Genetic diversity of HIV-1 in Central province, Kenya-Pilot study

Sheila Kageha

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

2009

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University

Signed…… Date ……

Sheila Kageha

This thesis has been submitted for examination with our approval as University Supervisors.

Signature…….. Date……....

Dr. Samoel A. Khamadi

KEMRI, Kenya

Signature…… Date……..

Dr. Jane Ngaira

JKUAT, Kenya

Signature…… Date.....……

Dr. Joseph K. Gikunju

JKUAT, Kenya

DEDICATION

I dedicate this work to my beloved mother, the late Jane Vuhya Nadida, a source of inspiration, a strong woman she was, God rest her soul.

ACKNOWLEDGEMENTS

This work would not have seen the light of the day were it not for the help of many people. My sincere gratitude goes to my supervisors: Dr. Samoel A. Khamadi for his constant guidance, resourcefulness and support throughout my graduate research, Dr. Jane Ngaira and Dr. Joseph Gikunju for their key ideas that provided a framework for this thesis.

The Kenya Medical Research Institute especially the HIV laboratory for providing an enabling environment without which this project would not have been feasible.

I am also indebted to Dr. Matilu Mwau for his invaluable advice and financial support.

Dr. Maritim E. Songok for providing a place for sequencing.

My special thanks go to my friends Violet Asiqo, Cecely Meeme and Laura Wangai for their encouragement.

All those who have made significant investment and immeasurable contributions to this work.

Lastly my uncle and his family who provided a home throughout my course work.

Unto you all, I am forever indebted!

TABLE OF CONTENTS

LIST OF TABLES`

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

AIDS	Acquired Immune Deficiency Syndrome
ARV	Antiretroviral
BLAST	Basic local alignment search tool
CCR5	Cysteine-Cysteine linked chemikine receptor 5
CD4	Cluster of Differentiation Antigen number 4
CMV	Cytomegalovirus
CRFs	Circulating recombinant forms
CTL	Cytotoxic-T-lymphocyte
ddH_2O	Double Distilled Water
DNA	Deoxyribonucleic Acid
DNTPs	Deoxyribonucleotide Triphosphate.
EBV	Epstein Barr virus
EDTA	Ethylenediaminetetraacetic Acid
Env gp 41	Envelop glycoprotein 41
FeLV	Feline leukemia virus
Gag	Group-specific antigen
Gp120	A sugar containing glycoprotein of approximately 120 daltons
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immunodeficiency Virus
HLA	Human Leukocyte Antigen
HTLV	Human T-lymphotropic virus
JKUAT	Jomo Kenyatta University of Agriculture and Technology

KAIS	Kenya AIDS Indicator Survey
KDH	Kiambu District Hospital
KEMRI	Kenya Medical Research Institute
KGDH	Kerugoya District Hospital
KMH	Kikuyu Mission Hospital
KNMH	Kieni Mission Hospital
LAV	Lymphadenopathy Associated Virus
LTR	Long terminal repeat
MDH	Maragua District Hospital
$MgCl_2$	Magnesium Chloride
MHC	Major Histocompatibility Complex
NaOAc	Sodium Acetate
Nef	Negative factor gene
NCBI	National Center for Biotechnology Information
NPGH	Nyeri Provincial General Hospital
P10	Protein 10
P17	Protein 17
P24	Protein 24
P32	Protein 32
P64	Protein 64
PBMCs	Peripheral Blood Mononuclear Cells
PCR	Polymerase chain reaction
Pol	HIV polymerase gene
Rev	Regulator of virion gene

RNA	Ribonucleic Acid
RT	Reverse transcriptase
SsRNA	Single stranded ribonucleic acid
Tat	Trans-Activator of transcription
TDH	Thika District Hospital
TMH	Tumutumu Mission Hospital
UNAIDS	United Nations Programme on HIV/AIDS
URFs	Unique Recombinant Forms
Vif	Viral infectivity factor
Vpr	Viral protein regulatory gene
Vpu	Viral protein unknown gene
Vpx	Viral protein x gene
WHO	World Health Organization

ABSTRACT

Reports on emergence of new strains characterize HIV epidemic in Sub-Saharan African countries. The distribution of these diversified strains is thought to follow geographic and population dynamics. These HIV variants have been reported to have distinctively regionalized endemicity. In Kenya, studies on subtype distribution have been done in Western, Nyanza, North-Eastern, Rift Valley, Nairobi and Coastal Provinces. However, this information in Central Province is lacking. This work reports genetic diversities of HIV-1 circulating in the Central Province of Kenya. HIV positive blood samples were collected from 8 health facilities in the province namely Nyeri, Maragua, Kiambu, Thika, Kerugoya, Kikuyu, Tumutumu and Kieni Mission hospital. A total of 96 samples were collected, 12 from each health facility. Part of the proviral HIV-1 *env* gene (*gp41*) was PCR amplified and directly sequenced. Subtypes were determined by sequence analysis using geographically diverse subtype reference sequences as well as sequences of known subtypes from Kenya. Sixty eight percent (68 %) of the samples analysed were subtype A-1, 10% were subtype C, 12% subtype D, while the rest were circulating recombinant forms (CRFs) mainly between C, D, G and A. This analysis of HIV-1 strains demonstrated the predominance of HIV-1 subtype A-1. Understanding the genetic diversity of HIV-1 in this region is important in monitoring the spread of infection and developing effective control strategies. This pilot study provides the basis for more expansive studies in determining HIV genetic diversity in Central province of Kenya. It also sets the benchmark for future studies in viral phylogeography.

CHAPTER ONE

1.0 INTRODUCTION

1.1 HIV AND AIDS

The Human Immunodeficiency Virus (HIV) is the etiological agent of Acquired Immunodeficiency syndrome (AIDS). The earliest documented case of HIV infection occurred in 1959 with the epidemic phase becoming apparent in 1981 (Gallo *et al*., 1984).AIDS is defined by tremendous decline in CD4+ cells (Coffin *et al*., 1986; Gallo and Montagnier, 2003) followed by onset of opportunistic infections and malignancies such as Kaposi's sarcoma (Ascher and Sheppard, 1988).

Since its discovery in 1983, HIV has continued unabated to trap its victims in to the vicious cycle of disease and death. According to the latest statistics, it is estimated that 33.2 million people worldwide are infected and currently living with HIV, resulting in about 2.1 deaths annually (WHO, 2007) and thereby placing HIV as a major threat to global health and prosperity (Greener, 1998).

1.2 HIV Origin and History

The origin of HIV has remained a subject of contention over the years since its first discovery in the human populations. Various theories have been put forward in a bid to explain the emergence of HIV, the most infamous one involved the alleged use of HIV contaminated oral polio vaccine in Africa during the late 1950s (Curtis, 1992; Worobey *et al.*, 2004). This was later disputed following findings showing that the Congo region plays a pivotal role in the genesis of HIV (Gottlieb *et al*., 1981; Masur *et al*., 1981; Rambaut *et al.*, 2001).

Two types of HIV exist in the world, HIV-1 and HIV-2. The two are similar in their genomic structures but can be distinguished by their sequences particularly in the envelope glycoprotein. Circumstantial evidence suggests that HIV-1 emerged from Africa and spread worldwide (Gilbert *et al*., 2007).

HIV-1 is believed to have entered the human population through cross-species transmission, linked to human consumption of nonhuman primates in Africa (Gao *et al*., 1999; Nerrienet *et al*., 2005; Sharp *et al*., 2001). HIV-1 is closely related to Simian Immunodeficiency virus (SIVcpz) infecting chimpanzees (Huet *et al*., 1990) whereas HIV-2 seems to have originated from Simian Immunodeficiency virus (SIVsm) infecting sooty mangabees (Hirsch *et al*., 1989). Recombinant viruses emerged later. Nevertheless, none of these theories have been fully substantiated and thus the outbreaks of HIV -1 and HIV-2 infections in the last century have not been conclusively explained.

In the early 1980s physicians observed an unusual increase in the incidence of Kaposi's sarcoma, *Pneumocystis carinii* pneumonia and persistent lymphadenopathy of unknown origin in homosexual men who had otherwise been healthy (Modlin *et al*., 1983; Picard *et al*., 1982). These patients shared significant loss of CD4+ helper T cells that led to an impairment of cellular immune responses with increased susceptibility to opportunistic infections. The causative agent of the AIDS epidemic was originally unknown and it was speculated that Cytomegalovirus (CMV) or Epstein Barr virus EBV) were responsible (Heng, 1983; Mathe, 1983). By 1983 the causative agent for AIDS was thought to be a retrovirus with similar characteristics to Feline leukemia virus (FeLV)

and Human T-lymphotropic virus (HTLV-1 and 2) (Evatt *et al*., 1983; Gallo, 1983). Like these recently discovered retroviruses, it was transmitted through sexual contact, contact with blood products and body fluids and led to immunosuppression.

In early 1983, Luc Montagnier's group isolated a new retrovirus from a cervical lymph node of a homosexual patient which they named Lymphadenopathy Associated Virus (LAV) (Chermann *et al*., 1983). This virus was cytopathic for lymphocytes, T-helper cell tropic and caused cell death. Robert Gallo's group successfully grew CD4 T cells from a man suffering from AIDS. These cells yielded both HTLV-1 and what later came to be known as HIV-1. In the course of 1984 a virus with the same properties was described and named HTLV-III (Gallo, 1983; Gallo *et al*., 1984). HTLV-III antibodies could be isolated almost always in people suffering from AIDS. Workers in San Francisco isolated a similar virus that they named Aids Related Virus in the same year (Levy *et al*. 1984).

In the same year HIV proteins were analysed and characterized, and by 1985 the nucleotide sequences of these new retroviruses were confirmed to be similar (Marx, 1985; Ratner *et al*., 1985). In 1986 the viruses were named the Human Immunodeficiency Virus (HIV) by the International Committee of Viral Taxonomy (Coffin *et al*., 1986). At this time, it was not clear whether HIV was a new virus or a virus that had suddenly become more virulent, but since then it has been established that HIV may have originated as a zoonosis in the non-human primates living in the rain forests of Central and West Africa (Sharp *et al*., 2001).

1.3 THE HIV STRUCTURE AND GENOME

Based on morphologic, genetic and biological properties, HIV is classified in a subgroup of retroviruses called the lentiviridae (Garry, 1989; Levy, 1993). HIV possesses a complex genome, approximately 9.8 kb long (Marx, 1985) (**Fig.1.1** gives a graphical representation of the HIV genome).

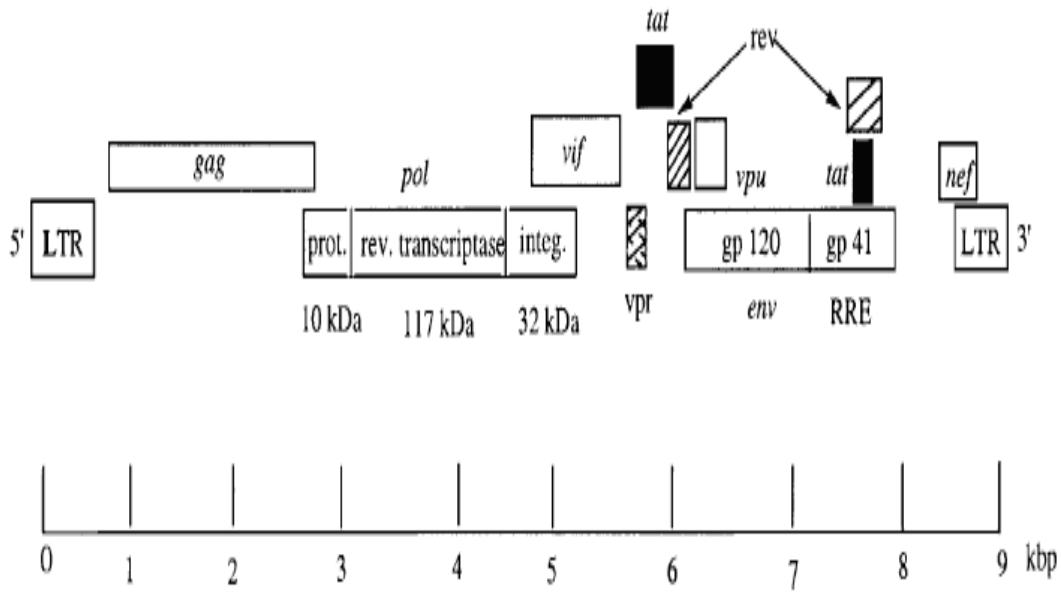

Figure 1.1: Proviral HIV Genome. Adapted from www.medguide.org.zm/aids/aids.html

HIV viral particle is composed of two copies of positive single-stranded RNA (ssRNA) that codes for the virus' nine genes classified into 3 functional groups: Gag and Env are structural genes that code for structural proteins; Pol genes coding for the three enzymes (Reverse transcriptase, integrase and protease); Tat and Rev are regulatory genes; Vpu, Vpr, Vif, and Nef are accessory genes. An open reading frame, *vpu* separates the *pol*

and *env* regions (Garry, 1989). The HIV provirus is a double stranded DNA integrated in the host cell genome.

The Envelope glycoprotein is composed of the two non-covalently linked subunits, SU (surface glycoprotein) and TM (transmembrane glycoprotein) largely responsible for host cell recognition and entry respectively (**Fig 1.2**). This glycoprotein complex enables the virus to attach to and fuse with target cells to initiate the infectious cycle (Chan *et al*., 1997). These surface proteins have been considered as targets for future treatments or vaccines against HIV.

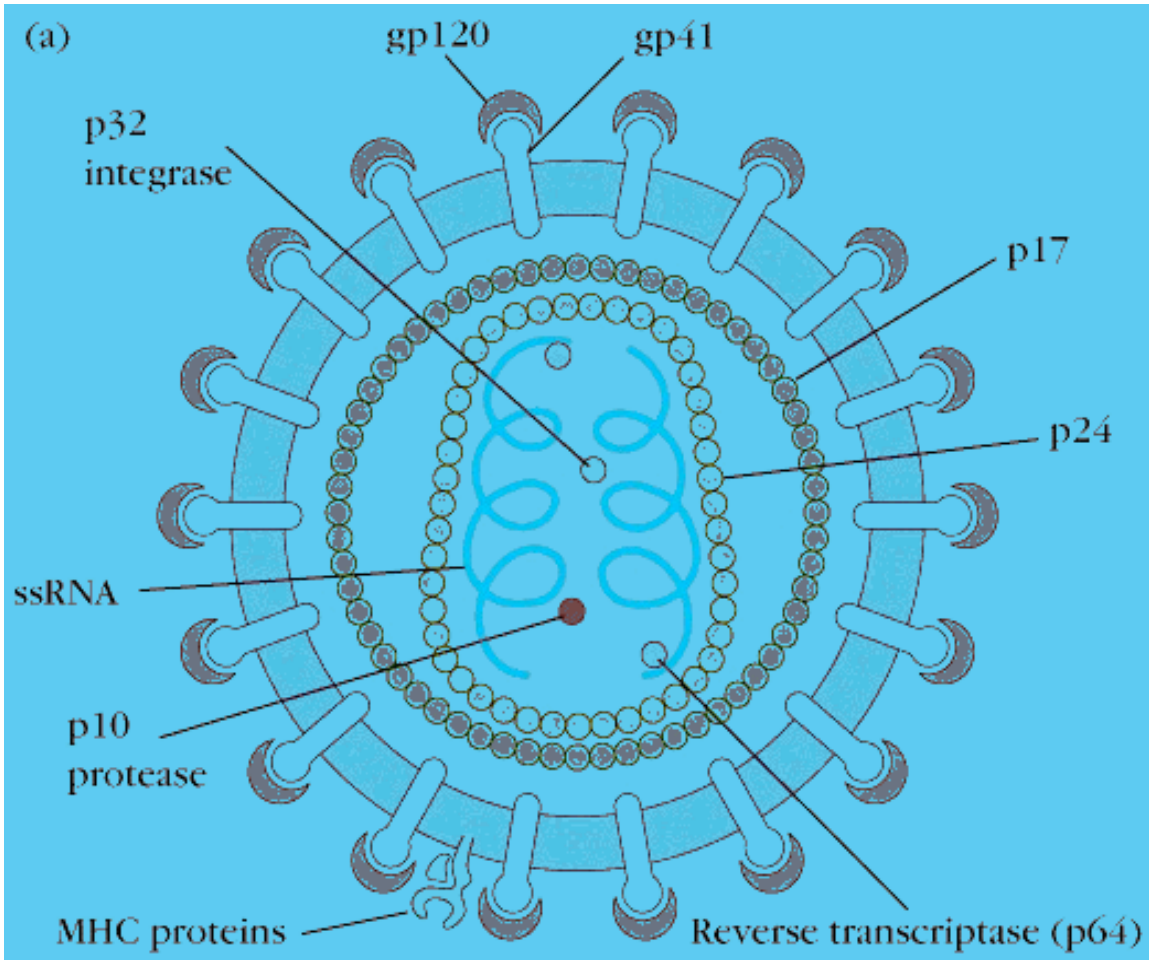

Figure 1.2: The Human Immunodeficiency Virus particle. Adapted from Duane (2002).

1.4 HIV REPLICATION CYCLE

HIV replication involves reverse transcription of viral RNA into a DNA copy and subsequent integration of the viral genome into the host cell chromosome. Thus the virus-infected cell is an important reservoir for HIV infection in the host. HIV genetic information is carried as RNA. The HIV envelope surface glycoproteins are strongly attracted to the CD4+ surface receptor expressed on T lymphocytes. HIV binding to CD4+ surface receptor activates other proteins on the cell's surface, allowing the HIV envelope to fuse to the outside of the cell. After the binding process, the viral capsid (the inside of the virus which contains the RNA, reverse transcriptase, integrase and protease) is released into the host cell. The HIV infectious virion once inside a host cell reverse transcribes its genetic material into DNA for integration into the host cell's own genome (Chan and Kim, 1998). This new DNA is called "proviral DNA."

Once HIV's genetic material is inside the cell's nucleus, it directs the cell to produce new HIV. The strands of viral DNA in the nucleus separate and special enzymes create a complementary strand of genetic material called messenger RNA which gives instructions for making new viral proteins.

As each mRNA strand is processed, a corresponding string of proteins is made. This process continues until the mRNA strand has been transformed or "translated" into new viral proteins needed to make a new virus. Finally, a new virus is assembled. Long strings of proteins are cut up by a viral enzyme (protease) into smaller proteins. These

proteins serve a variety of functions; some become structural elements of new HIV, while others become enzymes, such as reverse transcriptase.

Once the new viral particles are assembled, they bud off the host cell, and create a new virus. This virus is then able to infect new cells. Each infected cell can produce a lot of new viruses. **Figure 1.3** below illustrates how the virus attaches to host cell via co-receptor and its replication producing many virions that bud off to infect new cells

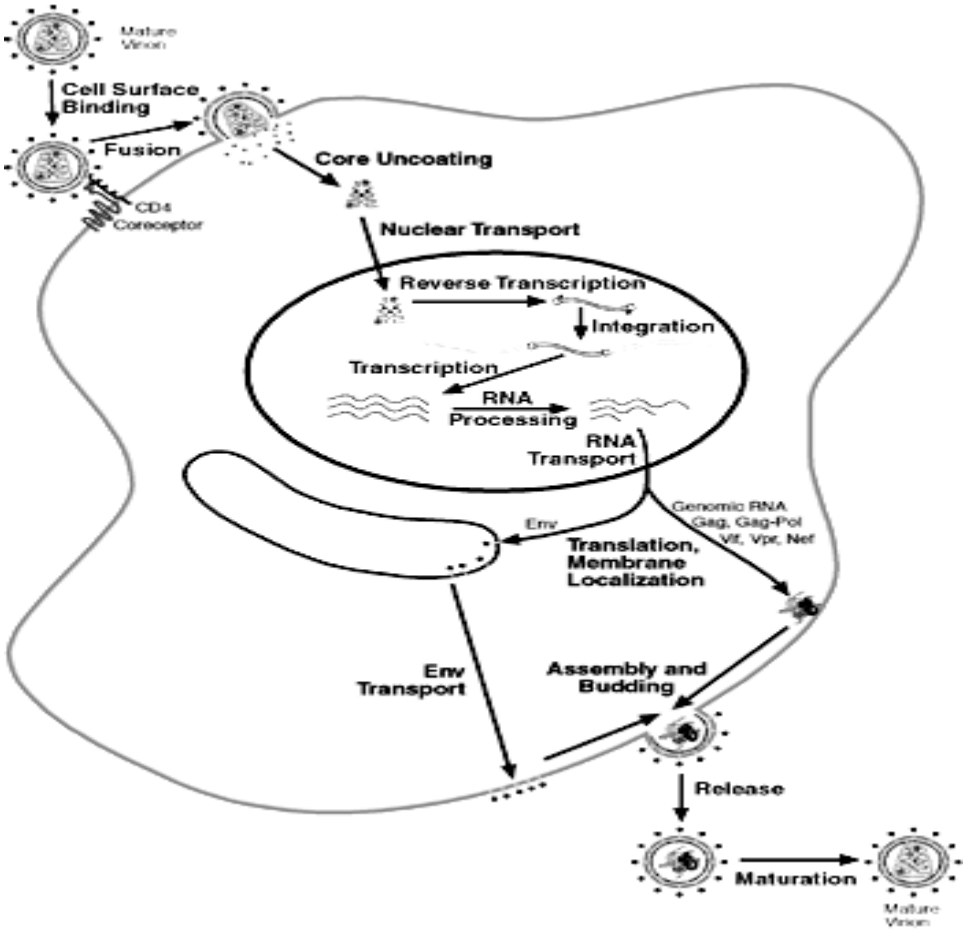

Figure 1. 3: HIV entry in to the host cell and transportation. Adapted from www.brookscole.com/.../ch_HIVLifecycle.html.

1.5 HIV Epidemiology

The HIV pandemic remains the most serious of infectious disease challenges to public health. The statistics describing the worldwide spread of HIV/AIDS are staggering. The global prevalence of HIV infection has been increasing steadily, although the global number of persons living with HIV is increasing because of ongoing accumulation of new infections with longer survival times, partly attributable to the scaling up of antiretroviral treatment services (WHO, 2007). A joint UNAIDS and WHO report estimated the number of persons living with HIV worldwide in 2007 to be 33.2 million. (UNAIDS, WHO. 2007). **Table 1** gives the global summary of the AIDS epidemic.

Sub-Saharan Africa remains the most seriously affected region, with AIDS remaining the leading cause of death. The estimated number of deaths due to AIDS in 2007 was 2.1 million [1.9–2.4 million] worldwide, of which 76% occurred in sub-Saharan Africa. Total number of people living with the virus is approximately 20.9–24.3 million. According to the Kenya AIDS Indicator Survey, more than 1.4 million (7.4%) Kenyans are currently living with HIV/AIDS (KAIS report, 2007). Geographic distribution of HIV infections vary greatly across Kenya with Nyanza Province recording the highest prevalence. Central province recorded a prevalence of 3.8%.

Table 1: Global summary of the AIDS epidemic. Adapted from WHO (2007).

Number of people living with HIV in 2007	
Total	**33.2 million [30.6–36.1 million]**
Adults	30.8 million [28.2–33.6 million]
Women	15.4 million [13.9–16.6 million]
Children under 15 years	2.5 million [2.2–2.6 million]
People newly infected with HIV in 2007	
Total	**2.5 million [1.8–4.1 million]**
Adults	2.1 million [1.4–3.6 million]
Children under 15 years	420 000 [350 000–540 000]
AIDS deaths in 2007	
Total	**2.1 million [1.9–2.4 million]**
Adults	1.7 million [1.6–2.1 million]
Children under 15 years	330 000 [310 000–380 000]

1.6 PROBLEM STATEMENT

HIV-1 subtypes are dynamic and new divergent strains are evolving. Transmission and distribution of these diversified HIV-1 strains follow geographic and socio-demographic characteristics. Studies on HIV-1 genetic diversity have been carried out in other provinces within the country including Western (Songok *et al*., 2003), Northern (Khamadi *et al*., 2005), Nairobi (Lihana *et al*., 2006) , Nyanza (Dowling *et al*., 2002), Coast (Khamadi *et al*., 2008), According to these studies, HIV-1 subtype A has been found to be the most common followed by other strains in significant proportions. The findings have shown considerable genetic diversities in regard to HIV-1 strains circulating in Kenya. However there is no published data on the circulating HIV-1 subtypes and their recombinants in Central Province. This presents a challenge in regard to HIV control and management strategies and therefore the need to monitor the circulating subtypes.

1.7 JUSTIFICATION

Demographic structures in the population, screening techniques, increasing resistance to the prevalent antiretroviral drugs, a vaccine which seems to be the only hope but still elusive are some of the factors compounding the scientific challenge of outsmarting an organism fully equipped with nature's set of survival tools. The relationship between genetic diversification, immune response, and disease progression is not clearly understood.

Currently there is no vaccine or cure for HIV/AIDS and intervention programs are facing enough challenges as to adherence, monitoring and emerging drug resistance. Biological consequences of HIV genetic variation and its implications with regard to epidemiology, diagnostics, classification and vaccine design needs to be well understood. Therefore, classification based on genetically defined subtypes provides an important framework for making advances on understanding viral biology and immunology.

Diagnostic tests, HIV vaccines, and antiretroviral drugs are mainly designed and developed in the industrialized countries of Europe and North America. In these parts of the world, the B subtype is by far the most common HIV-1 subtype (Perrin *et al*., 2003). Tests, vaccines, and drugs based on research on subtype B viruses might perform sub-optimally in populations where non-B subtypes are circulating, typical in Sub-Saharan Africa.

Because HIV infection is diagnosed with serological tests of antigen/antibody reactions, even subtle changes in antigenic structure may affect the sensitivity of these tests. In particular, several serological antibody assays have been reported to lack sensitivity to antibodies to some group O viruses, which also frequently produce indeterminate Western blot results (Loussert-Ajaka *et al*., 1994; Schable *et al*., 1994).

The strains reiterate the importance of implementing proper diagnostic policies to protect the blood bank beneficiaries. It is therefore necessary to keep HIV molecular surveillance programs, checking systematically the distribution of the HIV-1 subtypes, as well as the introduction of new variants. It is for these reasons that the information

obtained from this investigation will be useful, in the future, in an attempt to design cross-clade HIV-1 vaccine to reflect viral diversity (Gaschen *et al*., 2002), and predict future complexity of regional and global HIV-1 genetic diversity.

1.8 HYPOTHESES

1.8.1 Null Hypothesis

There is no difference in HIV-1 subtypes circulating in Central province of Kenya from those already identified in other provinces in the country.

1.8.2 Alternative Hypothesis

HIV-1 subtypes circulating in Central Province of Kenya are different from those already identified in other Provinces in the country.

1.9 OBJECTIVES

1.9.1 General Objective

To determine genetic diversity of HIV-1 strains circulating in Central Province of Kenya.

1.9.2 Specific Objectives

i). To determine the subtypes of HIV-1 circulating in the central province of Kenya.

ii).To determine phylogenetic relationships of identified HIV subtypes in Central Province

iii).To compare the subtypes with the ones already identified in other Provinces in Kenya.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Distribution and Classification of HIV

The global distribution of HIV is complex and dynamic with regional epidemics harboring only a subset of the global diversity. HIV-1 is extraordinarily variable and this diverging trend of HIV subtypes, mosaic viruses and potential recombinants poses a major challenge in the design and testing of HIV vaccines. HIV infections in the early years have been very low until the last 20 years when the rate began to rise exponentially. Multiple transmissions have since emerged with vastly different outcomes; some having caused major global pandemics while others appear to have generated almost no appreciable spread in humans. HIV-1 undergoes mutations that allow escape from cytotoxic T lymphocytes (CTL)-mediated restriction (Bailey *et al*., 2006).

