HIV-1 SUBTYPES AND CORECEPTOR USAGE IN A POPULATION FROM NORTHERN KENYA

MURIUKI JOHN MARTIN MWANGI (B.Sc.)

I56/11255/2004

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE (IMMUNOLOGY) IN THE SCHOOL OF PURE AND APPLIED SCIENCES OF KENYATTA UNIVERSITY

JUNE 2011
DECLARATION

I John Martin Mwangi Muriuki declare that this thesis is my original work and has not been presented for a degree or other award in other university or any other award.

Signature.................................. Date...................................

This thesis is submitted for examination with our approval as the university supervisors.

Dr. Michael M. Gicheru
School of pure and applied sciences
Department of Zoological sciences
Kenyatta University
Signature.................................. Date...................................

Prof. Zipporah Ng'ang'a
Institute of Tropical Medicine and infectious Diseases
Jomo Kenyatta University of Agriculture and Technology
Signature.................................. Date...................................

Dr. Samoel A. Khamadi
Kenya Medical Research Institute
KEMRI/ Walter Reeds HIV Laboratory, Kericho
Signature.................................. Date..................................
DEDICATION

To the graduate students in the third world countries who struggle with studies without sponsorship and all the students whose academic dreams are elusive due to lack of finances.
ACKNOWLEDGEMENTS

I am grateful for the support of my supervisors Dr. Michael M. Gicheru of Kenyatta University, Prof. Zipporah Ng'ang'a formerly of Kenyatta University but now of Jomo Kenyatta University Kenyatta University of Agriculture and Technology and also Dr. Samoe A. Khamadi of Kenya Medical Research Institute, KEMRI/ Walter Reeds Project HIV laboratory, in development and execution of this Project. Gratitude to the staff of Centre of virus Research, especially the HIV laboratory, for their cooperation and support during the period of learning and execution of the laboratory assays.

I would also wish to acknowledge the moral and financial support from my parents David and Jennifer Muriuki without which this work would have been impossible. Special thanks to my brothers Ken, Tony, and Peter for their prayers and encouragement. I am greatly indebted to my fiancée Sarah for her continuous support, patience and encouragement throughout the project.

This project was conducted under the auspices of Kenya Medical Research Institute (KEMRI).
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Declaration</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dedication</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>iv</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>Abbreviations and Acronyms</td>
<td>x</td>
</tr>
<tr>
<td>Abstract</td>
<td>xii</td>
</tr>
</tbody>
</table>

1. INTRODUCTION

1.1 Background information
   1.1.1 Global prevalence and impacts of HIV 3

1.2 Study justification 6

1.3 Problem statement 7

1.4 Research questions 8

1.5 Null hypothesis 8

1.6 Objectives
   1.5.1 General objective 9
   1.5.2 Specific objectives 9

2. LITERATURE REVIEW

2.1 Structure of Human Immunodeficiency Virus 10

2.2 Classification of HIV 14
   2.2.1 HIV-1 subtype determination 15
   2.2.2 Distribution of HIV subtypes 16

2.3 Life cycle of HIV 17

2.4 Pathogenesis of HIV 19

2.5 Immune responses to HIV 23

2.6 Chemokine and chemokine receptors 27
2.6.1 Classification and nomenclature of chemokine and chemokine receptors  
2.6.2 Physiological and inflammatory roles of the chemokine system  
2.6.3 Immune subversion through the chemokine system  

2.7 Chemokine receptors as HIV co receptors  
  2.7.1 Coreceptor usage in Disease pathogenesis and transmission  
  2.7.2 Relationship between coreceptor switch and the immune profiles  
  2.7.3 Relationship between HIV-1 subtypes and coreceptor usage  
2.8 Determination of viral coreceptor usage  

3. MATERIALS AND METHODS  
3.1 Study Population  
  3.1.1 Ethical clearance  
  3.1.2 Inclusion and exclusion criteria  
  3.1.3 Sample size determination  
3.2 Sample collection and processing  
  3.2.1 Sample screening and handling  
  3.2.2 Extraction of Deoxyribonucleic Acid (DNA)  
  3.2.3 Nested Polymerase Chain Reaction (PCR)  
  3.2.4 Deoxyribonucleic Acid (DNA) sequencing  
  3.2.5 Deoxyribonucleic Acid (DNA) sequence analysis  
  3.2.6 Determination of HIV-1 subtypes  
  3.2.7 Determination of HIV-1 coreceptor usage  
3.3 Data analysis
LIST OF TABLES

Table 2.1: Classification of chemokines and chemokine receptor families 29
Table 4.1: Distribution of circulating recombinant forms in Northern Kenya 51
Table 4.2: Coreceptor usage by HIV-1 Variants in Northern Kenya 52
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Basic structure of Human Immunodeficiency Virus</td>
<td>10</td>
</tr>
<tr>
<td>2.2</td>
<td>Genome map of Human Immunodeficiency Virus type 1</td>
<td>11</td>
</tr>
<tr>
<td>2.3</td>
<td>Life cycle of Human Immunodeficiency Virus type 1</td>
<td>18</td>
</tr>
<tr>
<td>4.1</td>
<td>Distribution of the HIV-1 subtypes circulating in Northern Kenya</td>
<td>50</td>
</tr>
<tr>
<td>4.2</td>
<td>Frequencies of coreceptors usage among different HIV-1 genetic subtypes</td>
<td>54</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>ABI</td>
<td>Applied Biosystems International</td>
<td></td>
</tr>
<tr>
<td>ADCC</td>
<td>Antibody Dependent Cell-mediated Cytotoxicity</td>
<td></td>
</tr>
<tr>
<td>AIDS</td>
<td>Acquired Immune Deficiency Syndrome</td>
<td></td>
</tr>
<tr>
<td>ANC</td>
<td>Antenatal clinics</td>
<td></td>
</tr>
<tr>
<td>CCR5</td>
<td>Beta Chemokine Receptor 5</td>
<td></td>
</tr>
<tr>
<td>CD4</td>
<td>Cluster of Differentiation 4</td>
<td></td>
</tr>
<tr>
<td>CD4+ T cells</td>
<td>Cluster of Differentiation 4 Positive T cells.</td>
<td></td>
</tr>
<tr>
<td>CD8</td>
<td>Cluster of Differentiation 8</td>
<td></td>
</tr>
<tr>
<td>CD8+ T cells</td>
<td>Cluster of Differentiation 8 Positive T cells</td>
<td></td>
</tr>
<tr>
<td>CDNA</td>
<td>Complementary Deoxyribonucleic Acid</td>
<td></td>
</tr>
<tr>
<td>CVR</td>
<td>Centre for Virus Research</td>
<td></td>
</tr>
<tr>
<td>CXCR4</td>
<td>Alpha Chemokine Receptor 4</td>
<td></td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
<td></td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylene Diamine Tetra-acetic Acid</td>
<td></td>
</tr>
<tr>
<td>env</td>
<td>Envelop gene region</td>
<td></td>
</tr>
<tr>
<td>gag</td>
<td>Group-specific antigen gene region</td>
<td></td>
</tr>
<tr>
<td>gp</td>
<td>Glycoprotein</td>
<td></td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
<td></td>
</tr>
<tr>
<td>HIV-1</td>
<td>Human Immunodeficiency Virus Type-1</td>
<td></td>
</tr>
<tr>
<td>HIV-2</td>
<td>Human Immunodeficiency Virus Type-2</td>
<td></td>
</tr>
<tr>
<td>HTA</td>
<td>Heteroduplex Tracking Assay</td>
<td></td>
</tr>
<tr>
<td>KEMRI</td>
<td>Kenya Medical Research Institute</td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>M viruses</td>
<td>Major viruses</td>
<td></td>
</tr>
<tr>
<td>NSI</td>
<td>Non-Syncytium Inducing</td>
<td></td>
</tr>
<tr>
<td>O viruses</td>
<td>Outlier viruses</td>
<td></td>
</tr>
<tr>
<td>PBMCs</td>
<td>Peripheral Blood Mononuclear Cells</td>
<td></td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
<td></td>
</tr>
<tr>
<td>pol</td>
<td>Polymerase gene region</td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td>Chemokine Receptor-5 Using Variants</td>
<td></td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>Syncytium Inducing</td>
<td></td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetic acid EDTA</td>
<td></td>
</tr>
<tr>
<td>V3</td>
<td>Third variable region</td>
<td></td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
<td></td>
</tr>
<tr>
<td>X4</td>
<td>Chemokine Receptor-4 Using Variants</td>
<td></td>
</tr>
</tbody>
</table>
ABSTRACT

Human immunodeficiency virus type 1 (HIV-1) phenotype variability plays an important role in HIV-1 transmission and pathogenesis of Acquired Immune Deficiency Syndrome (AIDS). The basis of heterogeneity is known to be the differential use of chemokine receptors as coreceptors for viral entry. Beta chemokine receptor 5 (CCR5) using variants are associated with acute infections, macrophage tropism, non-syncytium inducing (NSI) phenotype and slow progression to AIDS. Alpha chemokine receptor 4 (CXCR4) using variants evolve later in infection in about 50% of the patients and are associated with T cell line tropism, syncytium induction, accelerated T cell depletion and rapid progression to AIDS. Determinants of coreceptor usage and biological phenotypes are mapped in the third variable region (V3) of the gp120 envelope gene. The aim of this study was to determine viral coreceptor usage in a population of HIV-1 infected individuals in Northern Kenya and investigate presence of relationship with the specific HIV-1 subtypes. In this study, 135 whole blood samples were collected from HIV-1 positive patients in Northern Kenyan. Erythrocytes were lysed and proviral DNA extracted from lymphocytes using DNeasy reagent and ethanol precipitation. The V3 region was amplified by nested PCR using C2-V3 region specific primers. Amplicons were sequenced directly using BigDye technology (sequencing kit from Applied Biosystems) to generate gp120 sequences that were used to determine the HIV-1 subtype after comparison with the Los Alamos reference sequences. Three circulating HIV-1 subtypes were observed in the following proportions: Subtype A1 (54%), C (27%), D (10%) and 9% of these Subtypes recombinant Forms. The properties of these sequences, amino acid on certain key positions and their net charge, were analyzed by bioinformatics tools to determine coreceptor usage. CXCR4 usage was most predominant (69.2%) compared to CCR5 (21.8%) and dual coreceptor (9.0%) usage. Most infected patients were inferred to be in late stage of infection based on coreceptor usage data. Coreceptor usage was observed to be strongly associated with HIV-1 subtypes. From the observations in this study, specific HIV-1 subtypes needs to be taken into consideration during HIV-1 management and interpretation of coreceptor usage data. Sequences generated from HIV-1 strains circulating in Northern Kenya indicates high genomic divergence hence need to consider a cocktail of antigen in vaccine development and need to develop genomes analysis software based on locally circulating HIV-1 subtypes.
CHAPTER ONE: INTRODUCTION

1.1 Background Information

Human immunodeficiency virus (HIV) is the causative agent of Acquired Immune Deficiency Syndrome (AIDS; Callaway et al., 1999). HIV is a single stranded diploid positive sense RNA virus about 100-120 nm in diameter (Roitt et al., 2001). Its surface structure is made up of the host derived viral membrane and the surface glycoprotein (gp160). The HIV-1 160kDa envelope precursor (gp160) is cleaved by cellular protease into two sub units: the membrane spanning gp41 and the soluble gp120 (Wyatt et al., 1998; Earl et al., 1990; 1991). The internal structure is composed of a genome and the structural proteins. The basic gene structure has the group-specific antigen gene (gag) that encodes the core protein, polymerase gene (pol) that codes for the enzyme that participates in integration and replication (Reverse transcriptase, Integrase and Protease), and the envelope gene (env) that encode the envelope proteins/surface glycoprotein (Lori et al., 2005). Additional genes regulate the viral protein synthesis (Wilhelm et al., 2000). These transactivation and accessory genes products (tat, rev, vif, vpu, vpr, and nef.) are essential for HIV replication and pathogenesis.

HIV infection occurs when contaminated body fluids come into contact with cuts and abrasions on the skin or mucosae, mainly during sexual intercourse (Sehga et al., 2005). Mother to child transmission during pregnancy, childbirth and lactation is also very common (Kwick et al., 2005). The virus is also transmitted through use of contaminated needles, blades, toothbrushes and surgical equipments (Bobkov et al., 2005; Gerst et al., 1990). While in developed countries HIV is primarily spread by homosexual contact and
intravenous drug use (Braine et al., 2005). In developing countries including sub Saharan Africa HIV is spread primarily by heterosexual contacts, with mother to child transmission during pregnancy, child birth and lactation being important modes as well (van’t Hoog et al., 2005). Female commercial sex workers (prostitutes) are a significant core transmitter group for HIV and sexually transmitted diseases (Plummer et al., 1991). The infectivity of HIV infected persons is increased by presence of sexually transmitted diseases (STI) such as gonorrhea, chancroids and herpes that cause ulcerations of the genital tract and lead to increased viral shedding (Gadkari et al., 1998; Wright et al., 2001). Less frequently blood and blood products carrying the virus are infused into recipients (Spada et al., 2005).

Human Immunodeficiency Virus type-1 uses CD4 and a chemokine receptor to mediate membrane fusion and entry into the target cells (Berger et al., 1999). The chemokine receptors CCR5 and CXCR4 are the main HIV-1 coreceptors (Zhang et al., 1998a). Non Syncytium inducing (NSI) or macrophage tropic viruses mainly use CCR5 (R5 viruses), while syncytium Inducing (SI) or T cell-line tropic viruses preferentially use CXCR4 (X4 viruses; Alkhatib et al., 1996; Feng et al., 1996; Deng et al., 1996; Dragic et al., 1996). Early in the infection, the R5 viruses predominate, whereas X4 viruses are detected in later stages in approximately 40-50% of patients (Zhu et al., 1993; Schuitemaker et al., 1992; Connor et al., 1997; Koot et al., 1993). The detection of X4 viruses has been associated with rapid progression of disease (Scarlatti et al., 1997). Infections with dual tropic (R5/X4) viruses have also been associated with T cell depletion and disease progression (Yu et al., 1998b).
Subtype-specific differences in coreceptors usage have been controversial (Zhang et al., 1996b; Tscheming et al., 1998). A possible correlation between a specific genetic subtype and biological phenotypes has been studied (de Wolf et al., 1994; Tscheming et al., 1998; Bjourndal et al., 1999). While, the D subtype has been shown to exhibit SI phenotype more frequently than the other subtypes, (de Wolf et al., 1994; Tscheming et al., 1998) the C subtype express mainly one phenotype, NSI (Tscheming et al., 1998; Bjourndal et al., 1999). Despite exhibition of only NSI phenotype in C subtype, C subtype infected individuals’ progress to AIDS in a similar fashion as individuals infected by other subtypes. In subtype C disease progression might therefore be explained by factors other than coreceptors usage and the emergence of SI phenotype.

