-C., S.-L. Yu, and W.-J. Syu, *Organization of HIV-1pol is critical for pol polyprotein processing.* J Biomed Sci, 1999. **6**(5): p. 333-341.
147. Hehl, E.A., et al., *Interaction between human immunodeficiency virus type 1 reverse transcriptase and integrase proteins.* Journal of virology, 2004. **78**(10): p. 5056-5067.
148. Gulnik, S., J.W. Erickson, and D. Xie, *HIV protease: enzyme function and drug resistance.* Vitamins & Hormones, 2000. **58**: p. 213-256.
149. Brik, A. and C.-H. Wong, *HIV-1 protease: mechanism and drug discovery.* Organic & biomolecular chemistry, 2003. **1**(1): p. 5-14.
150. Hu, W.-S. and S.H. Hughes, *HIV-1 reverse transcription.* Cold Spring Harbor perspectives in medicine, 2012. **2**(10): p. a006882.
151. Sarafianos, S.G., et al., *Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition.* J Mol Biol, 2009. **385**(3): p. 693-713.
152. Chiu, T.K. and D.R. Davies, *Structure and function of HIV-1 integrase.* Current topics in medicinal chemistry, 2004. **4**(9): p. 965-977.
153. Chen, J.C.-H., et al., *Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding.* Proceedings of the National Academy of Sciences, 2000. **97**(15): p. 8233-8238.
154. Llano, M., et al., *Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75.* J Mol Biol, 2006. **360**(4): p. 760-773.
155. Weber, J., et al., *A Novel Method for the Simultaneous Quantification of Phenotypic Resistance to Maturation, Protease, Reverse Transcriptase, and Integrase HIV Inhibitors Based on 3' Gag (p2/p7/p1/p6)/PR/RT/INT-Recombinant Viruses: A Useful Tool in the Multi-Target Era of Antiretroviral Therapy.* Antimicrob Agents Chemother, 2011: p. AAC. 00396-11.
156. Chen, L.F., J. Hoy, and S.R. Lewin, *Ten years of highly active antiretroviral therapy for HIV infection.* Medical Journal of Australia, 2007. **186**(3): p. 146.

157. Stevceva, L., et al., *Immune responses to HIV Gp120 that facilitate viral escape.* Curr HIV Res, 2007. **5**(1): p. 47-54.
158. Liu, J., et al., *Molecular architecture of native HIV-1 gp120 trimers.* Nature, 2008. **455**(7209): p. 109-113.
159. Wyatt, R., et al., *The antigenic structure of the HIV gp120 envelope glycoprotein.* Nature, 1998. **393**(6686): p. 705-711.
160. Buzon, V., et al., *Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions.* 2010.
161. Barre-Sinoussi, F. 2013; Available from: http://www.nature.com/nrmicro/journal/v11/n12/fig_tab/nrmicro3132_F2.html.
162. Labrosse, B., et al., *Cooperation of the V1/V2 and V3 domains of human immunodeficiency virus type 1 gp120 for interaction with the CXCR4 receptor.* Journal of virology, 2001. **75**(12): p. 5457-5464.
163. Cormier, E.G. and T. Dragic, *The crown and stem of the V3 loop play distinct roles in human immunodeficiency virus type 1 envelope glycoprotein interactions with the CCR5 coreceptor.* Journal of virology, 2002. **76**(17): p. 8953-8957.
164. Gallo, S.A., et al., *The HIV Env-mediated fusion reaction.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2003. **1614**(1): p. 36-50.
165. Blumenthal, R., S. Durell, and M. Viard, *HIV entry and envelope glycoprotein-mediated fusion.* Journal of Biological Chemistry, 2012. **287**(49): p. 40841-40849.
166. Eckert, D.M. and P.S. Kim, *Mechanisms of viral membrane fusion and its inhibition.* Annual review of biochemistry, 2001. **70**(1): p. 777-810.
167. Boyd, M. and S. Pett, *HIV fusion inhibitors: a review.* Australian Prescriber, 2008. **31**(3): p. 66-69.
168. Bowman, M.-C., et al., *Inhibition of HIV fusion with multivalent gold nanoparticles.* J Am Chem Soc, 2008. **130**(22): p. 6896-6897.
169. Pang, W., S.-C. Tam, and Y.-T. Zheng, *Current peptide HIV type-1 fusion inhibitors.* Antiviral Chemistry and Chemotherapy, 2009. **20**(1): p. 1-18.
170. Wendeler, M., J.T. Miller, and S.F. Le Grice, *Human immunodeficiency virus reverse transcriptase*, in *Viral Genome Replication*2009, Springer. p. 403-427.
171. Delelis, O., et al., *Integrase and integration: biochemical activities of HIV-1 integrase.* Retrovirology, 2008. **5**(1): p. 114.
172. Martinez, J., P. Coplan, and M.A. Wainberg, *Is HIV drug resistance a limiting factor in the development of anti-HIV NNRTI and NRTI-based vaginal microbicide strategies?* Antiviral Res, 2006. **71**(2): p. 343-350.
173. Pommier, Y., A.A. Johnson, and C. Marchand, *Integrase inhibitors to treat HIV/AIDS.* Nature Reviews Drug Discovery, 2005. **4**(3): p. 236-248.
174. Goff, S.P., *Intracellular trafficking of retroviral genomes during the early phase of infection: viral exploitation of cellular pathways.* The journal of gene medicine, 2001. **3**(6): p. 517-528.
175. Zhou, H., et al., *Genome-scale RNAi screen for host factors required for HIV replication.* Cell host & microbe, 2008. **4**(5): p. 495-504.
176. Blissenbach, M., et al., *Nuclear RNA export and packaging functions of HIV-1 Rev revisited.* Journal of virology, 2010. **84**(13): p. 6598-6604.
177. Chamond, N., N. Locker, and B. Sargueil, *The different pathways of HIV genomic RNA translation.* Biochemical Society Transactions, 2010. **38**: p. 1548-1552.
178. Sundquist, W.I. and H.-G. Kräusslich, *HIV-1 assembly, budding, and maturation.* Cold Spring Harbor perspectives in medicine, 2012. **2**(7): p. a006924.
