GENETIC CHARACTERIZATION OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 IN SOUTH AFRICA

Wilhelmina Jacoba Bredell

Degree awarded with distinction on 2 December 1999

A dissertation submitted to the Faculty of Medicine,
University of the Witwatersrand, Johannesburg,
for the Degree of Master of Science in Medicine.

Johannesburg, 1999

ABSTRACT

A study was conducted to investigate the genetic diversity of HIV-1 in South Africa. Blood was collected from 172 HIV-1 seropositive individuals, representing different risk groups, and subtyped by heteroduplex mobility assay (HMA) using an ~700-bp *env* region. HIV-1 subtype C predominated and was present in 97% of heterosexually infected individuals. Ninety three percent of individuals infected homosexually and 89% of recipients of contaminated blood products were infected with subtype B viruses. Four *env* subtype A viruses were also identified. In order to screen for recombinant viruses, 117 of these samples were also subtyped using a *gag* HMA based on an ~540-bp *gag* region. There was concordance in 113 samples with both *gag* and *env* HMA identifying the same subtype. Four recombinant viruses were found which were mosaics between subtype C and regions belonging to either subtypes A or B. Thus, the prevalence of recombinant viruses in this study is relatively low (3%).

A second study examined the phylogenetic relationships between 44 HIV-1 isolates from 43 infected subjects employed by three adjacent South African gold mines. The patients were migrant workers originating from the rural areas of South Africa and neighbouring countries of Lesotho, Botswana, Swaziland and Mozambique. HIV-1 was subtyped by *env* HMA and sequence analysis was used to determine phylogenetic relationships between isolates. All isolates were identified as *env* subtype C. These isolates did not show a distinct phylogenetic relatedness based on the geographic origins of the migrant workers or show close homology to other subtype C sequences from southern Africa or India. The heterogeneity among these samples suggests multiple introductions of HIV-1 subtype C into this population. However, five clusters of closely related sequences were identified, mainly involving miners of disparate geographic origins, suggesting possible epidemiological linkage in these few cases.

In summary, subtype C is the predominant HIV-1 subtype in Gauteng associated with heterosexual transmission. Globally, subtype C has become the most prevalent subtype and it is likely that future vaccine strategies will be directed towards this subtype. Continual genetic characterization of HIV in South Africa is important to monitor the ongoing evolution of this subtype and to document the emergence of new subtypes and inter-subtype recombinants.

DECLARATION

I declare that this dissertation is my own, unaided work. It is being submitted for the degree of Master of Science in Medicine at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other university.

Wilhelmina Jacoba Bredell

15 day of November, 1999

To my husband, Riaan, with sincere thanks for his love and encouragement during the writing of this dissertation.

ACKNOWLEDGEMENTS

My sincere thanks go to the following people:

My supervisor, Dr L Morris, for her continuous guidance, encouragement, insight and interest shown throughout this study.

The Poliomyelitis Research Foundation (PRF) and Medical Research Council (MRC) of South Africa for providing the funding for this study.

Drs P Sonnenberg, B Morgan, D Spencer, R Sher, SF Lyons, G Gray, D Britain, S Abdool-Karim, G Ramjee, and L Sacks and their staff for providing specimens and demographic data, and also allowing me to personally interview some of their patients.

Dr P Louw and Gold Fields West Hospital for allowing us to conduct this study.

Dr CT Tiemessen and Ms GM Hunt for developing the *gag* HMA.

Ms S Nyoka for performing the $CD4^+$ T-cell counts.

Mrs J Croft for automated sequencing.

Dr C Williamson for assistance with the phylogenetic analysis.

Mrs H Saevitzon and Mrs S Holmes, the librarians of the National Institute for Virology.

Dr D Martin and Prof B Schoub for supporting this work.

My family, friends and colleagues for constant support and encouragement.