1.1.1 Global Prevalence and impacts of HIV

HIV is a global public health concern, 33 million people have died in two decades and 40 million more people are infected (www.unaids.org). The number of people living with HIV worldwide continued to grow in 2008, reaching an estimated 33.4 million [31.1 million–35.8 million]. The total number of people living with the virus in 2008 was more than 20% higher than the number in 2000, and the prevalence was roughly threefold higher than in 1990 (UNAIDS 2009). The continuing rise in the population of people living with HIV reflects the combined effects of continued high rates of new HIV infections and the beneficial impact of antiretroviral therapy (UNAIDS 2009). As of December 2008, approximately 4 million people in low and middle income countries were receiving antiretroviral therapy, a ten fold increase over five years (WHO, UNCF, UNAIDS, 2009) Sadly, 2.7 million more people were reported to have been newly
infected and an estimated 2 million AIDS related illness deaths worldwide in 2008 (UNAIDS 2009).

Worst hit is sub-Saharan Africa with 22.4 million people infected; HIV/AIDS has destroyed communities, health care systems and cast a shadow upon the future of entire countries (www.who.int/entity). Sub-Saharan Africa has just over 10% of the world’s population, but is home to about 67% of all people living with HIV ((UNAIDS 2009). Three quarters (1.4 million) of all the AIDS deaths occurred in sub Sahara Africa and an estimated 1.9 million people in the region became newly infected in 2008. The rate of infection and the number of death is however estimated to have declined with 25% and 19% respectively (UNAIDS 2009).

In Kenya, the estimated HIV prevalence rates in the year 2005 among those aged between 15-49 years was 6.7% with a total of approximately 1,200,000 adults and children living with HIV/AIDS, 1,100,000 of them being adults, 720,000 women (www.unaids.org). In 2007, the Kenya AIDS Indicator Survey (KAIS) reported a prevalence of 7.8% for Kenyans aged between 15-49 years, an estimate of 2.2 million Kenyans living with HIV-1 infection (Kiptoo et al., 2009). Wide variations in HIV prevalence and epidemiological patterns within the country are apparent based on factors like Geographical location, sex, age, residence, marital status and level of education.

The distribution of HIV in Kenya varies greatly across the country. The HIV prevalence ranges from 0.8% in North Eastern province to 14.9% in Nyanza province (NASCOP 2008). Moreover, a higher proportion of women age 15-64 (8.7%) than men (5.6%) are infected with HIV that means three out of five HIV infected Kenyans are female.
Differences in prevalence across the life span have been reported where among youth aged between 15-24 years women are four times likely to be infected than men. A higher proportion of Kenyans aged 30-34 years are currently infected with HIV than any other age group and there is a decline in infection in women above the age of 34 years and men above age 44 years (NASCOP 2008). About three quarters of Kenyans live in rural areas of the country. Among those ages 15-64, 7 percent are infected with HIV while in urban areas, the prevalence is 9 percent. Marital status is an important risk factor with those in polygamous union, cohabiting, widowed, divorced and separated partners having high HIV prevalence. Men and women with higher education levels have lower HIV prevalence than those with less education (NASCOP, 2008).

The epidemic brings additional pressure to bear in the health sector. As the epidemic matures, the demand for care for those infected and the toll among the health workers is on the increase (www.avert.org/hivtreatment.htm). The impact of HIV/AIDS on household income is devastating with members diverting most time from income generating activities (Bechu, 1998) to take care of the infected and also to attend funerals. The coping strategies adopted among the affected households include savings being used up or assets sold and assistance received from other households.

The composition of households tends to change, with fewer adults of prime working age in the household. The burden of coping rest on women as there is increased demand for their income earning labor, household work, childcare and care for the sick. As men fall ill, women, often have to step into their roles outside the home (Kaiser, 2002). The trauma and hardship that children affected by HIV/AIDS are forced to bear world wide is
hard to be overemphasized. Not only do children lose their parents and guardians, but sometimes it means they lose their childhood as well. By 2003, 15 million children below 18 years of age had been orphaned by HIV/AIDS world wide. About 12 million of these live in sub-Saharan Africa (www.avert.org/aidsorphan.htm). As parents and family become ill, children take more responsibility to earn an income, produce food and care for the sick family members. A decline in school enrolment is one of the most visible effects of the epidemic (www.avert.org/aidsorphan.htm).

1.2 Study Justification

The HIV-1 Syncytium Inducing Phenotype and Coreceptor Usage are strong predictors of disease progression and much effort has been put to prediction of these phenotypic traits of HIV-1 (Koot et al., 1993; Schellekens et al., 1992; Tersmette et al., 1989). The bases of non Syncytium Inducing (NSI) and syncytium Inducing (SI) phenotypes have been shown to be their differential use of chemokine receptors for viral entry (Jensen et al., 2003a). However, because of relatively few studies, information about the change in coreceptor usage during infection in Africa HIV-1 patients is still limited (Begaud et al., 2003). With the finding of specific mutations associated with a switch in coreceptor usage, and increasing availability of sequence corresponding to experimentally determined phenotypes it is feasible to determine the biological phenotype using envelope sequence data (Jensen et al., 2003b). This study was meant to determine the relationship between HIV-1 subtypes and viral coreceptor usage in a population of HIV-1 infected patients in Northern Kenya by analysis of the gp120 V3 loop gene sequences of the proviral DNA extracted from HIV-1 infected individuals in Northern Kenya.
With no study on coreceptor usage or HIV prevalence in Northern Kenya, this study investigated the possibility of any variation in coreceptor usage switch from the documented pattern (Zhu et al., 1993; Schuitemaker et al., 1992; Connor et al., 1997; Koot et al., 1993) in relation to the circulating subtypes. Determination of viral phenotypes is also essential before antiretroviral therapy, potentially for making treatment decisions (Saliou et al., 2011) and for analysis of disease progression during highly active antiretroviral therapy (HAART; Ruiz-Metaos et al., 2004). Moreover, with the emerging strategy of antiretroviral therapy being based on inhibition of HIV-1 by blocking the host cell coreceptors, CCR5 and CXCR4 (Saliou et al., 2011, Strizki et al., 2005), with several promising inhibitors being developed, this study was vital for determination of the appropriate therapeutic candidates in Northern Kenya. While these antiretroviral drugs aim at preventing virus/coreceptor interaction by binding to host proteins, neutralizing antibodies directed against gp120 have attracted attention as possible vaccine candidates (Sirois et al., 2005). Thus, advances in rational drug and vaccine design rely heavily on improved insight into the relationship between genotype and phenotype, the evolution of coreceptor usage and ultimately, the structural biology of coreceptor usage and inhibition. The recent development of drugs that block the entry of HIV-1, such as CCR5 antagonists, has made it necessary to accurately determine HIV-1 coreceptor usage (Saliou et al., 2011). The gp120 V3 loop involved in these aspects (Sirois et al., 2005) was the major focus of this study.

1.3 Problem Statement

Determinants of coreceptor Usage and biological Phenotypes are mapped in the third variable region (V3) of the gp 120 envelop gene. Although it is possible to determine
coreceptor usage through analysis of the V3 region gene sequences and the viral subtypes through analysis of the gp 120 gene sequences, there is neither information on coreceptor usage among the Northern Kenya HIV-1 infected individual nor the relationship of coreceptor usage with the circulating HIV-1 variants. Moreover, to alleviate the HIV-1 disease burden, rational drugs and vaccine development rely on improved insight on the genome of the circulating variants and variation at different stages of infection whose information is limited in the Northern Kenya HIV-1 infected individuals.

1.4 Research Questions

i. Based on gp120, HIV-1 sequence analysis, what are the subtypes of HIV-1 circulating in northern Kenya?

ii. Which chemokine receptors are used by HIV-1 variants in a population of HIV-1 infected individuals in Northern Kenya?

iii. At what stage of infection are the Northern Kenyan HIV-1 infected individuals based on the coreceptor usage data?

iv. How does coreceptor usage differ among HIV-1 subtypes circulating in Northern Kenya?

1.5 Null Hypothesis

There is no difference in coreceptor usage among different HIV-1 subtypes from a population of HIV-1 infected individuals in Northern Kenya.
1.6 Objectives

1.6.1 General Objective

To determine the Circulating HIV-1 subtypes and coreceptor usage in a population of HIV infected individuals in Northern Kenya.

1.6.2 Specific Objectives

(a.) To generate the *gp120 V3* gene sequences of various HIV-1 variants from a population of HIV-1 infected individuals in Northern Kenya.

(b.) To determine the HIV-1 subtypes Circulating in Northern Kenya.

(c.) To determine the chemokine receptors that are used by the virus as entry coreceptors in a population of HIV-1 infected individuals in Northern Kenya.

(d.) To infer the stage of infection in the HIV-1 infected individuals in Northern Kenya using *gp120V3* loop sequences.

(e.) To determine whether coreceptor usage differs significantly among HIV-1 subtypes circulating in Northern Kenya.
CHAPTER TWO: LITERATURE REVIEW

2.1 Structure of Human Immunodeficiency Virus

The mature HIV virion is a roughly spherical (actually icosahedral) particle with a diameter of approximately 110 nm (Roitt et al., 2002, Baker et al., 1999). The host cell membrane derived lipid bilayer outer envelope, acquired during virion budding, (Nair, 2002) is studded with approximately 72 spikes formed by the 2 major viral envelope glycoproteins, gp120 and gp41 (Fig 2.1). The central core contains 4 viral proteins (p24—the major capsid protein, p17 a matrix protein, p9 and p7), two copies of RNA (to which p9 and p7 are bound), and three viral enzymes namely protease (P10), reverse transcriptase (P64) and integrase (P32; Fig 2.1). Reverse transcriptase is a heterodimer of p66/p51 sub unit that possess two enzymatic functions (reverse transcriptase activity and RNase activity) located on different domains of the protein (Wilhelm et al., 2000).

![Fig 2.1: Basic Structure of HIV (adopted from www.avert.org/virus.htm)](image-url)
The constituent genes of HIV-1 provirus encode the structural proteins (*gag*, *env* and *pol*), accessory proteins (*vpu*, *vpr*, *vif* and *nef*) and regulatory proteins (*rev* and *tat*). Human immunodeficiency virus type-2 has *vpx* instead of *vpu* (Fig 2.2).

![Fig 2.2: The genome map of HIV-1](http://en.wikipedia.org/wiki/image:hiv-genome.png)

The *gag* gene encodes the *pr55* precursor protein, further cleaved by protease to form matrix (*p17*), capsid (*p24*), nucleocapsid (*p9*) and *p6*. The capsid forms the core of the virus particle. The nucleocapsid is essential for recognition of the HIV packaging signal via the two zinc finger motifs necessary for inclusion of the heterologous viral RNA into HIV enzymes. The nucleocapsid also assists in reverse transcription. The *p6* protein facilitates the inclusion of *vpr* into virus particle. This contains a late domain necessary for viral budding (Yu *et al.*, 1998a).

The proteins coded by *pol* are only produced in conjunction with the *gag* proteins. This occurs due to a specific *cis-acting* motif that triggers frame-shifting events causing translation to proceed beyond the stop codon and continue in a different reading frame. The ribosomes shift toward the *pol* without stopping translation (Jacks *et al.*, 1988). Self-cleavage of the resulting fusion protein pr160 *gag-pol* leads to the release of HIV
enzymes: protease (p10), reverse transcriptase (p51), ribonuclease H (p15), and integrase (p32).

The env gene encodes the 160 KDa env poly-protein, which on cleavage yields the env protein gp120 (surface glycoprotein) and the gp41 (transmembrane glycoprotein). Gp120 and gp41 interact with each other non-covalently (Earl et al., 1990). Gp120 contains 5 domains that vary between HIV strains. The V3 loop is one of these five and is vital because it determines the range of cells that a strain can infect in a host as well as env interaction with the chemokine receptors CCR5 and CXCR4. Gp120 has a high affinity for CD4 receptors and surface glycoprotein DC-SIGN located on dendritic cells.

Tat and rev are transactivation genes. The tat gene encodes viral transactivator (tat protein) that belongs to a novel class of eukaryotic regulatory proteins (Sodroski et al., 1985). The tat protein binds a specific sequence near the 5' end of the nascent RNA and increases the stability of the polymerase which transcribes the viral genes, and permits the synthesis of full length transcripts to occur (Laspia et al., 1990; Marciniak et al., 1990). Tat may also increase the frequency of RNA initiation. The rev gene encodes rev protein, an 18Kda protein that binds to the rev response element present in all unspliced viral RNA transcripts. The bound rev protein regulates the amount of spliced and non-spliced mRNA as well as transport of viral RNA to the cytoplasm (Haseltine, 1991).

The accessory genes; nef, vpr, vif, and vpu are late regulatory genes that are not essential for viral replication but are vital for infectivity, budding, disease induction, T cell activation, and the inhibition of MHC class I. Negative factor (Nef) down regulates MHC class I and class II possibly providing an immunological escape for HIV (Mangasarian et al., 1990).

al., 1999; Stumptner-Cuvelette et al., 2001). Nef protein is also involved in downregulating of cell surface CD4 molecules and activation of T cell hence HIV spread (Garcia and Miller, 1991; Kestler et al., 1991; Miller et al., 1994; Simmons et al., 2000; Rhee and Marsh, 1994). Nef has been found to be essential for prolific replication of the virus in primary peripheral blood lymphocytes, primary CD4 T cells and primary monocytes and macrophage cultures (Terwilliger et al., 1991). Deletion of Nef leads to attenuation of SIV and HIV thus indicating it is crucial in pathogenesis (Novembre et al., 1996; Gulizia et al., 1997; Carl et al., 1999).