179. Bukrinskaya, A., *HIV-1 assembly and maturation.* Archives of virology, 2004. **149**(6): p. 1067-1082.
180. Murakami, T., et al., *Regulation of human immunodeficiency virus type 1 Env-mediated membrane fusion by viral protease activity.* Journal of virology, 2004. **78**(2): p. 1026-1031.

181. Wyma, D.J., et al., *Coupling of human immunodeficiency virus type 1 fusion to virion maturation: a novel role of the gp41 cytoplasmic tail.* Journal of virology, 2004. **78**(7): p. 3429-3435.
182. Salzwedel, K., D.E. Martin, and M. Sakalian, *Maturation inhibitors: a new therapeutic class targets the virus structure.* Aids Rev, 2007. **9**(3): p. 162-172.
183. Ganser-Pornillos, B.K., M. Yeager, and O. Pornillos, *Assembly and architecture of HIV*, in *Viral molecular machines*2012, Springer. p. 441-465.
184. De Oliveira, T., et al., *Variability at human immunodeficiency virus type 1 subtype C protease cleavage sites: an indication of viral fitness?* Journal of virology, 2003. **77**(17): p. 9422-9430.
185. online, M.b.
186. Freed, E.O., *The HIV–TSG101 interface: recent advances in a budding field.* Trends in microbiology, 2003. **11**(2): p. 56-59.
187. Raiborg, C. and H. Stenmark, *The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins.* Nature, 2009. **458**(7237): p. 445-452.
188. Freed, E.O., *HIV-1 assembly, release and maturation.* Nature Reviews Microbiology, 2015. **13**(8): p. 484-496.
189. Coleman, C.M. and L. Wu, *HIV interactions with monocytes and dendritic cells: viral latency and reservoirs.* Retrovirology, 2009. **6**(1): p. 51.
190. Langford, S.E., J. Ananworanich, and D.A. Cooper, *Predictors of disease progression in HIV infection: a review.* AIDS research and therapy, 2007. **4**(1): p. 11.
191. Lama, J. and V. Planelles, *Host factors influencing susceptibility to HIV infection and AIDS progression.* Retrovirology, 2007. **4**(1): p. 52.
192. Fauci, A., et al., *Introduction to the Immune System.* Harrisons Principles of Internal Medicine, 2005: p. 17.
193. Poropatich, K. and D.J. Sullivan, *Human immunodeficiency virus type 1 long-term non-progressors: the viral, genetic and immunological basis for disease non-progression.* Journal of General Virology, 2011. **92**(2): p. 247-268.
194. Olson, A.D., et al., *Evaluation of Rapid Progressors in HIV Infection as an Extreme Phenotype.* Journal of acquired immune deficiency syndromes (1999), 2014. **67**(1): p. 15.
195. Grabar, S., et al., *Prevalence and comparative characteristics of long-term nonprogressors and HIV controller patients in the French Hospital Database on HIV.* Aids, 2009. **23**(9): p. 1163-1169.
196. Okulicz, J.F. and O. Lambotte, *Epidemiology and clinical characteristics of elite controllers.* Current Opinion in HIV and AIDS, 2011. **6**(3): p. 163-168.
197. Okulicz, J.F., et al., *Clinical outcomes of elite controllers, viremic controllers, and long-term nonprogressors in the US Department of Defense HIV natural history study.* Journal of Infectious Diseases, 2009. **200**(11): p. 1714-1723.
198. Levy, J.A., *HIV and the Pathogenesis of AIDS*2007: ASM press Washington, DC, USA:.
199. Cohen, M.S., et al., *The detection of acute HIV infection.* Journal of Infectious Diseases, 2010. **202**(Supplement 2): p. S270-S277.
200. Wawer, M.J., et al., *Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda.* Journal of Infectious Diseases, 2005. **191**(9): p. 1403-1409.
201. McMichael, A.J., et al., *The immune response during acute HIV-1 infection: clues for vaccine development.* Nature Reviews Immunology, 2010. **10**(1): p. 11-23.
202. Deeks, S.G., S.R. Lewin, and D.V. Havlir, *The end of AIDS: HIV infection as a chronic disease.* The Lancet, 2013. **382**(9903): p. 1525-1533.
203. Pantaleo, G. and A. Fauci, *Immunopathogenesis of HIV infection 1.* Annual Reviews in Microbiology, 1996. **50**(1): p. 825-854.
204. DEFINITIONS, W.C., *HIV/AIDS Programme*, 2007.
205. Simon, V., D.D. Ho, and Q.A. Karim, *HIV/AIDS epidemiology, pathogenesis, prevention, and treatment.* The Lancet, 2006. **368**(9534): p. 489-504.

206. Ng, O.T., et al., *HIV Type 1 polymerase gene polymorphisms are associated with phenotypic differences in replication capacity and disease progression.* Journal of Infectious Diseases, 2014. **209**(1): p. 66-73.
207. Kaleebu, P., et al., *Relation between chemokine receptor use, disease stage, and HIV-1 subtypes A and D: results from a rural Ugandan cohort.* JAIDS Journal of Acquired Immune Deficiency Syndromes, 2007. **45**(1): p. 28-33.
208. DE WOLF, F., et al., *Syncytium-inducing and non-syncytium-inducing capacity of human immunodeficiency virus type 1 subtypes other than B: phenotypic and genotypic characteristics.* AIDS Res Hum Retroviruses, 1994. **10**(11): p. 1387-1400.
209. Tscherning, C., et al., *Differences in chemokine coreceptor usage between genetic subtypes of HIV-1.* Virology, 1998. **241**(2): p. 181-188.
210. Esbjornsson, J., et al., *Frequent CXCR4 tropism of HIV-1 subtype A and CRF02_AG during late-stage disease–indication of an evolving epidemic in West Africa.* Retrovirology, 2010. **7**(1): p. 23.
211. Morrison, C.S., et al., *Plasma and cervical viral loads among Ugandan and Zimbabwean women during acute and early HIV-1 infection.* AIDS (London, England), 2010. **24**(4): p. 573.
212. Campbell, M.S., et al., *Plasma viral loads during early HIV-1 infection are similar in subtype C–and non-subtype C–infected African seroconverters.* Journal of Infectious Diseases, 2013. **207**(7): p. 1166-1170.
213. Hu, D.J., et al., *Viral load differences in early infection with two HIV-1 subtypes.* Aids, 2001. **15**(6): p. 683-691.
