HIV TYPE 1 GENETIC DIVERSITY IN SELECTED COUNTIES OF KENYA

KITAWI ROSE CHARI

MASTER OF SCIENCE

(Epidemiology)

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

2016

HIV Type 1 Genetic Diversity In Selected Counties Of Kenya

Rose Chari Kitawi

A thesis submitted in partial fulfillment for the Degree of Master of Science in Epidemiology in the Jomo Kenyatta University of Agriculture and Technology

2016

DECLARATION

This thesis is my original work and has not been presented for a degree at any other university.

Signature…………………………………. Date………………………..

Rose Chari Kitawi

This thesis has been submitted to the University for Examination with our approval as supervisors:

Signature…………………………………. Date……….

Prof. Washingtone Ochieng'
KEMRI, KENYA

Signature…………………………………. Date………………………..

Prof. Anne W. T. Muigai,
JKUAT, KENYA

DEDICATION

To my loving parents who have sacrificed so much for me to reach where I am.

ACKNOWLEDGEMENT

I am really grateful for the 'help from above'. What would I do without You?

I am grateful to my parents, my brothers and my sisters for being there for me whenever I needed them. Thanks so much for your support.

I acknowledge the help of Professor Washingtone Ochieng and Professor Anne Muigai, my supervisors, for valuable guidance and input during my work and thesis write-up.

This research project would have come to nothing were it not for the support and valuable input of my colleagues Timothy, Maureen, Ruth, Geoffrey, Javan, Nancy, Joyceline, Lucy and all the KEMRI CVR staff who have really been resourceful in this project.

I am grateful to the patients who consented to having the study done on them. Without you, the project would not even have begun.

TABLE OF CONTENTS

DECLARATION .... **ii**

DEDICATION .... **iii**

ACKNOWLEDGEMENT .... **iv**

LIST OF FIGURES .... **ix**

LIST OF TABLES .... **x**

LIST OF APPENDICES .... **xi**

LIST OFABBREVIATIONS .... **xii**

DEFINITION OF OPERATIONAL TERMS .... **xv**

ABSTRACT .... **xvi**

CHAPTER ONE **1**

INTRODUCTION **1**

1.1Background Information 1

1.2 Statement of the Problem 3

1.3 Justification of the Study 4

1.4 Research Questions 5

1.5 Objectives 5

1.5.1 General Objective 5

1.5.2 Specific Objectives 5

CHAPTER TWO **6**

LITERATURE REVIEW **6**

2.1 The HIV-1 Structure and Genome Organization 6

2.2 HIV-1 Diversity 8

2.3 Epidemiology of HIV strains 9
2.4 The global HIV-1 genomic database and phylogenetic profiles 12
2.5 Kenyan HIV-1 Genetic Profile....13
2.6 Co-receptor Usage....16
2.6.1. HIV Phenotypes and Coreceptor Usage 16
2.6.2 Mechanism of viral entry into host cell and coreceptor switching.... 16
2.7 HIV Envelop Diversity and Potential N-Linked Glycosylation Patterns....17
CHAPTER THREE 20
MATERIALS AND METHODS 20
3.1 Study design 20
3.2 Study site....20
3.3 Study population 22
3.3.1 Inclusion Criteria.... 22
3.3.2 Exclusion criteria.... 22
3.4 Sampling....22
3.4.1 Sample Size Determination 22
3.4.2 Sampling Procedure 23
3.5 Ethical Considerations....23
3.6 Laboratory and Analytical Procedures 23
3.6.1 Sample processing.... 23
3.6.2 Viral Load Determination 24
3.6.3 Nucleic Acid Preparation 24
3.6.3.1 DNA Extraction Procedure.... 24
3.6.3.2 RNA Extraction Procedure.... 26
3.6.4 Choice of Primers for Polymerase Chain Reaction (PCR).... 27
3.6.5 Reverse Transcription Polymerase Chain Reaction (RT-PCR) and Round 1 PCR for RNA Samples 28
3.6.6 Nested PCR for RNA Samples.... 29
3.6.7 DNA PCR.... 29

3.6.8 Gel Electrophoresis 30
3.6.9 Purification and Sequencing 31
3.6.10: HIV-1 Subtyping and Subtype Analysis 31
3.6.11: Determination of Coreceptor Usage 32
3.6.12: Determination of envelope protein diversity 33
3.7 Data Management and Analysis 34

CHAPTER FOUR 35

RESULTS 35

4.1 Patient Characteristics 35
4.2 HIV-1 subtyping 39
4.2.1 Phylogenetic Analysis 39
4.2.1.1. Phylogenetic analysis for the envelope C2V3 Sequences 39
4.2.1.2 Phylogenetic Analysis of the Pol RT sequences 41
4.2.2 Comparative sensitivities of common genotyping tools used for HIV-1 subtyping 43
4.2.3 Genetic Heterogeneity of HIV-1 *Pol RT* gene from different blood compartments 45
4.2.4 Genetic heterogeneity of HIV-1 *C2V3 envelope gene* from different blood compartments 46
4.2.5 Comparative distribution of HIV genetic variants between blood compartments 47
4.2.6 Subtype Distribution by County 48
4.3 Co-receptor Usage 50
4.4 Patterns of Potential N-Linked Glycosylation Sites (PNGs) 52

CHAPTER 5 58

DISCUSSION 58

5.1 Patient Characteristics 58
5.2 HIV-1 subtype diversity 59
5.3 Co-receptor Usage 62
5.4 Number of Potential N-Linked Glycosylation (PNG) Sites 63
5.5 Conclusion 64
5.6 Limitations of the study 65

5.7 Recommendations 65

REFERENCES.. 66

APPENDICES.. 79

Appendix 1A: Consent Information for Study Participants (English) 79

Appendix1B: Fomu ya idhini kwa wanauhusika katika utafiti huu (Kiswahili version) 82

Appendix 2A: Parental/Guardian consent for children 13-17 years (English).....................85

Appendix 2B: Idhini ya Mzazi wa mtoto mwenye umri wa miaka 13-17 (Kiswahili Version)...........86

Appendix 3A: QUESTIONNAIRE (English version) 87

Appendix 3B: MAHOJIANO...89

Appendix 4: Analysis of concordance between various subtyping tools 91

Appendix 5: Coreceptor Usage and Number of Potential N-Glycosites 96

Appendix 6: Request for SSC and ERC waiver 99

Appendix 7: ERC Approval 100

Appendix 8: SSC Approval..101

LIST OF FIGURES

Figure 2.1: Diagram of mature HIV virion 7

Figure 2.2: The HIV-1 genome map based on HXB2 reference sequence.
HIV Immunology 2002 7

Figure 2.3: Regional distribution of HIV 1 subtypes and recombinants in 2000-2003 (a)
and 2004-2007 (b) 11

Figure 2.4: Neighbour joining phylogenetic tree for HIV-1 group M viruses 13

Figure 3.1: Map of Kenya Counties with corresponding HIV prevalence 21

Figure 4.1: Phylogenetic Relationships of HIV-1 envelope *C2V3 gene* isolates 40

Figure 4.2: Phylogenetic Relationships of HIV-1 envelope *pol RT gene* isolates 42

Figure 4.3: Proportion of sequences containing PNGs at specific amino acid positions.......56

LIST OF TABLES

Table 1.1: Worldwide Distribution of HIV (in Thousands). 2

Table 2.1: Subtype Distribution in Kenya 15

Table 3.1: Pool of HIV specific primers used for amplification of *env* and *pol RT* regions by polymerase chain reaction (PCR) 28

Table 4.1: Patient characteristics based on *pol RT* phylogenetic subtypes 37

Table 4.2: Patient characteristics based on *C2V3* phylogenetic subtypes 38

Table 4.3: Comparative concordance of various subtyping tools for assignment of HIV-1 *envC2V3* subtypes 44

Table 4.4: Comparison of concordance for various subtyping tools for the *pol RT* gene.... 45

Table 4.5: Distribution and compartmentalization of HIV-1 *Pol-RT* subtypes.................... 46

Table 4.6: Distribution and compartmentalization of *envelope C2V3* subtypes 47

Table 4.7 Comparative frequency of subtypes by genetic region and compartment 48

Table 4.8: Variability of *C2V3* HIV genotypes **within** a given geographic locations across Kenya 49

Table 4.9: Variability of *Pol RT* HIV genotypes **within** a given geographic locations across Kenya 50

Table 4.10: The distribution by number and (proportions), of viral tropism across HIV subtype variants 52

Table 4.11: The distribution and clustering of specific amino acid PNG sites according to viral tropism and NA source material 54

Table 4.12: Mean PNGs for 55 isolates by subtype, nucleic acid source and tropism 55

LIST OF APPENDICES

Appendix 1A: Consent Information for Study Participants (English) 79
Appendix 1B: Consent Information for Study Participants (Kiswahili) 82
Appendix 2A: Parental/Guardian Consent for Children 13-17years (English) 85
Appendix 2B: Parental/Guardian Consent for Children 13-17years (Kiswahili) 86
Appendix 3A: Questionnaire (English) 87
Appendix 3B: Questionnaire (Kiswahili) 89
Appendix 4: Analysis of concordance of various subtyping tools 91
Appendix 5: Coreceptor Usage and Number of Potential N-Glycosites...................... 96
Appendix 6: Letter to request waiver for ERC and SSC 99
Appendix 7: Letter for ERC Approval 100
Appendix 8: Letter for SSC Approval 101

LIST OFABBREVIATIONS

3TC	Lamivudine
ABC	Abacavir
AIDS	Acquired Immune Deficiency Syndrome
ARV/ART	Antiretroviral Therapy
AZT	Azidothymidine (Zidovudine)
CCC	Comprehensive Care Centre
cDNA	complement deoxy ribonucleic acid
COMET	COntext-based Modeling for Expeditious Typing
CREATES	Centre for Research in Therapeutic Sciences
CRFs	Circulating Recombinant Forms
CVR	Centre for Virus Research
d4T	Stavudine
DDBJ	DNA Databank of Japan
DNA	Deoxy-ribonucleic Acid
dNTPs	Deoxyribonucleotide triphosphate
DRC	Democratic Republic of Congo
EDTA	Ethylene Diamine Tetraacetic Acid
EFV	Efavirenz
EMBL-EBI	European Molecular Biology Laboratory’s European Bioinformatic Institute
ERC	Ethical Review Committee
FPR	False Positive Rate (used here to mark the cut-off for coreceptor usage).
HAART	Highly Active Anitretroviral Therapy
HIV	Human Immunodeficiency Virus
HIV-1	Human Immunodeficiency Virus Type 1
JPHMM	Jumping Profile Hidden Markov Model
KEMRI	Kenya Medical Research Institute
KDHS	Kenya Demographic and Health Survey
KNBS	Kenya National Bureau of Statistics

LANL	Los Alamos National Laboratory
MEGA	Molecular Evolutionary Genetic Analysis
$MgCl_2$	Magnesium Chloride
min	Minutes
ml	Millilitres
mM	Millimolar
NACC	National AIDS Control Council
NASCOP	National AIDS and STI Control Programme
NCBI	National Center for Biotechnology Information
NNRTI	Non Nucleoside Reverse Transcriptase Inhibitors
NSI	Non-Syncytium Inducing
nt	Nucleotide
NVP	Nevirapine
PBMCs	Peripheral Blood Mononuclear Cells
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PLWHA	People Living With HIV AIDS
pmol	Picomole
PNGs	Potential N-linked Glycosylation Sites
REGA	Subtyping tool of the Rega Institute
RNA	Ribonucleic Acid
RT	Reverse Transcriptase
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
s	Seconds
SDR	Subtype Diversity Ratio
SI	Syncytium Inducing
SOP	Standard Operating Procedures
SPSS	Statistical Package for Social Sciences
SSC	Scientific Steering Committee
TDF	Tenofovir disoproxil fumarate (Tenofovir)
URFs	Unique Recombinant Forms

UV Ultra violet

µg micrograms

µl Microlitres

DEFINITION OF OPERATIONAL TERMS

CCR5 G-protein coupled beta chemokine receptor type 5 in the C-C chemokine group

CXCR4 C-X-C chemokine receptor type 4

HXB2 This HIV virus is the most commonly used reference strain for many different HIV studies. All structural data related to the envelope region published so far translates residue numbers into the HXB2 numbering scheme.

NXS Sequon with N referring to Asparagine, X to any amino acid apart from proline and S to serine

NXT Sequon with N referring to Asparagine, X to any amino acid apart from proline and T to Threonine

Sequon A sequon is a sequence of consecutive amino acids in a protein that can serve as the attachment site to a polysaccharide, which is frequently an N-linked glycan. The polysaccharide is linked to the protein via the nitrogen atom in the side chain of asparagine (Brooks, Dwek, & Schumacher, 2002).

ABSTRACT

The human immunodeficiency virus (HIV) is one of the most highly genetically diverse viruses known. This extensive genetic variation has many implications on anti-retroviral drug response, viral transmission and vaccine and diagnostics designs and development. A few studies conducted over the last 10 years have shown that HIV-1A is the predominant subtype circulating in Kenya, although there is evidence of increasing emergence of subtypes C and D. This cross-sectional study was conducted in Kisumu and Homabay, Kajiado, Nakuru, Kiambu and Kilifi Counties as part of a countrywide genetic study of HIV-1 diversity. The aim was to describe HIV-1 subtype diversity in the designated counties and the tropism and potential N-linked glycosylation site (PNGs) patterns of the isolates obtained. Five ml of blood was obtained from consenting HIV-1 positive patients and processed to obtain plasma and cells. The plasma was used to obtain viral RNA from the free virus and cells were used to obtain viral DNA. The nucleic material obtained was amplified using polymerase chain reaction (PCR), purified and then sequenced. The sequences were used to conduct phylogenetic analysis so as to deduce evolutionary patterns of HIV-1 using distance matrix. Co-receptor usage and potential N-linked Glycosylation (PNG) patterns of HIV-1 envelop was examined using generic bioinformatics tool. The results showed that subtype A1 is the predominant subtype (70%) although the percentage of recombinants is increasing. With up to 30% recombinant strains observed in the *pol RT* region, recombination appeared to be highest in the cell-free compartment (in plasma)- 31.8%, than in cellular compartment– 10.5%. The partial envelope sequences were also examined for coreceptor usage and the patterns of PNGs. At three False Positive Rate (FPR) cut-off points for CXCR4 usage i.e. 5%, 10% and 20%, the relationship between subtype and viral tropism was found to be significant (p=0.037; p=0.016 and p=0.005 for FPR of 5%, 10% and 20% respectively) using χ^2 test. A significant difference was found between subtype and mean PNGs, using ANOVA, at specific amino acid positions - N296 (p=0.021), N302 (p=0.034) and N366 (p=0.016) but not for the specific PNG patterns examined (NXT, NXS and NNX(S)T). Although R5-tropic isolates had fewer PNGs, on average, than X4-tropic isolates, the difference was not significant (p=0.303). The trend of R5-tropic isolates having more PNGs than the X4-tropic isolates was observed for subtype A1 and not the other subtypes. A significant difference was found when PNGs were examined by nucleic acid source for the NXT pattern (p=0.016), and when total PNGs for all

patterns was considered (p=0.011); with the DNA isolates having more PNGs or average than the RNA virus. The PNG pattern results obtained call for more studies to be done on subtype A1 as it has unique characteristics in the PNG pattern. Continued surveillance of the prevailing subtypes in the different regions in the country will help monitor the transmission patterns of the virus and therefore help in coming up with adequate measures to reduce infection rates.

CHAPTER ONE

INTRODUCTION

1.1Background Information

The UNAIDS 2013 Global Fact Sheet estimates that 35 million people were living with HIV globally in 2013, of these, 3.2 million were children (UNAIDS, 2014). The greatest burden of HIV AIDS lies in sub-Saharan Africa, which has 70%, or 24.7 million of the global population of people living with HIV-AIDS (PLWHA). The region with the least number of people living with HIV AIDS is the Middle East and North Africa. The burden of new infections was also highest in Africa (1.5 million) and lowest in the Carribean (12,000). There were about 1.1 million AIDS related deaths in Sub-Saharan Africa, the global AIDS related deaths being about 1.5 million. The global regional epidemiologic statistics of HIV-1 is summarized in Table 1.1. So far four distinct lineages of HIV-1 have been characterized from sequences accumulated in the past 25-30 years and include groups M (Main), O (Outlier), N (non-M, non-O or New) and P (Putative) (Keele et al., 2006; Plantier, Leoz, & Dickerson, 2009). The majority of global HIV-1 infections are due to Group M viruses which display a tremendous amount of genetic variability (Worobey, 2007). Nine different (pure) subtypes (subtypes A, B, C, D, F, G, H, J, and K), over 55 circulating recombinant forms (CRFs) and several unique recombinant forms (URFs) have been described (Foley *et al.*, 2013). Currently, in the Los Alamos HIV database, over 70 CRFs have been described. Inter-subtype recombinant genomes are commonly found in dually infected individuals, while CRFs arise from HIV-1 recombinant genes that have infected three or more persons who are not related epidemiologically (R. Lihana, Ssemwanga, Abimiku, & Ndembi, 2012). Subtypes A and F have additional sub-subtypes A1, A2, A3, A4, A5, F1 and F2 based on genetic variation of between 15-20% (R. Lihana et al., 2012; Taylor & Hammer,

2008). Like other genetic variants, recombinant forms show discrete geographic distribution patterns and contribute up to 20% of all global HIV infections.

Table 1.1: Worldwide Distribution of HIV (in Thousands).

Region	**People Living with HIV 2013**		**New HIV infections 2013**		**AIDS related deaths 2013 (Total)**
	Total	*Children*	*Total*	*Children*	
Sub-Saharan Africa	24,700 (23,500 – 26,100)	2,900 (2,600 – 3,200)	1,500 (1,300-1,600)	210 (180–250)	1,100 (1,000-1,300)
Asia and the Pacific	4,800 (4,100 -5,500)	210 (190–270)	350 (250–510)	22 (18– 32)	250 (210–290)
Latin America	1,600 (1,400 – 2,100)	35 (27– 54)	94 (71–170)	1.8 (<1 – 7.4)	47 (39- 75)
Western and Central Europe and N.America	2,300 (2,000 – 3,000)	2.8 (2.3-3.6)	88 (44–160)	<0.5 (<0.2-<0.5)	27 (23– 34)
Eastern Europe and Central Asia	1100 (980–1300)	14 (13– 14)	110 (86–130)	1 (<1– 1.2)	53 (43– 69)
Carribean	250 (230–280)	17 (14– 20)	12 (9.4 – 14)	<1 (<0.5 - <1)	11 (8.3– 14)
Middle East and North Africa	230 (160–330)	16 (11– 22)	25 (14- 41)	2.3 (1.5-3.4)	15 (10– 21)
Global	35,000 (33,200 - 37,200)	3,200 (2,900-3,500)	2,100 (1,900 - 2,400)	240 (210, – 280)	1,500 (1,400 – 1,700)

Source: UNAIDS 2013 Global Fact Sheet (UNAIDS, 2014)

Phylogenetics employs molecular techniques to study evolutionary relationships among organisms or genes by a combination of molecular biology and statistical techniques (Li, 1997), and has been useful in distinguishing genetically related groups of organisms. Phylogenetic analysis of HIV strains reveals distinct clusters of HIV-1, with well-defined patterns and trends of global distribution. Using these methods for example, the global HIV-1 epidemic can be traced to the Democratic Republic of Congo (DRC). Studies by Rambaut and colleagues showed that the DRC strains appeared to be basal hence the hypotheses that each global subtype is a result of chance exportations of the strains from founder viruses from the Democratic Republic of Congo into susceptible risk groups followed by diversification (A Rambaut, Robertson, Pybus, Peeters, & Holmes, 2001). These founder effects are what have given rise to the long evolutionary branches observed for each of the subtypes (Worobey, 2007).

Circulating Recombinant Forms (CRFs) on the other hand are represented by mosaic genomes that arose from different parent genomes (Robertson, Sharp, McCutchan, & Hahn, 1995), hence unlike pure viral strains, their phylogenetic relationships are dependent on the section of the genome being examined. Unique recombinant forms are produced when two or more strains undergo recombination within an individual host, before the resulting virus spreads beyond that host (Hoelsher & et al., 2001).

Overall, HIV genetic variations affect anti-retroviral drug response, viral transmission and diagnostic and preventive vaccine strategies. Genetic variability is also important in projecting morbidity and mortality and other key outcomes of infection (Hemelaar, Gouws, Ghys, & Osmanov, 2011). For effective prevention, a HIV vaccine will have to be developed that offers protection against at least majority of the circulating virus strains (Thomson, Perez-Alvarez, & Najera, 2002).

1.2 Statement of the Problem

HIV-1 has a high mutation rate that is associated with continuous generation of divergent virus populations. These mutants may have different rates of transmission and virulence (Hemelaar et al., 2011; N. Kiwanuka et al., 2010).

Like the global epidemic, the Kenyan HIV-1 epidemic follows a discrete but differential genetic distribution pattern across regions of the country. Data shows that the variability of the epidemic continues to evolve at the genetic level in Kenya, with new distribution patterns of both pure and recombinant strains as well as declining dominance of HIV subtype A (HIV-1A). This diversity affects intervention measures. For example, a vaccine developed to prevent HIV-1 subtype A is unlikely to be effective against HIV-1 subtype C (Corey & McElrath, 2010). It is important to understand the present and evolving trends of HIV-1 genotypes if we are to provide reliable data for disease prognosis, treatment and prevention (Baeten *et al.*, 2007; Vasan et al., 2006). This study was conducted as part of an ongoing need to maintain a repository of knowledge on virus heterogeneity and generate data that will inform policy, scientific and medical strategies focused on controlling the spread and burden of HIV-1 and AIDS in Kenya.

1.3 Justification of the Study

An understanding of the molecular distribution of HIV-1 within Kenya is an important public health concern. Since HIV-1 subtypes vary widely between geographical regions and to some extent within localized populations, intervention strategies must be tailored to unique epidemic characteristics of those regions. The counties were chosen from facilities from the west of the country to the East. HIV-1 molecular epidemiological studies have not been carried out in most of the health facilities under study (Kitengela Health Centre, Naivasha District Hospital, Ndhiwa District Hospital, Ahero district hospital)

In order to develop and implement successful intervention strategies, it is important that the molecular trends and patterns of the HIV strains be constantly monitored and updated. In addition, knowledge of HIV-1 diversity is pertinent to successful anti-retroviral management and monitoring. HIV genotypic variability also affects viral tropism, which has been shown to affect disease progression and pathogenicity (Pollakis *et al.*, 2001). In spite of the great significance of viral tropism, there is paucity of data on HIV tropism in Kenya. Generating such data will significantly contribute to the understanding of disease course and burden among the population, as well as inform

alternative intervention strategies. This study avails new sequence information that describes HIV diversity at genetic and protein level, which sequences have also been deposited to the global databases to broaden scientific community's knowledge of local HIV-1 epidemic.

1.4 Research Questions

1. What is the current genotype distribution pattern of HIV-1 in the Kenyan Counties of Kisumu, Homabay, Kajiado, Nakuru, Kiambu, and Kilifi.
2. How does the genetic variability in HIV from these counties relate to HIV coreceptor usage and viral envelope protein characteristics?
3. How does this genetic variability relate with disease outcomes and phenotypic characteristics?

1.5 Objectives

1.5.1 General Objective

To describe the genetic diversity of HIV-1 within Kisumu, Homabay, Kajiado, Nakuru, Kiambu, and Kilifi counties of Kenya.

1.5.2 Specific Objectives

1. To describe the genotypic distribution pattern of circulating HIV-1 strains from Kisumu, Homabay, Kajiado, Nakuru, Kiambu, and Kilifi.
2. To determine the patterns of HIV-1 co-receptor usage of isolates from the study counties.
3. To describe HIV-1 envelope diversity based on the patterns of Potential N-linked glycosylation sites of different viral genotypes.

CHAPTER TWO

LITERATURE REVIEW

2.1 The HIV-1 Structure and Genome Organization

HIV-1 and HIV-2 belong to the genus Lentivirus, family of *Retroviridae* and subfamily *Orthoretrovirinae* (International Agency for Research on Cancer., 1996). Retroviruses transcribe their genome from RNA into DNA using the enzyme reverse transcriptase. The integrase enzyme then integrates the viral DNA into the host's genome, becoming part of the host's cellular DNA. The integrated virus genome subsequently undergoes replication during the natural cellular processes (Nisole & Saib, 2004).

Figure 2.1 shows a representation of a mature HIV virion. All lentiviruses are enveloped by a lipid bilayer, which is derived from the membrane of the host cell. The surface glycoproteins gp120 (labeled SU in the diagram; SU= surface glycoprotein or = sub-unit) are anchored to the virus *via* the transmembrane (TM) glycoprotein gp41. A matrix shell of the matrix-associated protein p17 (MA in Figure 2.1) lines the inner surface of the viral membrane. At the viral core is the capsid protein p24 (CA), located at the center of the virus. The capsid particle encapsulates two copies of the RNA viral genome. The viral genome is stabilized within the capsid particle by the nucleocapsid (NC) protein p7. The capsid particle also contains three essential virally encoded enzymes: protease (PR), reverse transcriptase (RT) and integrase (IN). Virus particles also package the accessory proteins, Nef, Vif and Vpr and three additional accessory proteins Rev, Tat and Vpu (Turner & Summers, 1999).