The product of virion infectivity gene, Vif, is made of singly spliced mRNA, which accumulates late in infection (Garrett et al., 1991). In the absence of Vif function, virus particles released from the cells are poorly infectious and the infected cells have decreased ability to transmit infections to other cells (Sodroski et al., 1986; Fisher et al., 1986; Strebel et al., 1987). The vif gene is needed for viral replication and maturation in the peripheral blood but the exact mechanism is not understood (Haseltine, 1991).

The Vpu gene encodes an 18KDa protein, viral protein-U that also assists in virus maturation (Cohen et al., 1989; Strebel et al., 1988). The Vpu protein facilitates the export of virus particles from the cells (Strebel et al., 1989; Terwilliger et al., 1989; Klimkiat et al., 1990). The Vpu may play an important role in the pathogenesis and transmission of HIV-1.

The Vpr gene encodes the viral protein R, a small protein packaged within the virus particle itself (Cohen et al., 1990; Yu et al., 1990). The vpr stops the proliferation of HIV infected cell during the G2 phase, which increases the time available for viral replication
(He et al., 1995). The vpr protein also speeds the replication of the virus (Ogawa et al., 1989; Cohen et al., 1990). In addition, vpr protein trans-activates a variety of viral and cellular promoters (Cohen et al., 1990).

2.2 Classification of Human Immunodeficiency Virus

There are different genotypic and phenotypic variants of HIV, the main variants being HIV-1 and HIV-2 whose difference in terms of transmissibility, pathogenesis and pattern of spread are well documented (De Cock et al., 1993; Kanki, 1994). HIV-2 is endemic in West Africa and appears to be less virulent (Mullins et al., 2004), while HIV-1, which is the most widely distributed and most studied, is endemic in other Sub-Saharan Africa regions and appears to be more virulent (Janssens et al., 1997).

HIV-1 can further be classified into subtypes/clades and strains. Since 1992, the env coding sequence has been used to classify globally prevalent viruses. The first five sequenced subtypes-A, B, C, D, and E (Myers et al., 1990)-thus identified and five more subtypes labeled F through J, of which J and I have been reported only once. These 10 subtypes constitute the group M (major) HIV-1 viruses. Again with each subtype, a plethora of HIV-1 variants exists (Janssens et al., 1997).

Some HIV-1 viruses isolated from Cameroonian patients were highly aberrant and could not be classified in any of the known group M subtypes, and were classified into a separate group of Outlier viruses; group O (Charneau et al., 1994).

Another HIV-1 subtype, group N, was suggested based on 2 strains separated from Cameroonian patients (Simon et al., 1998). HIV-1 group N infections are rare with only
eight documented cases. All group N infections have been found in Cameroon and there is no evidence of direct linkage between the infected patients (Yamaguchi et al., 2006).

2.2.1 HIV-1 Subtype Determination

There are several molecular techniques available for typing and subtyping of viruses. These include; Restriction Fragment Length Polymorphism (RFLP), Southern Blot Analysis, Oligonucleotide Fingerprint Analysis, Reverse Hybridization, DNA Enzyme Immunoassay, RNase Protection Analysis, Single Strand Conformation Polymorphism Analysis, Heteroduplex Mobility/ Tracking Assay (HMA/HTA), Nucleotide Sequencing, and Genome Segment Length Polymorphism Analysis (Arens, 1999). The experimental determination of the linear arrangement of the bases in a viral genome is the ultimate form of subtyping. Sequencing of a genome has the potential to distinguish even between parent and progeny if a single mutation has occurred during the replicative process (Arens, 1999). For example in a study conducted by Kitchen et al. (2004), sequence analysis using a Bayesian model was able to show that the R5 viruses present after the antiretroviral therapy arose form the predecessor R5 and not from the continuous evolution of the X4 strains.

Small portions or specific variable regions of viral genome can be sequenced for the purpose of subtyping. Here only the portion of genome that confers the subtype can be sequenced. In most cases the comparison will include only the regions that are subjected to immunological pressure by the host (i.e. major antigenic epitopes of the surface proteins). In other cases the un-translated region of the genome are unique to a subtype because they contain random mutations that may persist indefinitely in the complete absence of immunologic pressure. The use of PCR amplification in conjunction with the
recent development in automated sequencing has made it possible to obtain fully analyzed sequence of about one kilo base portion of a viral genome within about three days after receiving a clinical specimen (Arens, 1999). PCR has made it possible for amplification of minute amounts of DNA or RNA and then sequence directly or clone the product for sequencing.

In this study, direct sequencing procedure was used for subtyping of the HIV-1 positive isolates to determine the clades that were positive in the samples. The process involved amplifying a region of gp120 in the env region from the HIV-1 proviral DNA using specific primers designed from the C2V3 region (Takehisa et al., 1998). Generated amplicons were then sequenced directly.

2.2.2 Distribution of HIV Subtypes

The distribution of HIV subtypes around the globe varies in different parts of the world. Before 1992, HIV-1 strains were classified on the basis of their origin (Myers et al., 1990). The strains are now classified on the basis of their molecular differences (Janssens et al., 1997). Since HIV inter-subtype recombination occurs frequently, subtype naming should indicate the region of the gene analyzed. Most molecular epidemiology studies (Robbins et al., 1999) done in Kenya found majority of sequences analyzed to be subtype A (71-87%), subtype D (7-29%) and subtype C (7-17%). Full-length sequences indicates high levels of recombination especially in subtype C and D, with majority of non-recombinants being of subtype A (Dowling et al., 2002). In a recent study by Oyaro et al., (2010) analysis of the 122 sequences from Western Kenya on both gag and env regions, 51.4% were subtype A, 9.4% subtype D, 1.4% subtype C, 4.1% subtype G, and 33.7% were discordant and thus possible recombinants, including A1/C, A1/D, A1/A2,
and A2/C. In Northern Kenya, the circulating HIV-1 genetic subtypes include; subtype A (50%), subtype C (39%) and subtype D (11%). Among these subtypes high level of recombination (56%) was also reported (Khamadi et al., 2005).

2.3 Life Cycle of Human Immunodeficiency Virus.
Successful infection is facilitated when Human Immunodeficiency Virus (HIV) is introduced in the human body system by binding of viral gp120 to the host cell CD4 antigen (Klatzmann et al., 1984; Fig 2.3). This results in conformation changes in the virus envelope. The membrane fusion reaction results from conformational changes in the env that likely include the formation of a coiled coil structure with resulting exposure of the amino terminal fusion peptide in gp41 and the appropriate coreceptor, normally CCR5 or CXCR4. Additional conformational changes within the gp41/gp120 trimmer trigger the insertion of gp41 fusion peptide into the plasma membrane. This insertion results in the eventual fusion of the viral and cellular membrane and insertion of the viral core into the cellular plasma, a process referred to as uncoating (Haseltine, 1991; Wise et al., 2005; Fig 2.3).

Once the viral core has penetrated the cell membrane, the nucleocapsid is partially dissolved to release the virus RNA (Schim et al., 1999; Fig 2.3). The viral reverse transcriptase enzyme transcribes the single stranded RNA into a single stranded DNA copy and then into a viral DNA duplex (Muller and Varmus, 1994). The reverse transcriptase uses nucleotides from the cell cytoplasm as the building blocks for the viral DNA duplex (Wise et al., 2005).

The newly translated DNA is transported across the nuclear membrane into the nucleus where it is inserted into the host cell DNA. This process is called integration and is
facilitated by a virally coded enzyme integrase (Muller and Varmus, 1994). The viral DNA that is integrated and covalently linked to the host cell genome is called the HIV provirus (Miriano and Victor, 1997). Once integrated into the host cell genome as a proviral DNA, HIV can establish a virus producing infection or a latent infection depending on whether the infected cell is in activated or resting state respectively (Ho et al., 1987; Bagasra et al., 1992).

Fig 2.3: Life Cycle of HIV (Adopted from www.jyi.org/articleimage)
The latently infected cells provide a reservoir of HIV that is able to escape the immune defense and the antiretroviral treatment. Latently infected cells can be activated by antigens, mitogens, cytokines and also viral gene products to initiate transcription and translation of the HIV provirus DNA. This leads to production of necessary components for infectious virus progeny (Marsh and Thali, 2003). From the proviral DNA, strands of the RNA are transported out of the nucleus. These strands become the genetic material of the new viruses while some are processed and translated into viral subunits. The produced structural proteins are assembled with viral RNA to form new virions (Li and wild, 2005; Fig 2.3).

The viral protease enzyme cleaves and separates various viral subunits for the final assembly as new viruses (Ventose et al., 2005). The final step of virus life cycle is viral budding from the host cell. In this stage, the genetic materials enclosed in the nucleocapsid merges with the cells membrane to form the new viral envelope. With its genetic material tucked in its nucleocapsid and the new host cell membrane derived outer coat, the newly formed HIV virions pinch off and enter into circulation ready to start a new lifecycle (Li and wild, 2005; Fig 2.3).

2.4 Pathogenesis of Human Immunodeficiency Virus

HIV causes disease as a result of depletion of CD4+ T helper cells and a consequent profound immunosuppression hence inability to fight opportunistic infections (Kim et al., 1987; Roitt et al., 2001; Haynes et al., 1999). The HIV target cell lysis is not well understood but several different mechanisms have been proposed: accumulation of RNA and unintegrated DNA in the cell cytoplasm, and intracellular binding of CD4 and gp120; infected cells may bind to gp120-CD4 linkage, with multinucleate giant cell formation;
gp120 bound to the surface of uninfected CD4+ cell also make them vulnerable to antibody dependent cell mediated cytotoxicity (ADCC); while infected cells may be killed by gp120-specific cytotoxic T cells; HIV may act as super antigens resulting in vast expansion and then exhaustive depletion of cells; HIV may induce T cell apoptosis; and also viral budding may lead to cell membrane weakening and lysis (Blaak et al., 2000; Kim et al., 1987; Koot et al., 1993; Roitt et al., 2001).

Other cells tend to harbor and replicate the virus without lysis, in case of dendritic cells, they may concentrate viruses in the cell surfaces with little or no replication of the virus. Other cells infected by HIV are Natural killer cells, cytotoxic (CD8+) killer cells, and cells of the nervous system (Chehimi et al., 1991; Moses et al., 1994). Some CD4 negative cells like epithelial cells of vagina, rectum and intestines, endothelial cells of the brain capillary, and other cells of the central nervous system have been shown to be infected via galactocerebrosidase receptors (Bagasra et al., 1992). Macrophages also bind HIV gp120 via syndecan and via CD91 antigen (Sapphire et al., 2001).

The clinical course of HIV infection is highly variable among patients (Rodes et al., 2004). Some of the variations can be explained by genotypic differences among patient and differences in the capacity to elicit a protective immune response (Jensen et al., 2003). Studies suggest that progression to AIDS in Africa may be more rapid than in the industrialized countries (Morgan et al., 1997; 1998; Anzalla et al., 1995; Bwayo et al., 1995). This can be partially attributed to a 32 base pairs deletion mutation (delta 32) within the second extracellular loop encoding regions of the CCR5 gene, which result into a truncated and non-functional protein (Sheppard et al., 2002; Kalev et al., 2000; Oh et al., 2000; Kantor et al., 1999). Specifically, delta32/delta32 homozygotes are protected
against acquisition of HIV-1 by the mucosal route despite high-risk exposure, where as disease progression among CCR5/delta32 heterozygotes occurs more slowly. It can also be explained by the differences among the infecting virus (Jensen et al., 2003). There are different genotypic and phenotypic variants of HIV, the main variants being HIV-1 and HIV-2 whose difference in terms of transmissibility, pathogenesis and pattern of spread are well documented (De Cock et al., 1993; Kanki and De Cock, 1994).

A typical course of HIV infection begins with an acute HIV syndrome characterized by flu like illness and a burst of viraemia with a decrease in the number of CD4+ cells soon after the exposure (Pantaleo et al., 1993). Fever, fatigue, headache, and joint pains are common complains during this period. There is high level of virus shedding in the genitalia. The end of this phase is usually coincident with the first sign of an immune response against HIV (Koup et al., 1994; Safrit and Koup, 1995; Prince et al., 1997). Within a few days of infection, the cellular and humoral immunity begins to have an effect. The CD4 counts return to normal, viraemia decreases and a period of latency that differs in length between individuals begins.

An asymptomatic period follows primary infection which can vary from two to fifteen years or more years (Pantaleo and Fauci, 1996; Phair, 1994). Although no visible symptoms are present, the replication kinetics of the virus is extremely fast during this period (Ho et al., 1995; Wei et al., 1995; Perelson et al., 1996). Still, virus level change very little in this time and the immune system is thought to be the mechanism which controls the virus (Haynes et al., 1996; Mackewics et al., 1991; Haas et al., 1996; Goulder et al., 1997). Initially there are few if any clinical signs and symptoms, but
nonetheless in the absence of antiviral therapy the level of CD4 T cells continue to fall gradually (Pantalleo et al., 1993).

At the onset of AIDS (when the CD4+ T helper cell counts fall below 200 cells/µl) there is an overall weakness of the immune system which allows establishment of the opportunistic infections (Haynes et al., 1996). The median time for full blown AIDS is about 10 years but in some population can be as little as 4.4 years. Furthermore for a typical progressor, the decline to symptomatic AIDS is very rapid, with onset of symptoms and death in a few months. Long-term non progressors (LNTPs) are HIV infected people who have a low viral load, no disease and a stable or even increasing CD4+ cell counts over years (Easterbrook and Schrager, 1998). They represent less than 5% of all HIV infected people. Both host and viral factors appear to be important in determining the rate of disease progression (Clerici and Shear, 1996; Easterbrook and Schrager, 1998). Patients with certain CCR5 and CCR2 gene mutations progress slowly to AIDS (Huang et al., 1996; Smith et al., 1997). More frequently however, LNTPs seem to have vigorous HIV specific CD8+ and CD4+ T cell responses, and tend to produce high level of chemokines RANTES, MIP-1α and MIP-1β (Cocchi et al., 1995). The SDF1-3'A, which is a variant of the ligand for CXCR4, may also have protective effects against HIV infection (Winkler et al., 1998).