214. Ferdinandy, B., et al., *HIV Competition Dynamics over Sexual Networks: First Comer Advantage Conserves Founder Effects.* PLoS Comput Biol, 2015. **11**(1): p. e1004093.
215. Sovran, S., *Understanding culture and HIV/AIDS in sub-Saharan Africa.* SAHARA-J: Journal of Social Aspects of HIV/AIDS, 2013. **10**(1): p. 32-41.
216. Brandenberg, O.F., et al., *Different Infectivity of HIV-1 Strains Is Linked to Number of Envelope Trimers Required for Entry.* PLoS Pathog, 2015. **11**(1): p. e1004595.
217. Smith, R.J., et al., *Evaluating the potential impact of vaginal microbicides to reduce the risk of acquiring HIV in female sex workers.* Aids, 2005. **19**(4): p. 413-421.
218. Renjifo, B., et al., *Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D.* Aids, 2004. **18**(12): p. 1629-1636.
219. Foy, H., et al., *HIV-1 subtypes and male-to-female transmission in Thailand.* The Lancet, 1995. **345**(8957): p. 1078-1083.
220. Renjifo, B., et al., *Differences in perinatal transmission among human immunodeficiency virus type 1 genotypes.* Journal of human virology, 2000. **4**(1): p. 16-25.
221. Soto-Ramirez, L.E., et al., *HIV-1 Langerhans' cell tropism associated with heterosexual transmission of HIV.* Science, 1996. **271**(5253): p. 1291-1293.
222. John-Stewart, G.C., et al., *Subtype C is associated with increased vaginal shedding of HIV-1.* Journal of Infectious Diseases, 2005. **192**(3): p. 492-496.
223. Hollingsworth, T.D., R.M. Anderson, and C. Fraser, *HIV-1 transmission, by stage of infection.* Journal of Infectious Diseases, 2008. **198**(5): p. 687-693.
224. Chalmet, K., et al., *Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections.* BMC infectious diseases, 2010. **10**(1): p. 262.
225. Van Harmelen, J., et al., *An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa.* Aids, 1997. **11**(1): p. 81-87.
226. Sutthent, R., et al., *Shedding of HIV-1 subtype E in semen and cervico-vaginal fluid.* Journal of the Medical Association of Thailand= Chotmaihet thangphaet, 1997. **80**(6): p. 348-357.
227. Hu, D.J., et al., *What role does HIV-1 subtype play in transmission and pathogenesis? An epidemiological perspective.* Aids, 1999. **13**(8): p. 873-881.

228. Yoshino, N., et al., *Vertical transmission of human immunodeficiency virus type 1 in Japan, 1989–1997: Presence of two subtypes B and E with subtype E predominance.* Pediatrics International, 1998. **40**(5): p. 503-509.
229. Conroy, S.A., et al., *Changes in the distribution of HIV type 1 subtypes D and A in Rakai District, Uganda between 1994 and 2002.* AIDS Res Hum Retroviruses, 2010. **26**(10): p. 1087-1091.
230. Ball, S.C., et al., *Comparing the ex vivo fitness of CCR5-tropic human immunodeficiency virus type 1 isolates of subtypes B and C.* Journal of virology, 2003. **77**(2): p. 1021-1038.
231. Abraha, A., et al., *CCR5-and CXCR4-tropic subtype C human immunodeficiency virus type 1 isolates have a lower level of pathogenic fitness than other dominant group M subtypes: implications for the epidemic.* Journal of virology, 2009. **83**(11): p. 5592-5605.
232. Kahle, E., et al., *HIV-1 subtype C is not associated with higher risk of heterosexual HIV-1 transmission: a multinational study among HIV-1 serodiscordant couples.* AIDS (London, England), 2014. **28**(2): p. 235-243.
233. Rodriguez, M.A., et al., *High replication fitness and transmission efficiency of HIV-1 subtype C from India: Implications for subtype C predominance.* Virology, 2009. **385**(2): p. 416-424.
234. Blackard, J.T., et al., *HIV-1 LTR subtype and perinatal transmission.* Virology, 2001. **287**(2): p. 261-265.
235. Hussain, L. and T. Lehner, *Comparative investigation of Langerhans' cells and potential receptors for HIV in oral, genitourinary and rectal epithelia.* Immunology, 1995. **85**(3): p. 475.
236. McCoombe, S.G. and R.V. Short, *Potential HIV-1 target cells in the human penis.* Aids, 2006. **20**(11): p. 1491-1495.
237. Kawamura, T., et al., *The role of Langerhans cells in the sexual transmission of HIV.* Journal of dermatological science, 2005. **40**(3): p. 147-155.
238. Moore, J.P., et al., *The CCR5 and CXCR4 coreceptors-central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection.* AIDS Res Hum Retroviruses, 2004. **20**(1): p. 111-126.
239. Kiwanuka, N., et al., *HIV-1 viral subtype differences in the rate of CD4+ T-cell decline among HIV seroincident antiretroviral naive persons in Rakai district, Uganda.* Journal of acquired immune deficiency syndromes (1999), 2010. **54**(2): p. 180.
240. Kiwanuka, N., 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.* Journal of Infectious Diseases, 2008. **197**(5): p. 707-13.
241. Kaleebu, P., et al., *Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1—positive persons in Uganda.* Journal of Infectious Diseases, 2002. **185**(9): p. 1244-1250.
242. Vasan, A., et al., *Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype.* Clin Infect Dis, 2006. **42**(6): p. 843-52.
243. Laurent, C., et al., *No difference in clinical progression between patients infected with the predominant human immunodeficiency virus type 1 circulating recombinant form (CRF) 02_AG strain and patients not infected with CRF02_AG, in Western and West-Central Africa: a four-year prospective multicenter study.* Journal of Infectious Diseases, 2002. **186**(4): p. 486-492.
244. Senkaali, D., et al., *The relationship between HIV type 1 disease progression and V3 serotype in a rural Ugandan cohort.* AIDS Research & Human Retroviruses, 2004. **20**(9): p. 932-937.
245. McKinnon, L.R., et al., *HIV-1 Clade D Is Associated with Increased Rates of CD4 Decline in a Kenyan Cohort.* 2012.
246. Easterbrook, P.J., et al., *Impact of HIV-1 viral subtype on disease progression and response to antiretroviral therapy.* Journal of the International AIDS Society, 2010. **13**(1): p. 4.