CONTENTS

CHAPTER TWO

MATERIALS AND METHODS

CHAPTER THREE

HIV-1 SUBTYPES IN SOUTH AFRICA AND THE IDENTIFICATION OF INTER-SUBTYPE RECOMBINANTS

CHAPTER FOUR

GENETIC CHARACTERIZATION OF HIV-1 FROM MIGRANT WORKERS IN THREE SOUTH AFRICAN GOLD MINES

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

aa	amino acid
AIDS	acquired immunodeficiency syndrome
ATP	adenosine triphosphate
AZT	azidothymidine
bp	base pair
BWA	Botswana
CDC	Center for Disease Control
cDNA	complementary deoxyribonucleic acid
CIAA	chloroform:isoamylalcohol
CNS	central nervous system
CoMA	combinatorial melting array
CTL	cytotoxic T lymphocytes
dATP	deoxy-adenosine triphosphate
dCTP	deoxy-cytidine triphosphate
ddATP	dideoxy-adenosine triphosphate
ddCTP	dideoxy-cytidine triphosphate
ddGTP	dideoxy-guanosine triphosphate
ddI	didanosine
ddTTP	dideoxy-thymidine triphosphate
dGTP	deoxy-guanosine triphosphate
DIG	digoxigenin
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleoside triphosphates
DTT	dithiothreitol
dTTP	deoxy-thymidine triphosphate
EDTA	ethylenediaminetetra-acetic acid
ELISA	enzyme-linked immunosorbent assay
EthBr	ethidium bromide
H_2O	water
HCL	hydrochloric acid
HIV-1	human immunodeficiency virus type 1
HIV-2	human immunodeficiency virus type 2
HMA	heteroduplex mobility assay
IPTG	isopropylthio-β-D-galactoside
IVDU	intravenous drug users
KCL	potassium chloride
LB	Luria broth
LC	Langerhans cells
LSO	Lesotho
LTR	long-terminal repeat
M group	major group

IPTG	isopropylthio-β-D-galactoside
M	molar
mA	milliampere
mg	milligram
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulphate
ml	millilitre
mM	millimolar
Mn^{2+}	manganese
MOZ	Mozambique
mRNA	messenger-ribonucleic acid
NaAc	sodium acetate
NaCl	sodium chloride
NSI	non-syncitium-inducing
O group	outlier group
OD	optical density
PBMC	peripheral blood mononuclear cells
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEIA	peptide-binding enzyme immunoassays
PEP	post-exposure prophylaxis
pmol	picomol
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RNAses	ribonucleases
rpm	revolutions per minute
RRE	Rev response element
RT-PCR	reverse transcriptase-polymerase chain reaction
SDS	sodium dodecyl sulphate
SI	syncitium-inducing
SIV	simian immunodeficiency virus
SSCP	single-strand conformational polymorphism
STD	sexually transmitted disease
SWZ	Swaziland
TAR	transactivation response element
TB	tuberculosis
TEMED	N,N,N',N'-tetramethylethylenediamine
U	unit
UV	ultraviolet
WHO	World Health Organization
X-gal	5-bromo-4-chloro-3-idolyl-β-D-galactoside
ZA	South Africa
μg	microgram
μl	microlitre
μM	micromolar