At the intrahost and interhost levels, HIV exhibits rather different evolutionary rates. This dynamical relationship has two possible causes: (*i*) that intrahost evolution, in contrast to that occurring among hosts, is dominated by the positive selection of amino acid changes that facilitate immune escape and that elevate rates of evolutionary change over that expected under neutral genetic drift (Nielsen and Yang, 1998), and/or (*ii*) that most of the mutations that occur within hosts are purged at transmission to new hosts because of strong purifying selection in this new environment. For instance, a proportion of the HIV genome appears to be ‘‘reset’’ at interhost transmission because of mismatches between mutations that confer escape from host cytotoxic T lymphocyte

responses and the HLA type determining the specificity of that response (Li *et al*., 2007).

The different HIV-1 subtypes are generally geographically distinct with Sub-Saharan Africa containing all known HIV-1 subtypes (Peeters, 2001). While HIV-2 has been basically restricted to the African continent, HIV-1 is associated with the worldwide AIDS pandemic. The number of unique HIV-1 sequence in public databases has been steadily increasing every year (Altschul *et al*., 1997)

HIV is a member of the retroviridae family with two sub-families; HIV-1 and HIV-2. HIV-1 is more prevalent and more pathogenic than HIV-2. HIV-1 is responsible for infections globally, whereas HIV-2 is found predominantly in West Africa. HIV classification is based upon the variable viral structure and different strains of HIV-1 have been documented (Robertson *et al*., 2000). At a global level, the genetic diversity of the virus is manifest in distinct HIV subtypes or clades that cluster epidemiologically in distinct geographic regions (Perrin *et al*., 2003).

Phylogenetic analyses of virus samples from different geographic regions have revealed that HIV-1 can be divided in to three groups: O (outlier), N (New) and the M group (Major), with each lineage representing a separate transmission from chimpanzees in central Africa (Thomson *et al*., 2002). The major HIV pandemic is caused by HIV-1 group M strains which have diversified into genetic subtypes (Arien *et al*., 2005a)**.** M, the major HIV group, seems to have entered the human population probably around

1931 and the B clade common in the United States may have emerged around 1941 (Hillis, 2000; Korber *et al*., 2000).

Full-length genome sequencing shows that group M viruses can further be subdivided into at least nine distinct, non recombinant subtypes (A, B, C, D, F, G, H, J, and K) approximately equidistantly related with intrasubtype divergence of up to 20% and intersubtype divergence of between 25% to 35%, for the *env* amino acid sequences (Achkar *et al*., 2004; Robertson et al., 2000; Triques *et al*., 2000). Groups O and N are very rare and essentially limited to Cameroon in Central Africa. The HIV-2 epidemic remains limited, both in terms of prevalence and with respect to geographic spread and its strains are classified into five subtypes; A through E (Zhang *et al*., 2005).

From the documented 800 complete genome sequences,743 sequences represent 9 genetic subtypes,43 circulating recombinant forms (CRF) and a variety of unique recombinant forms (URF) (McCutchan, 2006). Approximately 30% of the amino acids in the gp120 protein differ between clades (Peeters, 2001; Peeters and Delaporte, 1999). Clade B is prevalent in the western countries. Clades A and D are commonly found in sub-Saharan Africa, clade C in India, and clade E in Southeast Asia.

2.2 Global Distribution of HIV-1 Subtypes

The distribution of HIV subtypes globally is varied. In America and Western Europe, subtype B predominates everywhere but in eastern South America, there are substantial proportions of BF recombinants in addition to subtype B (Arien *et al*., 2007), (Burke, 1997; Carr *et al*., 2005). In Europe, subtype B is predominant in homosexuals, while a

variety of subtypes are found in a relatively small number of people infected through heterosexual contact (Ortiz *et al*., 2000). HIV-1 subtype B has also been noted in Indonesia, the Philippines, and Taiwan (Ortiz *et al*., 2000). Subtype F occurs in Romania, and subtype C is found in a small proportion in Brazil (McCutchan *et al*., 2000).

In Eastern Europe, subtypes of A, B, and AB recombinant strains dominate the epidemic. Three different patterns have been observed in Asia: subtype C, a mixture of B, C, and BC recombinants, and a mixture of subtype B and CRF01_AE. The Australian HIV-1 epidemic is caused by subtype B. HIV-1 subtypes A and D are found mainly in Africa while C is found mainly in Africa and Asia (Walker *et al*., 2005).

Imbalances in the prevalence of HIV exist with Sub-Saharan Africa bearing the largest burden in terms of HIV diversity. Africa has all HIV-1 subtypes although A and C predominate (Nasioulas *et al*., 1999). HIV-1 subtype C dominates the South and East, except for significant foci of subtypes A and D (Thongcharoen *et al*., 2007). Subtype G has been documented in many West and Central African countries, whereas subtype H was found in central Africa (Peeters and Delaporte, 1999).

The intermixture of HIV-1 variants that circulate together within a geographical region provides an opportunity for recombination of virus strains within dually infected individuals. In certain populations and regions where multiple HIV-1 subtypes co-circulate, many combinations of intersubtype recombinant viruses have been documented (A/C, A/D, B/F, A/G/I/J) (Montavon *et al*., 1999), and even an intergroup

M/O recombinant virus has been recently isolated from a Cameroonian patient (Peeters and Delaporte, 1999).

In 2000, the last year in which an analysis of global subtype prevalence was made, 47.2% of infections worldwide were of subtype C, 26.7% were of subtype A/CRF02_AG, 12.3% were of subtype B, 5.3% were of subtype D, 3.2% were of CRF_AE, and the remaining 5.3% were composed of other subtypes and CRFs (Osmanov *et al*., 2002).

Subtype A finds its highest concentration in East Africa and in Central Asia and Eastern Europe. HIV-1 subtypes A and D have dominated Uganda from mid 1980s (Hu *et al*., 2000). In Southern Tanzania town of Mbeya, the predominant subtype is C (Hoelscher *et al*., 2001). This probably reflects introduction from Southern African countries where subtype C is predominant (De Baar *et al*., 2000). HIV-1 subtypes A and D have been isolated from other parts of Tanzania (Blackard *et al*., 1999). Western Africa harbours all the documented subtypes of HIV-1. West and West Central Africa harbours mainly CRF02_AG, alongside a complex array of other recombinants, each present at a low frequency. The most complex epidemic is in Central Africa, where rare subtypes and a wide variety of recombinant forms circulate without any discernable predominant strain.

2.3 HIV-1 Subtypes Circulating in Kenya

In Kenya, the circulating strains have exhibited an extraordinary degree of genetic diversity with population-level phylogenetic patterns reflecting both transmission dynamics and genetic change; thought to have accumulated due to selection pressure.

Subtype A predominates in both non-recombinants (55%) and recombinants based on near full-length sequences that have been generated (Dowling *et al*., 2002).

Genetic studies carried out in 2006 (Lihana *et al*., 2006) and 2008 (Land *et al*., 2008) in Nairobi indicated that subtype A1 was the most prevalent (70%), D (9%) and C (6%). Circulating recombinant forms were about 15%. A research study based on the analysis of the *env* C2-V3 region (Neilson *et al*., 1999), revealed that subtype A predominates (71-87%), with significant components of subtype D (7-29%) and subtype C (7-17%).

Subtype C and D occur as non-recombinant (2% each) but to a much lesser extent than subtype A. Also, a full-length subtype G has been found previously in Kenya (Carr *et al* 1998), with recombinants between A1, A2, and D; A2 and D; A1 and D; A1 and G; A1 and C; A1, C and D (Dowling *et al*., 2002). Genetic studies, based on partial *env* sequences, carried out in Northern Kenya demonstrated dominance of HIV-1 subtype A (50%), subtype C (39%) and subtype D (11%). Subtypes A and C were found to be the most dominant (Khamadi *et al*., 2005).

HIV-1 subtype A-1 is the dominant strain in Kenya followed by D, C, G and recombinants AD, AC, CRF02-AG , and CRF16-A2D. This trend suggests that the HIV-1 epidemic may be evolving toward more virulent and complex subtypes through transmission of complex recombinants due to viral mixing. This suggest that any use of ARVs may therefore require initial testing for de novo resistance before commencement of treatment and/or management (Burke, 1997; Lihana *et al*., 2006).

2.4 Subtype Diversity In The Face of Antiretroviral Therapy

Highly Active Antiretroviral therapy (HAART) has significantly reduced the morbidity and mortality associated with HIV by effectively controlling disease progression and prolonging survivorship. However, these benefits have been compromised by the development of drug resistance, a consequence of mutations that merge in the viral proteins targeted by antiretroviral agents. This has continuously posed the greatest challenge to management of HIV positive patients (Little *et al*., 2002). Several factors, such as the dynamics of HIV replication, host genetics, physician practice and drug-related factors have also led to the emergence of HIV resistant strains in patients.

The rapid development of drug resistant HIV strains has been linked to the high turnover of HIV. The most successful current antiviral therapy used to treat HIV infection is the combination therapy (triple therapy). This regimen of three drugs, usually two reverse transcriptase inhibitors and a protease inhibitor is used to suppress viral replication (Carpenter *et al*., 1998).

Combination therapy has helped to stabilize progression and rates of HIV-1 subtype B infections in many western countries. Worldwide epidemics with group M (non-B, A through J) and O subtypes are expanding (Louwagie *et al*., 1995; Myers, 1994). Standard ARV regimens used in the treatment of HIV-1 subtype B are effective against all group M subtypes. HIV-1 type O and HIV-2 are naturally resistant to non-nucleoside reverse transcriptase inhibitors (Descamps *et al*., 1997; Descamps *et al.,* 1995; Quinones-Mateu *et al*., 1998; Tantillo *et al*., 1994).

The prevalence of mutations already present in treatment-naïve patients differs among demographic regions. Resistance was primarily observed in recent infections with subtype B (Wensing *et al*., 2005). In a study in 40 cities in the United States of America, 14% of the 371 isolates of treatment naïve patients had at least one primary mutation (Ross *et al*., 2007). In San Francisco and in a Spanish study, the prevalence of resistance among patients with acute or recent infection was even higher, at 26% and 19% respectively (Blower *et al*., 2003, de Mendoza *et al*., 2002).

HIV infected individuals under therapy are assumed to be non infectious due to low viral loads (Vernazza *et al*., 2000). However, individuals under treatment engender drug-resistant strains at an annual rate of 25%, after which they become infectious again and spread resistant strains. After resistance has developed, there are no benefits from treatment.

2.5 Sources of HIV variation

HIV differs from many other viruses as it has very high genetic variability attributed to its fast replication cycle, coupled with a high mutation rate and recombinogenic properties of reverse transcriptase (Osmanov *et al*., 2002). This complexity leads to the generation of many variants of HIV in a single infected patient (Osmanov *et al.,* 2002; Robertson *et al*., 1995). This scenario is further compounded when a single cell is simultaneously infected by two or more different strains of HIV. This hybrid virion then invades a new cell where it undergoes replication (Osmanov *et al*., 2002).

HIV retroviral transcription enzyme, reverse transcriptase, exhibits a low level of fidelity resulting in an error rate of between 1/1000 and 1/10000 nucleotide misincorporations (Takeuchi *et al*., 1988). Given a 9000–10000 base pair genome indicates that progeny viruses contain between 1 and 10 polymorphisms per genome per round of replication. HIV-1 replicates to high levels in individual untreated patients (Ho *et al*., 1989) which coupled with the size of the HIV-1 pandemic results in viral quasi-species of HIV-1 within an infected individual and a global distribution of viruses that are genetically divergent (Wain-Hobson, 1993).

Recombination occurs because reverse transcriptase copies the two RNA molecules packaged into the virion alternately, generating a mosaic DNA genome. Recombination is a major force in viral evolution, not only in the individual but also in the global epidemic (Robertson *et al*., 1995; Zhang *et al*., 2005). Because of the inaccuracy of the replication machinery of the virus, new mutations are introduced into virtually every virion generated in an infected individual.

Compared to SIV a different behavior is exhibited, for instance, in its natural hosts, the retrovirus is present in high levels in the blood, but elicits only a mild immune response and does not lead to the development of simian AIDS. The SIV does not undergo the extensive mutation and recombination typical of HIV-1 (Baier *et al*., 1991).

The role of genetic variations in influencing HIV transmission and disease progression has attracted a great deal of research interest, with the suggestion that the rapidly spreading HIV epidemic could be due to variations in circulating HIV-1 strains or subtypes. According to studies carried out in Rakai, Uganda, it was found that the risk

of progression to death in persons infected with HIV-1 subtype D and recombinant subtypes or multiple subtypes was higher than those infected with subtype A (Kiwanuka *et al*., 2008).

2.6 Implications of HIV genetic variations

The implications of these variations are many, ranging from false serological or immunological diagnosis of HIV to difficulties in developing vaccines and other management strategies. The various strains of HIV are also believed to play major roles in influencing transmission, prognosis, disease progression and therapeutic intervention success. Observations that different HIV subtypes exhibit different rates of disease progression are a clear indication that efforts to contain the epidemic are incomplete without information on the circulating subtypes.

In contrast to HIV-1 infection, which is spread throughout the continents, the HIV subtypes are associated with specific modes of HIV infection and transmission rates. Also, according to studies, some subtypes such as subtype A is most transmissible (Bobkov *et al*., 2004a; Bobkov *et al*., 2004b; Zhu *et al*., 1998). HIV-2 is primarily restricted to West Africa and to population movements from or through this region.

Despite ongoing prevention efforts, HIV continues to spread unabated in many parts of the world. The development of a vaccine against HIV is recognized as one of the most promising and a cost-effective; in addition to the current arsenal of HIV control measures. The most effective type of vaccine would include immunogenic regions or epitopes of the HIV-1 genome that are highly conserved across clades and strains of

HIV-1 (Rowland *et al.,* 1998). The most challenging aspect of the biology of HIV for those attempting to develop an AIDS vaccine is the extraordinary genetic diversity of the virus (Johnston and Fauci, 2007; Malim and Emerman, 2001; Richman *et al*., 2003). This diversity is apparent both in a single infected individual and at a global level in geographically disparate infected people.

Until recently, discovery of conserved epitopes in the HIV-1 genome has been hampered by lack of effective tools that would enable researchers or vaccine developers to mine large HIV-1 protein sequence databases for vaccine components (Gaschen *et al*., 2002). Both cross subtype-specific immunity and subtype-specific immune responses have been reported.

However, vaccinating with a single HIV-1 strain may not be a successful means of protecting against challenge by strains belonging to other clades of HIV-1 Therefore A thorough analysis of the sequence of transmitted viruses and the structures of their envelopes may yield clues to help guide vaccine design. Such a vaccine would induce immune responses that prevent the establishment of HIV infection by clearing virus before latent viral reservoirs are produced (Margaret *et al*., 2007).

Success of any intervention being made currently to curb this virus will depend on knowledge on genetic diversity of this virus. Research to improve current treatments include decreasing side effects of current drugs, further simplifying drug regimens to improve adherence, and determining the best sequence of regimens to manage drug

resistance. HIV-2 and HIV-1 group O have shown natural resistance to non nucleoside reverse transcriptase inhibitors (Garrido *et al*., 2008; Tuaillon *et al*., 2004).

The goal of universal coverage of HIV-1 strains by diagnostic tests can be met by minimizing false negative test rates for the six globally prevalent HIV-1 group M strains and HIV-2, and by evaluating systematically the coverage of rare subtypes and recombinant forms (McCutchan, 2006). It is almost certain that new HIV genetic subtypes and CRFs are expected as virus recombination and mutation continue to occur. The current subtypes, CRFs and URFs will also continue to spread to new areas as the global epidemic continues, hence the need for continuous surveillance.

CHAPTER THREE

3.0 MATERIALS AND METHODS

3.1. Study design

This was a cross-sectional, experimental study.

3.2. Study location

The Central province covers the area around the Kenya's capital city (Nairobi), Nyeri to south west of Mt. Kenya (Appendix I). It's the ancestral home of the gikuyu, embu, meru communities and has an estimated population of about 3.7 million people. Samples for the study were collected from the following health facilities in Central province: Thika, Maragua, Kikuyu, Nyeri, Kiambu, Kieni, Kerugoya and Tumutumu. These study sites were selected on the basis of their close proximity to healthcare facilities and ease of transportation to Kenya Medical Research Institute in Nairobi where this work was conducted.

3.3. Study Population

The study population comprised of HIV positive adults attending comprehensive care clinic in the health care facility. With informed consent, blood samples were collected from HIV-seropositive patients above 18 years of age. Blood samples were not collected from adults who had not given consent or minors. No additional demographic data was collected from the patients attending these health facilities.

3.4. Sample size calculation

To get statistically significant sample size, the national prevalence of HIV-1 in Kenya was used. The sample size for this investigation was determined using the formula derived by Fisher *et al.*, 1998 as follows:

$$N=\frac{Z^2P(1-P)}{D^2}$$

Where:

N= Minimum sample size required

Z= 1.96 standard error

P=0.07 (7%) is the current prevalence of HIV in Central Province (KAIS, 2007).

D= 0.05 the inverse of 95% confidence limit (the allowable error)

Therefore:

$$\mathbf{N}=\frac{1.96^2(0.07)(1-0.07)}{(0.05)^2}=\mathbf{100\ samples.}$$

Therefore, the total working sample size was 100.

3.5 Ethical Approval

This study was carried out after getting ethical approval from the KEMRI Scientific Steering Committee and the Ethical Review Committee. The approval no. was SSC/2728. Appendix II gives detailed information on informed consent.

3.6 Laboratory procedures

3.6.1 Whole blood sample collection and processing

About three millilitres of venous blood was drawn by venepuncture from each individual subjects into a 10ml EDTA laced vacutainer tubes. The samples collected were transported in ice packed cool boxes on a daily basis to KEMRI HIV laboratory where they were stored at +4ºC awaiting peripheral blood mononuclear cells (PBMCs) extraction, DNA isolation, amplification and Sequencing.

3.6.2. Extraction of Peripheral blood mononuclear cells (PBMCs)

Peripheral blood mononuclear cells were isolated from whole blood (Boyum, 1968). To a 15ml falcon tube containing 10 ml of 0.84% ammonium chloride , 3mls of whole blood was added and vortexed for complete mixing. This was incubated at 37°C for 10 minutes after which the blood was centrifuged at 1000xg (Beckman®) for 10 minutes at room temperature. The supernatant was discarded. To the resulting pellet, another 10 ml of 0.84% ammonium chloride was added and the above procedure repeated three more times. After a complete wash, the resulting white PBMC pellet was drawn into a 1.5ml eppendorf tube and spun at 1000xg in a microcentrifuge at room temperature. The supernatant was pipetted off and the resulting pellet of PBMCs stored at -20°C for DNA extraction process.

3.6.3. Proviral HIV DNA extraction from PBMCs

To the PBMC pellet, 500µl of DNAzol genomic DNA extraction reagent (Gibco BRL®) was added and dissolved completely by pipetting the reagent-pellet mixture up and down with a pipette. Two volumes (1000µl) of chilled (cooled to 4°C) absolute ethanol was added to the dissolved pellet and mixed gently. This mixture was spun at 12000rpm in a microcentrifuge at 4°C for 15 minutes and supernatant discarded. 1000µl of 70% ethanol was added to the pellet and vortexed thoroughly. The 70% ethanol acted as a wash solution. Spinning was done again at 10000xg in a microcentrifuge at 4°C for 15 minutes and supernatant discarded. The pellet was dried at room temperature in a biosafety cabinet. DNAse/RNase free water was used to dissolve the DNA pellet and stored at -20°C for subsequent processes.

3.6.4. Polymerase Chain Reaction (PCR)

The starting template was proviral DNA extracted from PBMCs. All reagents for this procedure were thawed on ice. Depending on the number of samples to be amplified, a PCR master mix was made containing 2mM MgCl, 0.8mM dNTPs, 0.5 units Taq polymerase, 1x Buffer, 2ng of each primer (Table 2) and the DNA Template. A nested PCR was done using a set of env gp41specific primers for the first PCR and second PCR respectively.

Table 2: List of primers used in the study.

HIV-1Region	Primer Name	Sequence	Region of Amplification	Reference
Env gp 41	gp 41F1: forward 1	TCTTAGGAGCAG CAGGAAGCACTA TGGG	7789-7816	Carr *et al*., 1998
	gp 41 R1:reverse 1	AACGACAAAGGT GAGTATCCCTGC CTAA	8347-8374	Carr *et al.,* 1998
	gp 41 F2: forward 2	ACAATTATTGTC TGGTATAGTGCA ACAGCA	7850-7879	Carr *et al*., 1998
	gp41R1:rev erse 2	TTAAACCTATCA AGCCTCCTTACT ATCATTA	8281-8310	Carr *et al.,* 1998

3.6.5. Analysis of PCR products by Gel Electrophoresis

Once the 2nd PCR was done, all the amplified products were analysed by conventional agarose gel electrophoresis. A 2% agarose gel was prepared using 1x TBE. The PCR products were mixed with loading dye (bromophenol blue) and loaded in to the wells on the gel. A constant voltage of 100volts/cm was applied (Bio-Rad model) for 30 minutes to allow resolution of the PCR products. Staining was done on completion of electrophoresis using ethidium bromide solution (0.5µg/ml) for 15 minutes. A U.V transilluminator was used to visualize the location of amplified DNA products (Sharp *et al*., 1973) and the size estimated by comparing with molecular weight markers loaded alongside.

3.6.6. HIV-1 DNA Sequencing

DNA sequencing PCR was done using a reaction mixture containing 3.0µl of 5x Sequencing buffer, 2.0 µl BigDye® Terminator v 3.1, 1.5 µl Primers (Forward/reverse), 10.5 µl ddH2O and 3.0 µl Sample; making up a total reaction volume of 20 µl. The sequencing PCR was done under the following cycling condition: strand denaturation at 96°c for 5 min followed by 25 cycles of 96°C for 10 sec, 50°C for 5 sec and a final extension at 60°C for 4 min.

3.6.7. DNA Precipitation

The best results are obtained when unincorporated dye terminators are completely removed prior to electrophoresis. Excess dye terminators in sequencing reactions obscure data in the early part of the sequence and can interfere with base calling. To precipitate the tagged DNA amplicons, a cocktail of the following reagents was constituted according to the manufacturer's instructions (Applied BioSystems Inc).

1µl NaOAc (3M pH 5.2)
25 µl 95% Ethanol
} 1x reaction

26 µl of the cocktail was added to the cyclesequenced product. This was covered with adhesive seal, vortexed and spun briefly and incubated for 1hr at 4°C. Centrifugation was done at 3000xg for 30 min at 4°C. The supernatant was dumped on paper towels. The plate was inverted on folded paper towels and spun upside-down at 200xg for 1 min. 150 µl of 70% Ethanol was added and centrifuged at 3000xg for 10min and Ethanol dumped on paper towel. The plate was inverted on folded paper towels and

spun upside down at 200xg for 1 min again. This step was repeated twice. The pellet was resuspended in 20 µl HI-DI formamide. Wells were sealed with adhesive covers and vortexed. The plate was heated in a thermocycler at 95°C for 2 min to denature the amplicons and flash-cooled by immediately placing it on ice. A quick spin was done to get liquid to the bottom and the plate loaded on to ABI 3130xl Genetic Analyzer for sequencing.

3.5.8. HIV-1 DNA Sequence Analysis

Samples from the selected sites were coded as per site using identification numbers and double-entered into a computer. Data was analysed according to source and subtypes characterised using bioinformatics tools. Basecalling was facilitated by SeqScanner v.2.5 (Applied Biosystems). Generated sequences were analyzed by the use of a phylogenetic tree analysis program to determine their subtypes (Grasso *et al*., 2004, Saldanha, 2004).

Sequencing was done in order to determine the order of nucleotide bases in the amplified products. Using the Applied Biosystems Sequencing Cycle kits version 3.1, fluorescently labelled dyes are attached to ACGT extension products in the DNA sequencing reactions. The dyes are labelled as follows: red (labels Thymidine base), blue (Cytosine), black (Guanine) and green (Adenine). The dyes are incorporated using either 5'-dye label primers or 3'-dye label dideoxynucleotide terminators.

An initial BLAST (Basic Local alignment search Tool) was carried out on all the generated sequences to ascertain that the region sequenced was actually the HIV-1 env

glycoprotein (Altschul *et al*., 1990). This is an algorithm for comparing primary biological sequence information, such as the amino acid sequences of different proteins or the nucleotides of DNA sequences. It enables comparison of a query sequence with a library or database of sequences, and identifies library sequences that resemble the query sequence above a certain threshold.

Sequences were aligned with subtype reference sequences from the Los Alamos database by CLUSTAL W (version 1.81). Phylogenetic trees were constructed by the neighbor-joining method and reliability estimated by 1000 bootstrap replications in Molecular Evolutionary Genetic Analysis (MEGA) v3.2 using genetic distances between sequences (Kumar *et al*., 2001). The profile of the tree was visualized with Tree View PPC version 1.6.5 (Institute of Biomedical and Life Sciences, Scotland, UK).

CHAPTER FOUR

4.0 RESULTS

4.1 HIV-1 Proviral DNA Amplification

Results of the PCR were visualized by gel electrophoresis. **Figure 4.1** show positive bands for HIV-1 *env gp 41* amplicons.

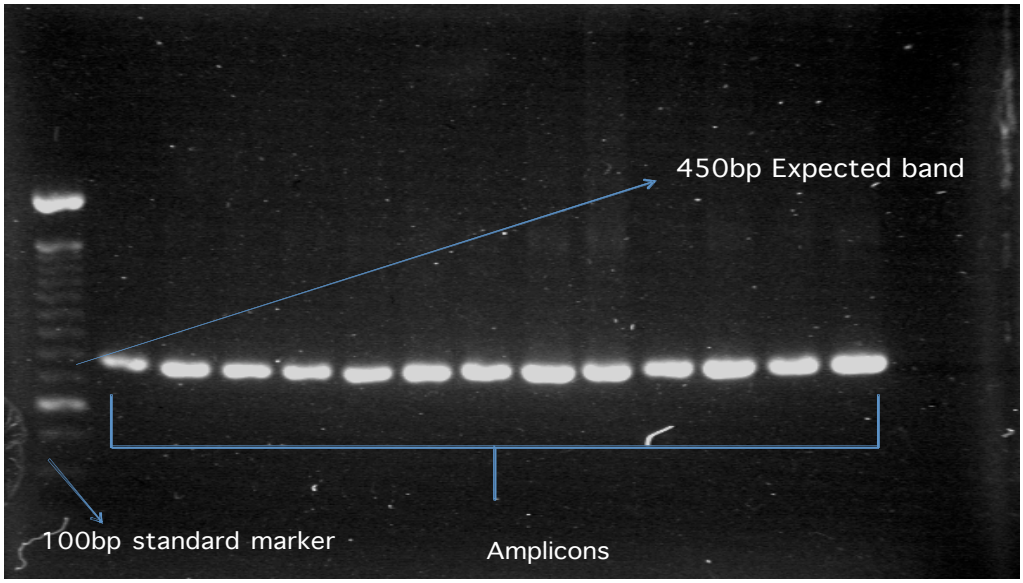

Figure. 4. 1: PCR gel picture showing successfully amplified products of 450bp

This gel picture represents samples from Nyeri (NPGH 001-0012) that amplified successfully by PCR. However some of the samples did not amplify at all while some amplified but did not have the expected fragment size (Gel picture not shown). Only the samples that successfully amplified proceeded to automated sequencing.