However, it appears that the HLA is the single most important factor in long-term non-progression. Patients with HLA-A1, HLA-A2, HLA-B14, HLA-B17 and HLA-B27, HLA-DR5, and HLA-DR6 are more likely to have a slow disease progression where as, those with HLA-B35, HLA-DR1, HLA-DR3 and HLA-DQ1 have a worse prognosis (Huang et al., 2011). One of the best determinants of disease progression remains the
baseline viral counts at the onset of an acute infection (Mellors et al., 1996; Duda et al., 2009). The viral set point determined by the interplay between the virus and the host immunity determine the clinical course. The higher the viral set point the worse the prognosis. In fact, the level of HIV RNA copies in blood is related to the clinical outcome independent of the CD4 T cell counts (Duda et al., 2009).

In addition to these factors, whether a particular person succumbs quickly to AIDS or not is dependent on age, concomitant infection, virulence of a particular strain of HIV, level of nutrition and psychological stress. The clinical progression of HIV disease correlates directly to the level of viraemia and inversely to the CD4+ T cell counts. In the course of infection the decrease in CD4 T cell counts have been attributed to the direct cytopathic effects of HIV sequestration in primary lymphoid organs and decreased production of new T cells (McCune, 2001). Throughout the infection the viral population is composed of a quasi species of strains that differ in, among other things, replication rate, cell tropism and ability to cause target cells to form large multinucleate bodies known as syncytia (Huang et al., 2011; Mellors et al., 1996).

2.5 Immune responses to HIV

The human immune system raises protective responses against HIV infection (Koup et al., 1994; Safrit and Koup, 1995). Small cohorts of commercial sex workers, some infants born to HIV positive mothers, some medical workers and various other small groups of people exposed to HIV show some resistance to HIV infection (Rowland-Jones et al., 1995; 1998). The exact mechanism of protection has not been established, but cell
mediated immune responses and secretory antibodies are believed to be important (Safrit and Koup, 1995; Koup et al., 1995; Rowland-Jones et al., 1995; 1998; Price et al., 1997; Marsh and Thali, 2003). Immunity against HIV is both innate and adaptive.

The innate immune system consists of both cellular and soluble mediators that recognize pathogens by specific patterns displayed on the surface antigens. These are; dendritic cells, macrophages, interferon (IFN) producing cells, Natural Killer cells, granulocyte, Natural Killer T-cells, B1 cell, cytokines, chemokines, complement mannose binding lectin, defensins and acute phase proteins. Innate immunity is non-specific, has limited antigen repertoire, lack capacity of memory and is activated within hours of antigen contact (Roitt et al., 2001). Immature dendritic cells known to be the interferon producing cells are vital in the control of HIV infection. Besides producing copious amounts of type 1 interferon, they mature to dendritic cells and enhance Th2 responses (Yonezawa et al., 2011). Interferons block HIV replication by activating a number of genes including a 67Kda protein kinase that inhibit Elongation factor 2 enzyme phosphorylation hence block protein translation; 2'5'-oligodenylate synthetase that activate latent rate endonuclease RNase-1 responsible for viral RNA degradation; Mx proteins that inhibit RNA virus transcription. Interferons also enhance MHC class 1 and B7 expression of antigen presenting cells (APCs), stimulate CD4+ T cells to produce IFNγ and promote Th1 type responses (Petryshyn et al., 1996). Dendritic cells phagocytose infected cells and produce large amounts of type 1 interferons which enhance Th2 responses.

Macrophages are professional antigens presenting cells (APCs); in addition they express co-stimulatory molecules CD80 which potentate adaptive immune responses. For viral infections including HIV, macrophages produce cytokines and chemokines that inhibit
the virus directly (such as IL-1, IFNα/β, TNFα, and nitric acid), activate virus specific cellular responses (IL-1, IL-12) and recruit inflammatory cells (MIP-1α/β, RANTES) (Meylan et al., 1993).

Natural killer (NK) cells are believed to eliminate HIV infected cells either directly or through antibody dependent cellular cytotoxicity. Going by their activity in other viral infections, they may produce GM-CSF, interferons and β-chemokines. The γδT cells recognize non-peptide antigens in mucosal surfaces and lyse HIV infected cells, produce antiviral chemokines and cytokines that enhance both Th1 and Th2 responses. B1 cells produce antibodies in recognizing polysaccharide antigens (Safrit and Koup, 1995).

Soluble factors in innate immunity bind to HIV directly or opsonise HIV infected cells to facilitate phagocytosis. Chemokines recruit cells to the sites of HIV infection. They determine whether the immune response will be predominantly Th1 or Th2 type. Chemokines can also inhibit HIV replication directly (Scarlatti et al., 1997; Cocchi et al., 1995). Cytokines produced by innate immune system stimulate the adaptive immune response. For example, stromal derived factor, α-chemokine, has chemotactic properties for T-lymphocytes and developmental role in B-lymphocyte maturation (Bleul et al., 1996; Oberlin et al., 1996). Moreover β-chemokines- RANTES, MIP-1α and MIP-1β-produced by CD8+ T lymphocytes are reported as potent inhibitors of HIV-1 mainly during asymptomatic stage (Scarlatti et al., 1997; Cocchi et al., 1995).

Immunity against viral infections, including HIV is mainly adaptive. Cytotoxic T cells, helper T cells and specific antibodies secreted by B cells are the major effectors of the adaptive immune response. Unlike innate, adaptive immunity is highly specific, has great
capacity of memory and has extensive range of antigens repertoire achieved through comprehensive gene rearrangements during B and T-cell development.

During HIV infection there is a strong antibody response to envelope proteins, but this response is ineffective against mature virions (Marsh and Thali, 2003; Safrit and Koup, 1995). Several studies have shown those antibodies are not enough to offer prophylactic or therapeutic benefits against HIV when infused (Safrit and Koup, 1995; Koup et al., 1995; Rowland-Jones et al., 1995; 1998; Price et al., 1997; Marsh and Thali, 2003). With the gp120 highly conserved and exposed receptor binding sites which are larger than the antibody footprint, HIV elicits abundant envelope antibodies. It is thus paradoxical how these antibodies have little neutralizing capacity. However, the major causes of envelope refractoriness to neutralizing antibodies is that gp120 antibodies binding leads to conformational masking, a phenomenon where several amino acids become hidden or buried making neutralization difficult (Marsh and Thali, 2003). Some antibodies generated against HIV may be harmful. This has been demonstrated in HIV infected individuals as well as gp120 immunized volunteers where antibodies enhance HIV infection. This act by Fc / complement mediated receptor endocytosis of primary isolates or T cell adapted virus. This is the limitation to success of many vaccine initiatives that have aimed to raise antibody mediated immunity (Marsh and Thali, 2003).

Cell mediated immunity is very vital against any viral infections. With most antibody response against HIV being thymus dependent, CD4 T-cells are essential for antibody class switching and affinity maturation; recruitment and activation of macrophage. CD4 T cells also help in induction of CD8+ cytotoxic T cells (McCune, 2001). Although CD4
T cells are considered to be the most important in HIV infection; they are the target cells of infection and it is their loss that leads to immune suppression (Kim et al., 1987).

Cytotoxic T-lymphocytes (CTL) therefore play an essential role in immunity against HIV. The CTL targets conserved regions in the internal viral proteins in the context of HLA class I molecules (Borrow et al., 1994; Huang et al., 2011). The CTL effector function is calcium dependent perforin/granzyme cell lysis or calcium independent Fas ligand cell lysis (Apoptosis) (Rowland-Jone et al., 1998). Other effector mechanisms include production of cytokines and chemokines (Duda et al., 2009). These are IFNγ and TNFα that inhibit viral replication (Meylan et al., 1993) and chemokines MIP-1α, MIP-1β, and RANTES that inhibit CCR5 coreceptor and HIV gp120 interaction necessary for primary infections.

2.6 Chemokines and Chemokine Receptors

Chemokines are a large family of proteins that regulate recruitment of leucocytes to sites of inflammation and coordinate their traffic throughout the body (Mackay, 2001). Chemokines are a group of small (8-14Kda), mostly basic, structurally related molecules that regulate cell trafficking of various types of leukocytes through interaction with a subset of seven transmembrane, G protein-coupled receptors (Zlotnik and Yoshie, 2000). Chemokines have a wide range of effects in many different cell types beyond the immune system including, various cells of central nervous system (Ma et al., 1998) and endothelial cells, where they result in either angiogenic or angiostatic effects (Strieter et al., 1995). In the immune system, chemokines mainly act on neutrophils, monocytes, lymphocytes and eosinophils and play a pivotal role in host defense mechanisms. Chemokines mediate leukocyte function by binding to and activating specific G protein-
coupled receptors (GPCRs) expressed by these cell populations (Rossi and Zlotnik, 2000).

### 2.6.1 Classification and nomenclature of chemokines and chemokine receptors

Chemokines are divided into two major subfamilies on the basis of the arrangement of N-terminal cysteine residues, CXC and CC, depending on whether the first two cysteine residues have an amino acid between them (CXC) or are adjacent (CC). Two other classes of chemokines have been described namely, Lymphoactin (C) and fractalkine (CX3C). The former lacks cysteine one and three of the typical chemokine structure (Zlotnik and Yoshie, 2000) while the later exhibit three amino acids between the first two cysteine and is also the only membrane-bound chemokine through a mucin like stalk (Bazan et al., 1997). The genes for these families are designated SCY (small secreted cytokines) - a through d, for CC, CXC, XC and CX3C respectively (Tab 2.1).

The chemokine nomenclature is based on the chemokine receptor nomenclature, which uses CC, CXC, XC, and CX3C followed by R (for receptor) and then a number. Thus we have the following chemokine receptors; CCR1-9, CXCR1-5, XCR1 (the lymphoactin receptor), and CX3CR1 (the fractalkine receptor) (Zlotnik and Yoshie, 2000; Tab 2.1). Actually, the chemokine ligands nomenclature just replaces R with L (ligand instead of receptors). Importantly, the biological effects of all these ligands are mediated by seven transmembrane domain receptors that represent a subset of the G protein-coupled receptor (GPCR) super family (Rossi and Zlotnik, 2000). Differences between the mouse and human chemokine not only reflect evolutionary patterns within this gene super family but also offer clues to understanding their biological functions.
Table 2.1: Classification of Chemokine and Receptor Families

<table>
<thead>
<tr>
<th>CXC Chemokine/Receptor Family</th>
<th>Systematic Name</th>
<th>Mouse Ligand</th>
<th>Chemokine Name</th>
<th>Receptor(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CXCL1</td>
<td>GRO/KC?</td>
<td>CXCR2</td>
<td>CCL1</td>
<td>CCR8</td>
</tr>
<tr>
<td>CXCL2</td>
<td>GRO/KC?</td>
<td>CXCR2</td>
<td>CCL2</td>
<td>CCR2</td>
</tr>
<tr>
<td>CXCL3</td>
<td>GRO/KC?</td>
<td>CXCR2</td>
<td>CCL3</td>
<td>CCR1, CCR5</td>
</tr>
<tr>
<td>CXCL4</td>
<td>PF4</td>
<td>Unknown</td>
<td>CCL4</td>
<td>CCR5</td>
</tr>
<tr>
<td>CXCL5</td>
<td>LIX?</td>
<td>CXCR2</td>
<td>CCL5</td>
<td>CCR1, CCR3</td>
</tr>
<tr>
<td>CXCL6</td>
<td>CKa-3</td>
<td>CCR1, CCR2</td>
<td>CCL6</td>
<td>CCR5</td>
</tr>
<tr>
<td>CXCL7</td>
<td>Unknown</td>
<td>CXCR2</td>
<td>(CCL9/10) MCP-2, MIP-1g</td>
<td>Unknown</td>
</tr>
<tr>
<td>CXCL8</td>
<td>Unknown</td>
<td>CXCR1,</td>
<td>CCL11</td>
<td>CCR3</td>
</tr>
<tr>
<td>CXCL9</td>
<td>Mig</td>
<td>CXCR3</td>
<td>CCL12</td>
<td>CCR2, CCR3</td>
</tr>
<tr>
<td>CXCL10</td>
<td>IP-10</td>
<td>CXCR3</td>
<td>CCL13</td>
<td>CCR1</td>
</tr>
<tr>
<td>CXCL11</td>
<td>Unknown</td>
<td>CXCR3</td>
<td>CCL14</td>
<td>CCR1</td>
</tr>
<tr>
<td>CXCL12</td>
<td>SDF-1</td>
<td>CXCR4</td>
<td>CCL15</td>
<td>CCR1, CCR3</td>
</tr>
<tr>
<td>CXCL13</td>
<td>BLC/BCA-1</td>
<td>CXCR5</td>
<td>CCL16</td>
<td>CCR1</td>
</tr>
<tr>
<td>CXCL14</td>
<td>BRAK</td>
<td>Unknown</td>
<td>CCL17</td>
<td>CCR4</td>
</tr>
<tr>
<td>(CXCL15)</td>
<td>Lungkine</td>
<td>Unknown</td>
<td>CCL18</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCL19</td>
<td>CCR7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCL20</td>
<td>CCR6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCL21</td>
<td>CCR7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCL22</td>
<td>CCR4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCL23</td>
<td>CCR1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCL24</td>
<td>CCR3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCL25</td>
<td>CCR9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCL26</td>
<td>CCR3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCL27</td>
<td>CCR10*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CC Chemokine/Receptor Family</th>
<th>Systematic Name</th>
<th>Mouse Ligand</th>
<th>Chemokine Name</th>
<th>Receptor(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCL1</td>
<td>TCA-3, P500</td>
<td>CCR8</td>
<td>CCR2</td>
<td></td>
</tr>
<tr>
<td>CCL2</td>
<td>JE?</td>
<td>CCR2</td>
<td>CCR1, CCR5</td>
<td></td>
</tr>
<tr>
<td>CCL3</td>
<td>MIP-1a</td>
<td>CCR1, CCR5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL4</td>
<td>MIP-1b</td>
<td>CCR5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL5</td>
<td>RANTES</td>
<td>CCR1, CCR3,</td>
<td>CCR5</td>
<td></td>
</tr>
<tr>
<td>(CCL6)</td>
<td>C10, MRP-1</td>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL7</td>
<td>MARC?</td>
<td>CCR1, CCR2,</td>
<td>CCR3</td>
<td></td>
</tr>
<tr>
<td>CCL8</td>
<td>MCP-2?</td>
<td>CCR3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CCL9/10)</td>
<td>MRP-2, MIP-1g</td>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL11</td>
<td>Eotaxin</td>
<td>CCR3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CCL12)</td>
<td>MCP-5</td>
<td>CCR2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL13</td>
<td>Unknown</td>
<td>CCR2, CCR3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL14</td>
<td>Unknown</td>
<td>CCR1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL15</td>
<td>Unknown</td>
<td>CCR1, CCR3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL16</td>
<td>LCC-1</td>
<td>CCR1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL17</td>
<td>TARC</td>
<td>CCR4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL18</td>
<td>Unknown</td>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL19</td>
<td>MIP-3b/ELC</td>
<td>CCR7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL20</td>
<td>MIP-3a</td>
<td>CCR6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL21</td>
<td>6Ckine/SLC</td>
<td>CCR7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL22</td>
<td>ABCD-1</td>
<td>CCR4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL23</td>
<td>Unknown</td>
<td>CCR1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL24</td>
<td>Unknown</td>
<td>CCR3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL25</td>
<td>TECK</td>
<td>CCR9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL26</td>
<td>Unknown</td>
<td>CCR3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCL27</td>
<td>ALP</td>
<td>CCR10*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Adapted from Zlotnik and Yoshie, 2000)
Mouse chemokines are therefore very important in immunology for experimental models and have their nomenclature based on the human chemokine as the standard (Zlotnik and Yoshie, 2000). The redundancy and binding promiscuity between many ligands and human chemokine receptors is an intriguing feature of chemokine super family (Tab 2.1). This may reflect their ability to regulate many different leukocyte subpopulations especially in complex microenvironments such as acute and chronic inflammatory responses (Ma et al., 1998). Nevertheless, other chemokines that play pivotal roles in homeostasis as well as in inflammatory responses do not exhibit redundancy—cannot be replaced with other chemokines (Tab 2.1; Zlotnik and Yoshie, 2000).