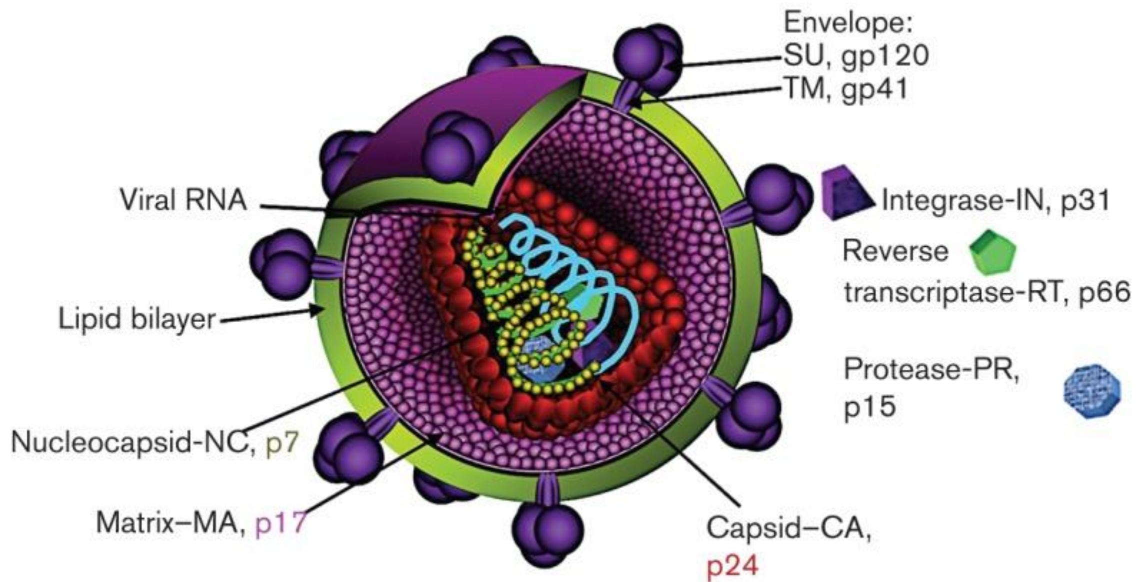


Figure 2.1: Diagram of mature HIV virion (Steckbeck, Kuhlmann, & Montelaro, 2013)

The HIV genome consists of genes – *env, gag and pol* that code for structural proteins (Figure 2.2) . The genome also has six regulatory genes - *tat, rev, nef, vif, vpr,* and *vpu*- that code for proteins that control the ability of HIV to infect a cell, produce new copies of virus, or cause disease (Bradac, Charles, Kristina, Catherine, & James, 2002).

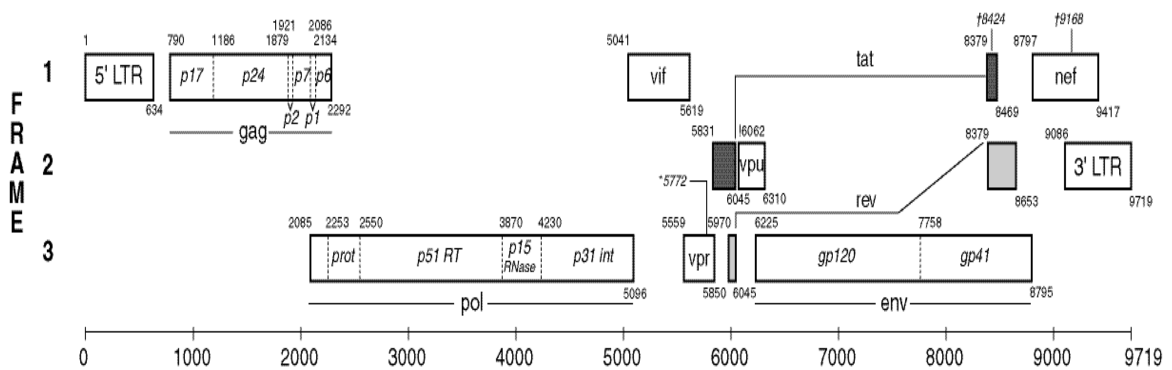


Figure 2.2: The HIV-1 genome map based on HXB2 reference sequence. HIV Immunology 2002 (Bradac *et al.*, 2002).

2.2 HIV-1 Diversity

The Human Immunodeficiency Virus is an RNA virus of the retrovirus family, with two RNA strands making up its genome. Its genetic diversity is believed to derive from three different sources; firstly, through multiple introductions of HIV into the human population, secondly through the low fidelity and high recombinogenic power of the viral reverse transcriptase enzyme and finally because of the high virus turnover (Castro-Nallar, Keith A Crandall, & Pérez-Losada, 2012). This, together with high positive selection pressure either due to the drugs or the host immune system results in significant viral genetic variability (Ariën *et al.*, 2005).

HIV-1 mutation rates average approximately 0.1-0.3 mutations per genome per replication cycle (Abram, Ferris, Shao, Alvord, & Hughes, 2010), and is as a result of high replication errors associated with the reverse transcriptase (RT) enzyme. During reverse transcription, the enzyme undergoes a highly inefficient and error-prone template switching between its two RNA genomic strands (Skar, Charlotte, & Jan, 2011). Genomic recombinations occur during this process in the order of about 0.14 recombinations per genome per replication cycle (Skar *et al.*, 2011). These molecular events produce quasi-viral species that proceed to superinfect the same cell or adjacent host cells during host cell replication. Both recombination and RT-dependent replicative error contribute enormously to HIV-1 diversity (Castro-Nallar *et al.*, 2012). Although selection forces in the host act to eliminate some of these viral mutants, purifying/negative selection accompany these selections, generating persistent mutants that are more potent and continue to propagate the epidemic at the population level (Skar *et al.*, 2011).

The evolutionary rate of HIV-1 has been shown to depend on the rate and mode of spread (I. B. Maljkovic et al., 2007). Anti-retroviral treatment further exacerbates evolution of HIV diversity through the selection of drug resistant mutants, particularly individuals who adhere less to treatment or in settings where treatment is not optimal (B. I. Maljkovic et al., 2009). In Kenya, HIV-1 infected people have continued to gain increasing access to Antiretroviral (ARV) treatment (ART). This unprecedented access

to treatment poses significant challenges with treatment adherence as has recently been reported among long-term ART patients (Ochieng *et al.*, 2015). Under sustained ART, more HIV-1 variants emerge either due to non-adherence or selective pressure that produces drug resistant mutants (R.W. Lihana et al., 2011; Lwembe et al., 2007). These 'new' strains emerge from treated individuals and are propagated and transmitted across population as new or sub-epidemics making future treatment of newly infected persons with the same drugs impossible. Other studies have suggested that Highly Active Antiretroviral Therapy (HAART) reduces the rate of evolution of HIV-1 Subtype B and probably other subtypes, largely because therapy can significantly inhibit viral replication and diminish accumulation of mutants (B. I. Maljkovic et al., 2009). As treatment becomes more available in Kenya, the potential for viral evolution and diversity become even more real.

2.3 Epidemiology of HIV strains

The Human Immunodeficiency Virus can be classified into two groups: HIV-1 and HIV-2. It has become clear that HIV resulted from zoonotic transfers of viruses infecting primates in Africa (Hahn, Shaw, De Cock, & Sharp, 2000). HIV-2 is mainly restricted to Western Africa, and is not associated with AIDS epidemic (Santiago *et al.*, 2005). The HIV-1 group M pandemic is believed to have originated in the Congo River Basin and has been traced by evolutionary analyses to having originated from there between the years 1915- 1941 (Korber *et al.*, 2000). HIV-1 is responsible for the global AIDS epidemic and exists in diverse genetic forms distributed distinctly across the globe (Hemelaar *et al.*, 2011).

Group M of HIV-1 contributes to about 90% of HIV infections in the world (Spira, Wainberg, Loemba, Turner, & Brenner, 2003). Hemelaar *et al.* carried out a global HIV-1 epidemiology study on samples obtained between 2004 and 2007. Subtype B viruses were found to be predominant in North America, the Caribbean, Latin America, western and Central Europe and Australia. The epidemic in Eastern Europe and central Asia is dominated by a mix of subtypes A and B. CRF01_AE predominates south and south-east Asia regions while CRF07_BC, CRF08_BC, CRF01_AE and subtype B dominate

East Asia. The Middle East and North Africa are dominated by subtype B and CRFs while West Africa features a highly heterogeneous epidemic with all the subtypes represented but predominated by CRF02_AG and subtype G. East Africa on the other hand has predominantly subtype A in circulation, with subtypes D, C and URFs also emerging or increasing. Ethiopia, Southern Africa and India are predominated by subtype C (Hemelaar *et al.*, 2011). Overall, HIV-1 subtype B accounts for about 11% of the global HIV burden, subtype C accounts for 48%, subtype A 12%, CRF02_AG (8%), CRF01_AE (5%), subtype G (5%) and D (2%). Subtypes F, H, J and K together cause less than 1% of infections worldwide. Other CRFs and URFs are each responsible for 4% of global infections, bringing the combined total of worldwide CRFs to 16% and all recombinants (CRFs along with URFs) to 20% (Hemelaar *et al.*, 2011).

The explanation for the current global HIV-1 spread and diversity are multi-factorial: population growth, founder effects, urbanization, migration and improved transport links all have played major part (Hemelaar *et al.*, 2011). It is thought that subtype C was introduced into Southern Africa, India, China and Brazil through a preferential introduction into a high risk heterosexual group (Walker, Pybus, Rambaut, & Holmes, 2005). In a study investigating the historical and evolutionary trends of HIV-1 in East Africa, it was found that HIV-1A and D were introduced into the region after 1950 and spread exponentially during the 1970s (Gray et al., 2009). The rapid growth and divergence of the epidemic can be attributed to the strong interconnectivity between population centers across the East African region.

The global distribution of HIV-1 is shown in Figure 2.3.

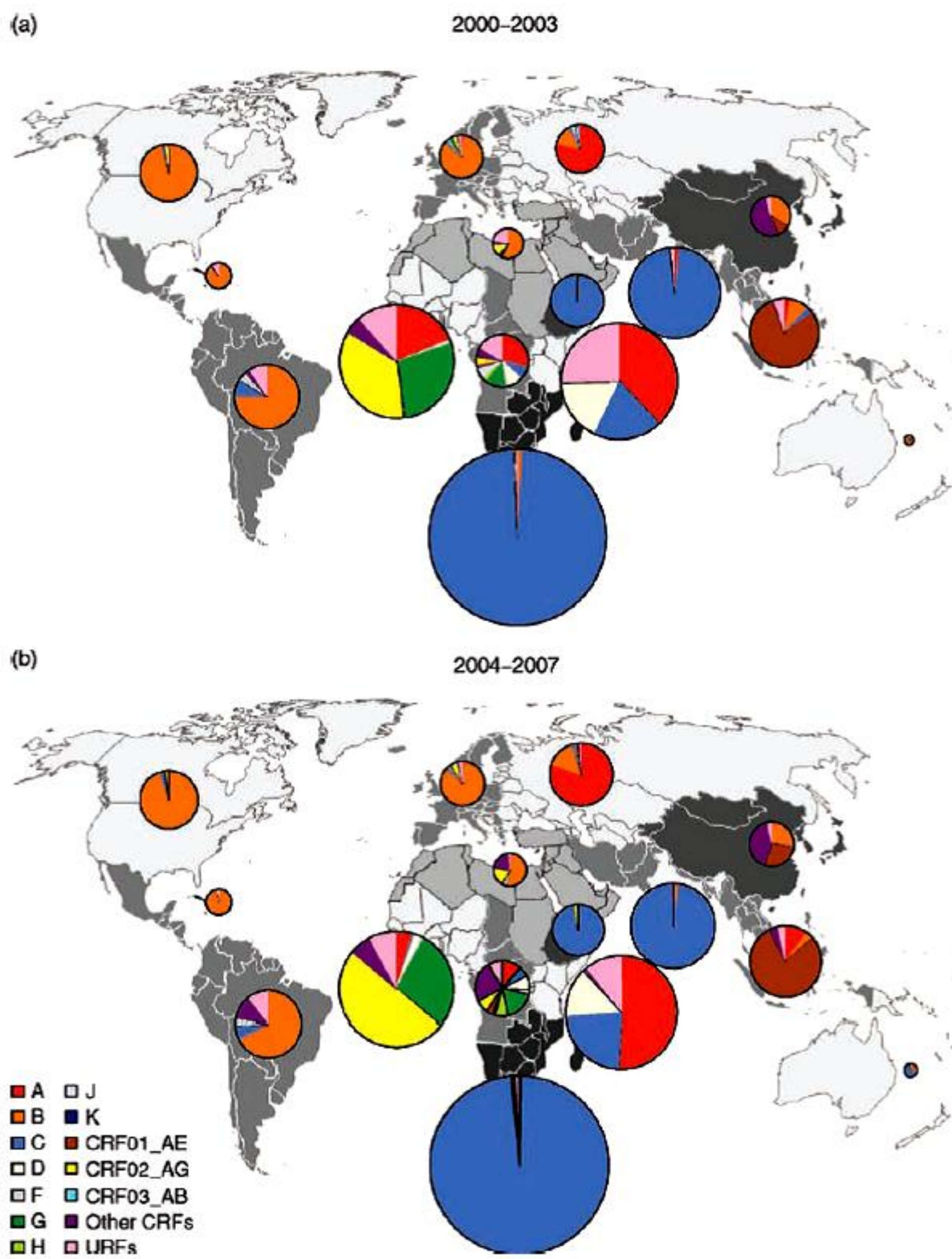


Figure 2.3: Regional distribution of HIV 1 subtypes and recombinants in 2000-2003 (a) and 2004-2007 (b) (Hemelaar *et al.*, 2011).

2.4 The global HIV-1 genomic database and phylogenetic profiles

The HIV databases contain data on HIV genetic sequences, immunological epitopes, vaccine trials and drug resistance-associated mutations. These platforms are useful when analyzing HIV viral sequences for purposes of identification of specific viral genetic and protein properties as well as deciphering epidemiologic trends. Various databases used in HIV subtyping include the Los Alamos HIV database (lanl) {http://www.hiv.lanl.gov/content/index}, The Stanford Drug Resistance Database (http://hivdb.stanford.edu/), Geno2pheno (http://www.geno2pheno.org), Rega Subtyping tool (http://www.bioafrica.net/rega-genotype/html/subtypinghiv.html) amongst others.

The criteria used for the current subtype classification is that subtypes are approximately equidistant from one another in the envelope gene (*env*). Two or more samples are required to define a new sequence subtype (Woutter, Buve, & Nkengasong, 1997). Evolutionary analysis of HIV is then performed on the sequences by phylogenetic inference, which captures the structure and genetic diversity of lineages using phylogenetic trees. While the phylogenies depict major groups in data format, the tree topologies show the evolutionary history of sequences and how these groups are related in time. This far, phylogenetic inference has become useful in categorizing the major groups and subtypes of HIV and their relationships (A. Rambaut, Posada, Crandall, & Holmes, 2004).

Figure 2.4, adapted from (Archer, 2008) shows in (A) neighbor joining phylogenetic tree reconstructed using global group M gp160 sequences sampled from LANL HIV Sequence Database with the lettering representing the subtypes and * representing bootstrap values greater than 90%. (B) shows the relationship between subtypes diversity ratio and the quality of clustering as shown in (A Rambaut et al., 2001). The group M phylogenetic tree can be described as having a double starburst appearance.

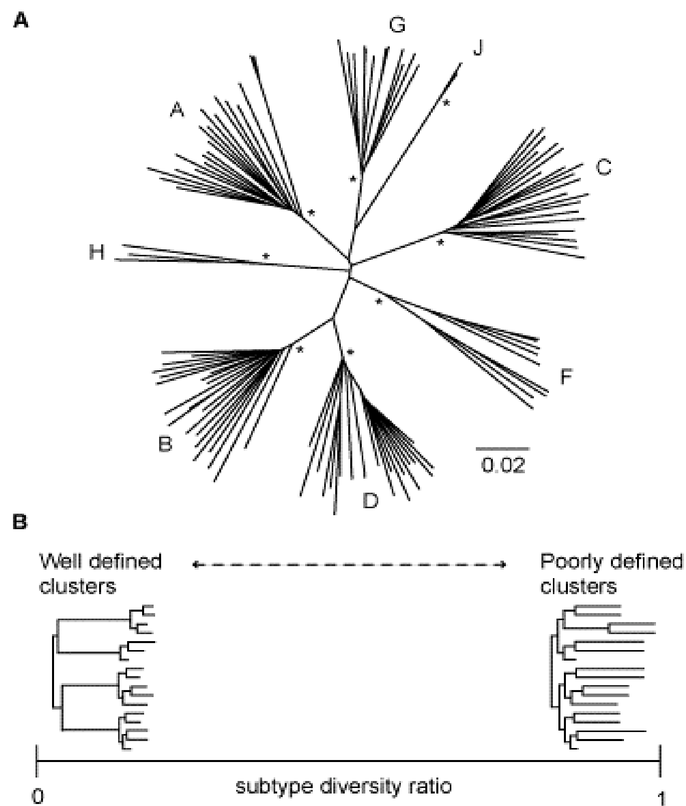


Figure 2.4: Neighbour joining phylogenetic tree for HIV-1 group M viruses (Archer, 2008)

2.5 Kenyan HIV-1 Genetic Profile

According to the National AIDS Control Council (NACC) report 2014, the national prevalence of HIV in Kenya is 6.04%, with the prevalence in women (7.6%) being higher than that in men (5.6%). The epidemic is geographically diverse with prevalence ranging from 25.7% in Homabay to 0.2% in Wajir (NACC., 2014). The predominant subtype in Kenya is HIV-1A, although there is evidence of increasing prevalence of subtypes C and D especially in regions around the northern border with Ethiopia (S. A. Khamadi et al., 2005) and lately in the central regions (Kageha *et al.*, 2012). There have also been reports of a rise in inter-subtype recombinants in parts of Kenya (Nyagaka et al., 2012; Songok et al., 2004). Subtype C has been found to be less fit (hence lower disease progression) than all the other subtypes of Group M but has higher

transmissibility (Abraha et al., 2009; Asamoah-Odei, Garcia Calleja, & Boerma, 2004; Gordon et al., 2003; Koch et al., 2001). The high transmission rate could be due to the fact that subtype C viruses are predominantly R5-tropic; R5-tropic viruses having greater transmissibility than the X4- tropic viruses (Abraha *et al.*, 2009). Patients infected with subtype C seem to have a high viral load but their CD4 counts reduce at a lower rate than those of Subtype A and D therefore explaining their slower disease progression. Slower disease progression and high viral load also means greater opportunity for transmission (Abraha *et al.*, 2009). On the other hand subtype D is associated with fast disease progression and lower infectivity rates (Kaleebu et al., 2007; N. Kiwanuka et al., 2010). One study evaluated the importance of recombinant forms and showed that A/D recombinants are effectively transmitted from mother to child (Steain, Wang, & Saksena, 2006).

Table 2.1 gives a summary of results from different studies within Kenya that give subtype distribution in various regions. In practically all regions, subtype A is the major subtype. One also observes regional differences in subtype distribution.

Table 2.1: Subtype Distribution in Kenya

Region in Kenya	Common Subtype (in %)				
	A	**C**	**D**	**Others**	**Recombinants**
Northern (S. A. Khamadi et al., 2005)	57	27	9	7§	
Coastal (S. A. Khamadi et al., 2009)	86	5	8	1*	
Nairobi(R.W Lihana et al., 2009c)	65	2	2	2*	CRF02_AG - 7 A1/C - 9; A1/D - 7; C/D - 2; A1/A2 - 2; A1/D - 2
North Rift Valley (Nyagaka *et al.*, 2012)	70	7	11	3- A2 1*	A1/D - 4 A1/C - 3 A2/C - 1
Western Kenya (Micah *et al.*, 2011)	51.4	1.4	9.4	4.1*	All recombinants – 33.7
Kericho (Arroyo *et al.*, 2009)	56	5	10		All recombinants - 29
Kilifi (Hue S, 2011)	59	8	10	<1*	All recombinants -22
Kisumu (Yang *et al.*, 2004)	67.4 A2-2.0	6.1	20.4	1.7*; 2.4§	Gag/Env: D/A-8.7; A/D-5.9, C/A-2.4; A/C-1.7
Nairobi (Kebira, Waihenya, & Khamadi, 2009)	50	35.9	10.25		A1/D -3
Southern Kenya (Dowling *et al.*, 2002)	56.9	2.4	2.4		A/D-14.6; A/C-7.3; A/G-2.4; A/G- 2.4 C/D-2.4; A/C/D-2.4
Nairobi (Khoja *et al.*, 2008)	56.5	10.1	18.8	2.9*	AD/CRF01_AE-4.34; AC- 1.45; AD -1.45
Central Kenya (Kageha *et al.*, 2012)	69.8	11.5	18.7		

Notes: §, Unclassified; * Subtype G

2.6 Co-receptor Usage

2.6.1. HIV Phenotypes and Coreceptor Usage

Entry of the HIV virus into the host cell requires cooperate binding and fusion of the viral surface envelope (env) glycoprotein trimer consisting of gp120 and gp41 heterodimers to host cell CD4 receptor. Binding to the CD4 receptor induces a conformational change in Env that enables the engagement of a specific seven-transmembrane chemokine receptor family on the cell surface (Craig B. Wilen, Tilton, & Doms, 2012). HIV-1 specifically uses the CC β-chemokine receptor CCR5, and the CXC α-chemokine receptor CXCR4. These are the predominant means of viral entry into target CD4 T-cells (Pollakis *et al.*, 2001). Based on co-receptor usage, HIV can be characterized as CCR5 (or R5) tropic or CXCR4 (or X4) tropic strains. HIV-1 can further be characterized phenotypically into synctium inducing (SI) and non synctium inducing (NSI) viruses according to differences in cellular tropisms (Koot *et al.*, 1992). NSI viruses are majorly macrophage tropic and are normally observed in the early stages of infection while SI viruses are T-Cell tropic and arise during late stages of disease progression or infection (Archer, 2008; Berger et al., 1998). X4-tropic viruses are largely of SI phenotype while R5-viruses are of NSI phenotype. Thus, co-receptor usage has important association and link to both infection and disease outcome.

2.6.2 Mechanism of viral entry into host cell and coreceptor switching

The HIV-1 *env* gene can be divided into alternating constant (C) regions (C1 - C5) and variable (V) regions (V1 -V5) (Starcich *et al.*, 1986). The *env* region is important because it contains an important target for immune response, determines coreceptor specificity and shows variability that is highly phylogenetically informative (Shankarappa *et al.*, 1999). Crystal structure of gp120 shows that the constant and variable regions can be organized into 4 structural domains i.e. the inner and outer domains, a 4-stranded bridging sheet and the V1/V2. On binding to the CD4 receptor, the inner domain of gp120 undergoes major structural changes that allow for formation of the bridging sheet while the majority of the outer domain appears to remain

unchanged. Subsequent co-receptor binding at areas on the bridging sheet to the outer domain of the virus V3 loop causes additional gp120 conformational changes that result in the ultimate dissociation of gp120 from gp41 and transition of gp41 into a structural form that allow for viral-host membrane fusion (Wang *et al.*, 2013). The virus V1V2 region has been shown to provide the genotypic pattern for R5X4 dual tropism, although it does not affect CCR5 utilization (Pollakis *et al.*, 2001). The overall amino acid charge of the V3 region has an effect on the biological phenotype with a higher positive charge favouring the SI phenotype and a higher utilization of CXCR4 coreceptor. The loss of an N-linked glycosite in the V3 region together with the higher positive charge has been shown to result in strong CXCR4 tropism and is thought to contribute to virus switching from R5 to the X4 phenotype (Pollakis *et al.*, 2001).

The co-receptors are useful target for antiretroviral drugs and candidate therapeutics (Kalinina, Pfeifer, & Lengauer, 2013). So far only CCR5 coreceptor blockers (entry inhibitors) have been developed, for example Maraviroc, which act by blocking the CCR5 coreceptor hence inhibiting entry. Although some candidate compounds have been shown to block viral entry via CXCR4 co-receptor, these have not been developed into antiretroviral agents because CXCR4 is critical for haematopoesis, cardiac function and cerebellar development hence using it as a target can lead to serious side effects (Archer, 2008). Coreceptor usage can be measured or deduced by using either phenotypic assays or genotypic assays (Mulinge *et al.*, 2013). This study used the genotypic method, relying on the Geno2Pheno platform which integrates position matrix analyses for CCR5 and CXCR4 at specific amino acids and glycosylation sites (Sing, Beerenwinkel, & Lengauer, 2004).