4.2 General Picture of HIV-1 subtype diversity in Central Kenya

A total of 96 HIV-1 positive samples, twelve from each study site, were investigated for subtype designation and phylogenetic relationships. The analysis demonstrated the predominance of HIV subtype A, 65 (68%), C 10 (10%), D 11 (12%) and recombinants which accounted for 9(9%). One percent was unknown (**Table 3**, **Fig 4.2** and **Fig. 4.3** shows the summary of results**).**

Table 3: Overall HIV-1 Subtypes in Central Province

Health Facility/HIV-1 Subtype	**A**	**C**	**D**	**AD**	**AG**	**CD**	**Unknown**	**Total**
NPGH	10	1	0	0	1	0	0	12
MDH	7	2	1	1	1	0	0	12
TDH	7	1	2	2	0	0	0	12
KDH	6	2	1	3	0	0	0	12
KGDH	9	1	2	0	0	0	0	12
KMH	11	0	1	0	0	0	0	12
TMH	8	2	2	0	0	0	0	12
KNMH	7	1	2	0	0	1	1	12
Total	**65**	**10**	**11**	**6**	**2**	**1**	**1**	**96**

KEY:

NPGH: Nyeri Provincial General Hospital

MDH: Maragua District Hospital

TDH: Thika District Hospital

KDH: Kiambu District Hospital

KGDH: Kerugoya District Hospital

KMH: Kikuyu Mission Hospital

TMH: Tumutumu Mission Hospital

KNMH: Kieni Mission Hospital

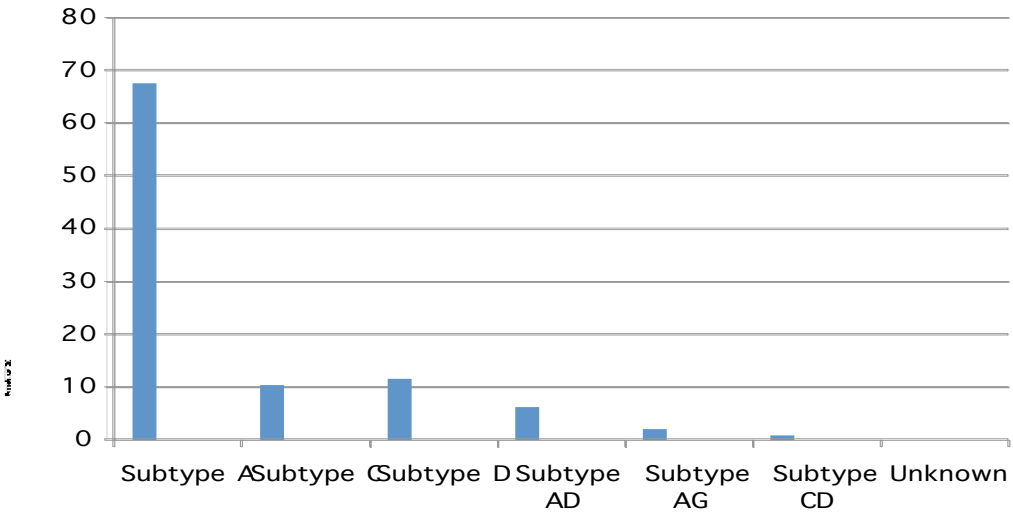

Figure 4.2: HIV-1 Subtype Prevalence in Central Province of Kenya.

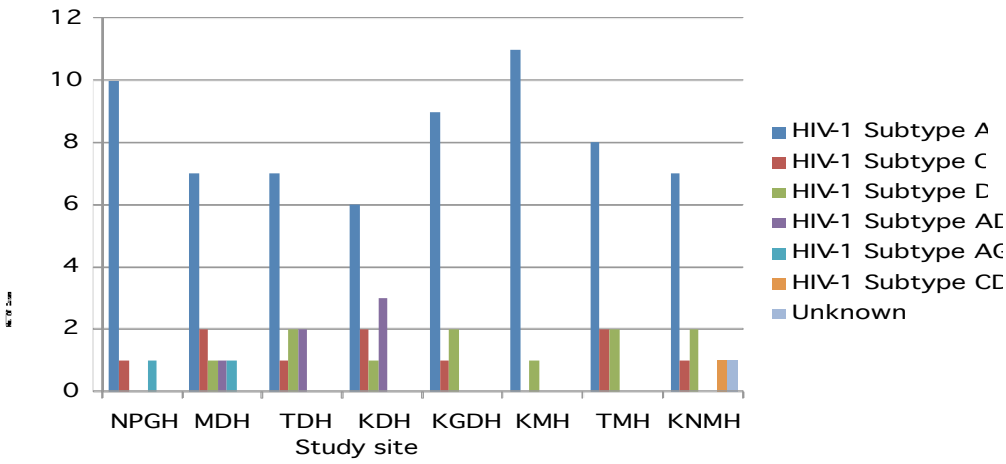

Figure 4. 3: Summary of HIV-1 subtype diversity per study site.

Key:

NPGH: Nyeri Provincial General Hospital

MDH: Maragua District Hospital

TDH: Thika District Hospital

KDH: Kiambu District Hospital

KGDH: Kerugoya District Hospital

KMH: Kikuyu Mission Hospital

TMH: Tumutumu Mission Hospital

KNMH: Kieni Mission Hospital

Maragua and Kieni had the highest HIV-1 Subtype diversity followed by Nyeri, Kiambu, Kerugoya, Tumutumu and Thika while Kikuyu and had the least divergence. Pure HIV-1 subtypes accounted for 90% of the results while 10% were circulating recombinants (**Figure 4. 4**)

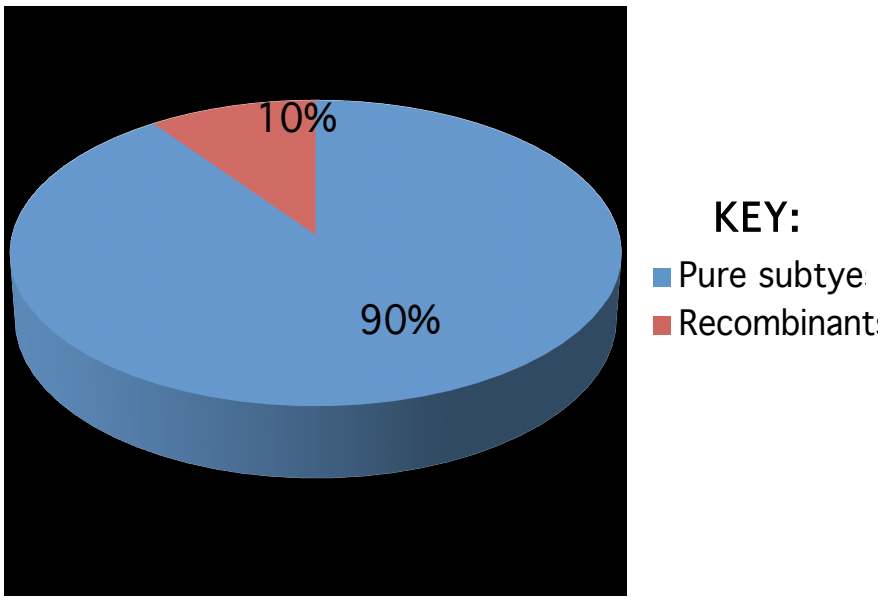

Figure 4.4: General Picture of HIV-1 Inter-subtype Recombination

4.3 Phylogenetic Relationship Analysis and Distribution of HIV-1 Subtypes by study Site

4.3.1 Analysis HIV-1 samples collected from Nyeri

A total of 12 HIV-1 samples were successfully sequenced from the region and the generated sequences analysed (Refer to **appendix III** for sequences generated from HIV-1 positive samples from this site). BLAST results indicated that 10 samples belonged to HIV subtype A, 1 sample was subtype C and the remaining sample was a recombinant between the main subtype A and G. Phylogenetic tree analysis showed that subtype A sequences clustered with reference sequences from Kenya, Rwanda, Tanzania and Belgium. HIV subtype C sequences clustered with subtype C references

from Kenya and Zambia while the recombinant sequence clustered with a reference sequence from Italy.

The results are as illustrated in table 4 and Figure 4.5

Table 4: Circulating HIV-1 Subtypes in Nyeri.

Sample code	**Length of generated Sequence (bp)**	**HIV-1 Subtype**
NPGH_001	460	A1
NPGH_002	454	AG
NPGH_003	437	A1
NPGH_004	455	A1
NPGH_005	459	A
NPGH_006	456	A1
NPGH_007	384	A
NPGH_008	403	A1
NPGH_009	394	C
NPGH_010	465	A1
NPGH_011	441	A1
NPGH_012	464	A1

Key:

NPGH: Nyeri Provincial General Hospital

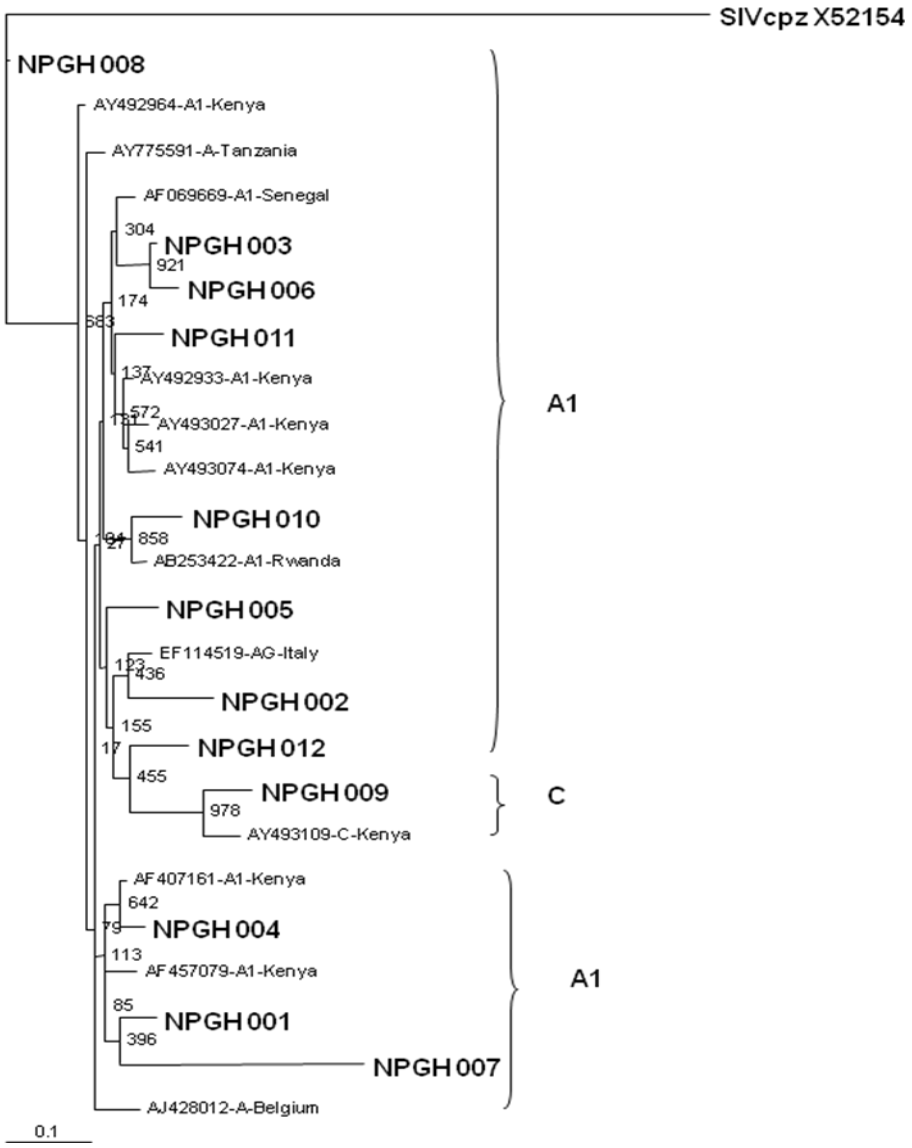

Figure 4.5: Phylogenetic analysis of sequences from HIV-1 samples from Nyeri.

HIV-1 Subtype A sequences clustered with reference sequences from Kenya, Belgium, Rwanda, Senegal. Recombinant AG clustered with a reference sequence from Italy.

4.3.2 Analysis HIV-1 samples collected from Maragua

Phylogenetic analysis of 12 sequences from Maragua showed that 7 (58%) belonged to HIV subtype A, 2(17%) belonged to subtype C, 1(8%) subtype D, 1(8%) recombinant AG and the remaining belonged to recombinant AD. This region showed the highest diversity in terms of pure subtypes and recombinant viruses. The sequences generated (Refer to appendix III for sequences generated from HIV-1 positive samples from this site) clustered with previous reference sequences generated locally and reference sequences from South Africa, Cameroon, Nigeria, Uganda and U.S.A. (**Table 5 and Figure 4.6**).

Table 5: Circulating HIV-1 Subtypes in Maragua.

Sample code	Length of generated Sequence (bp)	HIV-1 Subtype
MDH_001	407	D
MDH_002	432	A1
MDH_003	421	A1
MDH_004	461	A1
MDH_005	429	C
MDH_006	461	AD
MDH_007	433	A1
MDH_008	436	A1
MDH_009	453	AG
MDH_010	295	C
MDH_011	173	A1
MDH_012	241	A1

Key:

MDH: Maragua District Hospital

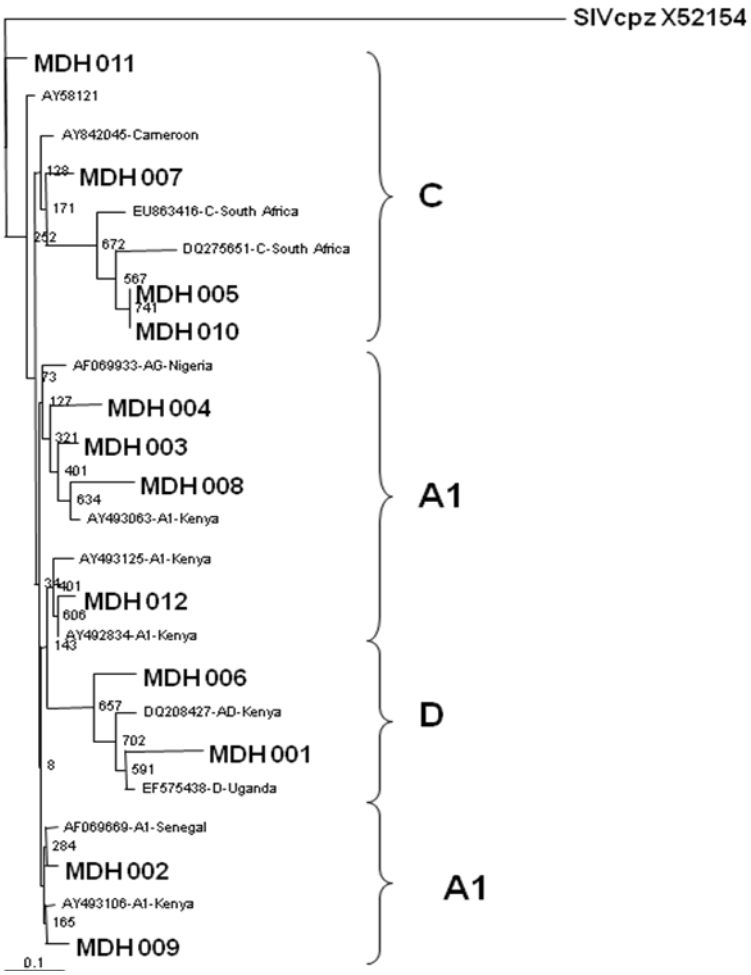

Figure 4.6: Phylogenetic analysis of sequences from HIV-1 samples from Maragua. The generated subtype A sequences clustered with reference sequences from Senegal, Nigeria, Kenya. Subtype C sequences clustered with reference sequences from Cameroon and South Africa.

4.3.3. Analysis HIV-1 samples collected from Thika

Twelve sequences (**Appendix III**) were phylogenetically analysed from this site. 7(58%) were HIV subtype A, 2(17%) were subtype D and AD respectively while 1 (8%) were HIV subtype C (**See table 6**). Subtype A sequences clustered with reference sequences from Kenya and Uganda. Subtype C and D sequences clustered with sequences from Kenya and Uganda while subtype AD clustered with reference sequences both from Kenya and Tanzania (**Figure.4.7**).

Table 6: Circulating HIV-1 Subtypes in Thika.

Sample code	**Length of generated Sequence (bp)**	**HIV-1 Subtype**
TDH_001	321	A
TDH_002	400	A1
TDH_003	390	A1
TDH_004	450	A1
TDH_005	436	D
TDH_006	445	AD
TDH_007	353	D
TDH_008	301	C
TDH_009	340	A1
TDH_010	244	A1
TDH_011	378	A1
TDH_012	354	A1D

Key:

TDH: Thika District Hospital

Figure 4.7: Phylogenetic analysis of sequences from HIV-1 samples from Thika. Generated sequences clustered with reference sequences from Kenya and Uganda.

4.3.4. Analysis HIV-1 samples collected from Kiambu

Twelve HIV-1 sequences (**Appendix III**) were analysed from this site. The phylogenetic analysis showed that 6 (50%) were subtype A, 3(25%) were subtype AD, 2 (17%) were subtype C and subtype D was only one sample which accounted for 8% (**See table 7**). Phylogenetic tree analysis with reference sequences portrayed HIV Subtype A, D and AD clustering with env sequences from Kenya, Subtype C sequences with env sequences from Senegal and Brazil (**Figure 4.8**).

Table 7: Circulating HIV-1 Subtypes in Kiambu.

Sample code	Length of generated Sequence (bp)	HIV-1 Subtype
KDH_001	471	D
KDH_002	441	A1
KDH_003	467	C
KDH_004	433	AD
KDH_005	340	AD
KDH_006	220	A1
KDH_007	331	A1
KDH_008	450	AD
KDH_009	440	A1
KDH_010	461	A1
KDH_011	402	C
KDH_012	191	A1

Key:

KDH: Kiambu District Hospital

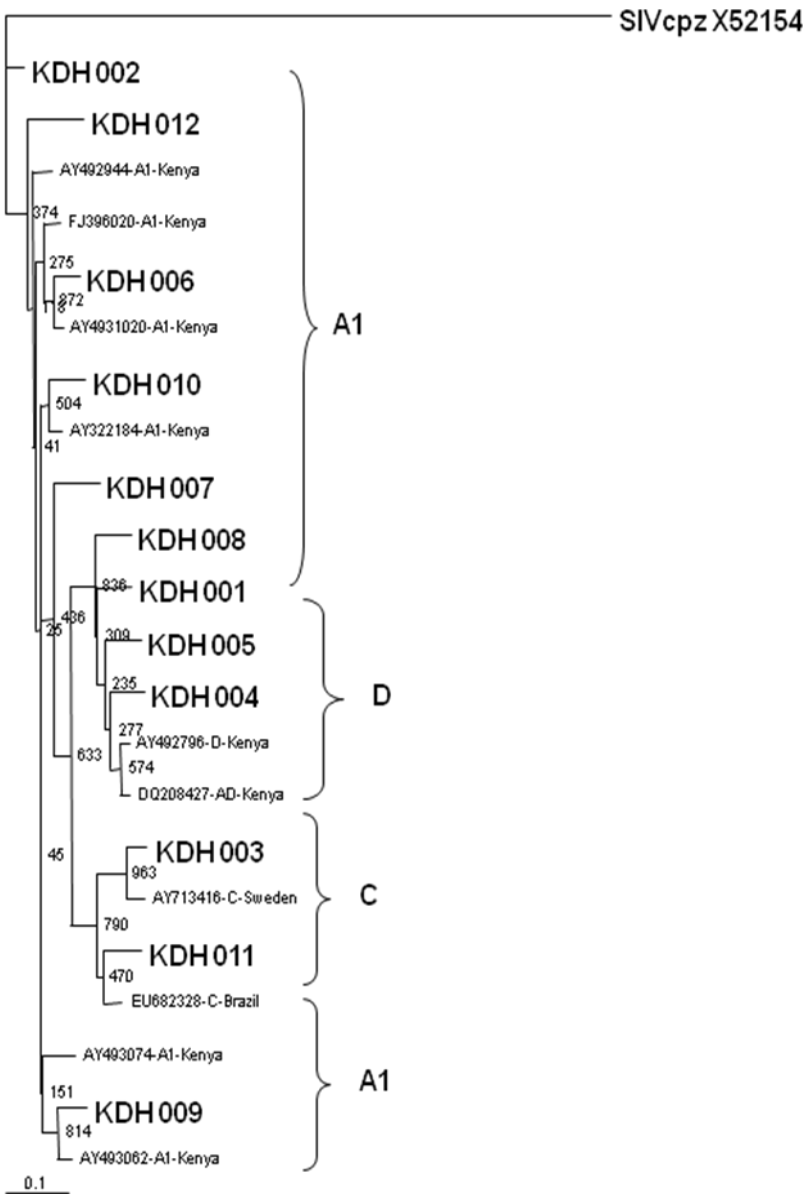

Figure 4.8: Phylogenetic analysis of sequences from HIV-1 samples from Kiambu. HIV-1 subtype A and D clustered with reference sequences from Kenya, subtype C with reference sequences from Sweden and Brazil.

4.3.5. Analysis HIV-1 samples collected from Kerugoya

From this site, 12 sequences (Appendix II) were analysed. HIV Subtype A predominated with 9 (75%) out of the twelve. Subtype D was represented by 2 (17%) whereas 1 (8%) was HIV Subtype C (**table 8**). This region portrayed pure subtypes with no intersubtype recombination. The A Subtypes corresponded to reference sequences from Kenya, Saudi Arabia and Uganda. HIV Subtype D clustered with a reference sequence from South Africa and Subtype C clustered with those from Botswana (**See figure.4.9**).

Table 8: Circulating HIV-1 Subtypes in Kerugoya.

Sample code	Length of generated Sequence (bp)	HIV-1 Subtype
KGDH_001	381	D
KGDH_002	333	A1
KGDH_003	324	D
KGDH_004	372	A
KGDH_005	272	A1
KGDH_006	438	A
KGDH_007	445	A1
KGDH_008	439	A1
KGDH_009	429	A1
KGDH_010	228	A1
KGDH_011	224	A1
KGDH_012	317	C

Key:

KGDH: Kerugoya District Hospital

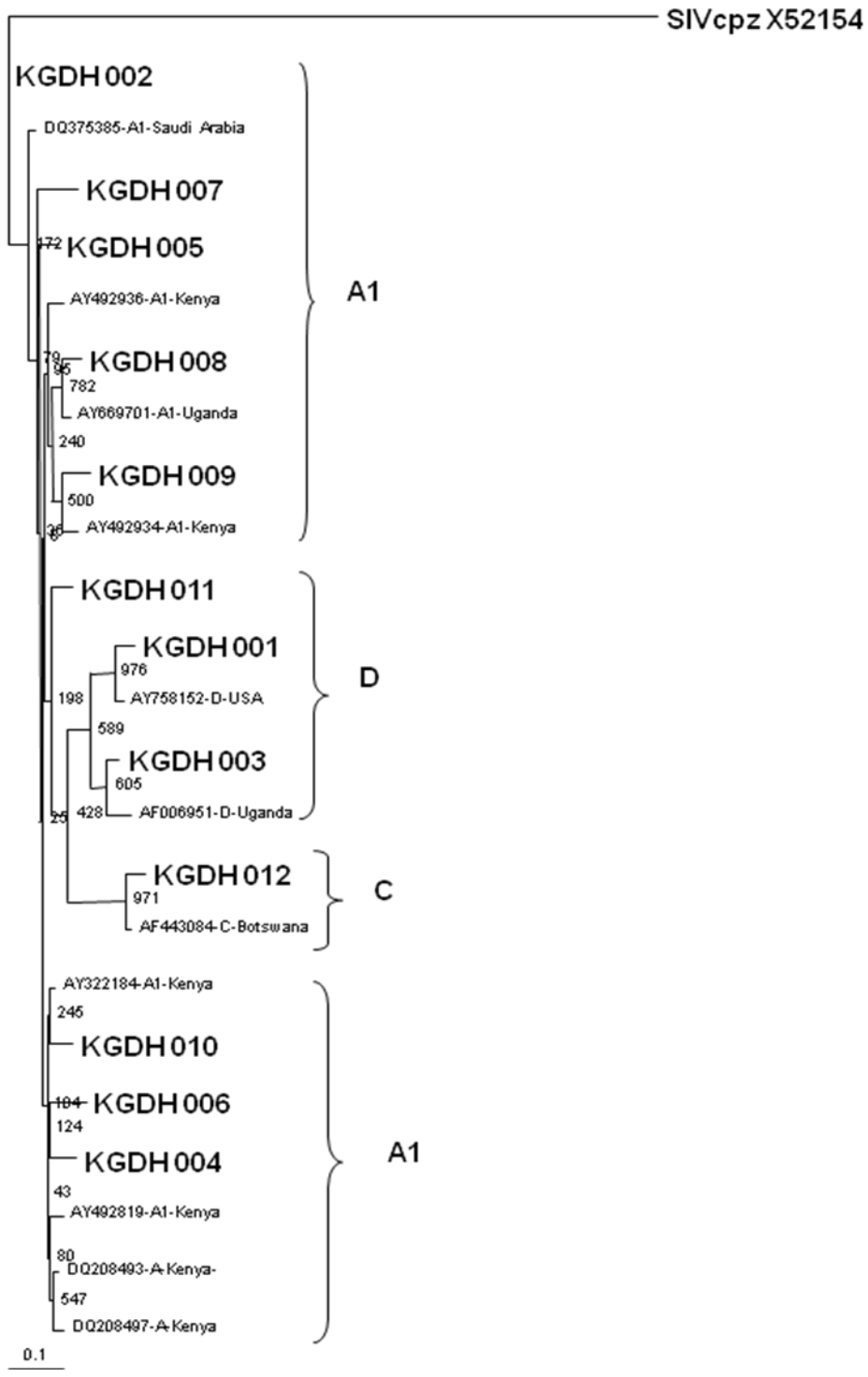

Figure 4.9: Phylogenetic analysis of sequences from HIV-1 samples from Kerugoya. HIV-1 Subtype A generated sequences clustered with reference sequences from Saudi

Arabia, Kenya and Uganda. Subtype C clustered with reference sequence Botswana whereas subtype D clustered with reference sequences from U.S.A and Uganda.

4.3.6. Analysis HIV-1 samples collected from Kikuyu

From 12 sequences analysed, HIV-1 subtype A were 11(92%) and D 1(8%). HIV subtype A sequences clustered with reference sequences from Kenya, Uganda and Tanzania while HIV subtype D clustered with a reference sequence from Uganda (**Table 9 and Figure. 4.10**).

Table 9: Circulating HIV-1 Subtypes in Kikuyu.

Sample code	Length of generated Sequence (bp)	HIV-1 Subtype
KMH_001	455	A1
KMH_002	438	A1
KMH_003	455	A
KMH_004	332	D
KMH_005	377	A
KMH_006	425	A1
KMH_007	431	A1
KMH_008	432	A1
KMH_009	457	A1
KMH_010	341	A1
KMH_011	316	A1
KMH_012	340	A1

Key:

KMH: Kikuyu Mission Hospital

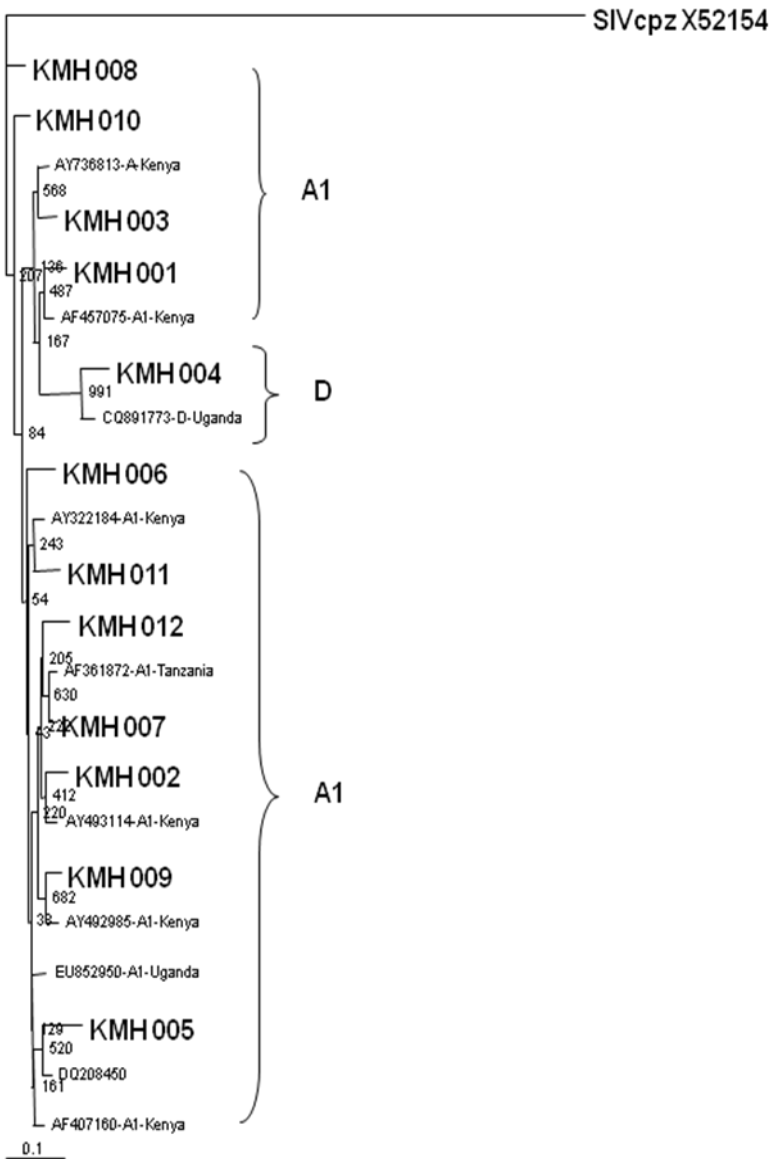

Figure 4.10: Phylogenetic analysis of sequences from HIV-1 samples from Kikuyu. The generated subtype A sequences clustered with reference sequences from Kenya, Tanzania and Uganda while subtype D with reference sequences from Uganda.

