SEQUENCE DIVERSITY OF HIV-1 SUBTYPE C ACCESSORY GENES: *VIF*, *VPR* AND *VPU*

MSc (Med) in Medical Virology

K M MOTHAPO

2010

SEQUENCE DIVERSITY OF HIV-1 SUBTYPE C ACCESSORY GENES: *VIF*, *VPR* AND *VPU*

by

Mr K M Mothapo

RESEARCH DISSERTATION

Submitted in fulfillment of the requirements for the degree of

MASTER OF MEDICAL SCIENCE

in

MEDICAL VIROLOGY

School of Pathology, Faculty of Health Sciences

at the

University of Limpopo, Medunsa Campus

Supervisor: Prof M J Mphahlele

Co-supervisors: Mr A M Musyoki and Dr S G Selabe

Department of Virology

March 2010

DECLARATION

I, **KHUTSO MANONE MOTHAPO,** hereby declare that the work for which this dissertation is based is original, and has not been submitted to any other institution for the purpose of any qualification.

This dissertation is being submitted in fulfillment of the MSc (Med) in Medical Virology, in the Department of Virology, School of Pathology, Faculty of Health Sciences, Medunsa Campus, University of Limpopo.

---------- ----------

Initials and Surname (Title) Date

Student Number

DEDICATION

I dedicate this work to my family; Mr and Mrs Mothapo for all the support they gave me. To my brother Thabo, I like to say everything is possible with a brother like you. To Mmapelo, James and Lehlogonolo, you guys really know how to make me laugh through stressful times, so thanks for that. Thanks to all my family for giving me all your love, I love you all.

ACKNOWLEDGEMENTS

I would like to acknowledge the following:

The Department of Virology at the University of Limpopo (Medunsa Campus) for letting me be part of their scientific family.

My fellow MSc students, for making the last two years seem to be very short and that just goes to say I enjoyed every moment that I have spent with you.

The Hepatitis and HIV Research Unit (HHRU), M P Gededzha, L R Lebelo, A Lukhwareni, L A Malinga, J N Rakgole and A M Musyoki, for making our laboratory a good work place.

Mr A M Musyoki for being the best supervisor and a friend.

Prof M J Mphahlele and Dr S G Selabe for grooming us to be where we are right now and for showing that fatherly and motherly support through tough times.

The National Research Foundation (NRF) and the Poliomyelitis Research Foundation (PRF) for financial assistance.

ABSTRACT

OBJECTIVES: To date there is no effective and safe vaccine to stop the spread of human immunodeficiency virus (HIV) and provide cross protection among different subtypes. HIV accessory genes were overlooked for many years and recently they are becoming candidates for development of new anti-HIV drugs and vaccines. This is supported by their ability to elicit cytotoxic T lymphocyte response. To date, there are limited studies on accessory genes (*nef*, *vif, vpr* and *vpu)* on South African HIV strains. This study sought to amplify and analyse the sequences of HIV-1 subtype C accessory genes (*vif*, *vpr* and *vpu*) to assess the genetic diversity as well as the motifs and residues associated with key biological functions of these genes. This study further sought to compare the degree of genetic diversity between the accessory and structural genes.

METHODS: The study was an exploratory study using stored (-70ºC) HIV positive plasma samples. The study population comprised of 25 HIV positive plasma samples which were already sequenced in the *gag* and *env* genes in another study. The samples were drawn from the neighbouring townships of Pretoria: Ga-Rankuwa, Soshanguve, Mamelodi, Laudium, Kalafong, Jubilee and Mabopane. For the purpose of this study, the same samples were amplified, sequenced and characterised in the *pol* and accessory (*vif*, *vpr* and *vpu*) genes in order to obtain near full length sequences of the HIV isolates from Pretoria region. Six samples were cloned for accessory genes. Five clones from each sample were selected. Sequence analysis was performed for all the PCR amplicons and clones. Base calling for the sequences generated was performed on Chromas Pro program. Computing of phylogenetic tree was performed with MEGA 4 program. ClustalW software was used for sequence alignment and translation of nucleotides to amino acids was performed with BioEdit. The amino acid alignments were analysed on graphic view.

RESULTS: All 25 samples were successfully amplified for accessory genes (*vif*, *vpr* and *vpu*) and *pol* gene. All the 25 *pol* PCR amplicons were successfully sequenced, while all but one accessory PCR amplicons were successfully sequenced. A number of conserved motifs and residues were observed in all the four genes (*vif*, *vpr*, *vpu* and *pol*). *Vif* and *vpr* showed to harbour most of these conserved motifs and residues; 144-SLQYLA-149 and H71 respectively. In addition, the R77Q mutation associated with long term non-progressors was observed in the *vpr* gene of 15 sequences. Drug resistant mutations were evaluated in both protease and RT regions. Nine samples had one or two drug resistant mutations i.e T74S, L10I, V179D, E138A/D, Y318F,Y181C and K108N.

Phylogenetic analysis confirmed the 25 HIV positive samples to be HIV-1 subtype C in both structural and accessory genes. The genetic diversity of HIV-1 subtype C was compared between accessory (*vif*, *vpr* and *vpu*) and structural (*pol*, *gag* and *env*) genes. The gag and env sequences were available from a previous project (Musyoki, 2009). The *gag* and *vif* gene sequences were highly conserved (89% to 96% and 88% to 96%, respectively), as compared to *vpr* gene (84% to 94%), the *pol* gene (79% to 95%), the *env* gene (83% to 93%) and finally the *vpu* gene (73% to 92%).

CONCLUSION: This study found that amplification of clones was more sensitive as compared to direct samples and analysis of clone sequences was more clear than analysis of direct PCR products. Functional motifs and residues observed in all accessory genes were highly conserved. *Vif* was more conserved, followed by *vpr* and *vpu*, respectively. Genetic analysis of *pol* gene revealed that there were drug resistant strains in circulation. This indicates that the patients were infected with drug resistant viruses; this cannot be verified from the study population. And that most of the strains in this study had mutations associated with long term non-progressors (LTNP's). However, it is not known whether these patients were indeed LTNP's. Comparison of genetic diversity between structural and accessory genes

demonstrated that, *gag*, *vif* and *vpr* were more conserved than *pol*, *env* and *vpu.*

PRESENTATIONS FROM THIS DISSERTATION

Local meetings

KM Mothapo, AM Musyoki, SG Selabe and MJ Mphahlele. Title: “HIV-1 accessory genes (*vif* and *vpr*) diversity of patients from Dr. George Mukhari Hospital”. University of Limpopo’s Research Academic Day 2009 (Oral presentation). 13-14 August 2009

KM Mothapo, AM Musyoki, SG Selabe and MJ Mphahlele. Title: “Evaluation of functional motifs and residues within HIV-1 subtype C accessory genes”. University of Limpopo’s Research Academic Day 2010 (poster presentation). 11-12 August 2010

KM Mothapo, AM Musyoki, SG Selabe and MJ Mphahlele. Title: “The genetic diversity between HIV-1 subtype C accessory genes (*vif, vpr* and *vpu*) and structural genes (*pol, gag* and *env*); in patients from Pretoria and surrounding regions”. University of Limpopo’s Research Academic Day 2010 (Oral presentation). 11-12 August 2010

International meetings

KM Mothapo, AM Musyoki, Rakgole JN, SG Selabe, PO Bessong and MJ Mphahlele. Title: “Phylogeny and analysis of sequence motifs associated with biological activities of HIV-1 subtype C structural and accessory genes from South Africa”. AIDS vaccine meeting 2010 in Atlanta, Georgia (Poster presentation). 28 September - 1 October 2010

TABLE OF CONTENTS

TITLE PAGE…… **I**

DECLARATION…… **II**

DEDICATION…… **III**

ACKNOWLEDGEMENTS…… **IV**

ABSTRACT…… **V**

PRESENTATIONS FROM THIS DISSERTATION..... **VIII**

TABLE OF CONTENTS…… **IX**

LIST OF TABLES…… **XVII**

LIST OF FIGURES…… **XVIII**

LIST OF ABBREVIATIONS…….... **XX**

CHAPTER ONE

1. PROBLEM STATEMENT AND EXPERIMENTAL PROPOSAL……. 1

1.1 Background……..... 1

1.2 Problem statement…… 2

1.3 Experimental proposal…….. 2

1.3.1 Rationale……... 2

1.3.2 Aim…… 3

1.3.3 Objectives…….. 3

1.3.4 Expected significance of the study……... 4

CHAPTER TWO

2. LITERATURE REVIEW……. 6

2.1 Historical background of HIV……. 6

2.2 Classification of HIV strains……. 7

2.3 HIV structure…….. 8

2.3.1 The virion……. 8

2.3.2 The viral genome……. 9

2.3.3 HIV genes and their functions……. 10

2.3.3.1 Gag (Group Specific Antigen)……. 10

2.3.3.2 Pol (Polymerase)……. 10

2.3.3.3 Env (Envelope)……. 19

2.3.3.4 Tat (Trans-Activator gene) and rev (Regulator of virion)…….. 11

2.3.3.5 Vif (Virion Infectivity Factor)……. 11

2.3.3.6 Vpr (Virion protein R)……. 12

2.3.3.7 Vpu (Virion protein U)…….. 12

2.3.3.8 Nef (negative Regulatory factor)……. 13

2.3.3.9 LTR (Long terminal repeat)……. 13

2.4 Replication of HIV…….. 14

2.4.1 Attachment and entry……. 15

2.4.2 Reverse transcription……. 16

2.4.3 Integration……. 17

2.4.4 Replication and transcription……. 17

2.4.5 Viral assembly and release……. 17

2.5 Route of HIV transmission……. 18

2.5.1 Sexual transmission of HIV…….. 18

2.5.2 Transmission of HIV from mother to child…….. 18

2.5.3 Direct inoculation with HIV infected blood and blood products…. 19

2.6 Immune response…….. 19

2.7 Diagnosis of HIV……. 20

2.7.1 Screening tests..... 21

2.7.1.1 Antibody detection……... 21

2.7.1.2 Antigen detection……. 21

2.7.2 Confirmatory test..... 22

2.7.2.1 HIV nucleic acid detection……. 22

2.8 Treatment of HIV……. 22

2.8.1 When to start HIV treatment…….. 22

2.8.2 ARV's for HIV treatment……. 24

2.8.3 Choices in ARV groups……. 25

2.8.4 Combination recommendations……. 25

2.9 Prevention of HIV transmission and infection……... 26

2.9.1 Sexual transmission……. 26

2.9.2 Preventing direct inoculation……. 27

2.9.3 Preventing mother to child transmission……. 27

2.10 Epidemiology of HIV……. 28

2.11 Genetic diversity of HIV……... 28

2.11.1 Cause of genetic diversity……. 28

2.11.2 HIV recombination……. 29

2.11.3 HIV distribution……. 31

2.12 Important regions and mutations in different genes……. 32

2.12.1 Important conserved regions……. 32

2.12.2 Important mutations……. 33

2.13 HIV vaccine…….. 34

2.13.1 Challenges in developing an HIV vaccine…….. 35

2.13.2 Vaccines in clinical trials……. 35

2.14 Possible approaches for HIV intervention using accessory genes 37

CHAPTER THREE

3. MATERIAL AND METHODS…….. 38

3.1 Ethical approval……. 38

3.2 Study design…….. 38

3.3 Study site and population…..……. 38

3.4 Sample nomenclature…….. 40

3.5 Laboratory methods…….. 42

3.5.1 Storage of samples……. 42

3.5.2 RNA extraction……. 43

3.5.3 HIV cDNA synthesis…….. 43

3.5.4 HIV PCR assay……... 45

3.5.4.1 First round PCR assay for HIV accessory genes…….. 45

3.5.4.2 Second round PCR assay for HIV accessory genes…… 45

3.5.4.3 First round PCR assay for pol gene…… 47

3.5.4.4 Second round PCR assay for pol gene…….. 47

3.6 Detection of PCR products……... 47

3.7 Sequencing analysis…… 48

3.7.1 Direct sequencing of PCR products…… 48

3.7.2 Cloning and sequencing…….. 49

3.7.2.1 Rationale for cloning…….. 49

3.7.2.2 Principle of cloning…… 49

3.7.2.3 Ligation reaction…….. 50

3.7.2.4 Transformation reaction…… 51

3.7.2.5 Confirming of cloned fragments…… 52

3.7.2.6 Detection of cloned PCR amplicons…… 53

3.7.3 Analysis of sequences…….. 53

3.7.3.1 Sequence editing…… 53

3.7.3.2 Sequence alignment…….. 53

3.7.3.3 Phylogenetic construction…… 54

3.8 Avoiding PCR contamination…….. 54

CHAPTER FOUR

4. RESULTS…….. 55

4.1 Overview of the results…… 55

4.2 Overview of PCR and sequencing results…… 57

4.3 Detection of PCR amplicons…… 57

4.4 Sequence analysis of accessory genes…… 59

4.4.1 Direct sequencing of the vif genes…… 59

4.4.2 Vif cloned sequences…… 62

4.4.3 Direct PCR sequences of vpr gene…… 65

4.4.4 Vpr cloned sequences…… 67

4.4.5 Direct PCR sequences of the vpu gene…… 69

4.4.6 Vpu cloned sequences…….. 71

4.5 Sequence analysis of the pol gene…… 73

4.6 Phylogenetic analysis of accessory genes…… 80

4.6.1 Phylogeny of the vif gene…… 80

4.6.1.1 Direct sequences versus clones…….. 80

4.6.2 Phylogeny of the vpr gene…….. 83

4.6.2.1 Direct sequences versus clones…… 83

4.6.3 Phylogeny of the vpu gene…… 86

4.6.3.1 Direct sequences versus clones…….. 85

4.7 Phylogeny of the pol gene…….. 89

4.8 Comparison of genetic diversity between accessory and structural genes……... 91

CHAPTER FIVE

5. DISCUSSION……. 93

5.1 Overview of the study……. 93

5.2 Detection and molecular characterization of accessory genes……. 94

5.3 Conserved motifs and important mutations of accessory genes…... 95

5.4 Phylogenetic analysis…….. 99

5.5 Implications of the findings on HIV prevention and control…….. 100

CHAPTER SIX

6. CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS……. 104

6.1 Conclusions…….. 104

6.2 Limitations…….. 106

6.3 Recommendations……. 106

CHAPTER SEVEN

7. REFERENCES……. 107

CHAPTER EIGHT

8. APPENDICES…….. 118

8.1 Accessory genes (vif, vpr and vpu) nucleotide sequences……. 118

8.2 Sequence homology for vif, vpr and vpu…….. 130

8.3 Gag and env nucleotide sequences…….. 136

LIST OF TABLES

Table No.

2.1: Classification of HIV associated with clinical disease……. 23

2.2: Choice of ARV…….. 25

2.3: Recommended combinations……. 26

2.4: WHO guidelines for PMTCT drug regimens in resource-limited setting…….. 27

2.5: CRF's described……. 30

2.6: Summary of conserved regions and their functions……. 32

2.7: Summary of important mutations……. 33

2.8 HIV vaccines ongoing trials…….. 36

3.1: Summary of the study population…….. 39

3.2: Primers for accessory genes…….. 44

3.3: Primers for pol gene……. 46

3.4: Sequencing primers for accessory genes…….. 48

3.5: Sequencing primers for pol gene……. 49

4.1: Summary of mutations within the accessory genes functional motifs and residues…….. 73

4.2: Summary of drug resistant mutations……. 74

LIST OF FIGURES

2.1: HIV groups and subtypes classification…… 8

2.2: HIV structure…… 9

2.3: HIV genome…… 10

2.4: Schematic representation of HIV-1 replication…….. 15

2.5: HIV-1 disease stages…… 20

2.6: Schematic representation of copy choice and recombination generated by copy choice…… 29

2.7: Geographic distribution of HIV…… 31

3.1: South Africa map showing the location of townships in Pretoria…. 40

3.2: Schematic summary of laboratory methods…….. 42

3.3: Schematic representation of HIV genome with primer binding sites for accessory genes…….. 44

3.4: Schematic representation of HIV genome with primer binding sites for pol gene…… 46

3.5: The pGEM-T Easy vector circle map…….. 50

3.6: Schematic representation showing growth of recombinant plasmids on agar…… 52

4.1: Summary of PCR and sequencing results…… 57

4.2: Gel electrophoresis of PCR amplicons of accessory genes…….. 58

4.3: Gel electrophoresis of pol gene PCR products…… 58

4.4: Amino acid sequence of vif from 24 HIV-1 Subtype C isolates……. 61

4.5: Graphic view of the amino acid alignment of vif sequences from 30 HIV-1 subtype C clones…….... 64

4.6: Graphic view of the amino acid alignment of vpr sequences from 24 HIV-1 subtype C isolates…….. 66

4.7: Graphic view of the amino acid alignment of vpr sequences from 30 HIV-1 subtype C clones……. 68

4.8: Graphic view of the amino acid alignment of vpu sequences from 24 HIV-1 subtype C isolates……. 70

4.9: Graphic view of the amino acid alignment of vpu sequences from 30 HIV-1 subtype C clones……. 72

4.10: Graphic view of the amino acid alignment of pol (protease) sequences from 25 HIV-1 subtype C isolates……... 75

4.11: Graphic view of the amino acid alignment of pol (RT) sequences from 25 HIV-1 subtype C isolates…….. 79

4.12: Phylogenetic relationship of vif gene samples (direct PCR products)……. 81

4.13: Phylogenetic relationship of vif gene clones……. 82

4.14: Phylogenetic relationship of vpr gene samples (direct PCR products)……. 84

4.15: Phylogenetic relationship of vpr gene clones……. 85

4.16: Phylogenetic relationship of vpu gene samples (direct PCR products)……. 87

4.17: Phylogenetic relationship of vpu gene clones…….. 88

4.18: Phylogenetic relationship of pol gene samples ……. 90

4.19: Phylogenetic comparison between accessory and structural genes…….. 92

LIST OF ABBREVIATIONS

General

µl……. Microliter

3TC……..…. Lamivudine

aa……. Amino acid

ABC……..…… Abacavir

AIDS…….... Acquired Immunodeficiency Syndrome

ART…….. Antiretroviral therapy

ARV's……..... Antiretrovirals

AZT……..... Zidovudine

CCR5…….. C-C chemokine receptor 5

CD4…….. Cluster of differentiation 4

CD8…….. Cluster of differentiation 8

CDC…….. Centre for Disease Control and Prevention

cDNA……. Complementary deoxyribonucleic acid

CRFs……. Circulating Recombination Forms

CTL……... Cytotoxic T Lymphocyte

CXCR4……. C-X-C chemokine receptor 4

d4T……. Stavudine

ddI……. Didonosine

DNA…….. Deoxyribonucleic acid

dNTP……. Deoxynucleotide triphosphate

ds…….. Double stranded

EDTA……..…. Ethylenediamine

EFV…….. Efavirenz

EIA……... Enzyme Immunoassay

ELISA……. Enzyme Linked Immunosorbent Assay

Env……. Envelope

ER…….. Endoplasmic Reticulum

FDA……. Food and Drug Administration

FTC……. Emtricitabine

Gag……... Group specific antigen

gp……... Glycoprotein

HAART…….. Highly Active Antiretroviral Therapy

HIV……. Human Immunodeficiency Virus

HTLV-III……. Human T-cell Leukaemia Virus III

IAVI……. International AIDS Vaccine Initiative

IL……. Interleukin

IPTG…….. Isopropyl-β-D-thio-galactoside

LAV……. Lymphadenopathy Associated Virus

LB……..... Luria-Bertani

LTRs…….. Long Terminal Repeats

MCREC……. Medunsa Campus Research and Ethics Committee

MeCRU……. Medunsa Clinical Research Unit

MREC……..... Medunsa Research and Ethics Committee

mRNA……... Messenger RNA

MTCT……. Mother-To-Child Transmission

MW……. Molecular Weight

Nef……. Negative regulatory factor

NHLS……. National Health Laboratory Services

NNRTIs……. Non-nucleoside Reverse Transcriptase Inhibitors

NRTIs……. Nucleoside/nucleotide Reverse Transcriptase Inhibitors

NVP…….. Nevirapine

ORFs……. Open Reading Frames

PBMC……. Peripheral Blood Mononuclear Cell

Pbs……. Primer binding site

PCR……..... Polymerase Chain Reaction

PIs……. Protease Inhibitors

PMTCT……. Prevention of Mother-To-Child Transmission

Pol……. Polymerase

RNA…….. Ribonucleic acid

Rpm…….. Revolution per minute

RRE…….. Rev responsive element

RT……. Reverse transcriptase

RT-PCR……..... Reverse Transcriptase Polymerase Chain Reaction

SIV……. Simian Immunodeficiency Virus

TAR…….. Transactivation response elements

Tat……. Trans-activator of transcription gene

TBE…….. Tris/Borate/EDTA

TDF…….. Tenofovir

UNAIDS…….. United Nations Program on HIV/AIDS

V3……. Third Variable region

Vif……. Virion Infectivity Factor

Vpr…….. Viral protein Regulatory

Vpu……. Viral protein Unknown

WHO…….. World Health Organisation

X-Gal……. 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside

Nucleotides

A……. Adenine

C……..... Cytosine

T……. Thymine

G……..... Guanine

Amino acids

A……..... Alanine

C……..... Cysteine

D…….. Aspartate

E…….. Glutamic acid

F…….. Phenylalanine

G…….. Glycine

H…….. Histidine

I……..... Isoleucine

K…….. Lysine

L…….. Leucine

M……..… Methionine

N……..... Asparagine

P……..... Proline

Q…….. Glutamine

R…….. Arginine

S……..… Serine

T……..... Threonine

V…….. Valine

W……..... Tryptophan

Y…….. Tyrosine

CHAPTER ONE

1. PROBLEM STATEMENT AND EXPERIMENTAL PROPOSAL

1.1 Background

Human immunodeficiency virus (HIV) is a single stranded RNA virus with enveloped icosahedral capsid and belongs to a family *Retroviridae*. It is further classified into a genus called *Lentiviruses* (lenti-, means slow). There are two types of HIV: HIV-1 and HIV-2 and are both transmitted sexually, through blood and from mother to child. The two HIV types result in an important human disease called AIDS. HIV-2 takes some time to cause the disease and appears to be limited to West Africa (Papathanasoulos et al, 2003; Perrin et al, 2003).

HIV-1 is further divided into three groups; group M (M=major), group O (O=outer) and group N (N=non-M and non-O). HIV-1 is the cause of around 90% of HIV infections globally. Group O is found to be restricted to west central Africa and group N was discovered in Cameroon and is extremely rare. Group M is subdivided into 9 subtypes, A, B, C, D, F, G, H, J and K (Perrin et al, 2003). There are recombinants forms of HIV-1 found in human population and are called circulating recombinant forms (CRFs). HIV-1 subtype C is the predominating strain found in South Africa and other southern Africa countries (Papathanasoulos et al, 2003).

HIV-1 genome has 9749 nucleotides which codes for nine genes. Three genes (*gag* and *env*) contain information needed to make structural proteins, *pol* gene codes for functional enzymes; protease, reverse transcriptase and integrase, two genes (*tat* and *rev*) code for regulatory proteins and the other four genes (*nef, vif, vpr* and *vpu*) code for accessory proteins that control the ability of the virus to infect a cell, and thus are responsible for viral

pathogenesis. HIV-1 accessory genes (*vpr, vpu* and *vif*) are located between the *pol* gene and *env* gene (Hunt, 2008).

1.2 Problem Statement

HIV-1 is the leading cause of AIDS worldwide, and it is estimated that 33.2 million people worldwide are living with HIV, with 2.5 people newly infected per year in 2007. The sub-Saharan Africa is worst affected by HIV pandemic, with the adult prevalence of 5.2% by the end of 2008, (http://www.avert.org/worldstats.htm). In South Africa, it was estimated in 2006 that 5.5 million people were HIV positive and 320,000 people died of AIDS, bringing an estimate of around 1000 deaths every day (Knight, 2006). There are different rates of progression of HIV-1 infections to AIDS in different patients. This was also observed in Simian Immunodeficiency Virus (SIV) infected monkeys. Among other studies the different rate of progression to AIDS have been linked to HIV-1 accessory genes (*vif*, *vpr* and *vpu*) since they have been shown to impact on viral replication and pathogenesis of HIV-1. Thus, it is important to study the accessory genes and assess sequence diversity and evolution of the three genes. It has been shown that SIV *vpr* gene diversity is much like the SIV *tat* gene, where evolution of *vpr* demonstrated a significant inverse correlation with disease progression (Richard et al, 2007).

1.3 Experimental proposal

1.3.1 Rationale

To date, there are limited studies on genetic diversity of accessory genes (*nef*, *vif, vpr* and *vpu)* from South Africa. HIV-1 accessory genes are potential vaccine candidates. Three vaccine trials (HVTN 064, HVTN 065 and HVTN 067) in the USA included either *vpu*, *vpr* or both, as vaccine antigens. Trial HVTN 064 has both *vpu* and *vpr* (IAVI, 2008). Hence, it is important to study

genetic diversity of these accessory genes and assess the degree of genetic conservation or variability in different parts of the world.

The sequence diversity of *vif, vpr, vpu, pol, gag* and *env* may vary, with the *pol* and *env* genes revealing a higher degree of diversity. This may be because the *pol* gene is targeted by antiretrovirals (ARV's), thus it is prone to mutations. The envelope protein interacts more with the immune system; this may lead to *env* genes being more variable. *Vif* is expected to be highly conserved and *vpu* less conserved within the three accessory genes (*vif*, *vpr* and *vpu*). It is expected that accessory genes will be highly conserved compared to structural genes.

1.3.2 AIM

The primary aim was to analyse sequence diversity of HIV-1 subtype C accessory genes (*vif*, *vpu* and *vpr*) and *pol* gene from samples previously sequenced for *gag* and *env* in order to generate near full length HIV-1 subtype C sequences. The secondary aim was to compare the sequences of accessory genes (*vif*, *vpr* and *vpu*) to structural genes (*pol*, *gag* and *env*) from HIV-1 isolates recovered in Pretoria and surrounding areas.

1.3.3 OBJECTIVES

1. To perform sequence analysis of the accessory genes (*vif*, *vpr* and *vpu*) and *pol* gene from samples previously sequenced for *gag* and *env* in order to generate near full length HIV-1 subtype C sequences.

   Specific objectives were:

   - To optimize RT-PCR assay targeting HIV-1 accessory genes (*vif*, *vpu* and *vpr*).
   - To mplify HIV-1 accessory genes (*vif*, *vpu* and *vpr*) and *pol* gene on samples sequenced for *gag* and *env* genes.

- To perform direct sequencing and analysis of accessory genes and the *pol* gene.
- To clone and sequence selected interesting PCR products of accessory genes.

2. To assess motifs and residues associated with key biological functions of accessory genes (*vif*, *vpr* and *vpu*) and the *pol* gene sequences.

   Specific objectives were:

   - To identify motifs and residues associated with key biological functions on accessory genes and the *pol* gene sequences.
   - To assess conservation and/or variability within the identified motifs and residues sequences.

3. To compare genetic diversity of accessory genes (*vif*, *vpr* and *vpu*) and structural genes (*pol*, *gag* and *env*).

   Specific objectives were:

   - To obtain genetic sequences of gag and env from a previous study Musyoki, 2009: (MREC/P/136/2008:PG).
   - To perform phylogenetic analysis of accessory genes and *pol* gene to confirm the HIV subtype.
   - To obtain inter-sample sequence homology for each of the accessory and structural genes.
   - To compare inter-samples sequence homology range and phylogenetic analysis between the accessory and the structural genes.

1.3.4 Expected significance of the study

This study will evaluate the divergence and conservation of important motifs of HIV-1 accessory genes. The study will assist in understanding HIV-1

accessory genes from South African HIV strains as they are potential new anti-HIV therapeutic targets as seen with *vif* gene (Chiu and Greene, 2008).

CHAPTER TWO

2. LITERATURE REVIEW

2.1 Historical background of HIV

Some important events had occurred that led to the discovery of HIV and AIDS. Available data suggest that this pandemic started from the 1970s (www.avert.org/aids-history-86.htm). The awareness of HIV infection started in 1981 when there was emergence of Kaposi sarcoma and pneumocystis carinii pneumonia (now known as Pneumocystis jiroveci). In 1981 both Kaposi sarcoma and pneumocystis carinii pneumonia were rare diseases. HIV infection was also highlighted with opportunistic infections emerging in high numbers and a more aggressive form (Friedman-Kien, 1984; www.avert.org/aids-history-86.htm).

In 1982 the disease did not have a universal name, with different parties referring to it differently. The Centers for Disease Control and Prevention (CDC) referred to the disease as lymphadenopathy, which was the disease presenting in those patients infected with HIV. The lancet called it gay compromised syndrome as it first appeared in gay men. Names such as GRID (gay related immune deficiency), AID (acquired immunodeficiency disease) and gay cancer were also used. Some of these names like GRID and gay cancer were found to be irrelevant as later in 1982 the disease was found to infect Haitians and Haemophiliacs. The same year the acronym AIDS (acquired immunodeficiency syndrome) was suggested and by the end of that year the name was in use in news papers and scientific journals (www.avert.org/aids-history-86.htm).

Report of AIDS in women in 1983 suggested that it can be transmitted through heterosexual sex. In the same year a new virus named lymphadonopathy associated virus (LAV) was isolated by doctors in France. Dr Robert Gallo from United State Health isolated the virus that causes AIDS, named human T lymphotropic virus-III (HTLV-III). By 1985 it was comfirmed that LAV and HTLV-III was the same type of virus. In 1986 there was still a

disagreement between the French and Gallo's group about which name to use, French LAV or Gallo's group HTLV-III. The International Committee on the Taxonomy of Viruses took a decision to drop the two names and assigned a new name, HIV (www.avert.org/aids-history-86.htm).

2.2 Classification of HIV strains

HIV is a single stranded RNA virus with enveloped icosahedral capsid and belongs to a family *Retroviridae*. It is further classified into a genus called *Lentiviruses* (lenti-, meaning slow). There are two types of HIV: HIV-1 and HIV-2 and their route of transmission is through sexual intercourse, blood and from mother to child during delivery or breast feeding. The two HIV types are the causative agents of AIDS. HIV-2 appears to be limited to West Africa (Papathanasoulos et al, 2003; Perrin et al, 2003).

HIV-1 is further divided into three groups; group (M=major), group (O=outer) and group (N=non-M and non-O). HIV-1 is the cause of around 90% of HIV infections globally. Group O is mainly found to be restricted to west central Africa and a rare group N was discovered in Cameroon. Group M is subdivided into 9 subtypes, A, B, C, D, F, G, H, J and K (Perrin et al, 2003). Subtypes A and F are further subdivided into A1-A3 and F1-F2 respectively. There are recombinant forms of HIV-1 and are referred to as circulating recombinant forms (CRFs) (Figure 2.1). HIV-1 subtype C is found to be the dominating strain in South Africa and other southern Africa countries (Papathanasoulos et al, 2003).

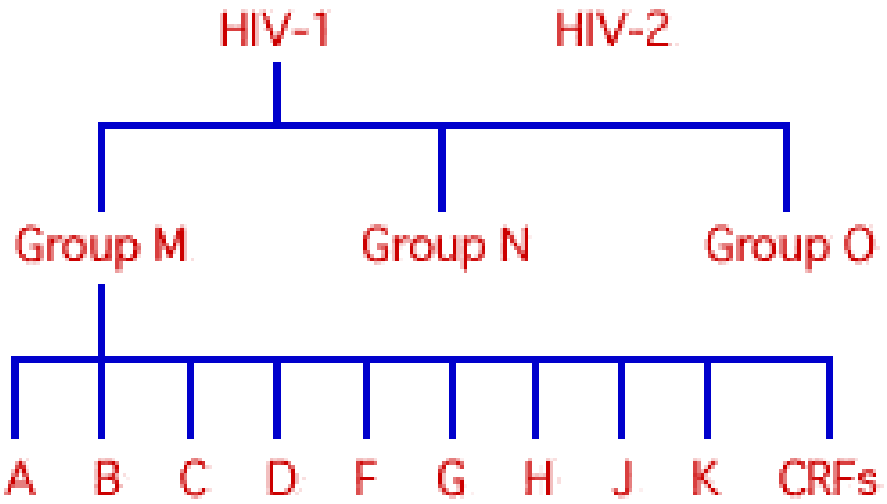


Figure 2.1: HIV groups and subtypes classification. Adopted from (www.avert.org/media/images/subtypes.gif) accessed 12 May 2008

2.3 HIV structure

2.3.1 The virion

HIV virion is about 100nm in diameter with the envelope as the outer most part of HIV comprising the spikes. The spikes comprised glycoprotein 120 (gp120) and glycoprotein 41 (gp41). The gp120 and gp41 are enzymatically cleaved from the precursor gp160. The gp120 have three surface domains and gp41 has three transmembrane domains. Binding of the virus to the target cell is facilitated by gp120 and gp41, these two glycoproteins mediate fusion of the viral membrane and cellular membrane (Chan, 1997). Following the envelope is the matrix protein p17, derived from the gag precursor polyprotein p55. It functions in nuclear exportation of unspliced viral RNA, incorporation of the HIV-1 envelope into virions and in particle assembly (Mammano, 1995).

The viral core is a bullet shaped structure formed by protein p24. It houses the genome of the virus and three important enzymes; reverse transcriptase,

integrase and protease (Figure 2.2). Inside the viral core are the diploid RNA molecules encoding all nine genes of HIV.

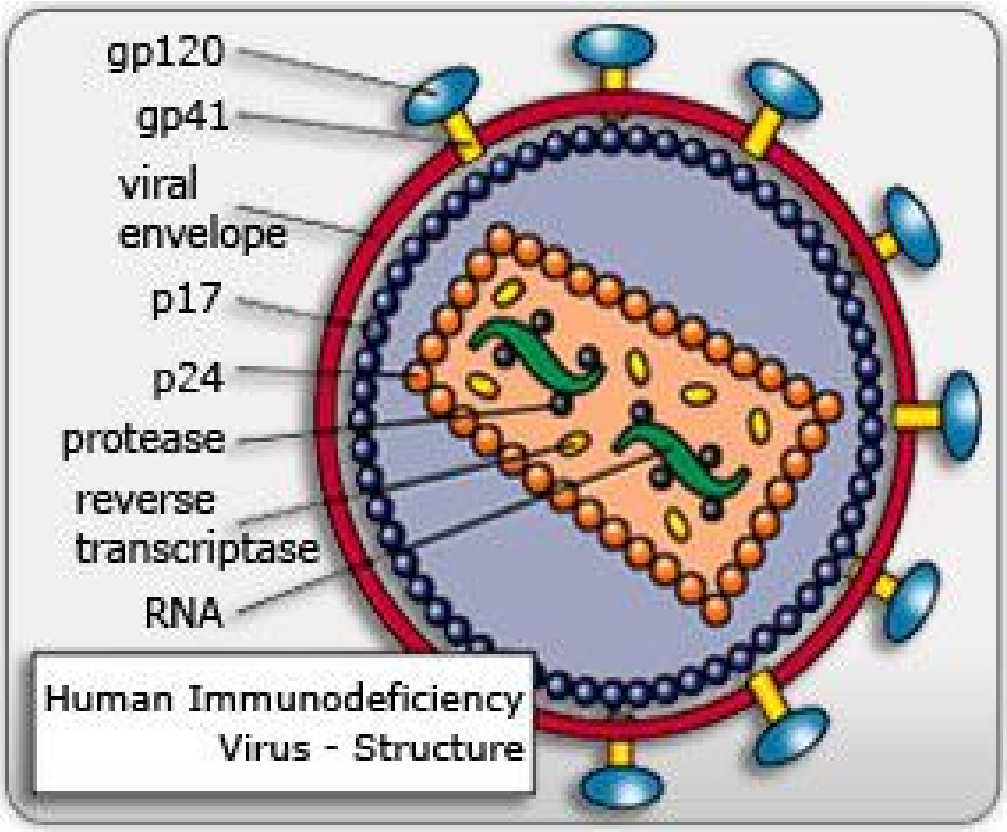


Figure 2.2: HIV structure. Adopted from (http://www.avert.org/virus.htm) accessed 04 March 2009

2.3.2 The viral genome

HIV-1 has a 9749 nucleotides long genome coding for nine genes. The long terminal repeats (LTR) flank the HIV-1 genome in both ends. Three genes (*gag, pol* and *env*) contain information needed to make structural proteins. The two genes (*tat* and *rev*) code for regulatory proteins. The remaining four genes (*nef, vif, vpr* and *vpu*) code for accessory proteins. The accessory proteins control the ability of the virus to infect a cell, and thus are responsible for viral pathogenesis. HIV-1 accessory genes (*vpr, vpu* and *vif*) are located in between the *pol* gene and *env* gene (Figure 2.3) (Hunt, 2008).

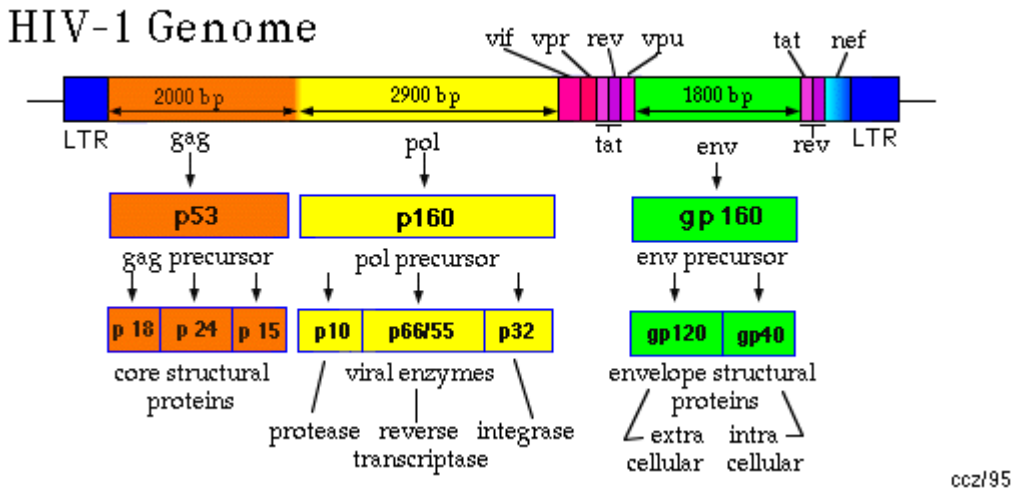


Figure 2.3: HIV genome. Adopted from (http://www.yale.edu/bio243/HIV/genome.html) accessed 12 May 2008

2.3.3 HIV genes and their functions

2.3.3.1 *Gag* (Group Specific Antigen)

Gag gene encode the *gag* precursor. The *gag* precursor is cleaved by viral protease to produce internal structural proteins; matrix, capsid, nucleacapsid and the three peptides (p1, p2 and p6) (Henderson et al, 1992).

2.3.3.2 *Pol* (Polymerase)

Pol gene encodes three enzymes; protease, reverse transcriptase and integrase. Protease functions to cleave gag and gag/pol polyproteins. Reverse transcriptase functions to convert the viral RNA to DNA while integrase integrates the viral DNA into the host genome (Chan et al, 1999).

2.3.3.3 *Env* (Envelope)

The *env* gene encodes the *env* precursor, gp160 which in turn is cleaved to gp120 and gp41. These glycoproteins (gp120 and gp41) are involved in viral binding and fusion to the target cell, respectively. HIV-1 binding requires the

viral gp120 and host cell CD4 receptors. HIV-1 fusion with the target cell requires a conformational change, which is achieved by coreceptors CXCR4 and CCR5 (Weng and Weiss, 1998; Dimitrov, 1997).

2.3.3.4 *Tat* (Trans-Activator of Transcription) and *rev* (Regulator of virion)

The *tat* and *rev* are regulatory genes, producing tat and rev proteins respectively. Tat and rev proteins play a positive role in promoting viral gene expression. In the absence of high level of *tat*, the host RNA polymerase initiates transcription properly at the LTR promoter. The transcription is usually prematurely terminated. This leads to the production of short, dead-end transcripts. *Tat* is a transcriptional activator that acts at a sequence near the beginning of the viral mRNA called transactivation response element (TAR). It recruits cellular proteins to the transcribing RNA polymerase that prevents premature termination of the proviral genome (Kuppuswamy et al, 1989).

Rev protein splices the mRNA (unspliced cellular transcripts are retained in the nucleus). The viral proteins that are made from spliced mRNAs are *tat, rev* and *nef*. The expression of vif*,* vpr and vpu proteins, are translated from the unspliced genome. The env*,* gag and pol polyproteins, are all made from single unspliced transcripts. *Rev* is necessary for transportation of incompletely spliced RNAs to the cytoplasm. The *rev* bind to a site on the viral RNA within the *env* gene called the rev-responsive element (RRE) to accomplish the transportation. The RNA bound *rev* is then exported to the cytoplasm through nuclear pore by normal machinery responsible for protein exportation from the nucleus (Dillon et al, 1990; Ryan and Ray, 2004).

2.3.3.5 *Vif* (Virion Infectivity Factor)

Vif is a protein of 192 amino acid and is essential for efficient viral replication in natural target cells. It functions to prevent action of APOBEC3G (cellular

inhibitor of HIV replication). Without *vif*, APOBEC3G is packaged into virions as a stable complex with the viral core. This affects the deamination of viral cytidines to uracil during the subsequent round of viral replication, thus leading to production of non-functional virions. Thus, *vif* bound APOBEC3G is degraded in the proteosome, permitting complete viral replication, thus increasing the infectivity of the HIV-1 (Mehle et al, 2007; Carr et al, 2008; Wang et al, 2005).

2.3.3.6 *Vpr* (Virion Protein R)

Vpr is an accessory protein of 96 amino acid found in HIV-1 and other *lentiviruses*. When this protein is expressed it causes G2 cell-cycle arrest, before the integration step of the viral DNA genome. It can do that in three ways; (i) by interacting with 14-3-3σ proteins, causing the hyperphosphorylation cytoplasmic localization of Cdc25c, stopping the cell-cycle at the G2 stage; (ii) by interacting with Rad3-related protein (ATR), which results in Chk1 phosphorylation and the subsequent inhibition of cdc25c; and (iii) by transactivating the *surviving* gene promoter, increasing its expression during G2. It can still induce G2 cell-cycle arrest while packaged into a virion. *Vpr* can also induce apoptosis, thus weakening the immune system by inducing CD4+ T cell apoptosis, and contributing to the pathogenesis of the HIV-1 (Le Rouzic and Benichou, 2005; Andersen et al, 2008). A study done in SIV, showed that *vpr* evolution is inversely proportional to the rate of disease progression (Richard et al, 2007).

2.3.3.7 *Vpu* (Virion Protein U)

V*pu* gene codes for an 81 amino acid protein. The *vpu* functions to enhance virion release by ion channel activity and down regulation of CD4+ expression from the surface of infected cells. The association of *vpu* with viral release was seen in an in vitro study with viruses lacking the functional *vpu*. The viral particles accumulated on the cell surface where they remain loosely attached, and the viral particles were released after vigorous shaking of the culture. This

suggested that *vpu* was involved in the last step of viral release. In addition, *vpu* expression has been associated with a reduction in syncytia formation of infected cells (Klimkait et al, 1990; Schubert and Strebel, 1994; Estrabaud et al, 2007). Vpu prevents the transport action of synthesized CD4 molecules for ubiquitin dependent proteosomal degradation at the endoplasmic reticulum (ER) through the affinity of the *vpu* C-terminal domain for β-transduction repeats-containing protein (βTrCP). The *vpu* inhibits βTrCP, impairing the proteosomal degradation of cellular Skp1-cullin-F-Box (SCF)-βTrCP substrates. This results in accumulation of substrates like β-catenin, IκBα, and ATF4, which have a role in innate immunity, cancer and several autoimmune disorders (Bell et al, 2007; Bour and Strabel, 2003; McCormick-Davis et al, 2000).

2.3.3.8 *Nef* (Negative Regulatory Factor)

Nef produces a protein that enhances virus production and virion infectivity. It also interferes with immune recognition of infected cells (Ryan and Ray, 2004). *Nef* down regulates the expression of cell surface CD4 receptors (Benson et al, 1993).

2.3.3.9 LTR (Long Terminal Repeat)

The terminal redundancy termed long terminal repeats (LTR) is a cis-acting sequence in HIV-1 that flank the HIV-1 genome. LTR is generated in the process of reverse transcription and LTR sequences are recognized by host cell machinery for transcription. There are two LTR, the 5' end LTR and 3' end LTR, with both serving different functions. The 5' LTR serve to initiate viral genome transcription and 3' LTR serves for formation of viral transcripts (cleavage and polyadenylation). Also 3' LTR has transcriptional activities but far less than that of 5' LTR. This was expected since both LTRs are identical in sequence. Both 5' LTR and 3' LTR function in integration of viral DNA into host genome (Klaver and Berkhout, 1994; Al-Harthi et al, 1998). HIV-1 LTR has three functional regions U3, repeat (R) and U5, with U3 region containing

basal, enhancer, modulatory promoter elements and negative regulatory element (Pereira et al, 2000). Transcriptional activities of the LTR are specifically in the U3 region, deletion of U3 region lead to no generation of transcripts (Choi and Faller, 1995).

2.4 Replication of HIV

The HIV replication involves a series of steps. The most noticeable steps are; 1) the attachment and entry of the virus; 2) reverse transcription of the viral genome; 3) integration of the viral genome in the host genome; 4) replication and transcription of the viral genome; 5) the viral assembly and release of mature virus. All these steps are important for viral replication (Figure 2.4).

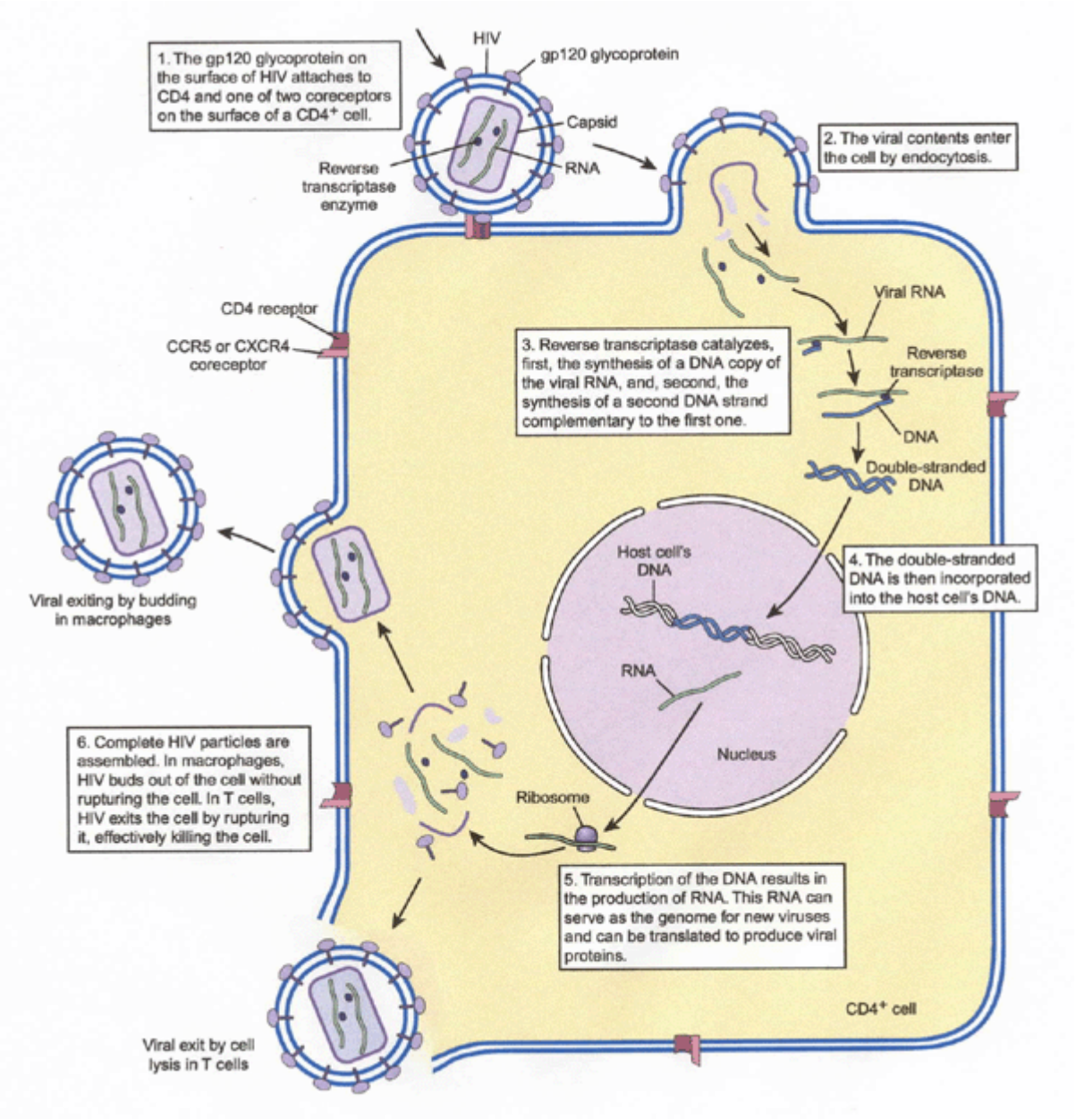


Figure 2.4: Schematic representation of HIV-1 replication. Adopted from (http://www.txtwriter.com/Backgrounders/Aids/HIVLifecycle.gif) accessed 12 January 2009

2.4.1 Attachment and entry

HIV-1 viral proteins gp120 and gp41 facilitate binding and fusion on the target cells, respectively (Chan, 1997). The mostly targeted cells are CD4 T lymphocytes and macrophages. Viral protein gp120 binds CD4 on T lymphocytes thus exposing binding site for coreceptors CCR5 or CXCR4. The

binding of coreceptor results in a conformational change in viral transmembrane protein gp41, leading to fusion of viral and target cell membranes (Lee et al, 1999). Both coreceptors CCR5 and CXCR4 can be used by macrophages (Simmons et al, 1998). There is not much known about the mechanism of HIV-1 viral core uncoating, but an in vitro study showed that lysate from activated CD4 lymphocytes can induce uncoating of HIV-1 viral core (Auewarakul et al, 2005). Since the nef protein is associated with proton pump; it is thought to mediate uncoating like that of influenza's M2 protein. *Nef* protein promotes an influx of hydrogen ions resulting in the change in the local pH and that results in uncoating (Lu et al,1998; Takeda, 2002).*Vif* binds to APOBEC3G (cellular inhibitor of HIV replication). The *vif* bound APOBEC3G is degraded in the proteosome, permitting complete viral replication, thus increasing the infectivity of the HIV-1 (Bell et al, 2007; Carr et al, 2008).