2.7 HIV Envelop Diversity and Potential N-Linked Glycosylation Patterns

The HIV-1 gp120 is heavily glycosylated by the infected host, with the glycan moieties accounting for up to 50% of its total mass (Wang *et al.*, 2013). N-Linked glycosylation sites are also referred to as sequons. The position and number of N-Linked oligosaccharides attached to a protein have a profound effect on the protein structure, expression and function (Kasturi, Chen, & Shakin-Eshleman, 1997). An amino acid

sequence with pattern N-X-S or N-X-T is required for glycosylation to take place. N in this context refers to Asparagine, X refers to any amino acid except proline, S refers to Serine and T refers to Threonine. A sequon will not be glycosylated if it contains Proline (Ming, Gaschen, & et al., 2004). Sequons are important because they have an effect on immunogenicity. A change of sequons in the HIV envelope protein gp41 for example can induce a conformational change in the associated HIV gp120 that will dramatically diminish the binding of specific gp120 antibodies (Si, Cayabyab, & Sodroski, 2001). These patterns also influence receptor binding and the phenotypic properties of the virus (Pollakis *et al.*, 2001). They could also cause T cell immune response escape (Botarelli et al., 1991; Ferris et al., 1999). Glycosylation of gp120 can change exposure of neutralizing antibody epitopes. A glycosylation site can therefore provide regional protection from antibodies the extent of which depends on the various viral subtypes. For example, the absence of a sequon near the base of the V3 loop in CRF01 strains of HIV has been correlated with rapid amino acid substitution and positive selection (Kalish *et al.*, 2002). Glycosylation at position 301 at the stem of the V3 loop is thought to play a role in stabilizing the V3 and in modulating coreceptor binding. Mutation at this point results in a drastic decrease in viral infectivity probably due to reduced coreceptor binding (Wang *et al.*, 2013). A study by Chohan and colleagues indicated that there is an apparent selection for HIV subtype A and C variants that are less glycosylated and with shorter V1-V2 loop sequences (Chohan *et al.*, 2005). It has also been found that potential N-linked glycosylation sites (PNGs) interact among themselves with negative or mutually exclusive forces occurring between sites of the same location and positive interaction occurring between sites separated by greater distances (Poon, Lewis, Kosakovsky, & Frost, 2007). The number of N-linked glycosylation sites have also been seen to be significantly different between the plasma and cellular compartments with the plasma compartment showing higher numbers of PNGs (Ho *et al.*, 2008). Thus, N-linked glycosylation of the HIV-1 envelope glycoprotein is one of the most protective mechanisms to overcome in order to develop an effective neutralizing antibody inducing vaccine (Poon *et al.*, 2007). This study also

aimed to map out the PNGs of various HIV-1 isolates and compare these patterns between cellular and acellular blood compartments.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study design

This was a cross-sectional study in which fresh blood samples and demographic information were collected from subjects at designated HIV comprehensive care centers around the country.

3.2 Study site

Subjects were recruited from HIV Comprehensive Care Centers in the counties of Kisumu (Ahero District Hospital), Homabay (Ndhiwa District Hospital) Nakuru (Naivasha District Hospital), Kajiado (Kitengela Health Centre), Kilifi (Malindi District Hospital) and Kiambu (Kiambu District Hospital). The sites were chosen to cover areas with the highest prevalence of HIV (Homabay and Kisumu) located in the Western region to regions with lower prevalence (Kiambu, Nakuru and Kajiado). Kilifi was chosen because the diversity of subtypes in this region and due to the fact that it is on the Eastern coast of Kenya. Kitengela Health Centre in Kajiado was also chosen because the region where it is located is an important stop-over for long distance drivers. The laboratory analysis was done at KEMRI-CVR, Strathmore CREATES (Centre for Research in Therapeutic Sciences) laboratory and at the African Centre for Clinical Trials. According to the National Aids Control Council 2014 report, the HIV prevalence in Kisumu is 19.3%, Homabay 25.7%, Nakuru 5.3%, Kajiado 4.4%, Kiambu 3.8% and Kilifi 4.4%. (NACC., 2014). Figure 3.1 gives the Kenya County HIV-1 prevalence.

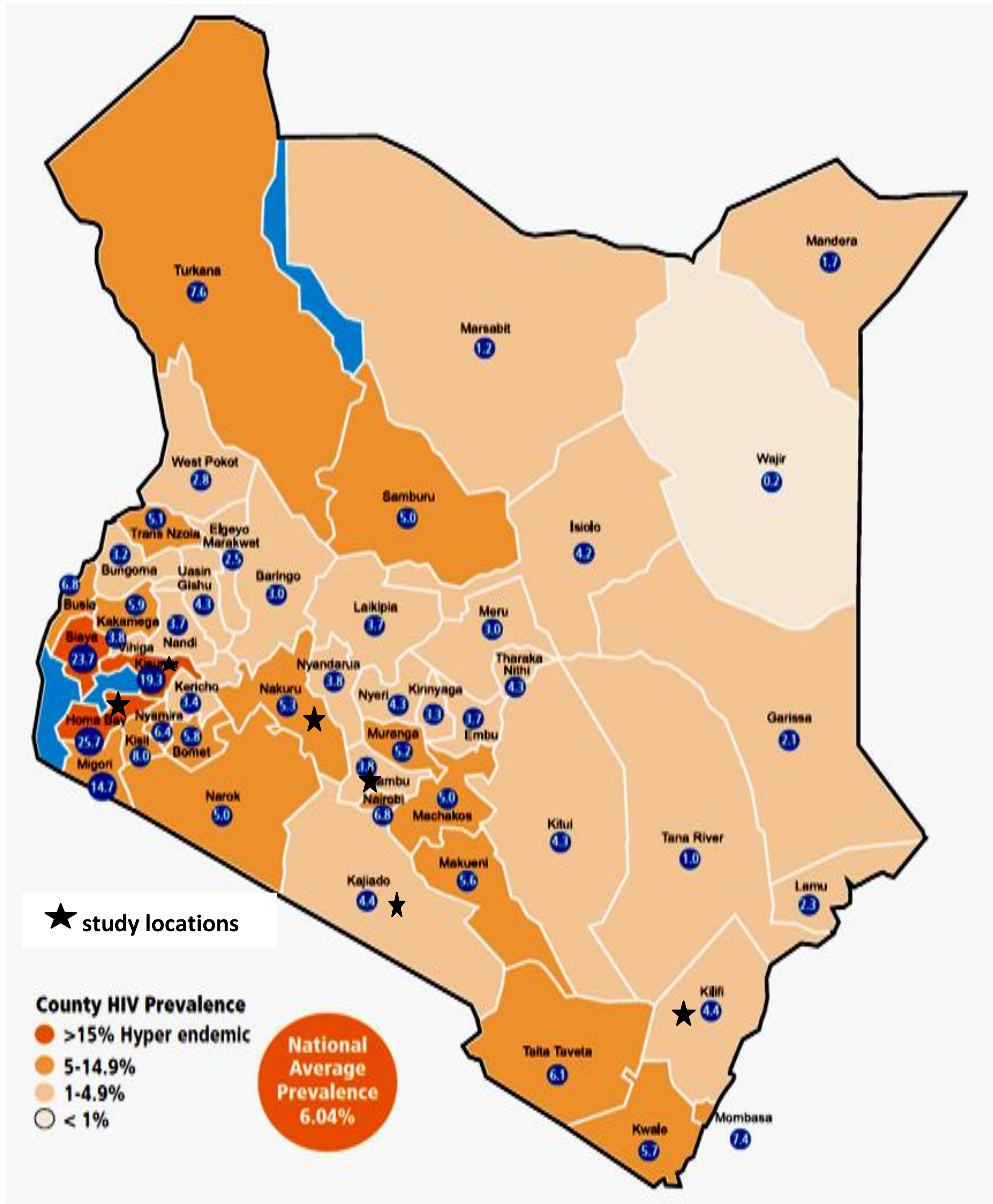


Figure 3.1: Map of Kenya Counties with corresponding HIV prevalence

Source: (NACC., 2014)

3.3 Study population

The study was on HIV-1 positive patients (see sample size determination) aged 13-64 years who were receiving antiretroviral treatment (ART) and care at comprehensive care centres (CCC) of the respective counties.. To get a more homogenous group, the paediatrics (12 years and below) and geriatrics (65 years and above) were excluded. The patients were consented and enrolled between April and August 2013.

3.3.1 Inclusion Criteria

Consenting individuals who:

1. Were receiving or not receiving anti-retroviral drugs and
2. Were aged 13-64 years and appearing at the participating facilities for care and
3. Had a determined HIV status and enrolled for care within the HIV-care facilities.

3.3.2 Exclusion criteria

1. All non-consenting patients
2. Patients whose place of residence was not within the respective counties.

3.4 Sampling

3.4.1 Sample Size Determination

This was a cross-sectional survey and therefore the following Cochran's formula was used.

$$n = \frac{z\alpha^2 * p * q}{d^2}$$

$z\alpha$= z score for α of 0.05

p= national prevalence of HIV = 5.6% from KAIS-2012 (NASCOP., 2014)

q= 1-p

d= precision (5% precision will be used)

$$n = \frac{1.96^2 * 0.056 * 0.944}{0.05^2} = \frac{0.2031}{0.0025} = 81.23$$

Minimum sample size needed for statistical representativeness of population was 81.

3.4.2 Sampling Procedure

Upon consent, a questionnaire was administered to all subjects to collect demographic and relevant clinical information including age, sex, sex partnership, occupation, antiretroviral drug use profile, ART regimen, and travel history. Patient files were also examined to collect information on CD4 counts and viral load where available.

3.5 Ethical Considerations

This study was conducted in accordance with the Helsinki Declaration of 1975, (revised in 2000) (World Medical Association, 2000) and after approval by the Scientific Steering Committee (SSC) and the Ethical Review Committee (ERC) of the Kenya Medical Research Institute (ERC/SSC protocol #2477). Participants were informed adequately about the study and their written consent obtained voluntarily before recruitment. Patient data was coded to ensure confidentiality

3.6 Laboratory and Analytical Procedures

3.6.1 Sample processing

Sample processing was done according to KEMRI CVR Standard Operating Procedures (SOPs). Five milliliters of venous blood was drawn into EDTA vacutainer tubes from each patient. Plasma was obtained by centrifuging the blood at 3000 revolutions per minute (rpm) or 900 x g for 10 minutes. Aliquots of the plasma were frozen at -80°C until needed for RNA processing. The cellular portion of the blood was lysed to obtain Peripheral Blood Mononuclear Cells (PBMCs). The lysis was done by adding 12 ml of 0.84% ammonium chloride to the plasma-free blood in a 15 ml falcon tube, vortexing and incubating for 10 minutes at 37°C. The tubes were then centrifuged at 900 x g for 10 min and the supernatant discarded. This was repeated 3 times. The cell pellets were washed by re-suspending in 1 ml of Phosphate Buffer Saline (PBS, pH 7.4) and centrifuging at 900 x g for 5 minutes in 1.5 ml eppendorf tubes. The resulting PBMCs were frozen as pellets at -80°C until needed for DNA processing.

3.6.2 Viral Load Determination

Only patient samples with viral load above 1000 RNA copies/ml were considered for RNA extraction and reverse transcriptase polymerase chain reaction (RT-PCR). Viral load was determined using m2000 Abbott RealTime HIV-1 assay following manufacturer's protocol (Abbott Laboratories, Abbott Park, Illinois). Briefly, internal control RNA was added to 200µl of plasma and loaded onto m2000sp instrument for RNA extraction. The range for the limits of detection was between 40 and 10,000,000 HIV-1 RNA copies. Undetectable VL was reported as 39 copies. All VL assays were conducted at the Early Infant Diagnosis laboratory of the Kenya Medical Research Institute (KEMRI).

3.6.3 Nucleic Acid Preparation

DNA was extracted from samples that had between undetectable viral load and up to a maximum of 1000 HIV-1 RNA copies/ml blood.

QIAamp DNA kit (Qiagen, Maryland USA) was used for DNA extraction and QIAamp RNA (Qiagen, Maryland USA) kit was used for RNA extraction. Before starting nucleic acid extraction, all samples and distilled water were equilibrated to room temperature. The oven was heated to 56°C and maintained at this temperature. Buffer AW1, Buffer AW2 were prepared by adding the indicated amount of absolute ethanol. All centrifugation steps were carried out at room temperature (15 – 25°C).

3.6.3.1 DNA Extraction Procedure

DNA extraction was done using Qiamp DNA Kit following instructions in the QIAamp® DNA Mini and Blood Mini Handbook. (Qiagen®, 2010). Briefly, Qiagen Protease (Qiagen, Maryland, USA) was prepared by adding the required amount of protease solvent to the tube containing the enzyme.

Twenty microlitres of QIAGEN Protease (or proteinase K) was pipetted into the bottom of a 1.5 ml microcentrifuge tube and 200 µl sample (PBMCs mixed with PBS pH 7.4) was added to the microcentrifuge tube. Two hundred microlitres of buffer AL was then

added to the sample and mixed by pulse-vortexing for 15 s. The mixture was then incubated at 56°C for 10 min. The 1.5 ml microcentrifuge tube was briefly centrifuged to remove drops from the inside of the lid.

Two hundred microlitres of ethanol (96 – 100%) was added to the sample and mixed again by pulse-vortexing for 15 s. After mixing, the 1.5 ml microcentrifuge tube was briefly centrifuged to remove drops from the inside of the lid. The mixture was then applied to the QIAamp Mini spin column (in a 2 ml collection tube) without wetting the rim. The cap was closed and the mixture centrifuged at 6000 x *g* for 1 min.

The QIAamp Mini spin column was placed in a clean 2 ml collection tube, and the tube containing the filtrate was discarded. The QIAamp Mini spin column was carefully opened and 500 µl Buffer AW1 was added without wetting the rim. The cap was closed and the column centrifuged at 6000 x *g* for 1 min.

The QIAamp Mini spin column was then placed in a clean 2 ml collection tube and the collection tube containing the filtrate discarded. The QIAamp Mini spin column was carefully opened and 500 µl of buffer AW2 added without wetting the rim. The cap was closed and the column centrifuged at full speed (20,000 x *g*) for 3 min.

The QIAamp Mini spin column was then placed in a new 2 ml collection tube and the old collection tube with the filtrate was discarded. The column was centrifuged again at full speed for 1 min. The QIAamp Mini spin column was then placed in a clean 1.5 ml microcentrifuge tube and the collection tube containing the filtrate was discarded. The QIAamp Mini spin column was carefully opened and 40 µl of distilled water (RNAse, DNAse free) was carefully pipetted to the centre of the membrane without touching it.

The column was incubated at room temperature (15 – 25°C) for 2 min, and then centrifuged at 6000 x *g* for 1 min. The column was discarded and the 1.5 ml

microcentrifuge tube contained extracted DNA was tightly closed. To prevent frequent thawing that affects the integrity of the DNA, the DNA extracted was divided into two portions and placed in two separate microcentrifuge tubes and stored at -80°C until required.

3.6.3.2 RNA Extraction Procedure

Extraction of RNA from plasma samples was carried following manufacturer's instructions using the QIAamp® Viral RNA Mini Handbook (Qiagen, 2012). Briefly, 310 microlitres of buffer AVE was added to the tube containing 310 μg lyophilized carrier RNA to obtain a solution of 1 μg/μl. The carrier RNA was dissolved thoroughly and divided into conveniently sized aliquots and stored at –20°C.

Carrier RNA was mixed with buffer AVL gently by inverting the tube 10 times, following manufacturer's instructions. Vortexing was not done to avoid foaming.
Five hundred and sixty microlitres of prepared buffer AVL containing carrier RNA was pipetted into a 1.5 ml microcentrifuge tube. One hundred and forty microlitres of plasma was added to the buffer AVL–carrier RNA in the microcentrifuge tube and mixed by pulse-vortexing for 15 s.

The mixture was then incubated at room temperature (15 – 25°C) for 10 min. The tube was briefly centrifuged to remove drops from the inside of the lid. Five hundred and sixty microlitres of ethanol (96 – 100%) was added to the sample, and mixed by pulse-vortexing for 15 s. After mixing, the tube was briefly centrifuged to remove drops from inside the lid.

Six hundred and thirty microlitres of the resulting lysate solution was carefully applied to the QIAamp Mini column (in a 2 ml collection tube) without wetting the rim. The cap was tightly closed to avoid cross-contamination during centrifugation. The column was centrifuged at 6000 x g for 1 min. The QIAamp Mini column was then placed into a

clean 2 ml collection tube and the tube containing the filtrate discarded. This step was repeated with the remaining lysate.

The QIAamp Mini column was carefully opened and 500 µl of Buffer AW1 added. The cap was closed and the column centrifuged at 6000 x *g* for 1 min. The QIAamp Mini column was placed in a clean 2 ml collection tube and the tube containing the filtrate discarded.

The QIAamp Mini column was carefully opened and 500 µl of Buffer AW2 added. The cap was closed and the column centrifuged at full speed (20,000 x *g*) for 3 min.
The QIAamp Mini spin column was then placed in a new 2 ml collection tube and the old collection tube with the filtrate was discarded. The column was centrifuged again at full speed for 1 min.

The QIAamp Mini spin column was again placed in a clean 1.5 ml microcentrifuge tube and the collection tube containing the filtrate was discarded. The QIAamp Mini spin column was carefully opened and 40 µl of distilled water (RNAse, DNAse free water) was carefully pipetted to the centre of the membrane without touching it. The column was incubated at room temperature (15 – 25°C) for 2 min, and then centrifuged at 6000 x *g* for 1 min. The column was discarded and the 1.5ml microcentrifuge tube was tightly closed.

To prevent frequent thawing that affects the integrity of the RNA, the RNA extracted was divided into three portions and placed in three separate microcentrifuge tubes and stored at -80°C until use.

3.6.4 Choice of Primers for Polymerase Chain Reaction (PCR)

Gene specific primers were designed that targeted a 701 base pair (bp) region of *reverse transcriptase* (*RT*) region and corresponding to nucleotides nt(nucleotide)2480- nt3180 of the HIV HXB2 *pol* gene reference sequence. For *envelope* gene, the primers targeted

a 546bp region of *C2V3* gene segment corresponding to nt6975-nt7520 on the HIV HXB2. The primer sequences are shown in Table 3.1.

Table 3.1: Pool of HIV specific primers used for amplification of *env* and *pol RT* regions by polymerase chain reaction (PCR)

HIV-1 region	**Primer Sequence**	**Purpose**	**Citation**
Envelope (C2-V3, nt 6975–7520 HXB2); 546bp	M5_5'- CCAATTCCCATACATTATTGTGCCCCAGCTGG -3' (F) M10_5'- CCAATTGTCCCTCATATCTCCTCCTCCAGG -3' (R)	Round 1PCR	(S.A Khamadi et al., 2008)
	M3_5'- GTCAGCACAGTACAATGCACACATGG-3' (F) M8_5'- CCTTGGATGGGAGGGGCATACATTGC-3' (R)	Nested PCR/ sequencing	
Pol-RT (nt 2480–3180 HXB2); 701bp	RT18_5'-GGAAACCAAAAATGATAGGGGGAATTGGAGG - 3' (F) KS104 _5'- TGACTTGCCCAATTTAGTTTTCCCACTAA-3' (R)	Round 1 PCR	(Lwembe *et al.*, 2007)
	KS101_5'- GTAGGACCTACACCTGTTCAACATAATTGGAAG-3' (F) KS102_5'- CCCATCCAAAGAAATGGAGGAGGTTCTTTCTGATG -3' (R)	Nested PCR/ sequencing	

Key: (F)- Forward Primer; (R)- Reverse Primer

3.6.5 Reverse Transcription Polymerase Chain Reaction (RT-PCR) and Round 1 PCR for RNA Samples

Plasma-derived RNA samples were reverse transcribed using Qiagen OneStep RT-PCR kit (Qiagen, Maryland, USA).

Briefly, each RT-PCR reaction included 1x PCR buffer that contained 1.5 mM $MgCl_2$, 0.4 mM dNTPs, 0.6 µM forward and reverse primers, 1 µl enzyme mix, 5 µl of RNA template, 5 units RNAseOut and nuclease-free water in a final volume of 25 µl. Reverse transcription was done at 50°C for 30 minutes followed immediately by PCR amplification cycles comprising denaturation at 95°C for 15 min and 38 PCR cycles of denaturation at 94°C for 30 s; annealing for 45 s at 58°C for *env* primers or 55°C for *pol-RT* primers and extension at 72°C for 1min. A final extension was done for 10 min at 72°C.

3.6.6 Nested PCR for RNA Samples

Nested PCR was carried out using HotStar Taq polymerase (New England Biolabs, Massachusetts, USA). The nested PCR components included the Taq enzyme (0.625units/25µl), 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM forward and reverse primers, 1x PCR buffer and nuclease-free water in a final reaction volume of 25 µl. Primers used for the nested PCR were M3 and M8 for *env* and KS101 and KS102 for *pol-RT* genes (Table 3.1). The PCR conditions were: Initial denaturation at 95°C for 5 min; 38 cycles of denaturation at 95°C for 30 s, annealing at 56°C for 45 s for both *env* and *pol RT* regions, and extension at 68°C for 45 s. Final extension was at 68°C for 10 min.

3.6.7 DNA PCR

Round 1 DNA PCR was carried out using HotStar Taq polymerase (New England Biolabs, Massachusetts, USA). The PCR components included Taq enzyme (0.625units/25µl), 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM forward and reverse primers (M5 and M10 for *envelope* and RT18 and KS104 for *pol-RT* genes) 1x PCR buffer, 1-3µl of the template and nuclease-free water in a final reaction volume of 25 µl. Initial denaturation was done at 95°C for 5 min; 38 cycles of denaturation at 95°C for 30 s, annealing at 58°C for 45 s for *env* and 55°C for *pol RT* regions, and extension at 68°C for 45 s. Final extension was at 68°C for 10 min.

Round 2 PCR was also carried out using HotStar Taq polymerase (New England Biolabs, Massachusetts, USA). The PCR components included the Taq enzyme (0.625units/25µl), 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM forward and reverse primers (M3 and M8 for *envelope* and KS101 and KS102 for *pol-RT* genes) 1x PCR buffer, 1-3µl of the round 1 amplicon and nuclease-free water in a final reaction volume of 25 µl. Initial denaturation was done at 95°C for 5 min; 38 cycles of denaturation at 95°C for 30 s, annealing at 56°C for 45 s for both *env* and *pol RT* regions, and extension at 68°C for 45 s. Final extension was at 68°C for 10 min.

3.6.8 Gel Electrophoresis

A 1.6% agarose gel matrix was made by mixing 0.64 g of agarose powder (Invitrogen, USA) with 40 ml of 1X Tris EDTA electrophoresis buffer. This was heated in a microwave until the powder had dissolved and then cooled to about 50°C. Two µl of ethidium bromide was then added to the mixture. Ethidium bromide intercalates into the DNA and fluoresces under UV light and hence allows visualization of nucleic acid after electrophoresis.

The solution was poured into a casting tray containing a sample comb and was polymerized for 20 min at room temperature. The sample comb was removed once the gel had polymerized and the gel was inserted into an electrophoresis chamber and covered with buffer solution. The first well was reserved for the DNA ladder. A total volume of 6 µl was used that comprised for the first well, 1 µl DNA ladder, 1 µl loading dye and 4 µl of distilled DNAse free, RNAse free water. The samples were loaded on the remaining wells by mixing 5 µl of the amplicon with 1 µl of loading dye. Since DNA is negatively charged, the gel was placed in the gel tank with the wells facing the cathode so that the current was run towards the anode (positive pole). A current of 100 volts was applied for 30 min.

The larger DNA fragments move more slowly through the agarose gel matrix than the smaller fragments and in this way, the DNA fragments are fractionated, The DNA ladder

helps one to identify the size of the DNA fragments. The gel was visualized under UV light.

3.6.9 Purification and Sequencing

Amplified products were purified using QIAquick PCR purification kit (Qiagen, Maryland USA) so as to remove primers, nucleotides, enzymes and other impurities from DNA samples that could interfere with sequencing. Briefly, binding buffer was added directly to the PCR sample and the mixture applied to the QIAquick spin column.

The binding buffer contains a pH indicator that allows for easy determination of the optimal pH for DNA binding. Nucleic acids were adsorbed to the silica membrane in the high-salt conditions provided by the buffer. Impurities were washed away and pure DNA was eluted with 40 µl of water. Two microlitres was examined for presence of DNA and purity using a nanodrop spectrophotometer using Nanodrop2000 looking out for A_{260}/A_{280} ratios of between 1.7 and 1.9. Due to the fact that the Macrogen sequencing facility in Amsterdam is efficient in giving high quality sequencing results in a short time and at a reasonable price, purified samples were sent to macrogen.

Twenty microlitres was packed for shipping to the Netherlands and sequenced following the macrogen standard sanger sequencing procedure using the ABI3730XL genetic analyzer. The sequencing required of 10 µl of 20ng/µl PCR product and 10 µl of 5pmol/µl nested PCR primers.