2.6.2 Physiological and Inflammatory Roles of the Chemokine System

The role of chemokines in the lymphoid system can best be visualized by dividing chemokines into two categories: Homeostatic and inflammatory chemokines. Homeostatic chemokines are constitutively expressed in a certain tissue or organ, suggesting a specific function involving cell migration. Moreover, chemokines play a vital role in lymphopoiesis. Experiments have shown that several chemokines are very essential in both B and T cell development (Nagasawa et al., 1996; Ma et al., 1998; Tachibana et al., 1998; Zou et al., 1998). Chemokines may also play an essential role in the thymic microenvironment (Campbell et al., 1999a). These observations strongly suggest a role for chemokine in the initiation of immune responses and more specifically suggest that some chemokines may have adjuvant properties (Zlotnic and Yoshie, 2000). However, inflammatory chemokines are strongly up regulated by inflammatory or immune stimuli in various cell types (macrophage, fibroblasts, T cells etc.) and are likely to participate in the development of immune or inflammatory reactions. The CCL3/MIP-
1α and CCL2/MCP-1 have been described as capable of inducing the differentiation of T-helper (Th) cells to either Th1 or Th2 cells (Zlontic and Yoshie, 2000). T-helper1 (Th1) cells have been shown to produce many more chemokine than Th2 cell (Orlofsky et al., 1994). Furthermore, the production of some chemokines is induced by Th2 cytokines like IL-4 (Orlofsky et al., 1994) or IL-10 (Zlontic and Yoshie, 2000).

2.6.3 Immune subversion through chemokine system.

Viruses have exploited Chemokine and chemokine receptor system mimics for immune subversion. One subset of mimic include actual chemokine and chemokine receptor, presumably acquired by virus in a manner of viral oncogenes, by gene transfer from host either as spliced or unspliced sequence, a clear example of divergent evolution (Tulman et al., 2000; Boshoff et al., 1997). The second subset includes molecule with structures that are unrelated to chemokines or chemokine receptors, but which bind to either thus distorting its function (Murphy, 2001). These are examples of convergent evolution. One of the most important achievements in this area has been to define the central role of chemokine receptors in HIV pathogenesis (Berger et al., 1999; Littman et al., 1998). The discovery of this concept has enabled new antiretroviral drugs and vaccine development strategies that have the usual features of exploiting the host proteins (Baba, 1999; La Casse et al., 1999). To date five distinct classes of mimics have been identified according to in vitro function; antichemokines, cell entry factors, cell growth factors, angiogenic factors, and leukocyte chemoattractants. Antichemokines subverts the immune system whereas other types of mimics exploit it. Antichemokines can be further sub classified into three groups according to structure and mechanism of action: chemokine receptor
homologs that acts as chemokine receptor antagonists; chemokine receptor homologs that function as plasma membrane expressed chemokine scavengers; and chemokine binding proteins with signal sequences which may have unique and unrelated structure and which function as secreted chemokine scavengers. Each mimic may carry out more than one function and mimics of more than one structural class may carry out each function. Viral chemokine agonists act in vitro as leukocyte chemoattractants of narrow specificity, possibly to focus preferred host cells at the site of infection and ensure viral spread (Zlotnic and Yoshie, 2000).

Examples of chemokine mimicry by virus include, human herpes virus type 8 (HHV8), also known as Kaposi’s sarcoma associated herpes virus (KSHV) and Human cytomegalovirus. Human cytomegalovirus (CMV), a severe opportunistic infection affecting multiple organs in immunocompromised hosts including patients with AIDS, encode the chemokine mimics; two CXC chemokine agonists (Penford et al., 1999) and a G protein-coupled chemokine receptor US28 (Gao and Murphy, 1994; Kledal et al., 1998). Human herpes virus type6 (HHV6), an opportunistic pathogen of HIV disease (Lusso and Gallo, 1995), has three chemokine system mimics; two are receptors and one chemokine agonist essential for viral spread from the initial site of infection in epithelia (Murphy, 2000).

Secreted chemokine scavengers, although less clearly related to HIV/AIDS, are of great conceptual and potential practical importance. Except for limitations of delivery and antigenicity, soluble chemokine scavengers may be well suited for clinical application in inflammation (Murphy, 2001). Several examples of chemokine and chemokine receptor mimics with unique structures, including β-defensin that bind to CCR6, have been
reported to be functionally active. Moreover, CCR5 and other chemokine receptors are reported to facilitate infections by myxoma, vaccinia and other poxviruses (Lalani and McFadden, 1999).

Due to immunosuppression such as that caused by HIV/AIDS, patients can be coinfected with several of these viruses, hence chemokine mimics produced by each virus might interact in a pathophysiologically significant manner with both host chemokine system and chemokine mimics encoded by other viruses (Murphy, 2001). Reported interactions include the ability of the CMV encoded chemokine receptor US28 to support HIV infection by functioning as a HIV coreceptor and the ability of the HHV8 chemokine mimic vMIP-II to block both the activity as well as HIV usage of the host leukocyte receptors CCR5 and CXCR4. Host chemokines can also block gp120 usage of host chemokine receptors for CD4 target cell entry. Other viral chemokine mimics can act as agonists at leukocyte chemokine receptors, possibly to promote viral dissemination. Interaction of HIV-1 gp120 with CCR5 is critical for efficient HIV transmission in exposed populations. Whether other potential interactions among viral mimics occur in vivo and affect pathogenesis is not yet known.

2.7 Chemokine receptors as HIV coreceptors

Human Immunodeficiency Virus uses CD4 and a chemokine receptor to mediate membrane fusion and entry into target cells (Berger et al, 1999). Fourteen chemokine receptors or structurally related molecules have been identified that can function as coreceptors for entry of HIV-1 in vitro (Van Rij et al., 2002; De Roda Husman et al., 1998). Moreover, several chemokine receptors have been identified which serve as coreceptors, along with CD4, for HIV-1 entry and fusion into CD4+ T lymphocytes and
macrophages (Van Rij et al., 2002). The CXC chemokine receptor 4 (CXCR4, Fusin, LESTR, HUMSTR) is a member of the seven transmembrane G-protein-coupled receptor family (Loetscher et al., 1994; Noetke et al., 1993) and acts as a cofactor for T cell tropic (T-tropic) HIV-1 strains (Berson et al., 1996; Feng et al., 1996). The CC chemokine receptor 5 (CCR5) acts as the major coreceptor for primary macrophage tropic (M-tropic) HIV-1 strains (Alkhatib et al., 1996; Deng et al., 1996; Dragic et al., 1996). Other chemokine receptors like CCR2b, CCR3, and STRL33 can also serve as coreceptors for HIV-1 (Choe et al., 1996; Doranz et al., 1996; Liao et al., 1997).

With the discovery of the chemokine receptor gene family as HIV-1 entry coreceptors, HIV-1 strains can also be classified by coreceptor use. Strictly NSI viruses primarily use the chemokine receptor CCR5 while SI viruses primarily use the chemokine receptor CXCR4 (Berger et al., 1998). However, most primary SI isolates use CXCR4 in conjunction with CCR5 and/or one of the minor coreceptors (McDonald et al, 2001). A nomenclature based on coreceptor use for CCR5, CXCR4 and a dual (both CCR5 and CXCR4) using variant has classified these variants as R5, X4, and R5/X4 respectively (Berger et al., 1998). The predominance of NSI/R5 viruses early in infection where as X4 SI viruses detection in later stages in approximately 40-50% of patients (Begaud et al., 2003) show that HIV strains transmitted in vivo generally use coreceptor CCR5 (R5 viruses) but strains using CXCR4 (X4 viruses) later evolve in the course of infection and emerge in approximately 50% of the infected individuals (Blaak et al., 2005; Kitchen et al., 2004). Therefore, X4 HIV variants evolve from R5 via the R5/X4 phenotype (Fransje et al., 2002; Tscherning et al., 1998).
2.7.1 Coreceptor Usage in Disease Pathogenesis and Transmission

The detection of the X4 SI viruses has been associated with the rapid progress of disease (Scarlatti et al., 1997). Infections of R5/X4 dual tropic viruses have also been associated with T cell depletion and disease progression (Yu et al., 1998). Therefore, coreceptor usage is thought to play a role in both HIV transmission and pathogenesis with an expanded coreceptor usage associated with more rapid disease outcome (Begaud et al., 2003). This conflicts with the lower pathogenicity of HIV-2 despite the observed broad coreceptor usage (Blaak et al., 2005). Expanded coreceptor usages described in HIV-1 strain have been reported but seem more limited. These include HIV-1 primary isolates that use GPR15 (BOB) to establish a productive infection in vitro (Xiao et al., 1998). Moreover viruses using BOB were isolated in late stages of disease suggesting a relationship between these viruses and HIV-1 pathogenesis (Pohlmann et al., 1999). A switch from R5 monotropic HIV-1 to R5/ Bonzo (CXCR6) was described in pregnant Cameroonian women who transmitted HIV-1 to their children (Tschering-Casper et al., 2000). It is interesting that after evolution of X4 variants, X4 and R5 variants co-exist and both may even expand (Koot et al., 1996). The coexistence of R5 and X4 variants can be explained by availability of separate cell niches within the pool of memory cells, defined by the differential expression of CCR5 or CXCR4 on subsets of memory CD4+ T cells (Fransje et al., 2002). Indeed, R5 HIV could be isolated from the CCR5+ memory CD4+ T cells population whereas X4 variants were isolated from the CXCR4+ memory subsets (Blaak et al., 2000).
2.7.2 Relationship between coreceptor usage switch and the immune profiles

Only a limited number of mutations are required for the NSI to SI phenotypic switch *in vitro* (Briggs *et al.*, 2000). It is unclear why X4 variants only develop in approximately half of the individuals, and only when CD4+ T cell numbers decline below approximately 400 cells per μl blood (Koot *et al.*, 1999). The appearance of X4 variants therefore suggests restraints on the ability to establish a productive infection or on the availability of susceptible target cells for X4 variants, rather than a difficulty to induce a switch *per se* (Fransje *et al.*, 2002). Evidence points to the possibility that the immune response against the SI strain is stronger than against the NSI strain (Schuitemaker *et al.*, 1992; Bouhabib *et al.*, 1994; Cornelissen *et al.*, 1995; Fauci, 1996; Koot *et al.*, 1999).

Since CXCR4 is expressed on many more CD4+ T cells in the body (including hematopoietic progenitor cells and thymocytes) than CCR5 (mostly found in memory T cells and macrophages), switching to CXCR4 usage potential allow the virus access to this large and critical pool of target cells, which may help to explain the accelerated CD4+ T cell decline associated with the R5 to X4 switch (Jensen *et al.*, 2003a). Before identification of chemokine receptors as coreceptors of HIV-1, the differences in pathogenicity of R5 and X4 HIV-1 was thought to be due to differences in cytopathicity and replication rates (Fransje *et al.*, 2002). X4 variants in general replicate more rapidly and to higher levels than R5 variants and X4 result in a more massive cell depletion *in vitro* (Van't Wout *et al.*, 1998; Fouchier *et al.*, 1996; Connor *et al.*, 1993; Asjo *et al.*, 1986). Now it is apparent that CCR5 and CXCR4 are not evenly distributed on the cells that have been used in these *in vitro* assays (Fransje *et al.*, 2002). Therefore, insights in cytopathicity based on these models need to be redefined in the context of the available
target cells. Indeed, R5 and X4 HIV-1 variants were equally cytopathic for the target cells expressing the appropriate coreceptors, resulting in depletion of the cognate target cells (Grivel et al., 1999; Kwa et al., 2001). The enhanced CD4+ T cell decline associated with X4 variants may therefore not be due to merely broad target cell range and more extensive replication of X4 viruses but rather to the infection and killing of naïve T cells by which the exponential clonal expansion of a progeny memory T cells daughter population is prevented (Blaak et al., 2000). This may further interfere with T cell renewal and contribute to the enhanced CD4+ T cell decline associated with X4 HIV-1 variants (Fransje et al., 2002).