2.4.2 Reverse transcription

Viral RNA is converted to DNA by reverse transcriptase (RT) producing the first DNA strand and then a complementary strand resulting in double stranded DNA. Reverse transcription is initiated using a molecule of tRNA that bind to the primer binding site (pbs), and DNA synthesis start from 3' to 5' end of the RNA generating a DNA/RNA hybrid. The RNase H degrade the RNA portion of the hybrid, this generates a DNA fragment called minus-strand strong stop DNA. The minus-strand strong stop DNA then jumps from 5' to the 3' end of the genome and the process is called first strand transfer. Plus strand is synthesised using a primer fragments of RNA remaining from minus-strand synthesis. The RNase H removes tRNA from pbs allowing second strand synthesis (Lavigne et al, 2001; Freed, 2001).

2.4.3 Integration

The linear double stranded (ds) viral DNA enters the nucleus through the nuclear pore as a pre integration complex (ds DNA bound reverse transcriptase, matrix, *vpr* and integrase). In the nucleus only integrase remains attached to the ds DNA and the ds DNA is integrated into the host DNA by integrase. First integrase cleaves off nucleotides at the 3' end of both the strands generating a molecule of double stranded DNA with 3' recessed ends. In the nucleus integrase cleaves host DNA making staggered cut. The recessed end of viral DNA is joined to host cleaved DNA and integration is complete after the cellular repair enzymes filled the gaps (Freed, 2001).

2.4.4 Replication and transcription

HIV-1 transcription is initiated in the LTRs which have cis-acting elements required for RNA synthesis. U3/R junction is where transcription is initiated in the LTR and U3 has elements needed to direct binding of RNA polymerase II. *Tat* acts at a sequence near the beginning of the viral mRNA called TAR to recruit cellular proteins to the transcribing RNA polymerase, this prevents premature termination of the proviral genome and this greatly increases RNA synthesis (Ryan and Ray, 2004; Freed, 2001).

2.4.5 Viral assembly and release

After elongation spliced mRNA encoding *nef*, *tat* and *rev* leaves the nucleus to the cytoplasm. The large unspliced mRNA remains in the nucleus, and they encode the structural enzymatic accessory proteins and viral genome needed in viral assembly. *Rev* is needed for transportation of this incompletely spliced RNA to the cytoplasm and this is accomplished by *rev* binding to a site specific on the viral RNA within the *env* gene called the rev responsive element (RRE). RNA bound *rev* is then exported to the cytoplasm (Greene and Peterlin, 2002; Ryan and Ray, 2004).

In the cytoplasm protease cleaves this large mRNA into their small functional mRNA. Viral assembly take place and *vpu* enhance virion release by ion channel activity of infected cells and virions are released by budding (Bour and Strabel, 2003).

2.5 Routes of HIV transmission

HIV is transmitted through the mucous membrane by direct inoculation with infected blood or blood products. Mucosal transmission is mainly through sexual intercourse, both heterosexual and homosexual intercourses, and through breastfeeding. Direct inoculation with infected blood or blood products, through injecting drugs with contaminated needles and blood transfusion with contaminated blood. Any skin breaching activities are thought to be potential routes for HIV transmission (Garland et al, 2006). HIV infects mainly CD4 cells with CCR5 and CXCR4 co-receptors which aid in viral attachment and fusion with host cell (Chan et at, 1997).

2.5.1 Sexual transmission of HIV

Transmission of HIV-1 through sexual intercourse was reported to be the most common route; in the 1980s about 64% of HIV-1 infections were from homosexual intercourse. In the 1990s things changed with three third of HIV-1 infections being from heterosexual intercourse (Evans et al, 1992).

2.5.2 Transmission of HIV from mother to child

Another troubling route of HIV-1 transmission is shown to be from mother to child and this can be antepartum, peripartum, intrapartum or postnatal transmission. A study showed that women who breastfeed for more than two years, increase the chances of transmitting HIV to their children two folds. Postnatal transmission is worrying in sub-Saharan Africa as most women

depend on breastfeeding and they breastfeed for more than two years (WHO, 2004).

2.5.3 Direct inoculation with HIV infected blood and blood products

Direct inoculation with infectious materials may occur in many different ways. The common way being with intervenous drug who share needles. It occurs also in blood transfusion, health care worker who may experience needle stick injury and to a lesser extent in skin piercing (Correa and Gisselquist, 2006)

2.6 Immune response

About 4 to 11 days after an infection with HIV, the viral load increases sharply with increased numbers of CD8 T lymphocytes and a decrease in the numbers of CD4, and this continues for a few weeks. Then CD8 T lymphocytes returns to near normal and viral load decreases.

The viral load is controlled by CD4 T lymphocytes and it remained low: while the numbers of CD4 T lymphocytes continuously drops but CD8 T lymphocytes remains near normal. With decreasing CD4 T lymphocytes as years go by one starts experiencing opportunistic infections.

After years of infection with HIV the CD8 and CD4 T lymphocytes count decreases and the viral load increases. Until the immune system is not functioning and the patient reaches the AIDS state (Figure 2.5) (Hunt, 2009).

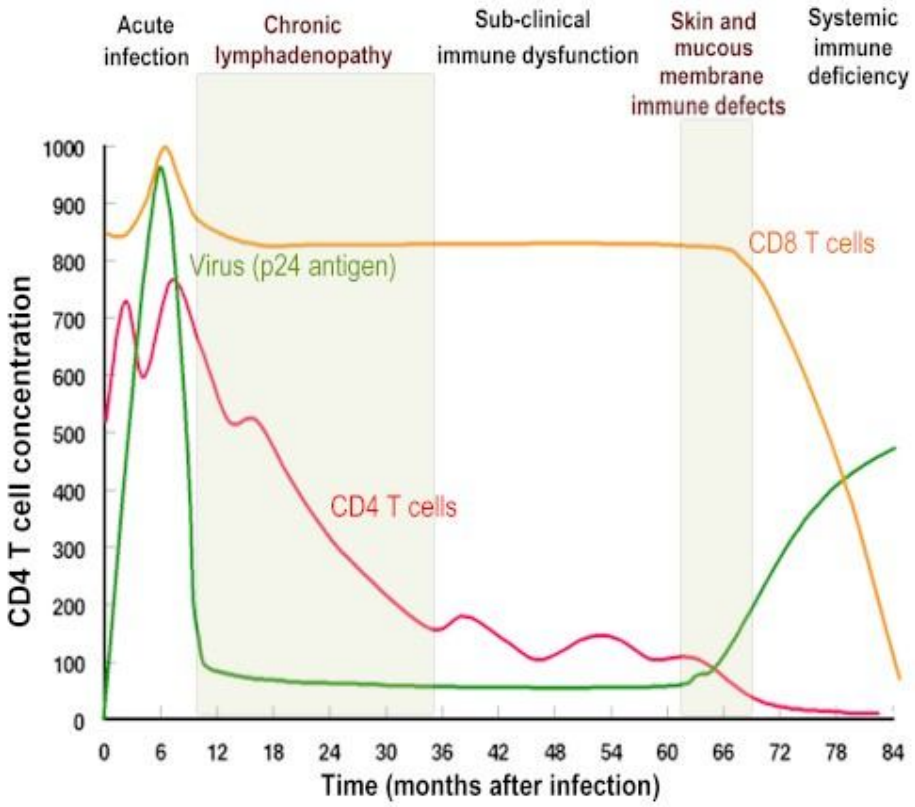


Figure 2.5: HIV-1 disease stages. Adopted from (http://pathmicro.med.sc.edu/lecture/hiv_time_course2.jpg) accessed 12 January 2009

2.7 Diagnosis of HIV

There are three main tests for detection of HIV infection, the antibody test, the antigen test and HIV nucleic acid tests. The antibody tests which are well established are enzyme linked immunosorbent assay (ELISA) and the western blot, were antibodies to HIV are detected. The antigen test used is the p24 test, it detect viral protein p24 which can be detected even before the appearance of antibodies. HIV nucleic acid tests can detect HIV RNA or DNA. The most preferred specimens is serum or plasma, but there are some antibody tests that can also detect HIV in whole blood, saliva/oral fluids, urine and dried blood spot (WHO, 2005).

2.7.1 Screening tests

2.7.1.1 Antibody detection

Rapid anti-HIV tests are available and are used mainly for screening of HIV infection. Some of these rapid tests have high sensitivity and specificity. Since false positive cases were seen from these rapid tests, two to three different ELISA tests are needed to confirm the positive results. The setback is that HIV antibody tests cannot be used in newborns born to HIV positive mothers, when the newborns are less than 18 months old. That is because the anti-HIV antibodies in newborn may be from the mother and they usually persist to up to 18 months in infants. The other reason is that infants immune system is not well developed and antibody to HIV will not be developed. HIV antigen and HIV nucleic acid tests are used in infants born to HIV positive mothers. Antibody testing will give negative test results in patients in window period. Window period is the period between infection and development of antibodies. HIV antibodies usually take three to four weeks to develop. To avoid false negatives results patients are recalled to do another antibody test after three months if the first antibody test results came negative (Aboud et al, 2006; WHO, 2005).

2.7.1.2 Antigen detection

The p24 antigen test employs detection of p24 protein and is useful diagnostic test that reduces the window period. It is mainly used in infants born to HIV positive mothers. Antigen test also have false negative results and false positive results due to the formation of antigen-antibody complex (WHO, 2005).

2.7.2 Confirmatory test

2.7.2.1 HIV nucleic acid detection

There are two kinds of HIV polymerase chain reaction (PCR); HIV DNA PCR and HIV RNA PCR. The HIV DNA PCR test is a molecular assay that determines the presence of HIV in peripheral blood mononuclear cells. HIV DNA PCR is used for early detection of HIV infection and detection of HIV in infants born to HIV positive mother. HIV RNA PCR test is mainly used in screening donated blood (WHO, 2005).

2.8 Treatment of HIV

2.8.1 When to start HIV treatment

WHO recommended that clinical and immunological assessments are used in making a decision on when to start antiretroviral therapy (ART). The assessment of patients' readiness to start therapy and understanding of its outcomes is also looked at, as patients need to understand that treatment is a lifelong commitment. It is also recommended that patients need to understand that adherence to treatment is important in reducing occurrence of resistance and that toxicity may occur during the therapy.

Table 2.1: Classification of HIV associated clinical disease (adopted from WHO, 2006).

Classification of HIV associated clinical disease	**Symptoms**	**WHO clinical staging**
Asymptomatic		1
Mild	Recurrent upper respiratory infections. Muco-cutaneous diseases: Seborrhiec dermatitis, prurigo, fungal nail infection, oropharyngeal ulcers and angular cheilitis.	2
Advanced	Pulmonary TB, severe bacterial infection, vulvovaginal candidiasis, oral candidiasis, oral hairy leukoplakia and chronic diarrhoea of more than one month.	3
Severe	HSV persistent for more than one month. Disseminated fungal infections: histoplasmosis. Candidiasis of the oesophagus, trachea etc. HIV wasting syndrome. CMV outside the liver: spleen or nodes.	4

According to WHO ART is initiated before the onset of symptoms or infection with opportunistic diseases. The immunological status (CD4 count) of patients is evaluated to assess if patients should start therapy. The patients with CD4 count below 200 cell/mm^3 are treated irrespective of clinical staging. The patients with CD4 count between 200-350 cells/mm^3 are treated to keep the CD4 count above as 200cell/mm^3 and those with CD4 above 350 cells/mm^3 do not initiate treatment (WHO, 2006).

Reports exhibit that it is importance to start treatment when it is necessary. The necessity of starting treatment is according to clinical staging and immunological assessment (CD4 test). If immunological assessment is not

available and the patient is in clinical stage 1 or 2, that patient does not need to start treatment. But if CD4 test is available and the CD4 cell count is below 200 cells/$mm^{3,}$ then treatment is initiated. If the patient is clinical stage 3 and CD4 test is not available, treatment is initiated. When the patient is in clinical stage 3 and CD4 test is available, treatment is start when CD4 cell count is below 350 cells/mm^3. When a patient is in clinical stage 4 treatment is initiated irrespective of the immunological assessment.

2.8.2 ARV's for HIV treatment

ART is a lifelong therapy since ARV's are not meant to cure but just to lower HIV viral load in the body. Highly Active Antiretroviral Therapy (HAART) is a combination therapy with three or more ARV's used at the same time. They are mostly given in combinations to reduce occurrence of resistance (WHO, 2006).

ARVs are categorized into the five groups targeting different sites and steps in HIV replication:-

1. Nucleoside Reverse transcriptase Inhibitors (NRTI's) are analogues interfering with the action of reverse transcriptase, which is needed to convert HIV RNA to cDNA in the first step of HIV replication. The first NRTI was approved for use in 1987.

2. Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTI's) directly inhibit the function of reverse transcriptase and stop the replication of HIV. The first NNRTI was approved for use in 1997.

3. Protease inhibitors (PI's) function to inhibit the protease enzyme in HIV replication. They inhibit cleaving of large non-functional mRNA's to their small functional mRNA. The first ARV in this group was approved for use in 1995.

4. Fusion or Entry inhibitors prevent binding or entry of HIV into host cell. The first ARV in this group was approved in 2003.

5. Integrase Inhibitors, interfere with integrase which functions to integrate HIV cDNA into host genome for viral replication. The first integrase inhibitor was approved for use in 2007 (http://www.avert.org/treatment.htm).

2.8.3 Choices in ARV groups

There is quite a number of ARV's available for the treatment of HIV infection. Table 2.2 summarises some of the ARV's approved by Food and Drug Administration (FDA).

Table 2.2: Choice of ARV (adopted from WHO, 2006; Johnson et al, 2008).

NRTI's	Lamivudine (3TC), emtricitabine (FTC), zidovudine (AZT), tenofovir (TDF), abacavir (ABC), didanosine (ddI) and stavudine (d4T)
NNRTI's	Nevirapine (NVP), etravirine and efavirenz (EFV)
Protease inhibitors	Ritonavir, atazanavir, darunavir, indinavir, lopinavir, nelfinavir, saquinavir, tipranavir and fosamprenavir
Integrase inhibitors	Raltegravir and elvitegravir
Entry inhibitors	Enfuvirtide and maraviroc

2.8.4 Combination recommendations

ARV's are used in combination to avoid development of resistant strains. There are clear guidelines for combination therapy from WHO (Table 2.3).

Table 2.3: WHO recommended combinations (adopted from WHO, 2006).

First-line regimen		**Second-line regimen**	
		RTI component	**PI component**
Standard Strategy	AZT or d4T + 3TC+NVP or EFV	ddI + ABC or TDF + ABC or TDF+3TC(± AZT)	Low-dose ritonavir
	TDF + 3TC+NVP or EFV	ddI + ABC or ddI +3TC (± AZT)	
	ABC + 3TC+NVP or EFV	ddI + 3TC (± AZT) or TDF+3TC(± AZT)	
Alternative Strategy	AZT or d4T + 3TC+TDF or ABC	EFV or NVP ± ddI	

2.9 Prevention of HIV transmission and infection

2.9.1 Sexual transmission

It is shown that abstaining from sexual intercourse, remaining faithful to your uninfected partner and by the use of condoms whenever engaging in sexual intercourse reduces the risk of contracting HIV through sexual intercourse. Circumcision is shown to reduce the risk of acquiring HIV. A study conducted in South Africa showed that circumcised men are 60% less likely to acquire HIV infection as compared to uncircumcised men (Auvert et al, 2005). Other studies done in Kenya and Uganda also show that there is reduction in HIV acquisition, with the Kenyan study showing 53% reduction and Uganda study showing 48% reduction (Rennie et al, 2007). Counselling and health education to individuals and the community about safe sexual behaviour is needed to reduce the risk of HIV transmission (WHO, 2009).

2.9.2 Preventing direct inoculation

The use of new disposable needles and new alcohol swabs are recommended every time to reduce the risk of direct inoculation with HIV. It is reported that blood and blood products to be transfused are screened for HIV infection (WHO, 2009). A report showed that post exposure prophylaxis (PEP) is a useful measure for HIV prevention, mainly used in health care setting. PEP is given to health care workers and patients who are exposed to potentially HIV positive specimen (blood or blood products). The ARVs for PEP are given within 72 hours post exposure. The ARV regimen should have two NRTI's, if resistance is suspected a protease inhibitor is added and given for 28 days as recommended by WHO (WHO, 2009).

2.9.3 Preventing mother to child transmission

Prevention of mother to child transmission is done by treatment with ARV drugs, delivery by caesarean section and avoiding breastfeeding. WHO reported that nevirapine is given to the mother on the onset of labour and to the infants after delivery. Combination treatment with AZT and nevirapine is used, as recommended by WHO table 2.4 (WHO, 2009).

Table 2.4: WHO guidelines for (preventing mother-to-child transmission) PMTCT drug regimens in resource-limited settings.

	Pregnancy	**Labour**	**After birth: mother**	**After birth: infant**
Recommended	AZT after 28 weeks	single dose nevirapine; AZT+3TC	AZT+3TC for seven days	single dose nevirapine; AZT for seven days
Alternative (higher risk of drug resistance)	AZT after 28 weeks	single dose nevirapine	-	single dose nevirapine; AZT for seven days
Minimum (less effective)	-	single dose nevirapine; AZT+3TC	AZT+3TC for seven days	single dose nevirapine
Minimum (less effective; higher risk of drug resistance)	-	single dose nevirapine	-	single dose nevirapine

2.10 Epidemiology of HIV

Almost 28 years after the discovery of HIV, it was estimated that around 33 million people were living with HIV/AIDS in 2007 and around 2.7 million were newly infected. Death caused by AIDS related illness was estimated to be 2 million in 2007 which is almost equal the number of people newly infected, as shown in table 1. The number of people living with HIV/AIDS has increased from 8 million in 1990 to 33 million in 2007. It is also shown that the rate of infection has dropped but still the number of people living with HIV is going up. Sub-Saharan Africa being the worst affected worldwide, with about 68% of adults living with HIV, 90% of the world's HIV infected children and 76% of all AIDS related deaths in 2007 (http://www.avert.org/worldstats.htm). In South Africa, it was estimated in 2008 that 5.2 million people were HIV positive and 250,000 people died of AIDS related illnesses, bringing an estimate of around 1000 deaths every day. The prevalence of HIV infection in South Africa is around 11% (http://www.avert.org/worldstats.htm).

2.11 Genetic diversity of HIV

2.11.1 Cause of genetic diversity

Due to their similarities, HIV is believed to have originated from SIV some decades ago. It is believed that SIV crossed the species barrier leading to different clades, groups and subtypes. The phylogenetic relationship of HIV-1 and HIV-2 reveals that they are closely related to SIV. The genetic variants in different individuals are due to the replication adaptability of HIV. The heterogeneity of HIV is due to rapid viral turnover, a high viral burden and the error prone nature of RT which lacks the proofreading activity.

The nucleotide sequence variation between HIV-1 and HIV-2 is 50%. While the nucleotide sequence variation between groups is 30%. Group M is divided into nine subtypes and sequence similarity between the subtypes is 70%-90% (Tebit et al, 2007).

2.11.2 HIV recombination

The recombination of retroviruses has been observed for many years. For HIV recombination to occur, the cell must be infected by two genetically distinct HIV viruses. RT synthesizes the first DNA strand from RNA template. The RNA template is degraded by RNase H, the DNA act as template for the second DNA strand. RT switches from one genomic RNA onto the other, during the synthesis of the first DNA strand. The process is referred to as copy choice, this switching results in genetic recombination (Figure 2.6) (Ramirez et al, 2008).

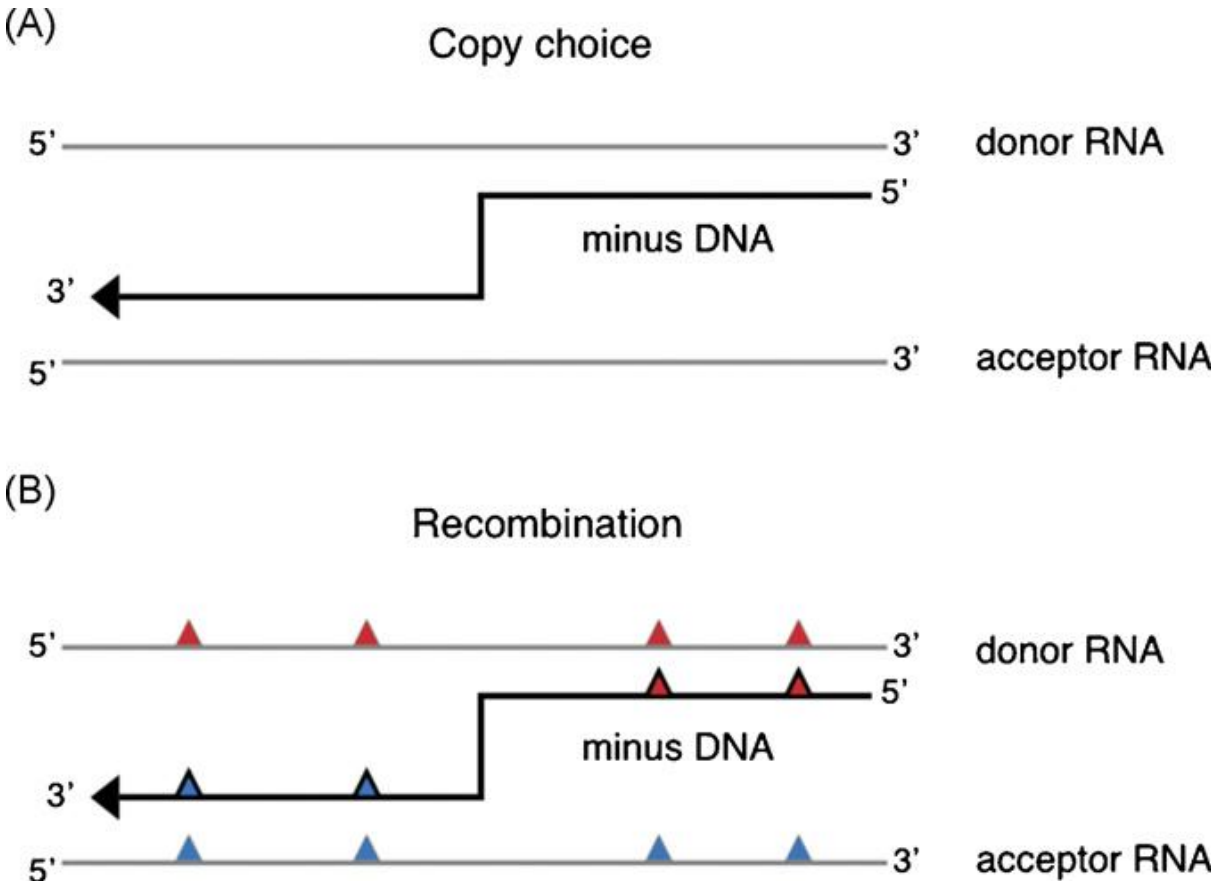


Figure 2.6: Schematic representation of copy choice (panel A) and of recombination generated by copy choice (panel B). The black arrow gives the direction of RT switching from one copy of genomic RNA onto the other. If the two RNAs present discordant residues, schematically indicated by red and blue triangles, copy choice results in recombination (panel B). Adopted from Ramirez et al, 2008

CRF's are formed from recombination of different viruses from different lineage within group M. So far there are at least 34 CRF described (Tebit et al, 2006; Ramirez et al, 2008). (Table 2.5)

Table 2.5: CRF's described (adopted from Tebit et al, 2007; Ramirez et al, 2008).

Name	Reference strain	Subtypes	Number of infections
CRF01_AE	CM240	A, E	1,303,689
CRF02_AG	IbNG	A, G	2,601,251
CRF03_AB	Kal153	A, B	6000
CRF04_cpx	94CY032	A, G, H, K, U	7500
CRF05_DF	VI1310	D, F	5920
CRF06_cpx	BFP90	A, G, J, K	100,332
CRF07_BC	CN54	B', C	108,000
CRF08_BC	GX-6F	B', C	126,900
CRF09_cpx	96GH2911	CRF02, A, U	14,100
CRF10_CD	TZBF061	C, D	17,640
CRF11_cpx	GR17	A, CRFO1, G, J	46,030
CRF12_BF	ARMA159	B, F	51,932
CRF13_cpx	96CM-1849	A, CRF01,G, J, U	27,260
CRF14_BG	X397	B, G	5550
CRF15_01B	99TH	CFR01, B	9860
CRF16_A2D	KISII5009	A2, D	NA
CRF17_BF	ARMA038	B, F	51,932
CRF18_cpx	CU76	A1,F, G, H, K, U	<1000
CRF19_cpx	CU7	A1, D, G	<1000
CRF20_BG	CB228	B, G	NA
CRF21_A2D	99KE_KER2003	A2, D	NA
CRF22_01A1	CM53122	CRF01, A1	NA
CRF23_BG	CB118	B, G	NA
CRF24_BG	CB378	B, G	NA
CRF25_cpx	02CM_1918LE	A, G, U	NA
CRF26_AU	Pending	A, U	NA
CRF27	Pending	Pending	NA
CRF28_BF	BREPM12609	B, F	51,932
CRF29_BF	BREPM16704	B, F	51,932
CRF30_0206	Pending	Pending	NA
CRF31_BC	Pending	B, C	NA
CRF32_06A1	EE0369	CRF06, A1	NA
CRF33_01B	Pending	CRF01, B	3631
CRF34_01B	Pending	CRF01, B	NA

A study on replication fitness suggests that CRF are more fit as compared to HIV-1 group M subtypes. The evolution of HIV leads to overcoming

transmission barriers, escaping the immune pressure and cause resistant to therapy.

2.11.3 HIV distribution

HIV strains are distributed unevenly throughout the globe. HIV-2 appears to be limited to West Africa. HIV-1 is distributed in different geographical regions. HIV-1 group M subtype A and D are prevalent in East and Central Africa. Subtype B is mostly prevalent in North America, Latin America and Australia. While subtype C is distributed in Southern Africa and India.

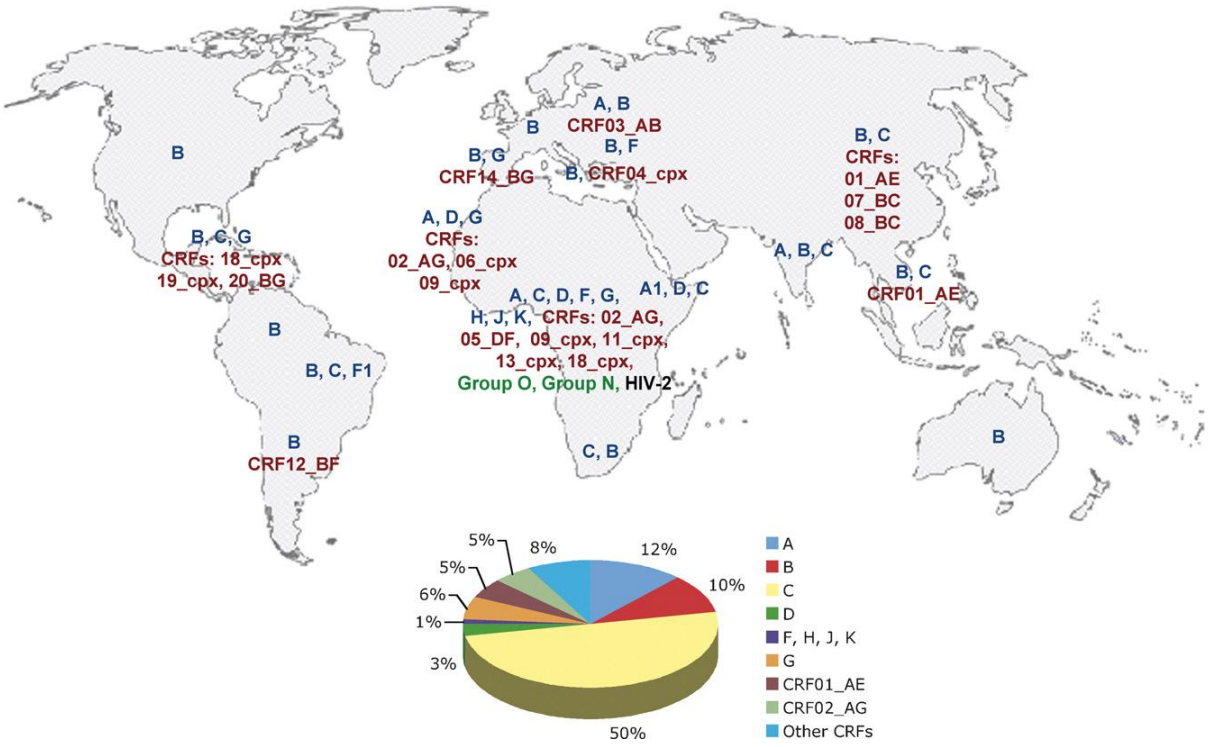


Figure 2.7: Geographic distribution of HIV. Adopted from Ramirez et al, 2008

2.12 Important regions and mutations in different genes

HIV is diverse, with different groups and subtypes. HIV genes have conserved regions which serve important functions. These regions are targeted for combating HIV infection. There are mutations that occur on HIV genes that have important implications. Some of these mutations are associated with drug resistance.

2.12.1 Important conserved regions

Table 2.6: Summary of conserved regions and their functions.

Gene	Conserved region	Function
Vif	88-EWRKKR-93	Thought to enhance expression of vif protein in the host cell.
	144-SLQYLA-149	A motif involved in APOBEC3 protein inactivation.
	H108, C114, C133, and H139	HCCH zinc motif: binds cellular cullin 5 which is subsequently used to degrade APOBEC3.
	W5, W11, W21, W38	Trytophans involved in recognition and suppression of APOBEC3G.
	90-RLRR-93	Nuclear localization inhibitory signal.
	161-PPLP-164	Vif dimerization site.
	88-EW-89	It enhances steady state expression of vif protein
Vpr	P35	It binds cyclophilin A
	71-FRIGCQHSRIGIL-84	Is the mitochondritoxic domain
	W54	It interacts with UNG2 for proteosomal degradation
Vpu	W28	Tryptophan involved in channel gating
	78-LRLL-81	Putative Golgi export signal
	57-DSNES-62	Casein kinase II phosphorylation site containing critical serines

2.12.2 Important mutations

Table 2.7: Summary of important mutations (adopted from Johnson et al, 2008).

Gene		**Mutation/s**	**Implication**
Pol	RT	K65R, L74V, Y115F, M184V	Cause resistance to abacavir
		K65R, L74V	Result in didanosine resistance
		K65R, M184V/I	Cause resistance to emtricitabine and lamivudine
		K65R, K70E	Result in resistance to tenofovir
		M41L, D67N, K70R, L210W, T215Y/F, K219Q/E	Cause resistance to stavudine and zidovudine
	Protease	V11I, V32I, L33F, I47V, I50V, I54M/L, T74P, L76V,I84V, L89V	Result in resistance of HIV to ritonavir boosted darunavir
	integrase	Y143R/H/C, Q148H/K/R, N155H	Result in resistance of HIV to raltegravir
Env		G36D/S, I37V, V38A/M/E, Q39R, Q40H, N42T, N43D	Results in HIV being resistant to enfuvirtide
Gag		A431V, K436R, I437V, L449F/V/H, S451T, R452S, P453L/A	Associated with protease inhibitors resistance
Vpr		R77Q	Associated with long term non progressors
		Q3R	Found in long term non progressors and associated with high viremia with no significant loss of CD4 lymphocytes
Vif		Mutation in SLQYLA motif (S147 and L148)	Results in less pathogenic virus

2.13 HIV vaccine

It has been more than 28 years since the HIV pandemic. To date there is no effective and safe vaccine to stop the spread of HIV. The predominant target on HIV that produces neutralizing antibodies in individuals infected with HIV is the envelope. Most HIV vaccine trial studies in human used subtype B using several adjuvant recombinant monovalent HIV envelope proteins, (e.g gp160 or the mature exterior portion gp120). The vaccine candidates induced neutralizing antibodies in almost all volunteers tested, but these antibodies did not have cross-reactivity against primary isolates of HIV (Johnston and Flores, 2001).

A study in goats demonstrated a vaccine that can prevent infection from mother to child through breast feeding. The vaccinated goats produced HIV antibodies in their milk. The antibodies were tested if they can neutralize the virus and they exhibited no neutralizing activities. This study was focused in preventing mother to child transmission through breast feeding in poor communities (Johnston and Flores, 2001; Dorosko et al, 2008). With another study showing that mutations in some conserved motifs (SLQYLA) of *vif* reduces the viral pathogenecity with low viral load and CD4 count maintained in almost the same levels. This study suggested that induced mutations targeting more than one functional and conserved region in *vif* may generate an effective immune responses and only permit for limited viral replication (Schmitt et al, 2009).

HIV-1 subtype C immune response was evaluated across different genes. The immune response was dominated by *nef* followed by *gag* then *pol*, thus making them the early stage of infection targets for controlling HIV infection (Masemola[a] et al, 2004). The immunogenicity of *vif*, *vpu* and *nef* was evaluated in mice. *Vif*, *vpu* and *nef* were evaluated as a fusion protein and the results showed that *vif*, *vpu* and *nef* induced an effective T helper-1 proliferation response. Also cytotoxic T lymphocyte (CTL) response was induced. *Vif*, *vpu* and *nef* genes were from subtype B and they also induced CTL response across clades D and E (Ayyavoo et al, 2000).

CTL recognition of *vif*, *vpr* and *vpu* was assessed and they showed CTL response against *vpr* (45%) was greatest followed by *vif* (33%) then *vpu* (2%). CTL epitopes were identified in *vif* and *vpr* (Altfeld et al, 2001; Addo et al, 2002). A study on structural and accessory genes CTL response showed that accessory genes have poor CTL response (Betts et al, 2002).

2.13.1 Challenges in developing an HIV vaccine

HIV-1 integrates in the long-lived memory CD4 T cells. This provide persistent reservoir of the virus that escape the immune system. The HIV infected memory CD4 T cells makes it hard to eradicate HIV infection.

There is lack of appropriate animal models to run HIV vaccine research trials. The only two animal models susceptible to HIV infection are chimpanzee (Pan troglodyte) and pigtail macaques (Macaca nemestrina). Both do not develop clinical symptoms and their viral loads remain low (Girard et al, 2006). HIV diversity is another one of the components that make development of a protective HIV vaccine hard.

The other challenge faced by HIV trails is that; the high risk populations are hard to recruit and retain in vaccine efficacy trials. The nfrastructure and trained investigators to conduct these trials in developing countries is limited. Social challenges such as volunteers safety has to be ensured. Others are financial and political support (Johnston and Flores, 2001).

2.13.2 Vaccines in clinical trials

Vaccines are effective way of controlling an infection. Besides the diverse nature of HIV, there are vaccines that are currently ongoing trials in different countries. Table 2.8 shows a number of trials taking place in the world.

Table 2.8: HIV vaccines ongoing trials (adopted from IAVI, 2008).

Phase	Countries	Trail number	Antigen (Clade)
III	Thailand (8)	RV 144	*Env* (B, E); *gag/pol*, *env* (B,E)
Iib	South Africa	HVTN 503 (Phambili)	pGA2/JS7 DNA/MVA/HIV62
II	South Africa (3), Uganda, Zambia	IAVI A002	*Gag*, PR, RT (C)
II	US (7), Brazil (2), South Africa (3), Haiti, Jamaica	HVTN 204	*Gag*, *pol*, *nef* (B), *env* (A,B,C); *pol* (B), *env* (A,B,C)
II	France (6)	ANRS VAC 18	5lipopeptides (CTL epitopes) from *gag, pol, nef* (B)
I/II	Thailand	NCHECR-AE1	*Gag, pol, tat/rev, env* (A,E)
I/II	Tanzania	HIVIS 03	*Env* (A,B,C), *gag* (A,B), RT (B), *rev* (B); *env* (E), *gag* (A), *pol* (E)
I/II	Kenya, Uganda, Tanzania	RV 172/WR 1218	*Gag*, *pol*, *nef* (B), *env* (A,B,C); *gag*, *pol* (B), *env* (A,B,C)
I/II	US (10)	HVTN 042/ANRS VAC 019	*Env*, *gag*, *pol*+CTL epitopes from *nef/pol* (B); CTL epitopes from *gag, pol, nef* (B)
I	Uganda	HPTN 027	*Env* (B,E)
I	US (3), Peru (2)	HVTN 064	*Env, gag, pol, vpu* (B); *gag, pol*, *vpr*, *nef* (A,B,C,D,F,G)
I	US (3)	HVTN 067	*Env, gag, pol, vpu* (B); *gag, pol, vpr, nef* (A,B,C,D,E,G)
I	Russian Federation	HVRF-380-131004	*Env, gag* (B)
I	India	IAVI D001	*Env, gag, tat, rev, nef,* RT (C)
I	Rwanda, Kenya	IAVI V001	*Gag, pol, nef* (B) *env* (A,B,C); *gag, pol* (B), *env* (A,B,C)

The numbers next to the countries represent number of those trails in that particular country.

2.14 Possible approaches for HIV intervention using accessory genes

HIV accessory genes were shown to be potential anti-HIV therapeutic targets. This is demonstrated by studies on *vif* gene. Zinc chaletor N,N,N,N-tetrakis-(2pyridylmethyl) ethylenediamine (TPEN) affects the ability of *vif* to recruit Cul5 to the E3 ligases. The viruses treated with TPEN inhibited *vif* mediated APOBEC3G degradation, making HIV sensitive to APOBEC3G activity (Xiao et al, 2007).

Another study showed that D128K mutation in human APOBEC3G confers resistance to the activity of *vif* and the mutant human APOBEC3G inhibits HIV replication (Xu et al, 2004). A small molecule RN-18 increases concentration of APOBEC3G and its uptake by budding virions. RN-18 molecule affects the action of vif protein by increasing APOBEC3G (Nathans et al, 2008).

CHAPTER THREE

3. MATERIAL AND METHODS

3.1 Ethical approval

The approval to conduct this study was sought from the Medunsa Research and Ethics Committee (MREC), and approved as project number MREC/P/136/2008: PG.

3.2 Study design

The study was an exploratory study using stored (-70ºC) HIV positive plasma which were previously used in another project (MCREC/P/01/2007:PG). In the previous project (MCREC/P/01/2007:PG), the samples were sequenced for both *gag* and *env* genes (Musyoki, 2009, Genetic characterisation of circulating HIV strains in Pretoria and surrounding areas, MSc dissertation, University of Limpopo, Medunsa Campus). For the purpose of this study, the same samples were sequenced and characterised in the *pol* and accessory (*vif*, *vpr* and *vpu*) genes in order to obtain near full length sequences of the HIV isolates from Pretoria region and compare the genetic diversity of HIV-1 subtype C between accessory (*vif*, *vpr* and *vpu*) and structural (*pol*, *gag* and *env*) genes.

3.3 Study site and population

The study population comprised of 25 HIV positive plasma samples which were, as already mentioned above, sequenced in the *gag* and *env* genes from a previous project. The samples were drawn from the neighbouring townships of Pretoria: Ga-Rankuwa, Soshanguve, Mamelodi, Laudium, Kalafong, Jubilee and Mabopane (Table 3.1 and Figure 3.1). The samples were all obtained from adult HIV patients enrolled for HIV treatment at various treatment centres around Pretoria, and were selected from the NHLS

Department of Virology diagnostic laboratory after routine testing for HIV viral load. Only samples with viral load of 100,000 copies/ml or greater were selected. It is well established that HIV-1 plasma viral load is highly linked with patients' infectivity. The HIV plasma viral load is highest in the early and late stages of the disease and the genetic diversity of this high infectivity stage was the target of the study.

Table 3.1: Summary of the study population

Sample Source region	**Number of study samples**
Ga-Rankuwa (GAR)	5
Soshanguve (SOS)	9
Mamelodi (MAM)	3
Laudium (LAU)	1
Kalafong (KAL)	2
Jubilee (JUB)	3
Mabopane (MAB)	2
Total	**25**

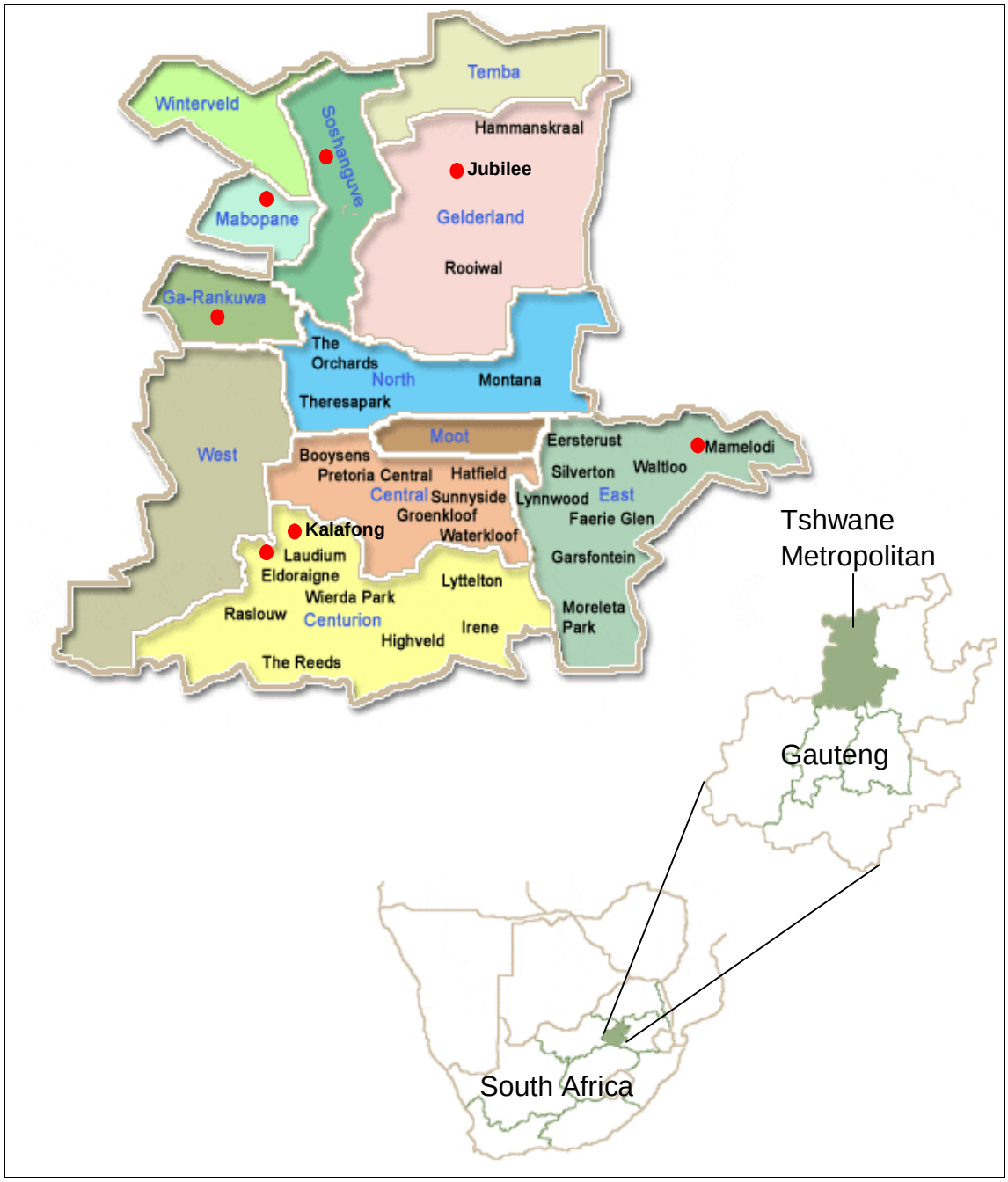


Figure 3.1: South African map showing the location of townships in Pretoria. The red dots represent regions where samples were collected.

3.4 Sample nomenclature

All samples were labelled with letters and numbers representing the country, the year the gene was characterised, the sample origin (township), sample study number (unique laboratory number) and the gene fragment targeted. A letter “c” was used to identify clones from direct samples. An example of a full

name of a labelled sequence is ZA.07.SOS.1.VIF, which represents the following:

ZA-Country (South Africa)

07-Year the sample was collected

SOS-Township of origin (Soshanguve)

1-Sample study number

VIF-Gene fragment

c-Clone

3.5 Laboratory Methods

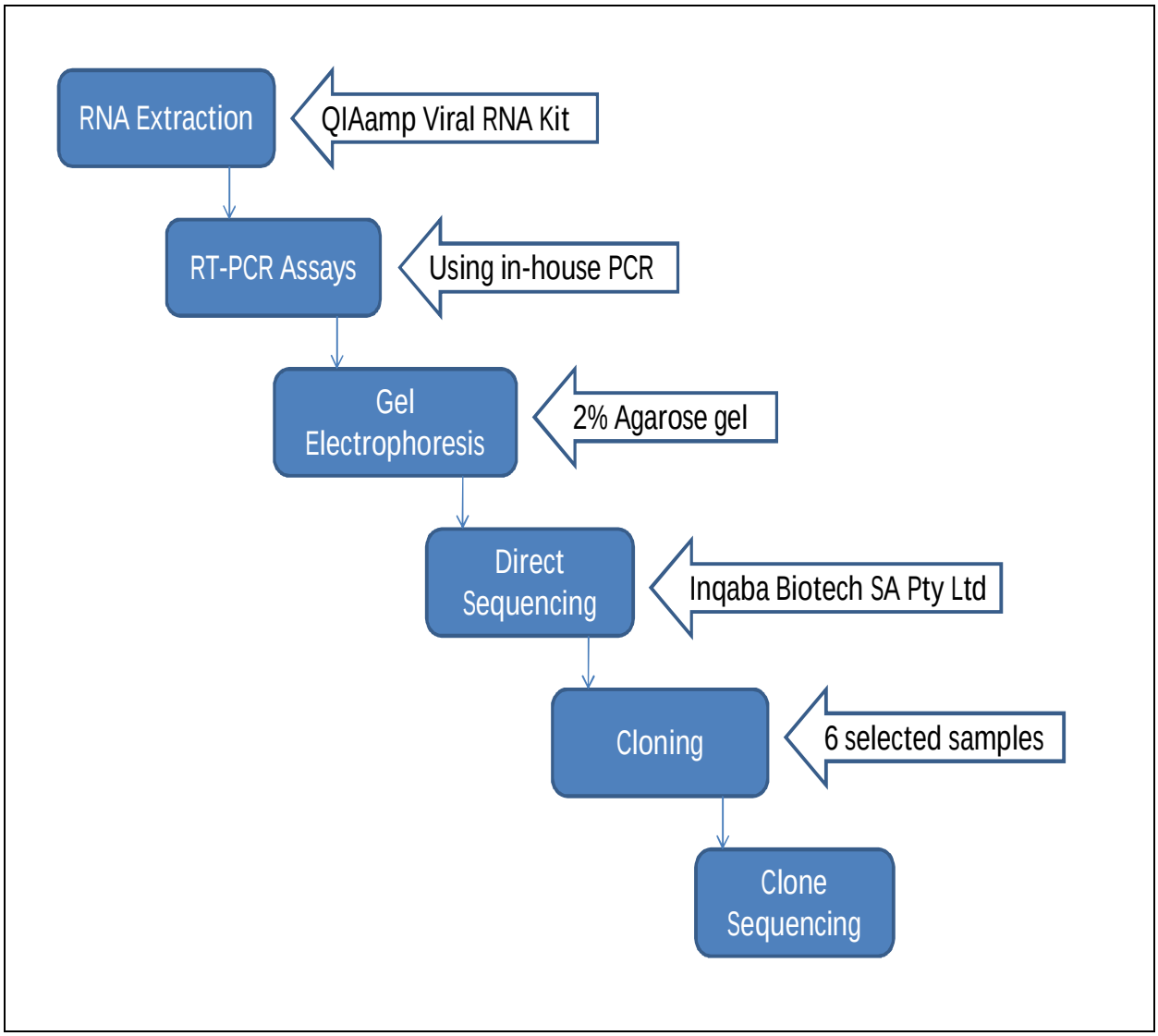


Figure 3.2: Schematic summary of laboratory methods.

3.5.1 Storage of samples

All plasma samples and their extracts were stored at -70ºC in ultra low freezer. The cDNA, first round PCR and second round PCR products were stored at -20ºC in the freezer.

3.5.2 RNA extraction

RNA extraction was performed with QIAamp assay (QIAGEN, Valencia, CA, USA). Briefly, carrier RNA was prepared by adding 310µl of AVE (elution) buffer to the tube containing 310µg RNA. The carrier RNA was added to AVL buffer to prepare working lysis solution (for one reaction 5.6µl of carrier RNA was added to 0.56ml of AVL (lysis) buffer. From the working lysis solution, 560µl was added to an Eppendorf tube. Then 140µl of sample was added and mixed by vortexing. The Eppendorf tube was incubated for 10 minutes at room temperature. The tube was centrifuged to settle the drops from the lid and 560µl of ethanol (96-100%) was added and mixed by vortexing.

A spin column with a collection tube was assembled and labelled. A 630µl of the mixture was transferred from the Eppendorf tube into the assembled spin column and centrifuged at 8000 × g for 1 minute. The spin column was moved into a new collection tube and the remaining 630µl of the mixture in the Eppendorf tube was added into the spin column. The spin column was centrifuged at 8000 × g for 1 minute. The spin column was moved into a new collection tube and 500µl of AW1 (wash) buffer was added and centrifuged at 8000 × g for 1 minute. The spin column was moved into a new collection tube and 500µl AW2 (wash) buffer was added. The spin column was centrifuged at 8000 × g for 4 minutes. The spin column was moved into a 1.5ml Eppendorf tube and 60µl of AVE buffer was added and centrifuged at 8000 × g for 1 minute. The spin column and all the collection tubes were discarded. The 1.5ml Eppendorf tubes with RNA were stored at -70ºC.