4.3.7. Analysis HIV-1 samples collected from Tumutumu

From the 12 sequences (Appendix III) analysed from this site, 8 (67%) were HIV Subtype A while 2 (17%) were subtype C and D respectively (**Table 10**). HIV Subtype A and D sequences clustered with reference sequences from Kenya, Tanzania and Uganda whereas HIV subtype C clustered with those from Kenya and Zambia (see phylogenetic tree **Figure. 4.11**).

Table 10: Circulating HIV-1 Subtypes in Tumutumu.

Sample code	Length of generated Sequence (bp)	HIV-1 Subtype
TMH_001	435	D
TMH_002	469	A1
TMH_003	425	C
TMH_004	458	A
TMH_005	474	A1
TMH_006	459	A1
TMH_007	442	C
TMH_008	451	A
TMH_009	445	A1
TMH_010	452	D
TMH_011	428	A1
TMH_012	455	A1

Key:

TMH: Tumutumu Mission Hospital

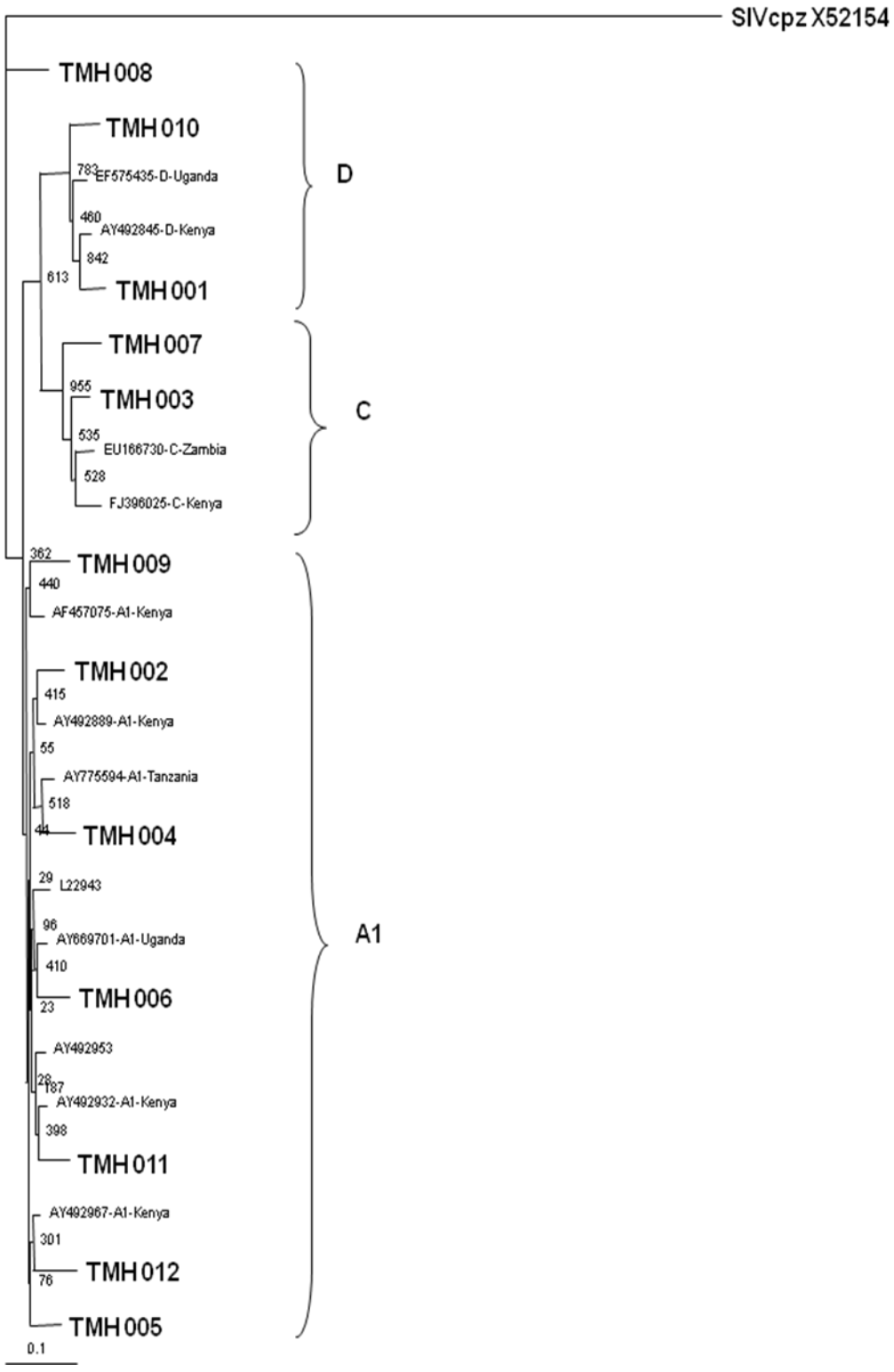

Figure 4.11: Phylogenetic analysis of sequences from HIV-1 samples from Tumutumu. Subtype A sequences clustered with reference sequences from Kenya Tanzania and

Uganda, subtype D with reference sequences from Kenya and Uganda while C with reference sequences from Kenya and Zambia.

4.3.8. Analysis HIV-1 samples collected from Kieni

A total of twelve sequences (Appendix III) were analysed from this site and from the analysis, 7 (58%) belonged to HIV subtype A, 2(17%) were subtype D while subtypes C and CD were represented by 1 (8%) each. These sequences clustered with reference sequences from Kenya and Uganda. It was difficult to establish the subtype of one sample, though, from phylogenetic tree analysis, it clustered with a genetic sequence from Equatorial Guinea (**see table 11 and Figure.4.12**).

Table 11: Circulating HIV-1 Subtypes in Kieni.

Sample code	Length of generated Sequence (bp)	HIV-1 Subtype
KNMH_001	434	D
KNMH_002	431	D
KNMH_003	433	A
KNMH_004	434	-
KNMH_005	438	A1
KNMH_006	463	A1
KNMH_007	455	A
KNMH_008	435	A1
KNMH_009	459	A1
KNMH_010	448	A1
KNMH_011	442	CD
KNMH_012	420	C

Key:

KNMH: Kieni Mission Hospital

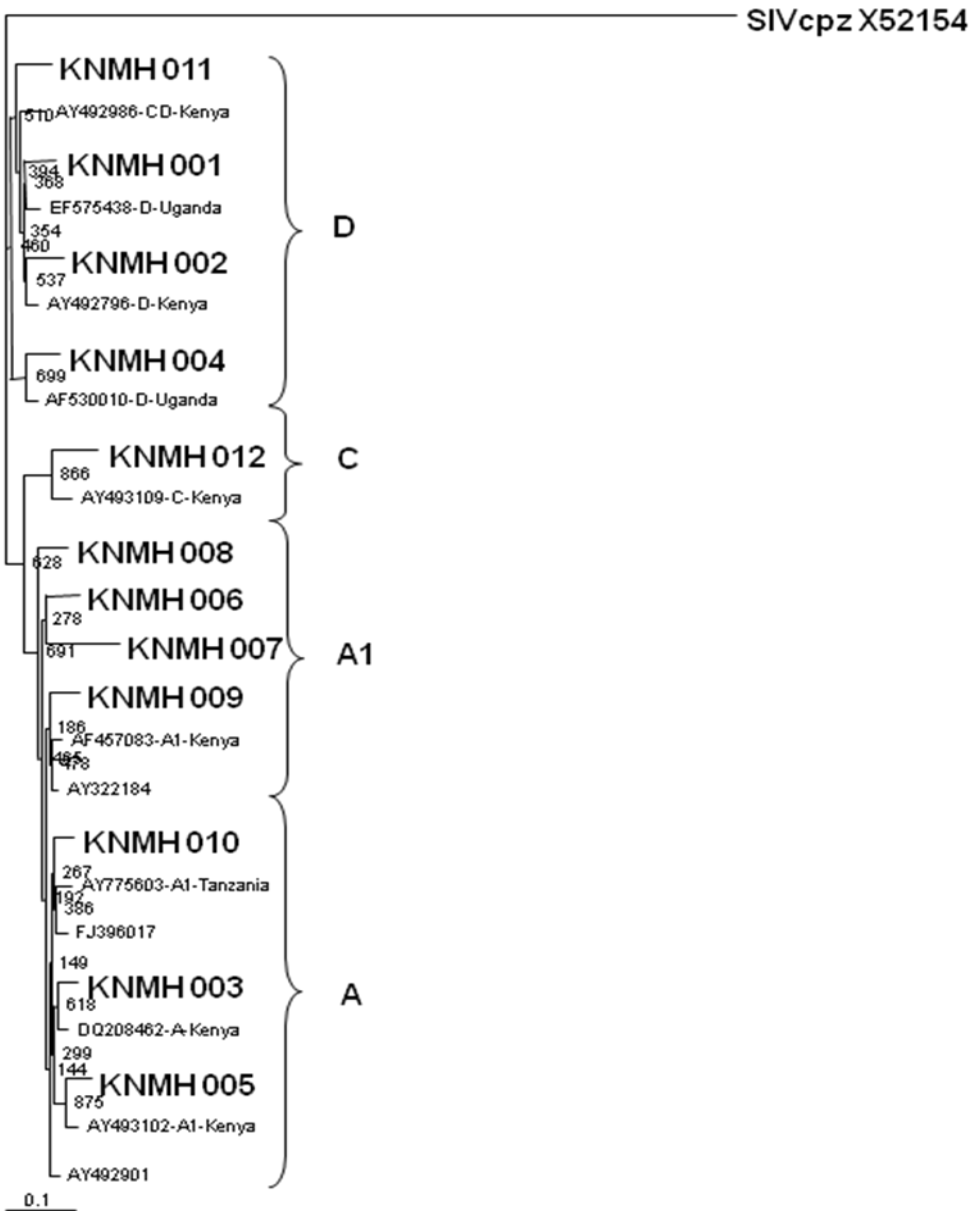

Figure 4.12: Phylogenetic analysis of sequences from HIV-1 samples from Kieni. HIV-1 subtype A C and D sequences clustered with reference sequences from Kenya, Tanzania and Uganda.

CHAPTER FIVE

5. DISCUSSION

HIV subtype surveillance studies have been carried out in most parts of the country including Western (Songok *et al.*, 2003), Southern Kenya (Dowling *et al*., 2002), Northern Kenya (Khamadi *et al*., 2005) and Nairobi(Land *et al*., 2008 Lihana *et al*.,2006 and Neilson *et al*., 1999). HIV-1 diversity surveillance studies have not been conducted in Central Province region over the years thus; this study provides information on HIV-1 subtypes circulating in this province.

BLAST analysis showed that 68% of the sequences analysed were HIV subtype A, 11% HIV-1 subtype D, 10% HIV-1 subtype C and 9% HIV-1 circulating recombinants. One of the sequences generated from Kieni (KNMH 004) was unknown from the BLAST result, However, this sequence clustered with HIV env glycoprotein reference sequence from Equatorial Guinea which had not been designated a subtype. This results concurs with results done in Western province (Songok *et al*., 2003) where two samples were unclassified. Most common recombinant subtype was AD with a proportion of 6% followed by AG with 2% and CD with a percentage of one.

These findings are in tandem with related surveillance studies conducted in Nyanza, Coast, Nairobi and Rift Valley provinces except for Northern Kenya which had the highest percentage of HIV-1 subtype C, 27%. This was attributed to the fact that the northern Kenya borders Ethiopia whose predominant subtype is C and cross-border

movements could have had an enormous impact on the circulating HIV-1 subtypes in the region amongst other reasons (Khamadi *et a*l., 2005). The findings show that HIV-1 subtype A is still by far the most dominant subtype circulating in Kenya. However the study also shows that there is an increase in the circulating recombinant forms. This may be indicative of the fact that the picture of HIV-1 genotypes in Kenya is subject to change with time (Refer to Table 12 for summary findings).

Table 12: Summary of findings from other studies done in Kenya.

Study	HIV-1 Subtypes (%)					Study site
	A	**D**	**C**	**G**	**CRFs**	
Kageha *et al*.,2009	68	12	10	0	9	Central
Khamadi *et al*.,2009	86	8	5	1		Coast
Land *et al*., 2008	70	9	6	0	15	Nairobi
Lihana *et al*.,2006	64	17	9	1	9	Nairobi
Khamadi *et al*.,2005	57	9	27	0	7	Northern
Songok *et al*.,2003	66.7	6.7	3	3	19.6	Western
Dowling *et al*., 2002	56	2	2	0	40	Southern
Neilson *et al*.,1999	70.3	20	6.9	0.3	2.2	Nairobi

Among the eight sites from where the samples were collected, it was observed that three of them: Kikuyu, Kerugoya and Tumutumu portrayed clear-cut HIV-1 subtypes with no circulating recombinants. This could be an indication that the development of subtype recombination in this area is in its initial stages. Kikuyu showed the least diversity despite its being near Nairobi with quite diverse HIV-1 subtypes. Two possibilities can explain this observation: i) limited cross infections with diversified subtypes and ii) Sampling limitations. This calls for more open and expansive studies to rule out the possibilities of sampling from the same households.

The results showed phylogenetic relatedness based on the geographic origins. Most of the sequences analysed clustered genetically with the subtype A strains common in other regions such as Kenya, Tanzania, Uganda and Rwanda. Only a few (1%) clustered with reference sequences from Saudi Arabia, Senegal, Rwanda, Belgium and U.S.A. Subtype D sequences generated clustered with reference sequences from Kenya and Uganda. HIV-1 subtype C sequences clustered with reference sequences from South Africa, Senegal, Brazil, Botswana, Zambia and Sweden. In this case, the root was difficult to trace because these geographic regions are vast and diversified therefore one can only conclude that the sequences are from a more divergent origin.

Recombination is an important evolutionary strategy geared towards the survival of an organism in an otherwise non-conducive environment. Homologous recombination occurs when a cell is co-infected with two different but related strains (Burke, 1997). HIV evolves by rapid mutations and recombination; both processes actively contribute to its genetic diversity. Rapid development of genetic diversity through recombination coupled with mutations may contribute to the HIV pathogenicity as it may allow escape from immune surveillance and/or from suppression of virus replication (Alexander *et al*., 2000; Arendrup *et al*., 1992; Arien *et al*., 2005a; Arien *et al*., 2005b; Bailey *et al*., 2006; Blackard *et al.*, 1999; Blackbourn *et al*., 1996).

In the present study, most recombinants were between the main HIV-1 subtype A and either D or G subtypes. Only 1% was between HIV-1 subtype C and D. The AD recombinants clustered with those from Kenya and Tanzania, AG with reference sequences from Nigeria and Italy while recombinant CD clustered with a reference

sequence from Kenya based on the *env gp41* region. Recombinant CD though rare was also observed in the studies carried out in southern Kenya (Dowling *et al*., 2002).

HIV subtypes have differing effects on transmission, development of resistance (Garrido *et al*., 2008; Tuaillon *et al*., 2004) and rates of disease progression. From research studies carried out in Nairobi, infection with subtype C was associated with more rapid disease progression exhibited by high viral loads and a decrease in CD4 T cells (Neilson *et al*., 1999). Other studies carried out in Rakai, Uganda showed that disease progression in patients infected with subtype D was more rapid than those infected with subtype A (Kiwanuka *et al*., 2008).

Transmission rates have also been shown to vary across the HIV-1 subtypes with clade A showing the highest transmissibility as compared to the rest. HIV strains capable of using chemokine coreceptor CCR5 are more frequently transmitted than strains that use the CXCR4 coreceptors. Some strains like D are dual tropic and thus more easily transmitted and are associated with rapid disease progression (Chan and Kim, 1998).

Some samples did not amplify during PCR while others amplified but did not show the expected size of the amplicon. A similar experience was encountered in a study conducted in the USA. The study reported the isolation of a new strain of HIV-2, HIV2-NWK08F, from an immunodeficient Sierra Leone immigrant (Smith *et al*., 2008). The patient's serum was repeatedly reactive by serological testing with ELISA kits containing HIV-1 and HIV-2 antigens but negative on western blot for HIV-1. His HIV-1 viral load was <48 copies and polymerase chain reaction (PCR) for HIV-1 proviral

DNA was negative. An HIV-2 immunoblot was positive. The presumptive diagnosis was that this patient had an HIV-2 infection. However, a PCR assay from a commercial laboratory for HIV-2 proviral DNA was negative.

From the study, two suggestions were given for failure: i) the proviral load could have been below the limit of detection of the assay or ii) the virus was too divergent from known HIV-1 epidemic groups to be amplified by the *gp 41* primers based on epidemic subtype consensus sequence. These possibilities could explain the PCR failure and therefore requiring further investigation.

Biological consequences of HIV genetic variation and its ramifications with regard to epidemiology, diagnostics, classification and vaccine design (McMichael *et al*., 2002) cannot be underscored. Classification based on genetically defined subtypes provides an important framework for making advances on understanding viral biology and immunology, and for vaccine development (Butler *et al*., 2007). HIV as a pathogen mutates extensively, presenting significant challenges to effective monitoring and disease control. This underscores the need for effective surveillance to track HIV variants and to direct research and prevention activities (Hu *et al.*, 1996).

5.2 Conclusion

From this investigation, a total of 96 sequences from the HIV-1 *env gp41* region were analysed and phylogenetic relationships determined. The findings demonstrated the predominance of HIV subtype A (68%), 12% were HIV Subtype D, and 10% HIV Subtype C**.** It was apparent that there exists some level of discordance especially from studies carried out in Northern Kenya. HIV Subtype A was 50%, 36% were subtype C and 14% were HIV-1 Subtype D. However, these findings are in congruence with results from previous studies carried out in Western, Nairobi and Nyanza, which showed that HIV-1 Subtype A was between 67-93%, subtype C and subtype D were between 7-15%.

This study was able to show levels of inter-subtype recombination (10%). Most of the observed recombinations were between the major HIV-1 Subtype A with D, 6%, A with G, 2% and C with D, accounting for 1%. 90% were pure HIV-1 subtypes. These preliminary findings demonstrate diversity in HIV-1 subtypes in Central Province.

5.3 Study limitations

This being a pilot study, only a hundred samples were analysed. Therefore the results do not adequately address the HIV subtype diversity in Central province of Kenya. A bigger surveillance study with a much larger sample is needed in order to make generalised deductions.

The samples that failed during amplification required more analysis but due to financial limitations this was not achieved. However, this study serves as a baseline for further studies.

5.4 Future prospects

Currently, candidate HIV vaccines are derived from isolates with hope that they will be sufficiently cross-reactive to protect against circulating strains of the virus (Gaschen *et al*., 2002; Gotch *et al*., 2000; Johnston and Fauci, 2007; Lau, Velasco, and Johnston, 2007). The importance of exploring innovative HIV vaccine strategies is crucial and therefore a deeper understanding of the implications of HIV variations for both antibody and cellular responses is necessary.

Knowledge on HIV-1 genetic diversity shall be applied in future research in developing more sensitive and specific diagnostic kits taking into consideration the HIV-1 antigen recombinants and various HIV-1 clades in circulation.

HIV management involves use of ARVs whose mechanism of action involves suppression of HIV replication. This has however been hindered by the development of drug resistant mutations. The transmission of drug-resistant strains increases despite of all prevention efforts, raising major public health concerns (Wensing and Boucher, 2003). These studies can help in identifying such mutations before treatment is initiated to avoid drug failure and psychological trauma due to adverse side effects.

Phylogenetic analysis will also help in determining possible HIV origins, transmission patterns and future course of AIDS (Clayton, 1996).

5.5 Recommendations

i) More research needs to be carried out terms of expanding the sites and number of samples analysed in order to determine the true picture of HIV genetic diversity in Central province.

ii) Precise mapping of recombination should be carried out using full-length genome sequencing.

iii) Information generated from this study should be used to design more sensitive test kits and candidate vaccines.

iv) Samples that did not amplify successfully by PCR should be analysed further using more sensitive primers to determine whether they are diverged strains of HIV-1 or other viruses.

REFERENCES

Achkar, J. M., Burda, S. T., Konings, F. A., Urbanski, M. M., Williams, C. A., Seifen, D., Kahirimbanyi, M. N., Vogler, M., Parta, M., Lupatkin, H. C., Zolla-Pazner, S., and Nyambi, P. N. (2004). Infection with HIV type 1 group M non-B subtypes in individuals living in New York City. J Acquir Immune Defic Syndr 36(3), 835-44.

Alexander, L., Weiskopf, E., Greenough, T. C., Gaddis, N. C., Auerbach, M. R., Malim, M. H., O'Brien, S. J., Walker, B. D., Sullivan, J. L., and Desrosiers, R. C. (2000). Unusual polymorphisms in human immunodeficiency virus type 1 associated with nonprogressive infection. J Virol 74(9), 4361-76.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol 215(3), 403-10.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17), 3389-402.

Arendrup, M., Nielsen, C., Hansen, J. E., Pedersen, C., Mathiesen, L., and Nielsen, J. O. (1992). Autologous HIV-1 neutralizing antibodies: emergence of neutralization-resistant escape virus and subsequent development of escape virus neutralizing antibodies. J Acquir Immune Defic Syndr 5(3), 303-7.

Arien, K. K., Abraha, A., Quinones-Mateu, M. E., Kestens, L., Vanham, G., and Arts, E. J. (2005a). The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. J Virol 79(14), 8979-90.

Arien, K. K., Troyer, R. M., Gali, Y., Colebunders, R. L., Arts, E. J., and Vanham, G. (2005b). Replicative fitness of historical and recent HIV-1 isolates suggests HIV-1 attenuation over time. Aids 19(15), 1555-64.

Arien, K. K., Vanham, G., and Arts, E. J. (2007). Is HIV-1 evolving to a less virulent form in humans? *Nat Rev Microbiol* **5**(2)**,** 141-51.

Ascher, M. S., and Sheppard, H. W. (1988). AIDS as immune system activation: a model for pathogenesis. Clin Exp Immunol 73(2), 165-7.

Baier, M., Dittmar, M. T., Cichutek, K., and Kurth, R. (1991). Development of vivo of genetic variability of simian immunodeficiency virus. Proc Natl Acad Sci U S A 88(18), 8126-30.

Bailey, J. R., Williams, T. M., Siliciano, R. F., and Blankson, J. N. (2006). Maintenance of viral suppression in HIV-1-infected HLA-B*57+ elite suppressors despite CTL escape mutations. J Exp Med 203(5), 1357-69.

Blackard, J. T., Renjifo, B. R., Mwakagile, D., Montano, M. A., Fawzi, W. W., and Essex, M. (1999). Transmission of human immunodeficiency type 1 viruses with intersubtype recombinant long terminal repeat sequences. Virology 254(2), 220-5.

Blackbourn, D. J., Mackewicz, C. E., Barker, E., Hunt, T. K., Herndier, B., Haase, A. T., and Levy, J. A. (1996). Suppression of HIV replication by lymphoid tissue CD8+ cells correlates with the clinical state of HIV-infected individuals. Proc Natl Acad Sci U S A 93(23), 13125-30.

Blower, S. M., Aschenbach, A. N., and Kahn, J. O. (2003). Predicting the transmission of drug-resistant HIV: comparing theory with data. *Lancet Infect Dis* **3**(1)**,** 10-1.

Bobkov, A. F., Kazennova, E. V., Selimova, L. M., Khanina, T. A., Ryabov, G. S., Bobkova, M. R., Sukhanova, A. L., Kravchenko, A. V., Ladnaya, N. N., Weber, J. N., and Pokrovsky, V. V. (2004a). Temporal trends in the HIV-1 epidemic in Russia: predominance of subtype A. J Med Virol 74(2), 191-6.

Bobkov, A. F., Kazennova, E. V., Sukhanova, A. L., Bobkova, M. R., Pokrovsky, V. V., Zeman, V. V., Kovtunenko, N. G., and Erasilova, I. B. (2004b). An HIV type 1 subtype A outbreak among injecting drug users in Kazakhstan. AIDS Res Hum Retroviruses 20(10), 1134-6.

Boyum, A. (1968). Isolation of mononuclear cells and granulocytes from human blood. Isolation of monuclear cells by one centrifugation, and of granulocytes by combining centrifugation and sedimentation at 1 g. Scand J Clin Lab Invest Suppl 97, 77-89.

Burke, D. S. (1997). Recombination in HIV: an important viral evolutionary strategy. Emerg Infect Dis 3(3), 253-9.

Butler, I. F., Pandrea, I., Marx, P. A., and Apetrei, C. (2007). HIV genetic diversity: biological and public health consequences. Curr HIV Res 5(1), 23-45.

Carpenter, C. C., Fischl, M. A., Hammer, S. M., Hirsch, M. S., Jacobsen, D. M., Katzenstein, D. A., Montaner, J. S., Richman, D. D., Saag, M. S., Schooley, R. T., Thompson, M. A., Vella, S., Yeni, P. G., and Volberding, P. A. (1998). Antiretroviral therapy for HIV infection in 1998: updated recommendations of the International AIDS Society-USA Panel. Jama 280(1), 78-86.

Carr J.K., Foley B.T.,Leitner T., Salminen M, Korber B, McCutchan F. (1998). Reference sequences representing the principal genetic diversity of HIV-1 in the pandemic. Human Retrovirus and AIDS, Pp.111-1-111-8

Carr, J. K., Nadai, Y., Eyzaguirre, L., Saad, M. D., Khakimov, M. M., Yakubov, S. K., Birx, D. L., Graham, R. R., Wolfe, N. D., Earhart, K. C., and Sanchez, J. L. (2005). Outbreak of a West African recombinant of HIV-1 in Tashkent, Uzbekistan. J Acquir Immune Defic Syndr 39(5), 570-5.

Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. Cell 89(2), 263-73.

Chan, D. C., and Kim, P. S. (1998). HIV entry and its inhibition. Cell 93(5), 681-4.

Chermann, J. C., Barre-Sinoussi, F., Dauguet, C., Brun-Vezinet, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a new retrovirus in a patient at risk for acquired immunodeficiency syndrome. Antibiot Chemother 32, 48-53.

Clayton, A. J. (1996). Possible Origins, Transmission Patterns, and Future Course of AIDS. J Travel Med 3(4), 231-234.

Coffin, J., Haase, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., and Vogt, P. (1986). Human immunodeficiency viruses. Science 232(4751), 697.

Curtis, T. (1992). Possible origins of AIDS. Science 256(5061), 1260-1.

De Baar, M. P., De Ronde, A., Berkhout, B., Cornelissen, M., Van Der Horn, K. H., Van Der Schoot, A. M., De Wolf, F., Lukashov, V. V., and Goudsmit, J. (2000). Subtype-specific sequence variation of the HIV type 1 long terminal repeat and primer-binding site. AIDS Res Hum Retroviruses 16(5), 499-504.

de Mendoza, C., Gallego, O., and Soriano, V. (2002). Mechanisms of resistance to antiretroviral drugs--clinical implications. *AIDS Rev* **4**(2)**,** 64-82.