CHAPTER ONE

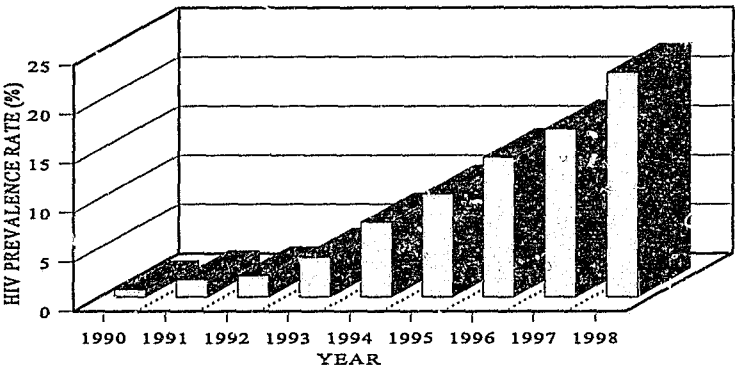
LITERATURE REVIEW

1.1 The HIV epidemic

In 1981 a number of previously healthy, young homosexual men were treated for *Pneumocystis carninii* pneumonia and Kaposi's sarcoma (CDC, 1981a, CDC, 1981b). The underlying disorder in these patients was found to be a severe immune deficiency with a dramatic decline in $CD4^+$ cells, which was subsequently given the name acquired immunodeficiency syndrome (AIDS). After two years of intensive research, Montagnier and coworkers identified a retrovirus, which grew to high titres in $CD4^+$ cells and was cytopathic, as the etiological agent of AIDS (Montagnier *et al.*, 1984). In 1986, the International Committee on Taxonomy of Viruses recommended the name, human immunodeficiency virus (HIV) (Coffin *et al.*, 1986). Soon after the discovery of HIV, a second immunodeficiency virus was identified in West Africa (Clavel *et al.*, 1986). This virus was designated HIV-2, and the first was subsequently renamed HIV-1. Both viruses belong to the family *Retroviridae*, subfamily *Lentivirinae*, which are slow acting, cytopathic viruses associated with long incubation periods and infection of the central nervous system (CNS) (Hehlmann, 1991).

Since then, nearly 31 million people have become infected with HIV and almost 12 million people around the world have died of AIDS. Over two-thirds of all these people live in sub-Saharan Africa, and 83% of the world's AIDS deaths have been in this region. HIV in sub-Saharan Africa is mostly spread through heterosexual transmission which means that women and children are more heavily burdened in Africa than in other regions (UNAIDS/WHO, 1998). It has been estimated by the WHO that by the year 2000 there will be 40 million people infected with HIV; more than 90% will be in developing countries such as Asia and sub-Saharan Africa (WHO, 1995). In developed countries such as the USA, the routine use of increasingly intensive anti-retroviral therapies has resulted in the dramatic decrease in morbidity and mortality among HIV-infected patients (Palella *et al.*, 1998). However, HIV-1 continues to spread in developing countries which are least able to afford anti-retroviral therapy, causing nearly 16,000 new infections a day (UNAIDS/WHO, 1998).

South Africa currently has the fastest growing epidemic in the world. In 1990, less than 1% of the women attending ante-natal clinics were infected, however, since then the prevalence has doubled roughly every one to two years (Figure 1.1). By the end of 1998, 22.8% of pregnant women were HIV-infected with KwaZulu/Natal province being the worst affected and Western Cape province the least affected. Based on these figures it is estimated that 3.6 million people in South Africa are infected with HIV-1 (Department of Health, 1999).



Redrawn from Department of Health Report (Department of Health, 1999)

Figure 1.1 HIV prevalence trends in South Africa (Based on the national surveys in antenatal clinic attenders: 1990-1998)

1.2 Modes of transmission

1.2.1 Sexual

HIV-1 is transmitted by both heterosexual and homosexual activity and the likelihood of becoming infected is related to the number of sexual partners and to different sexual practices (Anderson & May, 1988). In developed countries, such as the United States, anal intercourse between men is the most common mode of transmission, whereas in developing countries such as in Africa and Asia, heterosexual vaginal intercourse is the dominant mode of transmission.

Male-to-female transmission is approximately eight times more efficient than female-to-male transmission (Padian *et al.*, 1997). The vagina is more vulnerable to infection as large volumes of infected semen remain in contact with a substantial area of susceptible vaginal and cervical epithelium for long periods of time (Schoub, 1994). Transmission is likely to occur through cell-associated virus as this is present in higher numbers than cell-free virus in both semen and vaginal secretions (Levy, 1994). Rectal intercourse has a greater risk of transmitting the virus as the rectal mucosa is very delicate compared to the vaginal mucosa which is multi-layered and relatively robust. HIV gains entry in the vagina or the anal canal by infecting the epithelial cells (M-cells in the rectal mucosa, and $CD4^+$-bearing Langerhans cells in the vaginal mucosa) directly or through abrasions and genital ulcers in the mucosa. Uncircumcised men are significantly more susceptible to HIV infection than those who are circumcised, due to retention of infected cells under the foreskin. In both homosexual and heterosexual intercourse, the recipient partner is of greater risk than the insertive partner (Schoub, 1994). Sexually transmitted diseases (STD), associated with large numbers of inflammatory cells containing virus-infected cells, and often involving ulcerations or discharges, increases the risk of transmission by 5 to 20 times (Royce *et al.*, 1997). Successful treatment of STDs results in a significant reduction in vaginal shedding of HIV-1 which may reduce HIV-1 transmission (Htun *et al.*, 1999).