3.6.10: HIV-1 Subtyping and Subtype Analysis

Each partial sequence was manually inspected, aligned and edited using Bioedit platform (http://www.mbio.ncsu.edu/bioedit/ from North Carolina State University) (Hall, 1999). Initial pairwise alignment was done to join the reverse and forward sequences. The joined sequences were subsequently subjected to multiple sequence alignment against standard reference sequences from the Los Alamos National Laboratory HIV database using Clustal W (Larkin *et al.*, 2007). At least one reference

sequence for each HIV-1 subtype and the HXB2 reference sequence were included with the query sequences to generate the multiple sequence alignments. A consensus sequence was generated of all the alignments and used together with HIV HXB2 reference sequence, to facilitate further editing. Aligned sequences were transferred to MEGA 6 (Molecular Evolutionary Genetic Analysis 6) sequence analysis platform (Koichiro, Glen, Daniel, Alan, & Sudhir, 2013) for phylogenetic analysis. Phylogenetic trees were generated using the neighbor joining method, the test of phylogeny being the bootstrap method using 1000 replications. The substitution model used was the Kimura 2-parameter method. The generated trees were used to infer and assign HIV-1 subtypes.

For purposes of concordance analysis of different subtyping tools, multiple additional subtyping tools were used and compared with the phylogenetic method. These were: NCBI (http://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi), Stanford for *Pol RT* (Rhee et al., 2003; Shafer, 2006); Geno2Pheno (for *Pol RT*) (Lengauer & Sing, 2006); COMET (Struck, Ternes, Schmit, & Bercoff, 2014); jpHMM – jumping profile hidden markov model (Schultz *et al.*, 2009); REGA (from Rega Institute) (Alcantara et al., 2009; de Oliveira et al., 2005) and SQUEAL (Delport, Poon, Frost, & Pond, 2010).
It was also hypothesized that HIV-1 genotypes could vary genetically across the different blood compartments to reflect the uniqueness of these environments. Therefore, compartmentalization of HIV genetic variants was examined, using the cellular and cell-free compartments as reference. Virus isolates from cellular (DNA) and cell-free (plasma RNA) blood were sequenced on the *pol-RT* and *env* genes, and distribution of genotypes compared between the two compartments. Finally, the geographic distribution of subtypes by county was determined.

3.6.11: Determination of Coreceptor Usage

To determine co-receptor usage, edited sequences that cover the *V3* region were input into the relevant section of the geno2pheno phenotype prediction tool (Lengauer, Sander, Sierra, Thielen, & Kaiser, 2007). Available genotypic tools were developed primarily for detection of subtype B HIV viruses, which are common in Europe and the

U.S.A, and consequently are biased against non-clade B isolates (Garrido *et al.*, 2008). Geno2Pheno was used for determination of tropism in this study because it has previously been found to be sensitive to non clade B X4 variants (Garrido *et al.*, 2008).

To determine virus co-receptor usage or viral tropism, a false positive rate (FPR) cut-off of 10% (FPR10) was first applied, being the standard recommendation of the 'European Consensus Group on clinical management of HIV-1 tropism testing'. Since only a single sequence for each virus isolate was generated, analysis was repeated with FPR set at 20% (FPR20) to increase fidelity for correctly assigning X4-phenotype, and at 5% FPR (FPR5) based on recent findings from subtype C HIV-1 phenotyping that revealed an increased confidence at identifying CXCR4-usage in both CXCR4-exclusive and dual tropic variants (Crous, Krishna, & Travers, 2012). Variations in these strategies were finally compared. In all cases, values below set cut-off were considered to be predictive of an X4 virus.

3.6.12: Determination of envelope protein diversity

The number of potential N-Glycosylation sites (PNGs) was determined from *env* sequences using the Los Alamos HIV database platform (Zhang *et al.*, 2004). Briefly, viral sequences generated were aligned and edited using BioEdit as described in section 3.4.7. The curated sequences were then applied to the Los Alamos database. The nucleic acid sequences were first translated into amino acids sequence using the TRANSLATE tool within the Los Alamos Database (http://www.hiv.lanl.gov/content/sequence/TRANSLATE/translate.html). The generated amino acid sequences were then aligned to HXB2 reference sequence using HIVAlign tool (http://www.hiv.lanl.gov/content/sequence/VIRALIGN/viralign.html) and then applied to the N-Glycosite tool (Zhang *et al.*, 2004) to generate an array of PNG. The number of PNGs was assigned based on the NXT, NXS and the contiguous (NNXT(S)) patterns. PNG site patterns were compared across different viral strains.

3.7 Data Management and Analysis

A computerised data entry system was used to enter and store all subject information using unique subject identity codes. Initial entries were accomplished at the point of data collection by two independent personnel to ensure accuracy and security. Practical information on subjects' locality, unique identification number, occupation and age were coded and double entered into electronic storage servers and computers. Analysis of data was done using the statistical package for social sciences (SPSS).

CHAPTER FOUR

RESULTS

4.1 Patient Characteristics

In this study, 81 patients were recruited from six facilities around the country. From the 81 isolates, 19 amplified for both the *polRT* and *C2V3* genes. Patients were on the following treatment regimen: 1.2% (1/81) of the patients were on ABC+3TC+NVP, 33.3% (27/81) were on AZT+3TC+NVP/EFV, 19.8% (16/81) were on D4T+3TC+NVP/EFV and 24.7% (20/81) were on TDF+3TC+NVP/EFV, with NVP and EFV being used as alternate non-nucleoside reverse transcriptase inhibitors (NNRTI). Eight of the 81 patients (9.9%) were not on ART and ART regimen for 11.1% (9/81) of the patients was not recorded (NR). By gender, 32.1% (26/81) of the patients were males, 64.2% (52/81) were females and 3.7% (3/81) were not recorded. The distribution of the patients by age was 7.4% (6/81) below 25 years old, 33.3% (27/81) between 25 and 35 years, 44.5% (36/81) between 36 and 50 years, 11.1% (9/81) above 50 years and 3.7% (3/81) were not recorded. The treatment and demographic data are presented as part of table 4.1 for *pol RT* sequences and table 4.2 for *C2V3* sequences.

Overall, patients had been on treatment for a median length of 40 months (Range: 0 - 102 months). The CD4 count was obtained from patient records. CD4 T-cell count was less than 350/mm^3 for approximately 50% of the patients, with 27% having a CD4 count greater than 500/mm^3. About 60% of the patients had their viral load below detectable levels (39 HIV-1 RNA copies/ml blood). Of the 100 starting patient samples, DNA was extracted from PBMC from 48 patients while RNA was extracted from plasma from 33 patients for a total of 81 nucleic acid material. Fifty-nine of the 81 patient samples were successfully amplified with envelope-specific primers (Table 4.2), 42 of these being from PBMC derived DNA and 17 from plasma derived RNA. Another 41 of the 81

samples (Table 4.1) (19 derived from PBMC DNA and 22 from plasma derived RNA) were successfully PCR-amplified on the *pol-RT* gene. Nineteen samples amplified for both *pol RT* and *C2V3* regions.

Table 4.1: Patient characteristics based on *pol RT* phylogenetic subtypes

	Subtype (Number of Sequences)							
	Unassigned	**A1**	**A2**	**A2D**	**C**	**D**	**H**	**Total**
Region								
Homabay		1						1
Kisumu		2					1	3
Kajiado		7			1	1		9
Kiambu		4	1					5
Malindi	1	6		1	1	3		12
Naivasha		10				1		11
Total	**1**	**30**	**1**	**1**	**2**	**5**	**1**	**41**
Sex								
Male		11				1		12
Female	1	19	1	1	2	3	1	28
Not recorded						1		1
Total	**1**	**30**	**1**	**1**	**2**	**5**	**1**	**41**
ART Regimen								
ABC/3TC/NVP		1						1
AZT/3TC/EFV				1				1
AZT/3TC/NVP		7				2		9
D4T/3TC/NVP	1	6	1					8
None		4				1		5
NR		4				1		5
TDF/3TC/EFV		4			2	1		7
TDF/3TC/NVP		4					1	5
Total	**1**	**30**	**1**	**1**	**2**	**5**	**1**	**41**
Age								
Below 25		2					1	3
25-35		11	1	1	1	4		18
36-50	1	14			1			16
Above 50		3						3
Not recorded						1		1
Total	**1**	**30**	**1**	**1**	**2**	**5**	**1**	**41**

This table shows the HIV subtypes from 41 patients recruited from Kisumu, Kiambu, Kitengela, Malindi, Homa-Bay and Naivasha, derived from the *pol-RT* gene.

Table 4.2: Patient characteristics based on *C2V3* phylogenetic subtypes

		Subtype (No. of Sequences)					
		A1	**A2**	**C**	**D**	**N.D**	**Total**
Region							
Homabay		2			1		3
Kisumu		10	1	2	2	1	16
Kajiado		11		2		1	14
Kiambu		2				2	4
Malindi		6	1				7
Nakuru		10	2	1	2		15
	Total	**41**	**4**	**5**	**5**	**4**	**59**
Sex							
Female		25	1	3	4	3	36
Male		14	3	2	1	1	21
Not recorded		2					2
	Total	**41**	**4**	**5**	**5**	**4**	**59**
ART Regimen							
ABC/3TC/NVP							
AZT/3TC/EFV		3	1				4
AZT/3TC/NVP		11	1	2	3	2	19
D4T/3TC/NVP		6	1	1	2	1	11
None		5				1	6
Not recorded		6					6
TDF/3TC/EFV		2		1			3
TDF/3TC/NVP		8	1	1			10
	Total	**41**	**4**	**5**	**5**	**4**	**59**
Age							
Below 25		5					5
25-35		15	1	1	1	2	20
36-50		14	2	4	3	1	24
Above 50		5	1		1	1	8
Not recorded		2					2
	Total	41	4	5	5	4	**59**

This table shows HIV-1 genotypes based on *envelope* gene as obtained using Phylogeny for 59 patients who were recruited from Kisumu, Kiambu, Malindi, Homabay and Naivasha and Kitengela. Key: N.D, Not determined (these sequences were found to be too short by the publisher (AIDS Research and Human Retroviruses) and were not included in the Phylogenetic Analysis).

4.2 HIV-1 subtyping

4.2.1 Phylogenetic Analysis

4.2.1.1. Phylogenetic analysis for the envelope *C2V3* Sequences

From the 59 sequences, four were deemed to be too short by the publisher (AIDS Research and Human Retroviruses) and were not included in the phylogenetic analysis. Of the remaining 55 sequences (Figure 4.1), 41/55 (74.5%) of sequences clustered with subtype A1 and therefore were determined to be subtype A1 by phylogeny. Four of 55 (7.3%) clustered with subtype A2, 9.1% (5/55) clustered with subtype C and 9.1% (5/55) clustered with subtype D reference sequences, and were assigned corresponding subtypes.

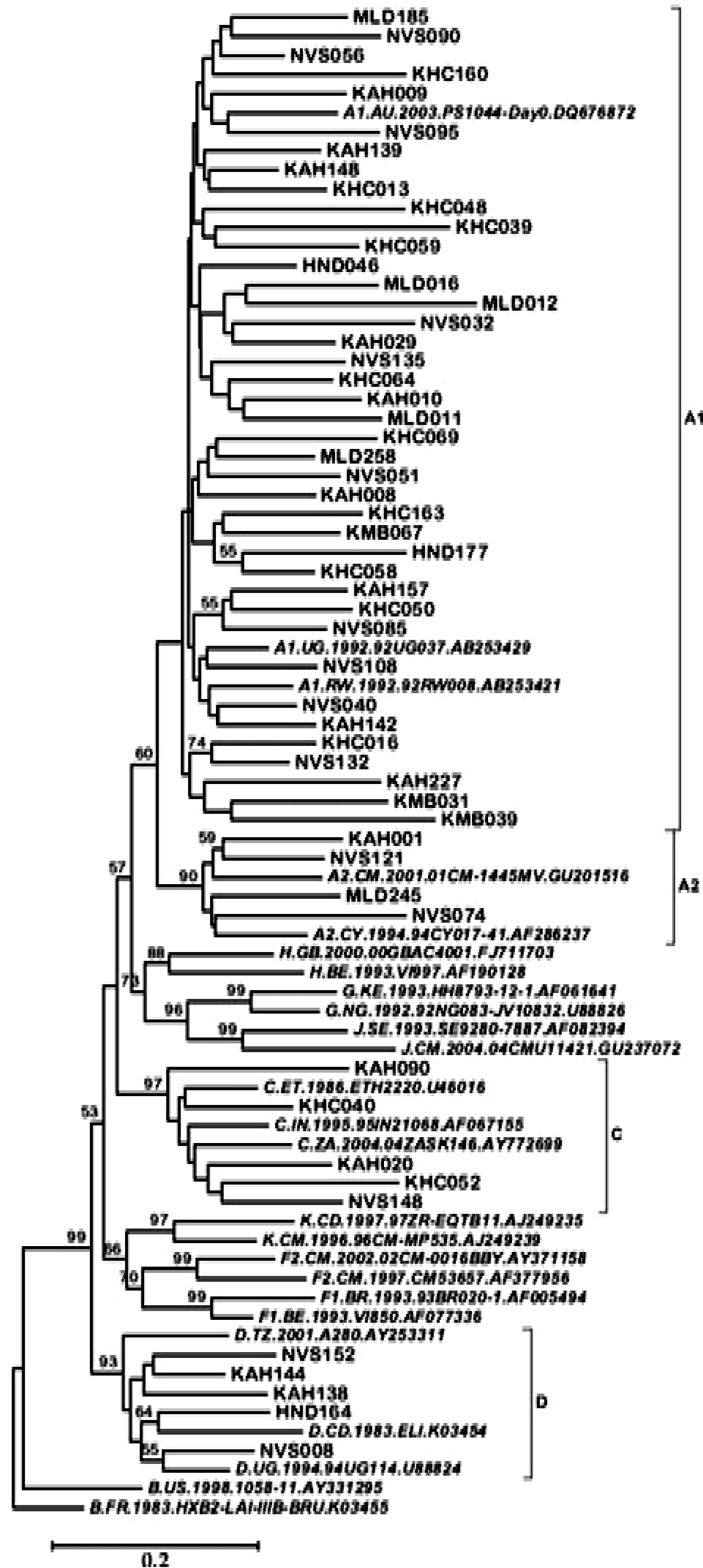


Figure 4.1: Phylogenetic Relationships of HIV-1 envelope *C2V3 gene*

The evolutionary distances of the various virus isolates were inferred using the Neighbour-Joining method. The evolutionary distances were computed using the Kimura 2-parameter method (Kimura, 1980). Node statistics corresponding to less than 50% bootstrap values were hidden to minimize clutter. Reference sequences are in italics

4.2.1.2 Phylogenetic Analysis of the *Pol RT* sequences

From the phylogenetic approach (Figure 4.2) most (30/41; 73.2%) of the sequences clustered with subtype A1 reference strains. One isolate, MLD245, was orthologous to A2D reference sequence from Kisii and was a likely recombinant of subtypes A and D viruses. One isolate clustered with A2 references sequences. Overall, 32/41 (78%) of all the isolates clustered with subtype A reference strains and were assigned as A viruses on the *pol-RT* gene by phylogeny. Another isolate, MLD548 did not cluster with any specific reference sequence and was unassigned by this method. KAH004 clustered together with subtype H reference sequence, which was unusual, and could be a recombinant strain. Subtype C accounted for 7.3% of the sequences (3/41) while Subtype D accounted for 9.8% of the sequences (4/41).

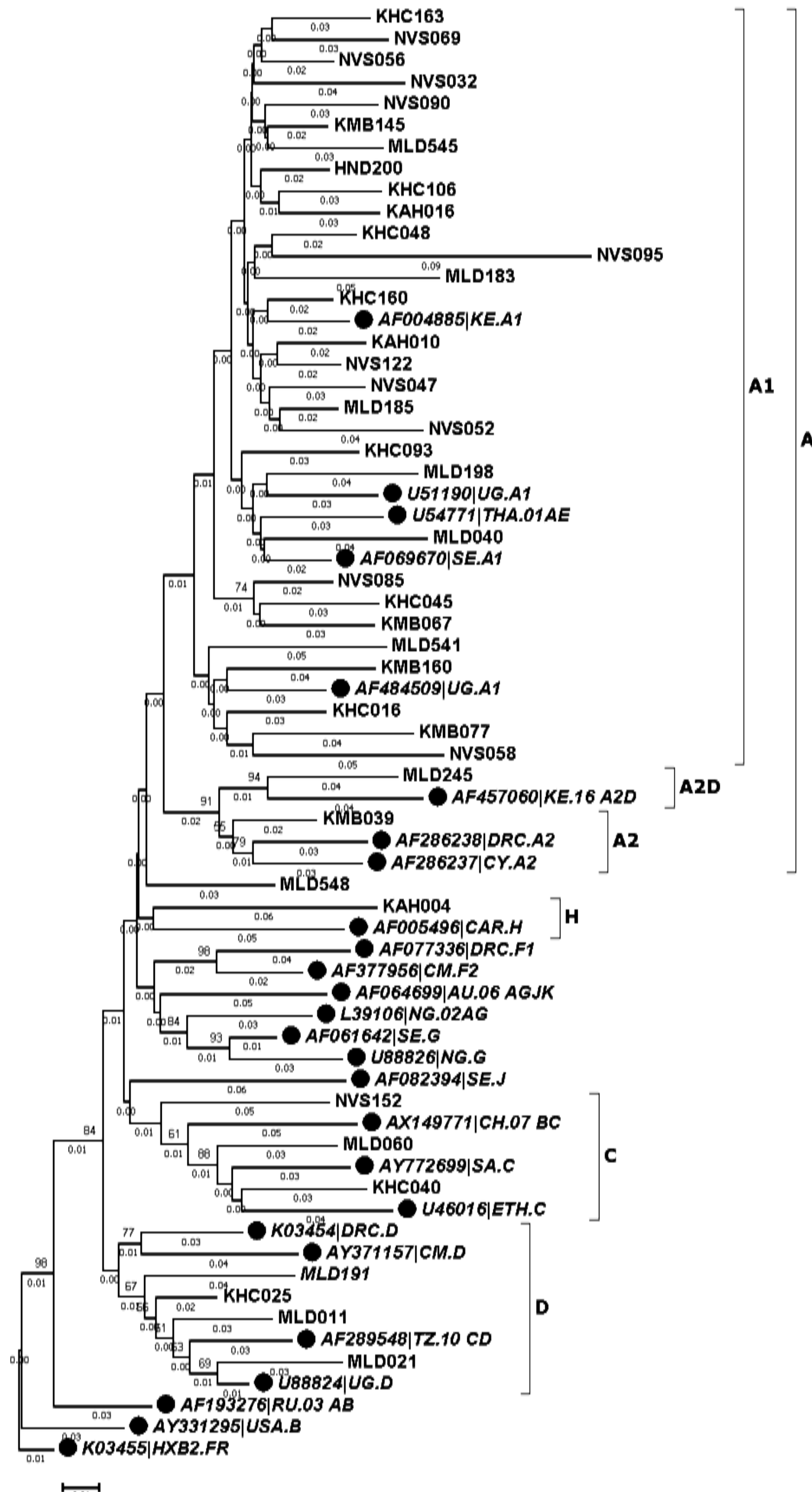


Figure 4.2: Phylogenetic Relationships of HIV-1 envelope *pol RT gene* isolates

The evolutionary history was inferred using the Neighbor-Joining method. The evolutionary distances were computed using the Kimura 2-parameter method (Kimura, 1980). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed and all positions with less than 50% site coverage were eliminated. Node statistics corresponding to less than 50% bootstrap values were hidden to minimize clutter. Reference sequences are in bold italics and filled circle nodes.

4.2.2 Comparative sensitivities of common genotyping tools used for HIV-1 subtyping

Multiple HIV genotyping tools are available for HIV subtype determination. The comparative sensitivities of multiple platforms were assessed, so as to correctly assign specific subtype to any HIV isolate. Four platforms were used for *envelope* gene and seven platforms for *pol-RT* gene for subtype comparisons. For empirical purposes, sensitivity was measured against phylogenetic method as a reference (and currently the most robust) genotyping method. Of the 59 *env C2V3* sequences, one sequence was not long enough to be determined by all the methods and was excluded from the analysis.

For this purpose, 58 *C2V3* and 41 *Pol RT* sequences were analysed using the respective platforms and the results compared for concordance with subtype assignment based on phylogenetic method. Concordance was described respectively as being either complete, partial or none concordant if the result obtained using the automated tool was similar, partly similar or completely different from that obtained by phylogeny respectively (Figure 4.2). Of the 58 C2V3 sequences, NCBI reported the proportions of subtypes A1, A2, C, D and recombinants as 23.7%, 0%, 5.1%, 0% and 71.2% respectively compared against 73.2%, 4.9%, 9.1%, 9.1% and 0% shown by phylogeny (Table 4.3). In terms of concordance, these translated into 29.3% (17/58), 65.5% (38/58) and 5.2% (3/58) complete, partial and none concordance respectively. Using the COMET platform, complete, partial and none concordance was found to be 79.3%, 13.8%, and 6.9% respectively. REGA on the other hand reported complete concordance for 41/58 of the samples (70.7%) and 17/58 (29.3%) as none concordant; with HIV1-A, C, D and unassigned subtypes being 57.6%, 6.8%, 8.5%, 27.1% respectively (Table 4.3). For JPHMM subtyping tool, complete and partial concordance was 84.5% (49/58) and 15.5% respectively; with HIV1-A1, A2, C, D and recombinants being 61.0%, 6.8%, 8.5%, 8.5% and 15.2% respectively. Through this comparative subtyping analysis approach using different tools, JPHMM was found to be most in agreement with the standard phylogenetic method in assigning HIV subtypes.

Table 4.3: Comparative concordance of various subtyping tools for assignment of HIV-1 *env C2V3* subtypes

%, (N)			
NCBI	**COMET**	**REGA**	**JPHMM**
Y- 29.3, (17)	Y- 79.3, (46)	Y- 70.7, (41)	Y- 84.5, (49)
P- 65.5, (38)	P – 13.8, (8)	P – 0 (0)	P- 15.5, (9)
N- 5.2, (3)	N - 6.9, (4)	N- 29.3, (17)	

Key: Y= Complete concordance; P= Partial Concordance; N= Not concordant. Concordance is based on phylogenetic method of subtyping as reference or standard method

Similar analytical approach was adopted for the *pol-RT* gene. For the 41 gene-specific sequences, complete, partial or none concordance was respectively, 17.1%, 75.6% and 7.3% for the NCBI tool (Table 4.4); 46.3%, 36.6%, and 17.1% for Stanford; 75.6%, 2.4%, 22% for REGA;65.9%, 19.5% and 14.6% for JPHMM; 82.9%, 9.8% and 7.3% for Geno2Pheno;75.6%, 19.5% and 4.9% for COMET and 56.1%, 29.3% and 14.6% for SQUEAL (Table 4.4). NCBI reported more CRFs than all other subtyping tools while REGA failed to identify recombinants (Table 4.4), with most unidentified sequences reported as complex (NA). Stanford reported higher percentages of CRF_01AE, majority of which were subtype A by other methods, and also assigned a pure subtype B that was an A2B recombinant by JPHMM (Appendix 4). In order to have a standard platform for the assigning of HIV-1 subtypes, JPHMM was used for subsequent analysis of both the *pol RT* and *env* gene.

Table 4.4: Comparison of concordance for various subtyping tools for the *pol RT* gene

NCBI N, (%)	STANFORD N, (%)	REGA N, (%)	JPHMM N, (%)	G2P N, (%)	COMET N, (%)	SQUEAL N, (%)
Y-7, (17.1)	Y-19, (46.3)	Y-31, (75.6)	Y-27, (65.9)	Y-34, (82.9)	Y-31, (75.6)	Y-23, (56.1)
P-31, (75.6)	P-15, (36.6)	P-1, (2.4)	P-8, (19.5)	P-4, (9.8)	P-8, (19.5)	P-12, (29.3)
N-3, (7.3)	N-7, (17.1%)	N-9, (22)	N-6, (14.6)	N-3, (7.3)	N-2, (4.9)	N-6, (14.6)

Key: Y= Complete concordance; P= Partial Concordance; N= None concordance

4.2.3 Genetic Heterogeneity of HIV-1 *Pol RT* gene from different blood compartments

In total, 19 sequences were generated from cellular DNA samples that had undetectable viral load and up to a maximum of 1000 HIV-1 RNA copies/ml blood. Subtyping was done using JPHMM. Out of these, 73.7% (14/19) were subtype A1, 10.5% (2/19) were subtype D, 5.3% (1/19) was subtype C and 10.5% (2/19) were recombinants (Table 4.5). A total of 22 patient RNA samples were sequenced from plasma samples with viral load above 1000 RNA copies/ml blood. Ten out of 22 (45.5%) were found to be subtype A1, 18.2% (4/22) subtype D, 4.5% (1/22) subtype C and 31.8% (7/22) were recombinants (Table 4.5).

The recombinants obtained from cell-free plasma compartment were A1H (n=1), A1D (n=1), A1C (n=2,), A2B (n=1), A1K (n=1) and DKD (n=1) and those obtained from cellular blood compartment were A1A2 (n=1), A1C (n=1). HIV *pol RT* gene was more diverse in the cell-free compartment than in the cellular compartment. Overall, there were more recombinants in the cell free compartment (31.8%) compared to 10.5% in the cellular compartment (Table 4.5).