Descamps, D., Collin, G., Letourneur, F., Apetrei, C., Damond, F., Loussert-Ajaka, I., Simon, F., Saragosti, S., and Brun-Vezinet, F. (1997). Susceptibility of human immunodeficiency virus type 1 group O isolates to antiretroviral agents: in vitro phenotypic and genotypic analyses. J Virol 71(11), 8893-8.

Descamps, D., Collin, G., Loussert-Ajaka, I., Saragosti, S., Simon, F., and Brun-Vezinet, F. (1995). HIV-1 group O sensitivity to antiretroviral drugs. Aids 9(8), 977-8.

Dowling, W. E., Kim, B., Mason, C. J., Wasunna, K. M., Alam, U., Elson, L., Birx, D. L., Robb, M. L., McCutchan, F. E., and Carr, J. K. (2002). Forty-one near full-length HIV-1 sequences from Kenya reveal an epidemic of subtype A and A-containing recombinants. Aids 16(13), 1809-20.

Duane W. Sears. (2003). The Human Immunodeficiency Virus particle. http://tutor.iscf.ucsb.edu/instdev/sears/immunology/. Retrieved on 15/06/09

Evatt, B. L., Stein, S. F., Francis, D. P., Lawrence, D. N., McLane, M. F., McDougal, J. S., Lee, T. H., Spira, T. J., Cabradilla, C., Mullens, J. I., and Essex, M. (1983). Antibodies to human T cell leukaemia virus-associated membrane antigens in haemophiliacs: evidence for infection before 1980. Lancet 2(8352), 698-700.

Fisher, I., Young, H., and Parzen, E. (1998). Sample determination in a randomized multicentric study. *Annals of statistic*(23)**,** 1631-1661.

Gallo, R. C. (1983). Human "T" leukemia virus still suspected in AIDS. Interview by John Maurice. Jama 250(8), 1015, 1021.

Gallo, R. C., and Montagnier, L. (2003). The discovery of HIV as the cause of AIDS. N Engl J Med 349(24), 2283-5.

Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J and Safai, B. (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science 224(4648), 500-3.

Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M., and Hahn, B. H. (1999). Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature 397(6718), 436-41.

Garrido, C., Zahonero, N., Fernandes, D., Serrano, D., Silva, A. R., Ferraria, N., Antunes, F., Gonzalez-Lahoz, J., Soriano, V., and de Mendoza, C. (2008). Subtype variability, virological response and drug resistance assessed on dried blood spots collected from HIV patients on antiretroviral therapy in Angola. J Antimicrob Chemother 61(3), 694-8.

Garry, R. F. (1989). Potential mechanisms for the cytopathic properties of HIV. Aids 3(11), 683-94.

Gaschen, B., Taylor, J., Yusim, K., Foley, B., Gao, F., Lang, D., Novitsky, V., Haynes, B., Hahn, B. H., Bhattacharya, T., and Korber, B. (2002). Diversity considerations in HIV-1 vaccine selection. Science 296(5577), 2354-60.

Gilbert, M. T., Rambaut, A., Wlasiuk, G., Spira, T. J., Pitchenik, A. E., and Worobey, M. (2007). The emergence of HIV/AIDS in the Americas and beyond. Proc Natl Acad Sci U S A 104(47), 18566-70.

Gotch, F., Rutebemberwa, A., Jones, G., Imami, N., Gilmour, J., Kaleebu, P., and Whitworth, J. (2000). Vaccines for the control of HIV/AIDS. Trop Med Int Health 5(7), A16-21.

Gottlieb, M. S., Schroff, R., Schanker, H. M., Weisman, J. D., Fan, P. T., Wolf, R. A., and Saxon, A. (1981). Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N Engl J Med 305(24), 1425-31.

Grasso, C., and Lee, C. (2004). Combining partial order alignment and progressive multiple sequence alignment increases alignment speed and scalability to very large alignment problems. *Bioinformatics* **20**(10)**,** 1546-56.

Greener, R. (1998). Why companies should intervene: a case study of the costs of HIV / AIDS to employers. Economic impact: Southern Africa. AIDS Anal Afr 8(2), 3-4.

Heng, M. C. (1983). Pathogenesis of the acquired immune deficiency syndrome (AIDS). Cutis 32(3), 255-7.

Hillis, D. M. (2000). AIDS. Origins of HIV. Science 288(5472), 1757-9.

Hirsch, V. M., Olmsted, R. A., Murphey-Corb, M., Purcell, R. H., and Johnson, P. R. (1989). An African primate lentivirus (SIVsm) closely related to HIV-2. Nature 339(6223), 389-92.

HIV entry in to the host cell and transportation. Adapted from www.brookscole.com/.../ch_HIVLifecycle.html. Retrieved on 21/06.2009

Ho, D. D., Moudgil, T., and Alam, M. (1989). Quantitation of human immunodeficiency virus type 1 in the blood of infected persons. *N Engl J Med* **321**(24)**,** 1621-5.

Hoelscher, M., Kim, B., Maboko, L., Mhalu, F., von Sonnenburg, F., Birx, D. L., and McCutchan, F. E. (2001). High proportion of unrelated HIV-1 intersubtype recombinants in the Mbeya region of southwest Tanzania. Aids 15(12), 1461-70.

Hu, D. J., Baggs, J., Downing, R. G., Pieniazek, D., Dorn, J., Fridlund, C., Biryahwaho, B., Sempala, S. D., Rayfield, M. A., Dondero, T. J., and Lal, R.

(2000). Predominance of HIV-1 subtype A and D infections in Uganda. Emerg Infect Dis 6(6), 609-15.

Hu, D. J., Dondero, T. J., Rayfield, M. A., George, J. R., Schochetman, G., Jaffe, H. W., Luo, C. C., Kalish, M. L., Weniger, B. G., Pau, C. P., Schable, C. A., and Curran, J. W. (1996). The emerging genetic diversity of HIV. The importance of global surveillance for diagnostics, research, and prevention. Jama 275(3), 210-6.

Huet, T., Cheynier, R., Meyerhans, A., Roelants, G., and Wain-Hobson, S. (1990). Genetic organization of a chimpanzee lentivirus related to HIV-1. Nature 345(6273), 356-9.

Johnston, M. I., and Fauci, A. S. (2007). An HIV vaccine--evolving concepts. N Engl J Med 356(20), 2073-81.

Joint United Nations programme on HIV/AIDS(UNAIDS) and World Health Organization(WHO). (2007). Global summary of the HIV/AIDS Epidemic. Newsletter http://www.unaids.org/en/KnowledgeCentre/HIVData/GlobalReport/

Khamadi, S. A., Ochieng, W., Lihana, R. W., Kinyua, J., Muriuki, J., Mwangi, J., Lwembe, R., Kiptoo, M., Osman, S., Lagat, N., Pelle, R., Muigai, A., Carter, J. Y., Oishi, I., Ichimura, H., Mwaniki, D. L., Okoth, F. A., Mpoke, S., and Songok, E. M. (2005). HIV type 1 subtypes in circulation in northern Kenya. AIDS Res Hum Retroviruses 21(9), 810-4.

Kiwanuka, N., Laeyendecker, O., Robb, M., Kigozi, G., Arroyo, M., McCutchan, F., Eller, L. A., Eller, M., Makumbi, F., Birx, D., Wabwire-Mangen, F., Serwadda, D., Sewankambo, N. K., Quinn, T. C., Wawer, M., and Gray, R. (2008). Effect of human immunodeficiency virus Type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. J Infect Dis 197(5), 707-13.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S., and Bhattacharya, T. (2000). Timing the ancestor of the HIV-1 pandemic strains. Science 288(5472), 1789-96.

Kumar, S., Tamura, K., Jakobsen, I. B., and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17(12), 1244-5.

Land, A. M., Luo, M., Pilon, R., Sandstrom, P., Embree, J., Wachihi, C., Kimani, J., Plummer, F. A., and Ball, T. B. (2008). High prevalence of genetically similar HIV-1 recombinants among infected sex workers in Nairobi, Kenya. AIDS Res Hum Retroviruses 24(11), 1455-60.

Lau, C. Y., Velasco, P. P., and Johnston, M. I. (2007). A new era in HIV vaccine development. Expert Rev Anti Infect Ther 5(2), 205-15.

Levy, J. A. (1993). Pathogenesis of human immunodeficiency virus infection. Microbiol Rev 57(1), 183-289.

Levy, J. A., A. D. Hoffman. (1984). "Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS." Science **225**(4664): 840-2

Li, B., Gladden, A. D., Altfeld, M., Kaldor, J. M., Cooper, D. A., Kelleher, A. D., and Allen, T. M. (2007). Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. J Virol 81(1), 193-201.

Lihana, R. W., Khamadi, S. A., Kiptoo, M. K., Kinyua, J. G., Lagat, N., Magoma, G. N., Mwau, M. M., Makokha, E. P., Onyango, V., Osman, S., Okoth, F. A., and Songok, E. M. (2006). HIV type 1 subtypes among STI patients in Nairobi: a genotypic study based on partial pol gene sequencing. AIDS Res Hum Retroviruses 22(11), 1172-7.

Little, S. J., Holte, S., Routy, J. P., Daar, E. S., Markowitz, M., Collier, A. C., Koup, R. A., Mellors, J. W., Connick, E., Conway, B., Kilby, M., Wang, L., Whitcomb, J. M., Hellmann, N. S., and Richman, D. D. (2002). Antiretroviral-drug resistance among patients recently infected with HIV. N Engl J Med 347(6), 385-94.

Loussert-Ajaka, I., Ly, T. D., Chaix, M. L., Ingrand, D., Saragosti, S., Courouce, A. M., Brun-Vezinet, F., and Simon, F. (1994). HIV-1/HIV-2 seronegativity in HIV-1 subtype O infected patients. Lancet 343(8910), 1393-4.

Louwagie, J., Janssens, W., Mascola, J., Heyndrickx, L., Hegerich, P., van der Groen, G., McCutchan, F. E., and Burke, D. S. (1995). Genetic diversity of the

envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. J Virol 69(1), 263-71.

Malim, M. H., and Emerman, M. (2001). HIV-1 sequence variation: drift, shift, and attenuation. Cell 104(4), 469-72.

Margaret I. Johnston, Ph.D., and Anthony S. Fauci, M.D. (2007) An HIV Vaccine — Evolving Concepts. N Engl J Med; 356:2073-81.

Marx, J. L. **(1985). AIDS virus genomes. Science 227(4686), 503.**

Masur, H., Michelis, M. A., Greene, J. B., Onorato, I., Stouwe, R. A., Holzman, R. S., Wormser, G., Brettman, L., Lange, M., Murray, H. W., and Cunningham-Rundles, S. (1981). An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. N Engl J Med 305(24), 1431-8.

Mathe, G. (1983). [AIDS and its association with human tumors and viruses. A viral and/or immunogenetic cause?]. Biomed Pharmacother 37(4), 153-9.

McCutchan, F. E. (2006). Global epidemiology of HIV. J Med Virol 78 Suppl 1, S7-S12.

McCutchan, F. E., Viputtigul, K., de Souza, M. S., Carr, J. K., Markowitz, L. E., Buapunth, P., McNeil, J. G., Robb, M. L., Nitayaphan, S., Birx, D. L., and Brown, A. E. (2000). Diversity of envelope glycoprotein from human immunodeficiency virus type 1 of recent seroconverters in Thailand. AIDS Res Hum Retroviruses 16(8), 801-5.

Modlin, R. L., Hofman, F. M., Meyer, P. R., Vaccaro, S. A., Ammann, A. J., Conant, M. A., Rea, T. H., and Taylor, C. R. (1983). Altered distribution of B and T

lymphocytes in lymph-nodes from homosexual men with Kaposi's sarcoma. Lancet 2(8353), 768-71.

Montavon, C., Bibollet-Ruche, F., Robertson, D., Koumare, B., Mulanga, C., Esu-Williams, E., Toure, C., Mboup, S., Saman, E., Delaporte, E., and Peeters, M. (1999). The identification of a complex A/G/I/J recombinant HIV type 1 virus in various West African countries. AIDS Res Hum Retroviruses 15(18), 1707-12.

Myers, G. (1994). Tenth anniversary perspectives on AIDS. HIV: between past and future. AIDS Res Hum Retroviruses 10(11), 1317-24.

Nasioulas, G., Paraskevis, D., Magiorkinis, E., Theodoridou, M., and Hatzakis, A. (1999). Molecular analysis of the full-length genome of HIV type 1 subtype I: evidence of A/G/I recombination. AIDS Res Hum Retroviruses 15(8), 745-58.

National AIDS and STI Control Programme, Ministry of Health, Kenya. July 2008. Kenya AIDS Indicator Survey 2007: Preliminary Report. Nairobi, Kenya

Neilson, J. R., John, G. C., Carr, J. K., Lewis, P., Kreiss, J. K., Jackson, S., Nduati, R. W., Mbori-Ngacha, D., Panteleeff, D. D., Bodrug, S., Giachetti, C., Bott, M. A., Richardson, B. A., Bwayo, J., Ndinya-Achola, J., and Overbaugh, J. (1999). Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. J Virol 73(5), 4393-403.

Nerrienet, E., Santiago, M. L., Foupouapouognigni, Y., Bailes, E., Mundy, N. I., Njinku, B., Kfutwah, A., Muller-Trutwin, M. C., Barre-Sinoussi, F., Shaw, G. M.,

Sharp, P. M., Hahn, B. H., and Ayouba, A. (2005). Simian immunodeficiency virus infection in wild-caught chimpanzees from cameroon. J Virol 79(2), 1312-9.

Nielsen, R., and Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148(3), 929-36.

Ortiz, M., Munoz, L., Bernal, A., Rodriguez, A., Zorraquino, A., Vadillo, J., Salas, A., Moreno, A., and Garcia-Saiz, A. (2000). Molecular characterization of non-B HIV type 1 subtypes from Africa in Spain. AIDS Res Hum Retroviruses 16(18), 1967-71.

Osmanov, S., Pattou, C., Walker, N., Schwardlander, B., and Esparza, J. (2002). Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. J Acquir Immune Defic Syndr 29(2), 184-90.

Peeters, M. (2001). The genetic variability of HIV-1 and its implications. Transfus Clin Biol 8(3), 222-5.

Peeters, M., and Delaporte, E. (1999). [Genetic diversity of HIV infection worldwide and its consequences]. Med Trop (Mars) 59(4 Pt 2), 449-55.

Perrin, L., Kaiser, L., and Yerly, S. (2003). Travel and the spread of HIV-1 genetic variants. *Lancet Infect Dis* **3**(1)**,** 22-7.

Picard, O., Gorin, I., Leibowitch, M., Laroche, L., Puchault, P., Escande, J. P., and Hewitt, J. (1982). [Kaposi's disease. 2 recent cases]. Nouv Presse Med 11(45), 3335-8.

Quinones-Mateu, M. E., Albright, J. L., Mas, A., Soriano, V., and Arts, E. J. (1998). Analysis of pol gene heterogeneity, viral quasispecies, and drug resistance in

individuals infected with group O strains of human immunodeficiency virus type 1. J Virol 72(11), 9002-15.

Rambaut, A., Robertson, D. L., Pybus, O. G., Peeters, M., and Holmes, E. C. (2001). Human immunodeficiency virus. Phylogeny and the origin of HIV-1. Nature 410(6832), 1047-8.

Ratner, L., Starcich, B., Josephs, S. F., Hahn, B. H., Reddy, E. P., Livak, K. J., Petteway, S. R., Jr., Pearson, M. L., Haseltine, W. A and Arya, S. K. (1985). Polymorphism of the 3' open reading frame of the virus associated with the acquired immune deficiency syndrome, human T-lymphotropic virus type III. Nucleic Acids Res 13(22), 8219-29.

Richman, D. D., Wrin, T., Little, S. J., and Petropoulos, C. J. (2003). Rapid evolution of the neutralizing antibody response to HIV type 1 infection. Proc Natl Acad Sci U S A 100(7), 4144-9.

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P. M., Wolinsky, S., and Korber, B. (2000). HIV-1 nomenclature proposal. Science 288(5463), 55-6.

Robertson, D. L., Sharp, P. M., McCutchan, F. E., and Hahn, B. H. (1995). Recombination in HIV-1. Nature 374(6518), 124-6.

Robertson, D. L., Hahn, B. H., and Sharp, P. M. (1995). Recombination in AIDS viruses. *J Mol Evol* **40**(3)**,** 249-59.

Ross, L., Lim, M. L., Liao, Q., Wine, B., Rodriguez, A. E., Weinberg, W., and Shaefer, M. (2007). Prevalence of antiretroviral drug resistance and resistance-associated mutations in antiretroviral therapy-naive HIV-infected individuals from 40 United States cities. HIV Clin Trials 8(1), 1-8.

Saldanha, A. J. (2004). Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**(17)**,** 3246-8.

Schable, C., Zekeng, L., Pau, C. P., Hu, D., Kaptue, L., Gurtler, L., Dondero, T., Tsague, J. M., Schochetman, G and Jaffe, H. (1994). Sensitivity of United States HIV antibody tests for detection of HIV-1 group O infections. Lancet 344(8933), 1333-4.

Sharp, P. A., Sugden, B., and Sambrook, J. (1973). Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose--ethidium bromide electrophoresis. *Biochemistry* **12**(16)**,** 3055-63.

Sharp, P. M., Bailes, E., Chaudhuri, R. R., Rodenburg, C. M., Santiago, M. O., and Hahn, B. H. (2001). The origins of acquired immune deficiency syndrome viruses: where and when? Philos Trans R Soc Lond B Biol Sci 356(1410), 867-76.

Smith, S. M., Christian, D., de Lame, V., Shah, U., Austin, L., Gautam, R., Gautam, A., Apetrei, C., and Marx, P. A. (2008). Isolation of a new HIV-2 group in the US. Retrovirology 5, 103.

Songok, E. M., Lihana, R. W., Kiptoo, M. K., Genga, I. O., Kibaya, R., Odhiambo, F., Kobayashi, K., Ago, Y., Ndembi, N., Okoth, F., Fujiyama, Y., Imanishi, J., and Ichimura, H. (2003). Identification of env CRF-10 among HIV variants circulating in rural western Kenya. AIDS Res Hum Retroviruses 19(2), 161-5.

Takeuchi, Y., Nagumo, T., and Hoshino, H. (1988). Low fidelity of cell-free DNA synthesis by reverse transcriptase of human immunodeficiency virus. *J Virol* **62**(10)**,** 3900-2.

Tantillo, C., Ding, J., Jacobo-Molina, A., Nanni, R. G., Boyer, P. L., Hughes, S. H., Pauwels, R., Andries, K., Janssen, P. A., and Arnold, E. (1994). Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase. Implications for mechanisms of drug inhibition and resistance. J Mol Biol 243(3), 369-87.

Thomson, M. M., Perez-Alvarez, L., and Najera, R. (2002). Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infect Dis* **2**(8)**,** 461-71.

Thongcharoen, P., Suriyanon, V., Paris, R. M., Khamboonruang, C., de Souza, M. S., Ratto-Kim, S., Karnasuta, C., Polonis, V. R., Baglyos, L., Habib, R. E., Gurunathan, S., Barnett, S., Brown, A. E., Birx, D. L., McNeil, J. G., and Kim, J. H. (2007). A phase 1/2 comparative vaccine trial of the safety and immunogenicity of a CRF01_AE (subtype E) candidate vaccine: ALVAC-HIV (vCP1521) prime with oligomeric gp160 (92TH023/LAI-DID) or bivalent gp120 (CM235/SF2) boost. J Acquir Immune Defic Syndr 46(1), 48-55.

Triques, K., Bourgeois, A., Vidal, N., Mpoudi-Ngole, E., Mulanga-Kabeya, C., Nzilambi, N., Torimiro, N., Saman, E., Delaporte, E., and Peeters, M. (2000). Near-full-length genome sequencing of divergent African HIV type 1 subtype F viruses leads to the identification of a new HIV type 1 subtype designated K. AIDS Res Hum Retroviruses 16(2), 139-51.

Tuaillon, E., Gueudin, M., Lemee, V., Gueit, I., Roques, P., Corrigan, G. E., Plantier, J. C., Simon, F., and Braun, J. (2004). Phenotypic susceptibility to nonnucleoside inhibitors of virion-associated reverse transcriptase from different HIV types and groups. J Acquir Immune Defic Syndr 37(5), 1543-9.

Vernazza, P. L., Troiani, L., Flepp, M. J., Cone, R. W., Schock, J., Roth, F., Boggian, K., Cohen, M. S., Fiscus, S. A., and Eron, J. J. (2000). Potent antiretroviral treatment of HIV-infection results in suppression of the seminal shedding of HIV. The Swiss HIV Cohort Study. Aids 14(2), 117-21.

Wain-Hobson, S. (1993). The fastest genome evolution ever described: HIV variation in situ. Curr Opin Genet Dev 3(6), 878-83.

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

Wensing, A. M., and Boucher, C. A. (2003). Worldwide transmission of drug-resistant HIV. AIDS Rev 5(3), 140-55.

Wensing, A. M., van de Vijver, D. A., Angarano, G., Asjo, B., Balotta, C., Boeri, E., Camacho, R., Chaix, M. L., Costagliola, D., De Luca, A., Derdelinckx, I., Grossman, Z., Hamouda, O., Hatzakis, A., Hemmer, R., Hoepelman, A., Horban, A., Korn, K., Kucherer, C., Leitner, T., Loveday, C., MacRae, E., Maljkovic, I., de Mendoza, C., Meyer, L., Nielsen, C., Op de Coul, E. L., Ormaasen, V., Paraskevis, D., Perrin, L., Puchhammer-Stockl, E., Ruiz, L., Salminen, M., Schmit, J. C., Schneider, F., Schuurman, R., Soriano, V., Stanczak, G., Stanojevic, M., Vandamme, A. M., Van Laethem, K., Violin, M., Wilbe, K., Yerly, S., Zazzi, M., and Boucher, C. A. (2005). Prevalence of drug-resistant HIV-1 variants in untreated individuals in Europe: implications for clinical management. J Infect Dis 192(6), 958-66.

Worobey, M., Santiago, M. L., Keele, B. F., Ndjango, J. B., Joy, J. B., Labama, B. L., Dhed, A. B., Rambaut, A., Sharp, P. M., Shaw, G. M., and Hahn, B. H. (2004). Origin of AIDS: contaminated polio vaccine theory refuted. Nature 428(6985), 820.

Zhang, M., Wilbe, K., Wolfe, N. D., Gaschen, B., Carr, J. K., and Leitner, T. (2005). HIV type 1 CRF13_cpx revisited: identification of a new sequence from Cameroon and signal for subsubtype J2. AIDS Res Hum Retroviruses 21(11), 955-60.

Zhu, T., Korber, B. T., Nahmias, A. J., Hooper, E., Sharp, P. M., and Ho, D. D. (1998). An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature 391(6667), 594-7.

APPENDIX I: STUDY SAMPLE COLLECTION SITES

http://www.nationsonline.org/oneworld/map/Kenya_map2.htm

APPENDIX II: INFORMED CONSENT INFORMATION

PART A; Consent Explanation

Study Title: Genetic Diversity of HIV-1 in Central Province of Kenya- A preliminary study

Information to Participant

HIV infections are acquired through unprotected sexual relations, blood and blood products or mother to child (MTCT) which occurs *in utero* during pregnancy and *intrapartum* at childbirth. HIV is highly divergent and has been classified in to sub-divisions known as subtypes. These subtypes are believed to exhibit different characteristics in terms of diagnosis, virulence and disease progression; posing a major challenge in regard to vaccine development and success of ARV therapy. Thus it is important to have knowledge about the prevailing subtypes so as to make informed decisions in terms of monitoring and implementation of control measures. It is for these reasons that we are interested in carrying out this investigation using blood samples from your facility.

Procedure to be followed

About 3ml of HIV positive blood already collected by technical staff for diagnosis will be passed on to me once the intended tests have been done.

Benefits of the study

The study will not have immediate benefits to the participants or your facility but the information obtained will be useful in improving medical attention offered.

Confidentiality

Information about the patient's name and background will not be asked for and results will not be linked to the individual facility.

Medical problems

Not expected. However should we observe any complications we will alert you for immediate attention

PART B;

Consent seeking form

You are encouraged to ask any questions that occur to you at this time or any time in the course of our interaction. If more information is required at a later date, you may call **Sheila Kageha** on telephone no.0722 587808 at any time.

Participation

You are free to accept or decline to participate in this study. You are also free to withdraw your consent to participate in this study at any time you choose. Your decision to withdraw will not affect your rights now or in the future.

I have read the information stated above and have had the opportunity to ask questions all of which were answered satisfactorily. I hereby give consent for my facility's participation as explained to me

Participant's signature……... Date…….

I, the undersigned have fully explained the relevant details of this study to the participant named above. I am qualified to perform the role of principal investigator in this study.

Signature……..Name……Date……
(Investigator)

Signature……Name…….Date……
(Co-investigator)

Address of Co-investigator……..