3.5.3 HIV cDNA synthesis

For the cDNA synthesis of HIV accessory genes (*vif*, *vpu* and *vpr*), 1µl of outer reverse primer (CATH3F; Figure 3.3 and Table 3.2) was added to 5µl of template and incubated at 70ºC for 5 minutes and chilled on ice for 1 minute. The master mix for one sample was prepared by adding 2µl of 5×buffer, 1µl of 25mM dNTPs, 0.5µl of revert Aid and 0.5µl ribolock. The master mix (4µl) was

transferred to the ice chilled mix and incubated at 42ºC for 1 hour and 70ºC for 10 minutes.

For the *pol* gene cDNA synthesis, the same procedure described above was followed using primer IN3 (Figure 3.4 and Table 3.3). The obtained HIV cDNA products were used for subsequent PCR reactions.

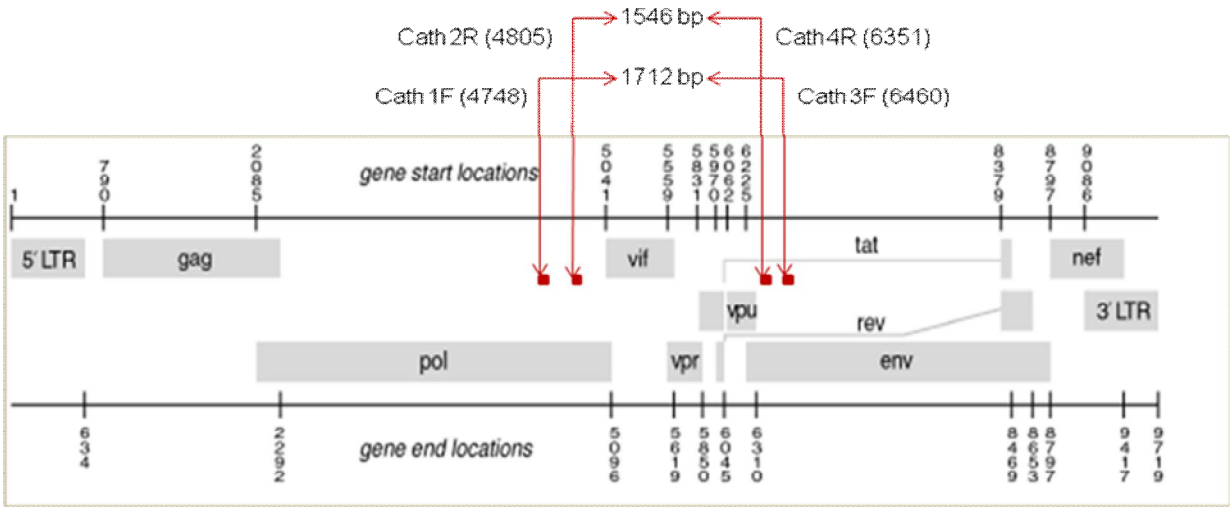


Figure 3.3: Schematic representation of HIV genome with primer binding sites for accessory genes, covering the whole region of *vif*, *vpr* and *vpu*.

Table 3.2: Primers for accessory genes (primer sequences were adopted from Bell et al, 2008).

Primer	Sequence	Polarity	Reaction	Product size
Cath1F	5'-GACAGCAGTACAAATGGCAG-3'	Sense	1st Round	1712bp
Cath3F	5'-GGGTCTGTGGGTACACAGGC-3'	Anti-Sense	1st Round and cDNA	
Cath2R	5'-GGGGTACAGTGCAGGGGAAAG-3'	Sense	2nd Round	1546bp
Cath4R	5'-GTACCCCATAATAGACTGTGACC-3'	Anti-Sense	2nd Round	

3.5.4 HIV PCR assay

3.5.4.1 First round PCR assay for HIV accessory genes

A master mix for one reaction contained 0.4µl of 25mM dNTPs, 2.5µl of 10×buffer, 0.5µl of 10mM CATH3F, 0.5µl of 10mM CATH1F (Figure 3.3 and Table 3.2), 0.2µl of expanded long template DNA polymerase and 16.9µl of distilled water. The DNA template (4µl) was added to the master mix. This brings the final volume for one PCR reaction up to 25µl. The PCR reaction was placed in a thermocycler. The cycling conditions were: 94ºC for 1 minute (initial denaturation), followed by 10 cycles of denaturation at 94ºC for 20 seconds, annealing at 57ºC for 30 seconds and elongation at 72ºC for 2.30 minutes, with an additional 25 cycles of denaturation at 94ºC for 20 seconds, annealing at 57ºC for 30 seconds and elongation at 72ºC for 2.30 minutes at 10 seconds increment. The final elongation was set at 72ºC for 10 minutes and the thermocycler was set to store finished PCR products at 4ºC until removing for forther use.

3.5.4.2 Second round PCR assay for HIV accessory genes

A master mix in the second round was prepared as described in first round PCR except that primer set CATH2R and CATH4R (Figure 3.3 and Table 3.2) was used. The amount of template was reduced to 2µl and distilled water was increased to 18.9µl. The cycling conditions and the number of cycles were the same as those in the first round except for the annealing temperature that was 58ºC. This produced a 1546 base pair long PCR product (Figure 3.3).

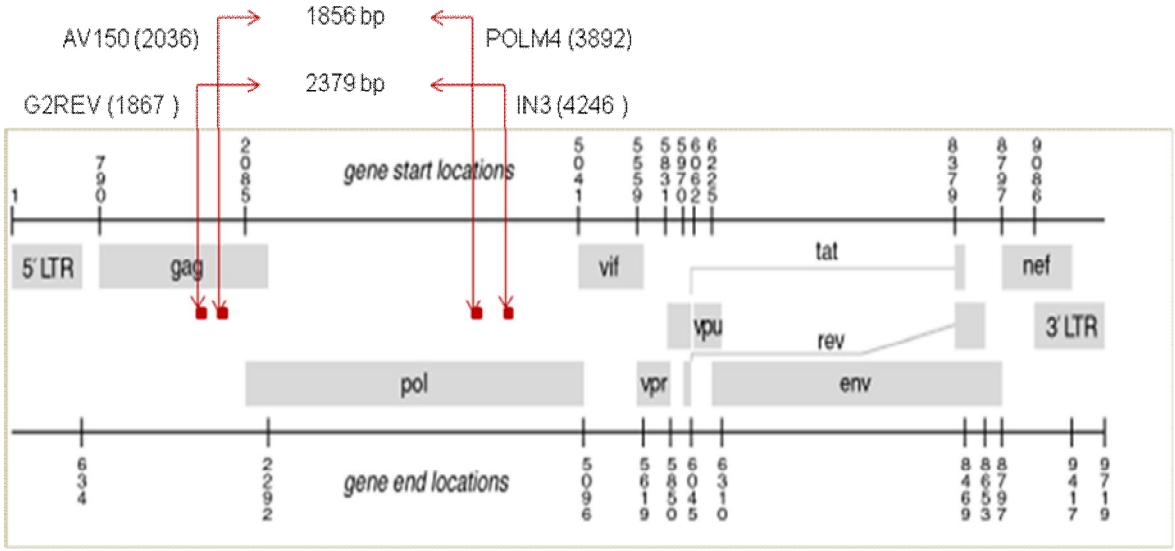


Figure 3.4: Schematic representation of HIV genome with primer binding sites for *pol* gene, encompassing the region of protease and reverse transcriptase.

Table 3.3: Primers for *pol* gene (primer sequences were adopted from Vergne et al, 2000).

Primer	Primer sequences	Polarity	Reaction	Product size
G2REV	5'-GCAAGAGTTTTGGCT GAAGCAATGAG-3'	Anti-sense	1st round	2379bp
IN3	5'-TCTATVCCATCTAAAA ATAGTACTTTCCTGATTC C-3'	Sense	1st round and cDNA	
AV150	5'-GTGGAAAGGAAGGAC ACCAAATGAAAG-3'	Anti-sense	2nd round	1856bp
POLM4	5'-CTATTAGCTGCCCCAT CTACATA-3'	Sense	2nd round	

3.5.4.3 First round PCR assay for HIV *pol* gene

A master mix for one reaction was prepared by adding 0.4µl of 25mM dNTPs, 2.5µl of 10×buffer, 0.5µl of 10mM G2REV, 0.5µl of 10mM IN3, 0.2µl of expanded long template DNA polymerase and 17.9µl of distilled water. 3µl of the template was transferred to the master mix. The PCR reaction was ran in a thermocycler. The cycling conditions were set at 92ºC for 2 minute (initial denaturation), followed by 10 cycles of denaturation at 92ºC for 20 seconds, annealing at 52ºC for 30 seconds and elongation at 68ºC for 2 minutes. This was followed by 30 cycles of denaturation at 92ºC for 20 seconds, annealing at 52ºC for 30 seconds and elongation at 68ºC for 2 minutes+5 seconds increment. The final elongation was at 68ºC for 10 minutes and stored at 4ºC for infinite.

3.5.4.4 Second round PCR assay for HIV *pol* gene

A master mix for the second round PCR reaction was prepared in the same manner as for the first round PCR assay, except that 0.5µl of the template (first round PCR product) was used and 20.4µl of distilled water and the primer set AV150 and PolM4 was used (Table 3.3). This produced a 1856 base pair long PCR product (Figure 3.4). The cycling conditions were the same as for the first round PCR reaction except the annealing temperature which was set at 49ºC of 30 seconds.

3.6 Detection of PCR products

The PCR products were viewed in 2% agarose gel, which was prepared by adding 2g of agar into 100ml of 1×TBE (Tris/Borate/EDTA) buffer and 1µl of ethidium bromide. The agarose wells were loaded with 2µl of base pair marker, 2µl of controls and 2µl of the samples. A loading dye was used to facilitate loading of samples and controls into the wells.

The loaded agarose gel was run submerged in 1×TBE buffer at 100 volt for 30 minutes. After 30 minutes, the results were viewed under UV light with protection glass.

3.7 Sequence analysis

3.7.1 Direct sequencing of PCR products

All 25 *pol* gene PCR products and 24 accessory genes PCR products were directly sequenced.

Principle

The sequencing assay was performed using Spectrumedix SCE 2410 genetic analyser from Inqaba Biotechnological Industry (Pty) (Ltd). The sequencing employed Sanger et al (1977) method, which uses the 2', 3'-dideoxynucleoside-5'-triphosphates (ddNTP's) analogues that act as chain-terminating inhibitors. The incorporation of ddNTP's (ddATP, ddCTP, ddGTP and ddTTP) causes chain-termination. The ddNTP's lacks 3'-OH group which is necessary for attachment of the subsequent nucleotide. One of the four ddNTP's was added with the enzyme (T7 DNA polymerase), 4 reactions each with a different ddNTP were performed. From all 4 reactions complete sequence information was acquired. The reaction was stopped by addition of 95% formalide, 20mM EDTA, 0.05% xylene cyanol and denatured by heating.

Table 3.4: Sequencing primers for accessory genes

Primer	Primer sequence	Polarity
Cath2R	5'-GGGGTACAGTGCAGGGGAAAG-3'	Anti-sense
Cath4R	5'-GTACCCCATAATAGACTGTGACC-3'	Sense

Table 3.5: Sequencing primers for *pol* gene

Primer	Primer sequence	Polarity
PolM0	5'-TCCCTCAGATCACTCTTTGGCA-3'	Sense
PolM1	5'-GTTAAACAATGGCCATTGACAGA-3'	Sense
PolM4	5'-CTATTAGCTGCCCCATCTACATA-3'	Anti-sense
PolM9	5'-ATTGAACTTCCCAGAAGTCTTGAGTT-3'	Anti-sense
PolM8	5'-CTGTATATCATTGACAGTCCAG-3'	Anti-sense

3.7.2 Cloning and sequencing

3.7.2.1 Rationale for cloning

Six samples were cloned to perform in-depth investigation of the genetic diversity of the accessory genes. Three of these samples (ZA.07.SOS.5, ZA.07.SOS.2 and ZA.07.GAR.13) were selected for cloning because their direct PCR sequences demonstrated insertions in one of the three accessory genes (*vif*, *vpr* and *vpu*). The other two samples (ZA.07.GAR.14 and ZA.07.KAL.26) were cloned to compare the genetic diversity of accessory genes with that of previously cloned *gag* and *env* genes. From each cloned sample, five clones were recovered by PCR assay and sequenced as described above.

3.7.2.2 Principle of cloning

The pGEM-T Easy Vector System is a convenient system for cloning PCR products. The vector is prepared by cutting the pGEM-5Zf (+) and pGEM-T Easy Vector with EcoRV and adding a 3' terminal thymidine to both ends (Figure 3.5). These single 3'-T overhangs at the insertion site improve the efficiency of ligation of a PCR product into the plasmids by preventing recircularization of the vector and providing a compatible overhang for PCR products thus by providing 5′ A overhangs.

The pGEM-T Easy Vectors contain T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the coding region for the α-peptide of β-galactosidase. Insertional inactivation of the α-peptide allows recombinant clones to be directly identified by colour screening on indicator plates containing X-Gal and IPTG.

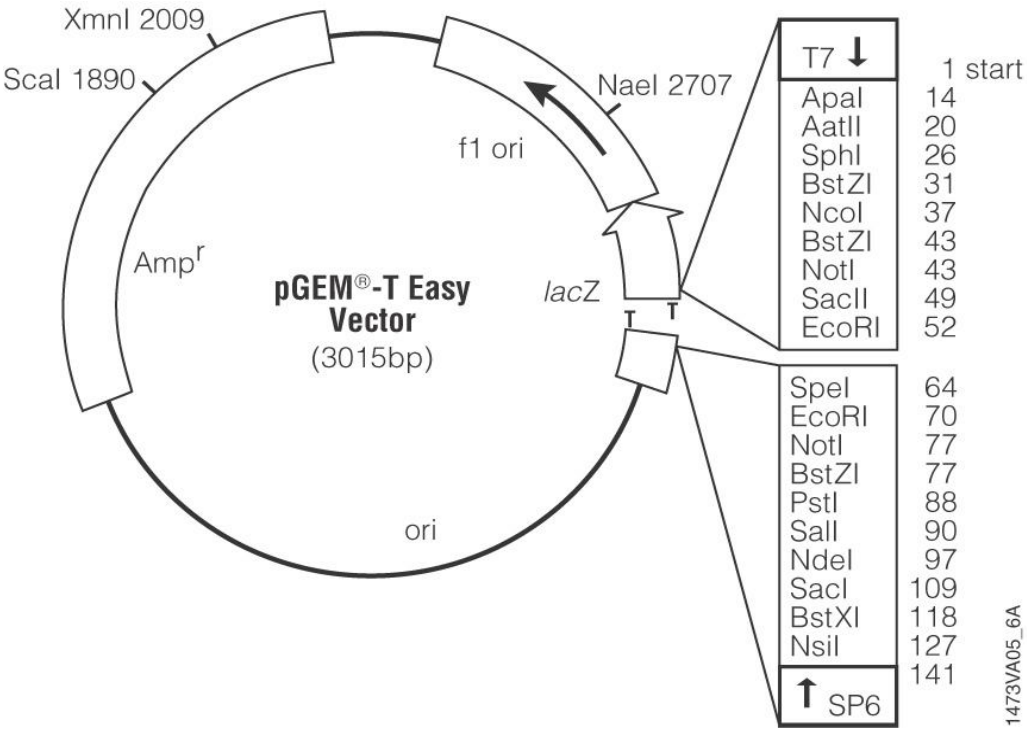


Figure 3.5: The pGEM®-T Easy vector circle map (Adapted from www.promega.com).

3.7.2.3 Ligation reaction

The pGEM-T Easy Vector and control insert DNA tube were briefly centrifuge to collect contents at the bottom of the tube. For each PCR product cloned, a separate ligation reaction was set-up. The 2×Rapid ligation buffer was vigorously vortexed each time it was used. For each PCR product ligation: 5µl of 2×Rapid ligation buffer, 1µl of pGEM-T easy Vector (50ng), 2µl of PCR product, 1µl of T4 ligase (3 Weiss units/µl) and 1µl of deionised water were added to 0.6ml tube. Positive and background controls ligation reactions were also prepared. The positive control was containing 2µl of control insert DNA

and a background lacked any DNA. The reaction mixtures were mixed by pipetting and incubated overnight at 4ºC.

3.7.2.4 Transformation reaction

Two LB/ampicillin/IPTG/X-Gal plates were prepared for each ligation reaction. The plates were cooled at room temperature before plating. The ligation reaction tubes were centrifuged and 2µl of each ligation reaction was collected at the bottom of the tube. The 2µl was transferred to a sterile falcon tube on ice. JM 109 high efficiency competent cells were placed in ice until just thawed and the cells were mix by gently flicking the tube. 50µl of cells were carefully transferred to each ligation reaction tube on ice. The tubes were mixed by flicking and incubated on ice for 20 minutes. The cells were heat-shocked for 45-50 seconds in a water bath at 42ºC and immediately returned to ice for 2 minutes. Then 950µl of SOC media (at room temperature) was added to each tube containing cells transformed with ligation reactions and incubated for 1.5 hours at 37ºC shacking at 150 rpm.

Two LB/ampicillin/IPTG/X-Gal plates were prepared for each transformation culture. 100µl of transformation culture was plated on each plate and incubated overnight at 37ºC. Ultimately, the recombinant vector were identified by their characteristic white colour, and vectors that did not pick up any DNA grew with a characteristic blue colour (Figure 3.6).

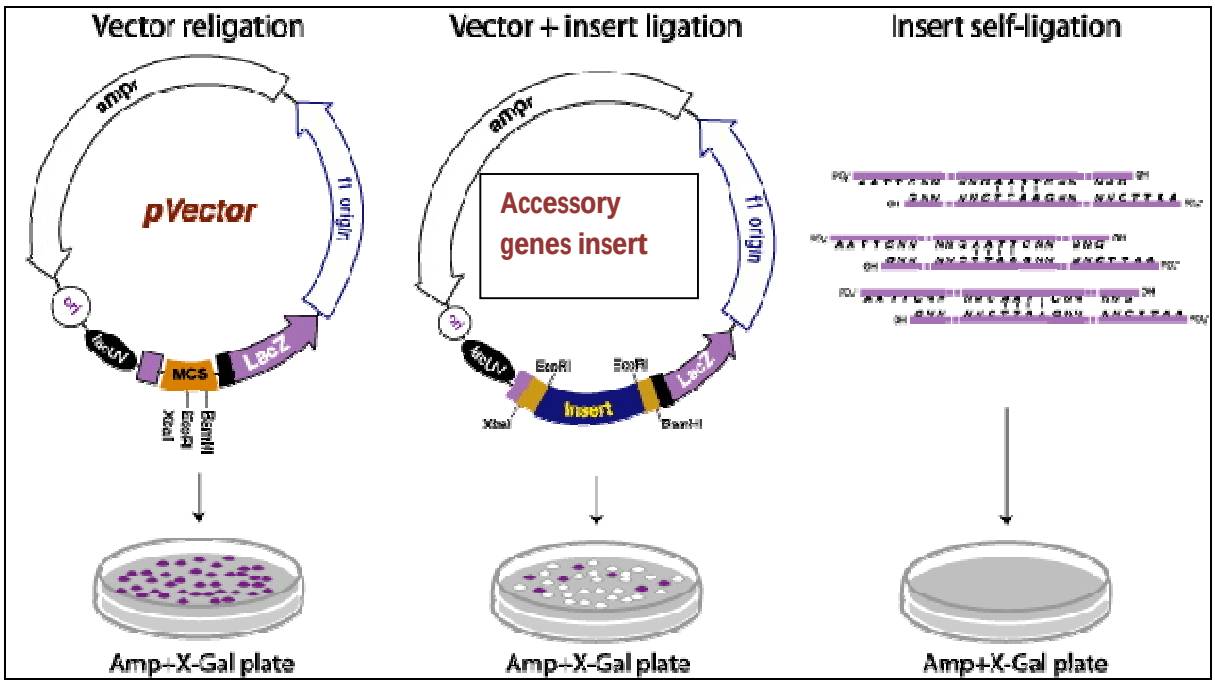


Figure 3.6: Schematic representation showing growth of recombinant plasmids on agar (Adapted from www.promega.com).

3.7.2.5 Confirmation of cloned fragments

Following transformation of the bacterial cells, the resulting bacterial colonies were assessed by PCR assay for the correct recombinant vector using specific primers (Table 3.2 and Figure 3.3) to amplify the insert. The amplicons were detected by agarose gel electrophoresis.

Briefly, a master mix for one reaction contained 0.4µl of 25mM dNTPs, 2.5µl of 10×buffer, 0.5µl of 10mM CATH2R, 0.5µl of 10mM CATH4R, 0.2µl of expanded long template DNA polymerase and 18.9µl of distilled water. A scoop of colony was used as DNA template. This brings the final volume for one PCR reaction up to 23µl. The PCR reaction was placed in a thermocycler. The cycling conditions were: 94ºC for 1 minute (initial denaturation), followed by 10 cycles each of denaturation at 94ºC for 20 seconds, annealing at 58ºC for 30 seconds and elongation at 72ºC for 2.30 minutes. There were additional 25 cycles each of denaturation at 94ºC for 20 seconds, annealing at 58ºC for 30 seconds and elongation at 72ºC for 2.30 minutes at 10 seconds

increment. The final elongation was set at 72ºC for 10 minutes and the thermocycler was set to store finished PCR products at 4ºC until manually switched off. The expected PCR product was 1549 base pairs long.

3.7.2.6 Detection of cloned PCR amplicons

The cloned PCR amplicons were detected using 2% agarose gel electrophoresis as described in section 3.6 earlier and the bands were viewed under UV light.

3.7.3 Analysis of sequences

Sequences generated from each primer for each sample were assembled on Chromas Pro version 1.4 to form contiguous sequence. Accessory genes (*vif, vpr* and *vpu*) were sequenced using two primers and five primers for the *pol* gene. The sequences generated from each primer were downloaded from Inqaba Biotechnological Industry (Pty) (Ltd) finch TV. Each gene was analysed separately for each sample. Sequence alignment was done using ClustalW and phylogenetic trees were constructed using Mega 4 program (Thompson et al, 1994; Tamura et al, 2007).

3.7.3.1 Sequence editing

The contiguous sequence was edited on Chromas Pro using reference sequence as a guide. The contiguous sequence was selected and saved in faster format. This was done for all samples and was saved on a notepad document.

3.7.3.2 Sequence alignment

The contiguous sequences on a notepad were opened with BioEdit together with reference sequences from GenBank. All sequences were selected and the opened window was set on edit mode. On the accessory option, ClustalW multiple alignment option was chosen to align all the sequences. The aligned sequences were then trimmed on both ends to form uniform length. The sequences were then saved as BioEdit faster format.

3.7.3.3 Phylogenetic construction

On Mega 4 program, the BioEdit faster file was opened and converted to mega format. This was done by clicking the convert to mega icon. The mega format file was saved. The mega format file was opened on Mega 4 program. The phylogenetic option was clicked and neighbour joining phylogenetic tree option was run. The resulting phylogenetic tree was copied and saved on Microsoft word.

3.8 Avoiding PCR contamination

To avoid false positive results that might be caused by contamination or carry over of amplified PCR products, the following measures were followed:

1. Working areas were cleaned with 70% ethanol and sodium hypochlorite before and after each PCR experiment.
2. Gloves and laboratory coats were worn during PCR procedures.
3. Nucleic acid extractions and PCR master mixes were prepared in different separate bio-safety cabinets.
4. Separate filter tips and pipettes for extraction and PCR master mixes were used.
5. Only sterile tips and tubes were used.
6. Positive and negative controls were included in every run.

CHAPTER FOUR

4. RESULTS

4.1 Overview of the results

A set of 25 HIV positive plasma samples were selected for phylogeny and analysis of motifs associated with key biological activities of HIV-1 structural (*env*, *gag* and *pol*) and accessory (*vif*, *vpr* and *vpu*) genes, in order to generate near full-length sequences of these 25 samples and to compare the genetic diversity of HIV-1 subtype C between accessory (*vif*, *vpr* and *vpu*) and structural (*pol*, *gag* and *env*) genes. The samples were drawn from the neighbouring townships of Pretoria: Ga-Rankuwa, Soshanguve, Mamelodi, Laudium, Kalafong, Jubilee and Mabopane. The study employed 25 samples which were sequenced for *gag* and *env* genes by a previous study. For the purpose of the current study, all 25 samples were successfully amplified for accessory genes (*vif*, *vpr* and *vpu*) and *pol* gene. All the 25 *pol* PCR amplicons were successfully sequenced, while all but one accessory PCR amplicons were successfully sequenced.

Inter-sample sequence homology was evaluated with ClustalW2 EBI software. The genetic diversity of HIV-1 subtype C was compared between accessory (*vif*, *vpr* and *vpu*) and structural (*pol*, *gag* and *env*) genes. The *gag* and *vif* gene sequences were highly conserved (89% to 96% and 88% to 96%, respectively), as compared to *vpr* gene (84% to 94%), the *pol* gene (79% to 95%), the *env* gene (83% to 93%) and finally the *vpu* gene (73% to 92%). The high variability of the *vpu* gene was seen towards the end of the gene, most probably due to primer binding region. We believe that the PCR amplicons for accessory genes could have contributed to the high variability of the *vpu* gene because the binding region of the reverse primer was very close to the end of the *vpu* gene, creating difficulties with interpretation of the sequence chromatograms of the *vpu*. In addition, a number of conserved regions were observed in all the six genes (*vif*, *vpr*, *vpu*, *gag*, *env* and *pol*).

The accessory genes were highly conserved at both nucleotide and amino acid levels. Direct observation of the *vif* sequences revealed a great degree of conservation, with amino acids W5, W11, W21 and W38 being highly conserved since they are involved in recognition and suppression of APOBEC3G. The HCCH zinc motif and *vif* dimerization site also demonstrated high conservation. The important motifs in the *vpr* gene demonstrated similar degree of conservation and included the P35 and H71 which are involved in cyclophilin A binding and *vpr* dimer stacking, respectively. In addition, the R77Q mutation associated with long term non-progressors was observed in the *vpr* gene of 15 sequences. The important sequence motifs in the *vpu* gene showed some conservation and included W28, 34-EYRKLL-39, 57-DSGNES-62 and 78-LRLL-81, which are involved in channel gating, hinge region containing salt-bridge, casein kinase II phosphorylation containing critical serines and putative Golgi export signal, respectively.

The studied samples were screened on BioAfrica for subtyping and phylogenetic analysis was used to confirmed that all 25 HIV positive samples were subtype C in all six (structural and accessory) genes, in comparison with HIV-1 group M reference sequences from the Los Alamos Sequence Database. All the sequences from all the six genes showed to be HIV-1 subtype C. There was no South African cluster that was observed in both the direct PCR sequences and cloned sequences. Regional clusters were observed with some of the samples in all the six genes. The genetic diversity was minimal in cloned sequences.

4.2 Overview of PCR and sequencing results

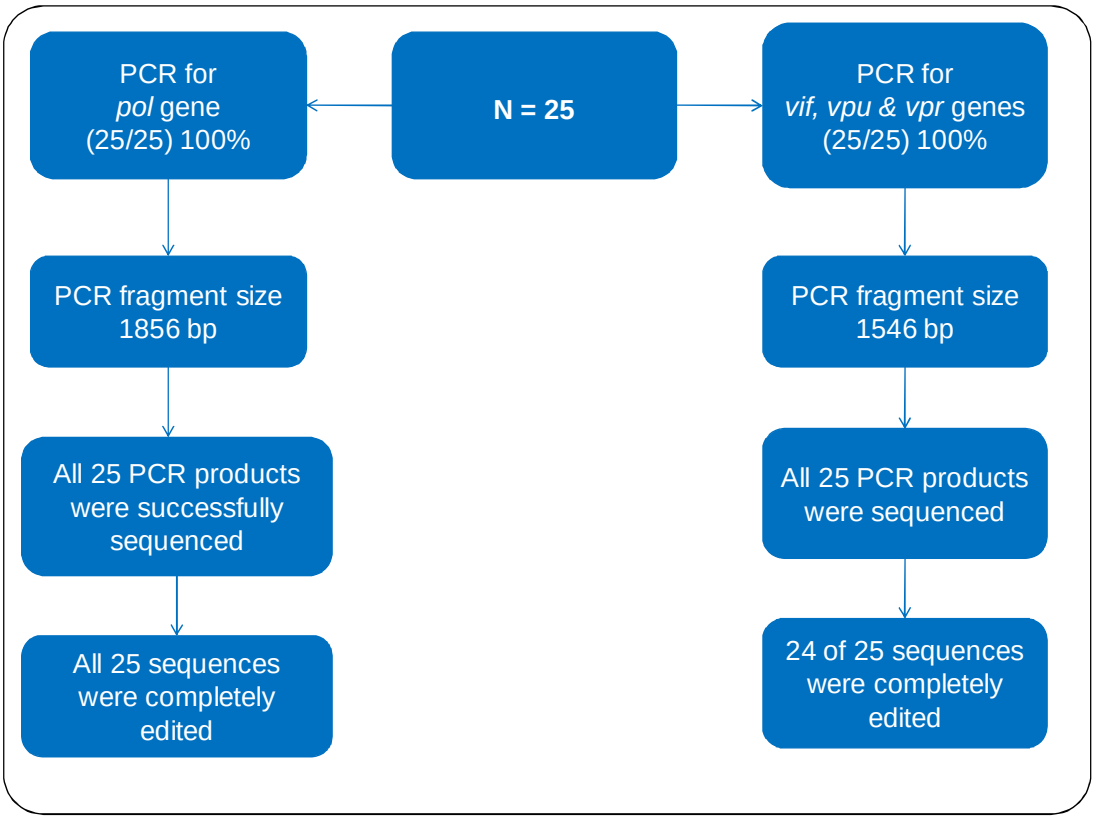


Figure 4.1: Summary of PCR and sequencing results.

All the 25 samples selected had been sequenced for both *gag* and *env* genes. The accessory gene primers covered a 1546bp region containing *vif*, *vpr* and *vpu,* while the *pol* primers targeted a 1856bp long region covering the protease and reverse transcriptase.

4.3. Detection of PCR amplicons

Two microlitres of the amplified PCR products were run on 2% agarose gel at 100 volts for 30 minutes. A 100bp molecular weight marker was used to identify the expected size of the amplicons. For every PCR run, positive and negative controls were included for quality control purpose. The results were considered to be positive when the 1546bp fragment of accessory gene and 1856bp fragment of *pol* gene were observed against the 100bp molecular weight marker (Figures 4.2 and 4.3).

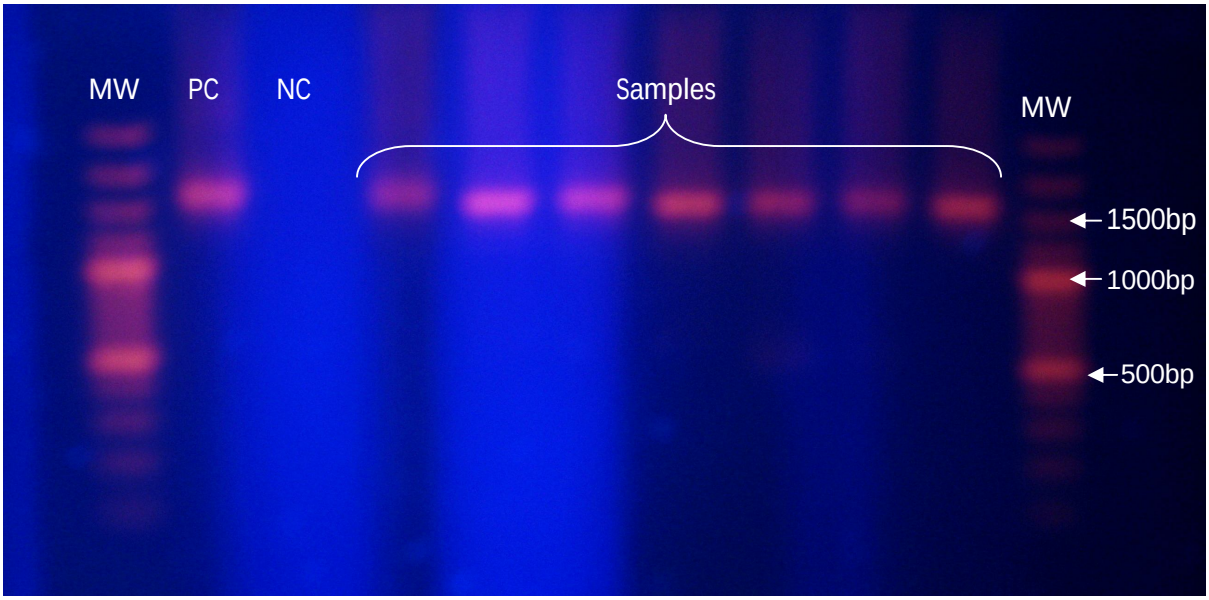


Figure 4.2: Gel electrophoresis of PCR amplicons of accessory genes. MW = 100bp molecular weight marker; PC = positive control; NC = negative control and study samples. Expected PCR product size = 1546bp

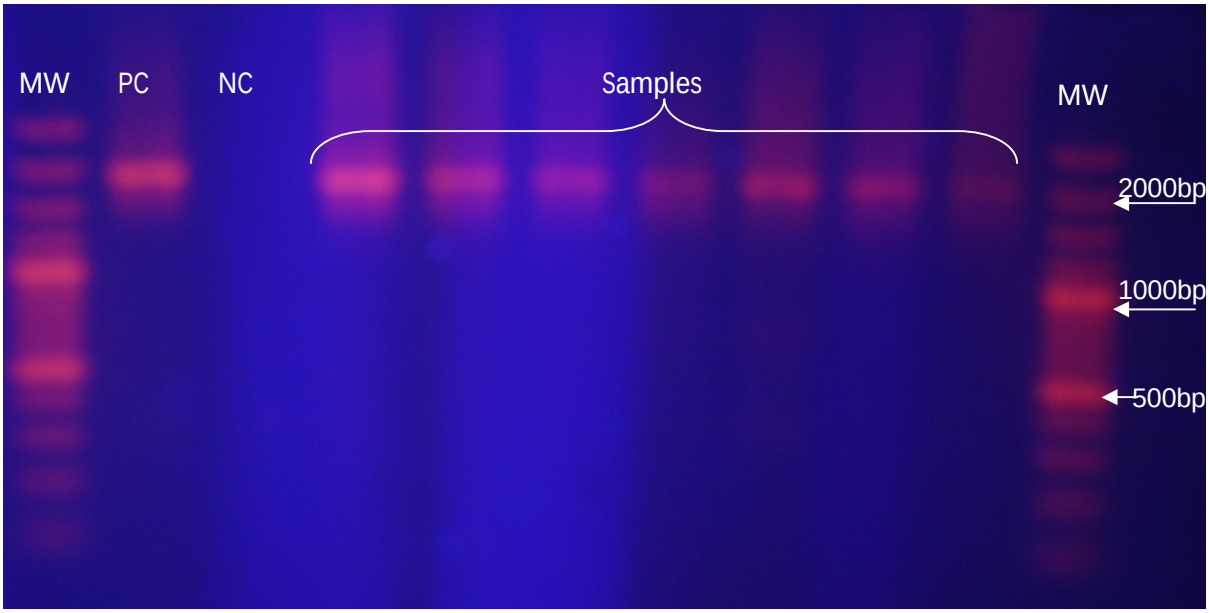


Figure 4.3: Gel electrophoresis of *pol* gene PCR products. MW = 100bp Molecular weight marker; PC = positive control; NC = negative control and study samples. Expected PCR product size = 1856bp

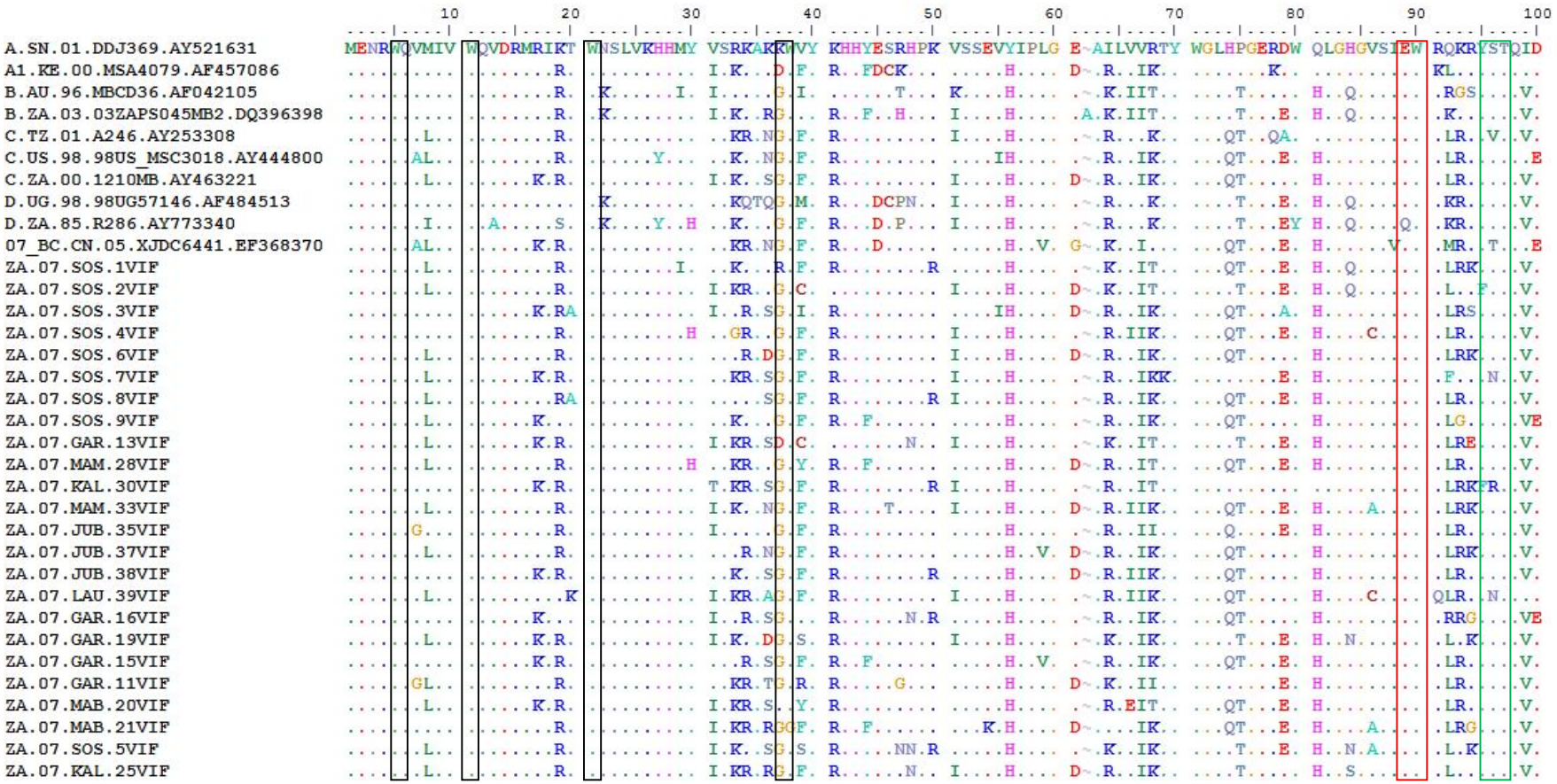
4.4 Sequence analysis of accessory genes

A total of 24 of 25 samples were successfully sequenced for accessory genes. Sequences from one sample, ZA.07.MAM.29 did not form a contiguous sequence. The 24 sequences were aligned together with HIV-1 with reference strains from the Los Alamos sequence Database by ClustalW program packaged in BioEdit software. The nucleotide sequences were translated to amino acids on BioEdit and saved in graphic view. The analysis of the accessory genes was performed, assessing the genetic diversity of accessory gene with special attention to functional motifs.

4.4.1 Direct sequencing of the *vif* gene

Analysis of the *vif* sequence alignments revealed a great degree of conservation in both nucleotide and amino acid levels (Figures 4.4 and 8.1). For example, amino acids W5, W11, W21and W38 (numbered according to HIV-1 subtype C strain from Bell et al (2007); GenBank accession number DQ185448) showed to be conserved. These amino acids are involved in recognition and suppression of APOBEC3G. Region 88-EW-89, involved in enhancing steady state expression of the virus. HCCH zinc motif and *vif* dimerization, also proved to be highly conserved. The SLQYLA motif also demonstrated some degree of conservation (Figure 4.4).

Inter-samples sequence homology was evaluated with ClustalW2 EBI software. The *vif* sequence homology ranged from 88% to 96%.


10
20
30
40
50
60
70
80
90
100
A.SN.01.DDJ369.AY521631 MENRWQVMIV WQVDRMRIKT WNSLVKHHMY VSRKAKKWVY KHHYESRHPK VSSEVYIPLG E~AILVVRTY WGLHPGERDW QLGHGVSIEW RQKRYSTQID
A1.KE.00.MSA4079.AF457086
B.AU.96.MBCD36.AF042105
B.ZA.03.03ZAPS045MB2.DQ396398
C.TZ.01.A246.AY253308
C.US.98.98US_MSC3018.AY444800
C.ZA.00.1210MB.AY463221
D.UG.98.98UG57146.AF484513
D.ZA.85.R286.AY773340
07_BC.CN.05.XJDC6441.EF368370
ZA.07.SOS.1VIF
ZA.07.SOS.2VIF
ZA.07.SOS.3VIF
ZA.07.SOS.4VIF
ZA.07.SOS.6VIF
ZA.07.SOS.7VIF
ZA.07.SOS.8VIF
ZA.07.SOS.9VIF
ZA.07.GAR.13VIF
ZA.07.MAM.28VIF
ZA.07.KAL.30VIF
ZA.07.MAM.33VIF
ZA.07.JUB.35VIF
ZA.07.JUB.37VIF
ZA.07.JUB.38VIF
ZA.07.LAU.39VIF
ZA.07.GAR.16VIF
ZA.07.GAR.19VIF
ZA.07.GAR.15VIF
ZA.07.GAR.11VIF
ZA.07.MAB.20VIF
ZA.07.MAB.21VIF
ZA.07.SOS.5VIF
ZA.07.KAL.25VIF

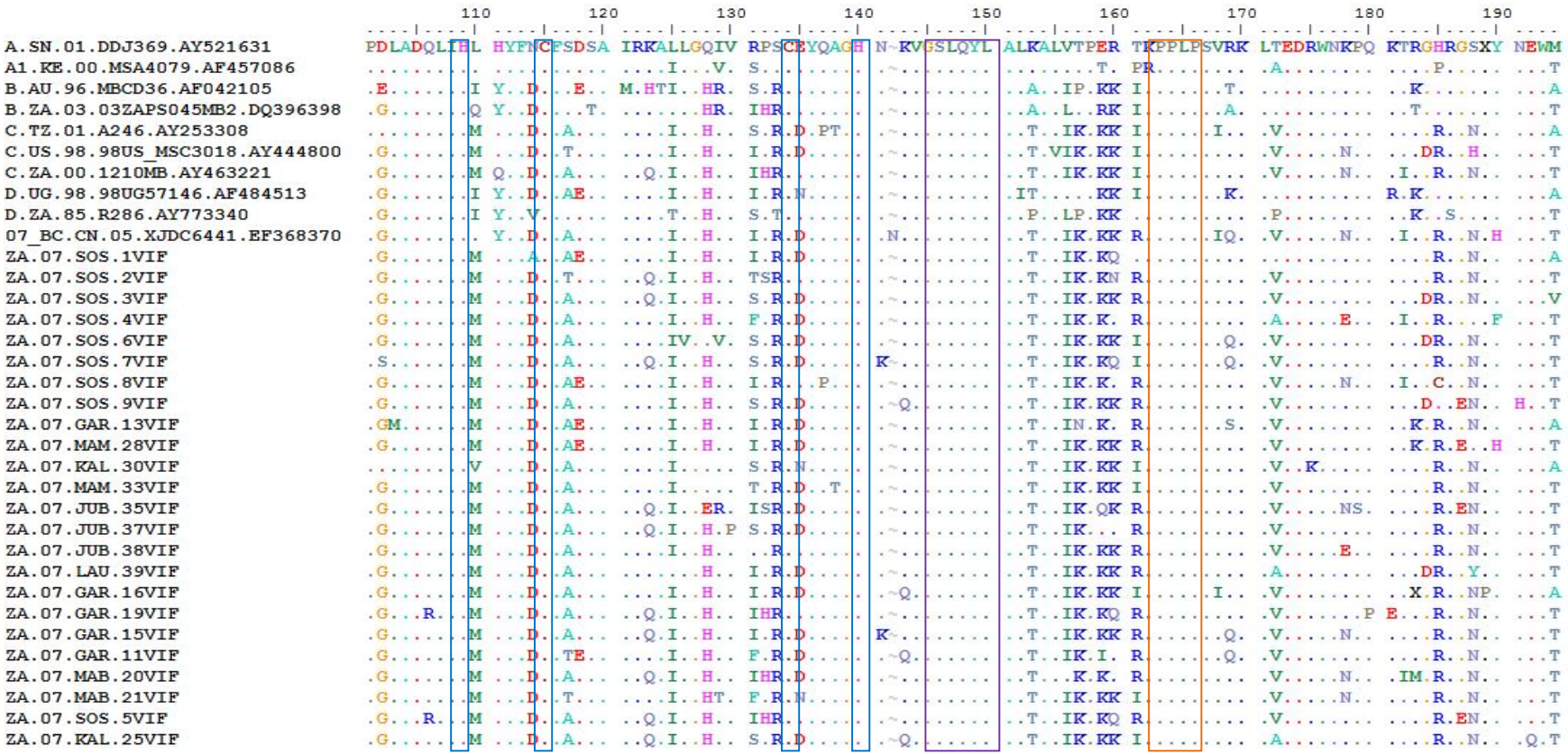


Figure 4.4: Amino acid sequence of *Vif* from 24 HIV-1 subtype C isolates (W5, W11, W21, W38 = tryptophan residues involved in recognition and suppression of APOBEC3G; 88-EW-89 = residues in central hydrophilic region, enhance steady state expression and S95 and T96. CKII and p44/42 MAPK phosphorylation sites. Also note the zinc binding motif [H108, C114, C133, H139. HCCH], the Viral BC box; elongin C bninding motif [144-SLQYLA-149] and finally the Vif dimerization site [161-PPLP-164]).

4.4.2 *Vif* cloned sequences

Five clones per sample were successfully amplified by PCR using specific primers. All clones PCR products were successfully sequenced and all their sequences formed contiguous sequences. The sequences were edited on Chromas Pro program and were aligned with HIV-1 subtype C reference strains from the Los Alamos sequence Database on clustalW. The nucleotide sequences were translated to their amino acids on BioEdit program.

Cloned sequences for *vif* showed degrees of conservation, with most of their important regions conserved. The clones from 1 sample (ZA.07.GAR.13VIF) showed to have the same characteristics, thus suggest related viruses or even from 1 virus (Figure 4.5).

Inter-clones sequence homology ranged from 91% to 99% for ZA.07.SOS.2.VIFc sequences; 91% to 99% for ZA.07.SOS.5.VIFc sequences; 97% to 100% for ZA.07.GAR.13.VIFc sequences; 99% to 100% for ZA.07.GAR.14.VIFc sequences; 95% to 98% for ZA.07.KAL.26.VIFc sequences; and 97% to 98% for ZA.07.MAM.29.VIFc sequences. Some of the clone sequences had a 100% sequence homology showing that the sequences originated from the same virus (Table 8.1 to 8.6).


10 20 30 40 50 60 70 80 90 100
A.SN.01.DDJ369.AY521631 MENRWQVMIV WQVDRMRIKT WNSLVKHHMY VSRKAKKWVY KHHYESRHPK VSSEVYIPLG E-AILVVRTY WGLHPGERDW QLGHGVSIEW RQKRYSTQID
A1.KE.00.MSA4079.AF457086 ....
B.AU.96.MBCD36.AF042105 ....
B.ZA.03.03ZAPS045MB2.DQ396398 ....
C.TZ.01.A246.AY253308 ....L..
C.US.98.98US_MSC3018.AY444800 ......AL..
C.ZA.00.1210MB.AY463221 ....L..
D.UG.98.98UG57146.AF484513 ....
D.ZA.85.R286.AY773340 ....I..
07_BC.CN.05.XJDC6441.EF368370 ......AL..
ZA.07.SOS.2.1VIFc ....L..
ZA.07.SOS.2.2VIFc ....L..
ZA.07.SOS.2.3VIFc ....L..
ZA.07.SOS.2.4VIFc ....L..
ZA.07.SOS.2.5VIFc ....L..
ZA.07.SOS.5.1VIFc ....L..
ZA.07.SOS.5.2VIFc ....L..
ZA.07.SOS.5.3VIFc .....R.L..
ZA.07.SOS.5.4VIFc ....L..
ZA.07.SOS.5.5VIFc ....L..
ZA.07.GAR.14.1VIFc ....L..
ZA.07.GAR.14.2VIFc V.....L..
ZA.07.GAR.14.3VIFc ....L..
ZA.07.GAR.14.4VIFc ....L..
ZA.07.GAR.14.5VIFc ....L..
ZA.07.KAL.26.1VIFc ....L..
ZA.07.KAL.26.2VIFc ....L..
ZA.07.KAL.26.3VIFc ....L..
ZA.07.KAL.26.4VIFc ....L..
ZA.07.KAL.26.5VIFc ....L..
ZA.07.GAR.13.1VIFc ....L..
ZA.07.GAR.13.2VIFc ....L..
ZA.07.GAR.13.3VIFc ....L..
ZA.07.GAR.13.4VIFc ....L..
ZA.07.GAR.13.5VIFc ....L..
ZA.07.MAM.29.1VIFc ....L..
ZA.07.MAM.29.2VIFc ....L..
ZA.07.MAM.29.3VIFc ....L..
ZA.07.MAM.29.4VIFc ....L..
ZA.07.MAM.29.5VIFc ....
ZA.07.SOS.2VIF ....L..
ZA.07.SOS.5VIF ....L..
ZA.07.GAR.13VIF ....L..