The trend in distribution showed fewer subtype A strains (45.5%) in cell-free compartment compared to cellular compartment which had 73.7% of subtype A isolates (Table 4.5). There was a larger percentage of Subtype D in the cell-free compartment (18.2%) compared to the cellular compartment (10.2%) (Table 4.5).

Table 4.5: **Distribution and compartmentalization of HIV-1 *Pol-RT* subtypes**

Subtype	Frequency in cellular blood DNA, N (%)	Frequency cell-free blood RNA, N (%)
A (A1)	14 (73.7)	10 (45.5)
D	2 (10.5)	4 (18.2)
C	1 (5.3)	1 (4.5)
All recombinants	*2 (10.5)	**7 (31.8)
Total	19 (100)	22 (100)

Key:*A1C, n=1; A1A2, n=1. **A1C, n=2; A1H, n=1; A2B, n=1; A1K, n=1; DKD, n=1; A1D, n=1.

4.2.4 Genetic heterogeneity of HIV-1 *C2V3 envelope gene* from different blood compartments

Similar analyses were conducted for the 59 *envelope* sequence isolates to assess compartmentalization of viral genetic variants (Table 4.6). Out of the 42 isolates from the cellular compartment, 57.1% (24/42) were subtype A1, 4.8% (2/42) were subtype A2, 11.9% (5/42) were subtype D and 11.9% (5/42) were subtype C (Table 4.6). Recombinant viruses accounted for 14.3% of the envelope DNA-derived virus strains with 7.1% (3/42) being A1A2, 4.8% or 2/42 being A1H, and 2.4% (1/42) being A1D. From cell-free plasma compartment, 70.6% (12/17) were pure envelope subtype A1 and 11.8% subtype A2 (2/17) (Table 4.6). The recombinants obtained from cell free compartment accounted for 17.6% (3/17) and included A1D (2/17) and A1A2D (1/17)) (Table 4.6. No pure subtypes D or C were identified based on envelope C2V3 gene from cell free RNA.

Overall, there was a higher percentage of subtype A in the cell free compartment (82.4%) compared to the cellular compartment (61.9%) (Table 4.6). There was a small difference in the percentage of recombinants in the cellular compartment (14.3%) compared to the cell-free compartment (17.6%) (Table 4.6).

Table 4.6: Distribution and compartmentalization of *envelope C2V3* subtypes

Subtype	Frequency in cellular blood DNA, N (%)	Frequency cell-free blood RNA, N (%)
A (A1,A2)	26 (61.9)	14 (82.4%)
D	5 (11.9)	
C	5 (11.9)	
All recombinants	*6 (14.3)	**3 (17.6)
Total	42 (100)	17 (100)

Key: *A1A2, n=3; A1H, n= 2; A1D, n= 1. **A1D, n= 2; A1A2D, n= 1

4.2.5 Comparative distribution of HIV genetic variants between blood compartments.

From the analyses reported in sections 4.2.3 and 4.2.4, the cellular compartment had more of the *pol-RT* subtype A (73.7%) than envelope subtype A (61.9%) - Table 4.7. The pattern was reversed for the plasma compartment which had more of the envelope *C2V3* (82.4%) than *Pol-RT* (45.5%) subtype A viruses (Table 4.7). Of the 61 sequences from the cellular compartment 40 (65.6%) were subtype A (A1, 62.3%; A2, 3.3%), 9.8% subtype C, 11.5% subtype D and 13.1% recombinants as compared to 61.5% subtype A (A1, 56.4%; A2, 5.1%), 2.6% of C, 10.3% of D and 25.6% of recombinants in the cell-free compartment (Table 4.7). No envelope-based subtypes C and D were identified from plasma compartment. The plasma compartment carried more recombinant *pol-RT* (31.8%) and envelope *C2V3* (17.6%) viruses than the proportion of recombinants found in the cellular compartment of similar viral gene segments (Table 4.7). Majority of the

subtype C isolates were dominantly identified in envelope sequences and from the cellular compartment (Table 4.7).

Table 4.7 Comparative frequency of subtypes by genetic region and compartment

	Cellular DNA, N (%)			Cell-free blood RNA, N (%)		
Subtype	*Pol-RT*	Env *C2V3*	**Total N (%)**	*Pol-RT*	Env *C2V3*	**Total N(%)**
A	14 (73.7)	26 (61.9)	**40 (65.6)**	10 (45.5)	14 (82.4)	**24 (61.5)**
D	2 (10.5)	5 (11.9)	**7 (11.5)**	4 (18.2)		**4 (10.3)**
C	1 (5.3)	5 (11.9)	**6 (9.8)**	1 (4.5)		**1 (2.6)**
All recombinants	2 (10.5)	6 (14.3)	**8 (13.1)**	7 (31.8)	3 (17.6)	**10 (25.6)**
Total	19 (100)	42 (100)	**61(100)**	22 (100)	17 (100)	**39 (100)**

4.2.6 Subtype Distribution by County

HIV has been shown to be distributed variably across different geographical settings (see Table 2.1). This section describes the distribution of the 81 unique virus isolates across the six Counties of Kenya, which data is part of two publications (Kitawi et al., 2015; Nzomo et al., 2015). Of all samples qualifying for this analysis, 32% were from Nyanza (Kisumu and Homa Bay Counties), 23.7% from Kajiado County, (Kitengela), 6.8% from Kiambu, 11.9% from Malindi and 25.4% from Nakuru County (Naivasha).

Because of differences in sample size between counties, subtype variability is assessed within each county and not between county to avoid bias. Table 4.8 summarizes the distribution of HIV *C2V3* genotypes within each of the 6 Counties. When the JPHMM subtyping tool was used, 67.8% of all subtypes were HIV-1A, 8.5% were subtype C, 8.5% were subtype D and 15.2% were recombinants (Table 4.8).

In Nyanza, of all the circulating genotypes, 68.4% were HIV-1A, 10.5% HIV-1C, 15.8% HIV-1D while recombinants were 5.3%. From Kitengela in Kajiado county, 57.1% were subtype A, 14.3% subtype C, 28.6% were recombinants while there was no subtype D. Kiambu had too few samples (n=4) to be representative (Table 4.8). However, 75% of these were subtype A and 25% were recombinants. Malindi in Kilifi County also had very few samples with 71.4% being subtype A and 28.6% being recombinants. A larger sample size from these counties would probably have changed the pattern seen. Naivasha in Nakuru County had most of the samples being subtype A (73.3%), 6.7% subtype C, 6.7% recombinants and 13.3% subtype D (Table 4.8).

Although sample distribution was not even across the sites, the data suggests that Kitengela in Kajiado County and Malindi in Kilifi County may carry the highest prevalence of recombinant envelope viruses, while Kiambu and Nakuru suggest a similar trend that may need corroborating with larger sample size.

Table 4.8: Variability of *C2V3* HIV genotypes within a given geographic locations across Kenya

	Number (%) in the region				
	A (A1&A2)	**C**	**D**	**Recombinants**	**Total**
Nyanza*	13 (68.4)	2 (10.5)	3 (15.8)	1 (5.3)	19 (100)
Kitengela	8 (57.1)	2 (14.3)		4 (28.6)	14 (100)
Kiambu	3 (75)			1 (25)	4 (100)
Malindi	5 (71.4)			2 (28.6)	7 (100)
Naivasha	11 (73.3)	1 (6.7)	2 (13.3)	1 (6.7)	15 (100)
Total	40 (67.8)	5 (8.5)	5 (8.5)	9 (15.2)	59 (100)

Key: * Ahero and Ndhiwa

The distribution of *pol-RT* subtypes was similarly assessed within Counties. Of all qualifying samples for the *pol RT* region, 9.7% of the samples were from Nyanza, 22% from Kitengela in Kajiado county, 12.2% were from Kiambu, 29.3% from Malindi in Kilifi county and 26.8% from Nakuru Naivasha in Nakuru county. Table 4.9 shows the

variability of subtypes within the counties. The four isolates from Nyanza, were evenly distributed by genotype at 50% subtype A and 50% recombinant strains. Kitengela in Kajiado county had 77.8% subtype A, 11.1% subtype C and 11.1% subtype D respectively (Table 4.9). No reverse transcriptase (*RT*) recombinants were found among the 9 isolates from Kajiado. Malindi in Kilifi county had 25% subtype A1 and D, 8.3% subtype C and 41.7% *RT* recombinants. Naivasha in Nakuru county had majority of the samples (81.8%) being subtype A1, 9.1% being subtype D and 9.1% being recombinants (Table 4.9). The overall pattern was that there were more recombinant isolates for the *pol RT* than *envelope* genes.

Table 4.9: Variability of *Pol RT* HIV genotypes within a given geographic locations across Kenya

	Number (%) in the region				
	A1	**C**	**D**	**Recombinants**	**Total**
Nyanza*	2 (50)			2 (50)	4 (100)
Kitengela	7 (77.8)	1 (11.1)	1(11.1)		9 (100)
Kiambu	3 (60)		1 (20)	1 (20)	5 (100)
Malindi	3 (25)	1 (8.3)	3 (25)	5 (41.7)	12 (100)
Naivasha	9 (81.8)		1 (9.1)	1 (9.1)	11(100)
Total	24 (58.5)	2 (4.9)	6 (14.6)	9 (22)	41(100)

*Sequences from Ahero and Ndhiwa

4.3 Co-receptor Usage

Two of the 59 sequences did not span the *V3* region and were thus deemed too short for coreceptor analysis. Four of the 59 (including 2 with short V3 region), had C2V3 region too short for phylogenetic subtyping, but were still able to be subtyped based on JPHMM and other online bioinformatics platforms. Of the 57 sequences with complete V3 region, 34 were subtype A1, 9 were recombinants of A1, 4 were A2, 5 were subtype C and 5 were subtype D.

Three different FPR cut-offs (5%, 10% and 20%) were applied as described under methods. Forty three of the 57 sequences (75%) were CCR5 (R5) tropic and 14 (25%) were CXCR4 (X4) tropic viruses at FPR10. No dual tropic (R5X4) virus isolates were found using the Geno2Pheno phenotyping algorithms. For all three FPR cut-off algorithms, there was a significant relationship between viral subtype and viral tropism (χ^2 p=0.037; p=0.016 and p=0.005 for FPR5, FPR10 and FPR20 respectively).

Clustered by viral subtype, approximately 80% of subtype A1 sequences were R5 tropic, against approximately 20% X4 viruses. Similarly, a larger proportion (approximately 75%) of A1 recombinant viruses were R5 tropic. Relative to subtype A1 viruses, there were however, comparably fewer subtype A2 viruses (n=4), 50% of them being either of X4 or R5 phenotype. None of the subtype C viruses were of X4 phenotype using all three FPR algorithms, compared to nearly all (80% at FPR10 and 100% at FPR20) of subtype D viruses being found to be X4-tropic. These analyses suggested a trend whereby more of the pure subtype A1 were likely to be assigned as R5 tropic at lower FPR than were assigned as X4 viruses, while more of the pure subtype D isolates tended to be assigned as X4 tropic at higher FPR. Tropism of pure subtype C and A2 viruses appeared unaffected across FPR algorithms, while that of recombinant strains either varied minimally or not at all. Table 4.10 shows the distribution of viral tropism across HIV subtype variants at the 3 FPR cut-offs.

Table 4.10: The distribution by number and (proportions), of viral tropism across HIV subtype variants.

	Viral Tropism at FPR cut-off (N, %)						
	FPR, 5%		FPR (10%)		FPR (20%)		
HIV-1 Subtype	**CCR5**	**CXCR4**	**CCR5**	**CXCR4**	**CCR5**	**CXCR4**	**TOTAL**
A1	30 (88.2%)	4 (11.8%)	28 (82.4)	6 (17.6)	26 (76.5)	8 (23.5)	34 (100)
*A1r.	7 (77.8)	2 (22.2)	7 (77.8)	2 (22.2)	6 (66.7)	3 (33.3)	9 (100)
A2	2 (50)	2 (50)	2 (50)	2 (50)	2 (50)	2 (50)	4 (100)
C	5 (100)	0 (0)	5 (100)	0 (0)	5 (100)	0 (0)	5 (100)
D	2 (40)	3 (60)	1 (20)	4 (80)	0 (0)	5 (100)	5 (100)
Total	46(81)	11 (19)	43 (75)	14 (25)	39 (68)	18 (32)	57 (100)

4.4 Patterns of Potential N-Linked Glycosylation Sites (PNGs)

The subtype of 55 out of 59 sequences could be defined phylogenetically (Kitawi *et al.*, 2015). These were the sequences that were examined for the number and pattern of N-linked glycosylation sites (PNGs) as described under methods. In terms of specific amino acids (A.A), the largest number of virus isolates were glycosylated at asparagine (N) position 277 (N277, 73%) and at N302 (71%) (Table 4.11). The proportion of sequences with PNGs at specific amino acid positions was 80% (44/55) for N262 and 98.2% (54/55) for N277. The rest were: N296 (38/55, 69.1%), N302 (54/55, 98.2%), N337 33/55 (60%), N345 (40/55, 72.7%), N366 (41/55, 74.5%), N399 (43/55, 78.2%) and N408 (40/55, 72.7%). These patterns remained largely true when data was further disaggregated by nucleic acid source as either being derived from DNA or RNA material and are presented in Table 4.11.

Nucleic Acid (NA) source material was used as a proxy for viral compartmentalization in blood, with plasma-derived RNA isolates being representative of extracellular blood

compartment and PBMC-derived DNA isolates representing cellular compartment. For the 55 sequences, the total number of PNGs considering all possible patterns was 615 (mean, 11.18, standard error –SE -, 0.274). Of these, 403 were of NXT type (mean, 7.33, SE 0.239) while 205 were of NXS type (mean, 3.73, SE 0.167) with only 7 being of the contiguous NNXS(T) pattern (mean, 0.13, SE 0.052). Table 4.12 summarizes the mean PNGs of isolates clustered according to subtype, NA source and viral tropism.

Table 4.11: The distribution and clustering of specific amino acid PNG sites according to viral tropism and NA source material

	Number of isolates; % possessing amino acid PNG at specified site									
Tropism	**N262**	**N277**	**N296**	**N302**	**N337**	**N345**	**N366**	**N399**	**N408**	**Total**
DNA source										
CCR5	22; 66.7	29; 72.5	20; 71.4	28; 71.8	18; 72	22; 73.3	26; 81.3	24; 68.6	24; 72.7	***29; 72.5***
CXCR4	11; 33.3	11; 27.5	8; 28.6	11; 28.2	7; 28	8; 26.7	6; 28.7	11; 31.4	9; 27.3	***11; 27.5***
All DNA	***33; 100***	***40; 100***	***28; 100***	***39; 100***	***25; 100***	***30; 100***	***32; 100***	***35; 100***	***33; 100***	**40; 72.7**
RNA source										
CCR5	9; 81.8	12; 85.7	9; 90	13; 86.7	7; 87.5	8; 80	7; 77.8	7; 87.5	5; 71.4	***13; 86.7***
CXCR4	2; 18.2	2; 14.3	1; 10	2; 13.3	1; 12,5	2; 20	2; 22.2	1; 12.5	2; 28.6	***2; 23.3***
All RNA	***11; 100***	***14; 100***	***10; 100***	***15; 100***	***8; 100***	***10; 100***	***9; 100***	***8; 100***	***7; 100***	**15; 27.3**
Total										
CCR5	***31; 70.5***	***41; 75.9***	***29; 76.3***	***41; 75.9***	***25; 75.8***	***30; 75***	***33; 80.5***	***31; 72.1***	***29; 72.5***	**42; 76.4**
CXCR4	***13; 29.5***	***13; 24.1***	***9; 23.7***	***13; 24.1***	***8; 24.2***	***10; 25***	***8; 19.5***	***12; 27.9***	***11; 27.5***	**13; 23.6**
All	**44; 80**	**54; 98.2**	**38; 69.1**	**54; 98.2**	**33; 60**	**40; 72.7**	**41; 74.5**	**43; 78.2**	**40; 72.7**	**55; 100%**

Key: NA – Nucleic Acid. R5 or X4 tropism on this analysis is based on false positive rate of 10%

Table 4.12: Mean PNGs for 55 isolates by subtype, nucleic acid source and tropism

		Mean PNGs (Standard Error)			
		All Patterns	NXT Pattern	NXS Pattern	NNXS(T) Pattern
Subtype	**A1**	11.18 (0.386)	7.24 (0.346)	3.79 (0.214)	0.15 (0.075)
	A1 R	10.57 (0.685)	6.71 (0.36)	3.71 (0.474)	0.14 (0.147)
	A2	10.75 (0.629)	7.75 (0.854)	3.00 (0.913)	
	C	11.2 (0.663)	7.4 (0.678)	3.80 (0.583)	
	D	12.4 (0.748)	8.4 (0.40)	3.80 (0.374)	0.2 (0.2)
Nucleic acid source	**DNA**	11.6 (0.288)	7.68 (0.252)	3.85 (0.198)	0.08 (0.042)
	RNA	10.07 (0.565)	6.40 (0.505)	3.40 (0.306)	0.27 (0.153)
Tropism	**CCR5**	11.02 (0.331)	7.21 (0.282)	3.64 (0.192)	0.17 (0.067)
	CXCR4	11.69 (0.429)	7.69 (0.444)	4.00 (0.340)	
Total		**11.18 (0.274)**	**7.33 (0.239)**	**3.73 (0.167)**	**0.13 (0.052)**

Clustered by blood compartment, there were 12 PNGs for each viral isolate derived from DNA (cellular) compartment compared to 10 PNGs from each isolate of RNA (extracellular) material. DNA derived isolates were still the most glycosylated at NXT sites. Analysis of variance (ANOVA) was next conducted to compare mean values of the specific PNG patterns across NA source. This analysis revealed a significantly different pattern and density of PNG between cellular and extracellular viral isolates for the NXT glycosylation pattern ($p=0.016$), and for all the patterns combined ($p=0.011$) - see Figure 4.3 below. Moreover considering specific amino acid PNG sites, there was

still a significant difference in glycosylation between cellular and extracellular isolates at A.A positions N399 (p=0.006) and N408 (p=0.007).

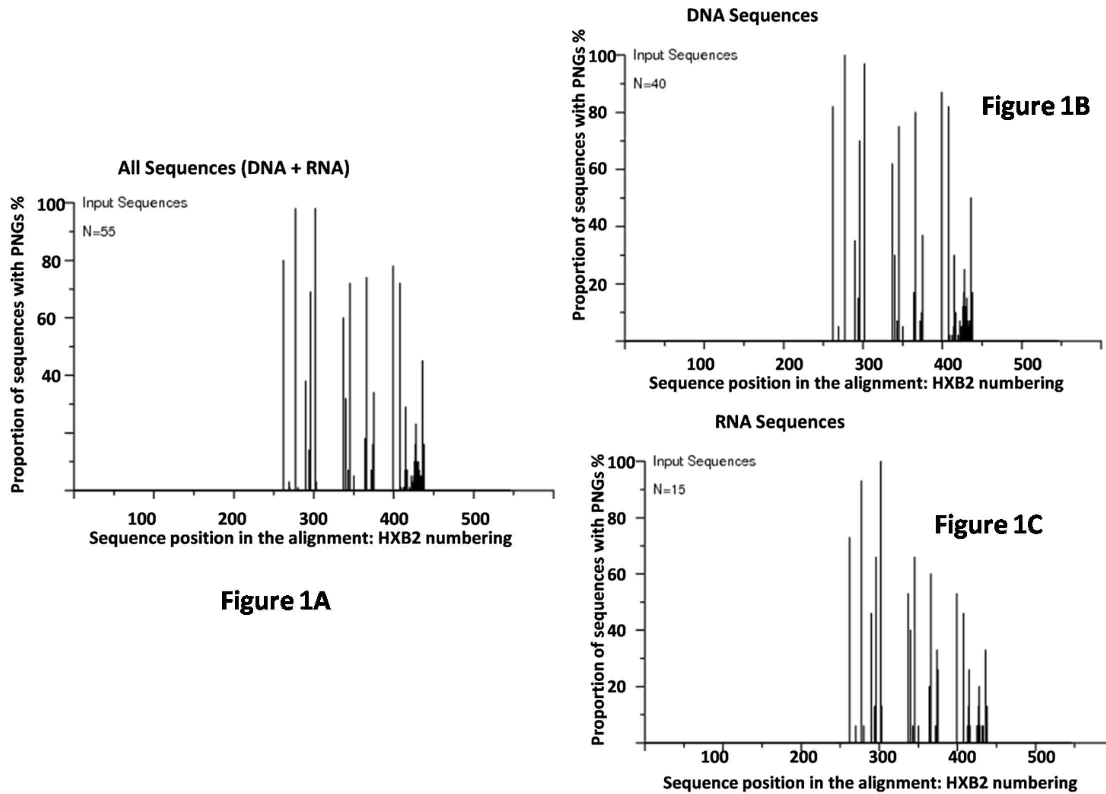


Figure 4.3: Proportion of sequences containing PNGs at specific amino acid positions. Figure 1A: all isolates; Figure 1B: isolates derived from DNA; Figure 1C: isolates derived from RNA.

When all the possible PNG patterns were considered, the mean PNGs for A1 was 11.18 (SE 0.386), A1 recombinants was 10.57 (SE 0.685), A2 was 10.75 (SE 0.629), C was 11.20 (SE 0.663) and D was 12.40 (SE 0.748). Thus subtype D isolates had the largest number of PNGs per isolate while recombinant strains had the least. Similarly, subtype D was the most abundantly glycosylated at NXT sites (mean of 7.33, SE, 0.239), with A1 strains being least glycosylated at these sites. Glycosylation density was comparable at NXS sites for subtypes A1 (mean 3.79; SE 0.214), subtype C (mean, 3.80; SE 0.583) and subtype D (mean, 3.80; SE 0.374) but much less for A1 recombinants (mean

3.00,SE 0.601) and subtype A2 (mean, 3.00; SE 0.913). Glycosylation at the contiguous NNXS(T) sites was the least common and was observed at a mean occurance of 0.15 per isolate for A1 viruses, 0.14 (SE 0.14) for A1 recombinants, and 0.2 (SE 0.2) for D viruses. None of the subtype C or A2 isolates were glycosylated at the NNXS(T) site. Table 4.12 shows the details for the average number of pattern-specific PNGs disaggregated by viral subtype, tropism and source material. An ANOVA test to compare mean PNGs for the different subtypes at specific amino acid position yielded significant results for position N296 (p=0.021), N302 (p=0.034) and N366 (p=0.016).

CHAPTER 5

DISCUSSION

5.1 Patient Characteristics

This study focused on investigating distribution and genetic diversity of HIV-1 in Homa-Bay, Kisumu, Kiambu, Nakuru, Kajiado and Kilifi counties of Kenya. Patients were recruited from six facilities located in these counties and sequence information was obtained from 81 patients. By gender, 32.1% (26/81) of the patients were males, 64.2% (52/81) were females and 3.7% (3/81) were not recorded. The Ministry of Health Kenya County HIV profiles showed a higher prevalence of HIV in Kenya among women (7.6%) than men (5.6%) (NACC., 2014). The distribution of the patients by age was 7.4% (6/81) below 25 years old, 33.3% (27/81) between 25 and 35 years, 44.5% (36/81) between 36 and 50 years, 11.1% (9/81) above 50 years and 3.7% (3/81) were not recorded.

The patients had been on treatment for a median length of 40 months. CD4 T-cell count was less than 350 cells/mm^3 for approximately 50% of the patients, with 27% having a CD4 count greater than 500. About 60% of the patients had their viral load (VL) below detectable levels (39 HIV-1 RNA copies/ml blood). The Guidelines for antiretroviral therapy in Kenya indicate that the viral load should be below detectable levels within six months of ART (NASCOP, 2011). Although there is a sharp decline in viral load within 6 months of treatment, the CD4 count increases slowly over time (NASCOP, 2011) and this would explain why 50% of the patients had their CD4 count below 350 despite a higher percentage (60%) having their viral load below detectable levels. HIV infection affects more women (59.1%) than men (40.9%) in Kenya (NACC & NASCOP, 2012), and this could be one reason why more women were enrolled in the study than men. About 9.9% of the patients were not on ART. Those on ART, were receiving the following treatment regimens (percentages given based on total number of patients): 1.2% on ABC+3TC+NVP, 33.3% on AZT+3TC+NVP/EFV, 19.8% on

D4T+3TC+NVP/EFV and 24.7% on TDF+3TC+NVP/EFV, with NVP and EFV being used as alternate non-nucleoside reverse transcriptase inhibitors (NNRTI). ART regimen for 11.1% of the patients was not recorded (NR). The recommended first line ART in Kenya is AZT+3TC+NVP/EFV or TDF+3TC+NVP/EFV(NASCOP, 2011). Although the recommended treatment was the most common among patients receiving ART, the continued use of D4T in ART was contrary to existing WHO recommendation for D4T phase-out due to its association with toxicity and intolerance (WHO, 2009.).

5.2 HIV-1 subtype diversity

Subtype diversity is influenced by multiple factors, including cross-border demographic patterns that affect transmission and gene flow (Hemelaar *et al.*, 2011). A number of studies have reported subtype A as the predominant subtype in Kenya (Table 2.1). In this study, subtype A was still the predominant subtype, majority of these being sub-subtype A1.