APPENDIX III: GENERATED SEQUENCES

These sequences were generated from HIV-1 positive samples from the Central province of Kenya. The sequenced region was HIV-1 *env gp41*

>NPGH_001
AgATTcTcAtatACACAtGCAGaTTGTTATGTTGAAcCAACTCCACAGATTTGCCCACTTGTCCgAT
GCCAATAAGTCCTGTTCATTCTTTTCCTGCTGGTTCTGCaATTCTTCAAGTAGACTATATATTA
TATTTGTGTAATTGCTAACTTCTTTATCCCATTGCAGCCAGGTCATGTTTTCCCATATCTCAtC
TAGGgATCTATTACTCCAaCTAGAGTtCCAGCGcACATTAGTGGTGCAGATGAGTTTTCCAGAG
CAGCCCCAAATTCCTAGGAGCTGTTGATCAcTTAGGTATCTTTCCaTAGCCAGGACTCTtGCCT
GgAGCTGTTTAATGCCCCaGACcGTGAGTCTcaACAgAtGCTGttgAgcCTcTATAGccCTcAGcAAA
TtgcTTtgctgtTGCaCTATACcAGAcAATaaTTgTAATGATAGTAGGAGgCtTgaataagtTtaAAAA

>NPGH_002
GATTgTTATATACACAGCATCTTGATATGTCAAACCAGTTCCACAGACTTGCCCATTTGTCAA
aTgCCAAtAACTCTTGTTCATTCTTTTCCTGCTGGATTTGCGATTCCCCAaTTAGTTGATATATG
TAGTCTGTGTAATTgCTAACTTCTTTATCCCATTGCaGCCATGTCATGTTGTTCCATATGTCCT
CAAAAGATTTATTACTCCAACTAGAGTTCCAGGGCACATTAGTGGTGCAGATGAGTTTTCCA
GAGCAACCCCACATTCCTAGGAGCTGTTGATCCTTTAGGtATCTTTCCACAGCCAGGACTCTT
GCCTGGAGCTGTTTAATGCCCCAgACCGtGAGTTTCAACAGGTGTtGTTGAGCCtCTATAGCCC
TCAGCAAATTGCTTTGctGTTGCACTATACCAGAAACTAATTGTAGGGATAGTgGGaGGctTgGA
GGgtCCTC

>NPGH_003
GATTCGTATATACACAGCAGTTTGATATGTCAACCAAtTCCACAGACTTGCCCATTTGTCCAA
TgCCAATAAaTCTTGtTCATTCTTTTCCTGCTGGTTCTGTGaTTcTTCAATTAGcTcgTATATTGTT
TGTGTGTAATTGCTAATTTCTcTATCCCATTGcAGCCAGGTCATGTTGTCCCATATCTCACTCT
GAGATTTAtTaCTCCAaCTAtcGTTCCAGgGCACATTAGTgGTGCAGATGAGTTTTCCAGAGCAa
CCCCAcATTCCTAGGAGCTGTtGATCCTTTAGcTATCTtTCCACAGCcAGGACTCTTGCCTGGAG
CTGTTTAATGCCCCAGACtGTGAGTTTCAACAGATGTTGTTGAGCCTCTATAGCCCTCAGCAA
ATTGCTTTGcTgTTgCACTATACCAGCAATAATTggTAGGGGGGGGGGGG

>NPGH_004
GCTTGTTGTGGtcACTGCaaAcATaTaGaACTGaCaAGTCCaCAgATTTGCCCACTTGTCaaTGCCtA
TAAGTCTTGTTCATTCTTTTCCTGCtGGaTCTGCgATTCTTCAAGTaGATcATATATTAACTGTGT
gTAATTGCTATTTCTTTaTCCCATTGCAGCCAGGTCATGTTTTcCCATATCTCATCCTGGGATTT
aTTAcTCCAACtAGAgTTCCAGGGCacAgcAGTGGTGcAGATGAGTtTTccAGAGCAgCCCCAAAtT
CCTAGgAgCTGTTGATCCTTTAaGgTATCTCTCCACAGCCCAGgACTcTTGGCcTGGAGCTGTTT
AAATGCCCCAGACCGtGGAGTTTTcAaCAGATGCtGtTTGAGCCCTCTATAGCcTTCaGcAAAATT
ATTTTTGCTGTTTGCACTAAtACCAGACAaTTaaTTgTTTtgATAGtAGGGaGGccTg

>NPGH_005
ATTCTTATATACACAGCAATTTGATATGTCAACCAATTCCACAGACTTGCCCACTTGTCCAGT
gCCaaTAAgTCTTGTTCATTCTTTTCCTGCTGGTTCTGTGATTCTTCAATTAgGTTATATATTATG
CTTGTGTAATTcccAATTTCTTTATCCCATTGCATCCAGGTCATGTTGTCCCATATCTCcTGCAT
AGATTTATTACTCCAACTAGAGTTCCAGGGCACACTAGTGGTGCAGATGAGTCTTCCAGAGC
AGCCCCAAATTCCTAAGAGCTGTTGGTCCTTTAGGTACCTcTCCACAGCCAGGACTCTTGCCT
GGAGCTGTTTAATGCCCCAGACCGTGAGTCTCAACAGgTGTTGTTGAGCCTCTATAGCCCTCA
GCAAATTGCTTTGCTGTTGCACTATACCAGcAAtTAATtGTACGtAGTaGGAGGCtCGAGAGGTT
AaaGGGGGGGGG

>NPGH_006
GtTGGTCGGgGTTActGCTAGcaTaATATGTCAaaCAATTcgCaGACTTGCCCAaTtGTCCgATGCCtA
TAAATCTTGtTCATTCTTTTCCTGCtGGtTCTGTGattCTTCAATTAGCTTGTATATTGTTTGTGTG
TAATTGCTAATTTCTCTATCCCATTGtAGCCAGGtCATGTTGtCCCAtATCTCACTTTGAgATTTA
CTACTCCAACTAGAGTTCCAGGGcACATTAGtGGtGCAGATGAGTTTTCCAGAgCAGCCCCAcA
TTCcTAGGAGCtGTTGATCCTTTAGGTATCTTTcCaCAGcCAGGACTCTTGCCTGGaGCTGTTTA
AtGCCCCAGACTGTGAGTTTCAACAGttGTTGTTGAGCCTCTATaGcCCTcAgCAAATTGcTTTgcT
gtTGCACTATACcagacaaTaaTTGTAAAGgtaGgaGGggGcTTtGtaaaGtTTaaAA

>NPGH_007
CTCCaTTtTCAGATTGCCCATTTatcAGGGGAAggAAggGgGtTCATTCTGtACctgCtggtTCtGGAActg
ggaagagacTATataTtaGccCtGgGtAAtTGcTAaTTtCTTTaTCCgggGGgaGgcgGGtCtTTTCCCcCcAtAtcT
CattCaGgGAtTTATTACTCCAaCTagcaTtCCaGGGcACaGtAGtGGtgCagAtgAgTtTtCCAGAGCaGCCc
CAAaTTCcTaGGcGCtgTTGatCCacTaGGTaTCTTTCCACaGCCAGgaCTCTTGTcTGGaGCTGTTTaA
tGCCCaAgACCGGGAgTTtcAACAAAATcTGTTGaGcCTTTaTTcCCtCAGaAATTgTTTTTcGgTTGC
CCAAACtttAtAAAAAAAAAA

>NPGH_008
GACGCTTTATatACTACAgGCAGTTTGAtATGTCTaCTAAGTCCACaGACTTgTCcATTTGtCaaAT
GCCACTaAGtCTTGtTCATTCTTTTCCTGCTGGgTCTGCGATTCTTCAATTAGGCcATATATTATg
cCTGTATAATTGCTAATTTTTTTCTCTtATTGTAGCCAGGTCATGTtTTCcCATATCTCAGCCTGA
GATtTATTACTTTACTAGTgCCTAGggcAgTTAGGAaAaATaGCgTGTCaGagCgCAtAAATTAGGG
AGACTTGtCTTAaTtCCCTCTcGGcGgCCttGGcGAAaaATTGGCcCCctGGcTGCCGGGGGGGcggcA
AGGGTGAAaAAcGTTCCTTtCtTAAAgTatTttTGTAgTgGGgTgtTgttaAtAAGGGGGGGGGGGGG

>NPGH_009
AAACCAATTCCACAGATTTTGCCACTTGTCCAaTGGGaAtAAATCCTTTTCATTGTTCTCCTGC
TGGATTTGCGAGTcTTCAaGCAaACTATATATTATTTTTGTGTAATTACTAATTTCTTTATCCCA
CTGCATCCAGGTCATGTTACCCCAAAtATCCTCTtGAGATTTaTTACTCCAACTAGtGTTCCAaG
GcACATTAGTGGtGCAGATGAGTTTTCCAgAGCAGCCCCAAATCcctAGGAGCtGTTGATCCCTT
AGGtATCTTTCTATAGCCAGGACTCTTGTCTGAAGCTGCTTAATGCCCCAgACTGTGAGTTGC
AACATATGCtGttgGCCTCTaTAGCCCTcAGCAAATTGCTTTctGTTGCgCTATacTTTTTTTtTaAtTgt
A

>NPGH_010
GtTTCATATATACCACaGCTaGTTTGagACGCTAAACCaATTCCACAGATCTGCCCACTTGtCTaA
TcTtAtTAAGgCTTGTTCATTCTTTTCCTGCTGGTTCtGCcATTCTTCaATTAGCCTATATATTAggT
TTGTGTAAtTGTtAAtTTCTTTATCCCAtTCcAACCAGGTCATGTTATCCCATATCTCAcTGTAAG
ATTTATTACTCCAACTAGAGTTCCAGGGCACAgTAGTGGTGCAGATGAGTTTTCCAGAGCAG
CCCCAAATTCCTAGGAGCTGTTGATCCCTTAGGTAGCTTTCCACAGCCAGGAcTCTTGCCTGG
AGCTGTTTAATGCCCCAGACCGTGAGTTTCAACAGATGTtGTTGAGCCTCTATAGCCCTCAGC
AAATTACTTTGCTGTTGCACTATACCAGAaaaTaaTTGTAAaGATAGTAGGAGgCtTtataaggtTTtA
AAAAGGGG

>NPGH_011
TCATtTcGcCTGcCATGACAACcGaCCaaaCGACGATTTctCaGATTTGCCCAtTTGTCcaTGCCtATA
AGTCTTGTTCATTCTTTTCCTGCTGGtTCTGCgAgTCTTCAAgTAGtTGATATATGTAATCTGTGT
AATTGCTAATTTCTTTATCCCATTGTAGCCAgGTCATGTTGTCCCATATCTCActCtGAGaTCTAT
TACTCCAACTAGAGTTCCAGGGCACATTAGtGGagcAGATGAGTTTTCCAGAGCAGCCCCAaAT
TCcTAGGAGCTGTTGATCCtTTAGGTaTCTTTCTATAGCCAGgaCTCTTGCCTGGAGCTGTTTAA
TGCCCcagActGTgAGcCTcAACAgATGttGTTGAGCCTCTATAGcCCTcAGCAAATTGCTTTGcTGT
TGCACTATACCAgAcAaTAATTGTAATGaTAGtAGGaggTTGg

>NPGH_012

GaTtTGTtTAtATACCACaGCATCTTGATATGTCaGAcCAATTCCACAGATTTGTCCATTTGTCTaA
TGCCAaTAAATCTTGtTCATTTTTCTCCTGcTGGTTCTGCgATTCTTCAATTAGGTTATATATTAT
GTCTGTGTAAtTgCTAATTTCTCTTTCCcaCTGTTGCcAGGTCATGTTTTCCCATATGTCCtCTaGA
GtTATATTATTACTCCaACTAGAGtTCCAAGGCACATTAGTGGTGCAGAtGAgTCTTCCAGAGC
AGCCCCAAATTCcTAGGAGCTGTTGATCCCTTAGGTATCTtTCCAcAGCcAGGATTCTTGCCTG
GAGCTGTTTAATGCCCCAGACTGTgAGtTTCAACAGATGTTGTTGAGCCTCTaTAGcCCTcaGCa
AATTGTTtTGCTGTtGCACTATaCCAGAcAaTaATTGTAATgaTAGTAGGAGgcTTGagaatggTTtAA
AA

>MDH_001
GaGccCTaaTGTCCgATtCTAATAGAtCTTTTTCGTTTTTGTCTTGCTGGATTtGCGAgtCTTCaAGtA
AGcTGTATATTAACtCTGTGTAATTGTCAATTTCTTTTTCCCACTGCATCCAGGTCATtTTATcC
CaAATCTCctCTAGAGATTTATTACTCCAaCTAGAGtTCCAGGGCACAGTaGTGGTGCATATGTG
TTTTCCAGAGCAACCCCAAACTCCTAGgAGCTGTTGATCCTTTAGGtATCTTTCCAcAGCcAGg
ATTCTTGCCTGGAGCTGTTTAATGCCCCAGACTGTGAGTtGCAACATATGCTGTTGCGCcTCTa
TAGcCCTcAGCaAATTGTTTTGCTGTTGCACTATACcAGAcAaTaATTGTaAtgaTAGtAGGAGGTT
ggtTGGGTTTAAA

>MDH_002
GatTTaTATATACCACAGCaCTTTGATATATCAACCAAGTCCACAGACTTGCCCACTTGTCCAA
TGCCaATAAGTCTTGTTCATTCTTTTCCTGCTGGATCTGTGATTCTTCAATTAGGCTATATATT
ATCTGTGTGTAATTGCTAATTTCTTTATCCCATTGTAgccAGGTCATGTTCTCCCATATCTCTTC
CTGAGATTTATTACTCCAACTAGAGTTCCAGGGCACACTAGTGGTGCAGATGAGTTTTCCAG
AGCAACCCCAAATTCCTAGGAGCTGTTGATCATGTAGGTATCTTTCTACAGCCAGGACTCTT
GCCTTGAGCTGTTTTATGCCCCAGACAGTGAGTTTCAACAGATGTTGTTGAGCCTCTATAGCT
CTCAGCAAATTGTTTTGCTGTTGCACTATACCAGaAAAttAaTAgatAAAGG

>MDH_003
GgaTtCaaTaCACAGCCATTtGATATGTCAACCAACTCCACAGACTTGcCCACTTGTCCAATGCCA
aTAAGTCTTGTTCATTCTTTTCCTGCTGGTTCTGcGAtctTTCAATTAGGtCATATATGTATTCTGT
GTAATTGCTAATTTCTTTATCCCATTGCAGCCAGGTCATGTTATCCCATATCTGATTGTAAGA
TTTATTgCTCCAACTAGAGTTCCAGGGCACATTAGTGGTGCAGATGAGTTTTCCAGAGCAGCC
CCAAATTCCTAGGAGCTGTTGATCTCTTAGGTATCTTTCTACAGCCAGGACTCTTGCCTGGAG
CTGTTTAATGCCCCAGACTGTGAGTCTCAACAGATGCTGTTGAGCCTCTATAGCCCTCAGCA
AATTGCTTTGCTGTTGCACTATACAAcAATGACTGTC

>MDH_004
AaCGtGTCATgTACCAcATGCAGAGaTaATATaTCaAacCaATtccaTAGaCtTGcCcACTCATTTaATG
CCAATAAGTCTTGTTCATTCTTTTCCtGCTGGtTCTGCGATTtCTCAAGTAGGTtATaTATTGTTT
GTGAGtAATTGCTAATTTCTTTATCCCaTTGTAGCCAGGTCATGTTGTCCCATATCTCTTCCtGA
GATTTATTACTCCAACTAGAaTTCCAGGGCACATTAGTGGTGCAGAtGAGTTTTCCAGAGCAA
CCCCAAATTCCtAGGAGCTGTTGATCCTTTAGGtATCTTTCCacAGcAAGGACTCGTGCCTGGA
GCTGTTTAAtGCCCcagAccGTGAGCCTcAACATCTGTtGTTgaGCcTcTaTAGcgATcAGCATAtTAc
TTTgctGTtGcaCTATaCCAGAcAaTaATTGTAAtGgTAGtAGGagGtTggttaggTTTAaAAA

>MDH_005
GATTGTCATATACCACAGCATTTTGTTATGGTAACCAATTCCATAGACTGTCCCACTTGTCCA
aTGCTAATAAATCTTTTTCATTTTTCTCCTGCTGGTTTTGCGaTTCTTCAAGCAAATGGTATATT
GTTTGTGTGTAATTACTAATTTCCTTATCCCACTGCATCCAGGTCGTGCGATTCCAAATTTCT
GTTTGATTTTTATTACTCCAACTGGAGTTCCAAGGCACACTAGTGGTGCAGATGAGTTTTCCA
GAGCAGCCCCAAAGTCCTAGGAGCTGTTGATCCTCTAGGTATCTTTCTATAGCCAGGATTCTT
GCCTGGAGCTGCTTAATGCCCCAGACTGTGAGTTGCAACATATGCTGTTggCCTCTATAGCTC
TCAGCAAATTGCTTTGCTGTTGCACTATACCAGACAaTAaTTgTAAtGa

>MDH_006

TgaTTTCaTtTaCACAGCATTTTGTTATGCTAAACcAATTCCACAAACTTGCCCATTTGTCTAATT
tgCAATAAtTCTTGTTCATTCTTTTCTTGCTGGATTTGCGAGTCTTCAATTAGGtCATATATTACA
CTTGTGTACTTGTCAATTTCTTTTTCCcACTCCATCcAGGTCATATTAttcCAAATgTacTCTAGAG
ATCTATTACTCcAGCTAGTGTTCCAGGGCACATTAGTGGTGCAAATGTGTTTTCCAGAGCAAC
cccTAATTTcTTAggAAGCtGTTGatcCCTTTAgGGTAGtTTTTcCtaCCgcCagGGAtCTCTTtGcCTGgag
CTTGtTTAATTGCCCCtAGACTGtGgAgtTGcAAcAtAtTgTtGgTgcgCCtccTaTTGCCCtTcagCAAAtT
TGtTTTgcTtgTTGCActatacCAgGcAaataatAtgTaTgGGAGGGGGGaGG

>MDH_007
CACTCTCTATATACCACAGCCATTGTGTTATGTCAACCAACTCCACAGACTGTcCCACTTGTC
CAAgGCCAATAAATCTTGTTCATTCTTTTCCTGCTGGTTCTGCgATTCTTCAATTAAgCTATAT
ATTaTGTgTGTGTAATTGCTAATTTCcTTATCCCATTGCAACCAGGTCATGTTGTCCCATATGgC
ATTGTAAGATTTATTACTCCAACTAGCGTTCCAGGGCACCGTAGTGGTGCAGATGAGTTTTC
CAgAGCAACCCCAAATTCCTAGGAGCTGTTGATCCCTcAGGTATCTTTCCACAGCCAGGACTC
TTGCCtGGAGCTGTTTAATGCCCCAGACCGTGAGTTTCAACAGATGTTGTTGAGCCTCTATAG
CCCTCAGCAAATTGCTTTGCTGTTGCACTATACCAGCAAtTAATTAGTAGGTG

>MDH_008
GaTTtCtTATACCACAGCATCTTGATATGTCAACCAATTCCACAGATTTGCCCACTGGTCaAAtG
CTAGTAAGTCTTTTTCATTGTTTTCCTGCTGGTCcTGcGCtTTTTCAATTAAGTCATATATCACTt
CTGTGTAACTGCTAATTTCTTTtTCCCATTGCAACCAGGTCATGTTCcCCCATATTTCCcCATGA
GATtTATTACTCCAACTAGAGTTCCAAGGCACAGTAGTGGTGCAGATGAGTTTTCCAGAGCA
GCCCCAAATTCCTAGGAGCTGTTGATCCCGtAGGTACCTTTCCACAGCCAGGACTCTTGCCTG
GAGCTGTTTAATGCCCCAGACCGTGAGtTTCAACAGATGTTGTTGAGCCTCTATAGCCTTCAG
CAAATTGcTTTGCTGTTGCACTATACCAGCAATaAaTTGTAaGGAGGGGGaG

>MDH_009
TtTTTTTcTgtgCcAtAGCATTTTGTTATGTCagAcCAGTTCcACAGACTTGCCCACTTGTCCaaTGCC
AATAAGTCCTTTTCATTCTTTTCCtGcTGGgtCTGCGATTCTTCAATTAGGTTATATATTATTTGT
GTGTAAtTgCTAaTTTCTTTATCCCATTGTaGccAGGTCATGTTATCCCAtATgTCACTCaGATTTT
TATTACTCCAACTAgaGTTCCAGGgCACAtTAGTGGGgCaGatGAGTTTTCCAGAGCAGCCCCAA
ATTCCTaGGAGCTGTTGATCCCTTAGGTATCTTTCtACAGCCAGGACTCGTGCCTGaGCTGTTT
AATGCCCcAgACCGTGAGTTTCAACAGATGTtGTtgAGCCTCTATAGCCCTCAGCaAATTGTTTt
GcTGTTGCACTAtaCcAGACAATaATTGTAATGGTAGtAGGagGGTgatagGttTAA

>MDH_010
CCTAGGACTTTGGGGCTGCTCTGGAAAACTCATCTGCACCAcTAGTGTGCCTTGgAACTCCAG
TTGGAGTAATAAAAATCAAACAGAAATTTGgAATCGCACGACCTGgATGCAGTGGGATAAGG
AAATTAGTAATTACaCACAAACAATATaCCATTTGCTTGAaGAATCGCAAAACCAGCAGGAG
AAAAATGAAAAAGATTTATTAGCATTGGACAAGTGGGACAGTCTATGGAATTGGTTTACCAT
AACAAAATGGCTGTGGTATATAAAAATATTCATAATgATAGTAGGA

>MDH_011
GGAAGTTAGCcAATACACAGACaCAATATATaaCCTAATTGAAGAATCaCAGAACCAGCAGGA
AAaGAATGAACAAGAcTTATTGGCATTGGAAAAGTGGGACAGTCTGTGGAATTGGTTTGATA
TAGTAAAGTGGCTGTGGTATATAAAAATATTTATAATGATAGTaGGAG

>MDH_012
GGAAaACTCATCTGCACCACTACTGTGCCCTGgAACTCTAGTTGGAGTAATAAaTCTtACaGtGA
GATaTGggATAaCATGaCCTGgTTGgAATGGGATAAAGAAATTAACAATTACACAAAccTaATaTa
TAGGCTAaTTGAaGAATCgCAGAaCCAGCAGGAAAAGAATGAACAAGACTTATTGGCATTAG
ACAAGTGGgCAGATCTGTGGAATTGGTTTAGCATCTCAAAATGgCTGT

>TDH_001

CTGGCTGTGGAAAGATACCTAAGGGATCAACAGCTCCTAGGAATTTGGGGCTGCTCTGGAA
AACtcaaccTcCcCCacTACTGTGCCctGGAACTCCAGTTGGAGTAATAAAtCCtAtGATGAAATATG
GGGGAACATGACCTggCTACaATGGGATAAAGAAGTTcACAATTACaCAGACATAAtATATAG
ACTAAtTgAAGAATCACAGAACCAGCAGGAaCAGAATGAAAAAGACTTATTGGCATTgGACA
AGTGGGCAGGTCTGTGGAATTGGTTTGAaATATCAAAtTGgCTGTGgTATATAAGAATATTCAT
AaT

>TDH_002
atatAccagTAaTTAAttGTaCaGTCCCAGATtGCCCTTgCGaGGGGGgAggAGcTTgTTCaTtttCtCCtGCT
GtTCTgGAatCgGCgagTAGATTATATatgATAcCTgtgtAGTattAaTTTCTTTAtCCCAtTGtgGcCAGGT
CaTTTGgcCcAgtAtCtccTtCaaGGATTTGTTAcTCCAACTAgaGTtccaGGGCaCagTAGtGGtgCagAtgA
aTTTtCCAAGAGCAGCCCCAAATTCCTAGgAtCTGTTGAtCCCTTAGctATCTTTCCACAgcCAggA
CTCtTGCCTGGaGCTGTTTAATGCCCCAgACCGTGAGTTTCaACAGTTGCtGttAGCCTCTATAGC
CCTCAgaAAATTATTTTGCTGgTgCtcTATTcCAGAAatAAtTttA

>TDH_003
ATCCGTTCCACAGAtTTGCCCATTTgTCCAGGGCagTAAGGCTTgtTCATtCTTTtCcTgCTGgtTCtG
CGAAtCTgCAagTAaAcTATATaTtAGCtCTGTgTAAtTgcTAATTTCTTTATCCCAtTGCAGCCAGGT
cATgTTgTCcCATATcTCaTtCAggGATTTATTaCTCCAACTAGAaTtCCaGGGCacaGTAGgGGtgCAg
AtGagTTTTCCAGAGCAgCCcCAAATTCCTAGGAGCTGTTGATCCacTAGGtATCTTTCCACAGCC
AGgACTcTTGCCTGGAGCTGTTTAATGCCCCAGACCGTGAGTTTCAACAgAtGCTGtTGAgCCTC
TATAGCCCTcAGcAAAtTGCTCTGCtGtgGCaCTaTagCaaAAaAAAAAAA

>TDH_004
AAATTtGccATATACCACAGCGAaTTTGaTATGTtAACCAACTCCACAGAtTTGCCCAcTTGTCcA
ATGCCAATAAGTCTTTtTCATTAATTTCCTGCTGGGTCTgCgAtTcTTCAATTAGTcTATATATTA
TGTCTGTGTAATTGCTAATTTCCTTATCCCATTGCAGCCAGGTCATgTTaTcCCATATcTCTcCC
TgAGATTTATTACTCCAACTAGAGTTCCAaGGCACATTAGTGGTGCAGATGAGTTTTCCAGAG
CAGCCCCAAATTCCTAGGAGCTGTTGATCCtTTAGGTATCTTTCCACAGCCAGGACTCTTGCC
TGGAGCTGTTTAATGCCCCAGACCGTGAGTTTCAACAGATGTTGTTGAGCCTCTATAGCCCTC
AGCAAATTGCTTTGTGTTGACTATACCAGCAATAATTGTAAGGAGTAGGaGGGCGGAGGCCC
TC

>TDH_005
GGATCTTTTAtTATGCaaGCTAtTGTGAtATGCTAACCAaTTCCACAAATTTGtCCACTTGTCCaAT
TCCaATAATTCTTTTTCATTCTTTTCTTGCTGgatCtGcgATTCTTCAATTAAGtTATATATAAaGCC
TGTGTAATTATcAAtTTCTcTTTCcCACTgCATCcAGGTCATGTTCTGCCAAATGTaATCTAaAGA
TTTATTACTCCAACTAGAaTTCCAGGGcACATTAGTGGTGCAAATGTGTCTTCCATAGCAACC
CCAAAGTCCTAGGAGCTGTTGATCCTTTAGGTATCTTTCCACAGCCAAAAtTCTTGCCCGGAG
CTGTTTAATGCCCCAGACCGTGAGTTGCAACAGATGTTGTTGCGCCTCTATAGCCCTCAGCA
AATTATTTTGCTGTTGCaCTATAcCAGAaAATAATTGTaTAAGGGCG

>TDH_006
GATgcTCATATACCACAGCATTgTGTTATGcTAACCAATTCCACAAACTTGCCCATTcGTCCAtT
TcCAATAGGTCTtGTTCATTCTTTTCTTGCTGGGTTTGCGATTCTTCAATTAAGCTGTATATaAA
CCCTGTGTAATTCTCAATTTCTCTTTCCCACTCTCTCCAGGTCATGTTACTATTCCAAATCTCC
GTGaAAGATTTATTACTCCAGcTAGAGTTCCAGGGCACATTAGTGGTGCAAATaTGTTTTCCA
GAGCAACCCCAgATTCCTAGGAGCTGTTGATCCTTTAGGTAGTTTTCCACAGCCAGGACTCTT
GCCTGGAGCTGTTTAATGCCCCAGACTGTGAGTTGCAACAGATGCTGTTGcGCCTCTATAGCC
CTCAGCAAATTATTTTGCTGTTGCACTATACCAGAcAAATAATTGTTAaGtAAaGGGGGcTT

>TDH_007
AAAtCtTGTTcATTcTttTCTTGctgtCCTGGAGAatgaggagTTaaGCTGTATaTtAACTCTGTGtAAttgtcaa
TTTCtCTTTCCcattCcagCcAGgttaTgTtATtCcAAAtaTCtcTtAGAGATTTATtACtCCAGCTaGAGTTC
CAAGGcACaTTAGtGGTGcAAATgTGTCTtCCAGAGCAACCCCAAAGtCctAGgAGCtGTTGatCCT

TtAGcATCTtTCCACAGCCAGgAtTCTTGCCTgGaGCTGTTTAACGCCCCAGaCTGTGAGTtGCaAC
AGATGTTGtTGgCCtCTATAGCCCTCAGCAAATTgTTttGCTGTtGCACTATACCAGACAATTaTT
GTA

>TDH_008
CCTAAGGGATCAACAGCTCCTAGGGATTTGGGGCTGCTCTGGAAAActCATCTGCaCCACTGC
TGTAAGTTGGAACTCTAGTTGGAGTAATAAATCATATGAGGAAATTTGGGGAAACATGACAT
GGATgCAgTGGGAGAAAGAAATTGATAATTATACAAACACAATATACAACTTGCTTGAAGAA
TCGCAAACCCAGCAGGAAAAGAATGAACaAGATTTACTAGCATTGGACAAATGGCAAAATT
TGTGGAATTGGTTTAACATATCAAATTGGCTATGGTATATAAaGATATTCATA

>TDH_009
CAAGAGTCCTGGCTGTGGAAAGATACCTAAGGgATCAACagcTTCTAGGAATTTGGGGCTGCT
CTGGAAAACTCaTCTGCACCaCTAATGTGCCCtGGAACTCTAGTTGGAGTAaTAAatCTTTCAAT
gAAATATGGGATAACATGACCTGgCTACaATGGGATAAAGAAATTAAtAATTACaCAGATATA
ATATATAgACTAATTGAAGAATCGCAGAACCaGCAGGAAAAGAATGaaCAAgccTTATTGGCAT
TGGACAAGTGGcCaAATCTGTGGAATTGGTTTGAAATATCAAAATGgTTGTGGTAtATAAAAA
TATTTATAATGATAGTAGgAGG

>TDH_010
TAATGTGCCtTGGAACTCTAGTTGGAGTAATAAatCTCTGaATGAGATaTGGGATAACATGACC
TGgCTGCAATGGGATAAAGAAATTAGCAaTTACACACAAATAATATATGAatTAATTGAaGAG
TCGCAGAACCAGCaGGAAAAGAATGAACAAGACTTATTGGCATTGGACAAGTGGgCAAATCT
GTGGAATTGGTTTgATATAACAAACTGgCTGTGGTACATAAGAATATTTATAAtG

>TDH_011
TGTTGAAGCTCACGGTCTGGGGCATTAAACAgCTCCAGGCAAGAGTCCTGGCTGTGGAAAGT
TACCTAAGGgATCAACAGCTCCTAGGGATTTGGGGCTGCTCTGGAAAACTCATCTGCACCAC
TAATGTGCCCTGGAACTCCAGTTGGAGTAATAAAACATATgAGGAgATATGGGAAAACATGA
CCTGGTTACaATGGGATAAAGAAATTAGCAATTACaCAGGCATAATATATAATCTACTTgAAA
TATCGCAAAACCAGCAGGAAaAGAATGAaCAAGACTTATtGgCATTaGACAAGTGGGCAAATC
TGTGGAATTGGTTTGATATATCAAAATGgCTGTGGTATATAAAAATATTCATAATGATAGtAG
GAG