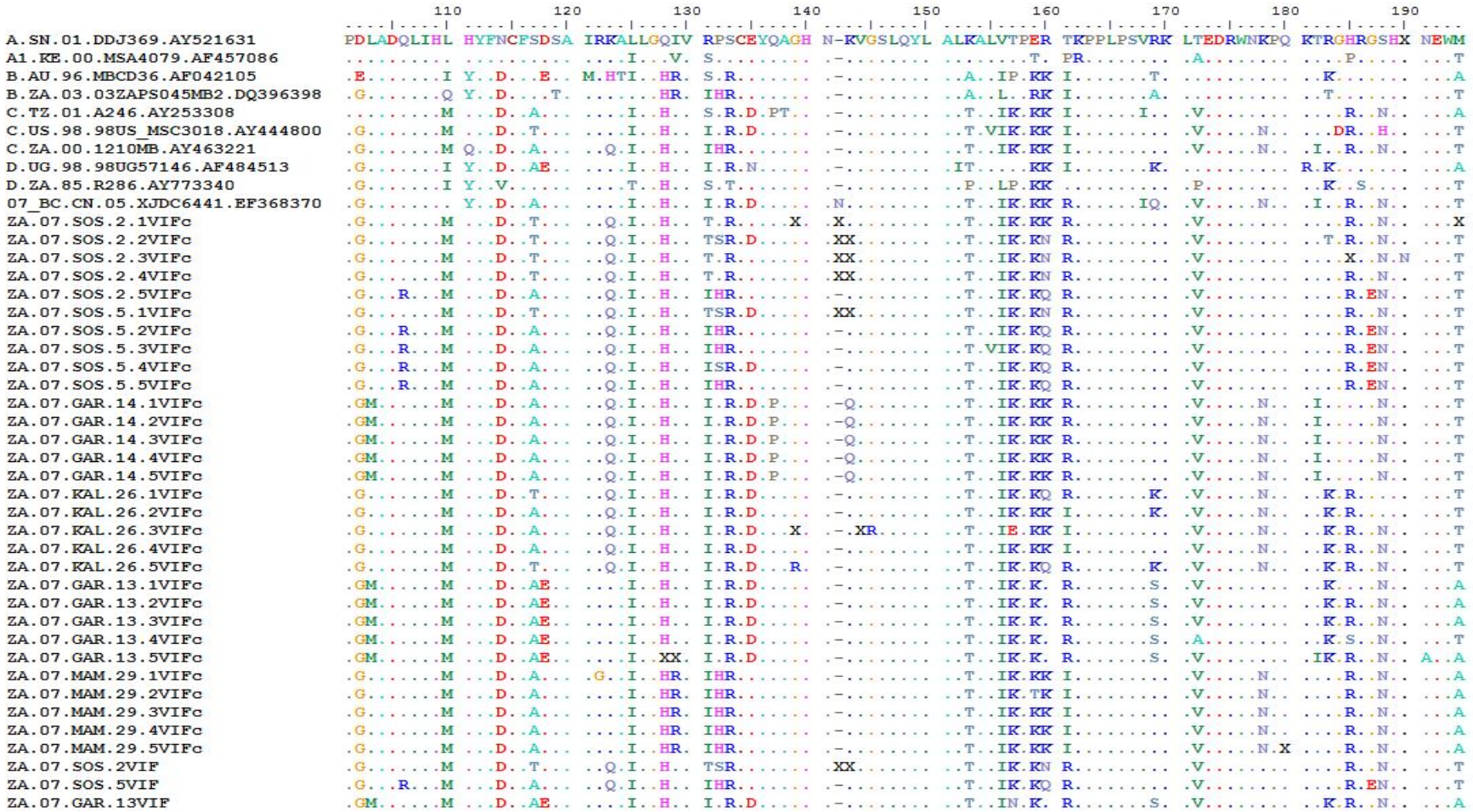


Figure 4.5: Graphic view of the amino acid alignments of *Vif* sequences from 30 HIV-1 subtype C clones. The sequences are compared to reference strains from the Los Alamos Sequence Database.

4.4.3 Direct PCR sequences of *vpr* gene

The important regions in the *vpr* gene showed same degree of conservation. P35 and H71 which are involved in cyclophilin A binding and *vpr* dimer stacking respectively. They also showed to be conserved. R77Q mutation in *vpr* gene was observed in 15 sample sequences; ZA.07.SOS.1.VPR, ZA.07.SOS.2.VPR,ZA.07.SOS.3.VPR, ZA.07.SOS.6.VPR, ZA.07.SOS.8.VPR, ZA.07.SOS.9.VPR,ZA.07.GAR.11.VPR,ZA.07.GAR.13.VPR,ZA.07.GAR.15.VPR,ZA.07.ODI.20.VPR,ZA.07.KAL.25.VPR,ZA.07.MAM.28.VPR,ZA.07.MAM.33.VPR, ZA.07.JUB.37.VPR and ZA.07.JUB.38.VPR. Patients with this mutation showed to be long term non-progressors. Sample ZA.07.SOS.5.VPR had an insertion of 6 amino acids at position 85 (Figure 4.6).

Inter-samples sequence homology was evaluated with clustalW2 EBI software. The inter-sample sequence homology ranged from 84% to 94%.

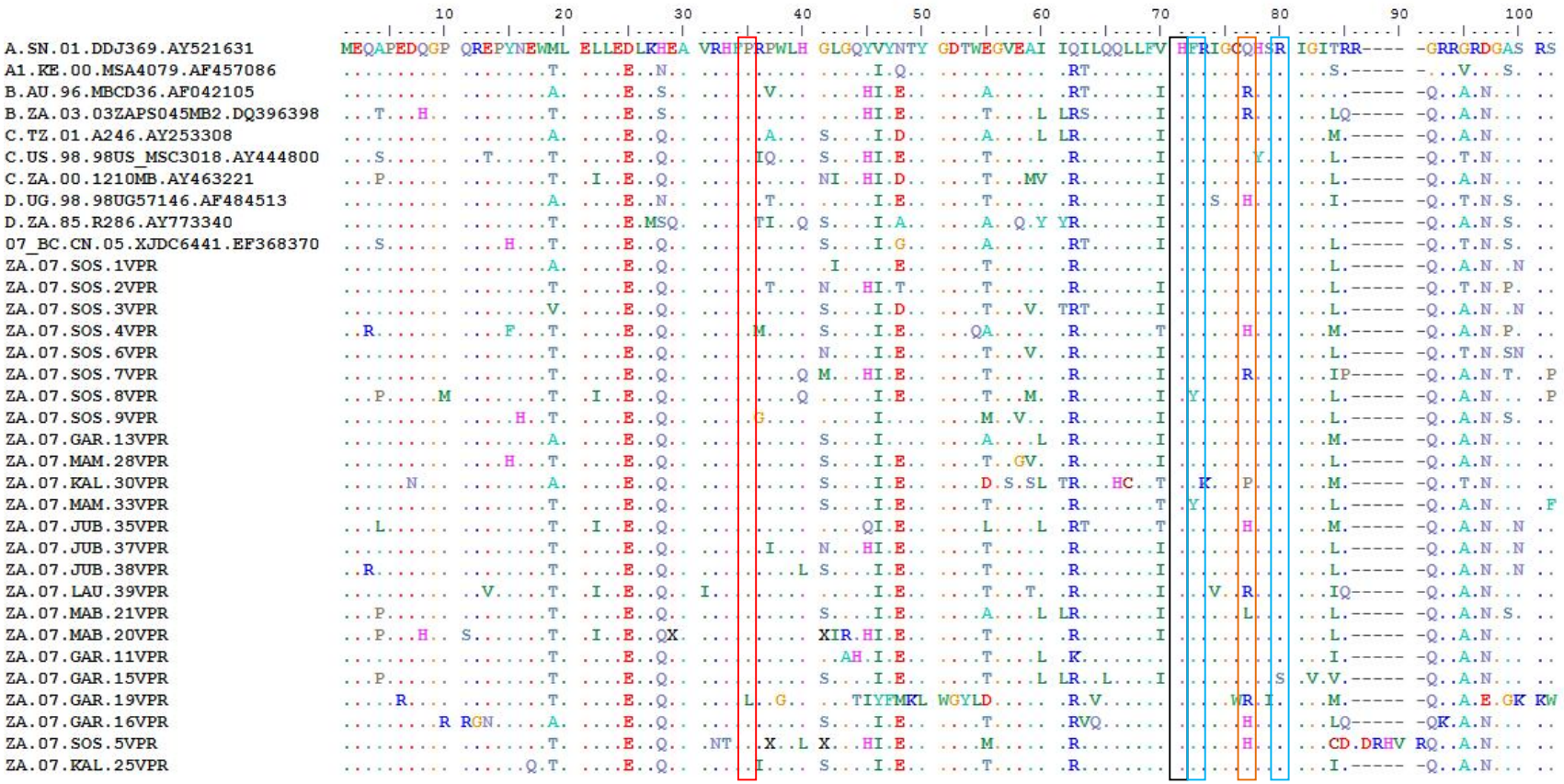


Figure 4.6: Graphic view of the amino acid alignments of *Vpr* sequences from 24 HIV-1 subtype C isolates. The sequences are compared to reference strains from the Los Alamos Sequence Database. (P35 = Cyclophilin A binding, H71 = Involved in vpr dimer stacking with W54 = R73 and R80. ANT protein binding and R77Q = Mutation involved in decreased pro-apoptotic activity).

4.4.4 *Vpr* cloned sequences

The clone sequences showed to be conserved and clone sequences from 1 sample exhibited the same characteristics. This suggesting that the clones are from a closely related viruses or the same virus. Clone sequences for sample; ZA.07.SOS.5 did not exhibit an insertion of 6 amino acids at position 86, which was observed with their direct PCR sequence (Figure 4.7).

Inter-clones sequence homology ranged from 88% to 99% for ZA.07.SOS.2VPRc sequences; 90% to 100% for ZA.07.SOS.5.VPRc sequences; 95% to 100% for ZA.07.GAR.13.VPRc sequences; 99% to 100% for ZA.09.GAR.14VPRc sequences; 97% to 100% for ZA.07.KAL.26.VPRc sequences; and 98% to 100% for ZA.07.MAM.29.VPRc sequences. Some of the clone sequences had a 100% sequence homology showing that the sequences originated from the same virus (Table 8.1 to 8.6).

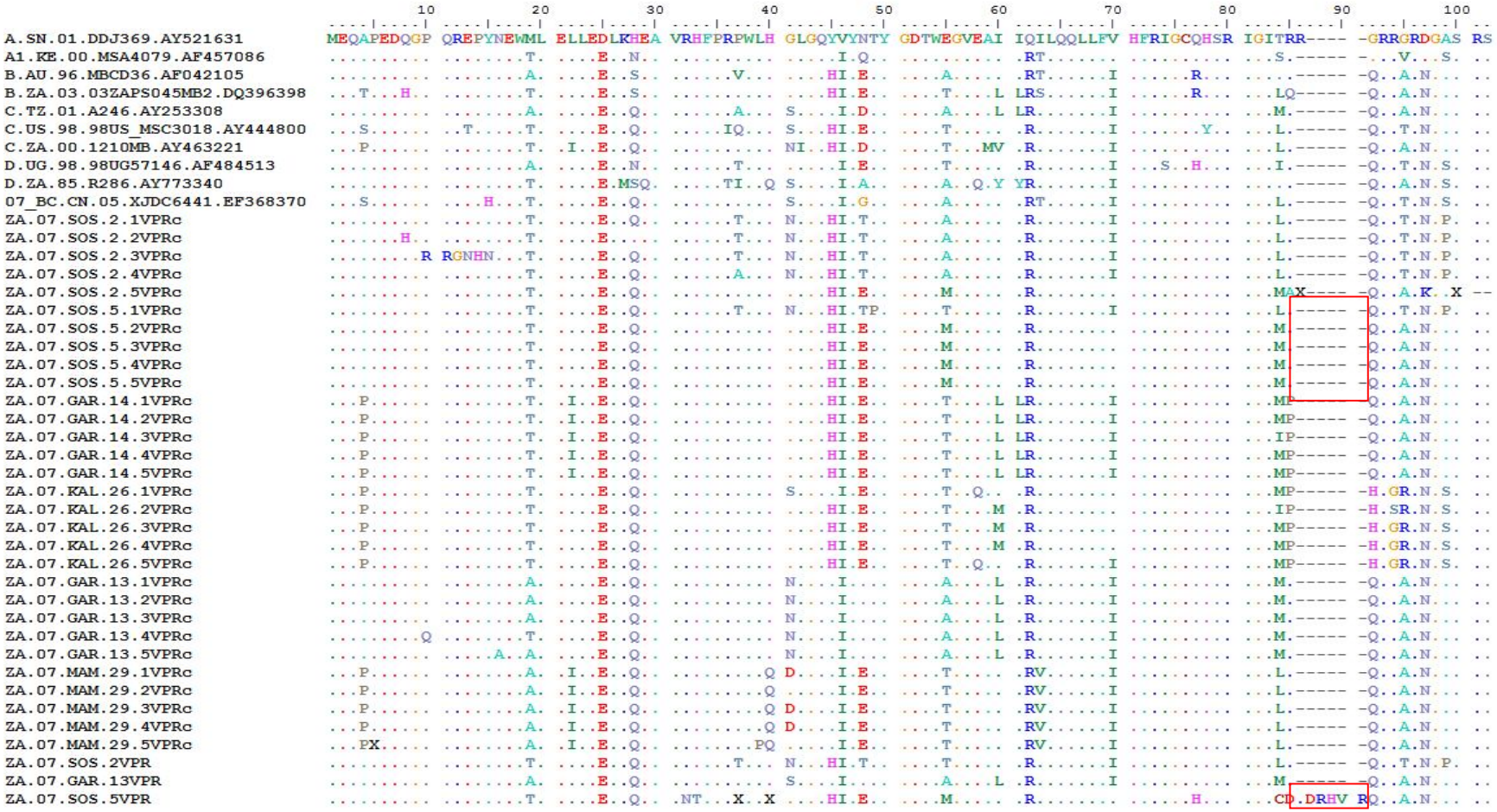


Figure 4.7: Graphic view of the amino acid alignments of *Vpr* sequences from 30 HIV-1 subtype C clones. The sequences are compared to reference strains from the Los Alamos Sequence Database. Amino acids insertion in sample ZA.07.SOS.5, which was not found in any of it clones.

4.4.5 Direct PCR sequences of *vpu* gene

The important sequence motifs in *vpu* gene also showed an overall degree of conservation. With these motifs; W28, 34-EYRKLL-39, 57-DSGNES-62 and 78-LRLL-81 showed to be conserved. Two sample sequences had insertions in their sequences. Sample ZA.07.SOS.2.VPU had an insertion of 3 amino acids at position 59 and this sequence also exhibited greater degree of divergence as compared to the reference strains and other sample sequences. Sample ZA.07.GAR.13.VPU had an insertion of 2 amino acids at position 2 (Figure 4.8).

Inter-samples sequence homology was evaluated with clustalW2 EBI software. The inter-sample sequence homology ranged from 73% to 92% for *vpu.* Sample ZA.07.SOS.2.VPU was not included when sequence homology was computed, as it did not give representative results of the studied population.

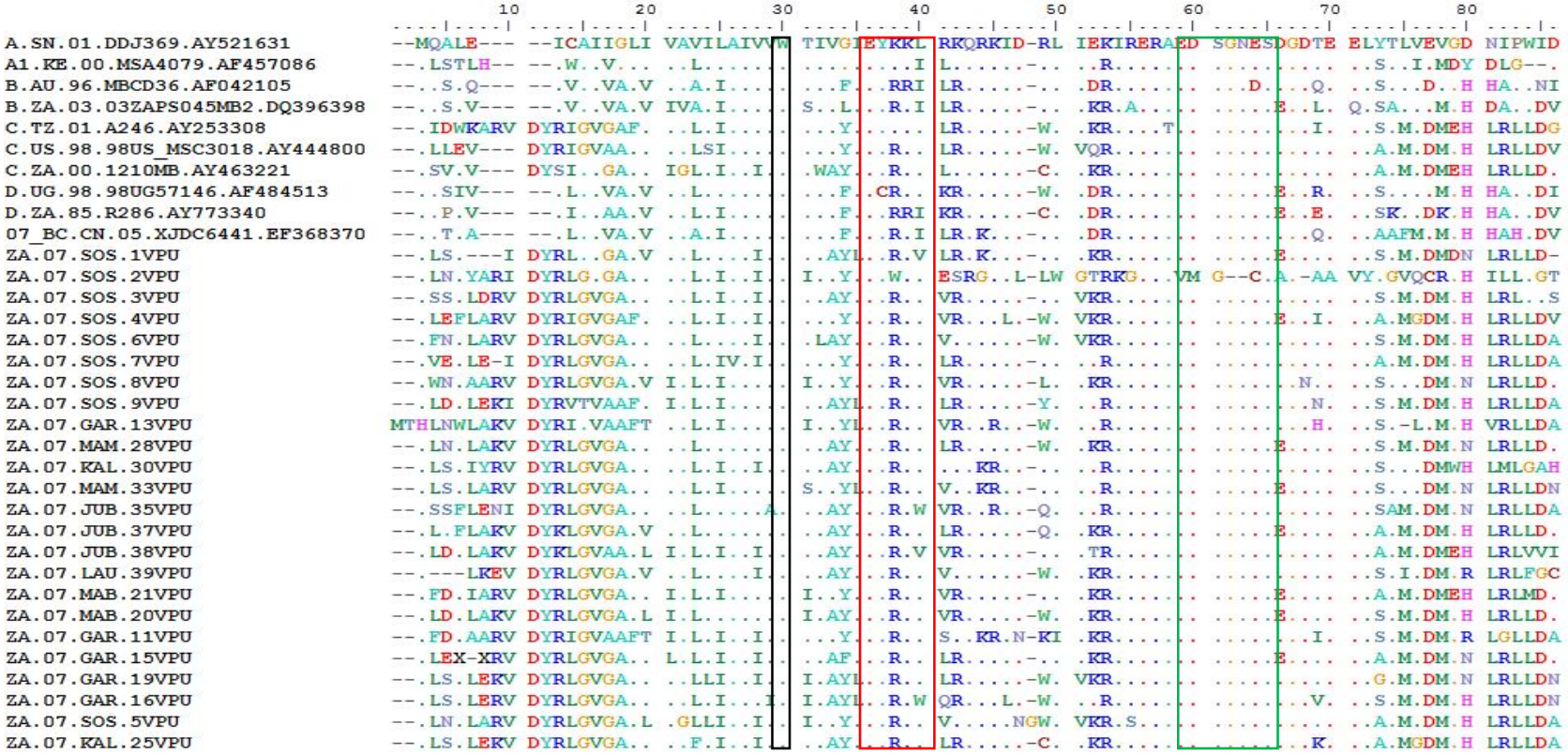


Figure 4.8: Graphic view of the amino acid alignments of *Vpu* sequences from 24 HIV-1 subtype C isolates. The sequences are compared to reference strains from the Los Alamos Sequence Database. (W28 = Tryptophan involved in channel gating, 34-EYRKLL-39 = Hinge region containing salt-bridge, 57-DSGNES-62 = Casein kinase II phosphorylation site containing critical serines and Putative Golgi export signal [78-LRLL-81]).

4.4.6 *Vpu* cloned sequences

Clone sequences for sample; ZA.07.SOS.2.VPU showed no high divergent characteristic that was observed in their direct PCR sequence. Three clone sequences of sample ZA.07.SOS.2.VPU had an insertion at position 5 of 5 amino acids, which were not present in direct PCR sequence. All clone sequences for sample; ZA.07.GAR.13.VPU had an insertion of 2 amino acids at position 2 like their direct PCR products. The direct PCR sequence had threonine and histidine insertion, which was also observed in 1 clone sequence. The other 4 clone sequences had serine and histidine insertion, this showed that the insertion is real for that sample (Figure 4.9).

Inter-clones sequence homology ranged from 83% to 100% for ZA.07.SOS.2.VPUc sequences; 86% to 99% for ZA.07.SOS.5.VPUc sequences; 96% to 98% for ZA.07.GAR.13.VPUc sequences; 98% to 100% for ZA.07.GAR.14.VPUc sequences; 96% to 100% for ZA.07.KAL.26.VPUc sequences; and 93% to 98% for ZA.07.MAM.29VPUc sequences. Some of the clone sequences had a 100% sequence homology showing that the sequences originated from the same virus (Table 8.1 to 86).

Figure 4.9: Graphic view of the amino acid alignments of *Vpu* sequences from 30 HIV-1 subtype C clones. The sequences are compared to reference strains from the Los Alamos Sequence Database. Sample ZA.07.SOS.2. The highly divergent amino acids sequence and amino acids insertion in sample ZA.07.GAR.13 and it clones.

Table 4.1: Summary of mutations within the accessory genes functional motifs and residues**.**

Gene	A A position	Mutations	Frequency
Vif	38	W38G	1/24
	95	S95N/R	3/24
Vpr	35	P35L	1/24
	73	R73K	1/24
	77	R77H/L/P	6/24
	80	R80S	1/24
Vpu	36	R36W	1/24
	38	L38V/W	4/24
	39	L39E/V/S/C/R	16/24
	58	S58A	1/24
	62	S62V	1/24
	78	L78V	2/24
	79	R79G/M	2/24
	81	L81W/M/C/V/P	4/24

4.5 Sequence analysis of the *pol* gene

The *pol* gene sequences were aligned together with reference strains from the Los Alamos sequence Database. The *pol* sequences were aligned on clustalW program. The *pol* gene was aligned in two sections (protease and RT regions). Drug resistant mutations were evaluated in both protease and RT regions. Nine sequences had one or two drug resistant mutations. Samples ZA.07.SOS.5.POL, ZA.07.SOS.7.POL and ZA.07.GAR.13.POL had a PI mutation T74S that are associated with potential low-level resistance to nelfinavir. Sample ZA.07.GAR.16.POL had L10I mutation (PI mutation) that is associated with PI resistance when present with other mutations and also had

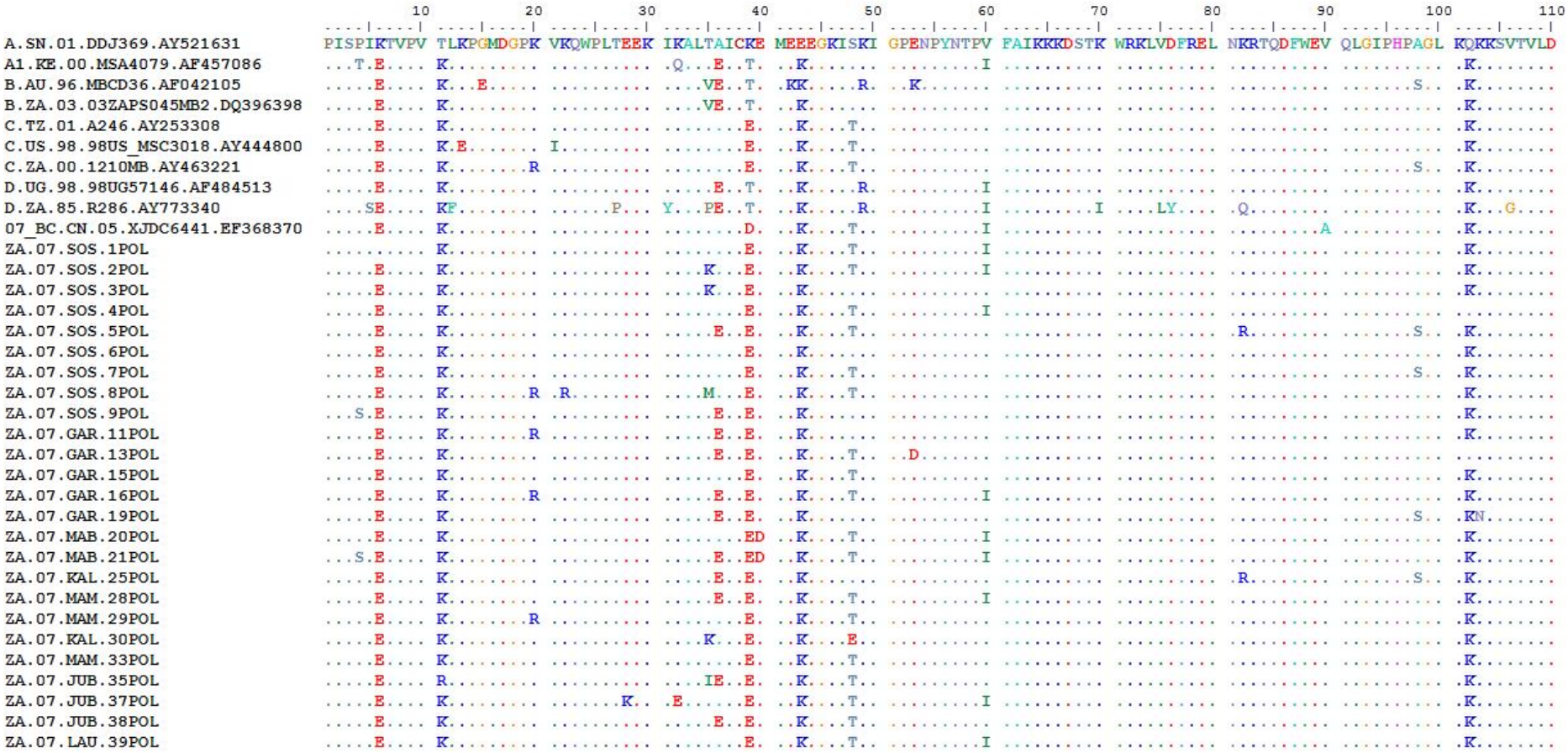
V179D mutation (NNRTI) that cause potential low-level resistance to delavirdine, efavirenz, etrvirine and nevirapine. Sample ZA.07.JUB.38.POL had E138A mutation which is associated with reduced etravirine response. Sample ZA.07.LAU.39.POL had a PI mutation that is associated with low-level resistant nelfinavir. Sample ZA.07.KAL.25.POL had Y318F mutation and it cause high level resistant to delavirdine,intermediate resistance to nevirapine and potential low-level resistance to efavirenz and etravirine. Sample ZA.07.GAR.19.POL had K103N and E138D mutations that cause potential low-level resistance to etravirine, high level resistance to delavirdine, efavirenz and nevirapine. Sample ZA.07.GAR.11.POL had Y181C mutation that cause high level resistance to delavirdine and nevirapine, and intermediate resistance to efavirenz and etravirine (Figure 4.10 and 4.11).

Table 4.2: Summary of drug resistant mutations

Resistant Mutations	**Number of samples**	**Affected Drugs**
T74S	3/25	Nelfinavir
L10I	1/25	It causes resistance when combined with other mutations
V179D	1/25	Delavirdine, efavirenz, etrvirine and nevirapine
E138A/D	2/25	Etravirine
Y318F	1/25	Delavirdine, nevirapine efavirenz and etravirine
Y181C	1/25	Delavirdine, nevirapine, efavirenz and etravirine
K103N	1/25	Etravirine, delavirdine, efavirenz and nevirapine

Inter-samples homology of the *pol* gene was evaluated and it ranged between 79% and 95%.

Figure 4.10: Graphic view of the amino acid alignments of *pol* (protease) sequences from 25 HIV-1 subtype C isolates. The sequences are compared to reference strains from the Los Alamos Sequence Database.


10 20 30 40 50 60 70 80 90 100 110
A.SN.01.DDJ369.AY521631 PISPIKTVPV TLKPGMDGPK VKQWPLTEEK IKALTAICKE MEEEGKISKI GPENPYNTPV FAIKKKDSTK WRKLVDFREL NKRTQDFWEV QLGIPHPAGL KQKKSVTVLD
A1.KE.00.MSA4079.AF457086 ...T.E.... K..... ..... .Q...E..T. ..K..... .....I .... .K.....
B.AU.96.MBCD36.AF042105 .....E.... K...E..... ....VE..T. .KK....R. ..K..... ....S.. .K.....
B.ZA.03.03ZAPS045MB2.DQ396398 .....E.... K..... ....VE..T. ..K..... .... .K.....
C.TZ.01.A246.AY253308 .....E.... K..... ....E. ..K...T.. .... .K.....
C.US.98.98US_MSC3018.AY444800 .....E.... K.E..... I..... .....E. ..K...T.. .... .K.....
C.ZA.00.1210MB.AY463221 .....E.... K.....R ....E. ..K...T.. ....S.. .K.....
D.UG.98.98UG57146.AF484513 .....E.... K..... ....E..T. ..K....R. .....I .... .K.....
D.ZA.85.R286.AY773340 ....SE.... KF..... .....P... Y...PE..T. ..K....R. .....I .....I ....LY.... .Q..... ..... .K...G....
07_BC.CN.05.XJDC6441.EF368370 .....E.... K..... ....D. ..K...T.. .....I ....A ..... .K.....
ZA.07.SOS.1POL ..... K..... ....E. ..K...T.. .....I .... .K.....
ZA.07.SOS.2POL .....E.... K..... ....K..E. ..K...T.. .....I .... .K.....
ZA.07.SOS.3POL .....E.... K..... ....K..E. ..K..... .... .K.....
ZA.07.SOS.4POL .....E.... K..... ....E. ..K...T.. .....I ....
ZA.07.SOS.5POL .....E.... K..... ....E..E. ..K...T.. .... .R..... .....S.. .K.....
ZA.07.SOS.6POL .....E.... K..... ....E. ..K..... .... .K.....
ZA.07.SOS.7POL .....E.... K..... ....E. ..K...T.. ....S.. .K.....
ZA.07.SOS.8POL .....E.... K.....R .R..... ....M..E. ..K...T.. .... .K.....
ZA.07.SOS.9POL ...S.E.... K..... ....E..E. ..K..... .... .K.....
ZA.07.GAR.11POL .....E.... K.....R ....E..E. ..K..... .... .K.....
ZA.07.GAR.13POL .....E.... K..... ....E..E. ..K...T.. ..D..... ....
ZA.07.GAR.15POL .....E.... K..... ....E. ..K...T.. .... .K.....
ZA.07.GAR.16POL .....E.... K.....R ....E..E. ..K...T.. .....I .... .K.....
ZA.07.GAR.19POL .....E.... K..... ....E..E. ..K..... ....S.. .KN.....
ZA.07.MAB.20POL .....E.... K..... ....ED ..K...T.. .....I .... .K.....
ZA.07.MAB.21POL ...S.E.... K..... ....E..ED ..K...T.. .....I .... .K.....
ZA.07.KAL.25POL .....E.... K..... ....E..E. ..K..... .... .R..... .....S.. .K.....
ZA.07.MAM.28POL .....E.... K..... ....E..E. ..K...T.. .....I .... .K.....
ZA.07.MAM.29POL .....E.... K.....R ....E. ..K...T.. .... .K.....
ZA.07.KAL.30POL .....E.... K..... ....K..E. ..K...E.. .... .K.....
ZA.07.MAM.33POL .....E.... K..... ....E. ..K...T.. .... .K.....
ZA.07.JUB.35POL .....E.... R..... ....IE..E. ..K...T.. .... .K.....
ZA.07.JUB.37POL .....E.... K..... .....K.. .E.....E. ..K...T.. .....I .... .K.....
ZA.07.JUB.38POL .....E.... K..... ....E..E. ..K...T.. .... .K.....
ZA.07.LAU.39POL .....E.... K..... ....E. ..K...T.. .....I .... .K.....


120 130 140 150 160 170 180 190 200 210 220
A.SN.01.DDJ369.AY521631 VGDAYFSVPL DEGFRKYTAF TIPSTNNETP GVRYQYNVLP QGWKGSPAIF QYSMTKILEP FRSQNPEIII YQYMDDLYVG SDL-EIG-QH RTKIEELRAH LLKWGLTTPD
A1.KE.00.MSA4079.AF457086
B.AU.96.MBCD36.AF042105
B.ZA.03.03ZAPS045MB2.DQ396398
C.TZ.01.A246.AY253308
C.US.98.98US_MSC3018.AY444800
C.ZA.00.1210MB.AY463221
D.UG.98.98UG57146.AF484513
D.ZA.85.R286.AY773340
07_BC.CN.05.XJDC6441.EF368370
ZA.07.SOS.1POL
ZA.07.SOS.2POL
ZA.07.SOS.3POL
ZA.07.SOS.4POL
ZA.07.SOS.5POL
ZA.07.SOS.6POL
ZA.07.SOS.7POL
ZA.07.SOS.8POL
ZA.07.SOS.9POL
ZA.07.GAR.11POL
ZA.07.GAR.13POL
ZA.07.GAR.15POL
ZA.07.GAR.16POL
ZA.07.GAR.19POL
ZA.07.MAB.20POL
ZA.07.MAB.21POL
ZA.07.KAL.25POL
ZA.07.MAM.28POL
ZA.07.MAM.29POL
ZA.07.KAL.30POL
ZA.07.MAM.33POL
ZA.07.JUB.35POL
ZA.07.JUB.37POL
ZA.07.JUB.38POL
ZA.07.LAU.39POL


230 240 250 260 270 280 290 300 310 320 330
A.SN.01.DDJ369.AY521631 KKHQKEPPFL WMGYELHPDK WTVQPIQ-LP DKDSWTVNDI QKLVGKLNWA SQIYPGIKVK QLCKLLRGAK ALTDVVVLTE EAELELAENR EILKDPVHGV YYDPSKDLVA
A1.KE.00.MSA4079.AF457086
B.AU.96.MBCD36.AF042105
B.ZA.03.03ZAPS045MB2.DQ396398
C.TZ.01.A246.AY253308
C.US.98.98US_MSC3018.AY444800
C.ZA.00.1210MB.AY463221
D.UG.98.98UG57146.AF484513
D.ZA.85.R286.AY773340
07_BC.CN.05.XJDC6441.EF368370
ZA.07.SOS.1POL
ZA.07.SOS.2POL
ZA.07.SOS.3POL
ZA.07.SOS.4POL
ZA.07.SOS.5POL
ZA.07.SOS.6POL
ZA.07.SOS.7POL
ZA.07.SOS.8POL
ZA.07.SOS.9POL
ZA.07.GAR.11POL
ZA.07.GAR.13POL
ZA.07.GAR.15POL
ZA.07.GAR.16POL
ZA.07.GAR.19POL
ZA.07.MAB.20POL
ZA.07.MAB.21POL
ZA.07.KAL.25POL
ZA.07.MAM.28POL
ZA.07.MAM.29POL
ZA.07.KAL.30POL
ZA.07.MAM.33POL
ZA.07.JUB.35POL
ZA.07.JUB.37POL
ZA.07.JUB.38POL
ZA.07.LAU.39POL

Figure 4.11: Graphic view of the amino acid alignments of *pol* (RT) sequences from 25 HIV-1 subtype C isolates. The sequences are compared to reference strains from the Los Alamos Sequence Database

4.6 Phylogenetic analysis of accessory genes

4.6.1 Phylogeny of the *vif* gene

4.6.1.1 Direct sequences versus clones

The construction of the phylogenetic tree was performed with Mega 4. The different HIV-1 strains from the Los Alamos sequence Database were used for subtyping. The phylogenetic analysis was performed with the Kimura two-parameter matrix, with 1000 bootstrap repeats.

All the study samples clustered with subtype C. The study samples did not form any South African cluster with South African references. The regional clusters in the South African samples were observed with samples ZA.07.JUB.37 and ZA.07.JUB.38. Samples ZA.07.SOS.2, ZA.07.SOS.3 and ZA.07.SOS.7 formed a cluster. Another regional cluster was seen with samples from mabopane; ZA.07.MAB.20 and ZA.07.MAB.21 (Figure 4.12).

The phylogenetic analysis of the clones showed that all belonged to HIV-1 subtype C. The clone sequences of the same sample have shown to form their own distinct clusters, with some demonstrating to be 100% identical (ZA.07.GAR.14). This confirmed that the clones originated from the same virus, as it was also shown by their homology. Higher degree of amino acid alignments conservation was observed in clones than direct PCR products (Figure 4.12 and 4.13).

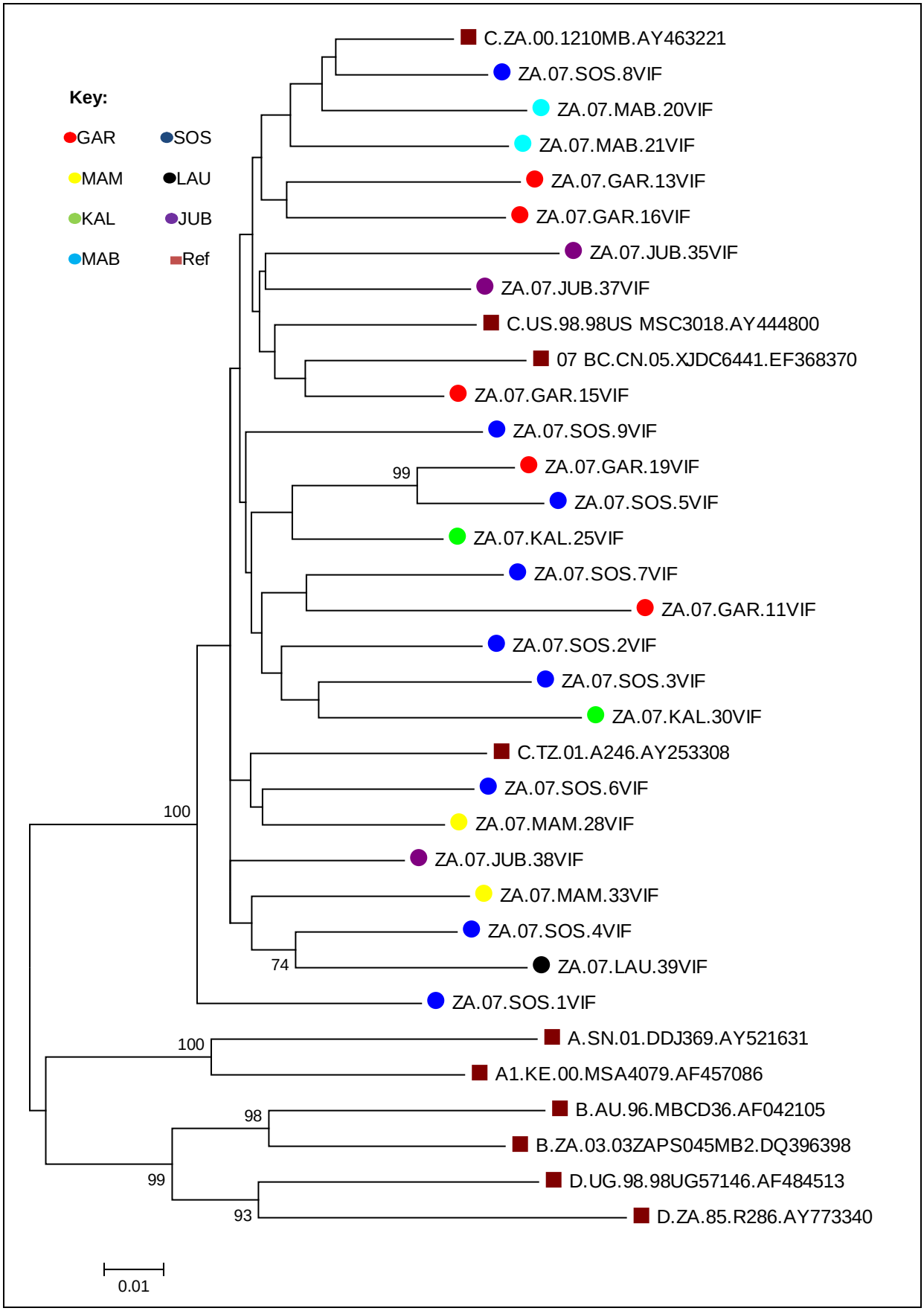


Figure 4.12: Phylogenetic relationship of *vif* gene samples (direct PCR products) with reference strains from the Los Alamos sequence Database.

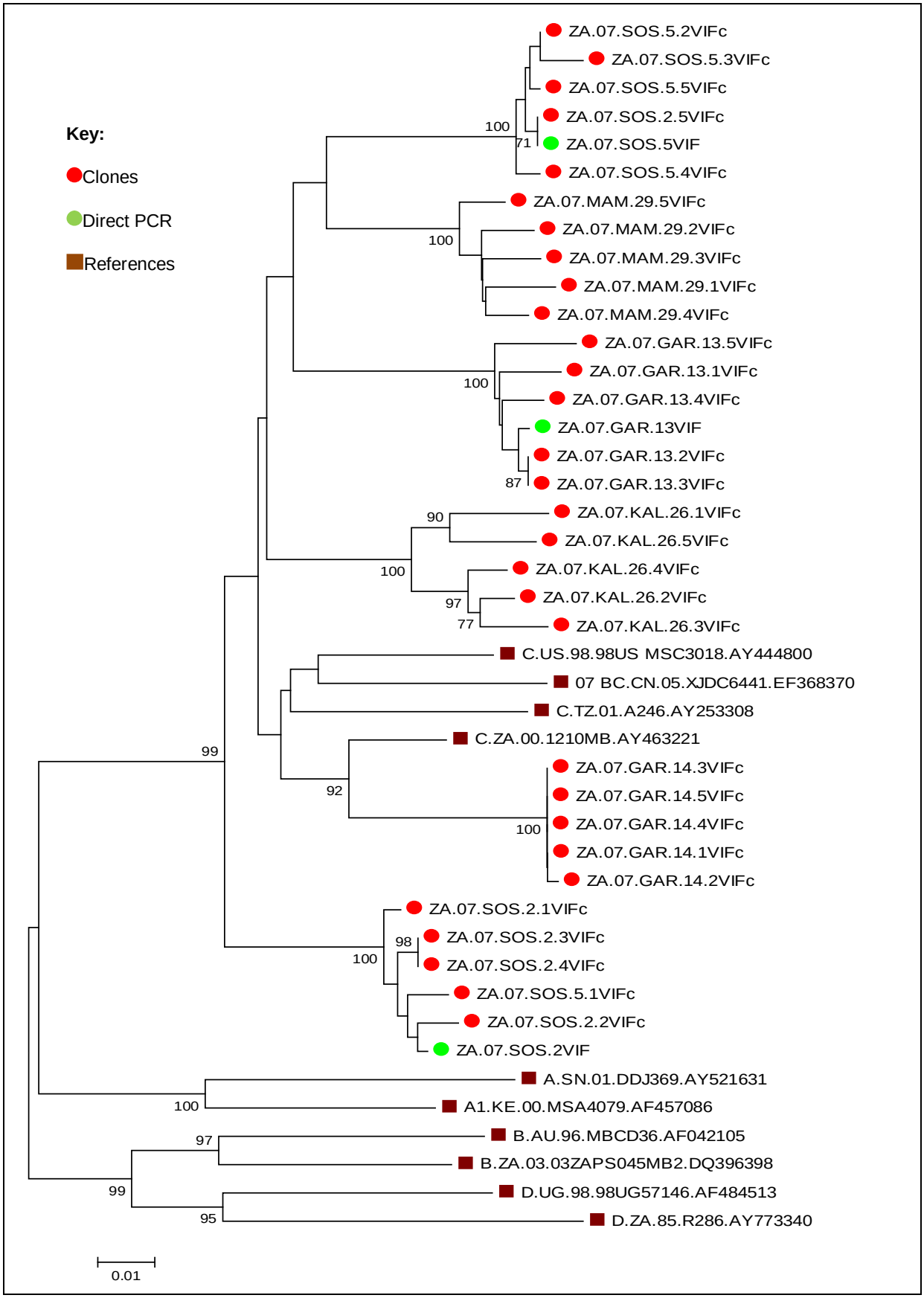


Figure 4.13: Phylogenetic relationship of *vif* gene clones with reference strains from the Los Alamos sequence database.

4.6.2 Phylogeny of the *vpr* gene

4.6.2.1 Direct sequences versus clones

All the study samples clustered with subtype C except sample ZA.07.SOS.4. Sample ZA.07.SOS.4 was blasted on bioAfrica and the blast results showed that is also subtype C. The study samples did not form any South African cluster. A couple of regional clusters in the South African samples were observed. Two samples from jubilee; ZA.07.JUB.37 and ZA.07.JUB.38 formed a cluster and samples from mabopane; ZA.07.MAB.20 and ZA.07.MAB.21 also formed a cluster (Figure 4.14).

The phylogenetic relationship confirmed that all the clones were HIV-1 subtype C. The cloned sequences of the same sample formed their own distinct clusters. Some of the clone sequences of the same sample showed to be 100% identical (ZA.07.SOS.5 and ZA.07.GAR.14). This showed that the clones originated from the same virus, as it was also shown by their homology. As expected, clones showed to be more conserved than direct PCR products (Figure 4.14 and 4.15).

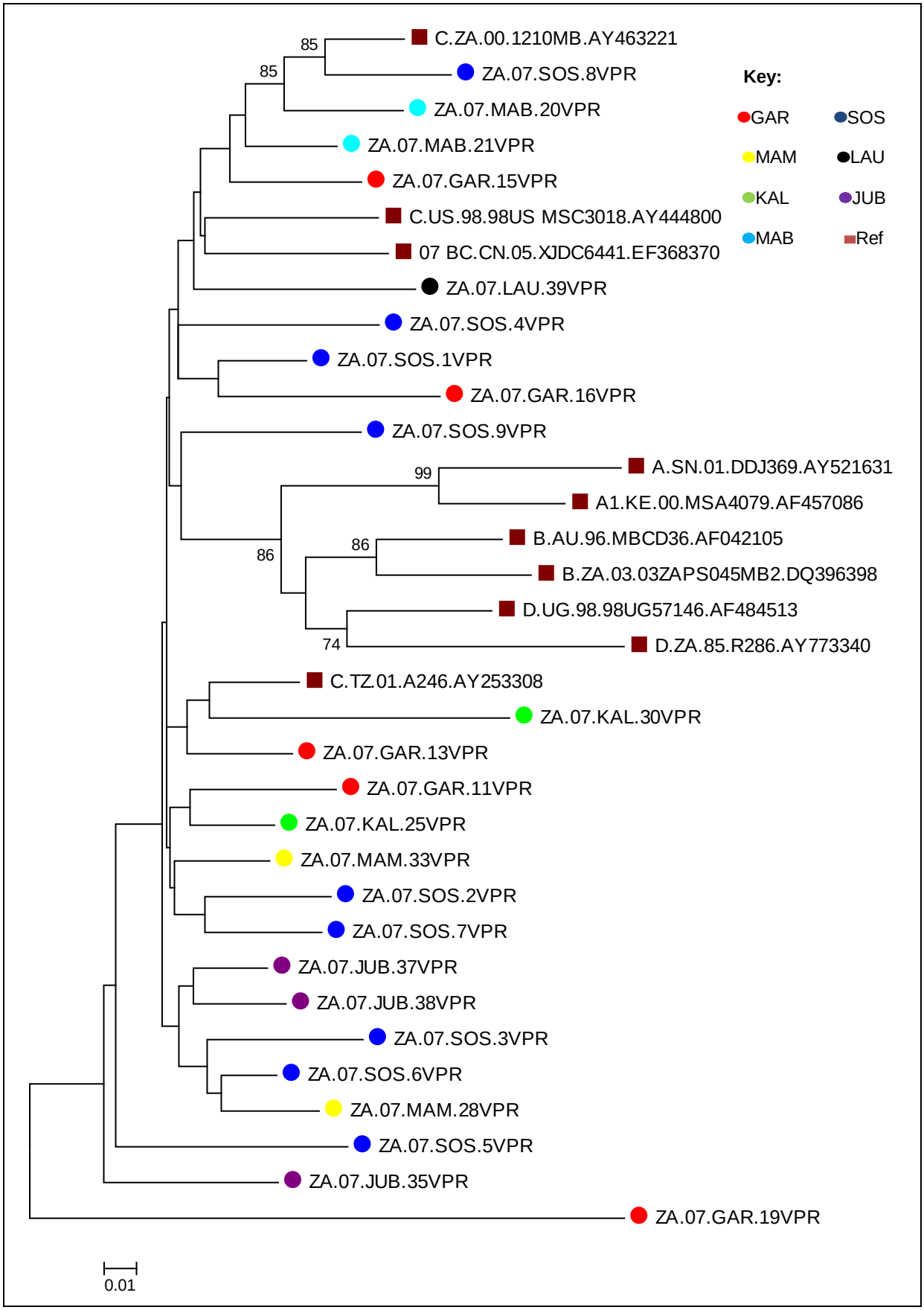


Figure 4.14: Phylogenetic relationship of *vpr* gene samples (direct PCR products) with reference strains from the Los Alamos sequence Database.