In relation to this observation, in vitro data have shown high expression of the CXCR4 coreceptor in immature thymocytes (Zeitseva et al., 1998) suggesting that the T tropic phenotype might be associated with poorer CD4+ T cell recovery after HAART because T-tropic strains use this coreceptor preferentially. For this reason it would be of special interest to study viral phenotype before treatment, predicting that the predominance of T tropic phenotype would be associated inversely with thymic function-related markers and with a poorer recovery of CD4+ T cell after HAART (Ruiz-Metaos et al., 2004). For HIV-1 the appearance of X4 variants and the acquired capacity to infect thymocytes and naïve T cells coincide with accelerated CD4+ T cell loss and disease progression (Berkowitz et al., 1998; Blaak et al., 2004; Koot et al., 1993). The natural course of disease progression however can be altered by antiretroviral therapy (Murphy et al., 2003; Ruiz-Metaos et al., 2004). Potent ART preferentially suppress X4 strains of HIV-1 in patients, shifting the viral population back to R5 after treatment (Ruiz-Metaos et al., 2004).
2.7.3 Relationship between HIV-1 subtypes and coreceptor usage

The prevalence of a specific HIV-1 subtype may have significant implications on coreceptors usage. This is due to the subtype-specific differences in coreceptors usage that have been controversial (Zhang et al., 1996; Tscheming et al., 1998). A possible correlation between specific genetic subtypes and biological phenotypes has been studied (de Wolf et al., 1994; Nkengasong et al., 1995; Tscheming et al., 1998; Bjourndal et al., 1999). The C subtype express mainly one phenotype, NSI, while the D subtype exhibit SI phenotype more frequently than the other subtypes. The basis of non syncytium inducing (NSI) and syncytium inducing (SI) phenotypes have been shown to be their differential use of chemokine receptors for viral entry (Jensen et al., 2003a).

Subtype C strains have been shown to express mainly the NSI phenotype and mostly use CCR5 as coreceptors, independent of the clinical status, CD4 counts or antiretroviral treatment contrary to other subtypes (Tscheming et al., 1998; Bjourndal et al., 1999). Despite this CCR5 phenotype predominance, subtype C infected individuals’ progress to AIDS in a similar manner as individuals infected with other subtypes. Therefore in the case of subtype C, disease progression might be explained by something else than the coreceptors usage and the emergence of the SI phenotype. One suggestion is the number of NFkB sites present in C subtype strains LTR-promoters; they seems to include frequently at least three functional NFkB sites, instead of only two sites present in other subtypes. Theoretically this could increase transcriptional activation from the viral promoter, perhaps leading to a higher replication rate (Abebe et al., 1999).

Subtype D strains were shown not to exhibit dual tropism for CCR5 and CXCR4, which is usually associated with the phenotypic switch for other subtype strains. It has been
suggested that dual coreceptor use is a transition stage and that ultimately differentiated virus would use only CXCR4 coreceptor (Tscherning et al., 1998). Instead D subtype strains were monotropic for CXCR4 after the switch (Tscherning et al., 1998). Again this was shown to be independent of clinical status, CD4 counts or antiretroviral treatment. It could also be due to faster diversification rate of this subtype, as the env V3 loop of D subtype strain has been shown to be significantly more divergent than other subtypes (Korber et al., 1994). Greater variability in the region responsible for cell tropism could allow the virus to switch to the SI phenotype rapidly.

2.8 Determination of Viral Coreceptor Usage

To study which chemokine receptor a HIV variant can use for cell entry, cell lines non-permissive for HIV-1 that are engineered to express CD4 antigen and a chemokine receptor of interest are generally used (Blaak et al., 2005). Derivatives of human glioma cell lines (U87) and human osteosarcoma line (HOS and GHOST) have predominantly been used for this purpose. Based on the capacity to productively infect these cell lines, it has been shown that most primary HIV-1 variants are restricted to use of CCR5, although variants exist that can also use CXCR4 and/or CCR3 (Connor et al., 1997). Only minority of HIV-1 variants have been shown to have capacity to infect cells expressing other coreceptors like, CCR2b, (Bjorndal et al., 1999; Connor et al., 1997) CCR8, (Dittmar et al., 1999; Zhang et al., 1998b) CXCR6 (Bonzo), or GPR15 (BOB) (Cecilia et al., 1998; Dittmar et al., 1999; Zhang et al., 1998a). In contrast many HIV-2 variants could infect a whole range of indicator cells expressing different coreceptors, indicating a more promiscuous nature of HIV-2 in respect to coreceptor usage (Blaak et al., 2005).
The MT2 cell fusion assay has been used for determination of SI/NSI phenotypes of HIV-1, a very reliable though laborious method to determine CXCR4 usage (Koot et al., 1992). In this assay, MT2 cells selectively support replication of SI variants only thus used as an indicator. Other methods include Heteroduplex Tracking Assay (HTA) targeting V1, V2 and V3 gp120 region and the genotype and sequence analysis (Freel et al., 2003).

The current study used the bioinformatics approach of sequence analysis based on the fact that determinants that govern the biological properties such as SI phenotype and cell tropism have mainly been mapped to the envelope gene (Groenink et al., 1993; O’Brien et al., 1990; Shioda et al., 1991; Chesebro et al., 1992) especially to the variable region, and in particular the third variable (V3) region, the V3 loop (Shioda et al., 1992; Fouchier et al., 1992; Groenink et al., 1993; Hwang et al., 1991; De Jong et al., 1992; Chesebro et al., 1992). Sequence analysis and V3 loop demonstrate the frequent presence of positively charged amino acids at position 11 and/or 25 in the V3 loop to confer SI variant (Fouchier et al., 1992). In addition overall charge of the V3 loop is higher in the SI variants (Shioda et al., 1992; Fouchier et al., 1992). Additional determinant for coreceptor usage map to V2 length and presence of glycosylation sites in other envelope regions (Groenink et al., 1993; O’Brien et al., 1990; shioda et al., 1991) but no single amino acid change have been pinpointed that determine phenotype. It appears V1/V2 changes are unlikely to cause coreceptor switch independent of V3 loop (Stamatatos et al., 1998). Instead V1/V2 changes may occur in concert with change in V3 that improve replicative capacity as switching evolves (Groenink et al., 1993; carrillo et al., 1996; Este et al., 1999).
CHAPTER THREE: MATERIALS AND METHODS

3.1 Study Population

Samples used in this study were collected from Lokichogio and Kakuma (Turkana district), Sololo and Moyale district hospital (Moyale district) and Mandera district Hospital (Appendix 1). The target population consisted of antenatal clinics (ANC) attendees, cases of sexually transmitted diseases (STD), blood donors, tuberculosis patients, and children born of HIV positive mothers. The samples were collected from healthcare outpatients and inpatients that were already tested HIV positive and counseled within the respective health facility framework. This was done after informed consent from the patients. In cases of children consent was sought from parents and guardians.

3.1.1 Ethical Clearance

Clearance to carry out the study was obtained from Kenyatta University. Kenya Medical Research Institute (KEMRI) Ethical Review Committee gave the approval and ethical clearance to carry out the study, being part of the ongoing projects being carried out in Northern Kenya.

3.1.2 Inclusion and Exclusion Criteria

The inclusion criteria included consenting persons presenting or admitted in the respective healthcare facility with determinate HIV status. Moreover, parental consent was sought in case of HIV positive children. HIV negative patients and non-consenting HIV positive individuals/patients were all excluded from the study.
3.1.3 Samples size determination

The samples for this study were randomly collected. The sample size was determined using the formula by Fisher et al. (1998). Using the HIV prevalence of 6.7 % in Kenya (NASCOP, 2005);

\[
N = \frac{Z^2 P (1-P)}{D/d^2}
\]

Where: N-minimum samples size

d- Allowable error (5%)

P- Expected prevalence (6.7 % ≈ 7 %)

Z- The standard error at 95 % confidence interval (1.96)

D- The design effect (1)

\[
N = 1.96^2 (0.07) (0.93) 1 / 0.05^2 = 100.03 \approx 100
\]

N=100 samples.

The Kenyan national HIV-1 prevalence was used to approximate the sample size in order to collect enough samples for analysis. The minimum sample size was approximated at 100 samples however, for the purpose of this study all the available samples (135 samples) were processed and used.

3.2 Sample Collection and Processing

3.2.1 Sample Screening and Handling

Five milliliters of whole blood samples was collected into 10mls EDTA tubes and screened for HIV antibodies. HIV positivity was confirmed using Particle Agglutination kit (KEMRI; Kenya) and an immunochromatography assay kit (Determine HIV-1/2;
Abbot: Japan) as described in the manufacturers manuals for each kit. The sample were then temporarily stored at \(-20^\circ\text{C}\) freezer in the Healthcare facility and transported by air in ice packed cool boxes to Nairobi within two weeks. The samples were then immediately stored at \(-80^\circ\text{C}\) at KEMRI-Center of Virus Research (CVR) laboratory until use.

### 3.2.2 Extraction of Deoxyribonucleic Acid (DNA)

Deoxyribonucleic acid (DNA) was extracted from whole blood according to the method described by Chomczynski et al. (1997). Five milliliters of whole blood sample was added to 10mls of 0.84 % ammonium chloride and vortexed to mix completely. This was incubated at \(37^\circ\text{C}\) for 10 minutes and then span at 10,000g for 10 minutes. The supernatant was then discarded and above procedure repeated on the resulting pellets thrice until the pellets appeared clear. The pellet was drawn into an Eppendorf tube and dissolved in 500\(\mu\)l of DNAzol genomic DNA extraction reagent (Gibco BRL\textsuperscript{®} Germany) in a 1.5ml Eppendorf tube. Two volumes of chilled absolute ethanol was then added and incubated at room temperature for 5 minutes to allow DNA to precipitate. The precipitated DNA was washed twice with 1ml of 70 % ethanol and then solubilized in 100\(\mu\)l of double distilled (DNase/RNase free) water.

### 3.2.3 Nested Polymerase Chain Reaction (PCR)

Two consecutive PCR were carried out using two different primers (the outer and the nested primer respectively) to enhance PCR specificity, a procedure known as Nested PCR. The first Polymerase chain reaction (PCR) was carried out using the proviral DNA extracted from the peripheral blood mononuclear cells (PBMC), as the templates,
following the method described by Carr et al. (1998) and Takehisa et al. (1998). The PCR was designed to amplify a region of gp120 in the *env* region using primers designated from the C2V3 region (Takehisa et al., 1998). The master mix was prepared to achieve the following final concentrations: 1× PCR buffer; 2.0 mM MgCl₂; 0.5 units of *Taq* polymerase; 2 ng of each primer; 2.0 mM deoxyribonucleoside triphosphates (DNTPs) and 2.0 ng of DNA template. The master mix was made depending on the number of samples to be processed and then aliquoted into PCR tubes where DNA templates were added. The PCR tubes were then vortexed to mix the template with the master mix and spin down briefly on a microfuge before being loaded on the thermocycler. The first PCR was carried out using the following conditions: Denaturation at 95 °C for 30 seconds; annealing at 55 °C for 30 seconds; extension at 72 °C for 1 minute for 35 cycles followed by a final extension of 72 °C for ten minutes. Using the products from the first PCR as the template and a master mix of the same concentration as the first PCR described above, the nested PCR was carried out using the same conditions as the first PCR above. The nested PCR uses different set of primers which are necessary to guarantee the specificity of the amplified region.

All PCR products were analyzed by conventional Agarose gel electrophoresis, to resolve the PCR products using 1× Tris acetic acid (TAE) Ethyl Diamine Tetra Acetic Acid (EDTA) following the method described by Sambrook et al. (1989). The Agarose gel was prepared with 1× TAE. The samples were then mixed with gel loading buffer (to increase the density and for visualization) before loading into the wells on the gel. Alongside the samples a molecular weight marker was loaded as a standard for estimation of the size for the sample amplicons. Electrophoresis was carried out at constant electron potential.
of 100 volts using bio Rad model 200/2 power supply source for one hour. On completion of the electrophoresis, the gel was stained with 0.5 μg/ml ethidium bromide for 15 minutes. The location of nucleic acid on the gel was determined by direct examination of the gel under UV light and the size estimated by comparing with the molecular weight marker loaded alongside.

3.2.4 Deoxyribonucleic Acid (DNA) Sequencing

Deoxyribonucleic Acid (DNA) Sequencing was carried out to determine the order of the bases in the PCR products using the Bigdye® Kit (Applied Biosystems: USA). Using this kit, fluorescently labeled dye is attached to the Deoxy Nucleotide Triphosphates (DNTPs; extension products) in DNA sequencing reactions. The dye comes in four colours: red for thymine, blue for cytosine, black for guanine and green labeling for adenine bases. Dye labels are incorporated using either 5'-dye labeled primers or 3'-dye dideoxyribonucleotide terminators. AmpliTaq ® polymerase was used for primer extension while the dye were used to perform sequencing PCR with the template being the PCR generated products to be sequenced.