1.2.2 Vertical

HIV transmission from an HIV-infected mother to her infant occurs via three pathways: a) pre-partum/intra-uterine (across the placenta), b) intra-partum/perinatal (exposure to an infected genital tract, maternal blood or vaginal secretions during delivery), and c) postpartum/postnatal, through breast milk (Ahmad *et al.*, 1995).

There are various factors which influence the transfer of HIV from mother to infant: poor clinical stage of the mother; low levels or absence of anti-V3 loop antibodies; reduced $CD4^+$ cell count; high viral load; and the presence of high-replicating, syncitium-inducing strains in the mother at the time of delivery (Husson *et al.*, 1995). Up to half of infant infections occur during birth through contact with the amniotic fluid, genital secretions, or blood of the mother. In the case of twins, the first born is at higher risk of becoming infected as it clears the vaginal birth canal during delivery (Goedert *et al.*, 1991). Caesarean deliveries have been associated with a reduction in risk of acquiring HIV infection (between 20 to 50%), although in other studies it did

not demonstrate any protective effect (Villari *et al.*, 1993). In the absence of any interventions, rates for perinatal transmission range from 15% - 35%. The highest rates are reported in Africa where the clinical stage of the mother may be more severe and breast-feeding is a common practice (Newell & Peckham, 1993). Recent studies have shown that the systematic use of azidothymidine (AZT) reduces the transmission rate to less than 10% (Masquelier *et al.*, 1999).

1.2.3 Blood and blood product transfusion

Before 1985, transmission by blood transfusion accounted for more than 90% of HIV-1 transmission. However, HIV-1 antibody screening of donated blood has significantly reduced HIV transmission through this route, which now accounts for less than 5% of all new cases (Schoub, 1994). Prolonged heat treatment (>60°C) of clotting factors VIII and IX has virtually eliminated HIV transmission via these products (Levy *et al.*, 1985).

1.2.4 Intravenous drug use

HIV is transmitted among injecting drug users (IVDU) through the sharing of contaminated needles and other equipment. Transmission is related to the frequency of needle sharing, the immunosuppressiveness of certain drugs, and the prevalence of HIV infection in the particular area. The inaccessibility of these groups to educational and needle-exchange programmes in countries where the use of drugs is illegal may propagate behaviour which increases transmission (Schoub, 1994). IVDU are the main risk group for HIV-1 infection in Spain (Registro de casos de SIDA, 1993), whereas, in South Africa very few cases have been reported.

1.2.5 Accidental needle-stick injuries

The risk for acquiring the virus through an accidental needle-stick injury is < 0.3%, but nevertheless remains an important and preventable route of transmission. The risk of transmission is dependent on the volume of infected blood injected, the clinical stage of the patient from whom the blood was taken as well as how immunocompromised the individual is who has been pricked (Schoub, 1994). Appropriate post-exposure prophylaxis (PEP) after per cutaneous exposure to infected blood may reduce the risk for HIV seroconversion by 79% and is now practised in many countries including South Africa (CDC, 1996).

1.3 The HIV virion

1.3.1 Structure

High-resolution electron microscopy has revealed the HIV virion as an icosahedral structure (Gelderblom *et al.*, 1987), 100 to 150 nm in diameter, containing 72 external spikes (Connor *et al.*, 1997, Green, 1991). Mature particles are characterized by an electron-dense cone-shape capsid, which consist of (i) the major capsid protein, p24; (ii) the nucleocapsid proteins, p7 and p9; (iii) the diploid single-stranded RNA genome; and (iv) the three viral enzymes, protease, reverse transcriptase, and integrase. The viral capsid is surrounded by a matrix protein, p17, located underneath the lipid bilayer envelope membrane that is acquired as the virion buds from an infected cell. The viral genomic RNA encodes a series of structural, enzymatic and regulatory proteins that are required to complete the viral life cycle efficiently (Barre-Sinoussi, 1996). Each of the 72 spikes are formed by the two major viral-envelope proteins, an external glycosylated protein gp120 and a transmembrane protein gp41, that are embedded in the lipid membrane (Montagnier & Clavel, 1994).