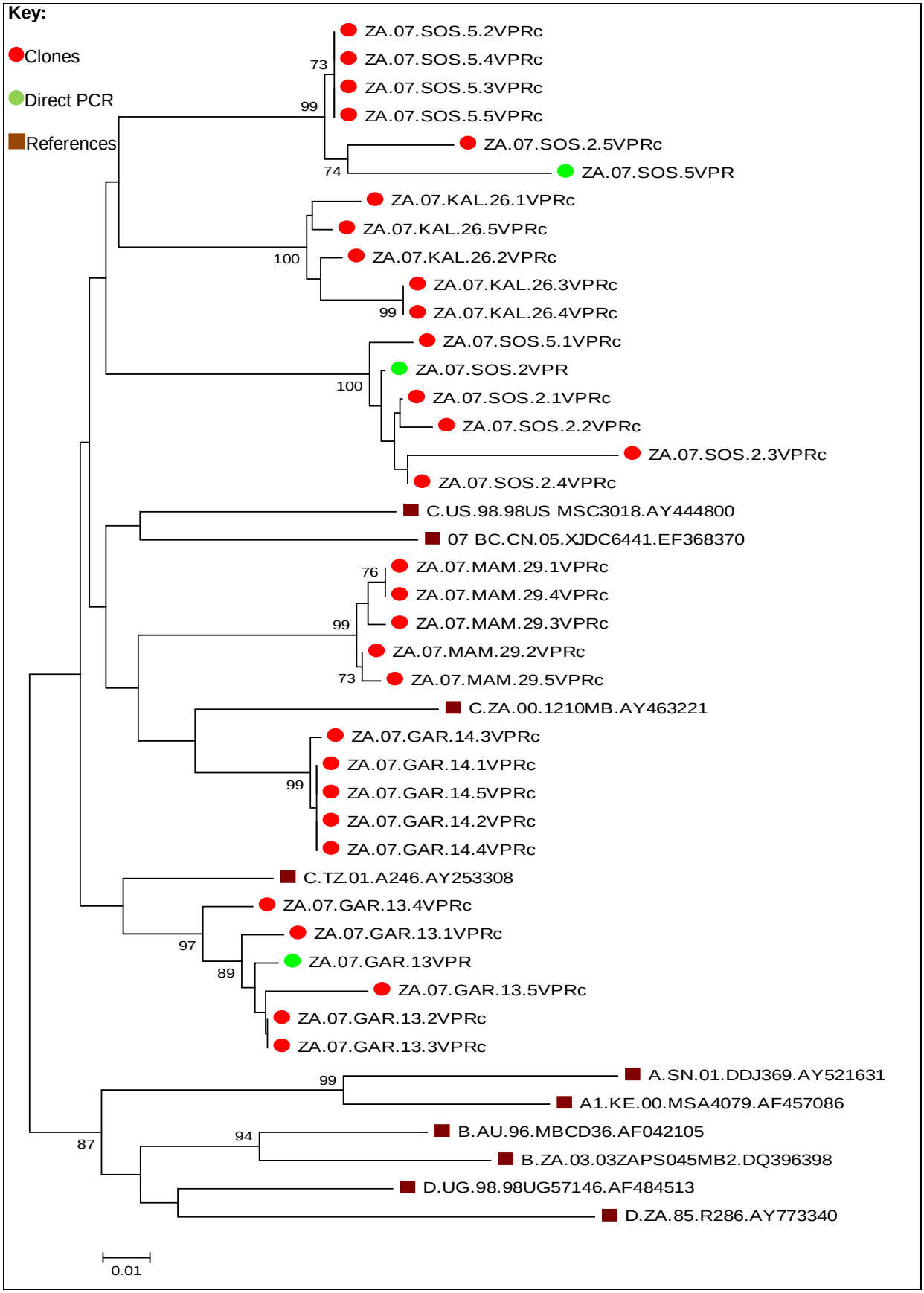


Figure 4.15: Phylogenetic relationship of *vpr* gene samples (clone products) with reference strains from the Los Alamos sequence Database.

4.6.3 Phylogeny of the *vpu* gene

4.6.3.1 Direct sequences versus clones

All the study samples clustered with HIV-1 subtype C. Sample ZA.07.SOS.2.VPU maid an out layer. The study samples did not form any South African cluster. Two regional clusters in the South African samples were observed with samples ZA.07.MAB.20.VPU and ZA.07.MAB.21.VPU. Samples; ZA.07.GAR.11.VPU and ZA.07.GAR.16.VPU also formed a regional cluster (Figure 4.16).

The cloned sequences of the same sample formed their own distinct clusters. Most of the *vpu* clones showed to be closely related. Some of the cloned sequences showed to be 100% identical, confirming that they originated from the same virus. The clones showed lesser diversity than direct PCR products (Figure 4.16 and 4.17).

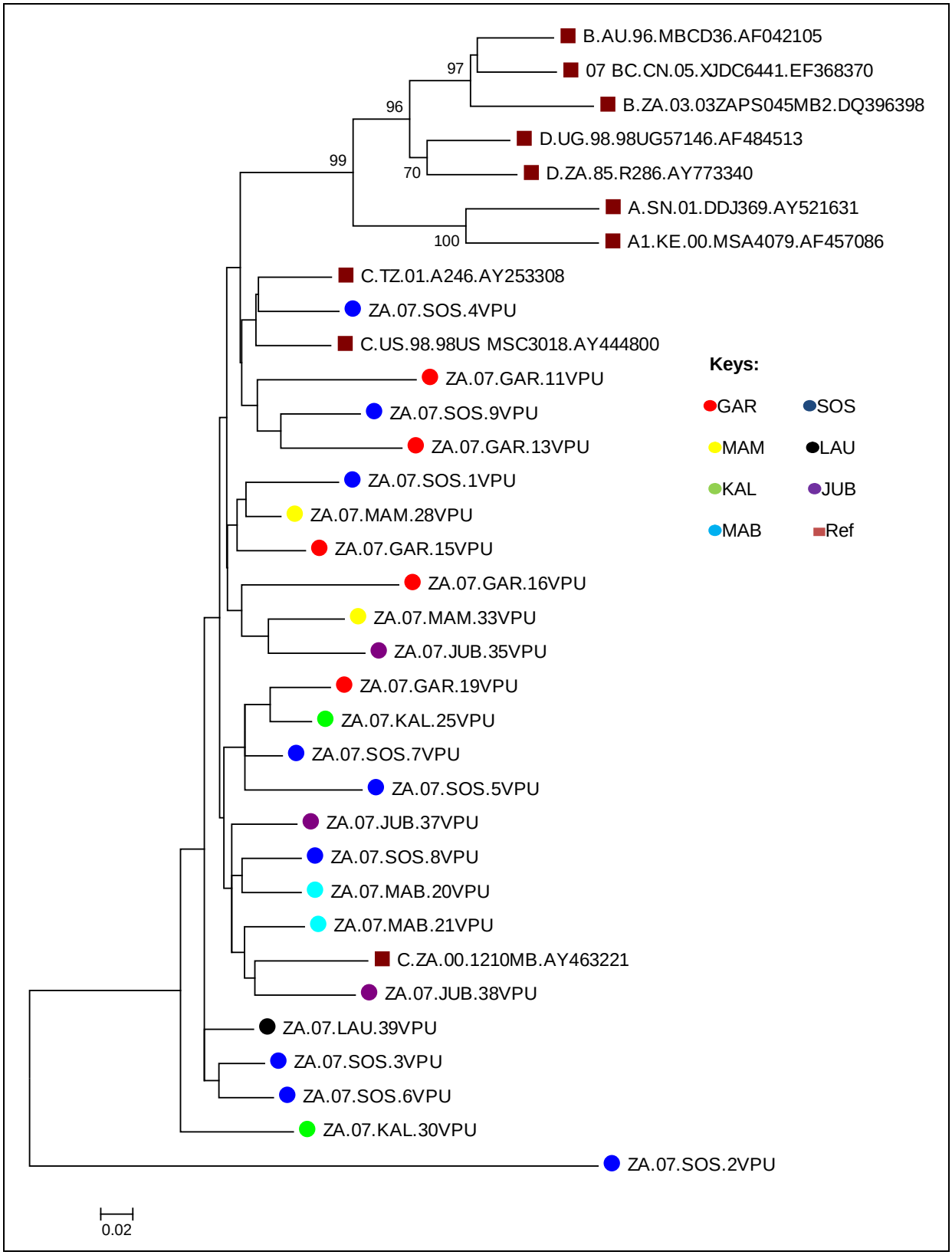


Figure 4.16: Phylogenetic relationship of *vpu* gene samples (direct PCR products) with reference strains from the Los Alamos sequence Database.

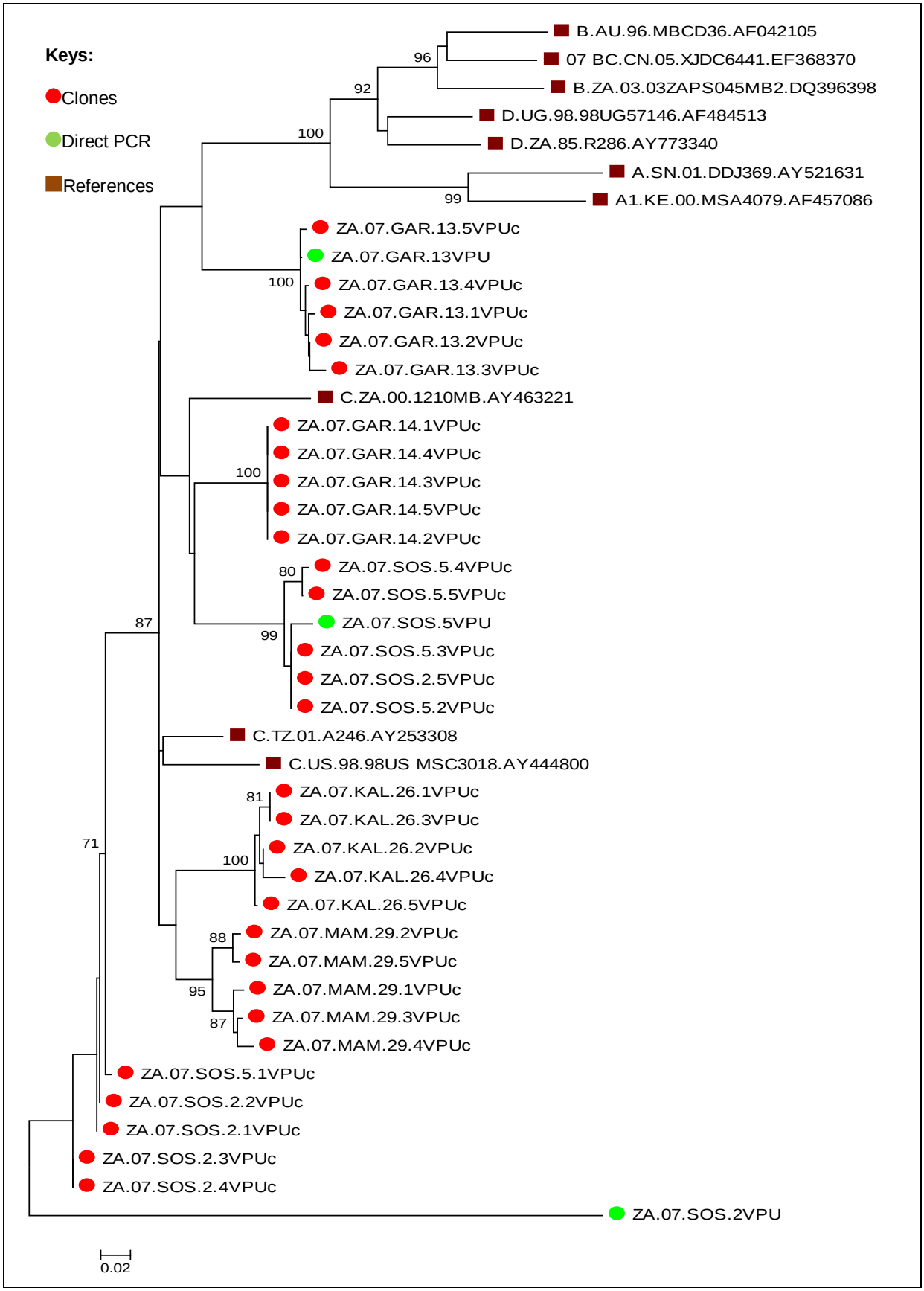


Figure 4.17: Phylogenetic relationship of *vpu* gene samples (clone products) with reference strains from the Los Alamos sequence Database.

4.7 Phylogeny of the *pol* gene

The study samples were evaluated with the reference strains from the Los Alamos Sequence Database for subtype. All the samples were subtype C by phylogenetic relationship.

The study samples did not form any South African cluster, i.e they intermingled with reference sequences from other parts of the world. Samples; ZA.07.SOS.5POL and ZA.07.SOS.7POL; ZA.07.MAB.20POL and ZA.07.MAB.21POL; ZA.07.SOS.4POL and ZA.07.SOS.6POL; ZA.07.SOS.2POL and ZA.07.SOS.3POL formed regional clusters in the South African samples (Figure 4.18).

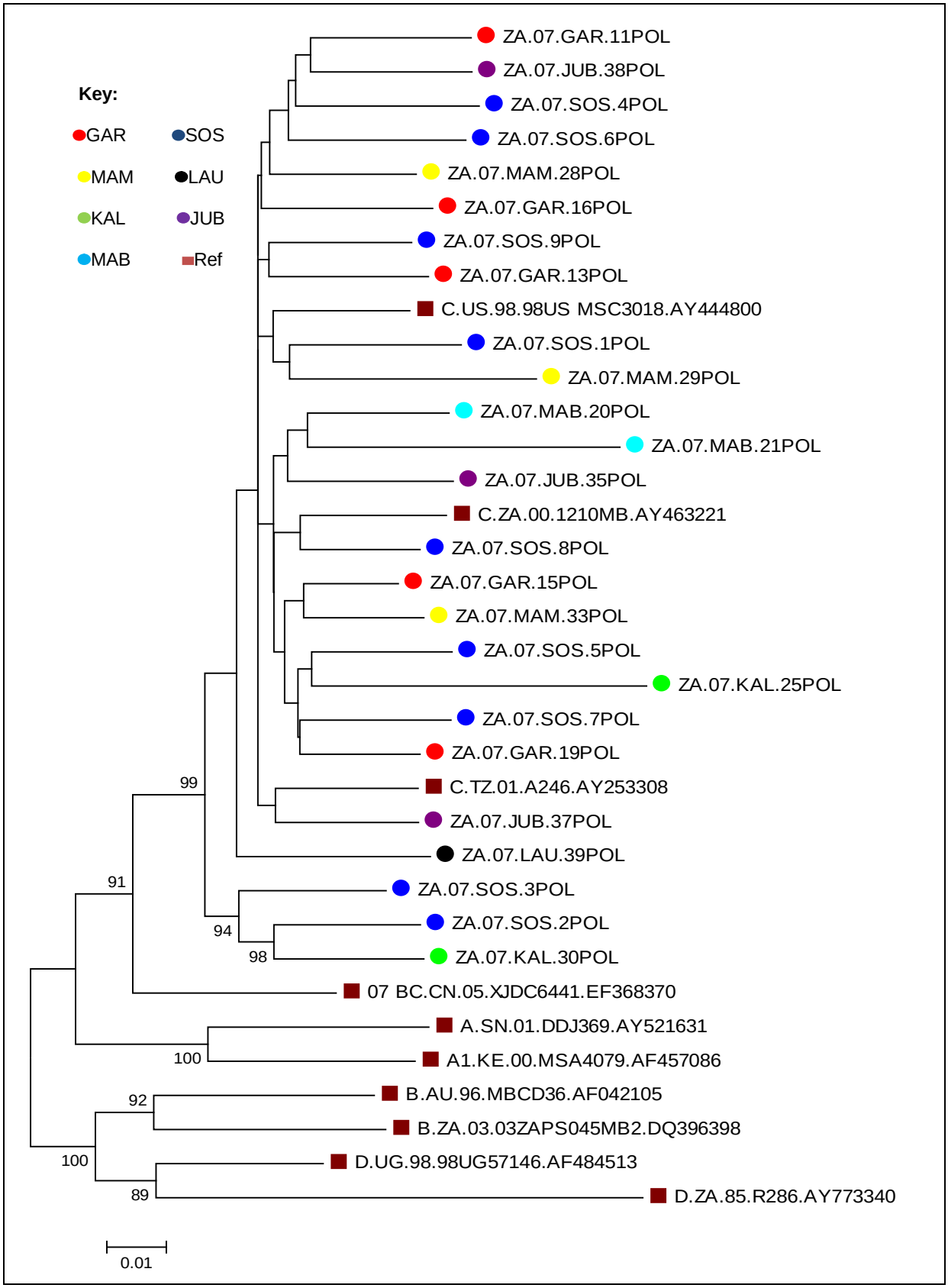


Figure 4.18: Phylogenetic relationship of *pol* gene samples with reference strains from the Los Alamos sequence Database.

4.8 Comparison of genetic diversity between accessory and structural genes

The genetic diversity of HIV-1 subtype C was compared between accessory (*vif*, *vpr* and *vpu*) and structural (*pol*, *gag* and *env*) genes. The *gag* sequences showed lesser genetic diversity (89% to 96%), as compared to the *vif* sequences (88% to 96%), the *vpr* sequences (84% to 94%), the *pol* sequences (79% to 95%), the *env* sequences (83% to 93%) and finally the *vpu* sequences (73% to 92%). Generally *gag*, *vif* and *vpr* showed to be more conserved. In addition, a number of conserved regions were observed in all the six genes (*vif*, *vpr*, *vpu*, *gag*, *env* and *pol*). The genetic diversity between the genes was clearly seen on the phylogenetic analysis, with the *vpu* phylogeny more branched as compared to the *env* phylogeny, the *pol* phylogeny, the *vpr* phylogeny, the *vif* phylogeny and lastly the *gag* phylogeny.

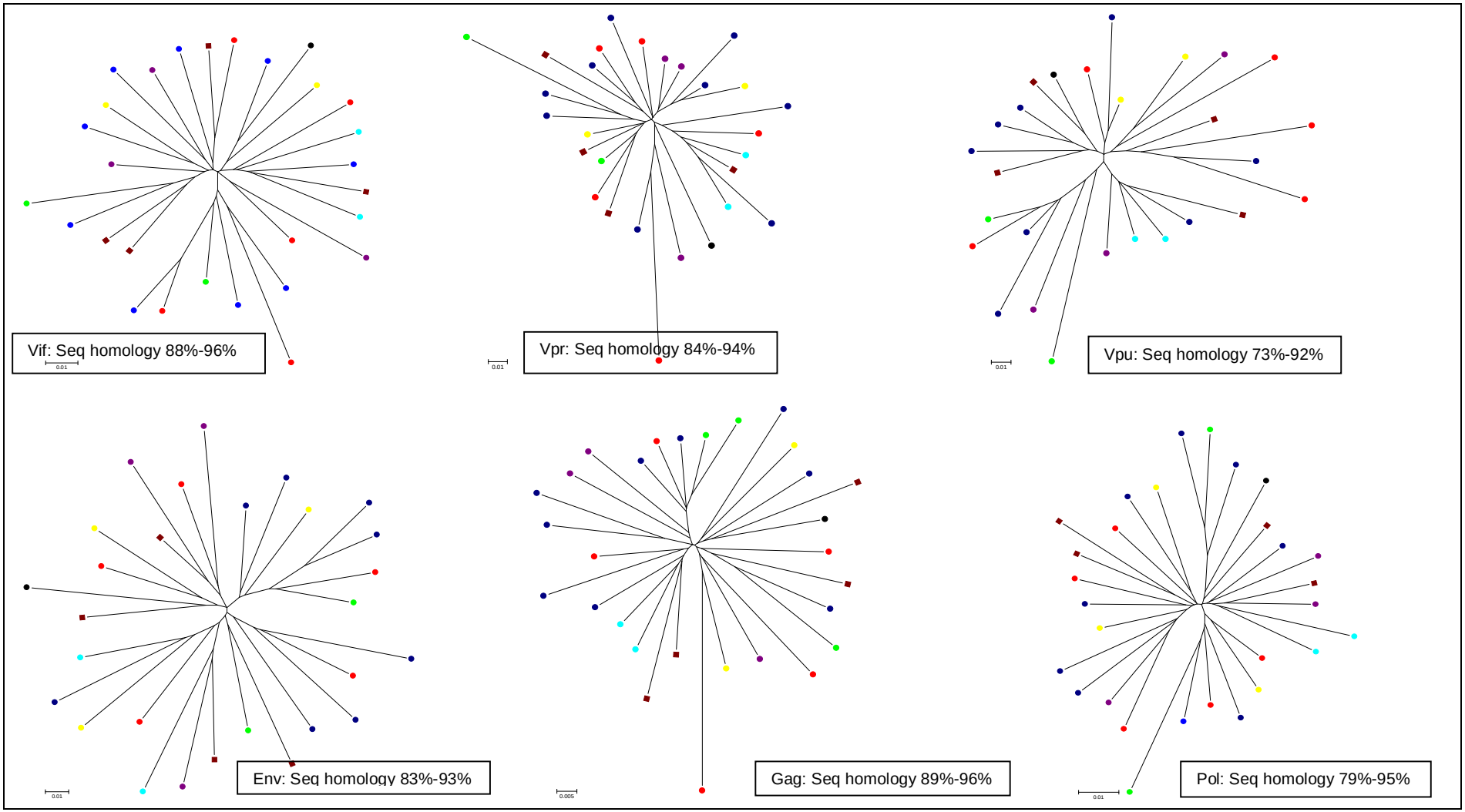


Figure 4.19: Phylogenetic comparison between accessory and structural genes. The studied samples are shown in round shapes and references in squares.

CHAPTER FIVE

5. DISCUSSION

5.1 Overview of the study

This study sought to amplify and analyse the genetic sequence diversity of HIV-1 subtype C accessory genes (*vif*, *vpr* and *vpu*) to assess motifs and residues associated with key biological functions on these genes. This study further sought to compare the degree of genetic diversity between the accessory gene and structural genes by evaluating their phylogenetic relationship and sequence homologies. A total of 25 samples which were previously sequenced for both *gag* and *env* genes were used in this study. All the 25 samples were amplified for accessory genes and *pol* gene. The rationale for amplifying and sequencing the *pol* gene was to obtain data for all three structural genes: *env*, *gag* and *pol*.

For this analysis, a 1546bp accessory genes and an 1856bp *pol* gene were amplified from each sample. Subtyping was achieved using phylogenetic inferencing and the REGA software on BioAfrica website. All the sequences were found to be subtype C. The inter-sample sequence homology ranged from 89% to 96% for *gag*, 88% to 96% for *vif*, 84% to 94% for *vpr*, 79% to 95% for *pol*, 83% to 93% for *env* and 73% to 92% for *vpu*. The inter-sample sequence homology showed that *gag* was the most conserved and *vpu* as the least conserved of the six genes.

Three samples that had insertions in their sequences were cloned for accessory genes. One sample whose primer sequences did not form a contiguous sequence and two samples which were previously cloned for *gag* and *env* were cloned. The phylogenetic relationship of the clones revealed that all clones were subtype C. There was no distinct South African clustering that was observed from the study samples and the reference sequences from South Africa. All the clones from the same sample formed distinct clusters

except for two clones ZA.07.SOS.2.5 and ZA.07.SOS.5.1. These clones were shown to have swapped clusters, with ZA.07.SOS.2.5c clustering with all the sample ZA.07.SOS.5 clones and ZA.07.SOS.5.1c clustering with all the ZA.07.SOS.2 clones (Figure 4.13, 4.15 and 4.17).

5.2 Detection and molecular characterisation of accessory genes

The nested PCR assay for accessory genes (*vif, vpr* and *vpu*) was optimized with primers from Bell et al (2007). The PCR assay covered the whole region of *vif, vpr* and *vpu*, producing a 1546bp fragment. Using the optimized PCR assay, accessory genes from all samples were successfully amplified. The comparison of the accessory genes PCR assay with *pol* PCR assay showed accessory genes PCR assay to be more sensitive, since *pol* PCR assay had to be repeated on some of the samples before successful PCR amplification.

Amplifying accessory genes as one fragment was easier, and proved to save time and reagents. The accessory genes PCR assay targeted *vif*, *vpr* and *vpu* of the same virus. Base calling for accessory genes saved time since it was performed on one fragment for *vif*, *vpr* and *vpu* by simply shifting the frame when editing overlapping genes, *vif* and *vpr.*

All the samples selected for cloning; ZA.07.SOS.2, ZA.07.SOS.5, ZA.07.GAR.14, ZA.07.GAR.13, ZA.07.KAL.26 and ZA.07.MAM.29 were successfully cloned for accessory genes and the required number of clones was successfully recovered from each sample. Samples ZA.07.KAL.26 and ZA.07.MAM.29 were previously cloned for *gag* and *env* genes. But not all the *env* gene clones recovered could be sequenced and analysed. This shows that the accessory genes and *gag* gene are easier to clone, and this may be due to the conserved nature of these genes.

The accessory genes were sequenced using second round primers (Table 3.4). The accessory genes were easily sequenced with two primers and the

chromatogram were usually clean. As compared to the *pol* sequencing that employed five primers (Table 3.5), contiguous sequences for accessory genes were much cleaner and base calling was performed with ease. Sample ZA.07.GAR.14 (accessory genes) did not form contiguous sequence for direct PCR but cloned sequences formed clean contiguous sequences. The accessory genes may have lesser variants as compared to the *pol* gene, which is why accessory genes chromatograms were more clear. Sequencing of the accessory genes as one fragment proved to be less cost effective.

5.3 Conserved motifs and important mutations of accessory genes

Vif gene

Vif is essential for efficient viral replication in natural target cells. It functions to prevent action of APOBEC3G (cellular inhibitor of HIV replication). Without *vif*, APOBEC3G is packaged into virions as a stable complex with the viral core. This affects the deamination of viral cytidines to uracil during the subsequent round of viral replication, thus leading to production of non-functional virions. Thus, *vif* bound APOBEC3G is degraded in the proteosome, permitting complete viral replication, thus increasing the infectivity of the HIV-1 (Wang et al, 2005; Mehle et al, 2007; Carr et al, 2008).

Direct observation of the *vif* sequence alignment in this study revealed a great degree of conservation of amino acids W5, W11, W21and W38 which are involved in recognition and suppression of APOBEC3G. Except for sample ZA.07.MAB.21VIF that had W38G mutation, the conserved motif 42-HH-43 is also conserved in these samples.

Important region 88-EW-89 residues were highly conserved in this study. The 88-EW-89 residues were shown to be involved in enhancing steady state expression of the virus. HCCH zinc motif and *vif* dimerization site revealed to be conserved. The SLQYLA motif in this study exhibited a degree of conservation and induced mutations in this motif have been shown to result in a less pathogenic virus as compared to the wild type. There was also reduced spreading of the virus to the visceral organs (Schmitt et al, 2009). The PPLP

motif was also conserved (Figure 4.4). This degree of conservation was reported by other studies (Wang et al, 2005; Bell et al, 2007). Mutations in these highly functional and conserved regions have a negative impact on the virus, suggesting these motifs are good candidates for HIV intervention. The amino acid sequence alighments revealed that intra-sample clones shared the same characteristics, clearly showing that the clones were from one virus or closely related virus. The conserved regions in *vif* gene were also shown to be conserved in all the clones, thus comfirming the importance of these motifs (Figure 4.5).

Vpr gene

When vpr protein is expressed, it causes G2 cell-cycle arrest, before the integration step of the viral DNA genome. *Vpr* can still induce G2 cell-cycle arrest while packaged into a virion. *Vpr* can also induce apoptosis of CD4+ T cells, thus weakening the immune system and contributing to the pathogenesis of the HIV-1 (Le Rouzic and Benichou, 2005; Andersen et al, 2008).

The important regions in the *vpr* gene exhibited some degree of conservation. P35 and H71 residues were highly conserved, and are involved in cyclophilin A binding and *vpr* dimer stacking respectively. This observation was also seen by Bell et al (2007), making these regions better candidates for HIV-1 subtype C interventions.

The Q3R mutation was found in two samples: ZA.07.SOS.4 and ZA.07.JUB.38. The Q3R mutation is associated with long term non-progression and is also associated with high viremia with no significant loss of CD4 lymphocytes (Somasundaran et al, 2002). These two samples (ZA.07.SOS.4 and ZA.07.JUB.38) could be from long term non-progressors. The R77Q mutation in *vpr* gene was observed in 15 sequences; ZA.07.SOS.1, ZA.07.SOS.2, ZA.07.SOS.3, ZA.07.SOS.6, ZA.07.SOS.8, ZA.07.SOS.9, ZA.07.GAR.11, ZA.07.GAR.13, ZA.07.GAR.15, ZA.07.ODI.20, ZA.07.KAL.25, ZA.07.MAM.28, ZA.07.MAM.33, ZA.07.JUB.37 and

ZA.07.JUB.38. Patients with R77Q mutation have been reported to be associated with long term non-progression (Lum et al, 2003). These results also revealed that Q3R and R77Q mutations which are both associated with long term non-progression, R77Q mutation was observed in 15 samples of which one (ZA.07.JUB.38) of the fifteen shown to have both Q3R and R77Q mutations (Figure 4.6).

Sample ZA.07.SOS.5 had an insertion of 6 amino acids at position 86, but the intra-sample clones did not show this insertion (Figure 4.7). This finding clearly supports that in one HIV positive patient there are quasispecies of HIV circulating. The amino acid sequence alighments revealed that intra-sample clones shared the same characteristics as parental samples (Figure 4.7).

Vpu gene

Vpu functions to enhance virion release by ion channel activity and down regulation of CD4+ expression from the surface of infected cells. The association of *vpu* with viral release was demonstrated in an *in vitro* study with viruses lacking the functional *vpu*. In addition, *vpu* expression has been associated with a reduction in syncytia formation of infected cells (Klimkait et al, 1990; Schubert and Strebel, 1994; Estrabaud et al, 2007). Vpu protein was associated with accumulation of substrates like β-catenin, IκBα, and ATF4, which have a role in innate immunity, cancer and several autoimmune disorders (McCormick-Davis et al, 2000; Bour and Strabel, 2003).

The important sequence motifs and residues in *vpu* gene revealed an overall degree of conservation. W28 residue and 57-DSGNES-62 were highly conserved. The conserved nature of these motifs suggests that they may be good candidates for HIV intervention.

The motifs, 34-EYRKLL-39 and 78-LRLL-81 were conserved with single or double amino acid(s) change in some samples. In 34-EYRKLL-39 motif mutations were common in position 39 and motif 78-LRLL-81 had more mutations in amino acid position 81 (Table 4.1). Premature stop codon

mutations in *vpu* gene did not occur in this study as was observed by Komoto et al (2005).

However two sequences (ZA.07.SOS.2 and ZA.07.GAR.13) had insertions. Sample ZA.07.SOS.2 had an insertion of 3 amino acids at position 59 and this sequence also exhibited greater degree of divergence as compared to the reference strains and other sample sequences. The intra-sample clones of sample ZA.07.SOS.2 did not show any of the characteristics that were observed in direct PCR sequence. This findings suggest that the hypermutated sequence that was sequenced may be ineffective, thus rendering it to suppression by the immune system. Sample ZA.07.GAR.13 had an insertion of 2 amino acids at position 2, this was also observed in all of its intra-sample clones. The quasispecies with this insertion seem to better survive the immune system in this patient as it was observed in all five intra-sample clones (Figure 4.8 and 4.9).

***Pol* gene**

The *pol* gene encodes three enzymes; protease, reverse transcriptase and integrase. Protease functions to cleave gag and gag/pol polyproteins. Reverse transcriptase functions to convert the viral RNA to DNA while integrase functions to integrate the viral DNA into the host genome (Chan et al, 1999).

The *pol* gene was aligned in two sections (protease and RT regions). Drug resistant mutations were evaluated in both protease and RT. Nine sequences had one or two drug resistant mutations in their sequences. Three sequences had T74S mutation that cause low-level resistance to nelfinavir and two sequences had E138A mutation which is associated with reduced etravirine, as reported by Johnson and colegues in “Update of the Drug Resistance Mutations in HIV-1:December 2008” (Bessong et al, 2006) (Figure 4.10 and 4.11). The National Antiretroviral Treatment Guidelines showed that nelfinavir and etravirine are not encluded in the public sector regimens, and finding resistance mutations for nelfinavir and etravirine suggest that the patients

might have acquired these resistance mutations from exposure to someone infected with mutant viruses. Or these mutations may have also been enduced by the regimens used in South African public sectors (Jacobs et al, 2008). Although the findings may not always reflect the treatment outcomes of patients, resistance testing is important before prescribing or switching ARV treatment, as resistant mutations lead to treatment failure in the long run. The results support the need for resistant testing before starting or switching ARV treatment.

5.4 Phylogenetic analysis

All 25 samples were subtype C by phylogenetic relationship and the REGA software on BioAfrica website. This further confirms that HIV-1 subtype C is the most prevalent subtype in South Africa. There was no distinct South African clustering from the study samples and the reference sequences from South Africa. There was regional clustering that was observed in the studied samples (Figure 4.12, 4.14 and 4.16).

The regional clustering was observed in all the six genes studied (*vif, vpr, vpu, pol, gag* and *env*). Two samples from Mabopane (ZA.07.MAB.20 and ZA.07.MAB.21) formed regional clusters in all the six genes. Although the number of samples forming regional clusters was small, there is a possibility of founder effect spread of genetically distinct viruses (Figures 4.12, 4.14, 4.16, 4.18, 8.5 and 8.7). *Gag, vif* and *vpr* are the most conserved genes of the studied genes by phylogeny and sequence homology (Figure 4.19).

The phylogenetic analysis of the clones showed that all the clones were subtype C. The cloned sequences of the same sample formed their own distinct clusters, with some depicting 100% identity. Most of the *vpu* clones were closely related, confirming the clones originated from the same virus (Figure 4.13, 4.15 and 4.17).

The comparison of accessory genes and structural genes revealed that generally accessory genes are more conserved than structural genes, with exception to the *gag* gene. *Vpu* appears to be more diverse compared to *vif* and *vpr* genes, as previously reported by Bell et al (2008). Similar results were observed in HIV-1 group O, where *vif* was more conserved, followed by the *vpr* gene and finally the *vpu* gene (Vallejo et al, 2003). *Gag* and *pol* genes are conserved compared to *env* gene, these findings support Seibert et al (1995) since *gag* and *pol* genes were shown to be subjected to purifying selection. Finally, the comparison of accessory genes and structural genes showed that *gag, vif* and *vpr* are more conserved with *pol*, *env* and *vpu* being the lesser conservered genes. The inter-sample sequence homology ranged from 89% to 96% for *gag*, 88% to 96% for *vif*, 84% to 94% for *vpr*, 79% to 95% for *pol*, 83% to 93% for *env* and 73% to 92% for *vpu*. Thus, inter-sample sequence homology confirmed that *gag* was the most conserved and *vpu* as the lesser conserved of the six genes. (Figure 4.19).

5.5 Implication of the findings on HIV prevention and control

Finding an HIV vaccine has been the centre of many studies since the discovery of HIV/AIDS. An epitope that elicits neutralizing antibodies against HIV and provides cross protection against many subtypes is the key factor in combating HIV/AIDS. Vaccines that target specific genes of HIV that shows to serve an important function in HIV pathogenesis like *pol* and *env*, have been targeted for development of HIV vaccine in many studies for many years. All these studies have paved the way for genes that were overlooked, that is, the accessory genes. It is then important to determine the genetic diversity of HIV-1 subtype C accessory genes. As they are suggested to be the next therapeutic intervention and vaccine candidates for HIV.

The *pol* gene diversity is assessed in drug resistance studies because its enzymes are targeted by ARV's. A study showed that mutations that increase polyprotein processing in the natural substrate not the enzymes represent a novel mechanism which develops resistance to PI's (Nijhuis et al, 2007). This finding shows the importance of evaluating HIV genes for their diversity.

A study on *gag* gene diversity revealed that mutations that occur in the highly conserved major homology region (amino acid 285-304) reduce membrane affinity *in vitro*. These mutations also showed to reduce efficiency of mature particle assembly in transfected mammalian cells (Ebbets-Reed et al, 1996). Another study illustrated a contribution of *gag* to the selection of HIV-1 resistance to PI's (Dam et al, 2009).

The accessory genes have shown to serve important roles in HIV pathogenesis. The studies where mutations were introduced in some of the important regions of accessory genes demonstrated a reduction in the virulence of the virus. These mutations also reduced the spread of the virus to many body organs. The accessory genes diversity is vital in understanding these genes.

Functional study on *vif* 144-SLQ-146 and 161-PPLP-164 motifs showed that when mutations are introduced in any of the two motifs, an inactive or non-functional proteins are produced. Mutations in these two motifs (144-SLQ-146 and 161-PPLP-164) reduced the activity of the protein by 10% compared to the wild type virus (Simon et al, 1999), hence they are highly conserved (Figure 4.4). The HCCH (H108, C114, C133, H139) motif was also highly conserved in the studied samples and mutations that were found in this motif have shown to negatively affect *vif* function. The Cys residues in HCCH motif are suggested to be involved in vif function (Paul et al, 2006). This suggest that any mutation that occur in conserved motifs or residues may have negative effect on the virus. In sample ZA.07.MAB.21.VIF, W38G mutation was observed and suggestion is that since W38 together with other tryptophans are involved in recognition and suppression of APOBEC3G, the W38G mutation may have negative effect on *vif* functions. The same can be said for S95N mutation that was observed in sample ZA.07.LAU.39.VIF.

The R90K mutation in *vpr* gene leads to vpr protein losing its suppressive power to IL-12, which mediates enhancement of the cytotoxic activity of natural killer (NK) cells and CD8 cytotoxic T lymphocytes. But the opposite can be said with other mutations like R90G or R90S as they result in

suppression of IL-12 to comparable levels with the wild type or even lower levels (Tchereparova et al, 2009). Mutation in R90 residue can have either positive or negative effect on the virus, although it is most likely that the mutations in the R90 residue have a negative effect, hence the highly conserved nature of this residue (Figure 4.6). The R77Q mutation appeared in 15 of the samples and is associated with long term non-progression. Other mutations R77H/L/P were present and are thought to have to same impact as R77Q mutation. Mutations in HXRXG motifs at positions 71, 73 and 75 resulted in a loss of cell cycle arrest in G2 stage (Mueller and Lang, 2002). One sample had an R73K mutation that may result in loss of cell cycle arrest or even reduce the apoptotic effect induced by *vpr* protein, thus resulting in increased CD4 T lymphocytes.

Mutations of serine residues in *vpu* at position 58 and 62 reduce degradation of CD4 T lyphocytes and partially affect virion release (Opela et al, 2006). And are highly conserved in these samples (Figure 4.8), this is because mutation in those residues will reduce degradation of CD4 T lymphocytes and even impair virion release. Overall, conserved residues and motifs on the accessory genes play critical roles in viral pathogenesis. Mutagenesis studies showed that mutations in most the conserved residues or motifs have negative effect on viral pathogenesis, thus rendering the virus less pathogenic. Looking at the functions, mutagenesis studies that were conducted on accessory genes and the degree of conservation of important residues and motifs, accessory genes are certainly good candidates for HIV-1 intervention. Most of the conserved residues and motifs are also common in other clades. This suggests that if they are used for ARV's development they may treat for different subtypes and resistance mutations are not likely to occur as commonly mutations in accessory genes have negative effects on the virus. Indeed, HIV-1 accessory genes are potential new anti-HIV therapeutic targets as seen with *vif* gene (Chiu and Greene, 2008).

These conserved epitopes and motifs also make accessoy genes good candidates for vaccine development that can protect against most clades.

These studies demostrated that HIV accessory genes can elicit cytotoxic T lymphocyte response (CTL), that making the accesory genes more attractive for inclusion in HIV vaccine. HIV-1 subtype C immune response was evaluated across all genes. The CTL response was dominated by *nef* then *gag* (Masemola et al, 2004). The immunogenicity of *vif*, *vpu* and *nef* was studied in mice. *Vif*, *vpu* and *nef* were studied as fusion protein and the observation showed that *vif*, *vpu* and *nef* induced an effective T helper-1 and CTL responses. The *vif*, *vpu* and *nef* genes were from subtype B but they induced CTL response in clade D and E (Ayyavoo et al, 2000). CTL response of *vif*, *vpr* and *vpu* was also evaluated by Altfeld et al (2001) and Addo et al (2002) and they showed that CTL response against *vpr* was the greatest followed by *vif* then *vpu*. And CTL epitopes within *vif* and *vpr* were also identified. Another study evaluated CTL response of structural and accessory genes and the results showed that accessory genes have poor CTL response. *Vpr* like in most studies showed to have great CTL response than other accessory genes (Betts et al, 2002). Vpr CTL response showed to provide cross-protection against different clades (Wilson et al, 2003).

The challenge is whether accessory genes elicit protective immune response or whether such vaccines will mainly be therapeutic vaccines. To date, three vaccine trials (HVTN 064, HVTN 065 and HVTN 067) in the USA included either *vpu*, *vpr* or both, as vaccine antigens. Trial HVTN 064 has both *vpu* and *vpr* (IAVI, 2008). The outcome of these trials may strengthen the focus on accessory genes as alternative targets for drug and vaccine development.

CHAPTER SIX

6. CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS

6.1 Conclusions

A set of 25 HIV positive plasma samples were selected for genetic analysis of motifs associated with key biological activities of HIV-1 structural (*env*, *gag* and *pol*) and accessory (*vif*, *vpr* and *vpu*) genes. Genetic diversity of HIV-1 subtype C accessory (*vif*, *vpr* and *vpu*) and structural (*pol*, *gag* and *env*) genes were compared and evaluated using their phylogenetic relationship and sequence homologies.

Inter-sample sequence homology showed that *gag*, *vif* and *vpr* were more conserved. *Pol*, *env* and *vpu* were the lesser conserved genes of the six. This showed the divergent nature of the six genes (*vif*, *vpr*, *vpu*, *gag*, *env* and *pol*) with *vpu* being the highly diverse and *gag* being the lesser diverse. The less diverse functional motifs on these genes offer better targets for vaccine and ARV development. In accordence with the objectives of this study, the following conclusions could be reached:

1. In establishing the methods to amplify HIV-1 subtype C accessory genes (*vif*, *vpr* and *vpu*), the following conclusions could be reached:

- The PCR assay for accessory genes was more sensitive than the PCR assay for the *pol* gene. There was 100% (25/25) successful amplification by both accessory genes and *pol* gene PCR assays, but *pol* PCR assay had to be repeated on some of the samples which were all successfully amplified by accessory genes PCR assay first time.
- Sequencing of larger regions was more sensitive when using many primers for sequencing.

- Sequencing of cloned products was more sensitive than sequencing of direct PCR products.

2. In assessing motifs and residues associated with key biological functions of accessory genes and *pol* gene.

- *Vif* was more conserved followed by *vpr* then *vpu.*
- Functional motifs and residues observed in all accessory genes were highly conserved. *Vif* functional motifs and residues were highly conserved followed by *vpr* then *vpu*.
- Genetic analysis of *pol* gene revealed that there are drug resistant strains in circulation; 9/25 (36%) patients had resistance mutations to PI's and/or RT inhibitors.
- Most of the strains in our study had mutations associated with long term non-progressors; however, it is not known whether these patients were indeed long term non-progressors.

3. In establishing genetic diversity of accessory genes and assessing the genetic diversity of the accessory genes in relation to structural genes (*gag*, *pol* and *env*);

- The phylogenetic relationship showed that all the samples were subtype C.
- By sequence homology, *gag*, *vif* and *vpr* were more conserved than *pol*, *env* and *vpu*. The *vif* and *vpr* were the most attractive for HIV interventions due to their conserved nature.

Sequence analysis revealed conserved regions in all the six genes. Most of these conserved regions serve very important functions on the viral pathogenesis, thus making them good candidates for HIV interventions.

6.2 Limitations of the study

The study was an exploratory study and did not obtain representative samples from the studied locations. Thus regional founder effect could not be ascertained. There was lack of clinical data on the samples used; hence correlation of clinical outcomes could not be matched with mutations detected.

6.3 Recommendations

The findings of this study support the development of transfection studies to clearly understand the impact of mutation in accessory genes. It is further recommended to monitor the genetic evolutionary characteristics of accessory genes, as to assess the background genetic diversity of accessory genes of circulating strains before widespread introduction of HIV vaccine candidates containing accessory genes as its antigens in various clinical trials. Such information will be vital for Medunsa Clinical Research Unit (MeCRU) as is a new HIV vaccine clinical trial sites.