247. Kaleebu, P., et al., *Relationship between HIV-1 Env subtypes A and D and disease progression in a rural Ugandan cohort.* Aids, 2001. **15**(3): p. 293-299.
248. Amornkul, P.N., et al., *Disease progression by infecting HIV-1 subtype in a seroconverter cohort in sub-Saharan Africa.* AIDS (London, England), 2013. **27**(17): p. 2775.
249. Kuritzkes, D.R., *HIV-1 subtype as a determinant of disease progression.* Journal of Infectious Diseases, 2008. **197**(5): p. 638-639.
250. Palm, A.A., et al., *Faster progression to AIDS and AIDS-related death among seroincident individuals infected with recombinant HIV-1 A3/CRF02_AG compared to sub subtype A3.* Journal of Infectious Diseases, 2013: p. jit416.
251. Gray, C.M., et al., *Short communication: viral dynamics and CD4+ T cell counts in subtype C human immunodeficiency virus type 1-infected individuals from southern Africa.* AIDS Research & Human Retroviruses, 2005. **21**(4): p. 285-291.
252. Kanki, P.J., et al., *Human immunodeficiency virus type 1 subtypes differ in disease progression.* Journal of Infectious Diseases, 1999. **179**(1): p. 68-73.
253. Lemonovich, T.L., et al., *Differences in Clinical Manifestations of Acute and Early HIV-1 Infection between HIV-1 Subtypes in African Women.* Journal of the International Association of Providers of AIDS Care (JIAPAC), 2013: p. 2325957413504827.
254. Lutalo, T., et al., *Survival of HIV-infected treatment-naive individuals with documented dates of seroconversion in Rakai, Uganda.* Aids, 2007. **21**: p. S15-S19.
255. Gaschen, B., et al., *Diversity considerations in HIV-1 vaccine selection.* Science, 2002. **296**(5577): p. 2354-2360.
256. Rambaut, A., et al., *The causes and consequences of HIV evolution.* Nature Reviews Genetics, 2004. **5**(1): p. 52-61.
257. Robertson, D., et al., *HIV-1 nomenclature proposal.* Science, 2000. **288**(5463): p. 55-55.
258. Seaman, M.S., et al., *Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies.* Journal of virology, 2010. **84**(3): p. 1439-1452.
259. Jeeninga, R.E., et al., *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): p. 3740-3751.
260. Roof, P., et al., *Differential regulation of HIV-1 clade-specific B, C, and E long terminal repeats by NF-κB and the Tat transactivator.* Virology, 2002. **296**(1): p. 77-83.
261. Baar, M.P.D., et al., *Subtype-specific sequence variation of the HIV type 1 long terminal repeat and primer-binding site.* AIDS Res Hum Retroviruses, 2000. **16**(5): p. 499-504.
262. van Opijnen, T., 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.* Journal of virology, 2004. **78**(7): p. 3675-3683.
263. Marozsan, A.J., et al., *Differences in the fitness of two diverse wild-type human immunodeficiency virus type 1 isolates are related to the efficiency of cell binding and entry.* Journal of virology, 2005. **79**(11): p. 7121-7134.
264. Iordanskiy, S., et al., *Subtype-associated differences in HIV-1 reverse transcription affect the viral replication.* Retrovirology, 2010. **7**(1): p. 85.
265. Velazquez-Campoy, A., et al., *Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes.* Proceedings of the National Academy of Sciences, 2001. **98**(11): p. 6062-6067.
266. Iwabu, Y., et al., *Differential anti-APOBEC3G activity of HIV-1 Vif proteins derived from different subtypes.* Journal of Biological Chemistry, 2010. **285**(46): p. 35350-35358.
267. Mann, J.K., et al., *Ability of HIV-1 Nef to downregulate CD4 and HLA class I differs among viral subtypes.* Retrovirology, 2013. **10**(1): p. 100.
268. Huang, W., et al., *Coreceptor tropism in human immunodeficiency virus type 1 subtype D: high prevalence of CXCR4 tropism and heterogeneous composition of viral populations.* Journal of virology, 2007. **81**(15): p. 7885-7893.
269. Church, J.D., et al., *HIV-1 tropism and survival in vertically infected Ugandan infants.* Journal of Infectious Diseases, 2008. **197**(10): p. 1382-1388.

270. Sucupira, M., et al., *Faster HIV-1 disease progression among Brazilian individuals recently infected with CXCR4-utilizing strains.* PLoS One, 2012. **7**(1): p. e30292.
271. Hansen, T.H. and M. Bouvier, *MHC class I antigen presentation: learning from viral evasion strategies.* Nature Reviews Immunology, 2009. **9**(7): p. 503-513.
272. Gulzar, N. and K.F. Copeland, *CD8+ T-cells: function and response to HIV infection.* Curr HIV Res, 2004. **2**(1): p. 23-37.
273. Goulder, P.J. and D.I. Watkins, *HIV and SIV CTL escape: implications for vaccine design.* Nature Reviews Immunology, 2004. **4**(8): p. 630-640.
274. Kiepiela, P., et al., *CD8+ T-cell responses to different HIV proteins have discordant associations with viral load.* Nat Med, 2007. **13**(1): p. 46-53.
275. Borrow, P., et al., *Antiviral pressure exerted by HIV-l-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus.* Nature medicine, 1997. **3**(2): p. 205-211.
276. Goulder, P.J., et al., *Evolution and transmission of stable CTL escape mutations in HIV infection.* Nature, 2001. **412**(6844): p. 334-338.
277. Draenert, R., et al., *Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection.* J Exp Med, 2004. **199**(7): p. 905-15.
278. Iversen, A.K., et al., *Conflicting selective forces affect T cell receptor contacts in an immunodominant human immunodeficiency virus epitope.* Nature immunology, 2006. **7**(2): p. 179-189.
279. Yokomaku, Y., et al., *Impaired processing and presentation of cytotoxic-T-lymphocyte (CTL) epitopes are major escape mechanisms from CTL immune pressure in human immunodeficiency virus type 1 infection.* Journal of virology, 2004. **78**(3): p. 1324-1332.
280. Tenzer, S., et al., *Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance.* Nature immunology, 2009. **10**(6): p. 636-646.
281. Leslie, A., et al., *HIV evolution: CTL escape mutation and reversion after transmission.* Nature medicine, 2004. **10**(3): p. 282-289.