A number of subtype A1 sequences clustered with reference sequence from Uganda, which shares the border with Kenya to the west (for *C2V3* region, KAH008, MLD258, KHC059, KHC069, KMB154, KAH203; and MLD198, KMB160 for *Pol RT* region). The clustering observed for sequences from Kisumu (prefix -KAH) with reference sequences from Uganda is probably due to the fact that Kisumu is located almost at the border with Uganda. A *pol RT* sequence from Malindi, MLD040, clustered with an A1 sequence from Senegal. Malindi being a tourist town, this could have been due to migration. The subtype C sequences clustered with reference sequences from South Africa (KHC052 and NVS148 for *C2V3*) and Ethiopia (KHC040 for both *C2V3* and *Pol RT*). South Africa and Ethiopia have predominantly subtype C (R. Lihana et al., 2012; Tully & Wood, 2010). MLD021 for *Pol RT* and NVS152 for *C2V3* clustered with subtype D reference sequences from Uganda and HND164 with a reference sequence from the Democratic Republic of Congo. For the *Pol RT* region, one of the sequences, KAH004 clustered with a subtype H from Cameroon which is an unusual subtype for Kenya. However, this sequence was comparatively short, which may reflect a compromised sequence quality and integrity and consequently, a biased subtype

assignment on the phylogenetic tree. A number of sequences were orthologous and in some cases had similar branch lengths showing that these sequences were phylogenetically closely related despite coming from different counties. The orthologous sequences for the *C2V3* region included KAH009 and NVS095; KAH001 and NVS121; HND046 and KHC013; KAH157 and KHC050; KHC052 and NVS148; KAH010 and KHC064, while those for the *Pol RT* region included KHC163 and NVS069; KMB145 and MLD545; KHC045 and KMB067; KMB077 and NVS058. These similarities suggest restricted evolutionary trend and virus transmission within the population across the country. A study with a larger sample size will probably bring out these dynamics to a greater extent.

It was hypothesized that the diversity of HIV in Kenya could be on the rise due to increasing demographic heterogeneity, human migration and ease of travel within Kenya; increasing influence of antiretroviral pressure in addition to generalized viral evolution. To capture this diversity, the plasma and cellular viruses were both assessed.
It has been shown that HIV subtypes vary between various compartments, with suggestion that inter-compartment recombination may anchor evolutionary mechanism that helps shape intra-host diversity of HIV-1 *env* gene in natural infection (Brown *et al.*, 2011). Compartmental differences in genetic diversity were observed in this study as well, particularly with the *pol RT* sequences. The cellular compartment had 10.5% of the sequences being recombinants while the cell-free compartment had 31.8% of the sequences being recombinants. This difference was less remarkable for the *C2V3* sequences, where recombinant viruses were 14.3% in the cellular compartment against 17.6% in the cell-free compartment. The *pol RT* region of HIV-1 is the target for current antiretroviral drugs (NASCOP, 2011). Constant exposure to these drugs and lack of adherence and compliance is associated with drug resistance mutations, which can further enhance variability (Ochieng *et al.*, 2015). Zidovudine, a Nucleoside Reverse Transcriptase Inhibitor (NRTI) for example, can enhance the mutation rate of HIV-1 by a factor of 7 for each replication cycle, and HIV-1 variants harbouring Zidovudine

resistant reverse transcriptase can have a mutation rate that is 3 fold that of the wild virus type (Martinez-Picado & Martinez, 2008).

Production of variants that are endowed with superior fitness that arises due to recombination has been observed in cases of dual or triple infection of a patient (Templeton et al., 2009; Van der Kuyl & Cornelissen, 2007). In this study, 19 isolates were sequenced for both the *pol RT* region and the *C2V3* region. Using JPHMM, it was found that the subtype was discordant for the *pol RT* and *C2V3* region (*polRT:C2V3*), for 7 isolates KAH004 (A1D:A1), KAH010 (A1:A1D), KHC045 (A1:A1H), KHC048 (A1:A1A2), MLD011 (D:A1D), MLD245 (A2B:A2), NVS032 (A1:A1H). Using phylogeny, subtype was discordant for 3 isolates KAH004 (H:A1), MLD011 (D:A1) and MLD 245 (A2D:A2). Phylogenetic method is not powered to detect recombinant strains, and there is a possibility that the unassigned isolates or strains in unusual clusters are recombinant viruses. Recombination observed in these instances, especially using phylogeny, is highly indicative of dual or triple infection in these patients. Recombination between variants that are highly similar occurs more frequently than with variants that are distant (Motomura, J. Chen, & Hu, 2008). Malindi, a town at the Coast of Kenya whose main economic activity is Tourism, was found to have many of these recombinants (prefix of sequence ID - MLD). Hue reported a similar pattern of recombinants at Kilifi, a town also located at the Coast of Kenya and about 100km from Malindi (Hué *et al.*, 2012).The recombinants they found in Kilifi using various methods include A1-AE, A1-A2, A1-A2-B, A1-A2-D, A1-C, A1-D, A2-B, A2-D, AE-C, B-C, C-D. It is highly probable too, that such variants are spread through intravenous drug use (IDU). The proportion of HIV infections linked to IDU at the Coast of Kenya is 17% compared to 4-5% in the whole country (Nieburg & Lisa, 2011). The high proportion of recombinants (in the *pol RT* region) in Malindi as compared to the other counties correlates well with what Hue found in Kilifi (Hué *et al.*, 2012).

The cellular compartment had a greater proportion and variety of pure subtypes than the acellular compartment, this observation being more remarkable with the *env* region than the *pol RT* region.

5.3 Co-receptor Usage

The analysis of *C2V3* sequences generated from this study using the Geno2pheno platform revealed that 75% of the viruses used CCR5 as their co-receptor of choice (R5-tropic) while 25% used CXCR4 as co-receptor (X4-tropic) for viral entry to CD4 T-cells at FPR10. No dual tropic sequences were found by the method used, which may be due to lack of sensitivity of the method to pick dual tropism. Site differences in tropism was not analysed due to differences in sample size that could bias the result. Subtype differences in co-receptor usage have been reported in different studies though few coreceptor studies have been done in Kenya. A few of the local studies with varied numbers of subjects have found majority (73%-84%) of HIV-isolates to be CCR5 tropic, although comparatively fewer samples were assessed (R.W. Lihana et al., 2009b; Nyamache, Muigai, Ng'ang'a, & Khamadi, 2013; Wambui et al., 2012) . In the current study, using a FPR of 10%, 81.8% of the A1 sequences were R5-tropic and 19.2% were X4-tropic sequences. These findings are in accord with the general trajectory of HIV-1 subtype A co-receptor tropism from the local population.

Subtype C has been shown to preferentially use CCR5 and to rarely induce syncytia (Abraha et al., 2009; Bessong et al., 2005). The current study found that all subtype C isolates were R5-tropic. CXCR4 use is generally associated with a more rapid decrease of CD4 counts and therefore faster disease progression, with subtype D viruses preferentially being X4-tropic (Noah Kiwanuka et al., 2008; Wambui et al., 2012). The study by Kiwanuka and colleagues attributed faster rate of disease progression in subtype D to a higher frequency of syncytium formation and tropism for CXCR4 than other subtypes. In the present study, a greater proportion i.e. 80%, of subtype D sequences were X4-tropic at FPR10 while at FPR20, all isolates were X4-tropic.

5.4 Number of Potential N-Linked Glycosylation (PNG) Sites

Glycans play various roles in the infective cycle of the virus. Any change in the glycosylation sites can affect protein folding and conformation. The envelope of HIV-1 becomes coated with so many glycans that the virus can become 'invisible' to (protected from) the immune response of the host (Poon *et al.*, 2007). In order to assess the relationships among PNG, HIV subtype and viral tropism, the number and proportion of PNGs were compared between coreceptor tropism (CCR5 versus CXCR4 tropic) and between the different genetic source materials (DNA versus RNA).

Compared by different blood compartment and genetic material, this study found a higher and significantly different frequency of PNGs in DNA (cellular blood) than RNA (cell-free blood). The difference was more exemplified by the NXT pattern of PNGs ($p=0.016$). It would be expected that RNA isolates would be more glycosylated as a way of counteracting the more intense selection pressure seen in this compartment due to host defense mechanisms or pressure from anti-retroviral therapy (Ho *et al.*, 2008). A possible explanation for the deviation from this expectation could be subtype differences. The study by Ho and colleagues comprised mainly subtype B viruses while majority of isolates in this study were of subtype A.

Comparing PNGs by tropism, X4 tropic viruses were more glycosylated in this study than the R5-tropic viruses both for the NXT or NXS pattern. Existing evidence from mainly subtype B and C viruses suggest that R5-tropic strains tend to be more glycosylated than their X4 counterparts (Kalinina et al., 2013; Zhang et al., 2004). The deviation from literature can be explained by the fact that majority of HIV-1 strains in Kenya and as shown from this study, are subtype A1. All the other subtypes in this study had more PNGs on average for the R5-tropic virus than the X4-tropic virus.

PNGs among other genetic and structural characteristics, have been associated with enhanced transmission virus (C.B. Wilen et al., 2011). Comparing PNGs by subtype variations, subtype A was found in a separate study to have greater selection for viruses

with shorter V1-V2 loops and fewer PNGs (Chohan *et al.*, 2005). The R5 tropic virus is preferentially transmitted compared to the X4 tropic virus (Pollakis *et al.*, 2001). The ease of transmissibility could also be enhanced for subtype A viruses (Chohan *et al.*, 2005), by having fewer PNGs. This study found that the R5 tropic subtype A viruses were less glycosylated than the X4 tropic virus, a characteristic that would enhance their transmissibility. It may be that as the disease progresses (and the tropism also shifts from R5 to X4 tropism), the virus adapts itself to fight the pressure from neutralizing antibodies by increasing its PNGs. X4-tropic viruses have been observed during later stages of infection and induce synctia formation that leads to faster disease progression (Berger *et al.*, 1998). It would be useful to do a review of the Subtype A1 viruses in Genbank, analyzing their N-linked glycosylation patterns and their coreceptor usage and comparing this with subtype A2, A1 recombinants and other subtypes.

The relationship between Total PNGS, NXT, NXS patterns and coreceptor usage though different on average was found to be not significant. The same applies with the relationship between the PNG patterns and subtype. Some glycosylation sites have been found to have a high frequency of glycosylation. The study also found a significant relationship, using chi square test, between subtype and glycosylation at N296 ($p=0.021$), N302 ($p=0.034$) and N366 ($p=0.016$). The significance of this is yet to be determined.

5.5 Conclusion

It can be concluded from this study that:

1. The major subtype in Kenya is still A1 although one can see different patterns in different regions and an increasing number of recombinants.
2. Comparison between the cellular and acellular compartments indicates that there is greater recombination in the acellular compartment and this can be explained by the fact that there is greater immune pressure within this compartment than the cellular compartment.
3. Seventy-five percent of the sequences were predicted to use the CCR5 coreceptor at the FPR of 10%. There were subtype differences in the use of coreceptor with

subtype A1 having 80% of the sequences being CCR5-tropic, subtype A2 being 50% CCR5-tropic, subtype C being 100% CCR5-tropic and subtype D being 20% CCR5-tropic. These differences in tropism have direct implications on disease transmission and progression.

4. N-linked glycosylation patterns showed inter-subtype variability with subtype A1 (which were the majority of the sequences) having less PNGs for the R5-tropic virus than the X4-tropic virus. The reverse pattern was seen for the other subtypes. This study also showed more PNGs from the isolates obtained from DNA than RNA for the subtype A virus.

5.6 Limitations of the study

1. In carrying out PCR, negative results were obtained severally due to contamination of the machinery and the laboratory itself. This meant that fewer samples than expected amplified for the required gene region.
2. In assessing coreceptor usage, phenotypic studies are more rigorous than genotypic studies. There wasn't sufficient capacity at the time of this study, to carry out phenotypic analyses.

5.7 Recommendations

1. It may be helpful, if resources are available, to derive samples from several health facilities in the same county so as to ensure that samples obtained are more representative of the entire population of that county. This will enable a more comprehensive analysis of HIV-1 diversity within the county.
2. Coreceptor studies that incorporate phenotypic analysis could be done so as to provide a more in depth characterization.
3. Carrying out an N-linked glycosylation analysis of the entire C1-V5 region of *env* may yield more useful information about the interaction of the different N-glycosite positions and subtype differences in the pattern of PNGs.

REFERENCES

Abraha, A., Nankya, I., Gibson, R., Demers, K., Tebit, D., Johnston, E., … & Arts, E. (2009). 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. *J Virology, 83*(11), 5592-5605.

Abram, M. E., Ferris, A. L., Shao, W., Alvord, W. G., & Hughes, S. H. (2010). The nature position and frequency of mutations made in a single cycle of HIV-1 replication. *J Virology, 84*, 9864-9878.

Alcantara, L. C., Cassol, S., Libin, P., Deforche, K., Pybus, O., Van Ranst, M., … & de Oliveira, T. (2009). A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. *Nucleic Acids Res, 37*(Web-server issue), W634-W642.

Archer, J. P. (2008). *The diversity of HIV.* (PhD), University of Manchester, Manchester.

Ariën, K. K., Abraha, A., Miguel E. Quiñones-Mateu, Luc Kestens, Guido Vanham, & Arts, E. J. (2005). The Replicative Fitness of Primary Human Immunodeficiency Virus Type 1 (HIV-1) Group M, HIV-1 Group O, and HIV-2 Isolates. *J Virology, 79*(14), 8979-8990.

Arroyo, M. A., Sateren, W. B., Foglia, G., Kibaya, R., Langat, L., Wasunna, M., … & McCutchan, F. E. (2009). HIV Type 1 Genetic Diversity among Tea Plantation Workers in Kericho, Kenya. *AIDS Res and Human Retro, 25*(11), 1061-1064.

Asamoah-Odei, E., Garcia Calleja, J. M., & Boerma, J. T. (2004). HIV prevalence and trends in sub-Saharan Africa: no decline and large subregional differences. *Lancet, 364*, 35-40.

Baeten, J. M., Chohan, B., Lavreys, L., Chohan, V., McClelland, R. S., Certain, L., … & Overbaugh, J. (2007). HIV-1 Subtype D Infection Is Associated with Faster Disease Progression than Subtype A in Spite of Similar Plasma HIV-1 Loads. *J Infectious Disease, 2007*(195), 1177-1180.

Berger, E. A., Doms, R., Fenyö EM, Korber BT, Littman DR, Moore JP, … & RA., W. (1998). A new classification for HIV-1. *Nature, 391*(240).

Bessong, P. O., Chikwelu, L. O., Tonie, C., Choge, I., Phoswa, M., Candice, P., … & Lynn, M. (2005). Characterization of Human Immunodeficiency Virus Type 1 from a Previously Unexplored Region of South Africa with a High HIV Prevalence. *AIDS Res and Human Retro, 21*(1), 103-109.

Botarelli, P., Houlden, B. A., Haigwood, N. L., Servis, C., Montagna, D., & Abrignani, S. (1991). N-glycosylation of HIV-gp120 may constrain recognition by T lymphocytes. *J Immunology, 147*, 3128-3132.

Bradac, J. A., Charles, C., Kristina, K., Catherine, M., & James, J. S. (2002). *HIV Molecular Immunology*. New Mexico, USA: Theoretical Biology and Biophysics, Los Alamos National Library.

Brooks, S., Dwek, M., & Schumacher, U. (2002). *Functional and molecular glycobiology*. Oxford UK: BIOS Scientific.

Brown, R. J., Peters, P. J., Caron, C., Gonzalez-Perez, M. P., Stones, L., Ankghuambom, C., … & Ball, J. K. (2011). Intercompartmental recombination of HIV-1 contributes to env intrahost diversity and modulates viral tropism and sensitivity to entry inhibitors. *J Virology, 85*(12), 6024-6037.

Castro-Nallar, E., Keith A Crandall, & Pérez-Losada, M. (2012). Genetic Diversity and Molecular Epidemiology of HIV Transmission. *Future Virology, 7*(3), 239-252.

Chohan, B., Lang, D., Sagar, M., Korber, B., Layreys, L., Richardson, B., & Overbaugh, J. (2005). Selection for Human Immunodeficiency Virus Type 1 Variants with Shorter V1-V2 loop Sequences Occurs During Transmission of Certain Genetic Subtypes and may Impact Viral RNA Levels. *J Virology, 79*(10), 6528-6531.

Corey, L., & McElrath, M. J. (2010). HIV vaccines: mosaic approach to virus diversity. *Nature Med*(16), 268-270.

Crous, S., Krishna, R. S., & Travers, S. A. (2012). Appraising the performance of genotyping tools in the prediction of coreceptor tropism in HIV-1 subtype C viruses. *BMC Infectious Diseases, 12*(203), 1-8.

de Oliveira, T., Deforche, K., Cassol, S., Salminem, M., Paraskevis, D., Seebregts, C., … & Vandamme, A.-M. (2005). An Automated Genotyping System for Analysis of HIV-1 and other Microbial Sequences. *Bioinformatics, 21*(19), 3797-3800.

Delport, W., Poon, A. F. Y., Frost, S. D. W., & Pond, S. L. K. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics, 26*(19), 2455-2457.

Dowling, W. E., Kim, B., Mason, C. J., Wasunna, K. M., Alam, U., Elson, L., … & Carr, J. K. (2002). Forty-one near full-length HIV-1 sequences from Kenya reveal an epidemic of subtype A and A-containing recombinants. *AIDS, 16*(13), 1809-1820.

Ferris, R. L., Hall, C., Sipsas, N. V., Safrit, J. T., Trocha, A., Koup, … & Siliciano, R. F. (1999). Processing of HIV-1 envelope glycoprotein for class I–restricted recognition: dependence on TAP1/2 and mechanisms for cytosolic localization. *J Immunology, 162*, 1324-1332.

Foley, B., Leitner, T., Apetrei C, Hahn B, Mizrachi I, Mullins J, … & Korber B. (2013). *HIV Sequence Compendium 2013*: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, LA-UR 13-26007.

Garrido, C., Roulet, V., Chueca, N., Poveda, E., Aguilera, A., Skrabal, K., … & de Mendoza, C. (2008). Evaluation of Eight Different Bioinformatics Tools To Predict Viral Tropism in Different Human Immunodeficiency Virus Type 1 Subtypes. *J Clinical Microbiology, 46*(3), 887-891.

Gordon, M. T., De Oliveira, K., Bishop, H. M., Coovadia, L., Madurai, S., Engelbrecht, E., … & Cassol, S. (2003). Molecular characteristics of human immunodeficiency virus type 1 subtype C viruses from KwaZulu-Natal, South Africa: implications for vaccine and antiretroviral control strategies. *J Virology, 77*, 2587-2599.

Gray, R. R., Tatem, A. J., Lamers, S., Hou, W., Laeyendecker, O., Serwadda, D., … & Salemi, M. (2009). Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *AIDS, 23*(14).

Hahn, B., Shaw, G., De Cock, K., & Sharp, P. (2000). AIDS as a zoonosis: Scientific and public health implications. *Science, 287*, 607-614.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series, 41*, 95-98.

Hemelaar, J., Gouws, E., Ghys, P. D., & Osmanov, S. (2011). Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS, 25*(5), 679-689. doi:10.1097/QAD.0b013e328342ff93

Ho, Y. S., Abecasis, A. B., Kristof, T., Koen, D., Dominic, E. D., Michael, C., … & Saksena, N. K. (2008). HIV-1 gp120 N-linked glycosylation differs between plasma and leukocyte compartments. *Virology J, 5*(14), 1-10.

Hué, S., Hassan, A. S., Nabwera, H., Sanders, E. J., Pillay, D., Berkley, J. A., & Cane, P. A. (2012). HIV Type 1 in a Rural Coastal Town in Kenya Shows Multiple Introductions with Many Subtypes and Much Recombination. *AIDS Res and Human Retro, 28*(2), 220-224.

International Agency for Research on Cancer. (1996). IARC Monograph-100B. *IARC Monograph*, 215-253.

Kageha, S., Lihana, R. W., Okoth, V., Mwau, M., Okoth, F. A., Songok, E. M., … & Khamadi, S. A. (2012). HIV type 1 subtype surveillance in central Kenya. *AIDS Res Hum Retro, 28*(2), 228-231. doi:10.1089/aid.2011.0089

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

Kalinina, O. V., Pfeifer, N., & Lengauer, T. (2013). Modelling binding between CCR5 and CXCR4 receptors and their ligands suggests the surface electrostatic potential of the co-receptor to be a key player in the HIV-1 tropism. *Retrovirology, 10*(130), 1-11.

Kalish, M. L., Korber, B. T., Pillai, S., Robbins, K. E., Leo, Y. S., Saekhou, A., … & Chan, R. (2002). The sequential introduction of HIV-1 subtype B and CRF01AE in Singapore by sexual transmission: accelerated V3 region evolution in a subpopulation of Asian CRF01 viruses. *Virology, 304*, 311-329.

Kasturi, L., Chen, H., & Shakin-Eshleman, S. H. (1997). Regulation of N-linked core glycosylation : use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. *J Biochemistry, 323*, 415-419.

Kebira, A. N., Waihenya, R., & Khamadi, S. (2009). HIV-1 genetic diversity among HIV drug naive populations of Nairobi, Kenya. *Retrovirology, 6*(3), 371.

Keele, B. F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M. L., … & Hahn, B. H. (2006). Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science, 313*(5786), 523-526.

Khamadi, S. A., Lihana, R. W., Mwaniki, D. L., Kinyua, J., Lagat, N., Carter, J. Y., … & Ochieng, W. (2008). HIV type 1 genetic diversity in Moyale, Mandera, and Turkana based on env-C2-V3 sequences. *AIDS Res Hum Retro, 24*(12), 1561-1564. doi:10.1089/aid.2008.0085

Khamadi, S. A., Lihana, R. W., Osman, S., Mwangi, J., Muriuki, J., Lagat, N., … & Okoth, F. A. (2009). Genetic diversity of HIV type 1 along the coastal strip of Kenya. *AIDS Res Hum Retro 25*(9), 919-923. doi:10.1089/aid.2009.0005

Khamadi, S. A., Ochieng, W., Lihana, R. W., Kinyua, J., Muriuki, J., Mwangi, J., … & Songok, E. M. (2005). HIV type 1 subtypes in circulation in northern Kenya. *AIDS Res Hum Retro, 21*(9), 810-814. doi:10.1089/aid.2005.21.810

Khoja, S., Ojwang, P., Khan, S., Okinda, N., Harania, R., & Ali, S. (2008). Genetic Analysis of HIV-1 Subtypes in Nairobi, Kenya. *PLoS ONE, 3*(9).

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Molecular Evolution, 16*(2), 111-120.

Kitawi, R. C., Nzomo, T., Mwatelah, R. S., Aman, R., Kimulwo, M., J. , Masankwa, G., … & Ochieng, W. (2015). HIV-1 Subtype Diversity Based on Envelope C2V3 Sequences from Kenyan Patients on Highly Active Antiretroviral Therapy. *AIDS Res and Human Retro, 31*(4), 452-455.

Kiwanuka, N., Oliver Laeyendecker, Robb, M., Kigozi, G., Arroyo, M., McCutchan, F., … & Gray, R. (2008). Effect of Human Immunodeficiency Virus Type 1 (HIV-1) Subtype on Disease Progression in Persons from Rakai, Uganda, with Incident HIV-1 Infection. *J Infectious Disease, 197*, 707-713.

Kiwanuka, N., Robb, M., Laeyendecker, O., Kigozi, G., Wabwire-Mangen, F., Makumbi, F. E., … & Whalen, C. C. (2010). HIV-1 Viral Subtype Differences in the Rate of CD4+ T-Cell Decline Among HIV Seroincident Antiretroviral Naive Persons in Rakai District, Uganda. *J Acquir. Immune Def. Syndr, 54*(2), 180-184.

Koch, N., Ndihokubwayo, J. B., Yahi, N., Tourres, C., Fantini, J., & Tamalet, C. (2001). Genetic analysis of HIV type 1 strains in Bujumbura (Burundi): predominance of subtype C variant. *AIDS Res and Human Retro, 17*, 269-273.

Koichiro, T., Glen, S., Daniel, P., Alan, F., & Sudhir, K. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evol, 30*, 2725-2729.

Koot, M., Vos A.H, Keet R.P, de Goede R.E., Dercksen M.W, Terpstra F.G, … & M., T. (1992). HIV-1 biological phenotype in long-term infected individuals evaluated with an MT-2 cocultivation assay. *AIDS, 6*(1), 49-54.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., … & Bhattacharya, T. (2000). Timing the Ancestor of the HIV-1 Pandemic Strains. *Science, 288*(5472), 1789-1796.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., … & Higgins, D. G. (2007). ClustalW and ClustalX version 2. *Bioinformatics, 23*(21), 2947-2948.

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

Lengauer, T., & Sing, T. (2006). Bioinformatics-assisted anti-HIV therapy. *Nature Reviews Microbiology, 4*(10), 790-797.

Li, W.-H. (1997). *Molecular Evolution*. Sunderland, Massachusetts: Sinauer Associates.