CHAPTER SEVEN

7. REFERENCES

Aboud S, Urassa W, Lyamuya E, Mhalu F and Biberfeld G. Evaluation of HIV antibody and antigen/antibody combination ELISAs for use in an alternative confirmation HIV testing strategy in Dar es Salaam, Tanzania. Journal of Virology Methods 2006; **135**: 192-196

Addo M M, Yu X G, Rosenberg E S, Walker B D and Altfeld M. Cytotoxic T lymphocyte (CTL) responses directed against regulatory and accessory proteins in HIV-1 infection. DNA and Cell Biology 2002; **21**: 671-678

AI-Harthi L, Spear G T, Hashemi F B, Landay A, Sha B E and Roebuck K A. A human immunodeficiency virus (HIV)-inducing factor from the female genital tract activates HIV-1 gene expression through the ĸB enhancer. Journal of Infectious Diseases 1998; **178**: 1343-1351

Altfeld M, Addo M M, Eldridge R L, Yu X G, Thomas S, Khatri A, Strick D, Phillips M N, Cohen G B, Islam S A, Kalams S A, Brander C, Goulder P J R, Rosenberg E S and Walker B D. Vpr is preferentially targeted by CTL during HIV-1 infection. Journal of Immunology 2001; **167**: 2743-2752

Andersen J L, Le Rouzic E and Plenelles V. HIV-1 vpr: Mechanism of G2 arrest and apoptosis. Journal of Experimental and Molecular Pathology 2008; **85**: 2-10

Auewarakal P, Wacharapornin P, Srichatrapimuk S, Chutipongtanate S and Puthavathana P. Uncoating of HIV requires cellular activation. Journal of Virology 2005; **337**: 93-101

Auvert B, Taljaard D, Ladarde E, Sobngwi-Tambekou J, Sitta R and Puren A. Randomized controlled intervention trial of male circumcision for reduction of HIV infection risk: The ANRS 1265 Trial. Journal of Medicine 2005; **2**: 1112-1122

Ayyavoo V, Kudchodkar S, Ramanathan M P, Le P, Muthumani K, Megalai N M, Dentchev T, Santiago-Barrios L, Mrinalini C and Weiner D B. Immunogenicity of a novel DNA vaccine cassette expressing multiple human immunodeficiency virus (HIV-1) accessory genes. AIDS 2000; **14**: 1-9

Bell C M, Connell B J, Capovilla A, Venter W D F, Stevens W S and Papathanasopoulos M A. Molecular characterization of the HIV Type 1 Subtype C Accessory Genes *vif*, *vpr*, and *vpu*. AIDS Research and Human Retroviruses 2007; **23**: 322-330

Benson R E, Sanfridson A, Ottinger J S, Doyle C and Cullen B R. Downregulation of cell-surface CD4 expression by simian immunodeficiency virus *nef* prevents viral super infection. Journal of Experimental Medicine 1993; **177**: 1561-1566

Bessong P O, Mphahlele J, Choge I A, Obi L C, Morris L, Hammarskjold M and Rekosk D M. Resistance mutation analysis of HIV type 1 subtype C among rural South Africa drug-naive patients prior to large-scale availability of antiretrovirals. AIDS Research and Human Retroviruses 2006; **22**: 1306-1312

Betts M R, Yusim K and Koup R A. Optimal antigens for HIV vaccines based on CD8 T response, protein length and sequence variability. DNA and Cell Biology 2002; **21**: 665-670

Bour S and Strabel K. The HIV-1 vpu protein: a multifunctional enhancer of viral particle release. Journal of Microbes and Infection 2003; **5**: 1029-1032

Carr M J, Coolen C, Davis A J, Burrell C J and Li P. Human Immunodeficiency Virus 1 (HIV-1) virion infectivity factor (vif) is part of reverse transcription complex and acts as an accessory factor for reverse transcription. Journal of Virology 2008; **372**: 147-150

Chan D C, Fass D, Berger J M and Kim P S. Core structure of gp41 from the HIV envelop glycoprotein. Journal of Cell 1997; **89**: 263-273

Chan Y, Yu S and Syu W. Organization of HIV-1 *pol* is critical for pol ployprotein processing. Journal of Biomedical Sciences 1999; **6**: 333-341

Chiu Y L and Greene W C. The APOBEC3 cytidine deaminases; an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annual Reviews Immunology 2008; **26**: 317-353

Choi S and Faller D V. A transcript from the long terminal repeats of a murine retrovirus associated with *trans* activation of cellular genes. Journal of Virology 1995; **69**: 7054-7060

Correa M and Gisselquist D. Route of HIV transmission in India: Assessing the reliability of information from AIDS case surveillance. Internation Journal of STD and AIDS 2006; **17**: 731-735

Dam E, Quercia R, Glass B, Descamps D, Launay O, Duval X, Krausslich H, Hance A J and Clavel F. Gag mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug experienced patients besides compensating for fitness loss. Journal of Pathogens 2009; **5**: 1-11

Dillon P J, Nelbock P, Perkins A and Rosen C A. Function of the human immunodeficiency virus types 1 and 2 rev proteins is dependent on their ability to interact with a structured region present in env gene mRNA. Journal of Virology 1990; **64**: 4428-4437

Dimitrov D S. How do viruses enter cells?: The HIV coreceptor teach us a lesson of complexity. Journal of Cell 1997; **91**: 721-730

Dorosko S M, Ayres S L and Connor R I. Induction of HIV-1 MPR649-684-specific IgA and IgG antibodies in caprine colostrums using a peptide-based vaccine. Journal of Vaccine 2008; **26**: 5416-5422

Ebberts-Reed D, Scarlata S and Carter C A. The major homology region of the HIV-1 gag precursor influences membrane affinity. Journal of Biochemistry 1996; **35**: 14268-14275

Estrabaud E, Le Rouzic E, Lopez-Verges S, Morel M, Belaidouni N, Benarous R, Transy C, Beelioz-Torrent C and Margottin-Goguet F. Regulated degradation of the HIV-1 vpu protein through a BTrCP-independent pathway limits the release of viral particles. Journal of Pathogens 2007; **3**: 995-1004

Evans B G, Noone A, Mortimer J Y, Gilbart V L, Gill O N, Nicoll A and Waight P A. Heterosexually acquired HIV-1 infection: Cases reported in England, Wales and Northern Ireland, 1985 to 1991. Communicable Disease Report 1992; **2**: 49-60

Freed E O. HIV-1 Replication. Journal of Somatic Cell and Molecular genetics 2001; **26**: 13-33

Friedman-Kien A E. Kaposi's Sarcoma: An opportunistic neoplasm. Journal of Investigative Dermatology 1984; **82**: 446-448

Garland S M, Ung L, Vujovic O V and Said J M. Cosmetic tattooing: A potential transmission route for HIV?. Clinical Scientific notes 2006; **46**: 456-462

Girard M P, Osmanov S K and Kieny M P. A review of vaccine research and development: The human immunodeficiency virus (HIV). Journal of Vaccine 2006; **24**: 4062-4081

Greene W C and Peterlin B M. Charting HIV's remarkable voyage through the cell: Besic science as a passport to future therapy. Nature Medicine Journal 2002; **8**: 673-680

Henderson L E, Bowers M A, Sowder R C, Serabyn S A, Johnson D G, Bess J W, Arthur L O, Bryant D K and Fenselau C. Gag proteins of the highly replicative MN strain of human immunodeficiency virus type-1: Posttranslational modifications, proteolytic processing and complete amino acid sequences. Journal of Virology 1992; **66**: 1856-1865

http://pathmicro.med.sc.edu/lecture/hiv_time_course2.jpg (Accessed 12 January 2009)

http://www.avert.org/treatment.htm (Accessed 2 April 2009)

http://www.avert.org/virus.htm (Accessed 04 March 2009)

http://www.avert.org/worldstats.htm (Accessed 27 May 2009)

http://www.txtwriter.com/Backgrounders/Aids/HIVLifecycle.gif (Accessed 12 January 2009)

http://www.yale.edu/bio243/HIV/genome.html (Accessed 12 May 2008)

Hunt R. Human Immunodeficiency virus and AIDS, Virology chapter seven part nine. http://pathmicro.med.sc.edu/lecture/hiv9.htm (Accessed on 12 May 2008)

Hunt R. Human Immunodeficiency virus and AIDS, Virology chapter seven part three. http://pathmicro.med.sc.edu/lecture/hiv3.htm (Accessed on 29 March 2009)

IAVI. Ongoing trials of preventive AIDS vaccines. www.iavireport.org/trialsdb (Accessed on 16 May 2008)

Jacobs G B, Laten A, Janse van Rensburg E, Bodem J, Weissbrich B, Rethwilm A, Preiser W and Engelbrecht S. Phylogenetic diversity and low level antiretroviral resistance mutations in HIV type 1 treatment–naive patients from Cape Town, South Africa. AIDS Research and Human Retroviruses 2008; **24**: 1009-1012

Johnson V A, Brun-Vezinet F, Clotet B, Gunthard H F, Kuritzkes D R, Pillay D, Schapiro J M and Richman D D. Update of the drug resistance mutations in HIV-1: December 2008. International AIDS Society 2008; **16**: 138-145

Johnston M J and Flores J. Progress in HIV vaccine development. Current Opinion in Pharmacology 2001; **1**: 504-510

Klaver B and Berkhout B. Comparison of 5' and 3' long terminal repeat promoter function in human immunodeficiency virus. Journal of Virology 1994; **68**: 3830-3840

Klimkait T, Strebel K, Hoggan M D, Martin M A and Orenstein J M . The human immunodeficiency virus type 1-specific protein Vpu is required for efficient virus maturation and release. Journal of Virology 1990; **64**: 621-629

Knight R. South Africa 2006: Population and HIV/AIDS. South Africa Delegation Briefing Paper 2006; 2

Komoto S, Tsuji S, Lee B, Iwabu Y, Kojima Y, Otake T, Taniguchi K and Ikuta K. Higher frequency of premature stop codon mutations at vpu gene of human immunodeficiency virus type 1 CRF01_AE compared with those of other subtypes. Journal of Microbes and Infection 2005; **7**: 139-147

Kuppuswamy M, Subramanian T, Srinivasan A and Chinnadurai G. Multiple functional domains of tat, the trans-activator of HIV-1, defined by mutational analysis. Molecular Biology 1989; **17**: 3551-3561

Lavigne M, Polomack L and Buc H. DNA synthesis by HIV-1 reverse transcriptase at the central termination site. Journal of Biological Chemistry 2001; **276**: 31429-31438

Le Rouzic E and Benichou S. The vpr protein from HIV-1: distinct roles along the viral life cycle. Journal of Retrovirology 2005; **2**: 1-2

Lee B, Sharron M, Montaner L J, Weissman D and Doms R W. Quantification of CD4, CCR5 and CXCR4 levels on lymphocyte subsets, dendritic cells and differentially conditioned monocyte-derived macrophages. Journal of Medical Sciences 1999; **96**: 5215-5220

Lu X, Yu H, Liu S, Brodsky F M and Peterlin B M. Interactions between HIV-1 nef and vacuolar ATPase facilitate the internalization of CD4. Journal of Immunity 1998; **8**: 647-656

Lum J J, Cohen O J, Nie Z, Weaver J G, Gomez T S,Yao X, Lynch D, Pilon A A, Hawley N, Kim J E, Chen Z, Montpetit M, Sanchez-Dardon J, Cohen E A and Badley A D. Vpr R77Q is associated with long-term nonprogressive HIV infection and impaired induction of apoptosis. Journal of Clinical Investigation 2003; **111**: 1547-1554

Mammano F, Kondo E, Sondroski J, Bukovsky A and Gottlinger H G. Rescue of human immunodeficiency virus type 1 matrix protein mutants by envelope

glycoprotein with short cytoplasmic domains. Journal of Virology 1995; **69**: 3824-3830

Masemola A M, Mashishi T N, Khoury G, Bredell H, Paximadis M, Mathebula T, Barkhan D, Puren A, Vardas E, Colvin M, Zijenah L, Katzenstein D, Musonda R, Allen S, Kumwenda N, Taha T, Gray G, Mclntyre J, Karim S A, Sheppard H W, Gray C M and HIVNET 028 study team. Novel and promiscuous CTL epitopes in conserved regions of gag targeted by individuals with early subtype C HIV type 1 infection from Southern Africa. Journal of Immunology 2004; **173**: 4607-4617

Masemola[a] A, Mashishi T, Kloury G, Mohube P,Mokgotho P, Vardas E, Colvin M, Zijenah L, Katzenstein D, Musonda R, Allen S, Kumwenda N, Taha T, Gray G, McIntyre J, Karim S A, Sheppard H W and Gray C M. Hierarchical targeting of subtype C human immunodeficiency virus type 1 proteins by CD8 T cell: Correlation with viral load. Journal of Virology 2004; **78**: 3233-3243

McCormick-Davis C, Dalton S B, Singh D K and Stephens E B. Comparison of vpu sequences from diverse geographical isolates of HIV type, identifies the presence of highly variable domains, additional invariant amino acids and a signature sequence motif common to subtype C isolates. AIDS Research and Human Retroviruses 2000; **16**: 1089-1090

Mehle A, Wilson H, Zhang C, Brazier A J, McPike M, Pery E and Gabuzda. Identification of an APOBEC3G binding site in human immunodeficiency virus type 1 vif and inhibitors of vif-APOBEC3G binding. Journal of Virology 2007; **81**: 13235-13241

Mueller S M and Lang S M. The first HXRXG motif in simian immunodeficiency virus mac239 vpr is crucial for G2/M cell cycle arrest. Journal of Virology 2002; **76**: 11704-11709

Musyoki A M. Genetic characterisation of circulating HIV strains in Pretoria and surrounding areas. MSc dissertation. University of Limpopo Medunsa Campus, 2009

Nathans R, Cao H, Sharova N, Ali A, Sharkey M, Stranska R, Stevenson M and Rana T M. Small molecule inhibition of HIV-1 vif. Nature Biotechnology 2008; **26**: 1187-1192

Nijhuis M, van Maarseveen N M, Lastere S, Schipper P, Coakley E, Glass B, Rovenska M, de Jong D, Chappey C, Goedegebuure I W, Heilek-Snyder G, Dulude D, Cammack N, Brakier-Gingras L, Konvalinka J, Parkin N, Krausslich H, Brun-Vezinet F and Boucher C A B. A novel substrate-Based HIV-1 protease inhibitor drug resistance mechanism. Journal of Medicine 2007; **4**: 152-163

Opella S J, Park S H, Lee S, Jones D, Nevzorov A, Mesleh M, Mrse A, Marassi F M, Oblatt-Montal M, Montal M and Strebel K. Structure and function of vpu from HIV-1. Viral Membrane Proteins 2006; **11**: 147-163

Papathanasopoulos M A; Hunt G M and Tiemessen C T. Evolution and Diversity of HIV-1 in Africa. Journal of Virus Genes 2003; **26**: 151-154

Paul I, Cui J and Maynard E L. Zinc binding to the HCCH motif of HIV-1 virion infectivity factor induces a conformational change that mediates protein-protein interactions. Biochemistry 2006; **103**: 18475-18480

Pereira L A, Bentley K, Peeters A, Churchill M J and Deacon N J. A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. Journal of Nucleic Acids Research 2000; **28**: 663-668

Perrin L; Kaiser L and Yerty S. Travel and the spread of HIV-1 genetic variants. The Lancet Infectious diseases journal 2003; **3**: 22-27

Ramirez B C, Simon-Loiere E, Galetto R and Negroni M. Implications of recombination for HIV diversity. Virus research journal 2008; **134**: 64-73

Rennie S, Muula A S and Westreich D. Male circumcision and HIV prevention; ethical, medical and public health tradeoffs in low-income countries. Journal of Medical Ethics 2007; **33**: 357-361

Richard J; Noel Jr and Anil K. SIV VPR evolution is inversely related to disease progression in a morphine-dependent rhesus macaque model of AIDS. Journal of Virology 2007; **359**: 397-404

Ryan K J and Ray C G; Medical Microbiology; 4th edition, 2004, 607-608

Sanger F, Nicklen S and Coulson A R. DNA sequencing with chain-termination inhibitors. Biochemistry 1977; **74**: 5463-5467

Schmitt K, Hill M S, Ruiz A, Culley N, Pinson D M, Wong S W and Stephens E D. Mutations in the highly conserved SLQYLA motif of Vif in a simian–human immunodeficiency virus result in a less pathogenic virus and are associated with G-to-A mutations in the viral genome. Journal of Virology 2009; **383**: 362-372

Schubert U and Strebel K. Differential activities of the human immunodeficiency virus type 1 encoded Vpu protein are regulated by phosphorylation and occur in different cellular compartments. Journal of Virology 1994; **68**: 2260-2271

Seibert S A, Howell C Y, Hughes M K and Hughes A L. Natural selection on the gag, pol and env genes of human immunodeficiency virus 1 (HIV-1). Molecular Biology and Evolution 1995; **12**: 803-813

Simmons G, Reeves J D, McKnight A, Dejucq N, Hibbitts S, Power C A, Aarons E, Schols D, de Clercq E, Proudfoot A E I and Clapham P R. CXCR4 as a functional coreceptor for human immunodeficiency virus type 1 infection of primary macrophages. Journal of Virology 1998; **72**: 8453-8457

Simon J H, Sheehy A M, Carpenter E A, Fouchier R A M and Malim M H. Mutational analysis of the Human Immunodecifiency Virus type 1 vif protein. Journal of Virology 1999; **73**: 2675-2681

Somasundaran M, Sharkey M, Brichacek B, Luzuriage K, Emerman M, Sullivan J L and Stevenson M. Evidence for cytopathogenicity determinant in HIV-1 vpr. Journal of Medical Sciences 2002; **99**: 9503-9508

Takeda M, Pekosz A, Shuck K, Pinto L H and Lamb R A. Influenza A virus M2 ion channel activity is essential for efficient replication in tissue culture. Journal of Virology 2002; **76**: 1391-1399

Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 2007; **24**: 1596-1599

Tcherepanova I, Starr A, Lackford B, Adams M D, Routy J, Boulassel M R, Calderhead D, Healey D and Nicolette C. The immunosuppressive properties of the HIV vpr protein are linked to a single highly conserved residue, R90. PlosONE 2009; **4**: 1-10

Tebit D M, Nankya I, Arts E J and Gao Y. HIV diversity, recombination and disease progression: How does fitness fit into the puzzle?. AIDS 2007; **9**: 75-87

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Research 1994; **22**: 4673-4680

Vallejo A, Gurtter L, Zekeng L and Hewlett I K. Nucleotide sequence analysis of the accessory genes of HIV-1 group O isolates. Virus Research 2003; **91**: 189-193

Vergne L, Peerters M, Mpoudi-Ngole E, Bourgeois A, Liegeos F, Toure-Kane C, Mboup S, Mulanga-Kabeya C, Saman E, Jourdan J, Reynes J and Delaporte E. Genetic diversity of protease and reverse transcriptase sequences in non-subtype B human immunodeficiency virus type-1 strains: Evidence of many minor drug resistance mutations in treatment-naive patients. Journal of Clinical Microbiology 2000; **38**: 3919-3925

Wang H, Sakurai A, Uchiyama T, Gu H, Adachi A and Mikako F. Unique characterization of HIV-1 vif expression. Journal of Microbes and Infection 2005; **7**: 385-390

Weng Y and Weiss C D. Mutational analysis of residues in the coiled-coil domain of human immunodeficiency virus type-1 transmembrane protein gp41. Journal of Virology 1998; **72**: 9676-9682

WHO. Antiretroviral therapy for HIV infection in adults and adolescents: Recommendations for a public health approach. 2006: 1-112

WHO. Guidelines for HIV diagnosis and monitoring of antiretroviral therapy. 2005: 1-75

WHO. HIV transmission through breastfeeding: A review of available evidence. 2004: 1-25

WHO. Priority interventions: HIV/AIDS prevention, treatment and care in the health sector. 2009; **1.2**: 1-126

Wilson C C, Mckinney D, Anders M, MaWhinney S, Forster J, Crimi C, Southwood S, Sette A, Chesnut R, Newman M J and Livingston B D. Development of a DNA vaccine designed to induce cytotoxic T lymphocyte responses to multiple conserved epitopes in HIV-1. Journal of Immunology 2003; **171**: 5611-5623

www.avert.org/aids-history-86.htm (Accessed 26 January 2009)

www.avert.org/media/images/subtypes.gif (Accessed 12 May 2008)

www.promega.com (Accessed April 2009)

Xiao Z, Ehrlich E, Luo K, Xiong Y and Yu X. Zinc chelaqtion inhibits HIV vif activity and liberates antiviral function of the cytidine deaminase APOBEC3G. The FASEB Journal 2007; **21**: 217-222

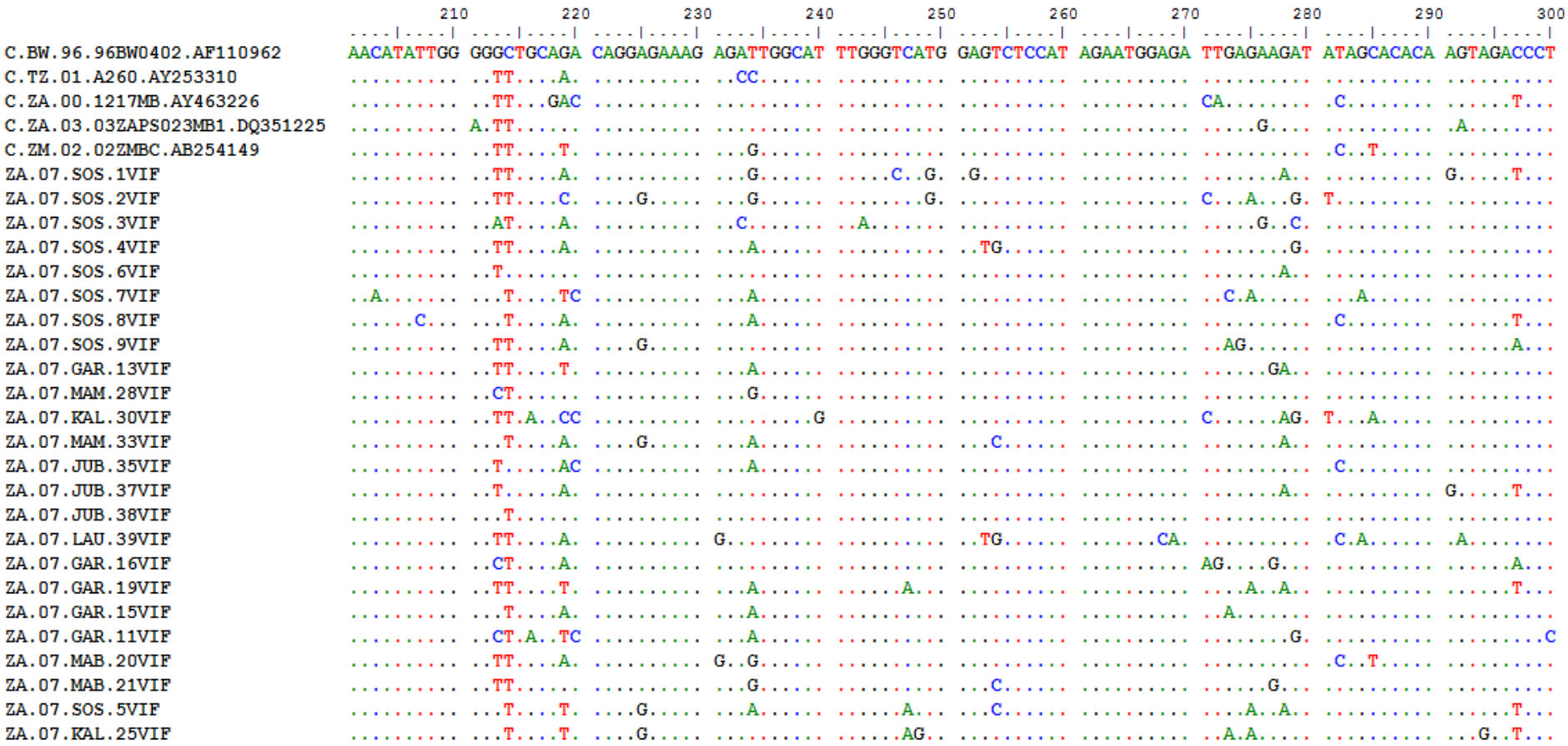
Xu H, Svarovskaia E S, Barr R, Zhang Y, Khan M A, Strebel K and Pathak V K. A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion. PNAS 2004; **101**: 5652-5657

CHAPTER EIGHT

8. APPENDICES

8.1 Accessory genes (*vif, vpr* and *vpu*) nucleotides sequences. (See subsequent pages).


110 120 130 140 150 160 170 180 190 200
C.BW.96.96BW0402.AF110962 AAGCTAATGG ATGGTTTTAC AGACATCATT ATGAAAGCAG ACATCCAAGA GTAAGTTCAG AAGTACACAT CCCATTAGGG GATGCTAGAT TAGTAATAAA
C.TZ.01.A260.AY253310
C.ZA.00.1217MB.AY463226
C.ZA.03.03ZAPS023MB1.DQ351225
C.ZM.02.02ZMBC.AB254149
ZA.07.SOS.1VIF
ZA.07.SOS.2VIF
ZA.07.SOS.3VIF
ZA.07.SOS.4VIF
ZA.07.SOS.6VIF
ZA.07.SOS.7VIF
ZA.07.SOS.8VIF
ZA.07.SOS.9VIF
ZA.07.GAR.13VIF
ZA.07.MAM.28VIF
ZA.07.KAL.30VIF
ZA.07.MAM.33VIF
ZA.07.JUB.35VIF
ZA.07.JUB.37VIF
ZA.07.JUB.38VIF
ZA.07.LAU.39VIF
ZA.07.GAR.16VIF
ZA.07.GAR.19VIF
ZA.07.GAR.15VIF
ZA.07.GAR.11VIF
ZA.07.MAB.20VIF
ZA.07.MAB.21VIF
ZA.07.SOS.5VIF
ZA.07.KAL.25VIF


210 220 230 240 250 260 270 280 290 300
C.BW.96.96BW0402.AF110962 AACATATTGG GGGCTGCAGA CAGGAGAAAG AGATTGGCAT TTGGGTCATG GAGTCTCCAT AGAATGGAGA TTGAGAAGAT ATAGCACACA AGTAGACCCT
C.TZ.01.A260.AY253310
C.ZA.00.1217MB.AY463226
C.ZA.03.03ZAPS023MB1.DQ351225
C.ZM.02.02ZMBC.AB254149
ZA.07.SOS.1VIF
ZA.07.SOS.2VIF
ZA.07.SOS.3VIF
ZA.07.SOS.4VIF
ZA.07.SOS.6VIF
ZA.07.SOS.7VIF
ZA.07.SOS.8VIF
ZA.07.SOS.9VIF
ZA.07.GAR.13VIF
ZA.07.MAM.28VIF
ZA.07.KAL.30VIF
ZA.07.MAM.33VIF
ZA.07.JUB.35VIF
ZA.07.JUB.37VIF
ZA.07.JUB.38VIF
ZA.07.LAU.39VIF
ZA.07.GAR.16VIF
ZA.07.GAR.19VIF
ZA.07.GAR.15VIF
ZA.07.GAR.11VIF
ZA.07.MAB.20VIF
ZA.07.MAB.21VIF
ZA.07.SOS.5VIF
ZA.07.KAL.25VIF


310 320 330 340 350 360 370 380 390 400
C.BW.96.96BW0402.AF110962 GGCCTGGCAG ACCAACTAAT TCATATGCAT TATTTTGATT GTTTTGCAGA CTCTGCCATA AGAAAAGCCA TCTTAGGACA GATAGTTAGC CCTAGGTGTG
C.TZ.01.A260.AY253310
C.ZA.00.1217MB.AY463226
C.ZA.03.03ZAPS023MB1.DQ351225
C.ZM.02.02ZMBC.AB254149
ZA.07.SOS.1VIF
ZA.07.SOS.2VIF
ZA.07.SOS.3VIF
ZA.07.SOS.4VIF
ZA.07.SOS.6VIF
ZA.07.SOS.7VIF
ZA.07.SOS.8VIF
ZA.07.SOS.9VIF
ZA.07.GAR.13VIF
ZA.07.MAM.28VIF
ZA.07.KAL.30VIF
ZA.07.MAM.33VIF
ZA.07.JUB.35VIF
ZA.07.JUB.37VIF
ZA.07.JUB.38VIF
ZA.07.LAU.39VIF
ZA.07.GAR.16VIF
ZA.07.GAR.19VIF
ZA.07.GAR.15VIF
ZA.07.GAR.11VIF
ZA.07.MAB.20VIF
ZA.07.MAB.21VIF
ZA.07.SOS.5VIF
ZA.07.KAL.25VIF


410 420 430 440 450 460 470 480 490 500
C.BW.96.96BW0402.AF110962 AATATCAAGC AGGACATAAC AAGGTAGGAT CTCTACAATA CTTGGCACTG ACAGCATTGA TAAAACCAAA AAAGAGAAAG CCACCTCTGC CTAGTGTTAG
C.TZ.01.A260.AY253310
C.ZA.00.1217MB.AY463226
C.ZA.03.03ZAPS023MB1.DQ351225
C.ZM.02.02ZMBC.AB254149
ZA.07.SOS.1VIF
ZA.07.SOS.2VIF
ZA.07.SOS.3VIF
ZA.07.SOS.4VIF
ZA.07.SOS.6VIF
ZA.07.SOS.7VIF
ZA.07.SOS.8VIF
ZA.07.SOS.9VIF
ZA.07.GAR.13VIF
ZA.07.MAM.28VIF
ZA.07.KAL.30VIF
ZA.07.MAM.33VIF
ZA.07.JUB.35VIF
ZA.07.JUB.37VIF
ZA.07.JUB.38VIF
ZA.07.LAU.39VIF
ZA.07.GAR.16VIF
ZA.07.GAR.19VIF
ZA.07.GAR.15VIF
ZA.07.GAR.11VIF
ZA.07.MAB.20VIF
ZA.07.MAB.21VIF
ZA.07.SOS.5VIF
ZA.07.KAL.25VIF

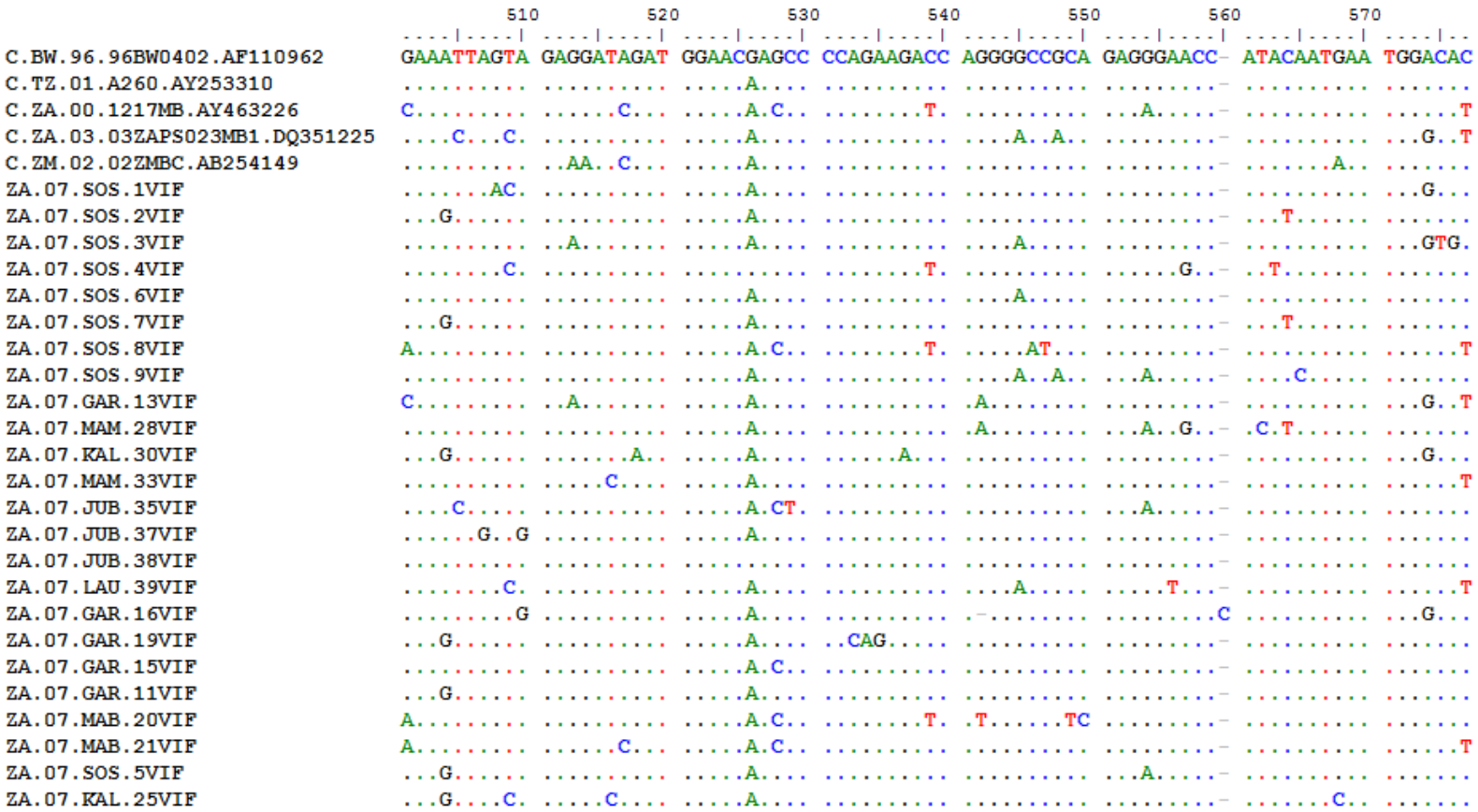
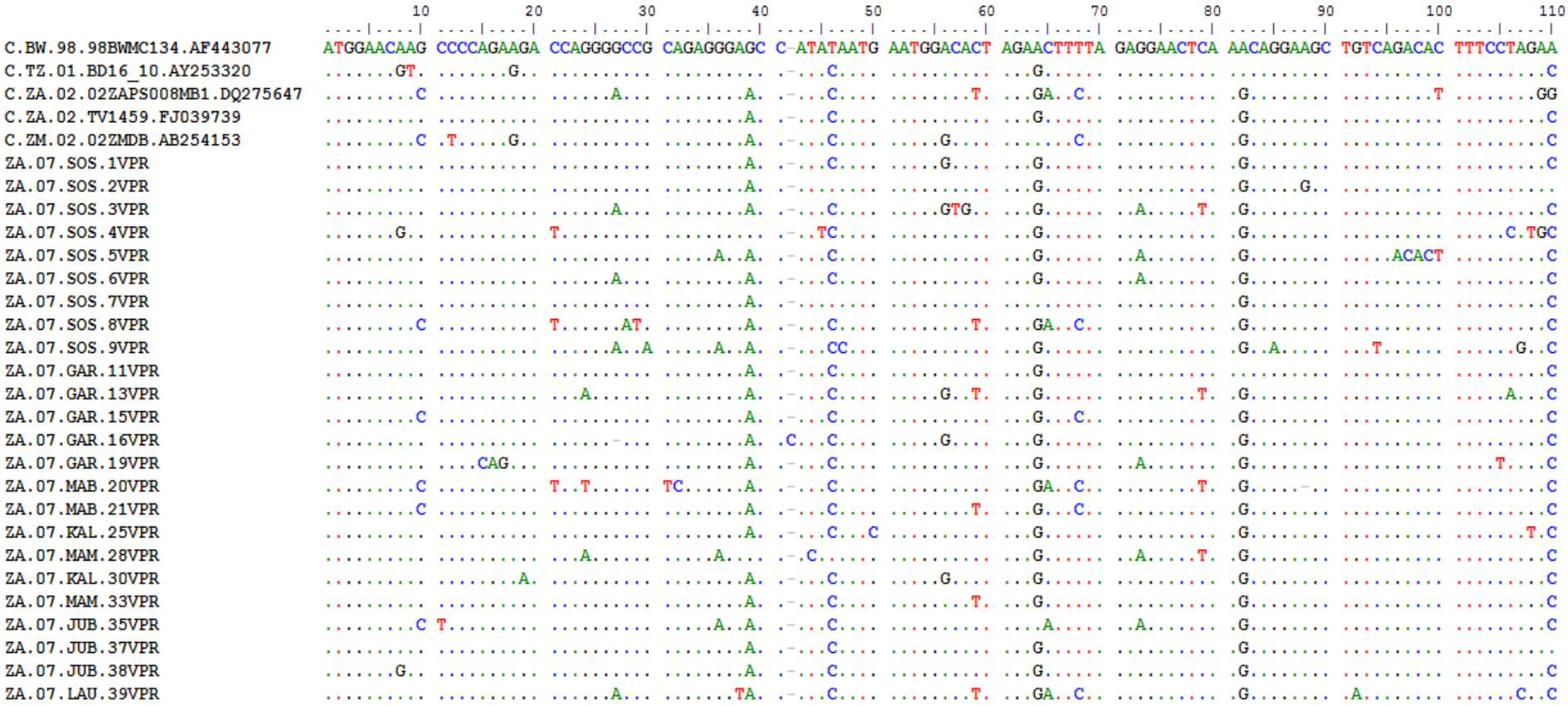


Figure 8.1: Nucleotide alignments of the *Vif* sequences from 24 HIV-1 subtype C isolates. The sequences are compared to reference strains from the Los Alamos Sequence Database.


10 20 30 40 50 60 70 80 90 100 110
C.BW.98.98BWMC134.AF443077 ATGGAACAAG CCCCAGAAGA CCAGGGGCCG CAGAGGGAGC C-ATATAATG AATGGACACT AGAACTTTTA GAGGAACTCA AACAGGAAGC TGTCAGACAC TTTCCTAGAA
C.TZ.01.BD16_10.AY253320 .....GT. .....G.. ....-...C.... ....G..... C
C.ZA.02.02ZAPS008MB1.DQ275647 .....C ....A... .....A. .-...C.... .....T. ...GA..C.. ..... .G..... .....T .....GG
C.ZA.02.TV1459.FJ039739 ....A. .-...C.... ....G..... ..... .G..... C
C.ZM.02.02ZMDB.AB254153 .....C .T.....G.. ....A. .-...C.... .....G.... .....C.. ..... .G..... C
ZA.07.SOS.1VPR ....A. .-...C.... .....G.... ...G..... ..... .G..... C
ZA.07.SOS.2VPR ....A. .-..... ....G..... ..... .G.....G.. ....
ZA.07.SOS.3VPR ....A... .....A. .-...C.... .....GTG.. ...G..... ..A.....T. .G..... C
ZA.07.SOS.4VPR .....G.. ..... T..... ..... .-..TC.... ....G..... ..... .G..... ....C.TGC
ZA.07.SOS.5VPR ....A..A. .-...C.... ....G..... ..A..... .G..... .....ACACT C
ZA.07.SOS.6VPR ....A... .....A. .-...C.... ....G..... ..A..... .G..... C
ZA.07.SOS.7VPR ....A. .-..... .... .G..... C
ZA.07.SOS.8VPR .....C ..... T.....AT. .....A. .-...C.... .....T. ...GA..C.. ..... .G..... C
ZA.07.SOS.9VPR ....A..A .....A..A. .-...CC.. ....G..... ..... .G..A..... ...T..... .....G..C
ZA.07.GAR.11VPR ....A. .-...C.... ....G..... C
ZA.07.GAR.13VPR ....A..... .....A. .-...C.... .....G..T. ...G..... .....T. .G..... ....A...C
ZA.07.GAR.15VPR .....C ....A. .-...C.... ....G...C.. ..... .G..... C
ZA.07.GAR.16VPR ....-... .....A. .C...C.... .....G.... ...G..... ..... .G..... C
ZA.07.GAR.19VPR ....CAG... ....A. .-...C.... ....G..... ..A..... .G..... ....T....C
ZA.07.MAB.20VPR .....C ..... T..T..... TC.....A. .-...C.... ....GA..C.. .....T. .G.....-.. C
ZA.07.MAB.21VPR .....C ....A. .-...C.... .....T. ...G...C.. ..... .G..... C
ZA.07.KAL.25VPR ....A. .-...C...C ....G..... ..... .G..... ....T.C
ZA.07.MAM.28VPR ....A..... .....A.... .-.C..... ....G..... ..A.....T. .G..... C
ZA.07.KAL.30VPR ....A. ....A. .-...C.... .....G.... ...G..... ..... .G..... C
ZA.07.MAM.33VPR ....A. .-...C.... .....T. ...G..... ..... .G..... C
ZA.07.JUB.35VPR .....C T..... ....A..A. .-...C.... ....A..... ..A..... .G..... C
ZA.07.JUB.37VPR ....A. .-...C.... ....G..... ..... .G..... ....
ZA.07.JUB.38VPR .....G.. ....A. .-...C.... ....G..... ..... .G..... C
ZA.07.LAU.39VPR ....A... .....TA. .-...C.... .....T. ...GA..C.. ..... .G..... .A..... .....C..C

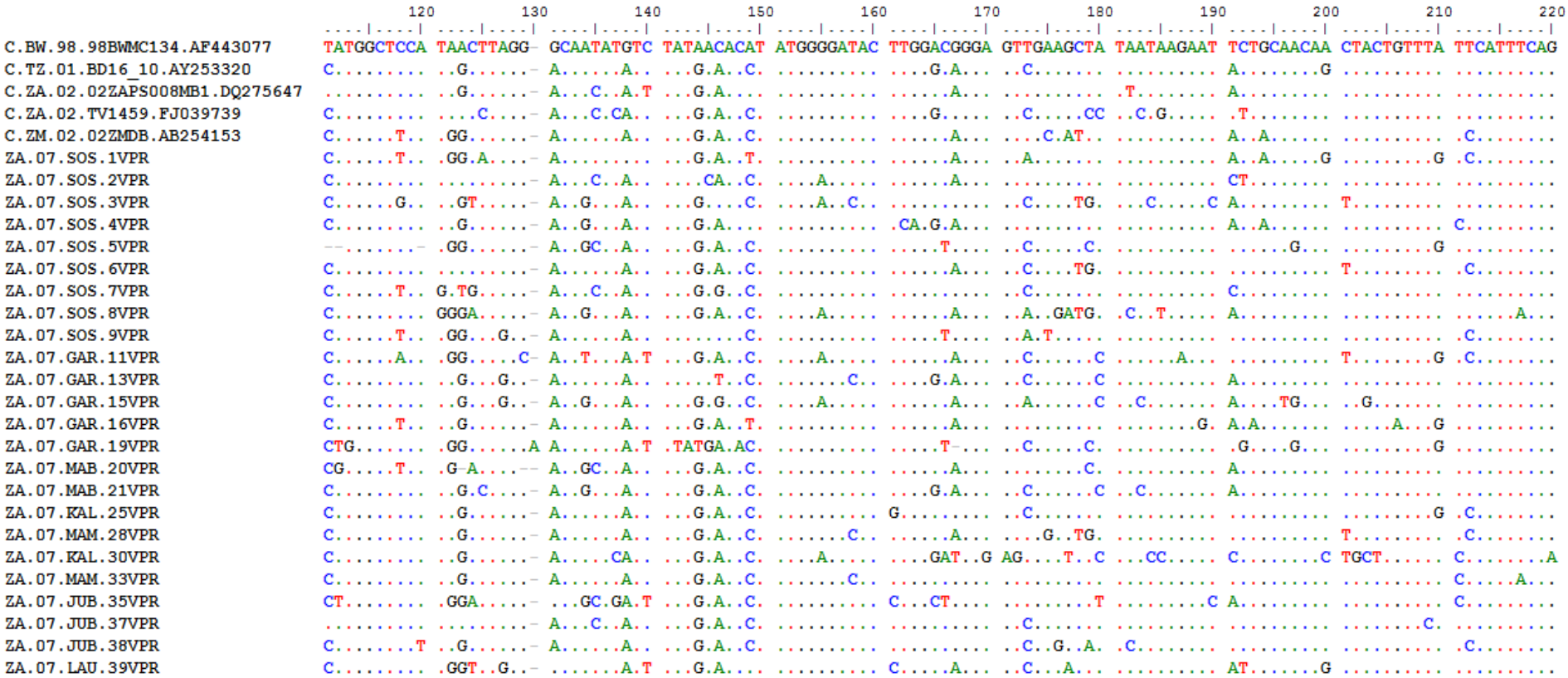
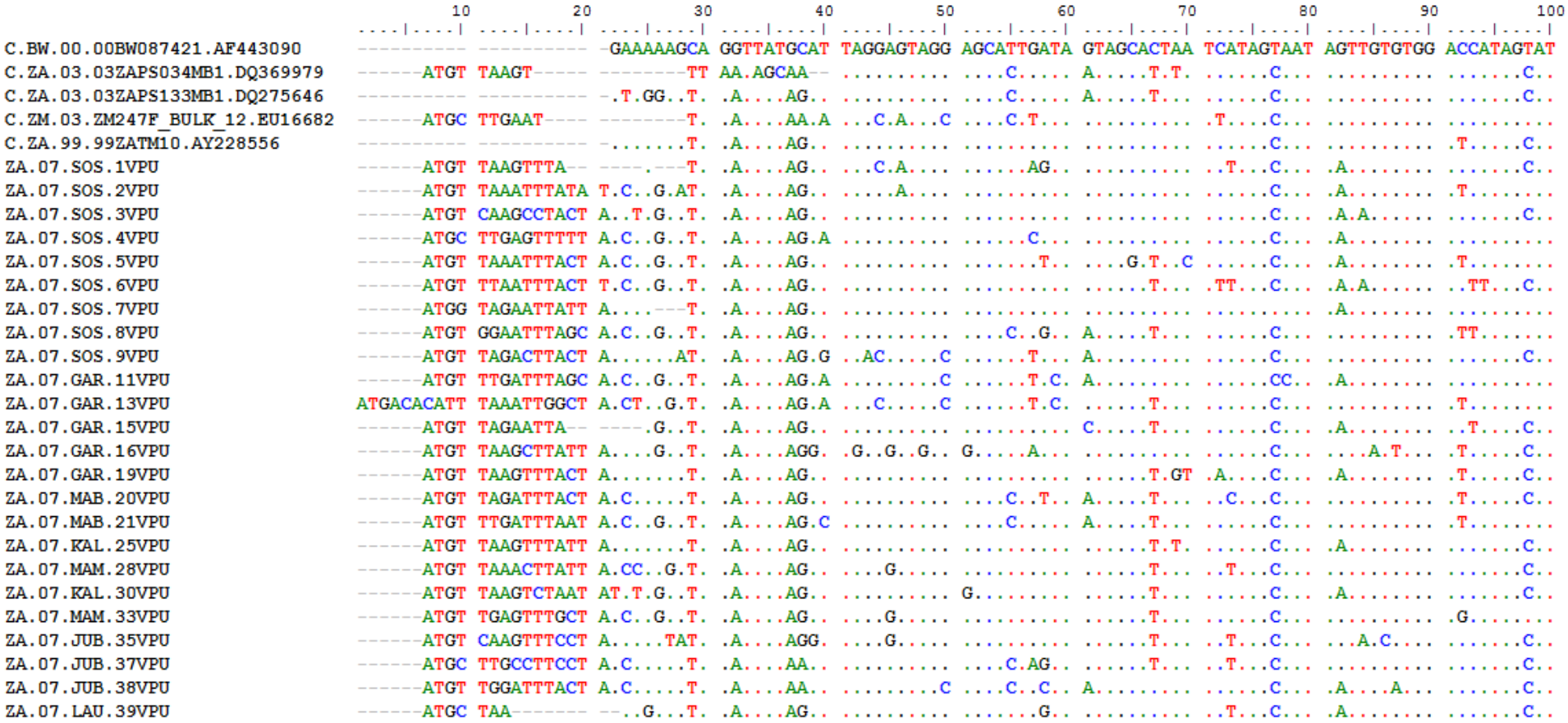
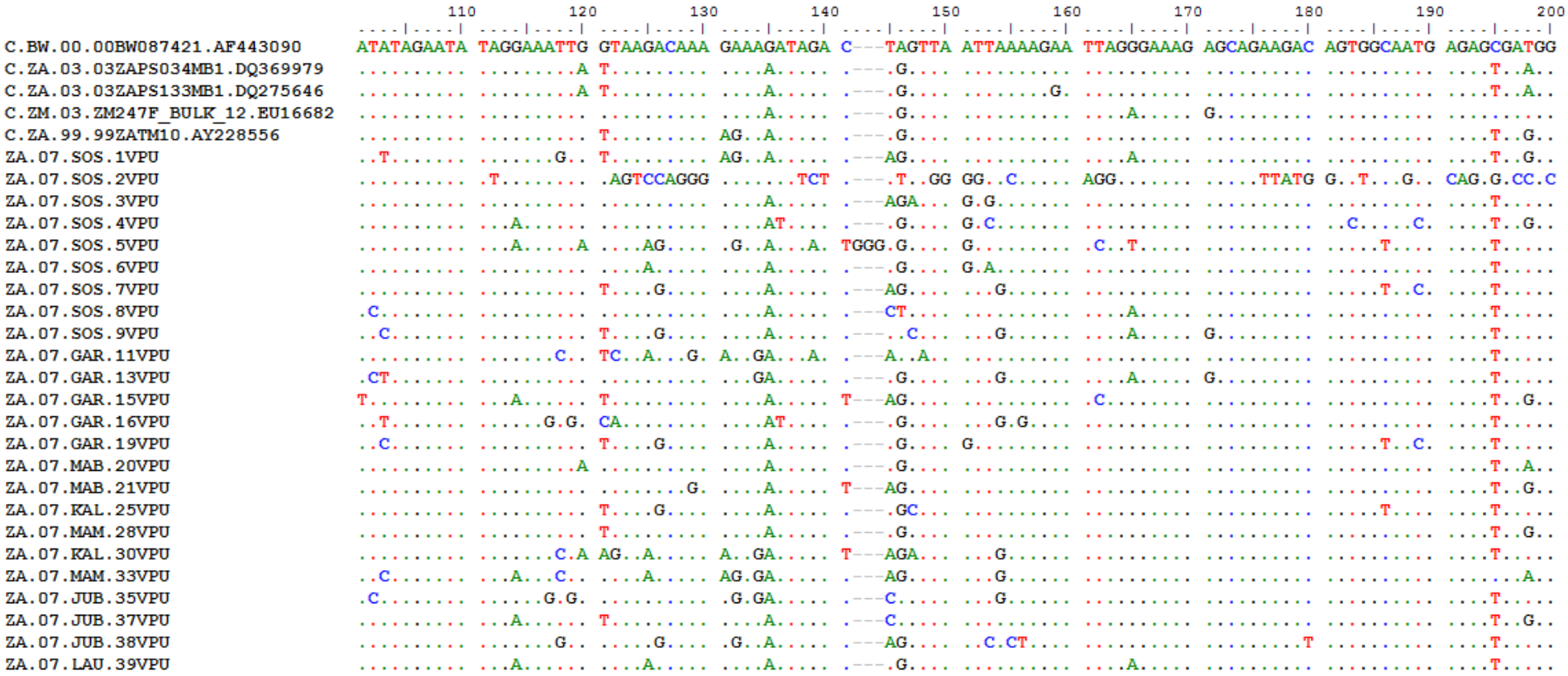

120 130 140 150 160 170 180 190 200 210 220
C.BW.98.98BWMC134.AF443077 TATGGCTCCA TAACTTAGG- GCAATATGTC TATAACACAT ATGGGGATAC TTGGACGGGA GTTGAAGCTA TAATAAGAAT TCTGCAACAA CTACTGTTTA TTCATTTCAG
C.TZ.01.BD16_10.AY253320 C......... ..G.....- A.....A.. ...G.A..C. ....G.A... ..C..... .... A.....G ....
C.ZA.02.02ZAPS008MB1.DQ275647 ....G.....- A...C..A.T ...G.A.... ....A... .... .T..... A..... ....
C.ZA.02.TV1459.FJ039739 C......... ....C....- A...C.CA.. ...G.A..C. ....G..... ..C.....CC ..C.G..... .T..... ....
C.ZM.02.02ZMDB.AB254153 C......T.. .GG.....- A.....A.. ...G.A..C. ....A... ....C.AT.. .... A..A..... .... .C.......
ZA.07.SOS.1VPR C......T.. .GG.A....- A......... ...G.A..T. ....A... ..A..... .... A..A.....G ....G .C.......
ZA.07.SOS.2VPR C......... ....- A...C..A.. ....CA..C. ....A..... .....A... .... CT..... ....
ZA.07.SOS.3VPR C.....G.. ..GT.....- A..G...A.. ...G....C. ....A..C.. ....C....TG. ...C.....C A..... T..... ....
ZA.07.SOS.4VPR C......... ..G.....- A..G...A.. ...G.A.... .... .CA.G.A... .... A..A..... .... C.......
ZA.07.SOS.5VPR --.....- .GG.....- A..GC..A.. ...G.A..C. ....T.... ..C.....C. ....G... ....G ....
ZA.07.SOS.6VPR C......... ....- A.....A.. ...G.A..C. ....A... ..C....TG. .... T..... .C.......
ZA.07.SOS.7VPR C.....T.. G.TG.....- A...C..A.. ...G.G..C. ....C..... .... C..... ....
ZA.07.SOS.8VPR C......... GGGA.....- A..G...A.. ...G.A..C. ....A..... .....A... ..A..GATG. .C..T..... A..... ....A...
ZA.07.SOS.9VPR C.....T.. .GG...G..- A.....A.. ....C. ....T.... ..A.T..... .... .C.......
ZA.07.GAR.11VPR C.....A.. .GG.....C- A..T...A.T ...G.A..C. ....A..... .....A... ..C.....C .....A... .... T.....G .C.......
ZA.07.GAR.13VPR C......... ..G...G..- A.....A.. .....T..C. ....C.. ....G.A... ..C.....C .... A..... ....
ZA.07.GAR.15VPR C......... ..G...G..- A..G...A.. ...G.G..C. ....A..... .....A... ..A.....C ..C..... A....TG... ..G..... ....
ZA.07.GAR.16VPR C.....T.. ..G.....- A.....A.. ...G.A..T. ....A... ....G. A.A..... .....A...G ....
ZA.07.GAR.19VPR CTG..... .GG.....A A.....A.T .TATGA.AC. ....T-... ..C.....C. .... .G...G... ....G ....
ZA.07.MAB.20VPR CG.....T.. .G-A....-- A..GC..A.. ...G.A..C. ....A... .....C. .... A..... ....
ZA.07.MAB.21VPR C......... ..G.C....- A..G...A.. ...G.A..C. ....G.A... ..C.....C ..C..... A..... ....
ZA.07.KAL.25VPR C......... ..G.....- A.....A.. ...G.A..C. .... G..... ..C..... ....G .C.......
ZA.07.MAM.28VPR C......... ..G.....- A.....A.. ...G.A..C. .....C.. .....A... ....G..TG. .... T..... .C.......
ZA.07.KAL.30VPR C......... ..G.....- A.....CA.. ...G.A..C. ....A..... ....GAT..G AG....T..C ...CC..... C.....C TGCT..... C.....A
ZA.07.MAM.33VPR C......... ..G.....- A.....A.. ...G.A..C. .....C.. .... C.....A...
ZA.07.JUB.35VPR CT..... .GGA.....- ...GC.GA.T ...G.A..C. .... C...CT.... .....T ....C A..... .... C.....
ZA.07.JUB.37VPR ....- A...C..A.. ...G.A..C. ....C..... ....C. .....
ZA.07.JUB.38VPR C.....T ..G.....- A.....A.. ...G.A..C. ....C..G..A. .C..... .... .C.....
ZA.07.LAU.39VPR C......... .GGT..G..- .....A.T ...G.A.... .... C....A... ..C...A... .... AT.....G ....

Figure 8.2: Graphic view nucleotide alignments of *Vpr* sequences from 24 HIV-1 subtype C isolates. The sequences are compared to reference strains from the Los Alamos Sequence Database.