282. Leslie, A., et al., *Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA.* J Exp Med, 2005. **201**(6): p. 891-902.
283. Zhang, S.C., et al., *Aminopeptidase substrate preference affects HIV epitope presentation and predicts immune escape patterns in HIV-infected individuals.* The Journal of Immunology, 2012. **188**(12): p. 5924-5934.
284. Yang, O.O., et al., *Determinants of HIV-1 mutational escape from cytotoxic T lymphocytes.* J Exp Med, 2003. **197**(10): p. 1365-1375.
285. Le Gall, S., P. Stamegna, and B.D. Walker, *Portable flanking sequences modulate CTL epitope processing.* J Clin Invest, 2007. **117**(117 (11)): p. 3563-3575.
286. Bailey, J.R., et al., *Evolution of HIV-1 in an HLA-B* 57-positive patient during virologic escape.* Journal of Infectious Diseases, 2007. **196**(1): p. 50-55.
287. Feeney, M., et al., *Immune escape precedes breakthrough human immunodeficiency virus type 1 viremia and broadening of the cytotoxic T-lymphocyte response in an HLA-B27-positive long-term-nonprogressing child.* Journal of virology, 2004. **78**(16): p. 8927-8930.
288. Geels, M.J., et al., *Identification of sequential viral escape mutants associated with altered T-cell responses in a human immunodeficiency virus type 1-infected individual.* Journal of virology, 2003. **77**(23): p. 12430-12440.
289. Edwards, B.H., et al., *Magnitude of functional CD8+ T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma.* Journal of virology, 2002. **76**(5): p. 2298-2305.
290. Zuñiga, R., et al., *Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control.* Journal of virology, 2006. **80**(6): p. 3122-3125.

291. Geldmacher, C., et al., *CD8 T-cell recognition of multiple epitopes within specific Gag regions is associated with maintenance of a low steady-state viremia in human immunodeficiency virus type 1-seropositive patients.* Journal of virology, 2007. **81**(5): p. 2440-2448.
292. Sáez-Cirión, A., et al., *Heterogeneity in HIV suppression by CD8 T cells from HIV controllers: association with Gag-specific CD8 T cell responses.* The Journal of Immunology, 2009. **182**(12): p. 7828-7837.
293. Schellens, I., et al., *Abundance of early functional HIV-specific CD8+ T cells does not predict AIDS-free survival time.* PLoS One, 2008. **3**(7): p. e2745.
294. Kawashima, Y., et al., *Adaptation of HIV-1 to human leukocyte antigen class I.* Nature, 2009. **458**(7238): p. 641-645.
295. Goulder, P.J. and B.D. Walker, *HIV and HLA class I: an evolving relationship.* Immunity, 2012. **37**(3): p. 426-440.
296. Study, T.I.H.C., *The major genetic determinants of HIV-1 control affect HLA class I peptide presentation.* Science (New York, NY), 2010. **330**(6010): p. 1551.
297. Apps, R., et al., *Influence of HLA-C expression level on HIV control.* Science, 2013. **340**(6128): p. 87-91.
298. Mothe, B., et al., *Virological, immune and host genetics markers in the control of HIV infection.* Disease markers, 2009. **27**(3-4): p. 105-120.
299. Gudmundsdotter, L., et al., *Cross-clade immune responses to Gag p24 in patients infected with different HIV-1 subtypes and correlation with HLA class I and II alleles.* Vaccine, 2008. **26**(40): p. 5182-5187.
300. MacDonald, K.S., et al., *Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection.* Journal of Infectious Diseases, 2000. **181**(5): p. 1581-1589.
301. MacDonald, K.S., et al., *The HLA A2/6802 supertype is associated with reduced risk of perinatal human immunodeficiency virus type 1 transmission.* Journal of Infectious Diseases, 2001. **183**(3): p. 503-506.
302. Borghans, J., et al., *HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24.* PLoS One, 2007. **2**(9): p. e920.
303. Altfeld, M., et al., *HLA alleles associated with delayed progression to AIDS contribute strongly to the initial CD8+ T cell response against HIV-1.* 2006.
304. Naruto, T., et al., *HLA class I-mediated control of HIV-1 in the Japanese population, in which the protective HLA-B* 57 and HLA-B* 27 alleles are absent.* Journal of virology, 2012. **86**(19): p. 10870-10872.
305. Chakrabarti, L.A. and V. Simon, *Immune mechanisms of HIV control.* Current opinion in immunology, 2010. **22**(4): p. 488-496.
306. Gesprasert, G., et al., *HLA-associated immune pressure on Gag protein in CRF01_AE-infected individuals and its association with plasma viral load.* PLoS One, 2010. **5**(6): p. e11179.
307. Kelleher, A.D., et al., *Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27–restricted cytotoxic T lymphocyte responses.* J Exp Med, 2001. **193**(3): p. 375-386.
308. Migueles, S.A., et al., *The differential ability of HLA B* 5701+ long-term nonprogressors and progressors to restrict human immunodeficiency virus replication is not caused by loss of recognition of autologous viral gag sequences.* Journal of virology, 2003. **77**(12): p. 6889-6898.
309. Kiepiela, P., et al., *Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA.* Nature, 2004. **432**(7018): p. 769-75.
310. Dai, B., et al., *HIV-1 Gag-specific immunity induced by a lentivector-based vaccine directed to dendritic cells.* Proceedings of the National Academy of Sciences, 2009. **106**(48): p. 20382-20387.

311. Kiepiela, P., et al., *Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA.* Nature, 2004. **432**(7018): p. 769-775.
312. Carrington, M. and S.J. O'Brien, *The Influence of HLA Genotype on AIDS*.* Annual review of medicine, 2003. **54**(1): p. 535-551.
313. De Luca, A., *The impact of resistance on viral fitness and its clinical implications.* 2006.
314. Dykes, C. and L.M. Demeter, *Clinical significance of human immunodeficiency virus type 1 replication fitness.* Clinical microbiology reviews, 2007. **20**(4): p. 550-578.
315. Wrin, T., *Natural variation of replication capacity measurements in drug-naive/susceptible HIV-1.* Antiviral therapy, 2001. **6**: p. 24-24.