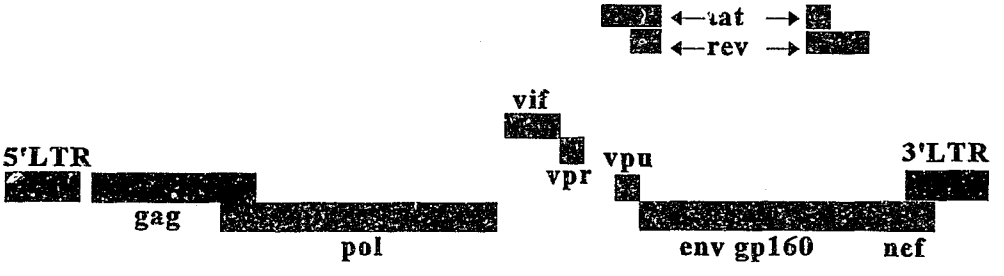
1.3.2 Genomic organization

The genome of HIV is approximately 10 kb in length with open reading frames coding for several viral proteins (Figure 1.2). Flanked by two long-terminal repeats (LTR), the genome consists of three main genes: *gag*, which encodes the virion structural components; *pol*, which encodes several viral enzymes; and *env*, which encodes the envelope glycoproteins. The HIV-1 genome also contains at least another six additional genes: *tat*, *rev*, *nef*, *vpr*, *vpu* (or *vpx* in HIV-2) and *vif*. Proteins encoded by these genes function both early and late in the replication cycle to regulate transcription and translation of viral proteins and to enhance assembly and release of particles from infected cells (Connor *et al.*, 1997).

1.3.3 Major proteins

The major core proteins of HIV are synthesized from *gag* as a large precursor protein (pr55), which is subsequently proteolytically cleaved by the HIV protease to yield p17, p24, p7 and p9 proteins. Both p7 and p9 are capable of binding genomic RNA and together form the nucleoid core (Green, 1991). A minus one nucleotide ribosomal frame shift occurs between the first and second long open reading frame of the HIV genome, resulting in the synthesis of a *gag*-

pol precursor protein (pr170) (Jacks *et al.*, 1988). The ratio of *gag-pol* to *gag* produced is approximately 1:20. Subsequent cleavage of this precursor protein produces the *gag* protein products and the retroviral enzyme protease (p11), reverse transcriptase-ribonuclease H (p51/p66), and integrase (p32). With the exception of HIV protease, these enzymes function in the early stages of the replication cycle to convert the viral genomic RNA to DNA and to mediate insertion of the viral genome into the host cell chromosome. By contrast, HIV-1 protease acts during assembly and release of virions to cleave the *gag* and *gag-pol* polyproteins (Kaplan *et al.*, 1994). The major protein product of the *env* gene is synthesized in the host cell cytoplasm as an 88 kd polypeptide (Willey *et al.*, 1988). After a series of post-translation glycosylation steps, primarily with N-linked sugars in the endoplasmic reticulum and Golgi apparatus, the final protein product is about 160 kd. A fraction of the newly synthesized gp160 is spliced by a cellular protease to generate the transmembrane (gp120) and surface (gp41) subunits (McCune *et al.*, 1988). Cleavage of the integral membrane proteins is of critical importance in the virus life cycle as mutations within the endoproteolytic cleavage site of the *env* gene result in the release of noninfectious particles. Following this cleavage step, gp120 remains non-covalently associated with gp41 on the cell surface and is incorporated into the envelope of budding virions as a trimeric complex in a typical spike-and-knob configuration (Ozel *et al.*, 1988).