Three micro liters of 5x buffer, 2μl of Bigdye and 1.5μl primers were added to 3 μl DNA sample. The sequencing PCR was carried out under the following conditions: 96 °C for 5 minutes, followed by 25 cycles of 96 °C for 10 seconds, 50 °C for 5 seconds and 60 °C for 4 minutes. After the sequencing PCR, the products were labeled by adding 2μl of 3M sodium acetate (pH 4.6) and 50μl of absolute ethanol to each sample. DNA was then washed in a 1.5ml Eppendorf tube using 250μl of 70 % ethanol then span at 15000 revolutions per minutes (rpm) and the supernatant aspirated from the tube before drying it
in a vacuum centrifuge for 10-15 minutes. After drying the sample completely, 25μl of template suppression reagent (TSR) was added to each sample, vortexed thoroughly then incubated at 95°C for 2 min, chilled on ice, vortexed thoroughly again, briefly span in a microfuge before transferring them to the automated sequencer.

During sequencing, the auto sampler successively brings each sample into contact with the cathode electrode and one end of the glass capillary filled with polymer. The polymer used in this study was pop6 polymer. An anode at the other end of the capillary is immersed in buffer. A portion of the sample enters capillary as current flows from the cathode to the anode a process called electrokinetic injection. The end of the capillary near the cathode is then placed in the buffer and the current applied again to continue electrophoresis.

When fluorescently labeled DNA nucleotides reach a detector window in the capillary coating, a laser excites the fluorescent dye labels. A CCD camera collects emitted fluorescence intensity at each data point. The software analyses the raw data to quantify the DNA fragments and determine the size of the fragment by comparing with the fragments contained in the size standard. The ABI PRISM® 310 data collection software (Applied Biosystems: USA) was used to collect the ‘raw data’ that was then analyzed by the ABI PRISM® DNA sequencing analysis software and the Genescan® version 2.0 analysis software program according to the method given by the manufacturer (Applied Biosystems: USA). The data collected was saved on a floppy disc and used for sequence analysis.
3.2.5 Deoxyribonucleic Acid (DNA) Sequence Analysis

Once the samples were sequenced, they were analyzed to determine their subtype using the Basic Local Alignment Search Tool (BLAST) references on the HIV sequence database following the method described by Altschul et al., (1997) and Karlin et al., (1993). These sequences were then analyzed using geno2pheno[XCORECEPTOR] version 1.0, a bioinformatics tool based on support vector machine for predicting HIV-1 coreceptor usage from the viral sequence information following the method described by Sing et al., (2004).

3.2.6 Determination of HIV-1 subtypes

All the generated DNA sequences were analyzed by basic local alignment search tool (BLAST) software following the method described by (Altschul et al., 1997) to determine their subtype and closely related sequences (Karlin et al., 1990; 1993). BLAST is a tool that is most frequently used for calculating sequence similarity. BLAST algorithm is a heuristic program, meaning that it relies on some smart shortcuts to perform the search faster. This is called the Local alignment approach as opposed to Global alignment approach that attempts to align two sequences over their entire length. BLAST comes in variation for use with different query sequences against different databases. In this study the Los Alamos HIV database was used to acquire references reported from all over the world for use in HIV-1 subtype determination (http://www.hiv.lanl.gov/content/hiv-db/mainpage.html). This procedure was used for analysis of all the sequences generated from the C2V3 region of the HIV-1 genome that were sequenced.
3.2.7 Determination of HIV-1 Coreceptor Usage

All the generated DNA Sequences were analyzed by geno2pheno\textsubscript{[CORECEPTOR]} version 1.0, a bioinformatics tool based on support vector machine for predicting HIV-1 coreceptor usage from the viral sequence information as described by Sing et al. (2004). To determine the coreceptors usage, V3 sequences were uploaded on geno2pheno [CORECEPTOR] input page and then the sensitivity and specificity levels adjusted at 86% and 87% respectively. This bioinformatics program evaluates the properties of the amino acids in specific key positions and the net charge of the amino acid sequence that are determinants of the coreceptor usage. The characteristics of these sequences were then compared with the reference sequences in the database for determination of CCR5 or and CXCR4 coreceptor usage.

3.3 Data Analysis

The data on coreceptor usage was entered into Ms Excel and imported into S-PLUS 6.0 Professional for analysis. Descriptive analysis using frequency distribution tables, charts and graphs were used to show the prevalence of coreceptor use in Northern Kenya. Inferential statistics using Fisher's Exact Chi Square test was used to test the significance and Marascuilo procedure for paired comparison of proportions was further used to determine the specific source of the difference in the relationship between specific subtypes and coreceptor Usage.
CHAPTER FOUR: RESULTS

4.1 Gene (gp120 V3) sequences of HIV-1 variants generated from Northern Kenya

One Hundred and Thirty Five samples that tested positive in HIV-1 serology assays were successfully processed to the level of gene sequencing (Appendix 2). The generated sequences were analyzed to determine their subtype using the Basic Local Alignment Search Tool version 2.0 (BLAST 2). The subtypes were determined by direct BLAST analysis (www.hiv.lanl.gov/content/hiv-db/BASIC BLAST/basic blast.html). Based on the direct blast analysis the circulating subtypes were found to belong to HIV-1 subtypes C, A1, D and circulating recombinant forms (CRFs). Thirty seven (27 %) of the samples belonged to HIV-1 subtype C, Twelve (9 %) to HIV-1 D subtype, Seventy three (54 %) to HIV-1 A1 subtype and thirteen (10 %) were circulating recombinant forms (Fig. 4.1).
Out of the thirteen circulating recombinant forms, four (30.8 %) were HIV-1 subtype A1D, five (38.4 %) of subtype A1C and other four (30.8 %) of subtype A1CD (Tab. 4.1).
Tab 4.1: Distribution of the circulating recombinant forms in the Northern Kenya

<table>
<thead>
<tr>
<th>Recombinant forms</th>
<th>HIV-1 Subtype A1D</th>
<th>HIV-1 Subtype A1C</th>
<th>HIV-1 Subtype A1CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of samples</td>
<td>4 (30.8 %)</td>
<td>5 (38.4 %)</td>
<td>4 (30.8 %)</td>
</tr>
</tbody>
</table>

4.2 Coreceptor Usage in Northern Kenya HIV-1 Variants

All the generated DNA Sequences were analyzed by geno2pheno[CORECEPTOR] version 1.0, a bioinformatics tool based on support vector machine for predicting HIV-1 coreceptor usage from the viral sequence information as described by Sing et al. (2004). geno2pheno[CORECEPTOR] version 1.0 software automatically aligns the query sequences to the reference V3 gene sequence and analyze for determination of coreceptor usage. Out of the 135 sequences analyzed, 78 sequences aligned well with the references and were successfully analyzed for determination of coreceptor usage. However, the results indicated possibility of absence of gp120 env V3 region in 57 sequences. All the 57 sequences defaulted to CXCR4 usage. To avoid attritions, only the 78 sequences that aligned properly with the references were considered. Out of the 78 properly aligned sequences 41 belonged to subtype A1, 19 belonged to subtype C, 11 belonged to subtype D, while 6 were circulating recombinant forms. Analysis of the V3 region for coreceptor usage was as follows; 17 (21.8 %) were CCR5, 54 (69.2 %) were CXCR4 while 7 (9 %) were dual tropic (both CCR5 and CXCR4 using). Among the subtype A1 variants, four
(9.8 %) were CCR5 using, thirty four (82.9 %) were CXCR4 using while three (7.3 %) were dual tropic. Analysis of subtype C variant for coreceptor usage showed ten (52.6 %) to be CCR5, seven (36.8 %) to be CXCR4 and three (15.8 %) to be dual tropic. All the eleven (100 %) subtype D variants were CXCR4 using. Among the circulating recombinant forms three (50 %) were CCR5 using, two (33.3 %) were CXCR4 using and one (16.7 %) was dual tropic (Tab.4.2).

Tab 4.2: Coreceptor Usage by HIV-1 variants in Northern Kenya based on gp120 V3 env gene sequence analysis

<table>
<thead>
<tr>
<th>Subtype</th>
<th>CCR5 Using</th>
<th>CXCR4 Using</th>
<th>Dual Tropic</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtype A1</td>
<td>4 (9.8 %)</td>
<td>34 (82.9 %)</td>
<td>3 (7.3 %)</td>
<td>41</td>
</tr>
<tr>
<td>Subtype C</td>
<td>10 (52.6 %)</td>
<td>7 (36.8 %)</td>
<td>3 (15.8 %)</td>
<td>19</td>
</tr>
<tr>
<td>Subtype D</td>
<td>0 (0 %)</td>
<td>11 (100 %)</td>
<td>0 (0 %)</td>
<td>11</td>
</tr>
<tr>
<td>CRF</td>
<td>3 (50 %)</td>
<td>2 (33.3 %)</td>
<td>1 (16.7 %)</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>17 (21.1 %)</strong></td>
<td><strong>54 (69.2 %)</strong></td>
<td><strong>7 (9.0 %)</strong></td>
<td><strong>78</strong></td>
</tr>
</tbody>
</table>
4.3 Relationship between Specific HIV-1 Genetic Subtypes and Coreceptor Usage

Based on the gp120 C2V3 region, three pure HIV-1 genetic subtypes and circulating recombinant forms were observed in Northern Kenya population in the following proportions: subtype A1 (54 %), subtype C (27 %), subtype D (9 %) and Circulating Recombinant Forms (10 %).

Coreceptor usage varied among different HIV-1 subtypes in the Northern Kenya region. The CCR5 Coreceptor use was observed to be 52.6 % in subtype C, 9.8 % in subtype A and 0 % in subtype D variants. A significantly high percentage of CCR5 coreceptor using variants was observed in subtype C (52.6 %) as compared to subtype A1 (9.8 %) and subtype D (0 %) observed in this study (P<0.001).

Subtype A1 Variants were observed to use all the three types of coreceptor with 4 (9.8 %) of the samples as CCR5 using, 34 (82.9 %) as CXCR4 using and 3 (7.3 %) as dual (CCR5 and CXCR4) coreceptor. Although CXCR4 using phenotype was higher (82.9 %) than the others, this shows a typical distribution of the coreceptor use where HIV-1 is normally transmitted as CCR5 using and eventually evolve to CXCR4 using via dual coreceptor (CCR5 and CXCR4) using phenotype. High proportion of CXCR4 using phenotype in subtype A1 suggests that most infections were in late stages.

Notably, all subtype D variants were CXCR4 using. None of the subtype D variants expressed dual tropism as compared to subtype A1 (7.3 %), subtype C (15.8 %) and the recombinants (16.7 %). Absence of any CCR5 using subtype from subtype D further indicates that all subtype D infections were in advanced stage thus having already
switched to CXCR4 usage. An association between the coreceptor usage and certain HIV-1 subtypes was reported. This is reflected in the proportions of HIV-1 variants using certain chemokine receptors as entry coreceptors. Subtype C variant was observed to express CCR5 coreceptor usage (Fig. 4.2).

![Graph showing frequency of coreceptors usage among different HIV-1 genetic subtypes]

**Fig 4.2: Frequency of coreceptors usage among different HIV-1 genetic subtypes**
CHAPTER FIVE: DISCUSSION

5.1 Genetic Subtypes of HIV-1 Circulating in Northern Kenya

Based on gp120 C2V3 gene sequence Phylogenetic analysis, three pure HIV-1 genetic subtypes were reported in Northern Kenya as follows: subtype A1 (54%), subtype C (27%) and subtype D (9%). Moreover, 10% of HIV-1 variants were reported to be circulating recombinant forms (CRF). The dominant HIV-1 genetic subtype in Northern Kenya as observed in this study was subtype A1. Previous studies in other parts of Kenya that involved HIV-1 gp120 C2V3 gene region reported the following percentages; subtype A1 (71-87%), subtype D (7-29%) and (7-17%) to be subtype C (Jensen et al., 1994; Robbin et al., 1999). Another study in Kenya by Dowling et al., (2002) that analyzed a near full length HIV-1 genome reported dominant variant to be subtype A1 (56%) others being subtype C and subtype D with only 2.4% each and a high proportion of recombinant forms that was above 40%. In Northern Kenya three HIV-1 genetic subtypes have previously been reported; subtype A1 (50%), subtype C (39%) and subtype D (11%). Among these subtypes high level of recombination (56%) was also reported (Khamadi et al., 2005). The data presented in this study is in agreement with that reported by Khamadi et al. (2005) on the HIV-1 subtypes circulating in Northern Kenya with minor differences in the percentage of the circulating subtypes. The differences in the percentages can be attributed to the time gap between the two studies hence change in percentage due to new infections and the possible death of some of the patients.

However, in the current study a lower level of recombination (10%) was reported compared to the previously reported studies (Dowling et al., 2002; Khamadi et al., 2005). Moreover cases of recombination have been reported to be on increase (Songok et al.,
The discrepancy between the current and the previous studies can be explained by the fact that the current study only analyzed the *env* gene region while Dowling *et al.*, (2002) and Khamadi *et al.*, (2005) analyzed near full length and three (*env*, *pol* and *gag*) regions respectively. It is argued that had Dowling *et al.* (2002) and Khamadi *et al.*, (2005) focused on the *gp120 env* region alone their result for recombination would have been 0% and 8% respectively (Khamadi, 2005). Considering this argument by Khamadi, (2005) the results of the current study are thus in line with the observation by Songok *et al.*, (2003) that cases of recombination are on the increase.

The *gp120 C2V3* gene sequences analyzed by the BLAST 2 analysis in this study clustered with reference sequences from Kenya, Uganda, Tanzania, Cameroon, Sudan, Zimbabwe, Ethiopia, South Africa, Burundi, Botswana, Yemen, Finland, Denmark and Sweden. This shows how divergent the HIV-1 strains circulating in Northern Kenya are. The close relationship resulting to clustering of the study sequences from Northern Kenya and reference sequences may suggest the origin of the Northern Kenya variants resulting from the interaction between the Northern Kenya population and the respective countries. This may be attributed to the cross border interaction between the Northern Kenya people and neighboring countries like Uganda, Sudan and Ethiopia (Appendix 1), cordial relationship between Kenya and the other countries like Tanzania, Burundi and Rwanda and interaction between Kenya and countries like Finland, Denmark and Sweden due to trade and tourism.