Lihana, R., Ssemwanga, D., Abimiku, A., & Ndembi, N. (2012). Update of HIV-1 diversity in Africa: A decade in review. *AIDS Rev*(14), 83-100.

Lihana, R. W., Khamadi, S. A., Lwembe, R. M., Kinyua, J. G., Muriuki, J. K., Lagat, N. J., … & Songok, E. M. (2009b). HIV-1 subtype and viral tropism determination for evaluating antiretroviral therapy options: an analysis of archived Kenyan blood samples. *BMC Infectious Diseases, 9*(215).

Lihana, R. W., Khamadi, S. A., Lwembe, R. M., Ochieng, W., Kinyua, J. G., Kiptoo, M. … & Songok, E. M. (2009c). The changing trend of HIV type 1 subtypes in Nairobi. *AIDS Res Hum Retro, 25*(3), 337-342. doi:10.1089/aid.2008.0228

Lihana, R. W., Lwembe, R. M., Bi, X., Ochieng, W., Panikulam, A., Palakudy, T., … & Ichimura, H. (2011). Efficient monitoring of HIV-1 vertically infected children in Kenya on first-line antiretroviral therapy. *J Clinical Virology, 52*(2), 123-128. doi:10.1016/j.jcv.2011.06.014

Lwembe, R., Ochieng, W., Panikulam, A., Mongoina, C. O., Palakudy, T., Koizumi, Y., … & Ichimura, H. (2007). Anti-retroviral drug resistance-associated mutations among non-subtype B HIV-1-infected Kenyan children with treatment failure. *J Medical Virology, 79*(7), 865-872. doi:10.1002/jmv.20912

Maljkovic, B. I., Athreya, G., Kothari, M., Daniels, M., Bruno, W. J., Korber, B., … & Leitner, T. (2009). The evolutionary rate dynamically tracks changes in HIV-1 epidemics: Application of a simple method for optimizing the evolutionary rate in phylogenetic trees with longitudinal data. *Elsevier, Epidemics, 1*, 230-239.

Maljkovic, I. B., Ruy Ribeiro, Moulik Kothari, Gayathri Athreya, Marcus Daniels, Ha Youn Lee, … & Leitner, T. (2007). Unequal Evolutionary Rates in the Human Immunodeficiency Virus Type 1 (HIV-1) Pandemic: the Evolutionary Rate of HIV-1 Slows Down When the Epidemic Rate Increases. *J Virology, 81*(19), 10625-10635.

Martinez-Picado, J., & Martinez, M. A. (2008). HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: a view from the clinic and ex vivo. *Virus Res, 134*(1-2), 103-123.

Micah, O., Mbithi, J., Oyugi, F., Laten, A., Anzala, O., & Engelbrecht, S. (2011). Molecular Characterization of HIV Type 1 Among HIV-Infected Respondents in a Cohort Being Prepared for HIV Phase III Vaccine Clinical Trials, Western Kenya *AIDS Res and Human Retro, 27*(3), 257-264.

Ming, Z., Gaschen, B., & et al. (2004). Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology, 14*(12), 1229-1246.

Motomura, K., J. Chen, & Hu, W. S. (2008). Genetic recombination between human immunodeficiency virus type 1(HIV-1) and HIV-2, two distinct human lentiviruses. *J Virology, 82*(4), 1923-1933.

Mulinge, M., Lemaire, M., Servais, J., Rybicki, A., Struck, D., da Silva, E., … & Bercoff, D. (2013). HIV-1 Tropism Determination Using a Phenotypic Env Recombinant Viral Assay Highlights Overestimation of CXCR4-Usage by Genotypic Prediction Algorithms for CRF01_AE and CRF02_AG. *PLoS ONE, 8*(5), 1-13.

NACC, & NASCOP. (2012). *Kenya Aids Epidemic Update 2011*. Government of Kenya,.

NACC. (2014). *Kenya HIV County Profiles*. Retrieved from Nairobi, Kenya:

NASCOP. (2011). *Guidelines for Antiretroviral Therapy in Kenya*. Government of Kenya; Nairobi, Kenya.

NASCOP. (2014). *Kenya AIDS Indicator Survey 2012: Final Report.* Retrieved from Nairobi:

Nieburg, P., & Lisa, C. (2011). *HIV Prevention Among Injection Drug Users in Kenya and Tanzania*. Retrieved from Washington DC:

Nisole, S., & Saib, A. (2004). Early steps of retrovirus replicative cycle. *Retrovirology, 1*(9), 1-20.

Nyagaka, B., Kiptoo, M. K., Lihana, R., Khamadi, S., Makokha, E., Kinyua, J., … & Songok, E. (2012). HIV type 1 gag genetic diversity among antenatal clinic attendees in North Rift Valley, Kenya. *AIDS Res and Human Retro, 28*(5), 523-526. doi:10.1089/AID.2011.0223

Nyamache, A. K., Muigai, A. W. T., Ng'ang'a, Z., & Khamadi, S. A. (2013). Profile of HIV Type 1 Coreceptor Tropism Among Kenyan Patients from 2009 to 2010. *AIDS Res and Human Retro, 29*(8), 1105-1109.

Nzomo, T. J., Kitawi, R. C., Mwatelah, R. S., Aman, R., Kimulwo, M. J., Masankwa, G., … & Ochieng, W. (2015). Genotypic variability of HIV-1 Reverse Transcriptase gene from long-term antiretroviral-experienced patients in Kenya. *AIDS Res and Human Retroviruses, 31*.

Ochieng, W., Kitawi, R., Nzomo, T., Mwatelah, R., Kimulwo, M., Ochieng, D., … & Aman, R. (2015). Implementation and Operational Research: Correlates of Adherence and Treatment Failure Among Kenyan Patients on Long-term Highly Active Antiretroviral Therapy. *J Acquir. Immune Def. Syndr, 69*(2), e49-56.

Plantier, J. C., Leoz, M., & Dickerson, J. E. (2009). A new human immunodeficiency virus derived from gorillas. *Nature Medicine, 15*(8), 871-872.

Pollakis, G., Kang, S., Kliphuis, A., Chalaby, M., Goudsmit, J., & Paxton, W. (2001). N-Linked Glycosylation of the HIV Type-1 gp120 Envelope Glycoprotein as a Major Determinant of CCR5 and CXCR4 Coreceptor Utilization. *J Biological Chemistry, 276*(16), 13433-13441.

Poon, A. F. Y., Lewis, F. L., Kosakovsky, P. S. L., & Frost, S. D. W. (2007). Evolutionary interactions between N-linked glycosylation sites in the HIV-1 envelope. *PLoS Computational Biology, 3*(1), 0110-0119.

Qiagen. (2012). *QIAamp® Viral RNA Mini Handbook*. Maryland.

Qiagen®. (2010). *QIAamp® DNA Mini and Blood Mini Handbook* (Third ed.). Maryland.

Rambaut, A., Posada, A., Crandall, K. A., & Holmes, E. C. (2004). The causes and consequences of HIV evolution. *Nature Reviews Genetics, 5*(1), 52-61.

Rambaut, A., Robertson, D. L., Pybus, O. G., Peeters, M., & Holmes, E. C. (2001). Human Immunodeficiency Virus. Phylogeny and origin of HIV-1. *Nature, 410*, 1047-1048.

Rhee, S.-Y., Gonzales, M. J., Kantor, R., Betts, B. J., Ravela, J., & Shafer, R. W. (2003). Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res, 31*(1), 298-303.

Robertson, D., Sharp, P., McCutchan, F., & Hahn, B. (1995). Recombination of HIV-1. *Nature, 374*(6518), 124-126.

Santiago, M., Range, F., Keele, B., Li, Y., Bailes, E., Bibollet-Ruche, F., … & Hahn, B. (2005). Simian immunodeficiency virus infection in free-ranging sooty mangabeys (Cercocebus atys atys) from the Taï Forest, Côte d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. *J Virology, 79*(19), 12515-12527.

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

Shafer, R. W. (2006). Rationale and Uses of a Public HIV Drug-Resistance Database. *J Infectious Diseases, 194*(Suppl 1), S51-58.

Shankarappa, R., Margolick, J., Gange, S., Rodrigo, A., Upchurch, D., Farzadegan, H., … & Mullins, J. (1999). Consistent Viral Evolutionary Changes Associated with the Progression of Human Immunodeficiency Virus Type 1 Infection. *J Virology, 73*(12), 10489-10502.

Si, Z., Cayabyab, M., & Sodroski, J. (2001). Envelope glycoprotein determinants of neutralization resistance in a simian-human immunodeficiency virus (SHIV-HXBc2P 3.2) derived by passage in monkeys. *J Virology, 75*, 4208-4218.

Sing, T., Beerenwinkel, N., & Lengauer, T. (2004). Learning mixtures of localized rules by maximizing the area under the ROC curve.

Skar, H., Charlotte, H., & Jan, A. (2011). HIV-1 evolution in relation to molecular epidemiology and antiretroviral resistance. *Annals NY Acad Sci, 1230*, 108-118.

Songok, E., Lwembe, R., Kibaya, R., Kobayashi, K., Ndembi, N., Kita, K., … & Ichimura, H. (2004). Active generation and selection for HIV intersubtype A/D recombinant forms in a coinfected patient in Kenya. *AIDS Res and Human Retro, 20*(2), 255-258. doi:10.1089/088922204773004996

Spira, S., Wainberg, M., Loemba, H., Turner, D., & Brenner, B. (2003). Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *J Antimicrobial Chemo, 51*(2), 229-240.

Starcich, B. R., Hahn Beatrice H. , George M. Shaw, Paul D. McNeely, Susanne Modrow, Hans Wolf, … & Wong-Staal, F. (1986). Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell, 45*(5), 637-648.

Steain, M. C., Wang, B., & Saksena, N. K. (2006). Analysis of HIV-1 sequences vertically transmitted to infants in Kisumu, Kenya. *J Clinical Virol, 36*(4), 298-302. doi:10.1016/j.jcv.2006.02.001

Steckbeck, J. B., Kuhlmann, A.-S., & Montelaro, R. C. (2013). C-terminal tail of human immunodeficiency virus gp41: Functionally rich and structurally enigmatic. *J General Virology, 94*(1), 1-19.

Struck, D., Lawyer, Glenn., Ternes, A.-M., Schmit, J.-C., & Bercoff, D. P. (2014). COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. *Nucleic Acids Res, doi: 10.1093/nar/gku739*.

Taylor, B. S., & Hammer, S. M. (2008). The challenge of HIV-1 subtype diversity. *N Engl J Med, 359*(18), 1965-1966. doi:10.1056/NEJMc086373

Templeton, A., Kramer, M., Jarvis, J., Kowalski, J., Gange, S., Schneider, … & Markham, R. (2009). Multiple infection and recombination in HIV-1 within a longitudinal cohort of women. *Retrovirology, 6*(article no. 54).

Thomson, M. M., Perez-Alvarez, L., & Najera, R. (2002). Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infectious Diseases, 2*(8), 461-471.

Tully, D., & Wood, C. (2010). Chronology and evolution of the HIV-1 subtype C epidemic in Ethiopia. *AIDS, 24*(10), 1577-1582.

Turner, B. G., & Summers, M. F. (1999). Structural Biology of HIV. *J Molecular Biology, 285*, 1-32.

UNAIDS. (2014). UNAIDS 2013 Global Fact Sheet. Retrieved from http://www.unaids.org/en/media/unaids/contentassets/documents/factsheet/2014/20140716_FactSheet_en.pdf

Van der Kuyl, A. C., & Cornelissen, M. (2007). Identifying HIV-1 dual infections. *Retrovirology, 4*, article no. 67.

Vasan, A., Renjifo, B., Hertzmark, E., Chaplin, B., Msamanga, G., Essex, M., … & Hunter, D. (2006). Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. *Clinical Infectious Disease*(42), 843-852.

Walker, P., Pybus, O., Rambaut, A., & Holmes, E. (2005). Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic growth. *Infection, Genetics, Evol, 5*(3), 199-208.

Wambui, V., Kiptoo, M., Kinyua, J., Odera, I., Muge, E., Muiruri, P., … & Songok, E. (2012). Predicted HIV-1 coreceptor usage among Kenya patients shows a high tendency for subtype d to be cxcr4 tropic. *AIDS Research Therapy, 9*(22), 1-7.

Wang, W., Nie, J., Prochnow, C., Truong, C., Jia, Z., Wang, S., … & Wang, Y. (2013). A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. *Retrovirology, 10*(1), 1-14.

WHO. (2009.). Rapid advice: antiretroviral therapy for HIV infection in adults and adolescents. Retrieved from http://www.who.int/hiv/pub/arv/advice/en/index.html

Wilen, C. B., Parrish, N. F., Pfaff J, M., Decker, J., Henning, E., Haim, H., … & Doms, R. (2011). Phenotypic and immunologic comparison of clade B transmitted/founder and chronic HIV-1 envelope glycoproteins. *J Virology, 85*(17), 8514-8527.

Wilen, C. B., Tilton, J. C., & Doms, R. W. (2012). Molecular Mechanisms of HIV Entry. *Viral Molecular Machines; Advances in Experimental Medicine and Biology, 726*, 223-242.

World Medical Association. (2000). *Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects*. Retrieved from Helsinki:

Worobey, M. (2007). The Origins and Diversification of HIV. *Global HIV/AIDS Medicine*, 13-21.

Woutter, J., Buve, A., & Nkengasong, J. N. (1997). The Puzzle of HIV subtypes. *AIDS, 11*, 705-712.

Yang, C., Li, M., Shi, Y., Winter, J., van Eijk, A., Ayisi, J., … & Lal, R. (2004). Genetic Diversity and High Proportion of Intersubtype Recombinants among HIV Type 1-Infected Pregnant Women in Kisumu, Western Kenya. *AIDS Res and Human Retro, 20*(5), 565-574.

Zhang, M., Gaschen, B., Blay, W., Foley, B., Haigwood, N., Kuiken, C., & Korber, B. (2004). Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology, 14*(12), 1229-1246.

APPENDICES

Appendix 1A: Consent Information for Study Participants (English)

Characterization of HIV Type 1 Genetic and Viral Diversity in Various Counties of Kenya

Rose Chari Kitawi; MSc Epidemiology, Jomo Kenyatta University of Agriculture and Technology (JKUAT)

Information to study participants

HIV can get into your blood through several ways such as when someone with the virus gives you there blood, accidents, or during or before birth when the mother has the virus. Anybody can get the virus! Although there are medicines to treat HIV, there is no cure. Available medicines also do not cure and sometimes they may cause other problems.

We want to know if drugs given to you are working well or they have failed to help you. If you accept, we will take a small amount of blood, the size of 2 full table spoons so as to look at how the drugs work and how the virus responds to those drugs. We are not giving you a different treatment, so you will have to continue taking the current medicine unless your doctor makes changes

Benefits of the study

If we use your blood to know why sometimes the medicines don't work, we can tell your doctor to give you better medicine and we can also use that knowledge to make better medicines which will help many people, including yourself in the future.

Confidentiality of your identity

We will not tell anybody anything about you. Nobody will know that you gave us blood. We will not even say something to anyone about your results.

Medical risks and problems.

When we take blood, you may experience some pain from the injection by needles. Sometimes the place may swell, but that will disappear shortly. We don't expect you to have any other problems because of the blood we take. But doctors will be available for you in case you feel different from before blood was taken.

Storing your blood for future studies

We will also keep your blood to look at it later.

Exportation of your samples:

Sometimes, we may send the blood out of Kenya where additional tests may be done if we do not have the equipment and chemicals to do it locally.

Obtaining additional information

If you have any questions at any time, please ask our Doctors and nurses and they help you. Their numbers are below, which you can also call.

Contact: Rose Kitawi 0733432175

Basis of participation

We will only take your blood if you agree. If you say no, we will not take blood. If you refuse, no one will know you talked to us and you will still be treated at this hospital as usual.

Tell us if you agree or if you don't agree to participate in this study

I agree ○ .... refuse ○..................to participate in this study.

Also tell us if you agree that we keep your blood to do tests later:

I agree ....○.....refuse ....○.

Also tell us if you agree that we can send your samples out of Kenya to be tested:

I agree to ◯ ....refuse ◯..................

Project Personnel to fill this section before the participant:

Unique study No.…………………Nationality/origin..………………………………………

Age……………… …………… ………… Sex ………………... ………………………..

Occupation…………………. ………………… Locality …………………………………..

Subject Recruited ◯ Not recruited to the study ◯ (Note: refusal by child overrides parental/guardian consent.

Appendix1B: Fomu ya idhini kwa wanauhusika katika utafiti huu (Kiswahili version)

Characterization of HIV Type 1 Genetic and Viral Diversity in Various Counties of Kenya

Rose Chari Kitawi; MSc. Epidemiology, Jomo Kenyatta University of Agriculture and Technology (JKUAT)

Taarifa kwa washiriki

Kuna njia nyingi ambazo unaweza kuambukuzwa ukimwi. Moja yao ni wakati wa kuzaliwa. Watoto haswa huambukizwa wakati wa kuzaliwa. Ijapokua kuna dawa za ukimwi, bado hakuna tibabu kamili au kinga.Virusi vinavyo ambukiza ukimwi vinaitwa HIV. Kuna aina mbalimbali ya HIV. Kuna aina zingine ambazo ni ngumu kutibu na huleta shida nyingi. Wakati mwingine HIV ya kawaida hubadilika na huzidi kuleta shida.

Kupitia uchunguzi huu, tunataka kuelewa zaidi jinsi HIV inavyobadilika na kuleta shida wakati mtu anapotumia dawa.

Tutakavyofanya Uchunguzi

Tutahitaji damu kiwango kidogo cha milliliter 10. Hiyo damu tutaipeleka kwa chumba cha utafiti ilitupime aina ya virusi vinavyopatikana kwa damu yako.

Ni faida gani utapata kutoka kwa uchunguzi huu?

Yale matokeo tutakayopata kutoka kwa upelelezi huu yatatumika kutengeneza dawa zilizo na nguvu ya kukabiliana na ukimwi. Matokeo hayo basi yatakuwa ya manufaa kwako na kwa wote walioambukizwa na virusi hivi.

Je, kuna wengine watakaojulishwa kuhusu ugonjwa wako au shida ulizonazo za kiafya?

La, hatutamwelezea mtu yeyote yule kuhusu ugonjwa wako au shida zozote ulizonazo za kiafya. Hata wanaofanya uchunguzi hawatajua jina lako kwasababu hatuandiki majina.

Kuna madhara gani kuhusianana uchunguzi huu?

Unaweza kuhisi uchungu kidogo tunapokutoa damu lakini hiyo ni kawaida. Ukihisi tatizo lolote unaweza kumfahamisha daktari wakati utakopotembelea kituo cha afya.

Kuna njia zingine zaidi tutatumia damu yako?

Ukikubali, tutahifadhi damu yako kwaajili ya uchunguzi wa baadaye. Pia huenda tukasafirisha damu yako nje ya Kenya tunapohitaji uchunguzi zaidi. Una uhuru kukubaliana na haya au unaweza kukataa

Maelezo zaidi

Utakapohitaji kuuliza maswali wakati wowote ule, wanautafiti wako tayari kukujibu. Pia unaweza kupiga simu wakati wowote kupitia nambari ifuatayo.

Rose Kitawi: 0733432175

Makubaliano

Tafadhalionyeshakwaalamahapachinichaguolako.

Unakubalia au unakataa kuhusika na Utafiti huu?

Ninakubali○ .... Ninakataa ○.................

Je unakubali kuhifadhiwa kwa damu yako?

Ninakubali○ .... Ninakataa ○.................

Je unakubali damu yako isafirishwe ngambo?

Ninakubali○ .... Ninakataa ○.................

Project Personnel to fill this section before the participant:

Unique study No.…………………Nationality/origin..………………………………………

Age……………… …………… ………… Sex ………………... ………………………..

Occupation…………………. ………………… Locality …………………………………..

Subject Recruited ◯ Not recruited to the study ◯ (Note: refusal of consent by children 13 to 17 overrides parental/guardian consent whereas refusal by guardian/parent nullifies child's consent).

Appendix 2A: Parental/Guardian consent for children 13-17 years (English).

About the study: We are doing a study to look at why some types of HIV cause more problems to treat than others. If we can answer that question, we can better understand how to treat HIV.

What about your child? We have talked to your child and explained about this study. Your child has the right to agree or refuse to participate and that choice will be respected. As a parent, you also have a right to refuse your child's participation. Although your child's choice will be respected, if they agree to participate but you refuse, we will not put them in the study.

Why are we telling you this? We want to know if you agree to have your child participate in this study or if you do not agree. Please make your choice next each section below

Do you agree that your child take part in this study?

I agree ○ ○ I do not agree

Do you agree that your child's blood may be kept for testing later?

I agree ○ ○ I do not agree

Do you agree that your child's blood sent outside Kenya to be tested?

I agree ○ I do not agree ○

Appendix 2B: Idhini ya Mzazi wa mtoto mwenye umri wa miaka 13-17 (Kiswahili Version).

Kuhusu uchunguzi huu

Huu uchunguzi utakagua jinsi virusi vya Ukimwi vinavyo kiuka madawa yanayo tibu ukimwi. Mtoto wako ameombwa kuhusika na utafiti huu. Kama mzazi, una haki ya kukataa mtoto asihusike. Tunakupatia hiyo fursa utuambie kama utamkubalia mtoto ahusike au la. Ikiwa mmoja wenu (mtoto au wewe kama mzazi) atakataa, basi mtoto hatahusika kamwe.

Tafadhali chagua moja kati ya chaguo mawili yanayofuata

Unakubali mtoto ahusike na uchunguzi huu?

Ndiyo, nakubali ○ La, nakataa ○

Unakubali damu ya mtoto ihifadhiwe kutumika baadaye inapotakikana?

Ndiyo, nakubali ○ La, nakataa ○

Unakubali damu ya mtoto itumwe ngambo ikaguliwe zaidi?

Ndiyo, nakubali ○ La, nakataa ○

Project Personnel to fill this section before the participant:

Unique study No.………………… Nationality/origin..……………………………

Age……………… Sex ……………….. Occupation………………….

Locality …………………………………..

Subject Recruited ○ Not recruited to the study ○ (Note: refusal of consent by children 13 to 17 overrides parental/guardian consent whereas refusal by guardian/parent nullifies child's consent).

Appendix 3A: QUESTIONNAIRE (English version)

Unique study No…………………… Nationality/origin..………… ………………

Age……………… Sex ………………….. Occupation………………….

Residence ………………………………….

1. Do you consider yourself as a Kenyan? Yes....... No.......
2. If No, what is your country of origin.......................?
    a. How long have you been in Kenya? ....
    b. When did you arrive? ....
    c. How often do you travel back and forth across the border
3. Do you travel often to countries other than Kenya? Yes....... No.......
    a. If Yes, which one(s)
    b. Which country did you last visit................when...........
4. Do you travel often to regions in Kenya outside where you live now? Yes....... No.......
    a. If Yes, which one(s)
    b. Which county/region did you last visit................when...........
5. Are you married? Yes....... No....... (if No, proceed to 6)
    a. If married, do you currently live together with your spouse? Yes..... .No.....
        i. Do you consider your relationship intimate? Yes.......No.......
    b. If not living together, do you ever come into contact Yes....... No.......
        i. How regular is the contact:
        ii. Is it an intimate contact:
6. Do you have sex partner(s) Yes..... .No.....
    a. If Yes, how many?...........
    b. Do you ever come into sexual contact Yes..... .No.....

c. If Yes, how regular?..................

7. Have you taken HIV test before? Yes..... .No.....If no, are you willing to take one?
8. Do you know your HIV status? Yes..... .No..... If no, do you want to know?
9. Does your partner/spouse know your HIV status? Yes..... .No.....
   a. If No, do you want them to know? Yes..... .No.....
   b. If Yes, do you need help to discuss with them your status? Yes..... .No.....
10. Has your partner/spouse taken HIV test before? Yes..... .No.....
    a. If Yes, do you know their status? Yes..... .No.....
    b. If No, do you continue to have sexual contact together? Yes..... .No.....
11. How long has it been since you first knew about your HIV status?
12. Are you taking any HIV drugs? Yes......No.....
    a. If Yes, how long since you started?
    b. Have you had to change the drugs because of complications?
       i. If Yes how many times?
    c. Have you had to stop taking the drugs temporarily for personal reasons
    d. Have you ever forgotten or postponed taking the drugs
       i. How often?
    e. Do you take them regularly now? Yes..... .No.....
13. If not taking HIV drugs, do you know you can get them free? Yes......No.....
    a. Do you want to start taking HIV drugs? Yes......No.....
14. Do you agree to extended storage of your blood/derivative samples for future analyses? Yes......No.....
15. Is there anything you want to tell me or questions you wish to ask?

Appendix 3B: MAHOJIANO

Nambari ya Utafiti…………………… Utaifa/Asili..………… …………

Umri…………… Mme/Mke ………………Kazi……………………………

Makao ………………………………….