316. Wright, J.K., et al., *Influence of Gag-protease-mediated replication capacity on disease progression in individuals recently infected with HIV-1 subtype C.* J Virol, 2011. **85**(8): p. 3996-4006.
317. Miura, T., 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): p. 140-9.
318. Boutwell, C.L., C.F. Rowley, and M. Essex, *Reduced viral replication capacity of human immunodeficiency virus type 1 subtype C caused by cytotoxic-T-lymphocyte escape mutations in HLA-B57 epitopes of capsid protein.* J Virol, 2009. **83**(6): p. 2460-8.
319. Brockman, M.A., 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.* Journal of virology, 2010. **84**(22): p. 11937-11949.
320. Wright, J.K., et al., *Gag-protease-mediated replication capacity in HIV-1 subtype C chronic infection: associations with HLA type and clinical parameters.* Journal of virology, 2010. **84**(20): p. 10820-10831.
321. Crawford, H., et al., *Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B* 5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection.* Journal of virology, 2007. **81**(15): p. 8346-8351.
322. Brockman, M.A., et al., *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): p. 134-42.
323. Fowke, K.R., et al., *Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya.* Lancet, 1996. **348**(9038): p. 1347-51.
324. Luo, M., et al., *HLA-A and HLA-B in Kenya, Africa: allele frequencies and identification of HLA-B*1567 and HLA-B*4426.* Tissue Antigens, 2002. **59**(5): p. 370-80.
325. Wright, J.K., 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): p. 10820-31.
326. Guindon, S., et al., *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.* Systematic biology, 2010. **59**(3): p. 307-321.
327. Primrose, S.B. and R. Twyman, *Principles of gene manipulation and genomics*2013: John Wiley & Sons.
328. Tsong, T.Y., *Electroporation of cell membranes.* Biophysical journal, 1991. **60**(2): p. 297.
329. Lurquin, P.F., *Gene transfer by electroporation.* Molecular biotechnology, 1997. **7**(1): p. 5-35.
330. Goddard, G., et al., *Single particle high resolution spectral analysis flow cytometry.* Cytometry Part A, 2006. **69**(8): p. 842-851.
331. Carlson, J.M., et al., *Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag.* PLoS Comput Biol, 2008. **4**(11): p. e1000225.

332. Wuensch, K.L., *Chi-Square Tests*, in *International Encyclopedia of Statistical Science*2011, Springer. p. 252-253.
333. Ng, O.T., et al., *Increased rate of CD4+ T-cell decline and faster time to antiretroviral therapy in HIV-1 subtype CRF01_AE infected seroconverters in Singapore.* 2011.
334. Nomura, S., 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): p. 1465-76.
335. Huang, K.-H., et al., *Progression to AIDS in South Africa is associated with both reverting and compensatory viral mutations.* PLoS One, 2011. **6**(4): p. e19018.
336. Han, Y., et al., *The role of protective HCP5 and HLA-C associated polymorphisms in the control of HIV-1 replication in a subset of elite suppressors.* Aids, 2008. **22**(4): p. 541-544.
337. Goulder, P.J. and D.I. Watkins, *Impact of MHC class I diversity on immune control of immunodeficiency virus replication.* Nature Reviews Immunology, 2008. **8**(8): p. 619-630.
338. Leslie, A.J., et al., *HIV evolution: CTL escape mutation and reversion after transmission.* Nat Med, 2004. **10**(3): p. 282-9.
339. Brockman, M.A., 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): p. 11937-49.
340. Huang, X., Y. Fang, and J. Wang, *Identification of functional rare variants in genome-wide association studies using stability selection based on random collapsing.* BMC Proc, 2011. **5 Suppl 9**: p. S56.
341. Blankson, J.N., et al., *Isolation and characterization of replication-competent human immunodeficiency virus type 1 from a subset of elite suppressors.* J Virol, 2007. **81**(5): p. 2508-18.
342. Bailey, J.R., et al., *Maintenance of viral suppression in HIV-1-infected HLA-B*57+ elite suppressors despite CTL escape mutations.* J Exp Med, 2006. **203**(5): p. 1357-69.
343. Blankson, J.N., *Viral reservoirs and HIV-specific immunity.* Curr Opin HIV AIDS, 2006. **1**(2): p. 147-51.
344. McKinnon, L.R., et al., *Clade-specific evolution mediated by HLA-B* 57/5801 in human immunodeficiency virus type 1 clade A1 p24.* Journal of virology, 2009. **83**(23): p. 12636-12642.
345. Lazaryan, A., et al., *Human leukocyte antigen B58 supertype and human immunodeficiency virus type 1 infection in native Africans.* Journal of virology, 2006. **80**(12): p. 6056-6060.
346. Martinez-Picado, J., et al., *Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1.* Journal of virology, 2006. **80**(7): p. 3617-3623.
347. Brockman, M.A., 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): p. 12608-12618.
348. Wright, J.K., et al., *Impact of HLA-B* 81-associated mutations in HIV-1 Gag on viral replication capacity.* Journal of virology, 2012. **86**(6): p. 3193-3199.
349. Carlson, J.M., et al., *Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag.* PLoS Comput Biol, 2008. **4**(11): p. e1000225-e1000225.
350. Lee, G.Q., et al., *Subtype-Specific HIV-1 Adaptation to Host HLA.* AIDS Res Hum Retroviruses, 2014. **30**(S1): p. A218-A218.
351. Wright, J.K., et al., *Impact of HLA-B*81-associated mutations in HIV-1 Gag on viral replication capacity.* J Virol, 2012. **86**(6): p. 3193-9.
352. Maguire, M.F., et al., *Changes in human immunodeficiency virus type 1 Gag at positions L449 and P453 are linked to I50V protease mutants in vivo and cause reduction of*

sensitivity to amprenavir and improved viral fitness in vitro. Journal of virology, 2002. **76**(15): p. 7398-7406.
353. Zennou, V., et al., *Loss of viral fitness associated with multiple Gag and Gag-Pol processing defects in human immunodeficiency virus type 1 variants selected for resistance to protease inhibitors in vivo.* Journal of virology, 1998. **72**(4): p. 3300-3306.
354. Hemelaar, J., *Implications of HIV diversity for the HIV-1 pandemic.* Journal of Infection, 2013. **66**(5): p. 391-400.
355. Castro-Nallar, E., et al., *The evolution of HIV: inferences using phylogenetics.* Molecular Phylogenetics and Evolution, 2012. **62**(2): p. 777-792.