The failure by some of the sequences to cluster with the reference sequences can be attributed to the small size of the generated sequences resulting to inadequate information for alignment with the reference sequences. It may also be attributed to HIV-1 error
prone reverse transcriptase enzyme and lack of a proof reading mechanism during reverse transcription of the viral RNA resulting to frequent incorporation of mutation in the viral genome (van Rij et al., 2000) hence variation from the reference sequences.

5.2 Coreceptor Usage in Northern Kenya HIV-1 Variants

The bases of non syncytium inducing (NSI) and syncytium inducing (SI) phenotypes have been shown to be their differential use of chemokine receptors for viral entry into the host cells (Jensen et al., 2003). Strictly NSI viruses primarily use the chemokine receptor CCR5 while the SI viruses primarily use CXCR4 and/or one of the minor coreceptors (McDonald et al., 2001). The HIV strains transmitted in vivo (especially through the mucosal route) use the CCR5 coreceptor but CXCR4 using strains later evolve in the course of infection in about 50 % of the infected individuals through CCR5/CXCR4 dual tropic phenotype (Fransje et al., 2002). Coreceptor usage and syncytium inducing phenotype are strong predictors of disease progression (Koot et al., 1993). A switch of coreceptor use from CCR5 monotropic to either CXCR4 monotropic or CCR5/ CXCR4 dual tropic are associated with T cell depletion and rapid disease progression (Scarlatti et al., 1997; Yu et al., 1998).

Out of the 135 samples collected from the Northern Kenyan population, 78 sequences aligned well with the references and were successfully analyzed for determination of coreceptor usage. However, the results indicated possibility of absence of gp120 env V3 region in 57 sequences. This was due to variation of the 57 test sequences from the reference sequences used to design the classifying software. The V3 sequences of CCR5 coreceptor using variants from different patients have been shown to be highly
homogeneous in contrast to the diversity seen in the V3 sequences of the CXCR4 coreceptor using isolates (Chesebro et al., 1992). During coreceptor usage determination by analysis of the V3 gene sequence all the divergent sequences default to CXCR4 usage (Sing et al., 2004) thus all the 57 sequences defaulted to CXCR4 usage. The variability of the 57 sequences can be attributed to the HIV-1 error prone reverse transcriptase enzyme and lack of a proof reading mechanism during reverse transcription of the viral RNA resulting to frequent incorporation of mutation in the viral genome (van Rij et al., 2000). This increasing diversity of HIV is considered to be one of the obstacles for development of a vaccine (Delwart et al., 1998). Moreover, the training data of the available classifying software is based on HIV-1 subtype B isolates predominant in the West (Sing et al., 2004) whose sequences may vary from the A1, C and D subtypes circulating in Northern Kenya. To avoid attrition due to these factors only the 78 properly aligned sequences were considered.

Analysis of the V3 gene region for the seventy eight properly aligned sequences showed 21.8% to be CCR5 using, 69.2% to be CXCR4 using and 9% to be dual/ CXCR4 and CCR5 using variants. The high proportion (78.2%) of the CXCR4 and CCR5/CXCR4 using variants indicated that majority of the infected individuals were most likely in their late stages of infection. The likely cases of acute infections were less than (21%) considering that approximately 50% of the HIV viruses do not switch coreceptor use (Connor et al., 1997). Positive behavior change and HIV control measure may have contributed to these result thus reduced HIV incidence rate hence few acute infections. This may further be reflected in the fall of the country’s HIV prevalence from 10% in the 90’s to the current 6.7% as indicated by the Kenya Demographic Health Survey 2003
(NASCOP, 2005). Most of these samples having been collected in the medical facilities could have majority of patient at late stages of HIV infection. It is however difficult to determine the stage of infection based on the fact that 50 % of the HIV viruses do not switch coreceptor use (Connor et al., 1997) and the subtype specific differences in coreceptor usage where subtype C predominantly express CCR5 usage while subtype D predominantly express CXCR4 coreceptor usage. Other studies suggest that the progression to AIDS in African population may be more rapid than in industrialized countries (Bwayo et al., 1995; Morgan et al., 1998). This is in line with the observation of this study where most of the isolates have already switched coreceptor usage indicating rapid progression to AIDS. However, because of relatively few studies, information about the change in coreceptor usage during infection in Africa HIV-1 patients is limited (Begaud et al., 2003).

5.3 Genetic Subtypes- Specific differences in Coreceptor Usage in Northern Kenya

Subtype-specific differences in coreceptors usage have been controversial (Zhang et al., 1996; Tscherning et al., 1998). A possible correlation between specific genetic subtypes and biological phenotypes has been studied (de Wolf et al., 1994; Nkengasong et al., 1995; Tscherning et al., 1998; Bjourndal et al., 1999). In the current study, a strong association between the coreceptor usage and HIV-1 subtypes was reported in Northern Kenya region. This is reflected in the proportions of HIV-1 variants using certain chemokine receptors as entry coreceptors. The CCR5 Coreceptor use was observed to be 52.6 % in subtype C, 9.8 % in subtype A and 0 % in subtype D variants. A high
percentage of CCR5 coreceptor using variants was observed in subtype C (52.6 %) as compared to subtype A1 (9.8 %) and D (0 %). This difference can be attributed to the fact that subtype C expresses mainly the NS1 phenotype and mostly use CCR5 as coreceptors independent of the clinical status of the patients (Tscherning et al., 1998; Bjournadal et al., 1999). It is also essential to note that 50 % of the recombinants were CCR5 using of which the majority (69.2 %) were subtype C recombinants. The high percentage of CCR5 coreceptor using isolates among the recombinant forms of which majority are Subtype C recombinants further indicate an association between CCR5 usage and subtype C.

Subtype A1 Variants were observed to use all classes of coreceptor with 4 (9.8 %) of the samples as CCR5 using, 34 (82.9 %) as CXCR4 using and 3 (7.3 %) as dual (CCR5 and CXCR4) coreceptor. Although CXCR4 using phenotype was significantly high (82.9 %) than the others, this shows a typical distribution of the coreceptor use where HIV-1 is normally transmitted as CCR5 using and eventually evolve to CXCR4 using via dual coreceptor (CCR5 and CXCR4) using phenotype (Fransje et al., 2002). High proportion of CXCR4 using phenotype in subtype A suggests that most infections were in advanced/late stages and rapidly progressing to AIDS.

Notably, all subtype D variants were CXCR4 using with none of the subtype D expressing dual tropism as compared to subtype A (7.3 %), subtype C (15.8 %) and the recombinants (16.7 %). The env V3 loop of subtype D has been shown to be significantly more divergent than other subtypes (Korber et al., 1994). Dual coreceptor usage is a transition stage where ultimately differentiated viruses are CXCR4 monotropic. Greater variability in the V3 region that is responsible for coreceptor use allows the virus to
switch to the syncytium inducing (SI) phenotype more rapidly hence bypassing the dual coreceptor usage stage. This is also in line with the previous observation that subtype D variants do not exhibit dual coreceptor use and are monotropic for CXCR4 after switch (Tscherning et al., 1998). Absence of any CCR5 using subtype from subtype D further indicates that all subtype D infections were in advanced stage thus having already switched to CXCR4 usage.
CHAPTER SIX: CONCLUSION AND RECOMMENDATIONS

6.1 Conclusions

a) The phylogenetic analysis of the gp120 C2V3 env region of the HIV-1 genome showed 27% of the viruses circulating in Northern Kenya to be subtype C, 54% to be subtype A1, 9% to be subtype D and 9.6% to be circulating recombinant forms. The dominant subtype circulating in Northern Kenya was reported to be HIV-1 subtype A1.

b) Analysis of the V3 region showed 21.8% to be CCR5 using, 69.2% to be CXCR4 using and 9% to be dual/ CXCR4 and CCR5 using variants. The dominant coreceptor used in Northern Kenya was that shown to be CXCR4.

c) Some of the HIV-1 variants circulating in Northern Kenya were shown to have major amino acid variations from the reference sequences indicating divergence in their env genome sequences.

d) The dominant coreceptor used was observed to be CXCR4 indicating most patients to be in advanced stages of infection.

e) The results also indicated strong association between HIV-1 subtypes and coreceptor usage. Subtype C had significantly high CCR5 coreceptor usage (52.6%) as compared to 9.8% in subtype A and 0% in subtype D. While subtype C and subtype A1 exhibited dual coreceptor usage (15.8% and 7.3% respectively), subtype D did not exhibit dual coreceptor usage and was predominantly CXCR4 monotropic.
6.2 Recommendations

a) Most of the patients were inferred to be in late stages of infections based on coreceptor usage data. Due to subtype specific differences in coreceptor usage, coreceptor data alone cannot be used to make treatment decisions hence, further investigations like CD4+ T cell counts and plasma viral load test should be carried to confirm their status in order to take the appropriate HIV management and therapeutic strategy.

b) Due to major amino acid variations from the reference sequences some sequences could not be analyzed with the available software. This shows that classifier softwares to be used in this region should be developed based on the circulating subtypes to facilitate analysis of variant amino acid sequences.

c) Having observed the different HIV-1 subtypes circulating in Northern Kenya and the divergence of their genome, vaccine development initiatives should develop a cocktail of antigens taking into account all circulating subtypes and their divergent genome.

6.3 Limitation of the study

The study was limited by lack of Clinical History, Demographic data and immune profiles data such as CD4 counts and Viral load that would have shed light on relationship between coreceptor use and Immune profiles as well as inter gender and age differences.

6.4 Hypothesis Testing

The null hypothesis- “There is no difference in coreceptor usage among different HIV-1 subtypes infected individuals from a population in Northern Kenya” - was not accepted.
REFERENCES


Sing, T., Beerenwinkel, N. and T. Lengauer (2004). Learning mixtures of localized rules by maximizing the area under the ROC curve. *16th European Conference on Artificial Intelligence (ECAI), Workshop on ROC Analysis in AI, 2004*


APPENDIX 1

Map of Kenya Showing the Study Areas (Adopted from www.map-zone.net)

KEY

- Areas of sample collection

Scale 1: 4 500 000
Appendix 2

Gp120 env sequences generated during the study indicating the V3 Region in Yellow highlight

> MADH012
TTTTGGGATTAAGCCGATGCTCAGTACTGCAATAGTTGTGGAATGGCAGTCTAGCAGAAGAAGATATAGTAATTAGGAACTGCAAATCTCACAAAATATAGATAGCTACATCATCAGGCAAGAGATCTGTAATTCGGATCTAAAAATCTGCCAGACAA

> MADH002
TACGCCGATGTATCACAATCAGCTACTGCTGTACATGGCAGTCTCGCAAGAAGAAGACATACTCATTAGGATCTGAAAATCTCACAAACAAATGCCTAATAGTTGGAATCTGGAATACATGTGCACAGGACAGTCTAATACCATCAGGCAAGAGATCTGAAATTCGGATCTAAAAATCTGCCAGACAA

> MADH005
GCCGATGTTCTCAACATCAACTCTACTGCTGTACATGGCAGTCTCGCAAGAAGAAGACATACTCATTAGGATCTGAAAATCTCACAAACAAATGCCTAATAGTTGGAATCTGGAATACATGTGCACAGGACAGTCTAATACCATCAGGCAAGAGATCTGAAATTCGGATCTAAAAATCTGCCAGACAA

> MADH007
ATTACCCAGTGGTGCAACTCAACTACTGCTGTACATGGCAGTCTCGCAAGAAGAAGACATACTCATTAGGATCTGAAAATCTCACAAACAAATGCCTAATAGTTGGAATCTGGAATACATGTGCACAGGACAGTCTAATACCATCAGGCAAGAGATCTGAAATTCGGATCTAAAAATCTGCCAGACAA

> MADH008
GATAACCAGTTTGCTACACTCAACTCTACTGCTGTACATGGCAGTCTCGCAAGAAGAAGACATACTCATTAGGATCTGAAAATCTCACAAACAAATGCCTAATAGTTGGAATCTGGAATACATGTGCACAGGACAGTCTAATACCATCAGGCAAGAGATCTGAAATTCGGATCTAAAAATCTGCCAGACAA

> MADH009
AGCCGATGTTCTCAACATCAACTCTACTGCTGTACATGGCAGTCTCGCAAGAAGAAGACATACTCATTAGGATCTGAAAATCTCACAAACAAATGCCTAATAGTTGGAATCTGGAATACATGTGCACAGGACAGTCTAATACCATCAGGCAAGAGATCTGAAATTCGGATCTAAAAATCTGCCAGACAA

> MADH010
AGCCGATGTTCTCAACATCAACTCTACTGCTGTACATGGCAGTCTCGCAAGAAGAAGACATACTCATTAGGATCTGAAAATCTCACAAACAAATGCCTAATAGTTGGAATCTGGAATACATGTGCACAGGACAGTCTAATACCATCAGGCAAGAGATCTGAAATTCGGATCTAAAAATCTGCCAGACAA

> MADH011
AGCCGATGTTCTCAACATCAACTCTACTGCTGTACATGGCAGTCTCGCAAGAAGAAGACATACTCATTAGGATCTGAAAATCTCACAAACAAATGCCTAATAGTTGGAATCTGGAATACATGTGCACAGGACAGTCTAATACCATCAGGCAAGAGATCTGAAATTCGGATCTAAAAATCTGCCAGACAA

> MADH012
AGCCGATGTTCTCAACATCAACTCTACTGCTGTACATGGCAGTCTCGCAAGAAGAAGACATACTCATTAGGATCTGAAAATCTCACAAACAAATGCCTAATAGTTGGAATCTGGAATACATGTGCACAGGACAGTCTAATACCATCAGGCAAGAGATCTGAAATTCGGATCTAAAAATCTGCCAGACAA

> MADH013
AGCCGATGTTCTCAACATCAACTCTACTGCTGTACATGGCAGTCTCGCAAGAAGAAGACATACTCATTAGGATCTGAAAATCTCACAAACAAATGCCTAATAGTTGGAATCTGGAATACATGTGCACAGGACAGTCTAATACCATCAGGCAAGAGATCTGAAATTCGGATCTAAAAATCTGCCAGACAA