10 20 30 40 50 60 70 80 90 100
C.BW.00.00BW087421.AF443090 ---------- ---------- -GAAAAAGCA GGTTATGCAT TAGGAGTAGG AGCATTGATA GTAGCACTAA TCATAGTAAT AGTTGTGTGG ACCATAGTAT
C.ZA.03.03ZAPS034MB1.DQ369979 ------ATGT TAAGT----- ---------TT AA.AGCAA-- ....C..... A.....T.T. ......C... ....C..
C.ZA.03.03ZAPS133MB1.DQ275646 ---------- ---------- -.T.GG..T. .A....AG.. ....C..... A.....T... ......C... ....C..
C.ZM.03.ZM247F_BULK_12.EU16682 ------ATGC TTGAAT---- ---------T. .A....AA.A ...C.A...C ....C.T... .... .T....C... ....
C.ZA.99.99ZATM10.AY228556 ---------- ---------- -.......T. .A....AG.. ....C... .... .T.....C..
ZA.07.SOS.1VPU ------ATGT TAAGTTTA-- -----.---T. .A....AG.. ...C.A.... ......AG.. ....T...C... .A........ ....C..
ZA.07.SOS.2VPU ------ATGT TAAATTTATA T.C..G.AT. .A....AG.. .....A.... ....C... .A........ .T........
ZA.07.SOS.3VPU ------ATGT CAAGCCTACT A..T.G..T. .A....AG.. ....C... .A.A...... ....C..
ZA.07.SOS.4VPU ------ATGC TTGAGTTTTT A.C..G..T. .A....AG.A ....C... ....C... .A........ ....
ZA.07.SOS.5VPU ------ATGT TAAATTTACT A.C..G..T. .A....AG.. ....T.. ....G.T..C ......C... .A........ .T........
ZA.07.SOS.6VPU ------ATGT TTAATTTACT T.C..G..T. .A....AG.. ....T... .TT...C... .A.A...... ..TT...C..
ZA.07.SOS.7VPU ------ATGG TAGAATTATT A.....--T. .A....AG.. .... .A........ ....
ZA.07.SOS.8VPU ------ATGT GGAATTTAGC A.C..G..T. .A....AG.. ....C..G.. A.....T... ......C... .... .TT.......
ZA.07.SOS.9VPU ------ATGT TAGACTTACT A......AT. .A....AG.G ..AC.....C ......T... A.......... ......C... ....C..
ZA.07.GAR.11VPU ------ATGT TTGATTTAGC A.C..G..T. .A....AG.A ....C ......T.C. A.......... ......CC.. .A........ ....
ZA.07.GAR.13VPU ATGACACATT TAAATTGGCT A.CT..G.T. .A....AG.A ...C.....C ......T.C. ......T... ......C... .... .T........
ZA.07.GAR.15VPU ------ATGT TAGAATTA-- -----.G..T. .A....AG.. .... C.....T... ......C... .A........ ..T....C..
ZA.07.GAR.16VPU ------ATGT TAAGCTTATT A....G..T. .A....AGG. .G..G..G.. G.....A... ....C... ....A.T... .T.....C..
ZA.07.GAR.19VPU ------ATGT TAAGTTTACT A.......T. .A....AG.. ....T.GT .A....C... .A........ .T.....C..
ZA.07.MAB.20VPU ------ATGT TAGATTTACT A.C.....T. .A....AG.. ....C..T.. A.....T... ..C...C... .... .T.....C..
ZA.07.MAB.21VPU ------ATGT TTGATTTAAT A.C..G..T. .A....AG.C ....C..... A.....T... ......C... .... .T........
ZA.07.KAL.25VPU ------ATGT TAAGTTTATT A.......T. .A....AG.. ....T.T. ......C... .A........ ....C..
ZA.07.MAM.28VPU ------ATGT TAAACTTATT A.CC..G.T. .A....AG.. ....G..... ....T... ..T...C... ....C..
ZA.07.KAL.30VPU ------ATGT TAAGTCTAAT AT.T.G..T. .A....AG.. .... G......... ......T... ......C... .A........ ....C..
ZA.07.MAM.33VPU ------ATGT TGAGTTTGCT A.C..G..T. .A....AG.. ....G..... ....T... ......C... .... .G........
ZA.07.JUB.35VPU ------ATGT CAAGTTTCCT A.....TAT. .A....AGG. ....G..... ....T... ..T...C... ...A.C.... ....C..
ZA.07.JUB.37VPU ------ATGC TTGCCTTCCT A.C.....T. .A....AA.. ....C.AG.. ......T... ..T...C... ....C..
ZA.07.JUB.38VPU ------ATGT TGGATTTACT A.C.....T. .A....AA.. ....C ....C..C.. A.......... ......C... .A....A... ....C..
ZA.07.LAU.39VPU ------ATGC TAA------- --..G...T. .A....AG.. ....G.. ....T...C... .A........ ....C..


110 120 130 140 150 160 170 180 190 200
C.BW.00.00BW087421.AF443090 ATATAGAATA TAGGAAATTG GTAAGACAAA GAAAGATAGA C---TAGTTA ATTAAAAGAA TTAGGGAAAG AGCAGAAGAC AGTGGCAATG AGAGCGATGG
C.ZA.03.03ZAPS034MB1.DQ369979
C.ZA.03.03ZAPS133MB1.DQ275646
C.ZM.03.ZM247F_BULK_12.EU16682
C.ZA.99.99ZATM10.AY228556
ZA.07.SOS.1VPU
ZA.07.SOS.2VPU
ZA.07.SOS.3VPU
ZA.07.SOS.4VPU
ZA.07.SOS.5VPU
ZA.07.SOS.6VPU
ZA.07.SOS.7VPU
ZA.07.SOS.8VPU
ZA.07.SOS.9VPU
ZA.07.GAR.11VPU
ZA.07.GAR.13VPU
ZA.07.GAR.15VPU
ZA.07.GAR.16VPU
ZA.07.GAR.19VPU
ZA.07.MAB.20VPU
ZA.07.MAB.21VPU
ZA.07.KAL.25VPU
ZA.07.MAM.28VPU
ZA.07.KAL.30VPU
ZA.07.MAM.33VPU
ZA.07.JUB.35VPU
ZA.07.JUB.37VPU
ZA.07.JUB.38VPU
ZA.07.LAU.39VPU

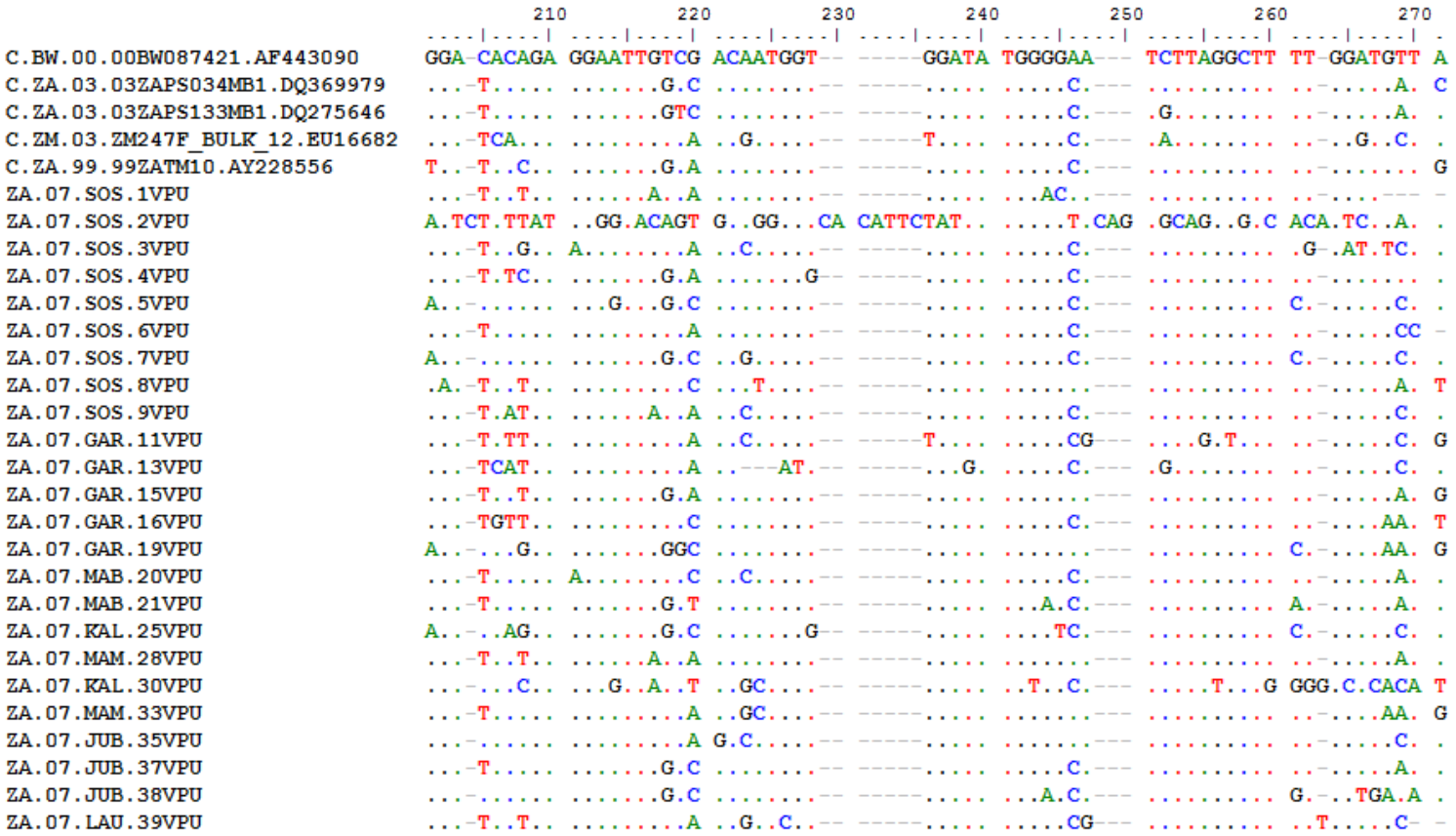


Figure 8.3: Graphic view nucleotide alignments of *Vpu* sequences from 24 HIV-1 subtype C isolates. The sequences are compared to reference strains from the Los Alamos Sequence Database.

8.2 Sequence homology for *vif*, *vpr* and *vpu*

Clone sequence homology

Table 8.1: Clones of sample ZA.07.SOS.2 (*vif*, *vpr* and *vpu*)

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ZA.07.SOS.2.1VIFc	574	2	ZA.07.SOS.2.2VIFc	576	98
1	ZA.07.SOS.2.1VIFc	574	3	ZA.07.SOS.2.3VIFc	576	98
1	ZA.07.SOS.2.1VIFc	574	4	ZA.07.SOS.2.4VIFc	576	98
1	ZA.07.SOS.2.1VIFc	574	5	ZA.07.SOS.2.5VIFc	576	92
2	ZA.07.SOS.2.2VIFc	576	3	ZA.07.SOS.2.3VIFc	576	98
2	ZA.07.SOS.2.2VIFc	576	4	ZA.07.SOS.2.4VIFc	576	98
2	ZA.07.SOS.2.2VIFc	576	5	ZA.07.SOS.2.5VIFc	576	91
3	ZA.07.SOS.2.3VIFc	576	4	ZA.07.SOS.2.4VIFc	576	99
3	ZA.07.SOS.2.3VIFc	576	5	ZA.07.SOS.2.5VIFc	576	92
4	ZA.07.SOS.2.4VIFc	576	5	ZA.07.SOS.2.5VIFc	576	92
1	ZA.07.SOS.2.1VPRc	288	2	ZA.07.SOS.2.2VPRc	288	99
1	ZA.07.SOS.2.1VPRc	288	3	ZA.07.SOS.2.3VPRc	288	99
1	ZA.07.SOS.2.1VPRc	288	4	ZA.07.SOS.2.4VPRc	288	99
1	ZA.07.SOS.2.1VPRc	288	5	ZA.07.SOS.2.5VPRc	282	88
2	ZA.07.SOS.2.2VPRc	288	3	ZA.07.SOS.2.3VPRc	288	98
2	ZA.07.SOS.2.2VPRc	288	4	ZA.07.SOS.2.4VPRc	288	98
2	ZA.07.SOS.2.2VPRc	288	5	ZA.07.SOS.2.5VPRc	282	87
3	ZA.07.SOS.2.3VPRc	288	4	ZA.07.SOS.2.4VPRc	288	99
3	ZA.07.SOS.2.3VPRc	288	5	ZA.07.SOS.2.5VPRc	282	88
4	ZA.07.SOS.2.4VPRc	288	5	ZA.07.SOS.2.5VPRc	282	88
1	ZA.07.SOS.2.1VPUc	274	2	ZA.07.SOS.2.2VPUc	256	96
1	ZA.07.SOS.2.1VPUc	274	3	ZA.07.SOS.2.3VPUc	276	97
1	ZA.07.SOS.2.1VPUc	274	4	ZA.07.SOS.2.4VPUc	275	98
1	ZA.07.SOS.2.1VPUc	274	5	ZA.07.SOS.2.5VPUc	261	85
2	ZA.07.SOS.2.2VPUc	256	3	ZA.07.SOS.2.3VPUc	276	96
2	ZA.07.SOS.2.2VPUc	256	4	ZA.07.SOS.2.4VPUc	275	95
2	ZA.07.SOS.2.2VPUc	256	5	ZA.07.SOS.2.5VPUc	261	89
3	ZA.07.SOS.2.3VPUc	276	4	ZA.07.SOS.2.4VPUc	275	100
3	ZA.07.SOS.2.3VPUc	276	5	ZA.07.SOS.2.5VPUc	261	84
4	ZA.07.SOS.2.4VPUc	275	5	ZA.07.SOS.2.5VPUc	261	83

Table 8.2: Clones of sample ZA.07.SOS.5 (*vif*, *vpr* and *vpu*)

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ZA.07.SOS.5.1VIFc	576	2	ZA.07.SOS.5.2VIFc	576	91
1	ZA.07.SOS.5.1VIFc	576	3	ZA.07.SOS.5.3VIFc	576	90
1	ZA.07.SOS.5.1VIFc	576	4	ZA.07.SOS.5.4VIFc	576	92
1	ZA.07.SOS.5.1VIFc	576	5	ZA.07.SOS.5.5VIFc	576	91
2	ZA.07.SOS.5.2VIFc	576	3	ZA.07.SOS.5.3VIFc	576	99
2	ZA.07.SOS.5.2VIFc	576	4	ZA.07.SOS.5.4VIFc	576	99
2	ZA.07.SOS.5.2VIFc	576	5	ZA.07.SOS.5.5VIFc	576	99
3	ZA.07.SOS.5.3VIFc	576	4	ZA.07.SOS.5.4VIFc	576	98
3	ZA.07.SOS.5.3VIFc	576	5	ZA.07.SOS.5.5VIFc	576	98
4	ZA.07.SOS.5.4VIFc	576	5	ZA.07.SOS.5.5VIFc	576	98
1	ZA.07.SOS.5.1VPRc	288	2	ZA.07.SOS.5.2VPRc	288	90
1	ZA.07.SOS.5.1VPRc	288	3	ZA.07.SOS.5.3VPRc	288	90
1	ZA.07.SOS.5.1VPRc	288	4	ZA.07.SOS.5.4VPRc	288	90
1	ZA.07.SOS.5.1VPRc	288	5	ZA.07.SOS.5.5VPRc	288	90
2	ZA.07.SOS.5.2VPRc	288	3	ZA.07.SOS.5.3VPRc	288	100
2	ZA.07.SOS.5.2VPRc	288	4	ZA.07.SOS.5.4VPRc	288	100
2	ZA.07.SOS.5.2VPRc	288	5	ZA.07.SOS.5.5VPRc	288	100
3	ZA.07.SOS.5.3VPRc	288	4	ZA.07.SOS.5.4VPRc	288	100
3	ZA.07.SOS.5.3VPRc	288	5	ZA.07.SOS.5.5VPRc	288	100
4	ZA.07.SOS.5.4VPRc	288	5	ZA.07.SOS.5.5VPRc	288	100
1	ZA.07.SOS.5.1VPUc	264	2	ZA.07.SOS.5.2VPUc	254	88
1	ZA.07.SOS.5.1VPUc	264	3	ZA.07.SOS.5.3VPUc	257	86
1	ZA.07.SOS.5.1VPUc	264	4	ZA.07.SOS.5.4VPUc	258	86
1	ZA.07.SOS.5.1VPUc	264	5	ZA.07.SOS.5.5VPUc	256	87
2	ZA.07.SOS.5.2VPUc	254	3	ZA.07.SOS.5.3VPUc	257	97
2	ZA.07.SOS.5.2VPUc	254	4	ZA.07.SOS.5.4VPUc	258	97
2	ZA.07.SOS.5.2VPUc	254	5	ZA.07.SOS.5.5VPUc	256	96
3	ZA.07.SOS.5.3VPUc	257	4	ZA.07.SOS.5.4VPUc	258	98
3	ZA.07.SOS.5.3VPUc	257	5	ZA.07.SOS.5.5VPUc	256	98
4	ZA.07.SOS.5.4VPUc	258	5	ZA.07.SOS.5.5VPUc	256	99

Table 8.3: Clones of sample ZA.07.GAR.13 (*vif*, *vpr* and *vpu*)

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ZA.07.GAR.13.1VIFc	576	2	ZA.07.GAR.13.2VIFc	576	98
1	ZA.07.GAR.13.1VIFc	576	3	ZA.07.GAR.13.3VIFc	576	98
1	ZA.07.GAR.13.1VIFc	576	4	ZA.07.GAR.13.4VIFc	576	97
1	ZA.07.GAR.13.1VIFc	576	5	ZA.07.GAR.13.5VIFc	574	97
2	ZA.07.GAR.13.2VIFc	576	3	ZA.07.GAR.13.3VIFc	576	100
2	ZA.07.GAR.13.2VIFc	576	4	ZA.07.GAR.13.4VIFc	576	98
2	ZA.07.GAR.13.2VIFc	576	5	ZA.07.GAR.13.5VIFc	574	98
3	ZA.07.GAR.13.3VIFc	576	4	ZA.07.GAR.13.4VIFc	576	98
3	ZA.07.GAR.13.3VIFc	576	5	ZA.07.GAR.13.5VIFc	574	98
4	ZA.07.GAR.13.4VIFc	576	5	ZA.07.GAR.13.5VIFc	574	97
1	ZA.07.GAR.13.1VPRc	288	2	ZA.07.GAR.13.2VPRc	288	98
1	ZA.07.GAR.13.1VPRc	288	3	ZA.07.GAR.13.3VPRc	288	98
1	ZA.07.GAR.13.1VPRc	288	4	ZA.07.GAR.13.4VPRc	288	97
1	ZA.07.GAR.13.1VPRc	288	5	ZA.07.GAR.13.5VPRc	288	97
2	ZA.07.GAR.13.2VPRc	288	3	ZA.07.GAR.13.3VPRc	288	100
2	ZA.07.GAR.13.2VPRc	288	4	ZA.07.GAR.13.4VPRc	288	97
2	ZA.07.GAR.13.2VPRc	288	5	ZA.07.GAR.13.5VPRc	288	97
3	ZA.07.GAR.13.3VPRc	288	4	ZA.07.GAR.13.4VPRc	288	97
3	ZA.07.GAR.13.3VPRc	288	5	ZA.07.GAR.13.5VPRc	288	97
4	ZA.07.GAR.13.4VPRc	288	5	ZA.07.GAR.13.5VPRc	288	95
1	ZA.07.GAR.13.1VPUc	260	2	ZA.07.GAR.13.2VPUc	265	96
1	ZA.07.GAR.13.1VPUc	260	3	ZA.07.GAR.13.3VPUc	252	98
1	ZA.07.GAR.13.1VPUc	260	4	ZA.07.GAR.13.4VPUc	262	98
1	ZA.07.GAR.13.1VPUc	260	5	ZA.07.GAR.13.5VPUc	254	98
2	ZA.07.GAR.13.2VPUc	265	3	ZA.07.GAR.13.3VPUc	252	97
2	ZA.07.GAR.13.2VPUc	265	4	ZA.07.GAR.13.4VPUc	262	96
2	ZA.07.GAR.13.2VPUc	265	5	ZA.07.GAR.13.5VPUc	254	98
3	ZA.07.GAR.13.3VPUc	252	4	ZA.07.GAR.13.4VPUc	262	98
3	ZA.07.GAR.13.3VPUc	252	5	ZA.07.GAR.13.5VPUc	254	98
4	ZA.07.GAR.13.4VPUc	262	5	ZA.07.GAR.13.5VPUc	254	98

Table 8.4: Clones of sample ZA.07.GAR.14 (*vif*, *vpr* and *vpu*)

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ZA.07.GAR.14.1VIFc	576	2	ZA.07.GAR.14.2VIFc	576	99
1	ZA.07.GAR.14.1VIFc	576	3	ZA.07.GAR.14.3VIFc	576	100
1	ZA.07.GAR.14.1VIFc	576	4	ZA.07.GAR.14.4VIFc	576	100
1	ZA.07.GAR.14.1VIFc	576	5	ZA.07.GAR.14.5VIFc	576	100
2	ZA.07.GAR.14.2VIFc	576	3	ZA.07.GAR.14.3VIFc	576	99
2	ZA.07.GAR.14.2VIFc	576	4	ZA.07.GAR.14.4VIFc	576	99
2	ZA.07.GAR.14.2VIFc	576	5	ZA.07.GAR.14.5VIFc	576	99
3	ZA.07.GAR.14.3VIFc	576	4	ZA.07.GAR.14.4VIFc	576	100
3	ZA.07.GAR.14.3VIFc	576	5	ZA.07.GAR.14.5VIFc	576	100
4	ZA.07.GAR.14.4VIFc	576	5	ZA.07.GAR.14.5VIFc	576	100
1	ZA.07.GAR.14.1VPRc	288	2	ZA.07.GAR.14.2VPRc	288	100
1	ZA.07.GAR.14.1VPRc	288	3	ZA.07.GAR.14.3VPRc	288	99
1	ZA.07.GAR.14.1VPRc	288	4	ZA.07.GAR.14.4VPRc	288	100
1	ZA.07.GAR.14.1VPRc	288	5	ZA.07.GAR.14.5VPRc	288	100
2	ZA.07.GAR.14.2VPRc	288	3	ZA.07.GAR.14.3VPRc	288	99
2	ZA.07.GAR.14.2VPRc	288	4	ZA.07.GAR.14.4VPRc	288	100
2	ZA.07.GAR.14.2VPRc	288	5	ZA.07.GAR.14.5VPRc	288	100
3	ZA.07.GAR.14.3VPRc	288	4	ZA.07.GAR.14.4VPRc	288	99
3	ZA.07.GAR.14.3VPRc	288	5	ZA.07.GAR.14.5VPRc	288	99
4	ZA.07.GAR.14.4VPRc	288	5	ZA.07.GAR.14.5VPRc	288	100
1	ZA.07.GAR.14.1VPUc	257	2	ZA.07.GAR.14.2VPUc	261	100
1	ZA.07.GAR.14.1VPUc	257	3	ZA.07.GAR.14.3VPUc	260	99
1	ZA.07.GAR.14.1VPUc	257	4	ZA.07.GAR.14.4VPUc	266	100
1	ZA.07.GAR.14.1VPUc	257	5	ZA.07.GAR.14.5VPUc	261	99
2	ZA.07.GAR.14.2VPUc	261	3	ZA.07.GAR.14.3VPUc	260	100
2	ZA.07.GAR.14.2VPUc	261	4	ZA.07.GAR.14.4VPUc	266	100
2	ZA.07.GAR.14.2VPUc	261	5	ZA.07.GAR.14.5VPUc	261	99
3	ZA.07.GAR.14.3VPUc	260	4	ZA.07.GAR.14.4VPUc	266	100
3	ZA.07.GAR.14.3VPUc	260	5	ZA.07.GAR.14.5VPUc	261	98
4	ZA.07.GAR.14.4VPUc	266	5	ZA.07.GAR.14.5VPUc	261	99

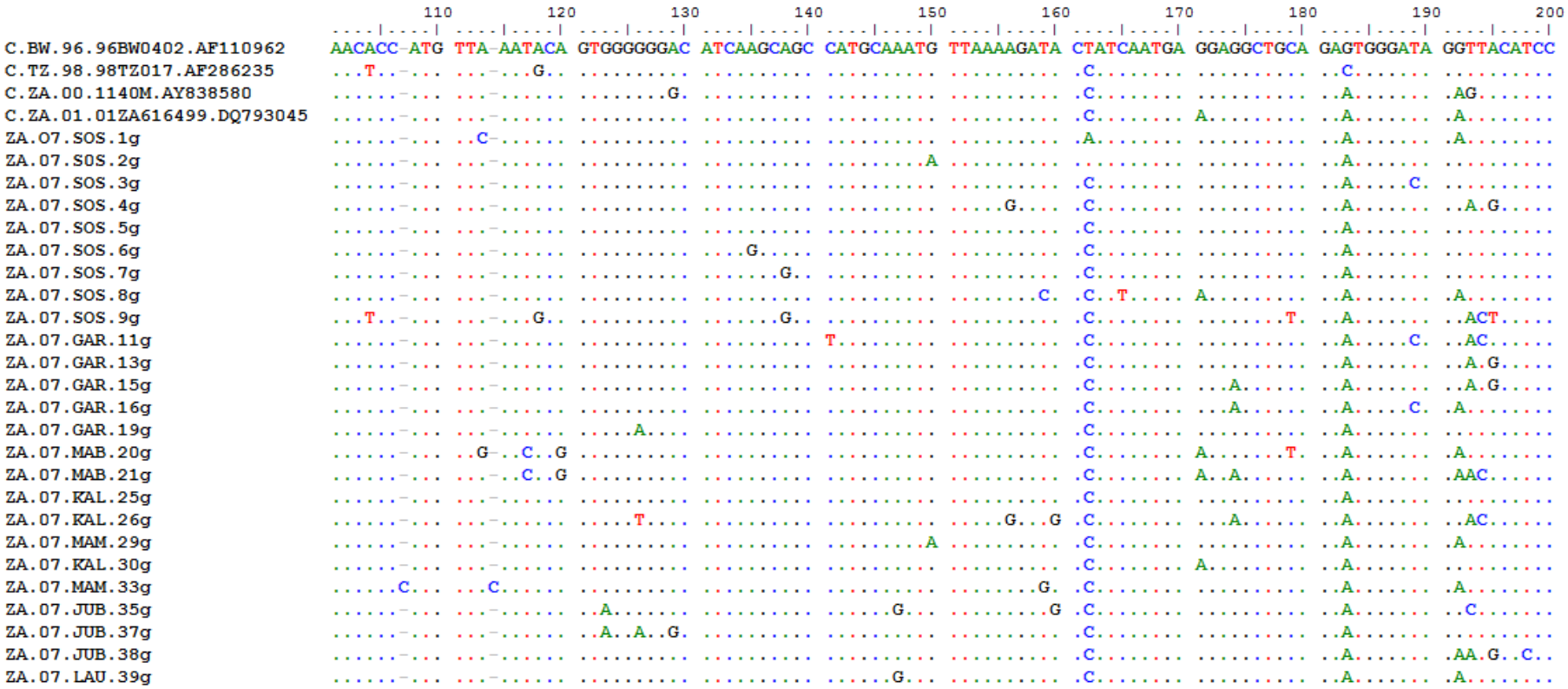
Table 8.5: Clones of sample ZA.07.KAL.26 (*vif*, *vpr* and *vpu*)

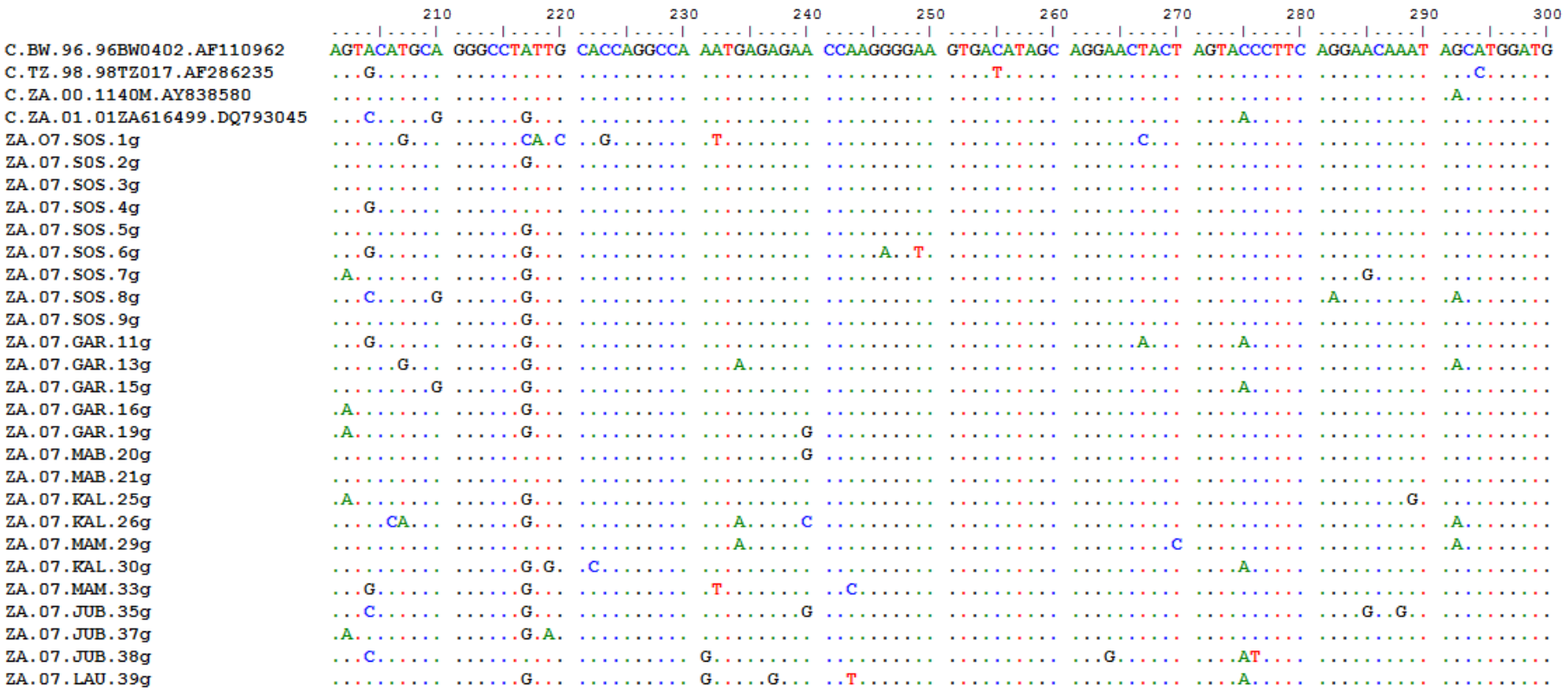
SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ZA.07.KAL.26.1VIFc	576	2	ZA.07.KAL.26.2VIFc	576	96
1	ZA.07.KAL.26.1VIFc	576	3	ZA.07.KAL.26.3VIFc	573	95
1	ZA.07.KAL.26.1VIFc	576	4	ZA.07.KAL.26.4VIFc	576	96
1	ZA.07.KAL.26.1VIFc	576	5	ZA.07.KAL.26.5VIFc	576	96
2	ZA.07.KAL.26.2VIFc	576	3	ZA.07.KAL.26.3VIFc	573	98
2	ZA.07.KAL.26.2VIFc	576	4	ZA.07.KAL.26.4VIFc	576	98
2	ZA.07.KAL.26.2VIFc	576	5	ZA.07.KAL.26.5VIFc	576	96
3	ZA.07.KAL.26.3VIFc	573	4	ZA.07.KAL.26.4VIFc	576	98
3	ZA.07.KAL.26.3VIFc	573	5	ZA.07.KAL.26.5VIFc	576	96
4	ZA.07.KAL.26.4VIFc	576	5	ZA.07.KAL.26.5VIFc	576	96
1	ZA.07.KAL.26.1VPRc	288	2	ZA.07.KAL.26.2VPRc	288	97
1	ZA.07.KAL.26.1VPRc	288	3	ZA.07.KAL.26.3VPRc	288	97
1	ZA.07.KAL.26.1VPRc	288	4	ZA.07.KAL.26.4VPRc	288	97
1	ZA.07.KAL.26.1VPRc	288	5	ZA.07.KAL.26.5VPRc	288	98
2	ZA.07.KAL.26.2VPRc	288	3	ZA.07.KAL.26.3VPRc	288	97
2	ZA.07.KAL.26.2VPRc	288	4	ZA.07.KAL.26.4VPRc	288	97
2	ZA.07.KAL.26.2VPRc	288	5	ZA.07.KAL.26.5VPRc	288	97
3	ZA.07.KAL.26.3VPRc	288	4	ZA.07.KAL.26.4VPRc	288	100
3	ZA.07.KAL.26.3VPRc	288	5	ZA.07.KAL.26.5VPRc	288	97
4	ZA.07.KAL.26.4VPRc	288	5	ZA.07.KAL.26.5VPRc	288	97
1	ZA.07.KAL.26.1VPUc	263	2	ZA.07.KAL.26.2VPUc	259	96
1	ZA.07.KAL.26.1VPUc	263	3	ZA.07.KAL.26.3VPUc	258	100
1	ZA.07.KAL.26.1VPUc	263	4	ZA.07.KAL.26.4VPUc	264	97
1	ZA.07.KAL.26.1VPUc	263	5	ZA.07.KAL.26.5VPUc	262	96
2	ZA.07.KAL.26.2VPUc	259	3	ZA.07.KAL.26.3VPUc	258	96
2	ZA.07.KAL.26.2VPUc	259	4	ZA.07.KAL.26.4VPUc	264	98
2	ZA.07.KAL.26.2VPUc	259	5	ZA.07.KAL.26.5VPUc	262	97
3	ZA.07.KAL.26.3VPUc	258	4	ZA.07.KAL.26.4VPUc	264	98
3	ZA.07.KAL.26.3VPUc	258	5	ZA.07.KAL.26.5VPUc	262	97
4	ZA.07.KAL.26.4VPUc	264	5	ZA.07.KAL.26.5VPUc	262	96

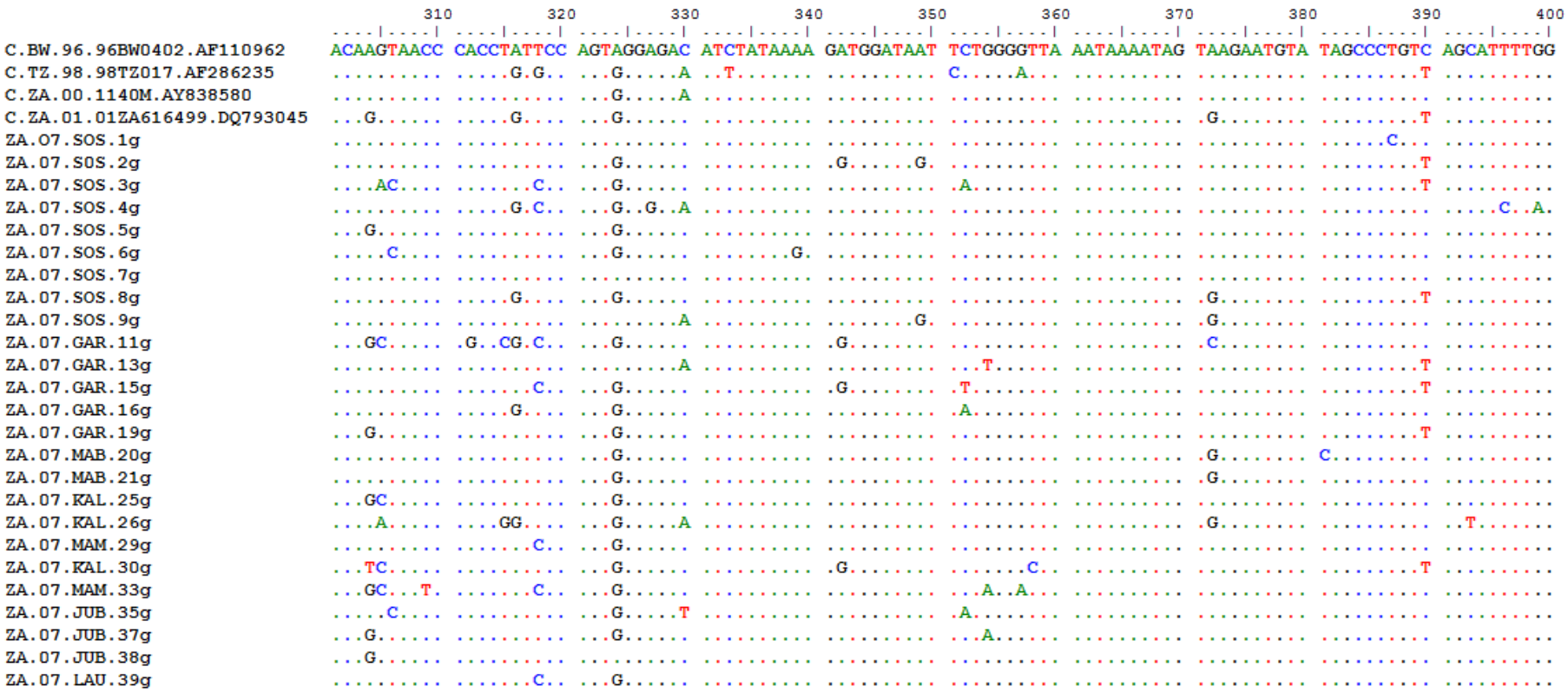
Table 8.6: Clones of sample ZA.07.MAM.29 (*vif*, *vpr* and *vpu*)

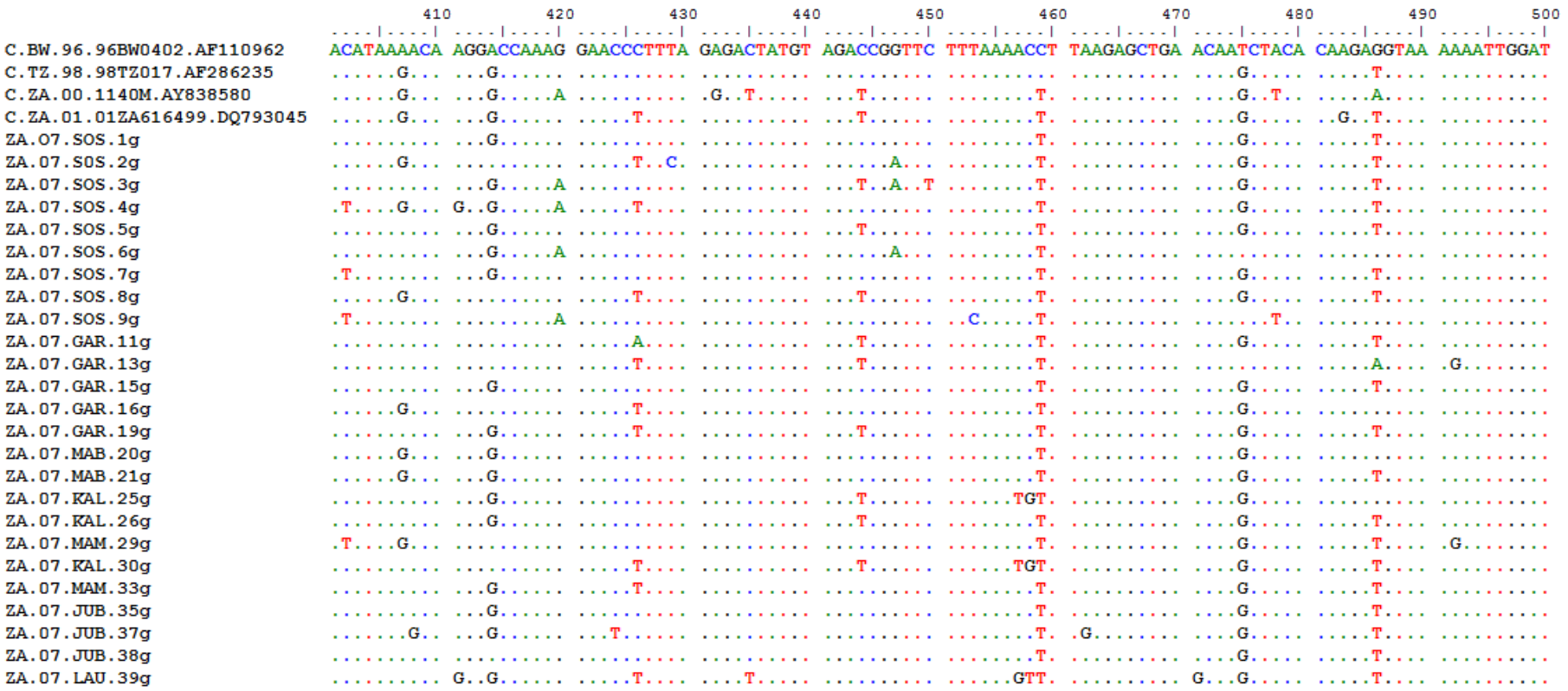
SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	ZA.07.MAM.29.1VIFc	576	2	ZA.07.MAM.29.2VIFc	576	97
1	ZA.07.MAM.29.1VIFc	576	3	ZA.07.MAM.29.3VIFc	576	97
1	ZA.07.MAM.29.1VIFc	576	4	ZA.07.MAM.29.4VIFc	576	98
1	ZA.07.MAM.29.1VIFc	576	5	ZA.07.MAM.29.5VIFc	575	97
2	ZA.07.MAM.29.2VIFc	576	3	ZA.07.MAM.29.3VIFc	576	98
2	ZA.07.MAM.29.2VIFc	576	4	ZA.07.MAM.29.4VIFc	576	98
2	ZA.07.MAM.29.2VIFc	576	5	ZA.07.MAM.29.5VIFc	575	98
3	ZA.07.MAM.29.3VIFc	576	4	ZA.07.MAM.29.4VIFc	576	98
3	ZA.07.MAM.29.3VIFc	576	5	ZA.07.MAM.29.5VIFc	575	97
4	ZA.07.MAM.29.4VIFc	576	5	ZA.07.MAM.29.5VIFc	575	98
1	ZA.07.MAM.29.1VPRc	288	2	ZA.07.MAM.29.2VPRc	288	99
1	ZA.07.MAM.29.1VPRc	288	3	ZA.07.MAM.29.3VPRc	288	99
1	ZA.07.MAM.29.1VPRc	288	4	ZA.07.MAM.29.4VPRc	288	100
1	ZA.07.MAM.29.1VPRc	288	5	ZA.07.MAM.29.5VPRc	287	98
2	ZA.07.MAM.29.2VPRc	288	3	ZA.07.MAM.29.3VPRc	288	99
2	ZA.07.MAM.29.2VPRc	288	4	ZA.07.MAM.29.4VPRc	288	99
2	ZA.07.MAM.29.2VPRc	288	5	ZA.07.MAM.29.5VPRc	287	99
3	ZA.07.MAM.29.3VPRc	288	4	ZA.07.MAM.29.4VPRc	288	99
3	ZA.07.MAM.29.3VPRc	288	5	ZA.07.MAM.29.5VPRc	287	98
4	ZA.07.MAM.29.4VPRc	288	5	ZA.07.MAM.29.5VPRc	287	98
1	ZA.07.MAM.29.1VPUc	253	2	ZA.07.MAM.29.2VPUc	256	95
1	ZA.07.MAM.29.1VPUc	253	3	ZA.07.MAM.29.3VPUc	253	98
1	ZA.07.MAM.29.1VPUc	253	4	ZA.07.MAM.29.4VPUc	252	97
1	ZA.07.MAM.29.1VPUc	253	5	ZA.07.MAM.29.5VPUc	250	96
2	ZA.07.MAM.29.2VPUc	256	3	ZA.07.MAM.29.3VPUc	253	93
2	ZA.07.MAM.29.2VPUc	256	4	ZA.07.MAM.29.4VPUc	252	93
2	ZA.07.MAM.29.2VPUc	256	5	ZA.07.MAM.29.5VPUc	250	98
3	ZA.07.MAM.29.3VPUc	253	4	ZA.07.MAM.29.4VPUc	252	98
3	ZA.07.MAM.29.3VPUc	253	5	ZA.07.MAM.29.5VPUc	250	96
4	ZA.07.MAM.29.4VPUc	252	5	ZA.07.MAM.29.5VPUc	250	95

8.3 *Gag* and *env* nucleotide sequences. (See subsequent pages).