>TDH_012
aaGAGTCCTgGCTGTGGAAAGATACCTAAaGGAtCatCAGCTCCTAGGAATtTGGGGaTGCTCTG
GAAAAcTCaTCTGCACCACTAATGTGCCCTGGAACtcTAGtTGGAGtAaTAAATCTCAgAGTGAG
aTATGGGACAACATGACCTGgCTgCAATGGGATAaAGAaATTAGCAATTACaCACAAAcAATAT
AtaAcCTAATTGAAaAATCACAGAACCAGCAGGAAAAGAATGAaCAAGAtTTATtGGCATTGGA
CAAATGGGCAAGTCTGTGGAATTGGTTTGACATATCAAACTGgCTGTGGTATATAAaAATATT
tATAATGATAGTAGGAGGCTTTGAGGGATCCAA

>KDH_001
CTTTGTtaTACCTCAGCCATTTCTGTTATGTCAACCAATTCCACAAACTTGCCCAACTGTCCAA
TttTcAATAaCTCTTGTTCATTCTTTTCTTGCTGGAGTTGCgATtcTTCAATTAAGCTGTATATTAA
gCCTGTGTAATTGTCAATTTCTCTTTCCCACTCCATCCAGGTCATGTTaTcCCAAATtTCATTTAc
AGATtTATTACTCCAGCTAGAATTCCAAGGCACAGAAGTGGTGCAAATGTGTTTTCCAGAGC
AACCCCAAATTCCTAGGAGCTGTTGATCCCTTAGGTATCTTTCCACAGCCAGGACTCTTGcCT
TGAGCTGTTTAATGCCCCAGACTGTGAGTTGCAACAAATGCTGTTGCGCCTCTATAGCCCTC
AGCAAATTGTTTTGCTGTTGCACTATACcAGCAAaTaAATtGTAATGtTAGTaGGaGGCTCatAAg
GGTTAaTaCAgGGGGGGGGGG

>KDH_002
CtCTCTTATATACCACAGCCAtTTTGTTATGTCAACCAAcTCCACAGACtTGCCCACTTGTCTaA
TgACAATAAGTCTTGTTCATTCTTTTcctGCTGGTTcTgCGaTTcTTCAAgTAGtCTATATATTaTGC

CTGTGTAATTgTCAATTTCTTTATCCCATTGCAGCCAGGTCATGTTTcCCCAtATTTcTtcCATAga
TTTATTACTCCAACTCTCATTCCAGGGTACATTAGTGGTGCAGATGAGTTTTCCAGAGCAaCC
CCAAATTCCTAGGAGCTGTTGATCCTGTAGGTATCTTTCCACAGCCAGGACTCTTGCCCGGA
GCTGTTTAATGCCCCAGACCGTGAGTTTCAACAGATGTTGTTGAGCCTCTATAGCCCTCAGC
AAATTGCTTTGCTGTTGCACTATAccAGAAAtAAATTgTAAAaTGAttaGgaGG

>KDH_003
TATTTTTcATATACCACAGCAATTTGATATGTtAACCAAtTCCACAGATTGTTCCAACTGTCCaA
TGCTAATAAATCTTTTTCATTTTTCTCTTGCTGGATTTGCgAGTCTTCAAGCAAAgTGTAGATT
ATGCTTGTGtAAtTaCTAATTTCTCTATCCCACTGCATCCAGGTCATGTTATCCCAAATTTCTTC
TTGAGATTTATTACTCCAACTaGtGTTCCAAGGCACAGCaGTGGTGCAGATGAGTTTTCCAGA
GCAGCCCCAAATCCCTAGGAGCTGTTGATCCTTTAGGTATCTTTCTATAGCCAGGACTCTTGC
CTGGAGCTGCTTAATGCCCCAGACTGTGAGTTGCAACATATGCtGTTGCGCCTCTATGGCTCT
CAGCAAATTGCTTTGCTGTTGCACTATAcCAGACAAaTAATTgtAATGATAGTAGGAGGCtTGA
TAagGtttAAAAaGGGGGGG

>KDH_004
GaCaaTTCaTaTaCACAGCCATTTTGTTATGTTAACCAATTCCACAGACTTGCCCACTGGTCCAac
TgCAgtAGGTCTTGTTCATTTTTGTCTTGCTGGAtTTGCGACTCTTCAATTAAGgTGTATATTgTA
ccTGTGTAATTGTCAATtTCTCTTTCCCAcTCCAtCCAGGTCATGTTATTCCAAATCTCTTcTAGA
gATTTATTCTTACTCCAGCTAGAGTTCCAgGGCACATTAGTGGTGCAAATGTGTTTTCCAgAGC
AACCCCAAATTCCTAGGAGCTGTTGATCCTTTAGGTAGCTTTCCACAGCCAGGACTCTTGCCT
GGAGCTGTTTAATGCCCCAGACTGTGAGTTGCAACAAATGcTGTTGCGCCTCTATAGCCCTCA
GCAAGTTGTCTTGCTGTTGCACTATAccAgCAAAatAaTtgTTAA

>KDH_005
GGCAAGAgTCCTGGcTGTGGAAAGCTACCTAAGGGATCAACAGCTCCTAGGAATTTGGGGTT
GCtCTGGAAAaCACATTTGCACCACTAATGTGCCCTGGAACAcTAGCTGGAGTAACAAATCTC
TAgACAAGATTTGGAATAATATGACCTGGATGGaGTGGGAAAggGAAATTGAAAATTACACA
GaCTTAATATATAAcTTAATtgAAGAATCGCAAACCCaGCAAGAAAAGAATGAACAAGAATTA
TTGCAATTgGACcAATGGGCAAGTTTGTGGAATTGGTTTAGCATAaCACAATGGcTGTGGTAT
ATAAAAATATTCATAATGATAGTaGGA

>KDH_006
CTGCACCACTAATGTGCCCTGgaACTCTAGTTGGAGTAATAAaTCaTTCAATGaCATaTGggaCA
ACATGACCTGgCTGCaATGGGATAAAGAGGTTGGcAATTACACACaAATaATATATGGCCTGaT
TGAaGAATCGCAGAaCCAGCAGgAAAAGAATGAACAAGACTTATTGGCATTGGACAAGTGGG
CAAGTCTGTGGAATTGGTTTGACATATCA

>KDH_007
AGAATCCTGGcTGTGGAAAGATACCTAAGGgATCAACAGCTCCTAGGAATTTGGGGCTGCtCT
GGAAGAcTCATCTgcACCACTAATGTGCCTTGGAACTCTAGTTGGAGTAATAATATAACTCTA
GAGGACATATGGGAAAACATGACCTGGCAACaGTGGGAAAGAGAAATTAGCAATTACaCAG
ACATAATATATAACCTAATTgAAGAATCGCAGAACCAGCAGGAGAAAAATGAACAAGATTT
ATTGGCATTAGACAAATGGACAAATCTGTGGAATTGGTTTGACATATCAAGATGgCTGTGgT
ATATAAAAATATTCATAATGA

>KDH_008
CTTTcTTATATACACAGCATcGTGATATGTCAACCAATTCCACAAACTTGCCCATTTATCCAAT
TgCAATAGTTCtTGTTCATTCTTTTCtTGCTGGGTTTGCGATTCTTCAATTAGAcTGTATATTAAa
CCTGTGTAGTTGTCAATTTCtTTTTCCCACTgCATCCAGGTCATGTTaTCCCAAATCTCCTGTAT
AGATTTATTACTCCAGCTAGAGTTCCAGGGCACTGTaGTGGtgcAAATGTGTCTTCCAgAGCAA

CCCCAAAGTCCTAGGAGCTGTTGATCCTTTAGGTATcTTTCCACCGCCaGGACTCTcGCCTGGA
GctGTTTAATGCCCCAGACTGTGAGTTGCAACAgATGTtGTtGCGCcTCTATaGCCCTCAGCAAA
TTGTTTTGCTGTTGCACTATACCAGAAAaTAATTGTTtgGAGgGGGGGgttGtGAtaTAAAA

>KDH_009
GTTtGaaTCCACAGCCACTTctGaTATctaAAACCAATTTCCACAGATTTGCCCACTTGTCCAATGC
CAATAAGTCtTGtTCATTCTTTTCCTGCTGGTTcTGcGATTcTTCTATTAGGGTATATATTAGTTG
TGTGTAATTGttAATTTCCTTATCcCATTGCATCCAGGTCATGtTAtCCCAtATCTCcttATAAcTTTT
GTTACTCCaACtAgAgTTCCaGGGCaCatTAgTGGTGCagAtGAgTTTTCCagAaCaaCCCcaAATTcCT
AgGAaCtgtTGATCCCTTagGTaTctTTcCaaAGCCAGGACTCTTGCCTGGAGCTGTTTAATGCCCCA
GACcGTGAGTTTCAACATATGTtGTTGAGCCTCTaTAGCCCTCAccAAATTgCTTTGCTGTtGCaC
TATAccaAAAAaaAAattggTAtAACCCAAATGG
>KDH_010
TTTGTCGtTTatACCACAGCAaTtTGATATGttAGACCAATTCCACAGATTTGCCCACTTGTCCAA
TGCcAACAAaTCTTGTtCATTTTTTTCcTGCtGGtTCTGTgATTCTTCCAGTAGCAAAtATaTTATgT
tTGTGtAAtTgCTAATTTCTTTATCCCATTgCAGCCAGgTCATgTTCTCCCATATCtGTTCCTGAGA
TTtATTACTCCaACTAGAGTTCCAGGgCACATTAGTGGTGCAGATGAGTTTTCCAGAGCAACC
CCAAATTCCTAGGAGCTGTTGATCTTTTAGgtATCTTTCCACAGCAAGgACTCTTGCCTGgAGC
TGTTtAATGCCCCAGACcGtGAGTTTCAACAaATGCtGTtGAGCCTCTATAGCCCTCAGCAAATtG
CTTTGCtGTTGCaCTaTaCCAGACAATAATTGTAATGATAGTaGGAGGCTTGAtAaTgTtTAAAaG
G

>KDH_011
CtgcTTtCCattgTaCACAGCATTTGTTATGTCAACCAATTCCACAgAtTTTGCCAcTTGTCCAATTC
TAATAAtTcCTTTTCATTTTcTTCCTGCTGGaTTTGCGATTcTcCAAGCAACTGGTATATTaTGTcT
GTGTAATTtTcAATTTCCcTATCCCACTGCATCCAGGTCATGTTACcCCAAATcTACcCTaGAGAT
TTATTACTCCAACTAGAGTTCCaAgGcaCaTTAGtGGtgaGATGAGTTTTCCaGaGCAGCCCCAAA
TCCCTAGGAgCTGTTGaTCCTGTAGGTaTCTTTCTATAGCCAGGACTCTTGCCTGaAGCTGCTaA
TGCCCAgACGtGaGTGCACATATGCTGTTCgCCTCTaTAGTCTCAGAATTGTcTgTGTGaCTAtAC
AaAA

>KDH_012
TGTGCCCtGgAACTCTAGTTGGAGtAATAAAtCTTTTGAGgaCATaTggaACAACATGACATGcCtG
CaATGGgATAAAGAAGTTAGCAATTACACAGACTACATATATCAACTAATTGGGGAATCGCA
AATCCaGCAGGAAAAGAATGAACAAgAgTTATTGGCATTtGACAAATGGcCAAGTcTGTGGA

>KGDH_001
CATCTGTTGCAACTCACGGTCTGGGGCATTAAACAgTtCCGGGCAAGAaTTTTGGCTGTGGAA
AGATaCCTAAAGcaccaaCagCtCCtAGgAcTTTGGGGTTGCtATGgAAgACACaTTTGCACCAcTAAT
GTGCCCtGgAAtTCTAGTTGGAGTAATAAatCTgTAGATtACATTTGgCAgAACATGACCTGGAtGc
aGtGGgAAAGAGAAATTGATAaTTACACAGGCcTTATaTaTAACTTAaTTGAAGAATCgCAGAtC
CaGCAAGAAAAGAATGAAAAAGAATTATTGGAATTGGACAAGTGGACAAATTTGTGGAATT
GGTTTAaCATATCACAAtGgCTGTCATATATAAAAATATTCATAAtGATAGTaGGA

>KGDH_002
TGCTGTGGAAAGATACCTAAAAGatCtACAGCtCCTAGGAATTTGGGGTTGCTCTGGAAAACTC
aTCTGCaCCACTAATGTGCCCtGGAACTCTAGTTGGAGTAATAAATCTCAGGAACAGATATGG
gAgAACATGACCTGGctGcaAtGGGATAAAGAAATTAGCAATTACaCAaAcATAATATATTTGCT
ACtGGAAGAaTCACAGAACCaGCAgGAAAAAAATGAACAAGATTTGTTGGCATTGGACAAGT
GGgCAAATCTGTGGAATTGGTTTaACATATCAAACTGGCTGTGGTATATAAGAATATTCATAA
TGATAGTAGGAGGCTT

>KGDH_003
TATGTTGcACTCACAGTTTGGGGCATTAAACAGCTCCAGGcAAGAATCCTGGCTGTGGAAAG
CTACCTAAGGGATCAaCaGCTCCTAGGAGTTTGGGGTTGCtCTGGAAAACACATTTGCaCCACT

AATGTGCCTTGGAACTCTAGCTGGAGTAAtAAATCCCTAGATgAGATTTGGGAGAACATGAC
CTGgATGGATTGGGAAAAAGAAATTGCCAaTTACACAGGGTTAATATATGAATTAATTGaAG
AATCGCAAACCCAGCAAGAAAAGAATGAAAAAgAACTATTGGAATtGGACAAATGGgCaAGT
TTGTGGAATTGG

>KGDH_004
ACTCACAGTCTGGGGCATTAAACAGTtCCAGGCAAGAGTCCTGGCTGTGGAAAGATACCTAA
GGgatCAACAGCTCCTAGGAATTTGGGGCTGCTCTGGAAAACTCATCTGCCCCACTACTGTGC
CCTGGAACTCTAGTTGGAGTAATAAAaCCCAGaAGGAGATATGGGAAAACATGACCTGgCTA
CAATGGGATAGAGAAATTGACAATTACACACAGATAATATATAACCTACTTGAAGAaTCGCA
GAACCAGCAGGAAAAGAATGAACAAGATTTATTGGCATTGGACAAGTGGgCAAGCCTgTGG
AATTGGTTTGACATATCAAAATGGCTGTGGTATATAAgAATaTTTATAATGATAGTAGgAGG

>KGDH_005
TGGAAAACTCATCTGCACCACTAATGTGCCCTGgAACTCTAGTTGGAGtAATAAATCTCAGGA
gGAGATAtGGGACAACATGACCTGgCTaCAATGGGATAAAGAAATTAACAATTACaCAGACaT
AATATaTAAACTACTTGAAGAATCGCAGAACCAGCaGGACAaGAATgAACAAGACTTATtGGC
ACTGGACAAGTGGGCAAGTTTGTGGAATTGGTTTAGCaTAACAAATTGgCTGTGGTATATAA
AAATATTTATAATGATAgTAG

>KGDH_006
GGACCACAGCTAtCtTGcTATGTcgTaCAagTCCACAGACTtGCaCACTTGTCCAATGCTAATAAG
gCTTgTTCATTTTTCTCCtGCTGGtTCTGCgATTCTTCAAGTAGAtTATATATtATATtTGTGTAGTT
GgTAATTTCTTTATCCcaTTGTAGCCAGGTCATGTTGTCCCAtATCTCAcTCtgGGATTTGTTACT
CCAAcTAGAGTTCCAGGGCACATTAGTGGTGCAGATGARTTTTCCAGAGCAGCCCCAAATTC
YTAGGAKCKGTTGATCCCTTAGGTATCTTTCCACAGCCAGGACTCTcGCCTGGAGCTGTTTAA
TGCCCCAGACtGTGAGTTTCAACAgTTGCTGTTGAGCCTCTATAGCCCTCAGCAAaTTatTTGc
TGTTGCACTATACcAGAcAaTaAttaGAAAtGgTAGtAGGAGGctTgG

>KGDH_007
TTTTTTTTTaatACCaCaGCATtTTGAATATCAgACCAATTCtAcAGATTTGCCCACTTGTcTAATGc
AGaAGtgTTGTTCATTCTtTTCCTGCTGGTTTTGCGATATTTCAAGtAGATTATATATTATGCCTG
TGTAATTGCTAATTTCTTTATCCCATTGTAACCAGGTCATGTTTTCCCATATcTCCTCATATGT
TTTATTACTCcAAcTGGAGTTCcaGGGCACATTAGtGGTGCAGATGAGTTTTCCAGAGCAGCCC
CAAATCCCTAGGAGCTGTTGATCCCTTAGGTAACTTTCCAcAGCCAGGACTcTTGcCTGGAGcT
GtTTaAtGCCCCaGAccGtGAGCtTcAACATCTGCtGttGAGCTTcTATaGCCCTCAGctAAtTGCTTTGc
TGTTGcaCTaTaCCAgTcaTtTaAtTGTTAtGGTaGaAGGgGGCtTGg

>KGDH_008
ATCCTGaTtTcCACAgCTaGTTTGATaTgTCAACCaaTTCCACAGACTTGCCCACTTGTCCaaTgCcta
TAAGtCTTGTtCATTCTTTTCCTGCtGgTTCTgCgATTCCTCAATTAGCCTATATATTAgTTCTGTG
TAATTGTTAATTTCTTTATCCCATTGCTGCcAGGTCATGTTATCCCATATGTCATTATAAGATT
TATTACTCCAACTGGAGTTCCAGGGCACATTAGTGGTGCAGATGAGTTTTCCAGAGCAGCCC
CAAATTCCTAGGAGCTGTTGATCCCTTAGGTATCTTTCCACAGCCAGGACTCTTGCCTGGAGC
TGTTTAATGCCCCAGACCGTGAGTCTCAACAAATGTTGTTGAGCCTCTATAGCCCTCAGCAA
ATTGCTTTGCTGTTGCACTATACCAGaaATTaATTGTAGGgAGtAGGGGCTT

>KGDH_009
TcTTTTTTAAtACCACAGCCACTTTGATATATCAACCAATTCCACAGATTTTCCTACTTGTCCA
ATGCAAGgAGTCttGTTCATTCTTTTCCTGCTGGTTCTGCgATTCCTCAATTAgTCCATATATTAT
GTTTGTGTAATTGCTAATTTCTTTCTCCCATTGCAGCCAGGTCATATTCCCCCATATCTCACTT
TGAGATTTATTACTCCAACTAGAGTTCCaGGGCACATtAgtGGTGCAgAtGAgTTTCCtAgAgCAg
CCCCaAatCCCTAgGAGCtGTTGATCCTTTAgGTATCtTTCCACAgcCAGGATTCTTGCCTGGAGC

TGTTTAATGCCCCAgACTGTGAgTTTCAACaGATGTTGTTGAGCCTCTAtAGCCCTCAgcAaATT
GCTTTGctGTTGCactATACCaGAcaATTaTTgTAGTA

>KGDH_010
cTCTAGTTGGAGTAATAAATCCCAGggTgAGATATGGaAAAACaTGACCTGGaTGCaATGGGAtA
AAGAAATTAaCAATTACACACAGtTAATATATaAtCTAcTTGAAGAATCGCAGATCCaGCAGGA
AAAGAATGAACAAGACTTATTGGCATTTGACAAGTGGgCaAATtTGTGGACTTGGTTTGACAT
ATCAAACTGgCTGTGGTATATAAAAATATTTATAAT

>KGDH_011
CTAaTGTGCCCTGGAACTCTAGTTGGAgtAaTAAATCAAAGgAAGATaTATGGgACAACATGAC
CTGGTTGCAATGGGATAAAGAAATTCACAATTACaCAGAAACAATATAtAaATTAATTGAAGA
ATCgCAGAaCCAGCAGGAAAAGAATGAaCAaGACTTATTGgCATTGGACAAGTGGGCaAATCT
GTGGAATTGgTTTGATATAACaAACTGgCTGTGG

>KGDH_012
CTATAGAAAGATACCTAAAGgATCAACAGCTCCTAGGGATTTGGGGCTGCTCTGGAAAACTC
ATCTGCACCACTGCTGTGCCTTGGAACTCcAGTTGGAGTAATAAATCTCAAGAAGAAATTTG
GGATAACATGACCTGGATGCaGTGGGATAGAGAAATTAGTAATTACaCAAACAcAATcTACAg
gTTGCTTGAaGACTCGCAAATCCaGCAAGAgAAAAATGAAAAAgATTTATTAGCATTGGACAG
TTGGAAAAATCTGTGGAcTTGGTTTgACATATCAAATTGGCTGTGGTATATAAAAATATTCAT
AATG

>KMH_001
GAATcTGTAtATACTCAGCAGTTTGTTATGTCAACCAATTCCCACAGATTTGCCCACTTGTCCA
ATGCCaaTAAGTCTTGTTCATTCTTTTCCTGCTGGTTCTGCGATTCTTCAATTAGccTATATATT
ATGtTTGTGTAATTGTCAATTTCTTTATCCCACTGCAGCCAGGTCATGTTATcCCATATTTCCTC
AAAAGATTTATTACTCCAACTAGAATTCCATGGCACATTAGTGGGGCAGATGAGTTTTCCAG
AGCAgCCCCAAATtCCTAGGAGCTGTTGATCCCTTAGGTATCTTTCCAGAGCCAGGACTCTTG
CCTGGAGCTGTTTAATGCCCCAGACCGTGAGTTTCAACAGATGCTGTTGAGCCTCTATAGCC
CTCAGCAAATTGCTTTGCTGTTGCACTATACCAGAAATTAATTGTTTGGAGTaGGAGGCttgGaa
GggCCTCAG

>KMH_002
ACTGATATaCCACaGCTATTGTGTTATGTCAAGCCAATTCCCACAGAcTTGCCCACTTTGtCCgA
TGCCAATAAgTCTtGTTCATTCTTTTCCTGCTGGTTCTGCGaTTCCTCaATTAGGgtATATATTGT
TTCTGTGTAATTGCtAATTTCTTTATCCCAtTGcAGCcATGTCATGTTGTCCCATATTTCACTCTG
AGATTTGTTACTCCAACTAGAGTTCCAGGGCACATTAGTgGGGCAGATGAgTTTTCCAGAGCA
gCCCCAAATTCCTAGGAGCTGTTGATCCTTTAGGTATCTTTCCACAGCCAGGAcTCTTGCCCG
GAGCTGTTTAATGCCCCAGACCGTGAGTCTCAACATATGTtGTTGAGCCTCTATAGCCCTCAG
CAAATTGcTTTgTgTTGACTATACcAaAAAttatTtgtAaGGGGGGaaGGGG

>KMH_003
TtgTTGTtaTAtACcACAGCaTTTTGATATGttAACCAATTCCACAAaTTTgCCCACTTGTCCAATGC
TAaaAGgCTTGTTCATTCTTTTCCTGCtGGTTCTGCgATTCTtCAAGCAAAtTaTATATTAATTGTG
TGTAATTGTCAATTTCTTTaTCCCaTTGCAACCAGGTCATGTTATCCCATATTTCATCaTGAGAT
TTATTACTCCAACtAGAGtTCCAGGGCACATTAGTGGtGCAGATgAGTTTTCCAGAGCAGCCCC
AAATTCCTaGGAGCTGTTGATCCCTTAGGTATCTcTCCAGAGCCAGGACTCgTGCCTGGAGCT
GTTTAATGCCCCAGACCGTGAGTTTCAACAGTTGTTGTTGAGCCTCTATAGCCCTCAGCAAAT
TGCTTtgtGttGCACTATACCAgcgaTTaATTgTAAGtAGTAGGaGGcTTGatAggTTTaAAtGG

>KMH_004
cTAATGTGCCTTGgAACTCTAGCTGGAGTAATAAAtCTCTAaAtgAcaTTTGGAATAACATGACC
TGGATGGAATGGGAAAGAGAAATTGACAatTaCACAGAGTTAATaTACAGcTTAATTGcAGAAT

CGCAAGGCCAGCAAGAAAAGAATGAACAAGAtCTATTgAAATTGGACCAATGGGCAAGTCTG
TGGAATTGGTTtAGCATATCAAATTGgCTGTGGTATATAAGA

>KMH_004
AGTCTGGGGCATTAAACAGCTCAAGGCAAGAGTCCTGGCTGTGGAAAGATACCTAAGGgAT
CAaCagCTCCTAGGAATTTGGGGTTGCtCTGGAAAACACATTTGCaCCACTTCTGTGCCTTGgAA
TTCTAGCTGGAGTAATAaatCTgTAAATgAaATTTGGaAtAACATGACCTGGATGGAGTGGGAAA
GAGAAATTGACAATTACaCaGGcTTAATATACAGCTTAATTGAAGtATCGCAaCTCCaGCAAGA
AAAGAATGAACAAGAGtTATTGAAATTGGACAGTTGGgCAAGTTTGTGgAATTGGTTTGACAT
AACAAAATGGCTGA

>KMH_005
tTGAAACTCACGGTCTGGGGCATTAAACAccTCCAGGCACGAGTCCTGGCTCTGGAAAGATAC
CTAaGGGATCAACAGCTCCTAGGAATTTGGGGaTGCTCTGGAAAACTCATCTGCACCACTAAT
GTGCCCTGGAAtaATAGTTGGAGTAATAAAAaCCtGACTGAcATATGGaACAACATGACCTGGT
TGCAATGGGATAAAGAAaTTAGTAATTACACACAatTaATATATGATCTACTTGAgGAaTCACA
GAACCaGCAGGAAAAGAATGAACAAGAcTTATTGGCATTGGATAAGTGGgCAAATCTGTGGA
ATTGGTTtaaCATATCAAAtTGgCTATGGTATATAAgGATATTTATAATGATAGtaGGAGG
>KMH_006
AAgttttgtttaCCACAgCaAGtTaGTTATgTaAAACCAATTCCATAGATTTGCCCATTTGTCCAATGGt
AAAAtCTTGTTCATTCTTTTCCTGCTGGTTCTGCGATTCTTCAgGtAaACTATATATTAGcTCTGT
GTAATTGTTAATTTCTTTATCCCATTGCAGCcaGGTCATGTTgTcCCATATCTCATTCAGGGATT
TATTACTCCAACTAGAATTCCAGGGCACAGTAGTGGTGCAGAtGAGTTTTCCAGAGCAGCCC
CAAATTCCTAGGCGCTGTTGATCCACTAGGTATCTTTCCACAGCCAGGACTCTTGCCTGGAG
CTGTTTAATGCCCCAGACCGtGAGTTTCAACAGATGCTGTTGAGCCTCTATAGCCCTCAGCAG
ATTGCTCTGCTGTTGCACTATAccaGAcaTTaaTTgT

>KMH_007
ATTCTTATATACCACAGCAGTTTGATATGTCaACCAAGTCCacAGATTTGCCCaTTTATCCATG
CCAGTaAGTCTTGTTCATTCTTTTCCTGCtGGTTCTGCgATTtTTCAATTAGGTcATATATTATgtC
TGTGTAATTGcCAATTTCtCTATCCCATTGCATCCAGGTCATaTTGTCCCAtATCTCATGCTGAG
ATTTATTACTCCAACTAttGTTCCAGGcCACATTAGTGgtgCAGATGAGCcTTCCAGAGCAGCCC
CAaATTCCTAGgAGCTGtTgATCCCtTAGGtATCTTTCcACAGCCAGGACTCgtGCCTcGAGCTgCT
TAATGCCCCAGACCGTGAGTCTcAtCAGcTGTTGTTgAGCCTCTATAGCCCTCAGCAAATTGcT
TTGCTGAtgCaCTATAcCAgAcAATaAaTaGTAaCTGGC