Redrawn from J.A. Levy (Levy, 1994)

Figure 1.2 Genomic map of HIV-1

1.3.4 The life cycle of HIV

1.3.4.1 Attachment of the virion:

HIV-1 can infect many different human cells in culture, but *in vivo* $CD4^+$ T lymphocytes and monocytes are the major cellular targets (Fauci, 1988). The binding of gp120 to $CD4^+$ occurs through the interaction of conformationally defined domains. These molecular "twists" reveal additional attachment sites for the viral envelope to bind to a second molecule, a seven-transmembrane domain G-protein coupled chemokine receptor, either CXCR4 or CCR5 (Levy, 1996). Non-syncitium-inducing (NSI) macrophage-tropic HIV-1 strains use the CC chemokine receptor CCR5. This co-receptor is used by almost all primary HIV-1 isolates regardless of viral genetic subtype (Alkhatib *et al.*, 1996, Deng *et al.*, 1996, Dragic *et al.*, 1996). Syncitium-inducing (SI) T-tropic isolates use the CXCR4 co-receptor (Feng *et al.*, 1996), although many of them also use CCR5 (Simmons *et al.*, 1996).

1.3.4.2 Internalization of the virion:

Internalization of the receptor-bound HIV-1 virion by the cell takes place through either receptor-mediated endocytosis (Maddon *et al.*, 1986) or virus-mediated membrane fusion (Stein *et al.*, 1987), although the latter appears to be the dominant mechanism (Bedinger *et al.*, 1988). Chemokine-receptor binding (CCR5 or CXCR4) is believed to trigger additional conformational changes in the envelope glycoprotein trimer that leads to the exposure of the three N-terminal fusion peptides of the trimeric gp41. These fusion peptides interact with a still undefined cell-membrane (fusogenic) domain in the lipid bilayer of the host. Activation of this domain leads to fusion of the viral and host cell membranes which in turn promotes internalization of the virion (Kwong *et al.*, 1998).

1.3.4.3 Reverse transcription and integration:

After internalization, the HIV-1 virion is stripped of its lipid membrane (Varmus, 1988), and the capsid complex moves into the cytoplasm. Reverse transcription in initiated with the conversion of viral RNA to DNA mediated by the coordinated polymerase and ribonuclease H activities of the HIV-1-encoded reverse transcriptase. This is done by the synthesis of a complementary strand of DNA from the viral genomic RNA. Ribonuclease H selectively degrades the original RNA template while a second strand of DNA is synthesized, yielding a double-stranded DNA copy of the original viral RNA genome (Goff, 1990). This double-stranded

DNA replica is translocated to the nucleus and inserted randomly into the host genome by the viral integrase. Once integrated, this DNA copy of the retroviral genome is called a "provirus" (Barre-Sinoussi, 1996).

1.3.4.4 Expression of early and late genes:

After integration, cellular DNA-dependent RNA polymerases generate genomic-length viral pre-messenger RNA (pre-mRNA) transcripts using the provirus as a template. These early transcripts are transported to the cytoplasm as multiply spliced viral RNAs that are translated into various regulatory proteins including Tat, Rev and Nef. Thereafter, Tat and Rev shuttle back to the nucleus. Initially the level of viral-specific transcription is low in the absence of Tat. Tat binds directly to an RNA stem-loop structure, termed the transactivation response element (TAR), located at the 5' end of all viral mRNAs. It acts primarily at the level of transcription initiation and elongation, giving rise to full-length HIV-1 transcripts rather than the short, truncated transcripts. Transcription of early regulatory proteins predominates while a significant amount of Rev protein accumulates in the nucleus. In the presence of Rev, unspliced and singly-spliced viral RNA transcripts (late transcripts) encoding structural and enzymatic proteins, as well as the viral genomic RNA, are transported to the cytoplasm. Rev functions through a five stem-loop RNA structure, the Rev response element (RRE), located within the *env* gene of the mRNA species whose expression is regulated by Rev. Once transported, these late transcripts are translated to the products of the *env*, *gag* and *pol* genes (Montagnier & Clavel, 1994).