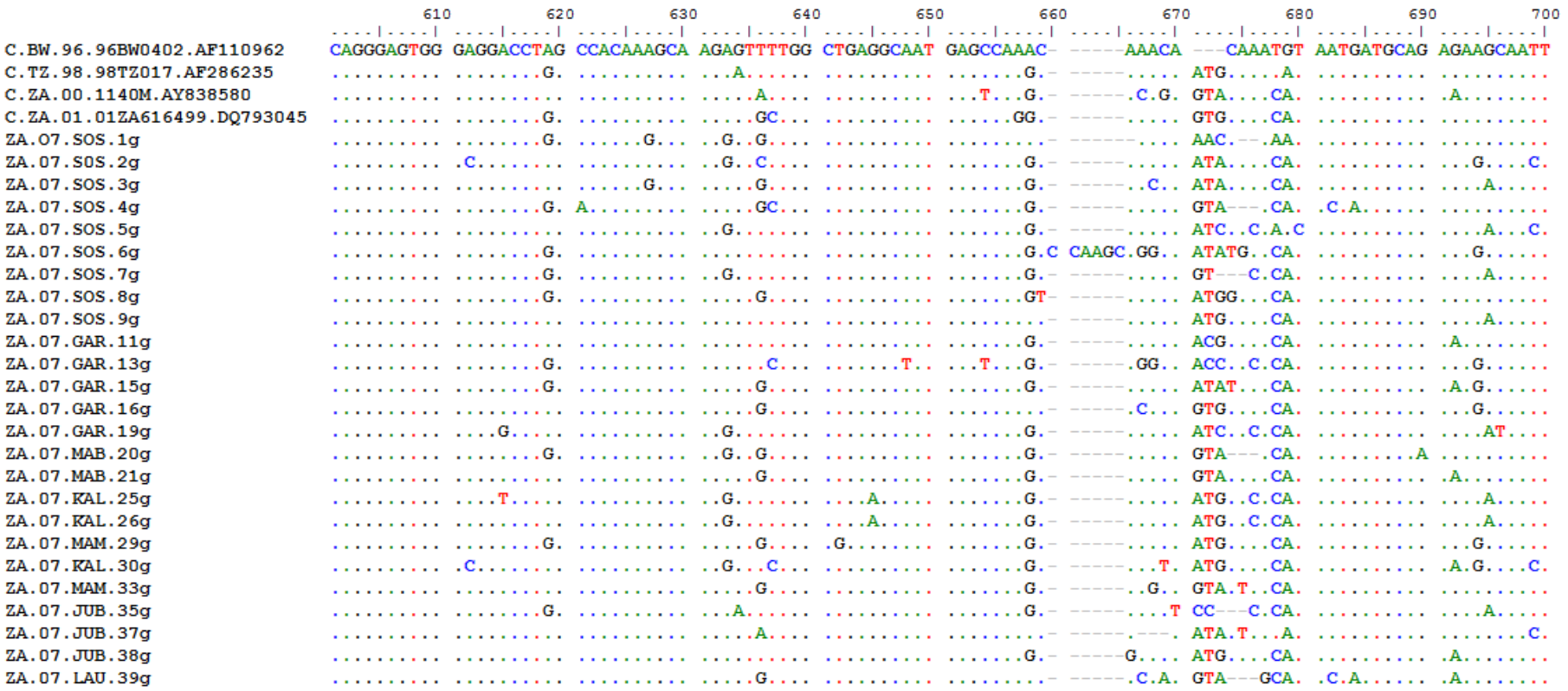


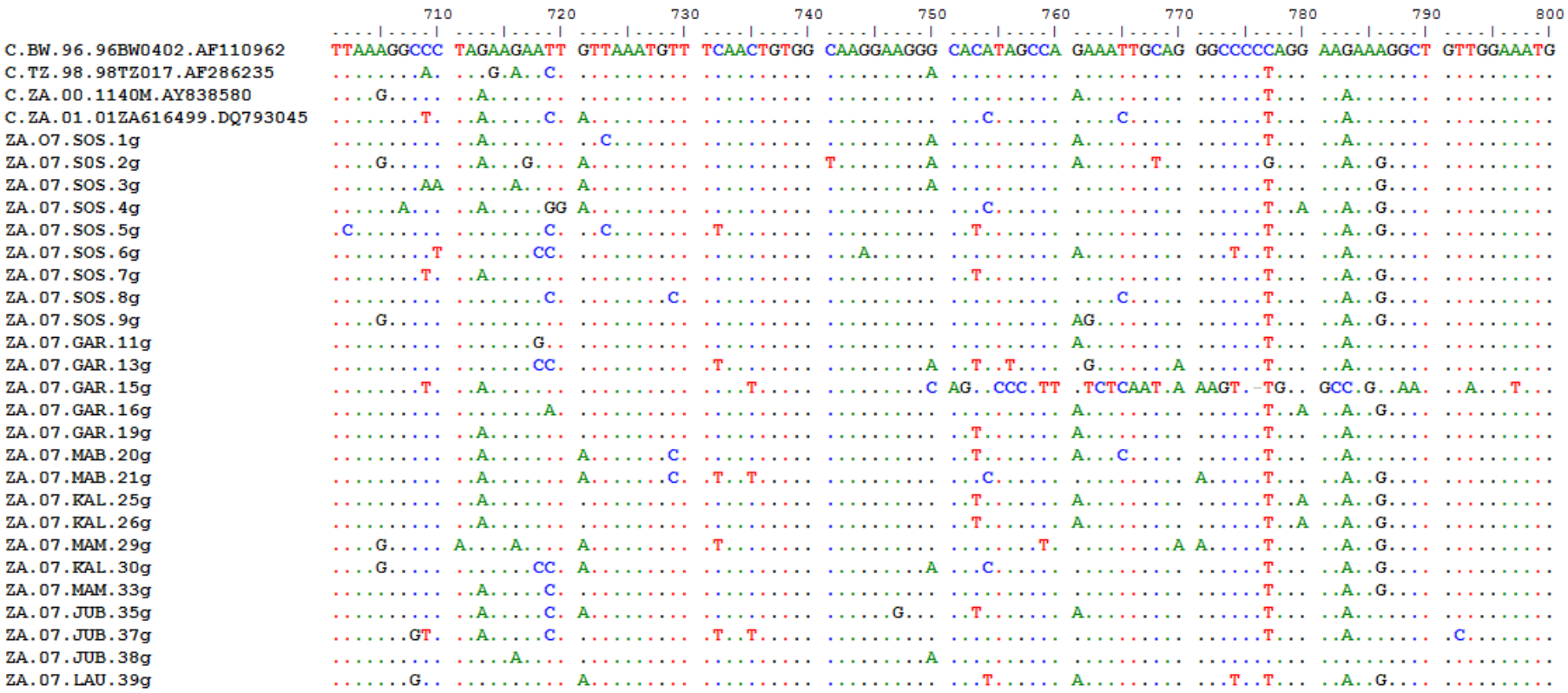

210 220 230 240 250 260 270 280 290 300
C.BW.96.96BW0402.AF110962 AGTACATGCA GGGCCTATTG CACCAGGCCA AATGAGAGAA CCAAGGGGAA GTGACATAGC AGGAACTACT AGTACCCTTC AGGAACAAAT AGCATGGATG
C.TZ.98.98TZ017.AF286235
C.ZA.00.1140M.AY838580
C.ZA.01.01ZA616499.DQ793045
ZA.07.SOS.1g
ZA.07.SOS.2g
ZA.07.SOS.3g
ZA.07.SOS.4g
ZA.07.SOS.5g
ZA.07.SOS.6g
ZA.07.SOS.7g
ZA.07.SOS.8g
ZA.07.SOS.9g
ZA.07.GAR.11g
ZA.07.GAR.13g
ZA.07.GAR.15g
ZA.07.GAR.16g
ZA.07.GAR.19g
ZA.07.MAB.20g
ZA.07.MAB.21g
ZA.07.KAL.25g
ZA.07.KAL.26g
ZA.07.MAM.29g
ZA.07.KAL.30g
ZA.07.MAM.33g
ZA.07.JUB.35g
ZA.07.JUB.37g
ZA.07.JUB.38g
ZA.07.LAU.39g

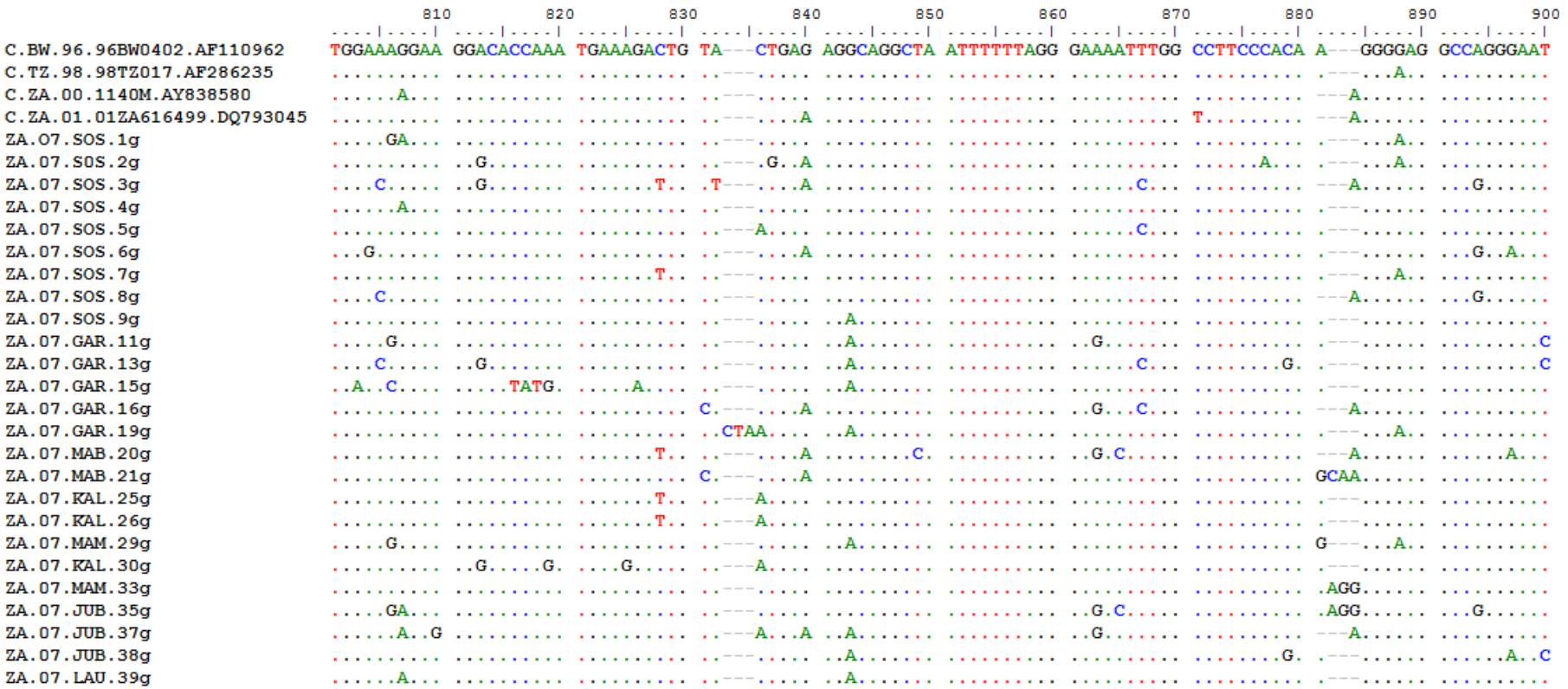

310 320 330 340 350 360 370 380 390 400
C.BW.96.96BW0402.AF110962 ACAAGTAACC CACCTATTCC AGTAGGAGAC ATCTATAAAA GATGGATAAT TCTGGGGTTA AATAAAATAG TAAGAATGTA TAGCCCTGTC AGCATTTTGG
C.TZ.98.98TZ017.AF286235
C.ZA.00.1140M.AY838580
C.ZA.01.01ZA616499.DQ793045
ZA.07.SOS.1g
ZA.07.SOS.2g
ZA.07.SOS.3g
ZA.07.SOS.4g
ZA.07.SOS.5g
ZA.07.SOS.6g
ZA.07.SOS.7g
ZA.07.SOS.8g
ZA.07.SOS.9g
ZA.07.GAR.11g
ZA.07.GAR.13g
ZA.07.GAR.15g
ZA.07.GAR.16g
ZA.07.GAR.19g
ZA.07.MAB.20g
ZA.07.MAB.21g
ZA.07.KAL.25g
ZA.07.KAL.26g
ZA.07.MAM.29g
ZA.07.KAL.30g
ZA.07.MAM.33g
ZA.07.JUB.35g
ZA.07.JUB.37g
ZA.07.JUB.38g
ZA.07.LAU.39g

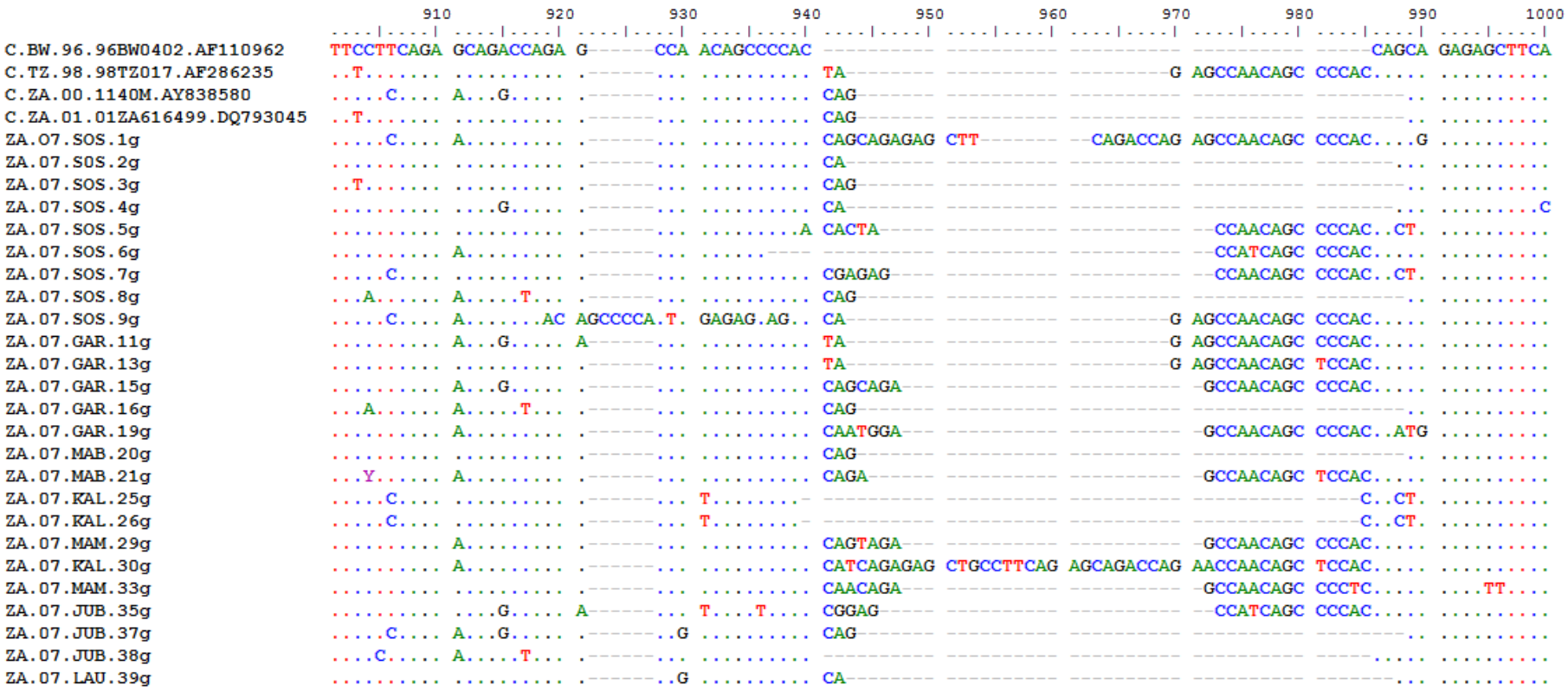

410 420 430 440 450 460 470 480 490 500
C.BW.96.96BW0402.AF110962 ACATAAAACA AGGACCAAAG GAACCCTTTA GAGACTATGT AGACCGGTTC TTTAAAACCT TAAGAGCTGA ACAATCTACA CAAGAGGTAA AAAATTGGAT
C.TZ.98.98TZ017.AF286235
C.ZA.00.1140M.AY838580
C.ZA.01.01ZA616499.DQ793045
ZA.07.SOS.1g
ZA.07.SOS.2g
ZA.07.SOS.3g
ZA.07.SOS.4g
ZA.07.SOS.5g
ZA.07.SOS.6g
ZA.07.SOS.7g
ZA.07.SOS.8g
ZA.07.SOS.9g
ZA.07.GAR.11g
ZA.07.GAR.13g
ZA.07.GAR.15g
ZA.07.GAR.16g
ZA.07.GAR.19g
ZA.07.MAB.20g
ZA.07.MAB.21g
ZA.07.KAL.25g
ZA.07.KAL.26g
ZA.07.MAM.29g
ZA.07.KAL.30g
ZA.07.MAM.33g
ZA.07.JUB.35g
ZA.07.JUB.37g
ZA.07.JUB.38g
ZA.07.LAU.39g

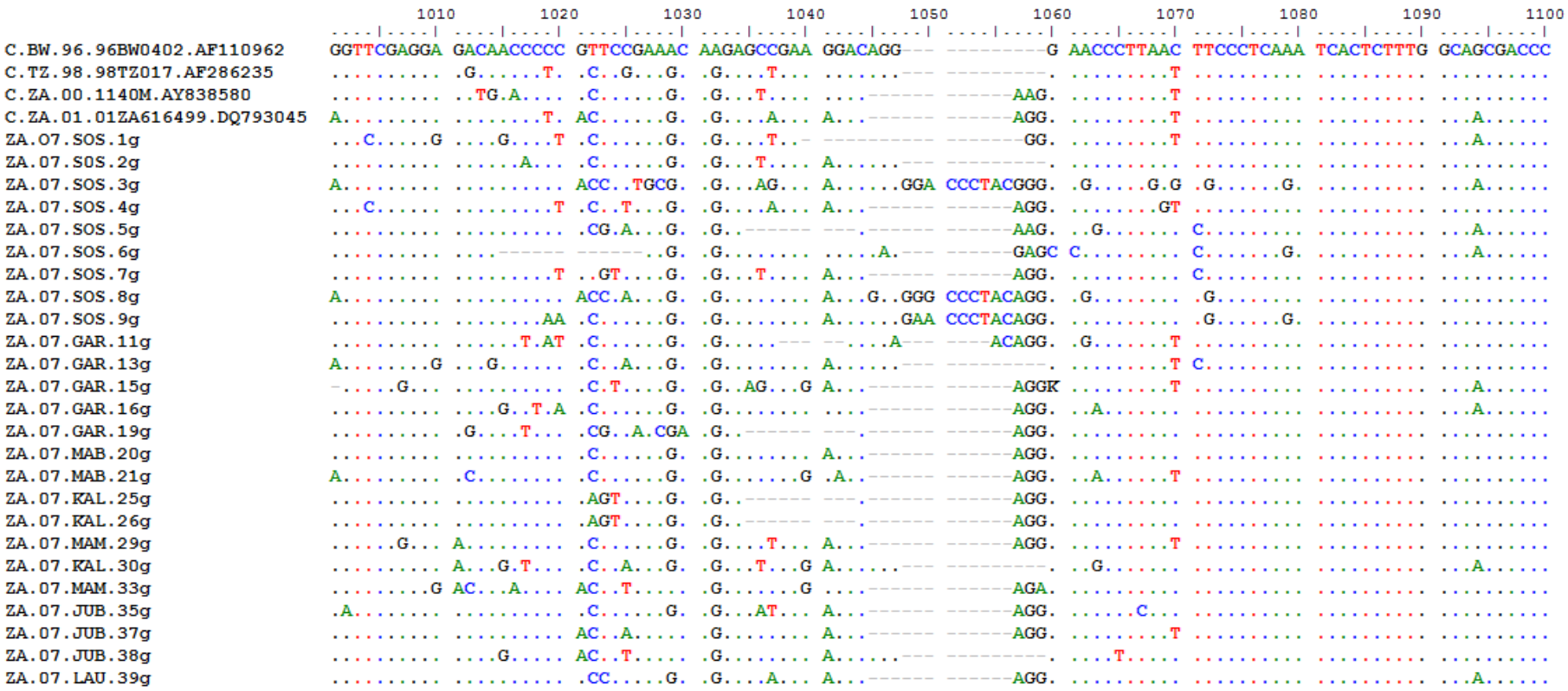

510 520 530 540 550 560 570 580 590 600
C.BW.96.96BW0402.AF110962 GACAGACACC TTGTTGGTCC AAAATGCGAA CCCAGATTGT AAGACCATTT TAAGGGCATT AGGACCAGGG GCTTCATTAG AAGAAATGAT GACAGCATGT
C.TZ.98.98TZ017.AF286235
C.ZA.00.1140M.AY838580
C.ZA.01.01ZA616499.DQ793045
ZA.07.SOS.1g
ZA.07.SOS.2g
ZA.07.SOS.3g
ZA.07.SOS.4g
ZA.07.SOS.5g
ZA.07.SOS.6g
ZA.07.SOS.7g
ZA.07.SOS.8g
ZA.07.SOS.9g
ZA.07.GAR.11g
ZA.07.GAR.13g
ZA.07.GAR.15g
ZA.07.GAR.16g
ZA.07.GAR.19g
ZA.07.MAB.20g
ZA.07.MAB.21g
ZA.07.KAL.25g
ZA.07.KAL.26g
ZA.07.MAM.29g
ZA.07.KAL.30g
ZA.07.MAM.33g
ZA.07.JUB.35g
ZA.07.JUB.37g
ZA.07.JUB.38g
ZA.07.LAU.39g


610 620 630 640 650 660 670 680 690 700
C.BW.96.96BW0402.AF110962 CAGGGAGTGG GAGGACCTAG CCACAAAGCA AGAGTTTTGG CTGAGGCAAT GAGCCAAAC- ----AAACA ---CAAATGT AATGATGCAG AGAAGCAATT
C.TZ.98.98TZ017.AF286235 .....G. .....A..... .....G. -----..... ATG.....A. ....
C.ZA.00.1140M.AY838580 ....A.... .....T...G. -----.C.G. GTA....CA. ..... .A.....
C.ZA.01.01ZA616499.DQ793045 .....G. ....GC... .....GG. -----..... GTG....CA. ....
ZA.07.SOS.1g .....G. .....G... ..G..G.... .... -----.... AAC.---.AA. ....
ZA.07.SOS.2g .C..... ....G..C.... ....G. -----..... ATA....CA. ....G....C.
ZA.07.SOS.3g ....G... .....G.... ....G. -----..C.. ATA....CA. ....A.....
ZA.07.SOS.4g .....G. A..... .....GC... ....G. -----..... GTA---.CA. .C.A..... .....
ZA.07.SOS.5g ....G..... ....G. -----..... ATC..C.A.C ....A...C.
ZA.07.SOS.6g .....G. ....G.C CAAGC.GG.. ATATG..CA. ....G.....
ZA.07.SOS.7g .....G. ....G..... ....G. -----..... GT--C.CA. ....A.....
ZA.07.SOS.8g .....G. ....G.... ....GT -----..... ATGG...CA. ....
ZA.07.SOS.9g .... -----..... ATG....CA. ....A.....
ZA.07.GAR.11g ....G. -----..... ACG....CA. ..... .A.....
ZA.07.GAR.13g .....G. ....C... .....T.. ...T...G. -----.GG.. ACC..C.CA. ....G.....
ZA.07.GAR.15g .....G. ....G.... ....G. -----..... ATAT...CA. ..... .A.G.....
ZA.07.GAR.16g ....G.... .... -----.C... GTG....CA. ....G.....
ZA.07.GAR.19g ....G..... ....G..... ....G. -----..... ATC..C.CA. ....AT....
ZA.07.MAB.20g .....G. ....G..G.... ....G. -----..... GTA---.CA. .....A .....
ZA.07.MAB.21g ....G.... ....G. -----..... GTA....CA. ..... .A.....
ZA.07.KAL.25g ....T..... ....G..... ....A..... .....G. -----..... ATG..C.CA. ....A.....
ZA.07.KAL.26g ....G..... ....A..... .....G. -----..... ATG..C.CA. ....A.....
ZA.07.MAM.29g .....G. ....G.... .G..... .....G. -----..... ATG....CA. ....G.....
ZA.07.KAL.30g .C..... ....G...C... ....G. -----...T. ATG....CA. ..... .A.G....C.
ZA.07.MAM.33g ....G.... ....G. -----..G.. GTA.T..CA. ....
ZA.07.JUB.35g .....G. ....A..... ....G. -----....T CC---C.CA. ....A.....
ZA.07.JUB.37g ....A.... .... -----.---. ATA.T...A. ....C.
ZA.07.JUB.38g ....G. -----G.... ATG....CA. ..... .A.....
ZA.07.LAU.39g ....G.... .... -----.C.A. GTA---GCA. .C.A..... .A.....


710 720 730 740 750 760 770 780 790 800
C.BW.96.96BW0402.AF110962 TTAAAGGCCC TAGAAGAATT GTTAAATGTT TCAACTGTGG CAAGGAAGGG CACATAGCCA GAAATTGCAG GGCCCCCAGG AAGAAAGGCT GTTGGAAATG
C.TZ.98.98TZ017.AF286235
C.ZA.00.1140M.AY838580
C.ZA.01.01ZA616499.DQ793045
ZA.07.SOS.1g
ZA.07.SOS.2g
ZA.07.SOS.3g
ZA.07.SOS.4g
ZA.07.SOS.5g
ZA.07.SOS.6g
ZA.07.SOS.7g
ZA.07.SOS.8g
ZA.07.SOS.9g
ZA.07.GAR.11g
ZA.07.GAR.13g
ZA.07.GAR.15g
ZA.07.GAR.16g
ZA.07.GAR.19g
ZA.07.MAB.20g
ZA.07.MAB.21g
ZA.07.KAL.25g
ZA.07.KAL.26g
ZA.07.MAM.29g
ZA.07.KAL.30g
ZA.07.MAM.33g
ZA.07.JUB.35g
ZA.07.JUB.37g
ZA.07.JUB.38g
ZA.07.LAU.39g


810 820 830 840 850 860 870 880 890 900
C.BW.96.96BW0402.AF110962 TGGAAAGGAA GGACACCAAA TGAAAGACTG TA---CTGAG AGGCAGGCTA ATTTTTTAGG GAAAATTTGG CCTTCCCACA A---GGGGAG GCCAGGGAAT
C.TZ.98.98TZ017.AF286235
C.ZA.00.1140M.AY838580
C.ZA.01.01ZA616499.DQ793045
ZA.07.SOS.1g
ZA.07.SOS.2g
ZA.07.SOS.3g
ZA.07.SOS.4g
ZA.07.SOS.5g
ZA.07.SOS.6g
ZA.07.SOS.7g
ZA.07.SOS.8g
ZA.07.SOS.9g
ZA.07.GAR.11g
ZA.07.GAR.13g
ZA.07.GAR.15g
ZA.07.GAR.16g
ZA.07.GAR.19g
ZA.07.MAB.20g
ZA.07.MAB.21g
ZA.07.KAL.25g
ZA.07.KAL.26g
ZA.07.MAM.29g
ZA.07.KAL.30g
ZA.07.MAM.33g
ZA.07.JUB.35g
ZA.07.JUB.37g
ZA.07.JUB.38g
ZA.07.LAU.39g


910 920 930 940 950 960 970 980 990 1000
C.BW.96.96BW0402.AF110962 TTCCTTCAGA GCAGACCAGA G CCA ACAGCCCCAC CAGCA GAGAGCTTCA
C.TZ.98.98TZ017.AF286235 ..T..... .... TA G AGCCAACAGC CCCAC..... ....
C.ZA.00.1140M.AY838580 .....C.... A...G..... .... CAG .. ..
C.ZA.01.01ZA616499.DQ793045 ..T..... .... CAG .. ..
ZA.07.SOS.1g .....C.... A..... .... CAGCAGAGAG CTT CAGACCAG AGCCAACAGC CCCAC....G ....
ZA.07.SOS.2g .....C.... A..... .... CA ... .
ZA.07.SOS.3g ..T..... .... CAG .. ..
ZA.07.SOS.4g .....C.... ....G..... .... CA C
ZA.07.SOS.5g .....C.... A..... ....A CACTA CCAACAGC CCCAC..CT. ....
ZA.07.SOS.6g .....C.... A..... .... CCATCAGC CCCAC..... ....
ZA.07.SOS.7g .....C.... A..... .... CGAGAG CCAACAGC CCCAC..CT. ....
ZA.07.SOS.8g ...A..... A.....T... .... CAG .. ..
ZA.07.SOS.9g .....C.... A.......AC AGCCCCA.T. GAGAG.AG.. CA G AGCCAACAGC CCCAC..... ....
ZA.07.GAR.11g .....C.... A...G..... A .... TA G AGCCAACAGC CCCAC..... ....
ZA.07.GAR.13g .....C.... A..... .... TA G AGCCAACAGC TCCAC..... ....
ZA.07.GAR.15g .....C.... A...G..... .... CAGCAGA GCCAACAGC CCCAC..... ....
ZA.07.GAR.16g ...A..... A.....T... .... CAG .. ..
ZA.07.GAR.19g .....C.... A..... .... CAATGGA GCCAACAGC CCCAC..ATG ....
ZA.07.MAB.20g .....C.... A..... .... CAG .. ..
ZA.07.MAB.21g ...Y..... A..... .... CAGA GCCAACAGC TCCAC..... ....
ZA.07.KAL.25g .....C.... A..... .... T.....- C..CT. ....
ZA.07.KAL.26g .....C.... A..... .... T.....- C..CT. ....
ZA.07.MAM.29g .....C.... A..... .... CAGTAGA GCCAACAGC CCCAC..... ....
ZA.07.KAL.30g .....C.... A..... .... CATCAGAGAG CTGCCTTCAG AGCAGACCAG AACCAACAGC TCCAC..... ....
ZA.07.MAM.33g .....C.... A..... .... CAACAGA GCCAACAGC CCCTC..... ....TT....
ZA.07.JUB.35g .....C.... ....G..... A .... T....T.... CGGAG CCATCAGC CCCAC..... ....
ZA.07.JUB.37g .....C.... A...G..... .... ..G .... CAG .. ..
ZA.07.JUB.38g ....C..... A.....T... ....
ZA.07.LAU.39g .....C.... A..... .... ..G .... CA ... .


1010
1020
1030
1040
1050
1060
1070
1080
1090
1100
C.BW.96.96BW0402.AF110962
C.TZ.98.98TZ017.AF286235
C.ZA.00.1140M.AY838580
C.ZA.01.01ZA616499.DQ793045
ZA.07.SOS.1g
ZA.07.SOS.2g
ZA.07.SOS.3g
ZA.07.SOS.4g
ZA.07.SOS.5g
ZA.07.SOS.6g
ZA.07.SOS.7g
ZA.07.SOS.8g
ZA.07.SOS.9g
ZA.07.GAR.11g
ZA.07.GAR.13g
ZA.07.GAR.15g
ZA.07.GAR.16g
ZA.07.GAR.19g
ZA.07.MAB.20g
ZA.07.MAB.21g
ZA.07.KAL.25g
ZA.07.KAL.26g
ZA.07.MAM.29g
ZA.07.KAL.30g
ZA.07.MAM.33g
ZA.07.JUB.35g
ZA.07.JUB.37g
ZA.07.JUB.38g
ZA.07.LAU.39g

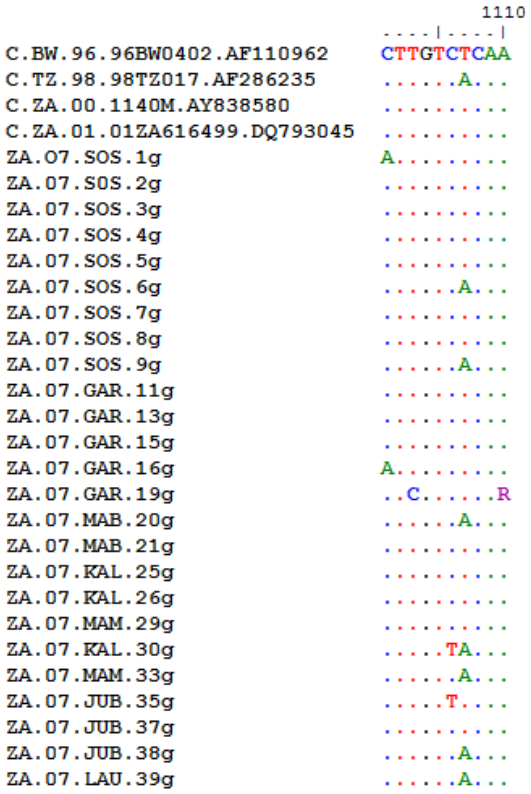
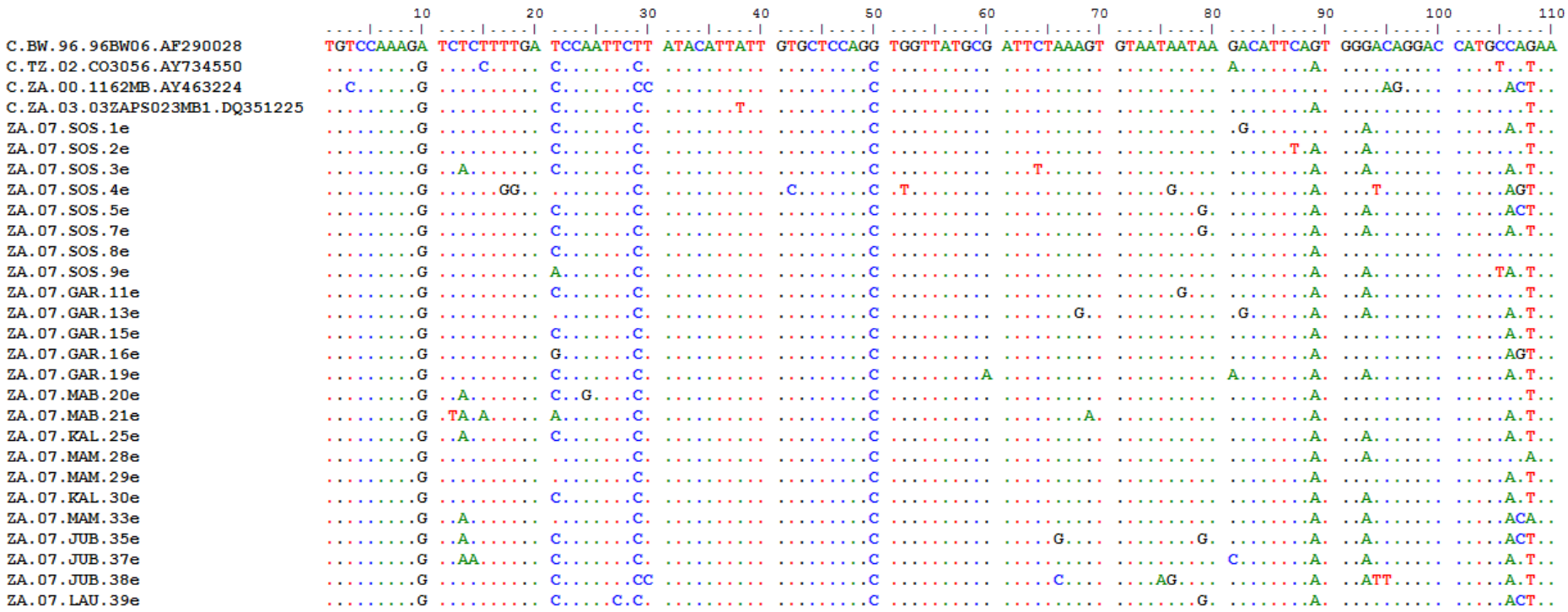
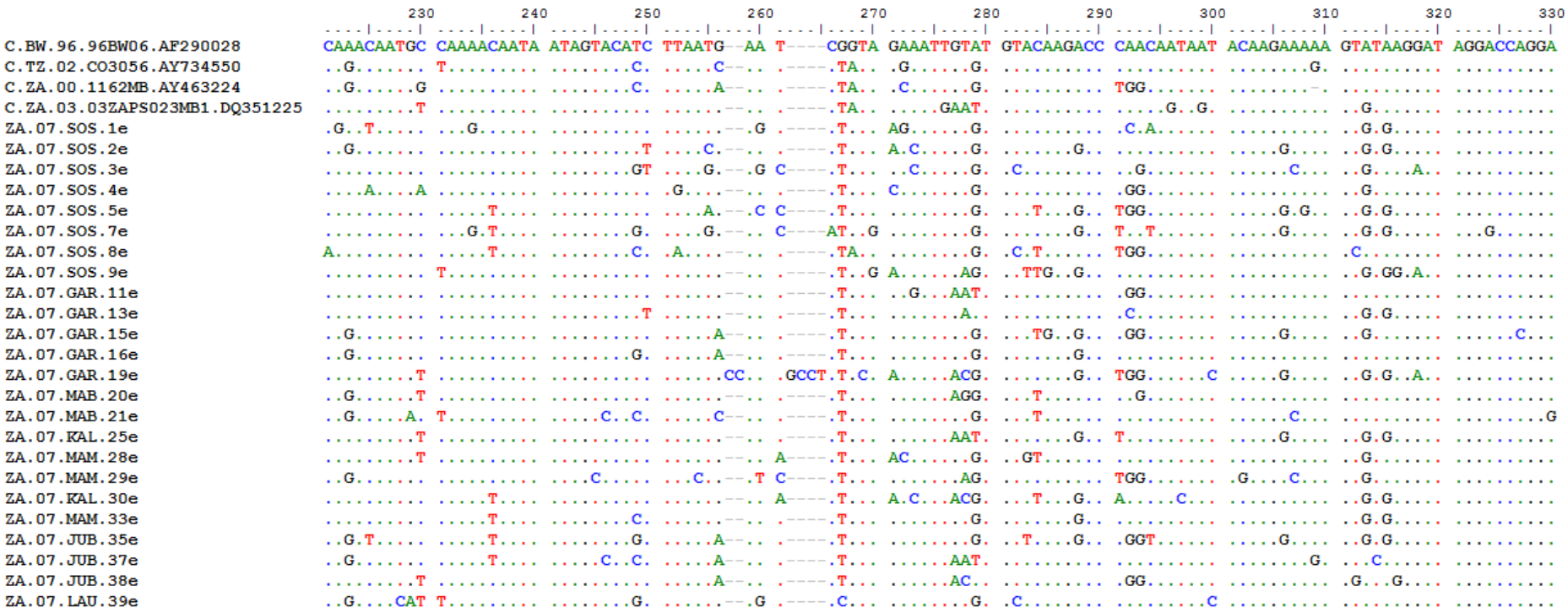
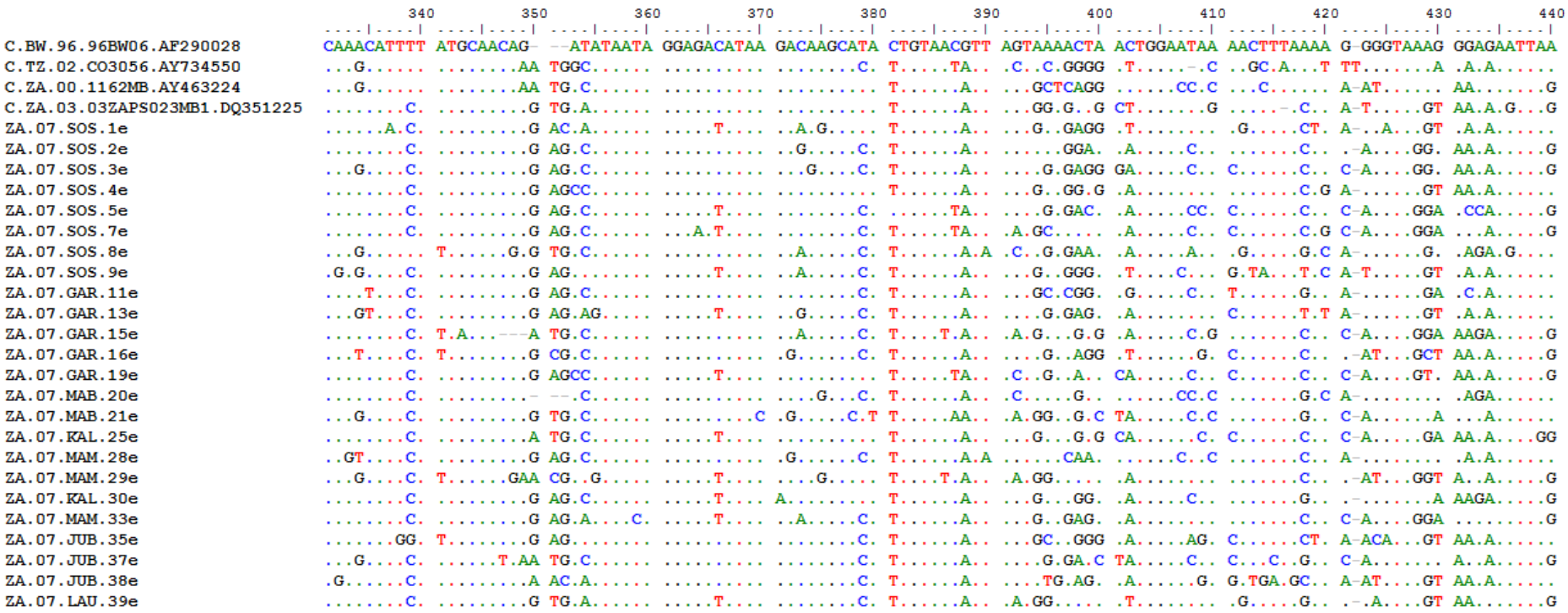


Figure 8.4: Graphic view of the amino acid alignments of *gag* sequences from 25 HIV-1 subtype C isolates. The sequences are compared to reference strains from the Los Alamos Sequence Database. (From Musyoki, 2009 MSc study). The sequences were used to generate phylogenetic tree on page 91.


10 20 30 40 50 60 70 80 90 100 110
C.BW.96.96BW06.AF290028 TGTCCAAAGA TCTCTTTTGA TCCAATTCTT ATACATTATT GTGCTCCAGG TGGTTATGCG ATTCTAAAGT GTAATAATAA GACATTCAGT GGGACAGGAC CATGCCAGAA
C.TZ.02.CO3056.AY734550 .....G ....C..... C.....C. ....C .... A.....A. ....T..T..
C.ZA.00.1162MB.AY463224 ..C.....G ..... C.....CC ....C ....AG.... .....ACT..
C.ZA.03.03ZAPS023MB1.DQ351225 .....G ..... C.....C. .....T.. .....C ....A. ....T..
ZA.07.SOS.1e .....G ..... C.....C. ....C .... .G..... ..A..... .....A.T..
ZA.07.SOS.2e .....G ..... C.....C. ....C ....T.A. ..A..... .....T..
ZA.07.SOS.3e .....G ..A..... C.....C. ....C ....T..... ....A. ..A..... .....A.T..
ZA.07.SOS.4e .....G .....GG.. ..... C..... .C.....C .T..... ....G.... .....A. ...T..... .....AGT..
ZA.07.SOS.5e .....G ..... C.....C. ....C ....G. .....A. ..A..... .....ACT..
ZA.07.SOS.7e .....G ..... C.....C. ....C ....G. .....A. ..A..... .....A.T..
ZA.07.SOS.8e .....G ..... C.....C. ....C ....A. ....
ZA.07.SOS.9e .....G ..... A.....C. ....C ....A. ..A..... ....TA.T..
ZA.07.GAR.11e .....G ..... C.....C. ....C ....G... .....A. ..A..... .....T..
ZA.07.GAR.13e .....G ....C. ....C ....G.. ..... .G.....A. ..A..... .....A.T..
ZA.07.GAR.15e .....G ..... C.....C. ....C ....A. ....A.T..
ZA.07.GAR.16e .....G ..... G.....C. ....C ....A. ....AGT..
ZA.07.GAR.19e .....G ..... C.....C. ....C .....A .... A.....A. ..A..... .....A.T..
ZA.07.MAB.20e .....G ..A..... C..G....C. ....C ....A. ....T..
ZA.07.MAB.21e .....G .TA.A..... A.....C. ....C ....A. ....A. ....A.T..
ZA.07.KAL.25e .....G ..A..... C.....C. ....C ....A. ..A..... .....A.T..
ZA.07.MAM.28e .....G ....C. ....C ....A. ..A..... .....A..
ZA.07.MAM.29e .....G ....C. ....C ....A. ....A.T..
ZA.07.KAL.30e .....G ..... C.....C. ....C ....A. ..A..... .....A.T..
ZA.07.MAM.33e .....G ..A..... ....C. ....C ....A. ..A..... .....ACA..
ZA.07.JUB.35e .....G ..A..... C.....C. ....C ....G.... .....G. .....A. ..A..... .....ACT..
ZA.07.JUB.37e .....G ..AA..... C.....C. ....C .... C.....A. ..A..... .....A.T..
ZA.07.JUB.38e .....G ..... C.....CC ....C ....C.... ....AG.... .....A. ..ATT..... .....A.T..
ZA.07.LAU.39e .....G ..... C.....C.C. ....C ....G. .....A. ....ACT..


120 130 140 150 160 170 180 190 200 210 220
C.BW.96.96BW06.AF290028 TGTCAGTACA GTACAATGTA CGCATGGAAT TAAGCCAGTG GTATCAACTC AATTACTGTT AAATGGTAGT CTAGCAGAAG AAAAGATAAT AATCAGATCT GGAAATGTGA
C.TZ.02.CO3056.AY734550
C.ZA.00.1162MB.AY463224
C.ZA.03.03ZAPS023MB1.DQ351225
ZA.07.SOS.1e
ZA.07.SOS.2e
ZA.07.SOS.3e
ZA.07.SOS.4e
ZA.07.SOS.5e
ZA.07.SOS.7e
ZA.07.SOS.8e
ZA.07.SOS.9e
ZA.07.GAR.11e
ZA.07.GAR.13e
ZA.07.GAR.15e
ZA.07.GAR.16e
ZA.07.GAR.19e
ZA.07.MAB.20e
ZA.07.MAB.21e
ZA.07.KAL.25e
ZA.07.MAM.28e
ZA.07.MAM.29e
ZA.07.KAL.30e
ZA.07.MAM.33e
ZA.07.JUB.35e
ZA.07.JUB.37e
ZA.07.JUB.38e
ZA.07.LAU.39e


230 240 250 260 270 280 290 300 310 320 330
C.BW.96.96BW06.AF290028 CAAACAATGC CAAAACAATA ATAGTACATC TTAATG--AA T----CGGTA GAAATTGTAT GTACAAGACC CAACAATAAT ACAAGAAAAA GTATAAGGAT AGGACCAGGA
C.TZ.02.CO3056.AY734550
C.ZA.00.1162MB.AY463224
C.ZA.03.03ZAPS023MB1.DQ351225
ZA.07.SOS.1e
ZA.07.SOS.2e
ZA.07.SOS.3e
ZA.07.SOS.4e
ZA.07.SOS.5e
ZA.07.SOS.7e
ZA.07.SOS.8e
ZA.07.SOS.9e
ZA.07.GAR.11e
ZA.07.GAR.13e
ZA.07.GAR.15e
ZA.07.GAR.16e
ZA.07.GAR.19e
ZA.07.MAB.20e
ZA.07.MAB.21e
ZA.07.KAL.25e
ZA.07.MAM.28e
ZA.07.MAM.29e
ZA.07.KAL.30e
ZA.07.MAM.33e
ZA.07.JUB.35e
ZA.07.JUB.37e
ZA.07.JUB.38e
ZA.07.LAU.39e


340 350 360 370 380 390 400 410 420 430 440
C.BW.96.96BW06.AF290028 CAAACATTTT ATGCAACAG- --ATATAATA GGAGACATAA GACAAGCATA CTGTAACGTT AGTAAAACTA ACTGGAATAA AACTTTAAAA G-GGGTAAAG GGAGAATTAA
C.TZ.02.CO3056.AY734550 ...G..... ....AA TGGC..... ....C. T.....TA.. .C..C.GGGG .T.....-.C ..GC.A...T TT.......A .A.A.....
C.ZA.00.1162MB.AY463224 ...G..... ....AA TG.C..... .... T.....A.. ...GCTCAGG .....CC.C ...C..... A-AT..... AA.......G
C.ZA.03.03ZAPS023MB1.DQ351225 ....C. ....G TG.A..... .... T.....A.. ...GG.G..G CT.......G .....-.C.. A-T.....GT AA.A.G...G
ZA.07.SOS.1e ......A.C. ....G AC.A..... .....T.... ..A.G..... T.....A.. ...G..GAGG .T..... .G.....CT. A-..A...GT .A.A.....
ZA.07.SOS.2e ....C. ....G AG.C..... ....G.....C. T.....A.. .....GGA. .A.....C.. .....C.. .-A....GG. AA.A.....G
ZA.07.SOS.3e ...G....C. ....G AG.C..... ....G....C. T.....A.. ....G.GAGG GA.....C.. C.....C.. C-A....GG. AA.A.....G
ZA.07.SOS.4e ....C. ....G AGCC..... .... T.....A.. ...G..GG.G .A..... .....C.G A-.....GT AA.A.....
ZA.07.SOS.5e ....C. ....G AG.C..... .....T.... ....C. .....TA.. ....G.GAC. .A.....CC. C.....C.. C-A....GGA .CCA.....G
ZA.07.SOS.7e ....C. ....G AG.C..... ...A.T.... ....C. T.....TA.. .A.GC..... .A.....C.. C.....C.G C-A....GGA ...A.....G
ZA.07.SOS.8e ...G..... T......G.G TG.C..... ....A.....C. T.....A.A .C..G.GAA. .A.....A.. .G.....G.C A-.....G. .AGA.G....
ZA.07.SOS.9e .G.G....C. ....G AG..... .....T.... ..A.....C. T.....A.. ...G..GGG. .T....C... G.TA...T.C A-T.....GT .A.A.....
ZA.07.GAR.11e ....T...C. ....G AG.C..... ....C. T.....A.. ...GC.CGG. .G.....C.. T.....G.. A-.....GA .C.A.....
ZA.07.GAR.13e ...GT...C. ....G AG.AG..... .....T.... ..G.....C. T.....A.. ....G.GAG. .A..... C.....T.T A-.....GT .A.A.....
ZA.07.GAR.15e ....C. T.A...---A TG.C..... ....A.....C. T....T.A.. .A.G...G.G .A.....C.G .....C.. C-A....GGA AAGA.....G
ZA.07.GAR.16e ...T....C. T........G CG.C..... .... .G.....C. T.....A.. ....G..AGG .T.....G. C.....C.. .-AT...GCT AA.A.....G
ZA.07.GAR.19e ....C. ....G AGCC..... .....T.... .... T.....TA.. .C..G..A.. CA.....C.. C.....C.. C-A....GT. AA.A.....G
ZA.07.MAB.20e ....C. ....- --.C..... ....G...C. T.....A.. .C.....G.. .....CC.C .....G.C A-..... .AGA.....
ZA.07.MAB.21e ...G....C. ....G TG.C..... ....C .G.....C.T T.....AA.. .A.GG..G.C TA.....C.C .....G.. C-A.....A ...A.....
ZA.07.KAL.25e ....C. ....A TG.C..... .....T.... .... T.....A.. ...G...G.G CA.....C. C.....C.. C-A.....GA AA.A....GG
ZA.07.MAM.28e ..GT....C. ....G AG.C..... .... .G.....C. T.....A.A .....CAA. .....C..C .....C.. A-..... .A.A.....
ZA.07.MAM.29e ...G....C. T......GAA CG..G..... .....T.... ....G..... T....T.A.. .A.GG..... .A..... .....C.. .-AT...GGT A..A.....G
ZA.07.KAL.30e ....C. ....G AG.C..... .....T.... A..... ..... T.....A.. ...G...GG. .A.....C.. .....G.. .-..... .A AAGA.....G
ZA.07.MAM.33e ....C. ....G AG.A....C. .....T.... ..A.....C. T.....A.. ...G..GAG. .A..... .....C.. C-A....GGA .....G
ZA.07.JUB.35e ....GG. T........G AG..... ....C. T.....A.. ...GC..GGG .A.....AG. C.....CT. A-ACA...GT AA.A.....
ZA.07.JUB.37e ...G....C. ......T.AA TG.C..... ....C. T.....A.. ....G.GA.C TA.....C.. C...C..G.. C-A..... A..A.....G
ZA.07.JUB.38e .G......C. ....A AC.A..... ....C. T.....A.. ....TG.AG. .A.....G. G.TGA.GC.. A-AT....GT AA.A.....
ZA.07.LAU.39e ....C. ....G TG.A..... .....T.... ....C. T.....A.. .A.GG..... .T..... .G.....G.. .-.A....GT AA.....G

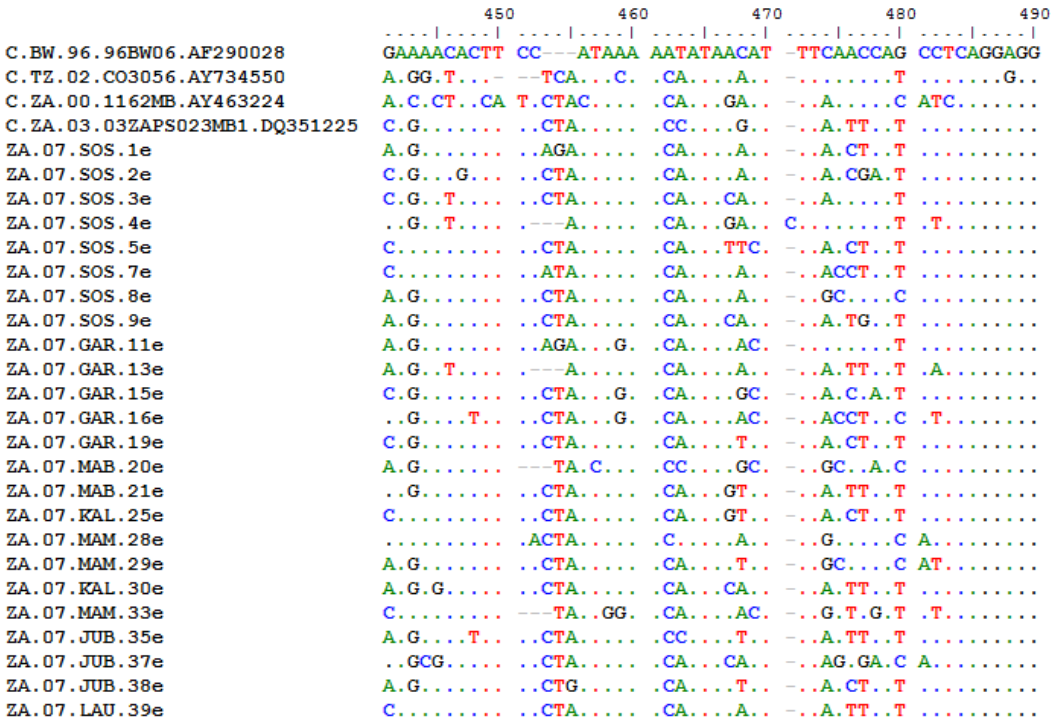


Figure 8.5: Graphic view of the amino acid alignments of *env* sequences from 25 HIV-1 subtype C isolates. The sequences are compared to reference strains from the Los Alamos Sequence Database. (From Musyoki, 2009 MSc study). The sequences were used to generate phylogenetic tree on page 91.