356. Freed, E.O., *HIV-1 replication.* Somatic cell and molecular genetics, 2001. **26**(1-6): p. 13-33.
357. Gottlinger, H., *HIV-1 Gag: a molecular machine driving viral particle assembly and release.* Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM, LA-UR, 2001: p. 02-2877.
358. Wu, Z., et al., *Total chemical synthesis of N-myristoylated HIV-1 matrix protein p17: structural and mechanistic implications of p17 myristoylation.* Proc Natl Acad Sci U S A, 2004. **101**(32): p. 11587-11592.
359. von Schwedler, U.K., et al., *The protein network of HIV budding.* Cell, 2003. **114**(6): p. 701-713.
360. Alce, T.M. and W. Popik, *APOBEC3G is incorporated into virus-like particles by a direct interaction with HIV-1 Gag nucleocapsid protein.* Journal of Biological Chemistry, 2004. **279**(33): p. 34083-34086.
361. Locker, N., N. Chamond, and B. Sargueil, *A conserved structure within the HIV gag open reading frame that controls translation initiation directly recruits the 40S subunit and eIF3.* Nucleic Acids Res, 2011. **39**(6): p. 2367-2377.
362. Schieffer, M., et al., *Induction of Gag-specific CD4 T cell responses during acute HIV infection is associated with improved viral control.* Journal of virology, 2014. **88**(13): p. 7357-7366.
363. Prince, J.L., et al., *Role of transmitted Gag CTL polymorphisms in defining replicative capacity and early HIV-1 pathogenesis.* 2012.
364. Alexander, L., et al., *Unusual polymorphisms in human immunodeficiency virus type 1 associated with nonprogressive infection.* Journal of virology, 2000. **74**(9): p. 4361-4376.
365. Jia, M., et al., *Preferential CTL targeting of Gag is associated with relative viral control in long-term surviving HIV-1 infected former plasma donors from China.* Cell research, 2012. **22**(5): p. 903-914.
366. Miura, T., et al., *HLA-associated viral mutations are common in human immunodeficiency virus type 1 elite controllers.* Journal of virology, 2009. **83**(7): p. 3407-3412.
367. Moses, S., et al., *Controlling HIV in Africa: effectiveness and cost of an intervention in a high-frequency STD transmitter core group.* Aids, 1991. **5**(4): p. 407-412.
368. Fowke, K.R., et al., *Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya.* The Lancet, 1996. **348**(9038): p. 1347-1351.
369. Chang, C.-H., et al., *HIV-infected sex workers with beneficial HLA-variants are potential hubs for selection of HIV-1 recombinants that may affect disease progression.* Sci Rep, 2015. **5**.
370. Gray, R.R., et al., *Spatial phylodynamics of HIV-1 epidemic emergence in east Africa.* AIDS (London, England), 2009. **23**(14): p. F9.
371. Rainwater, S., et al., *No evidence for rapid subtype C spread within an epidemic in which multiple subtypes and intersubtype recombinants circulate.* AIDS Research & Human Retroviruses, 2005. **21**(12): p. 1060-1065.
372. Powell, R.L., et al., *Longitudinal quasispecies analysis of viral variants in HIV type 1 dually infected individuals highlights the importance of sequence identity in viral recombination.* AIDS Res Hum Retroviruses, 2010. **26**(3): p. 253-264.

373. Muñoz-Nieto, M., et al., *HIV type 1 intersubtype recombinants during the evolution of a dual infection with subtypes B and G.* AIDS Res Hum Retroviruses, 2008. **24**(2): p. 337-343.
374. Ramos, A., et al., *Dual and recombinant infections: an integral part of the HIV-1 epidemic in Brazil.* Emerging infectious diseases, 1999. **5**(1): p. 65.
375. Jetzt, A.E., et al., *High rate of recombination throughout the human immunodeficiency virus type 1 genome.* Journal of virology, 2000. **74**(3): p. 1234-1240.
376. Peeters, M., *Recombinant HIV sequences: their role in the global epidemic.* HIV sequence compendium, 2000. **2000**: p. 1-39.
377. Quinones-Mateu, M.E., et al., *In vitro intersubtype recombinants of human immunodeficiency virus type 1: comparison to recent and circulating in vivo recombinant forms.* Journal of virology, 2002. **76**(19): p. 9600-9613.
378. Tatem, A.J., et al., *Spatial accessibility and the spread of HIV-1 subtypes and recombinants.* Aids, 2012. **26**(18): p. 2351-2360.
379. Neogi, U., et al., *Molecular epidemiology of HIV-1 subtypes in India: origin and evolutionary history of the predominant subtype C.* PLoS One, 2012. **7**(6): p. e39819.
380. Simon-Loriere, E., et al., *Molecular mechanisms of recombination restriction in the envelope gene of the human immunodeficiency virus.* PLoS Pathog, 2009. **5**(5): p. e1000418.
381. Galli, A., et al., *Patterns of human immunodeficiency virus type 1 recombination ex vivo provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication.* Journal of virology, 2010. **84**(15): p. 7651-7661.
382. Baird, H.A., et al., *Influence of sequence identity and unique breakpoints on the frequency of intersubtype HIV-1 recombination.* Retrovirology, 2006. **3**(1): p. 91.
383. Archer, J., et al., *Identifying the important HIV-1 recombination breakpoints.* PLoS Comput Biol, 2008. **4**(9): p. e1000178-e1000178.
384. Casado, G., et al., *Identification of a novel HIV-1 circulating ADG intersubtype recombinant form (CRF19_cpx) in Cuba.* JAIDS Journal of Acquired Immune Deficiency Syndromes, 2005. **40**(5): p. 532-537.
385. Carr, J.K., et al., *Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants.* Virology, 1998. **247**(1): p. 22-31.
386. Bagaya, B.S., et al., *Functional bottlenecks for generation of HIV-1 intersubtype Env recombinants.* Retrovirology, 2015. **12**(1): p. 44.
387. Quiñones-Mateu, M.E. and E.J. Arts, *Recombination in HIV-1: update and implications.* Aids Rev, 1999. **1**(2): p. 89-100.
388. Levy, D.N., et al., *Dynamics of HIV-1 recombination in its natural target cells.* Proceedings of the National Academy of Sciences, 2004. **101**(12): p. 4204-4209.