1. Je, wewe in Mkenya? Ndio....... La.......
2. Kama si Mkenya, Nchi yako ni gani.......................?
    a. Ume umeishi humu Kenya kwa mda gani? ....
    b. Tangu lini umeishi humu nchini? ....
    c. How often do you travel back and forth across the border
3. Je, wewe husafiri mara nyingi kwa nchi zinginezo isipokuwa Kenya? Ndio.......La.......
    a. Kama ndio, we husafiri wapi?
    b. Je, mwisho utembelea nchi gani?.................lini?...........
4. Je, wewe husafiri mara nyingi na Kenya nje ya unapoishi sasa? Ndio....... La.......
    a. Kama ndio, wapi?
    b. Nchi gani ama sehemu gani ulisafiria mwisho?.................lini?...........
5. Umeoa/Umeolewa? Ndio....... La....... (kama La, endelea na #6)
    a. Kama uko ndoani, je, kwa sasa unaishi pamoja na wenzi wako? Ndio....La....
        i. Je, wewe uhusiano wako na mwenzio ni ya ubinafsi? Ndio.......La.......
    b. Kama si pamoja, huwa mnawahi kuwasiliana kimwili? Ndio........La......
        i. Mara ngapi mna jiunga ama wasiliana kwa ubinafsi:
        ii. Je, ni muungano wa karibu wa kindoa?:
6. Je, una mpenzi/ wapenzi wa kindoa? Ndio..... .La.....
    a. Kama ndio, wangapi?............

   b. We waja nao pamoja kwa ubinafsi wa ndoa? Ndio..... La.....
   c. Kama ndio, mara ngapi?..................

7. Ume wahi kupimwa viini vya Ukimwi? Ndio..... .La.....Kama la, unakubali kupimwa?
8. Unajua hali yako kidamu ya ukimwi? Ndio..... .La..... Kama la, wataka kujua?
9. Je, Mwenzi/Mpenzi wako ajua hali yako ya ukimwi? Ndio..... La.....
   a. Kama la, unataka/unakubali wajulishwe? Ndio..... .La.....
   b. Kama ndio, unahitaji ushaidishi kuwajilisha hali yako? ndio..... .la.....
10. Je, Mpenzi wako amewahi kupimwa kwa ajili ya Ukimwi? Ndio.. La....
    a. Kama ndio, unajua hali yake/yao? Ndio..... La.....
    b. Kama la, wewe uendelea kuwa naye/nao pamoja kimwili? Yes.... No....
11. Ni mda gani sasa tangu ulijua kuhusu hali yako ya Ukimwa mara ya kwanza?
12. Je, unatumia dawa ya kutibi Ukimwi? Ndio......La.....
    a. Kama ndio, kwa mda gani sasa tangu uanze kuzitumia?
    b. Je, umewahi kubadilishiwa madawa haya kwa ajili yeyote?
       i. Kama ndiyo, mara ngapi umebadili madawa haya?
    c. Je, umewahi kusimamisha utumishi wa madawa haya kwa sababu yoyote?
    d. Umewahi kusahau ama kuahirishwa uimishi wa haya madawa popotepo?
       i. Mara ngapi?
    e. Je, kwa sasa unatumia dawa hizo mara kwa mara? Ndio..... La.....
13. Je, kama hutimii dawa ya Ukimwi, unajua waweza kuzipata bure? Ndio......La.....
    a. Je, unataka kuanza kutumia dawa ya kutibu Ukimwi? Ndio......La.....
14. Je, unakubaliana na kuhifadhi kwa damu/sampuli yako kwa ajili ya utumizi wa baadaye?
15. Je, kuna kitu unataka kuniambia au maswali unataka kuuliza?

Appendix 4: Analysis of concordance between various subtyping tools

a) Analysis for the Envelope C2V3 Region

Patient ID	NCBI	concord	COMET	concord	REGA	concord	jpHMM	concord	Phylogen
HND046	A1	Y	A1	Y	A1	Y	A1	Y	A1
HND164	D/CRF10	P	D	Y	D	Y	D	Y	D
HND177	A1	Y	A1	Y	A1	Y	A1	Y	A1
KAH001	A2/C	P	A2 (check for 02_AG)	P	N/A	N	A2	Y	A2
KAH004	A1/CRF02	P	A1	Y	A1	Y	A1	Y	A1
KAH010	A1	Y	unassigned_1;45_cpx, D	N	N/A	N	A1D	P	A1?
KAH020	CRF07/C	P	C	Y	N/A	N	C	Y	C
KAH029	A1/A2	P	A1	Y	A1	Y	A1	Y	A1
KAH090	C/CRF07	P	C	Y	C	Y	C	Y	C
KAH138	D	Y	D	Y	D	Y	D	Y	D
KAH139	A1/CRF02	P	A1 (check for 02_AG)	P	N/A	N	A1	Y	A1
KAH142	A1/C	P	A1	Y	N/A	N	A1	Y	A1
KAH144	D/CRF10	P	D	Y	D	Y	D	Y	D
KAH148	A1	Y	A1	Y	A1	Y	A1	Y	A1
KAH157	A1/U	P	A1 (check for 02_AG)	P	A1	Y	A1	Y	A1
KAH227	CRF11/CRF02/A1	P	A1	Y	N/A	N	A1	Y	A1
KAH008	A1	Y	A1	Y	A1	Y	A1	Y	A1
KAH009	CRF02/A1/CRF11	P	A1	Y	A1	Y	A1	Y	A1
KHC013	A1/CRF11/C	P	A1	Y	N/A	N	A1	Y	A1
KHC016	A1/CRF11	P	A1	Y	A1	Y	A1	Y	A1-UG
KHC039	A1/CRF11	P	A1	Y	A1	Y	A1A2	P	A1
KHC040	C/CRF08	P	C	Y	C	Y	C	Y	C
KHC045	A1/CRF11	P	A1	Y	A1	Y	A1H	P	A1
KHC048	A1/CRF0	P	A1	Y	A1	Y	A1A2	P	A1

Patient ID	NCBI	concord	COMET	concord	REGA	concord	jpHMM	concord	Phylogen
	2								
KHC050	A1/CRF02/CRF11	P	A1	Y	A1 (100)	Y	A1	Y	A1
KHC052	C/CRF08	P	C	Y	C	Y	C	Y	C
KHC058	A1	Y	A1	Y	A1	Y	A1	Y	A1
KHC059	A1/CRF11	P	A1	Y	A1	Y	A1	Y	A1
KHC064	A1/CRF10/CRF13/CRF02/	P	unassigned_2;02_AG, 10_CD	N	N/A	N	A1D	P	A1
KHC069	A1/CRF02/CRF11	P	A1	Y	N/A	N	A1	Y	A1
KHC160	A1	Y	A1	Y	A1	Y	A1	Y	A1
KHC163	A1/CRF02	P	A1	Y	A1	Y	A1	Y	A1
KMB031	CRF02/A1	P	A1	Y	A1	Y	A1	Y	A1-UG
KMB039	CRF02/B	N	A1 (check for 02_AG)	P	A1	Y	A1A2	P	A1-UG
KMB067	A1	Y	A1 (check for 02_AG)	P	A1	Y	A1	Y	A1
KMB154	A1	Y	A1	Y	N/A	N	A1	Y	A1
MLD011	A1/D	P	unassigned_1;D, A1, 02_AG	P	N/A	N	A1D	P	A1?
MLD012	A1/C	P	unassigned_1;D, A1	P	N/A	N	A1A2D	P	A1?
MLD016	A1/A2	P	A1	Y	N/A	N	A1	Y	A1?
MLD185	A1	Y	A1	Y	A1	Y	A1	Y	A1
MLD245	CRF16/A2	P	A1	N	N/A	N	A2	Y	A2
MLD258	A1/CRF11	P	A1	Y	A1	Y	A1	Y	A1
MLD301	CRF02	N	A1	Y	A1	Y	A1	Y	A1
NVS008	D	Y	D	Y	D	Y	D	Y	D
NVS032	A1	Y	A1	Y	A1	Y	A1H	P	A1-UG
NVS040	A1/CRF02	P	A1	Y	A1	Y	A1	Y	A1-UG
NVS051	A1	Y	A1	Y	A1	Y	A1	Y	A1
NVS056	A1	Y	A1	Y	A1	Y	A1	Y	A1-UG

Patient ID	NCBI	concord	COMET	concord	REGA	concord	jpHMM	concord	Phylogen
NVS074	A2/CRF16	P	A1	N	A1	N	A2	Y	A2
NVS085	A1	Y	A1	Y	A1	Y	A1	Y	A1?
NVS090	A1/CRF11	P	A1	Y	A1	Y	A1	Y	A1?
NVS095	A1/A2	P	A1	Y	A1	Y	A1	Y	A1
NVS108	A1/G	P	A1	Y	A1	Y	A1	Y	A1-UG
NVS121	A2/CRF13	P	A2	Y	N/A	N	A2	Y	A2
NVS132	CRF02/CRF13	N	A1 (check for 02_AG)	P	A1	Y	A1	Y	A1-UG
NVS135	A1/CRF11	P	A1	Y	A1	Y	A1	Y	A1?
NVS148	CRF07/CRF08/C	P	C	Y	C	Y	C	Y	C
NVS152	D	Y	D	Y	D	Y	D	Y	D
	Summary of Concordance								
		Y-17 P- 38 N- 3		Y- 46 P - 8 N -4		Y- 41 P-0 N-17		Y-49 P-9	

b) Analysis of concordance for the Pol RT region

Patient ID	NCBI	C	Stanford	C	REGA	C	jpHMM	C	G2P	C	COMET	C	Squeal	C	Phylogen
HND200	A1,CRF01	P	CRF01_AE (94.4)	P	A1 (96)	Y	A1	Y	A1 (100%)	Y	A1	Y	A1	Y	A1-KE
KAH004	CRF10/A1/U/CRF16	N	D (90.4)	P	N/A	Y	A1D	N	10_CD (100)	N	unassigned_2;02_AG, 10_CD	Y	Complex	Y	Unassigned
KAH010	A1/CRF01/CRF15	P	CRF01_AE (92.6)	P	A1 (97)	Y	A1	Y	A1 (100%)	Y	A1	Y	A-ancestral	P	A1-KE
KAH016	A1/CRF01/CRF15	P	CRF01_AE (93.4)	P	A1 (91)	Y	A1C	P	A1 (100%)	Y	unassigned_1;C, A1	P	Complex	N	A1-KE
KHC016	A1/CRF15/G	P	CRF01_AE (93.1)	N	A1 (93)	N	A1	N	A1 (100%)	N	A1	N	A-ancestral	N	G or AG
KHC025	D/CRF10	P	D (94.2)	Y	D (73)	Y	D	Y	D (100)	Y	D	Y	A4,D recombinant	P	D
KHC040	C	Y	C (93.2)	Y	C (100)	Y	C	Y	C (100)	Y	C	Y	C	Y	C
KHC045	A1/CRF15/CRF01	P	CRF01_AE (92.5)	P	A1 (98)	Y	A1	Y	A1 (100%)	Y	A1	Y	Complex	N	A1-UG
KHC048	A1	Y	A (92.5)	Y	A1 (97)	Y	A1	Y	A1 (100%)	Y	A1	Y	A1	Y	A1-KE
KHC106	A1/CRF01/CRF15	P	CRF01_AE (93.2)	P	A1	Y	A1	Y	A1	Y	A1	Y	A1	Y	A1-KE
KHC160	A1	P	CRF01_AE (93.3)	P	A1 (99)	Y	A1	Y	A1 (100%)	Y	A1	Y	A1	Y	A1-KE
KHC163	A1/CRF01/CRF16	P	CRF01_AE (93.1)	P	A1 (100)	Y	A1	Y	A1 (100%)	Y	A1	Y	A1	Y	A1-KE
KHC093	A1	Y	A (93.3)	Y	A1 (100)	Y	A1	Y	A1 (100)	Y	A1	Y	A1	Y	A1-KE
KMB039	A2/A1/CRF16	P	CRF01_AE (92.3)	N	A1 (94)	N	A1A2	P	A2 (100%)	Y	unassigned_1;A2, A1	P	A,A2 recombinant	P	A2
KMB067	A1/CRF01/CRF15	P	A (92.5)	Y	A1 (98)	Y	A1	Y	A1 (100%)	Y	A1	Y	A1	Y	A1-UG
KMB145	A1/CRF15	P	CRF01_AE (93.7)	P	A1 (97)	Y	A1	Y	A1 (100%)	Y	A1	Y	A1	Y	A1-KE
KMB160	A1/CRF15/CRF02	P	CRF01_AE (91.6)	P	A1 (98)	Y	A1	Y	A1 (100%)	Y	A1	Y	A1	Y	A1-UG
KMB077	CRF10/A1/CRF15	N	D (92.4)	N	NA	N	D	N	10_CD	P	unassigned_1;39_BF, A1	N	Complex	Y	C?
MLD011	D/CRF10	P	D (94.9)	P	D (100)	P	D	P	10_CD	Y	D	P	D	P	10_CD
MLD183	A1/H/CRF05	P	A (91.6)	Y	N/A	N	A1/H	P	A1 (100)	Y	unassigned_1;H, A1	P	A1,D recombinant	P	A1-KE
MLD185	A1/CRF15/CRF02	P	CRF01_AE (93.7)	P	A1 (93)	Y	A1	Y	A1 (100%)	Y	A1	Y	A1	Y	A1-KE
MLD191	D/CRF10	P	D (94.3)	Y	D (89)	Y	D	Y	D (100)	Y	D	Y	D	Y	D
MLD198	A1/CRF01	P	A (93.7)	Y	A1 (100)	Y	A1	Y	A1	Y	A1	Y	A-ancestral	Y	A1-KE
MLD021	CRF10/D/CRF05/B	P	D (100)	Y	D (100)	Y	D/K/D	P	10_CD	P	D	Y	A1,D recombinan	P	D

Patient ID	NCBI	C	Stanford	C	REGA	C	jpHMM	C	G2P	C	COMET	C	Squeal	C	Phylogen
													t		
MLD245	CRF16/A2/B	P	B (91.4)	N	N/A	N	A2/B	P	A2 (100%)	P	B (check for 16_A2D)	P	CRF16-like	P	A2?
MLD040	A1/CRF01/A2/C	P	CRF01_AE (93.5)	N	N/A	N	C/A1	P	A1 (100)	Y	unassigned_1;C, A1	P	A1,C recombinant	P	A1-KE
MLD541	CRF15/CRF01/U	P	CRF01_AE (92.4)	Y	N/A	N	A1/K	N	CRF01_AE	Y	A1	P	A-ancestral	N	01_AE
MLD545	A1	Y	A (93.0)	Y	A1 (100)	Y	A1	Y	A1 (100)	Y	A1	Y	A1	Y	A1-KE
MLD548	D/CRF10/A1	N	D (93.8)	N	N/A	Y	D	N	D (100)	N	unassigned_1;D, A1	Y	U	Y	Unassigned
MLD060	C	Y	C (94.1)	Y	C (100)	Y	C	Y	C (100)	Y	C	Y	C	Y	C
NVS032	A1/CRF15/CRF01	P	CRF01_AE (92.2)	P	A1 (100)	Y	A1	Y	A1 (100%)	Y	A1	Y	A,A1 recombinant	P	A1-KE
NVS047	A1/CRF01	P	CRF01_AE (92.3)	P	A1 (98)	Y	A1	Y	A1 (100%)	Y	A1	Y	A1	Y	A1-KE
NVS052	A1/CRF01/C	P	A (93.2)	Y	A1 (94)	Y	A1	Y	A1 (100%)	Y	A1	Y	Complex	N	A1-KE
NVS056	A1	Y	A (93.4)	Y	A1 (99)	Y	A1	Y	A1 (100%)	Y	A1	Y	A1	Y	A1-KE
NVS069	A1/CRF01/CRF15	P	CRF01_AE (91.7)	P	A1 (87)	Y	A1	Y	A1 (100%)	Y	A1	Y	A1	Y	A1-KE
NVS085	A1/CRF01/CRF15	P	A (93.4)	Y	A1 (100)	Y	A1	Y	A1 (100%)	Y	A1	Y	A,A1 recombinant	P	A1-UG
NVS090	A1/CRF01/CRF15	P	A (92.8)	Y	A1 (97)	Y	A1	Y	A1	Y	A1	Y	A1	Y	A1-KE
NVS095	A1/CRF01	P	A (91.9)	Y	A1 (98)	Y	A1	Y	A1 (100)	Y	A1	Y	A,A1 recombinant	P	A1-KE
NVS122	A1	Y	A (92.8)	Y	A1 (98)	Y	A1	Y	A1 (100%)	Y	A1	Y	A1	Y	A1-KE
NVS152	D/CRF10/C	P	D (91)	N	N/A	N	D	N	10_CD (100)	P	C	Y	C,D recombinant	P	C
NVS058	A1/CRF01/CRF09	P	A (92.3)	Y	NA	N	A1/C	P	A1	Y	unassigned_1;C, A1	P	Complex	N	A1-UG
	Summary of concordance														
	Y – 7		Y – 19		Y - 31		Y – 27		Y – 34		Y – 31		Y – 23		
	P – 31		P – 15		P – 1		P – 8		P – 4		P – 8		P – 12		
	N - 3		N - 7		N - 9		N - 6		N - 3		N - 2		N - 6		

Appendix 5: Coreceptor Usage and Number of Potential N-Glycosites

	Patient ID	Coreceptor	Total PNGs	NXT	NXS	NNXST
1	HND046	CCR5	**12**	8	4	0
2	HND164	CXCR4	**10**	7	3	0
3	HND177	CCR5	**8**	7	1	0
4	KAH001	CCR5	**12**	8	4	0
5	KAH004	CCR5	**6**	4	2	0
6	KAH008	CCR5	**12**	7	5	0
7	KAH009	CXCR4	**13**	11	2	0
8	KAH010	CCR5	**10**	7	2	1
9	KAH020	CCR5	**11**	6	5	0
10	KAH029	CCR5	**14**	10	4	0
11	KAH090	CCR5	**9**	6	3	0
12	KAH138	CXCR4	**12**	9	3	0
13	KAH139	CXCR4	**13**	8	5	0
14	KAH142	CCR5	**13**	9	3	1
15	KAH144	CCR5	**14**	9	4	1
16	KAH148	CCR5	**12**	9	3	0
17	KAH157	CCR5	**14**	9	5	0
18	KAH227	CXCR4	**12**	8	4	0
19	KHC013	CCR5	**13**	8	5	0
20	KHC016	CCR5	**11**	6	5	0
21	KHC039	CCR5	**8**	5	3	0
22	KHC040	CCR5	**12**	7	5	0
23	KHC045	CCR5	**6**	5	1	0
24	KHC048	CXCR4	**10**	6	4	0
25	KHC050	CCR5	**13**	10	3	0
26	KHC052	CCR5	**11**	9	2	0

	Patient ID	Coreceptor	Total PNGs	NXT	NXS	NNXST
27	KHC058	CCR5	**11**	7	4	0
28	KHC059	CCR5	**10**	4	6	0
29	KHC064	CCR5	**14**	8	6	0
30	KHC069	CXCR4	**10**	5	5	0
31	KHC160	CCR5	**8**	4	2	2
32	KHC163	CCR5	**10**	5	4	1
33	KMB031	CCR5	**7**	4	3	0
34	KMB039	CXCR4	**7**	6	0	1
35	KMB067	CCR5	**11**	6	4	1
36	MLD011	CCR5	**11**	7	4	0
37	MLD012	CCR5	**11**	7	4	0
38	MLD016	CCR5	**9**	7	2	0
39	MLD185	CCR5	**13**	8	5	0
40	MLD245	CXCR4	**11**	6	5	0
41	MLD258	CCR5	**12**	7	5	0
42	MLD301	CCR5	**7**	4	3	0
43	NVS008	CXCR4	**12**	8	4	0
44	NVS032	CCR5	**10**	7	3	0
45	NVS040	CCR5	**13**	9	4	0
46	NVS051	CXCR4	**13**	7	6	0
47	NVS056	CCR5	**13**	9	4	0
48	NVS074	CCR5	**11**	10	1	0
49	NVS085	CCR5	**8**	5	3	0
50	NVS090	CCR5	**12**	10	2	0
51	NVS095	CXCR4	**13**	9	4	0
52	NVS108	CCR5	**14**	9	5	0

	Patient ID	Coreceptor	Total PNGs	NXT	NXS	NNXST
53	NVS121	CXCR4	**9**	7	2	0
54	NVS132	CCR5	**11**	7	4	0
55	NVS135	CCR5	**9**	6	3	0
56	NVS148	CCR5	**13**	9	4	0
57	NVS152	CXCR4	**14**	9	5	0
Total			636	417	210	9

Appendix 6: Request for SSC and ERC waiver

THE DIRECTOR,

ITROMID, JKUAT,

P.O. BOX 62000-00200,

NAIROBI, KENYA.

3/2/2014

Dear Sir/Madam,

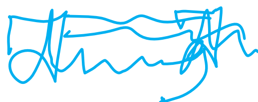
RE: MSc RESEARCH PROJECT FOR ROSE CHARI KITAWI; STUDENT REGISTRATION TM 306-0957/2012

The above named MSc. Medical Epidemiology Student is one of the students that I enlisted to complete her thesis project under my guidance. Rose's project is hived out of a larger study, SSC #2477, "Genetic Characterization of HIV-1 Strains and Determination of Known Mutations Conferring Resistance to Antiretroviral Drugs in Kenya". Both the Scientific Ethical committees of KEMRI already approve SSC #2477. Rose has started working on her part of the study, which is focused on generating the below data from Kisumu, Nakuru and Kajiado counties;

1. To determine circulating HIV strains and their genetic diversity in Kisumu, Nakuru and Kajiado Counties
2. To determine the patterns of HIV co-receptor usage in Kisumu, Kajiado and Nakuru Counties
3. To compare trends of HIV-1 and deduce circulating patterns (dispersal, transmission, spread) of the virus from Kisumu, Nakuru and Kajiado Counties.
4. To assess HIV-1 envelope diversity based on the patterns of Potential N-Glycosylation sites of different viral genotypes.

Please consider waiving any requirement that Rose's study be individually approved by the SSC and ERC based on the existing approval (enclosed) of the parent study to which hers is appended.

Sincerely



Washingtone Ochieng', Ph.D,
Centre for Virus Research, KEMRI

Appendix 7: ERC Approval

KENYA MEDICAL RESEARCH INSTITUTE

P.O. Box 54840-00200, NAIROBI, Kenya
Tel (254) (020) 2722541, 2713349, 0722-205901, 0733-400003; Fax: (254) (020) 2720030
E-mail: director@kemri.org info@kemri.org Website:www.kemri.org

KEMRI/RES/7/3/1 **April 15, 2013**

TO: **WASHINGTON OCHIENG (PRINCIPAL INVESTIGATOR)**

THROUGH : **DR. JUMA RASHID;**
DIRECTOR, CCR Forwarded 16/04/2013

RE: **SSC PROTOCOL NO. 2477 – REVISED (*RE-SUBMISSION*): GENETIC CHARACTERIZATION OF HIV-1 STRAINS CIRCULATING IN KENYA AND DETERMINATION OF KNOWN MUTATIONS CONFERRING RESISTANCE TO ANTIRETROVIRAL DRUGS**

Make reference to your letter dated 4th April, 2013, Received on April 4, 2013.

We acknowledge receipt of the Revised Study Protocol.

This is to inform you that the Ethics Review Committee (ERC) reviewed the document listed above and is satisfied that the issues raised at the 211th meeting held on 26th November, 2012 have been adequately addressed.

The study is granted approval for implementation effective this **15th day of April 2013**. Please note that authorization to conduct this study will automatically expire on **April 14, 2014**. If you plan to continue with data collection or analysis beyond this date, please submit an application for continuing approval to the ERC Secretariat by **March 4, 2014**.

Any unanticipated problems resulting from the implementation of this protocol should be brought to the attention of the ERC. You are also required to submit any proposed changes to this protocol to the ERC prior to initiation and advise the ERC when the study is completed or discontinued.

You may embark on the study.

Sincerely,

EAB

Dr. Elizabeth A. Bukusi,
ACTING SECRETARY,
KEMRI/ETHICS REVIEW COMMITTEE

Appendix 8: SSC Approval

KENYA MEDICAL RESEARCH INSTITUTE

P.O. Box 54840-00200, NAIROBI, Kenya
Tel (254) (020) 2722541, 2713349, 0722-205901, 0733-400003; Fax: (254) (020) 2720030
E-mail: director@kemri.org info@kemri.org Website:www.kemri.org

ESACIPAC/SSC/101264

13th December, 2012

Bernhards Ogutu

Thro'
Director, CCR
NAIROBI

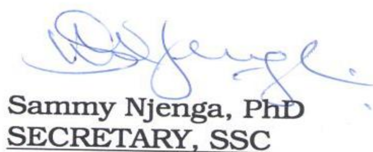
Forwarded 17/12/2012

REF: SSC No. 2477 (Revised) – Genetic characterization of HIV-1 strains circulating in Kenya and determination of known mutations conferring resistance to antiretroviral drugs

Thank you for your letter received on 7th December, 2012 responding to the comments raised by the KEMRI SSC.

I am pleased to inform you that your protocol now has formal scientific approval from SSC.

The SSC however, advises that work on the proposed study can only start after ERC approval



Sammy Njenga, PhD
SECRETARY, SSC

In Search of Better Health