>KMH_008
CtttttGGgGgGttgCaCAgCaaAaTcatcATAGTtAAtCAattCCACAcATTcGCcCaCTTGtCCAATGGgtAAa
AGGgGtTGtTCATTCTttaCCTGCtGGtTCTGCgAACCtTCAAttAGCCTGTATATTATtTCTGTATAAt
TGtCAaTTTCCTCATCCCATTGCCTCCAGGTCaTGTTatTCCAtAtGTCATCcaGAGAtTTATTACTC
CAACTAGAATTCCAGGgCACATTAGtGGtGcAGATaAgtTTTCCAGAGCAGCCCCAAATTCCTAG
GAGCTGTTGATCCCTTAGGTATCTTTCCATCGCCaGGaCTcTtGCCTGGaGCTGTTTaATGCCCaA
GACTGtGAGTTTCAACAtAtGCtGtTgAGCCTcTaTaGcCCTCAGcAAATTGTTTTgcTgTtGCaCTATa
tTtgATTTTTTtaaTgT

>KMH_009
ATTTcTctGttTAcTGCAGTTTGATCTTCTGATCAAgTCtACAGACTTGCCCAcTTGTCCaAGCTAA
TAAGTCTTGTTCATTTTTCTCCTGCTGGTTCTGcGATTCTTCAATTAGAggaTATATTATCctTGT
GTAATTgCTAATTTCTTTATCCcATTGTAGCCAaGTCATGTTgTCCCATATTTCAcTCTGGGATT
TATTACTCCAACTAGAGTTCCAGGGCACAGTAGTGGTGCAGATGAGTtTTCCAGAGCAGCCC
CAAATTcCTAGGAGCTGTTGAtCCTTTAGGTAcCTTTcCaCAGCCAGGACTCTTGCCTGGAGTTG
TtTAATGCCCCAGAcCGTGAGGcTcAACAGATGCTGtTGACCtTcTtatagCCTTaGCaAgtTgGTTTTG
cTGTTGcaCTAataCaagAAcAtTaaTTGtaAAtgaTAGtAGGaGGcTTgaaaGgTTtaAAA

>KMH_010

AGAGTCCTGGCTGTGGAAAGATACCTAAGGGatCAACAGCTCCTAGGAATTTGGGGATGCTC
TGGAAAACTTATATgCCCCACTAATGTGCCTTGgAACTCTAGTTGGAGTAATAAATCTCTGCA
TGAGATATGGGATAACATGACCTGGCGGCAATGGGATAAAGAAATTAACAaTTaCaCAgAAA
TAATATATAACTTACTTGAAGAATCACAGAACCaGCAGGAAAAGAATGAACAAGACTTATTG
GCATTGGACAAGTGGgCAAGTCTGTGGACTTGGTTTGACATATCAAACTGGCTGTGGTATAT
AAAAATATTTATAATGATAGTAGGAGGCTT

>KMH_011
ACCTAAGTGATCAACAGCTCCTAGGAATTTGGGGCTGCTCTGGAAAACTCATCTGCACCACT
AATGTGCGCtGGAACTCTAGtTGGAGTAATAGATCCCTAGATGAGATATGGGAAAACATGAC
CTGGCTGCAATGGGATAAAGAAGTTAGCAATTACACAAATATAATATATAGTCTACTTGAAG
AATCGCAGAACCAGCAGGAAAAGAATGAaCaGGAcTTATTGGCATTGGACAAGTGGgCAAAT
CTGTGGAGTTGGTTCAACATAACAAACTGGCTGTGGTATATAAGAATATTTATAATGATAGT
AGGAGG

>KMH_012
GCAAGAGTCCTGGCTATAGAAAGATACCTAAaGGATCAACAGCTCCTAGGAATTTGGGGCTG
CTCTGGAAAaCTCATCTGccCCACTAATGTGCCCtGGAACTCTAGTTGGAGTAATAgAaCTCAGa
gTGAGATATGGgACaACATGACcTGGCTACAATGGGATAAAGAAATTAGCAATTACACAGATT
ACATATATCAACTAcTTGAaGACTCGCAGAACCaGCAGGAAAAGAATGAACAAGACTTATTG
GCATTtGACAAaTGGGCAAATCTGTGGAATTGGTTTGaTATATCaAATTGGCTGTGGTATATAA
gAATATTCATAATGATAGTaGGAG

>TMH_001
GaATTTTTTTAATACCACAGCCAtTTTGATATGTCAACCAATTCCACAAACTTGCCCAGcTGTC
CAATTtCAATAGTTCTTGTTCATTCTTTTCTTGCAGGGTtTGCgAaTCTTCAAGcAAGCTATATAT
TAAACCTGTGTAATTGTCAATTTCTCTTTCCCACTCCATCCAGGTCATGTTtCCCCAAATcTCTT
CTAcAGATCTATTACTCCAGCTAGAGTTCCAtcgCACAGcAGTGGAGCAAATgTGTcTTCCAGA
GCAACCCCAAATTCCTAGGAGCTGTTGATCCCTTAGGTAACTTTCCACAGCCAGGAcTCTTGC
CTGGAGCTGTTTAATGCCCCAgACTGtGAGTTGCAgCAGATGCTGTTGCGCCTCTATAGCCcTC
AGCAAATTGTTTTGCTGTTGCACTATACCAgacAAaTAATTGtAAAACG

>TMH_002
AAGGggcATaTACCACAGCAGtTTGATATGTCaACCAATTCCACAGAGATGCCCACTTGTCCAA
GGCTAATAAGTCTTGTTCATTCTTTTCCTGCTGGTTCTGCgATTGTTCAATTAGGTCATATATT
TTTTGTGTGtAAtTATTGATTTCTTTATCCCATTgCAGCCAGGTCATGTTGTCCCATATCTCATT
ATAAGATTTATTACTCCAACTAGAGTtCCAGGgCACATTAGTGGtGCAGATGAGTTTCCCAGA
GCAGCCCCAAATTCCTAAGAGCTGTTGATCCTTTAGGTATCTCTCCACAGCTAGGACTCTTGC
CTGGAGCTGTTTAATGCCCCAGACCGtGAGTTTCAACAGTTGTtGTTGAGCCTCTATAGCCCTC
AGCAAATTGTTTTGctGTTGCaCTATACCAGACAATAaTTGTAATgAtAGTaGGaGgcGTgagAAttTt
tAAAAAAGGGGGGGGGG

>TMH_003
GATTTTTAATACCACAGCCATTGTGTTATGCTAACCAATTCCACAGATTTTGCCACTTGTCCA
AtGtatTAAATCCTTTTCATTGTTCTCCTGCTGGATTTGCgAgTcTTCAAGCAAACTATATATTAT
TTTTGTGTAATTACTAATTTCTTTATCCCACTGCATCCAGGTCATGTTACCCCAAATATCCTCT
TGAGATTTATTACTCCAACTAGTGTTCCAAGGCACATTAGTGGTGCAGATGAGTTTTCCAGA
GCAGCCCCAAATCCCTAGGAGCTGTTGATCCCTTAGGTATCTTTCTATAGCCAGGACTCTTGT
CTGAAGCTGCTTAATGCCCCAGACTGTGAGTTGCAACATATGCTGTTGCGCCTCTATAGCCCT
CAGCAAATTGCTTTGCTGTTGCACTATACCAGAAAATAATTGTT

>TMH_004
TCCCGCTcATgtaCCAcaGCATacTgTTATGTcaAaCCAaCTccACAGACcTGcCCACtTGTCTaATGcA
AtAAtGCTTTTTCATTTATTTCCTGCTGGTTTTGCgatcTCTCaaTTAGGTCATATATTACTTCTGTG
TAATTgCTAATTTCTTTCTCCCATTGTAGCCAGGTCATGTTTCCCCATATCTCATCATGAGATT

TATTACTCCAACTAGAGTTCCAGGGCACAGTAGTGGTGCAGATGAGTTTTCCAGAGCAGCCC
CAAATTCCTAGGAGCTGTTGATCCTTTAGGtATCTTTCCACAGCAAGGACTCTTGCCTGGAGC
TGTTTAATGCCCCAGACCGTGAGTTTCAACAGATGTtGTTGAGCCTCTATAGCCCTCAGCAAA
TTGCTTTGcTGTTGCaCTATAccAgACAATAATTGTaATGaTAGtAGGaGGcTTGattaaggTTTaAA

>TMH_005
CAAGGcTcaTgGACtAcAGCAGAgtTGaATATaTaaAaCCAATTCCACAGATTTgCCCACTTGTCCaA
TGCCAATAAGtCTTGTTCATTCTTTTCCtGCtGGgTCTGcgtTTCTTCAAtTAATctatATATTGTTTCT
gTGtAAtTgtGAATTTCTTTATCCCaTTgCaACCAGgTCATGTTGTCCCaTATATCTtCCTTTGATTT
ATTACTCCAACTAGAGTTCCAGGGCACATTAGTGGtGCaGATGAgTcTTCCaGAgCAGCCCCAA
ATTCCTAGgAGCTGTtGATCCCTTAGGTATCTTTCCACAGCAAGGACTCGTGCCTGGAGCTGT
TTAATGCCCCaGAcCGTGAGTCTcAACATATGTTGtTGAGCCTCTATAGcCcTCAGCaAATTGCTt
TGcTGTtGCACTATaCcagAcAaTAATTGTGATGATAGtAGGaGTCTTGAttagtaTTTAAAAAGGGG
GGGGGGG

>TMH_006
AATGCGTtATATACCACAGCAGTTTGtTATGCcaAACCAATTCCATAAATTTGCCCACTTGTCC
AATtcCAATAAGTCTTTTTCATTCTTTTCCTGCTGGtTCTGCgaTTgTtCaatTAGGTTATATATTGT
TTGTgTGtAAtTgttAATTTCTTGGTCCCATTGCAGCCAGGTCATGTTATCCCATATcTCACTCAA
AGATTTATTACTCCAAGTAGAGTTCCAAGGCACATTAGTGGTGCAGATGAGCTTTCCAGAGC
AGCCCCAAATTCCTAGGAGCTGTTGATCTcTTAGGTAACTTTCCAGAGCcAGGACTCTTGCCT
GGAGCTGTTTAATGCCCCAGACTGTGAGTCTCaaCAGCTGTTGTTGAGCCTCTATAGCCCTCA
GCAAATTGCTTTGCTGTTGCACTATACCAgACAATAATTgTAATGAtAGTAGGAGGCTTGAttAt
TgTTtAAA

>TMH_007
TAatGgTTAgTACCtaGCTATtTGATATGTTAaCCAATTCCACAAATTTTGCCATTTGTCCAATGCT
AGtAAAtCTTGTTCATTCTTTTCcTGCTGGGTTTGCGATTCTTCAAGCAAGTTGTATATTGTGTT
TGTATAATTATCAATTTCTTTCTCCCAcTGCATCCATGTCATGTTTCCCCAAATTTCCTCATAT
GATTTATTACTCCAACTAGAGTTCCAACTTACAGCAGTGGtGCAGATgAGTTTTCCAGAGCAG
CCCCAAAtCCCTAGGAGCTGTTGAtCCCTTAGGTATCTTTCTACAGCCAGgaCTCTtGCCCGGAG
CTGCTTAATGCCCCAGACCGTGAGTTGCAACAtATGTTGtTgCGCCTCTATaGCTCTCAGCAAA
TTGTTTTGctgttGCaCTATACCAgAcAAtttAAttggaATGAAACGTGAAAGG

>TMH_008
TATGATATaTACtACtGCTaGttTGAtATTTCAACCAATtCCACAGGATtGCCCACtTGTCcgTGgAA
GAAGTCTTGTtcATTCTTTTCCTGCtGGtTCTgCGACCCTCAAGTAGATtAtATATTAAGTGTGTGt
AAgTACTAAgCTcTTTtTCCcAtTgTAACCAGGTCATGTtaTCCcATATcTcAtTCTGAGATTtATTGc
TCCgACTAGAATTCCAaGGcaCaTTAGgGGggcAGATGAtTTTTCCaGAGCaGCCCCAAAtCCcTAG
gAGCTGttgAtCCTGTAGATATCTTTCCACaGCcaGGACTCTTGCCTGGAGCTGTTTAATGCCCCA
GACTGTgAGTTTCAACAGATGTTGTTGAGCCTCTaTAGCcCTcAGCAAATTGCTTtCTGTTGCAC
TATACTTTTTTTATaaTTGTAAagatagGGGgGGcTTtAtaggTTTAAA

>TMH_009
CCTTTGTCATaTACCACAGCAGTTTGATATGGTAACCAATTCCACAGACTTGCCCACTTGTCCa
aTGCTAATAAGtCTTGtTCATTCTTTTCCTGTtGGGTCTGCGATTCTTCTATTAGCCCATATATTA
TGCTTGTGTAATTgtCAATTTCTTTATCCCATTGTAACCAGGTCATGTTATCCCATATTTCATCA
TAAGTTTTATTACTCCAACTAGAGtTCCAAGGCACGTTAGGGgggcaGAtGAgTTTTCCAGAgCA
GCCCCAAATTCctAGGAgCTGTtGATCCTtTAGgtATCTtTCCaCAGCCAGtaCTCTtGCCTGgAGCT
GTtTAATGCCCCAGACCGTGAGTTTCagCAgATGTgGTtGAGCCTCTATAGCTCTCAGCAAATTT
CTTTGCTGTTGCaCTATACCAGAcATTAATTgTAATGATAGTAGGaGGctTa

>TMH_010
TCAgTCTTCaTgTACCACAGCAGTTTGATATATcgaACCAATTCCACAAACTTGCCCATTTGTCC
aaTTCCAATAGTTCTTTTTCATTCTTTTCTTGCTGGGTTTGCGAtTCTTCAATTAATTCaTATATT

AACCCTGTGtAATTggCAATTTCTTTTTCCCAATCCATCCAGGTCATgTTCTCCCAAATCTCATC
TAGGGATTTATTACTCCaGCTAGAGTTCCAAGGCACATTAGtGGTGaAAATGTGTTTTCCAGA
GCAACCCCAAACTCCTAGGAGCTGTTGATCCCTTAGGTAGCTTTCCACAGCcAGGATTCTTGC
CTGGAGCTGTTTAATGCCCCAAACTGTGAGTTGCaACATATGCtGTTGCGCCTCTaTAGCTCTC
AGCAAATtATTTTGCTGTTGCACTATACCAGACAATAATTGtAATGATAGTAGGAGgCTtGgTAg
G

>TMH_011
GtTTtGaTTTaCCACAGCAgTTTGTTATGctAACCAATTCCACAGACTTGCCCACTGaTCcAATgAC
AATAAGTCTTtTTCATTCTTTTCCTGCTGGTTCTGTGAaTCCTCAATTAGCCTATATATTaTGTC
TGTGTAATTgCTAATTTCTTTaTCCCATTGCAGCCAGGTCATATTATCCCATATCTCATTGTAA
GATTTATTACTCCAACTAGAGTTCCAGGGCACATTAGTGGTGCAGATGAGtcTTCCAGAGCAG
CCCCAAATTCCTAGGAGCTGTTGATCCTTTAGGTAatTTTCCACAGCCAGGACtCGTGCCTGGA
GCTGCTTAATGCCCCAGACCGTGAGTCTCAACAtATGcTGTTGAGCCTCTATAGCCCTCAGCA
AATTGCTTTGCTGTTGCACTATACAaAAAAATTTTTTtAGGT

>TMH_012
CtAGGGtgTGGTtAcTGCagctgATATaTCaGtCAaGtccACgGtTgTGCccATttATCCaTGCCAGTAAGTC
TTGTTCATTCTTTTCCTGCTGGtTcTGCgAtTcTtCAATTAGcacATATATTAagTCTGTGTAATTGT
CAATTTCTCTtTCCCATTGCATCCaGGTCATATTGTCCCATATCTCATGCTGAGATTTATTACTC
CAACTATTGTTCCAGGGCACATTAGTGGTGCAGATGAGCCTTCCAGAGCAGCCCCAAATcCC
TAGGAGCTGTTGATCCCTTAGGTATCTTTCCACAGCCAGGACTCGtGCCTGGAGCTGCTTAAT
GCCCCAGACCGTGAGTCTCAACAGaTGTtGTTGAGCCTCTATAGCCCTCAGCAAATtGCTTTGc
TGtTGcACTATACCAGACaATAaTTGTAAttaTaggAGGaggccTgattaaggTTtAAAA

>KNMH_001
TttttttgGctaTACAcAACAATttTtaTaAgTcaTACCAATTCCACAAAtTTGCCCATTTGTCCAATtGCAA
CaGtTCTTtTTCATTTTTGTCTTGCTGGATTTGCGATTCTTCAATTAAaGTGTATATTAcACCTGT
GTAATTGgCAATTTCTCTTTCCCACTCCATCCAGGTCATGTTATcCCAAATTTCCCCTAGAGAT
TTGTTACTCCAGCTAGAGTTCCAGGGCACATTAGTGGTGCAAATGtGTTTTCCAGAGCAACCC
CAaATTCCTAGGAGCTGTTGATCCCTTAGGTAGCTTTCtACAGCCAGgACTCTTGCCTGGAGCT
GTTTGATGCCCCAGACTGTGAGTTGCAACAGGTGCTGTTgCGCCTCTATAGCCCTCAGCAAAT
TGTTTTGCTGTTGCACTATACCAGACAATaAaTTgAAAGGG

>KNMH_002
CaacTCCaTaTaCACAGCCATTTtGaTATgTCAAACCAaTTCcACAAaCTTGCCCAgcTGTCCAATTtc
AATAgTTCtTgTTCATTCTTTTCTTGCaGGgTtTGCGAaTCTTCAAgcAAGCTATATATTaaAcCTGT
GTAATTGTCAATTTCTCTTTCCCACTCCATCCAGGtCATGTTTCCCCAAATTTCTTCTACaGATC
TATTACtcCaGCTAGaGTTCCAGCGCACAGAAGTGGcGcAAATAtGTTTTCCAGAGCAACCCCaA
ATTCCTAGGAgCTGTTGatCCCTtAGGTAAcTTTCCACAGCCAGGACTCTTGCCTGTAGCTGtTT
AATGCCCCAGACTGTgAGTTGCAGCAaATGCTGGTGCGCCTCTATaGCCCTCAGCAAAaTGtTT
TGCTGtTGCACTATACcAGAAaAAAATGtAAGCCA

>KNMH_003
GCagtCTTATaTACCACaGCAATTTGATATGcCAaACCAATTCCACAGACTTGCCCACTTGTCCA
ATGCcAATAATTCTTGTTCATTCTTTTCCTGCTGGTTCTGtgATTCTTCAATCAGTTGATATATT
ATGTGTGTGTAATTGCTAATTTCTCTATCCCATTGTAGCCAGGTCATGTTATCCCATATCTCA
CTCTGAGATTTATTACTCCAACTAGAGTTCCAaGGCACATTAGTGGTGCAGATGAGTTTTCCA
GAGCAGCCCCAAATTCCTAGGAGCTGTTGATCCcTTAGGTATCTTTCCAGAGCCAGGACTCTT
GCCTGGAGCTGTTTAATGCCCCAGACTGTGAGTTTCAGCAGATGTTGctGAGCCTCTATaGCCC
TCAGCAAATTGcTTTGctgttGCACTATACCAGAcaAtTaaaTTgTAATG

>KNMH_004
GagCCTTGtTATACCACAGCCATTTGTTATATCAACCAGTTCCACAAGCTTGCCACTTGTCCAA
TtCttATaattCCTTTTCATTCTTTTCTTGCTGgATCTGCGaTTCTTCAATTAAGcTATATATTGCAcC

TGTGTAATTGTCAATTTCCCTTTCCCACTGCATCCAGGTCATGTTCTGCCAAATGTAATCTAG
AGATTTATTACTCCAACTAGAGTTCCAGGGCACAGCAGTGGTGCAAATGTGTTTTCCAGAGC
AACCttAAATTCCTAGGAGCCgTTGATCCTGTAGGTATCTTTCCACAGCCAGGATTCTtGCCTTG
AGCTGTTTAATGCCCCAGACCGTGAGTTGCAACAGATGCTGTTGCGCCTCTATAGCTCTCAG
CAAATTATTTTGCTGTTGCACTATACCAaCAATAAAaTtgTAAAGGGCG

>KNMH_005
GtTTGTCATaTACCACAGCATCTTGATATGTCaaCCAATTCCACAGACTTGCCCACTTGTCCaAT
cCCAATAAGTCTTGTTCATTCTTTTCCTGCTGGtTCTGCgATTCTTCAatCaGGCCATATATTATT
TGTGTGTAATTGCCAACCTCTTTATCCCATTGCAGCCAGGTCATGTTGTCCCaTATGtCAtTGA
ATGAtTTATTACTCCaACTAGAGTTCCAGGgCaCATTAGtGGtGCAGATGAGtTTTCCAGAGCAA
CCCCAAAtTCCTAGGAGCTGCTgATCCCTtAGGTAACTTTCCAGAGCCaGGaCTCTTGCCTGGA
GCTGTTTAATGCCCCAGACTGTGAGTTTCAACAGATGCTGTTGAGCCTCAATAGCCCTCAGC
AAATTGCTTTGCTGTTGCACTATACCAGAcAATAATTgTAAaGATAGTAGGA

>KNMH_006
AtTTTTTATaTACCACAGCAGTTTGATATGTcAACCAATTCCACAGATTTGCCCACTTATCTAaT
gCCAATAAtTCTtGTTCATTCAgTTCCTGCTGGTTTTGTGACTCTTCAAGTAGATtATATATTATC
TGTGTGTAATTGTCAATTTCTTTATCCCATTGCAgCCAGGTCATGTTTCcCCATATCTCaTCCTG
GGATtTATTACTCCAACTAGAATTCCAGGGCACATTAGTGGTGCAGATGAGcTTTCCAGAGCA
GCCCCAAATTCCTAGGAGCTGTTGATCCTTTAGGTATCTCTCCACAGCCAGGACTCTCGCCTG
GAGCTGTTTAATGCCCCAGACTGTGAGTCTCAACAgaTGCTGTTGAGCCTcTATAGCtTTCAGC
AAATTGTTTTGCTGTTGCACTATACCAGAAAaTAATTgtTAGAgAGTaGGAGgCttgAtAAgggAAt
ATaAAGGGGG

>KNMH_007
ATtTcCtcTGtACcAcAGCAGTTTGATATgTCaaCCAATTCCACAgACTTGCCCACTTGTCCaCGCC
AATAAGTCtTtTTCATTTcTTTCCtGCTcTcTCTGTGCCTCTTCAAGCaGGGtATATaTTGTTgcTGT
GtAAtTgCcAATTTCTTtATCCcATTgCATCcAGGTCATGTTGcCCCATATCTCACTCTGAGATTTA
TTAtTCCgaCtAgAGgTCCAGGgCACAGtAGTgGTGCAGAtGAgtTTTCCaGAGCATCCCCAAATTC
CTAGgAGCTGCTGATCCTCttgTATCTCTCCTCAGCcGcGACTCTTGCCTGGAGCTGtTtAATGCC
CCAGACTGTgGTTTCAACAGAaGTTGTTGAGCCTCTaTgGTgGTCAGCAATTGCTTTTtGTTGtCT
ATAcTAgATGAtAATTGTaATAATAGTAGGAGGtTGaTagTATTTaTAAaGGGG

>KNMH_008
ATTCTTATATACACAGCCATtTTGTTATGtCAAACCAATTCCACAGATTTGCCCACTTGTCCAA
TgCcAATAAGTCTTGTTCATTCTTTTCCTGCTGGTTTTGTGAGaCTTCAAGTAGGtTATATATAT
ATTCTGTGTGATTGCTAATTTCTCTCTCCCATTCCAgCCAGGTCATGTTCTCCCATATTTCcTtC
AGAGATCTATTATTCTTATTCCAGCTAGAgTTCCAGGGCACATTAGTGGTGCAGATGAGTTTT
CCAGAGCAGCCCcAAATTCCTAGGAGCTGTTGATCCtTTAGGTATCTTTCCACAGCCAgGACTc
TTGCCTGGAGCTGcTTAATGCcCCAGACCGTGAGTtTCAACAGaTGCTGTTGAGCCTCTATAGC
CTTCAGCAAATTGcTTTGCTGTTGCACTATACCAGCAAATaAaTtgTAAC

>KNMH_009
TgTTTGtTTaTgtACcACAGCAGaTTGATATGTTAACCAATTCCATAGATTTGCCCACTTGTCCaac
GtCAATAAGtCTTGtTCATTCTTTTCCTGCTGGTTCTGCgACTCTTCAATTAACCCATATATTATC
TCTGTGTAATTGCTAATTTCTTTATCCCATTgCaGCCAGGTCaTgTCTttCCCATATCTGTTCCTG
GGTTTTATTACTCCAACTAGAGTTCCAGGGCACATTAGTGGTGCAGATGAGTCTtCCAGAGCA
GCCCCAAATTCCtAGGAGCTGTTGATCCTTTAGGTATCTTTCCACAGCCAGGACTCTTGCCTG
GAGCTGTTTAATGCCCCAGACCGTGAGTCTCAACATAtGCtGTTGAGCCTCTATAGcCTTCAGC
AAATTGCTTTGCTGTtGCACTATACCAGACAATAATTGTaATGATAGTAGGAGGCTTGATatgGt
tttAA

>KNMH_010

GTATgTATATACCACAGCACTTTGATATGTCaACCAAGTCCACAGATTTGCCCACTTGTCCAaT
gcCAGTAAGTCTTGTTCATTCTTTTCCTGCTGGtTCTgCgATTCTtCAATTAGTTTATATATTATTT
GTGTGTAATTGCTAATTTcTTTGTCCCATTGCAGCCAGGTCATGTTATCCCATATCTCAGCTTG
AGATTTATTACTCCAACTAGAGTtCCAAGGCACATTAGTGGGGCAGATGAGTTTTCCAGAGC
AGCCCCAAATTCCTAGGAGCTGTTGATCCTTTAGGTATCTTTCCACAGCCAGGACTCGTGCCT
GGAGCTGTTTAATGCCCCAGACCGTGAGCTTCAACAGATGTtGTTGAGCaTCTATAGCCCTCA
GCAAATTGCTTTGCtGTTGCACTATACCAGACAATAaTTGTAGTGATAGTAGGAGGCGTgGAG
GG

>KNMH_011
TAGTGTTATATACCACAGCATTTTGtTATGCTAACCAATTCCATAAGCTTGCCCATTCATCCA
ATTTcAATAGTTCtTGTTCATTTTTtTCTTGCTGGtTTTGCgATTCTTCAATTAAGCTATATATTA
AttCTGTGTAATTGTCAATTTCTCTTTCCCATTCCATCCAGGTCATGTTTcCCCAAATATAATCT
AGAGtTCTATTATTACTCCAGCTAtcGTTCCAAGGcACAGCAGTGGTGCAAATGAGTTTTCCAG
AGCAACCCCaAATTCCTAGGAGCTGTTGATCCtTTAGGTATcTTTCCACAgCCAGGACTCTTGC
CTGGAGCTGTTTAATGCCCCAGACTGTGAGTTGCAACAgAGACTGTTGcGCCTCTATAGCCCT
CAGcAAGTTGTTCTGCTGttGCACtATAccAgAAAAtAAAtaGTAAGTGGGGGGGTT

>KNMH_012
GCATCTTTAaCTACCaCaGCCaGTTTGTtATATTACACCAACTCCACAGACTTTGCcATTTGtCCA
ATGCTaGTAATTCCTGTTCATTTgTCtCCTGCtGgTTTTGCGAGTCTTCAAGCAACTGGTATATT
GTGTTTGTATaAtTgttAatTgctcTaTcAcATTTaAACAAGGTCATGTTACCCCAAATATAGTCCTTA
GATTTGTTACTCCAGCTAGAGTTCCAAGGCACAGCAGTGGTGCAGATGACTTTTCCAGAGCA
GCCCCAAATCCCTAGGAGCTGTTGATCCTTTAGGTATCTTTCTATAGCCAGGACTCTTGTCTG
AAGCTGTTTAATGCCCCAGACCGTGAGTTGCAACATATGTtGtTGCGCCTCTATaGCTCTCAGC
AAATTGCTTTgCTGTtGCACtATACcAgcAA