389. Tarosso, L.F., et al., *Short Communication: HIV Type 1 Subtype BF Leads to Faster CD4+ T Cell Loss Compared to Subtype B.* AIDS Res Hum Retroviruses, 2014. **30**(2): p. 190-194.
390. Vijay, N., et al., *Recombination increases human immunodeficiency virus fitness, but not necessarily diversity.* Journal of General Virology, 2008. **89**(6): p. 1467-1477.
391. Kouri, V., et al., *CRF19_cpx is an evolutionary fit HIV-1 variant strongly associated with rapid progression to AIDS in Cuba.* EBioMedicine, 2015. **2**(3): p. 244-254.
392. Campbell, T.B., et al., *Relationship between in vitro human immunodeficiency virus type 1 replication rate and virus load in plasma.* Journal of virology, 2003. **77**(22): p. 12105-12112.
393. Gijsbers, E.F., et al., *HIV-1 Replication Fitness of HLA-B* 57/58: 01 CTL Escape Variants Is Restored by the Accumulation of Compensatory Mutations in Gag.* 2013.
394. Ariën, K.K., G. Vanham, and E.J. Arts, *Is HIV-1 evolving to a less virulent form in humans?* Nature Reviews Microbiology, 2007. **5**(2): p. 141-151.

395. Travers, S.A., et al., *Timing and reconstruction of the most recent common ancestor of the subtype C clade of human immunodeficiency virus type 1.* Journal of virology, 2004. **78**(19): p. 10501-10506.
396. Osmanov, S., et al., *Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000.* Journal of acquired immune deficiency syndromes (1999), 2002. **29**(2): p. 184-190.
397. Gräf, T. and A.R. Pinto, *The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent.* Virology, 2013. **435**(1): p. 170-178.
398. Chibo, D., M. Kaye, and C. Birch, *HIV transmissions during seroconversion contribute significantly to new infections in men who have sex with men in Australia.* AIDS Res Hum Retroviruses, 2012. **28**(5): p. 460-464.
399. Renjifo, B. and M. Essex, *HIV-1 subtypes and recombinants*, in *AIDS in Africa*2002, Springer. p. 263-281.
400. Ndung'u, T., et al., *HIV-1 subtype C in vitro growth and coreceptor utilization.* Virology, 2006. **347**(2): p. 247-260.
401. Liu, D., et al., *Preexisting compensatory amino acids compromise fitness costs of a HIV-1 T cell escape mutation.* Retrovirology, 2014. **11**(1): p. 101-101.
402. Serwanga, J., et al., *Frequencies of Gag-restricted T-cell escape "footprints" differ across HIV-1 clades A1 and D chronically infected Ugandans irrespective of host HLA B alleles.* Vaccine, 2015. **33**(14): p. 1664-1672.
403. Sakai, K., 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**(1): p. 98.
404. Hendel, H., et al., *New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS.* The Journal of Immunology, 1999. **162**(11): p. 6942-6946.
405. Bailey, J.R., et al., *Maintenance of viral suppression in HIV-1–infected HLA-B* 57+ elite suppressors despite CTL escape mutations.* J Exp Med, 2006. **203**(5): p. 1357-1369.
406. McLaren, P.J., et al., *Fine-mapping classical HLA variation associated with durable host control of HIV-1 infection in African Americans.* Hum Mol Genet, 2012. **21**(19): p. 4334-4347.
407. Peterson, T.A., et al., *Diversity and frequencies of HLA class I and class II genes of an East African population.* Open Journal of Genetics, 2014. **4**(02): p. 99.
408. Koehler, R.N., et al., *Class I HLA-A* 7401 is associated with protection from HIV-1 acquisition and disease progression in Mbeya, Tanzania.* Journal of Infectious Diseases, 2010. **202**(10): p. 1562-1566.
409. Saina, M.C., 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): p. e0137140.
410. Matthews, P.C., et al., *HLA-A* 7401–mediated control of HIV viremia is independent of its linkage disequilibrium with HLA-B* 5703.* The Journal of Immunology, 2011. **186**(10): p. 5675-5686.
411. Tang, J., et al., *CD4: CD8 lymphocyte ratio as a quantitative measure of immunologic health in HIV-1 infection: findings from an African cohort with prospective data.* Front Microbiol, 2015. **6**.
412. Kløverpris, H.N., et al., *HLA-A∗ 68: 02-restricted Gag-specific cytotoxic T lymphocyte responses can drive selection pressure on HIV but are subdominant and ineffective.* Aids, 2013. **27**(11): p. 1717-1723.
413. Zhang, W., et al., *Frequency of HLA-A* 03 associates with HIV-1 infection in a Chinese cohort.* Science China Life Sciences, 2013. **56**(11): p. 1014-1019.
414. Vandiedonck, C. and J.C. Knight, *The human Major Histocompatibility Complex as a paradigm in genomics research.* Briefings in Functional Genomics, 2009: p. elp010.

415. Maguire, M.F., et al., *Changes in human immunodeficiency virus type 1 Gag at positions L449 and P453 are linked to I50V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro.* J Virol, 2002. **76**(15): p. 7398-406.
416. Sanford, B., et al., *Deletions in the fifth alpha helix of HIV-1 matrix block virus release.* Virology, 2014. **468**: p. 293-302.
417. Geels, M.J., et al., *Broad cross-clade T-cell responses to gag in individuals infected with human immunodeficiency virus type 1 non-B clades (A to G): importance of HLA anchor residue conservation.* Journal of virology, 2005. **79**(17): p. 11247-11258.
418. Peters, H.O., et al., *An integrative bioinformatic approach for studying escape mutations in human immunodeficiency virus type 1 gag in the Pumwani Sex Worker Cohort.* Journal of virology, 2008. **82**(4): p. 1980-1992.
419. Ball, S.C., et al., *Comparing the ex vivo fitness of CCR5-tropic human immunodeficiency virus type 1 isolates of subtypes B and C.* Journal of virology, 2003. **77**(2): p. 1021-1038.
420. Claiborne, D.T., J.L. Prince, and E. Hunter, *A restriction enzyme based cloning method to assess the in vitro replication capacity of HIV-1 subtype C Gag-MJ4 chimeric viruses.* J Vis Exp, 2014(90).