1.3.4.5 Assembly and budding:

Assembly of new virions occurs at the plasma membrane of the host cell. The process involves the aggregation of the ribonucleoprotein cores, composed of the HIV-1 RNA, and *gag* proteins, together with the various enzymes encoded by the *pol* gene, in the cytoplasm. Virions are formed when these cores bud through the plasma membrane during which they acquire their lipid membranes, complete with the external (gp120) and transmembrane (gp41) envelope glycoproteins (Connor *et al.*, 1997). These virions are then able to initiate new cycles of infection.

1.4 Genetic variability of HIV-1

In the absence of anti-retroviral therapy HIV-1 shows extremely high continuous turnover of virus with up to 10^9 virions produced per day (Ho *et al.*, 1995) resulting in a substantial viral burden (Piatak *et al.*, 1993). This high replicative capacity together with its ability to mutate has resulted in considerable genetic diversity in the HIV-1 genome. The *env* gene is the most variable region, where divergence between isolates of different subtypes can be up to 30% (Montagnier & Clavel, 1994). It is estimated that within an infected person HIV-1 sequences change by approximately 1% per year in the *env* gene (Myers *et al.*, 1993). Thus, there exists a swarm or quasispecies of related but distinct HIV variants which increase over time and affect the biological phenotype and possibly also virulence and transmissibility of the virus (Hahn *et al.*, 1986).

1.4.1 Mechanisms of variability

1.4.1.1 Errors during reverse transcription:

The molecular basis of variability in HIV-1 is partly due to the low fidelity of the reverse transcriptase enzyme. None of the three stages in retroviral replication, including reverse transcription, plus-strand DNA synthesis as well as transcription, are subjected to proof-reading (Ricchetti & Buc, 1990). It has been estimated that during this highly error-prone process, HIV-1 can mutate *in vitro* at a rate of about one nucleotide per genome per replication cycle. Therefore, a population of HIV-1 viruses will contain few if any identical genomes (Subbarao & Schochetman, 1996). The HIV-1 viral copies within a single individual (quasispecies) commonly vary by 1% - 6% in *env* (Hahn *et al.*, 1986, Goodenow *et al.*, 1989). In addition, retroviral genomes undergo insertions, deletions and frame shifts which contribute to genetic diversity (Dougherty & Temin, 1988, Pathak & Temin, 1990a, Pathak & Temin, 1990b). High rates of G-to-A transitions also occurs in the HIV-1 genome. This phenomenon, by which individual proviruses acquire several mutations in a single replication cycle, is called hypermutation (Vartanian *et al.*, 1991). G-to-A substitutions not only contribute to the high A content (34%) of the HIV-1 genome, but also to the genetic variability (Goodenow *et al.*, 1989).

1.4.1.2 Recombination:

Retroviruses are known to be highly recombinogenic. The enzyme's affinity for the two

RNA templates is low and is able to frequently switch templates during proviral DNA synthesis. Recombination results when a host cell is infected with two different viruses, either sequentially (super-infection) or simultaneously (co-infection). The error-prone reverse transcriptase jumps between the two RNA templates so that the newly synthesized DNA transcripts are a mosaic between that of the two parents. However, recombination can also occur when one RNA transcript from each virus is encapsidated into a single heterozygous virion. When such a virion subsequently infects a new host cell, the reverse transcriptase uses both strands as templates to generate a mosaic DNA transcript (Hu & Temin, 1990). Co-infection with highly divergent strains from the same subtype (Zhu *et al.*, 1995), from different HIV-1 subtypes within the major (M) group (Artenstein *et al.*, 1995, Van der Ryst, 1996, Xin *et al.*, 1995) or with members of group M and the highly divergent outlier (O) group (Heyndrickx *et al.*, 1996, Songok *et al.*, 1996) have been described.

Inter-subtype recombinant viruses are identified when different genes, or different regions within the same gene, group phylogenetically with different HIV-1 subtypes. One of the earliest characterized HIV-1 isolates from Africa was MAL which was found to cluster with subtype D in *env*, but with subtype A in *gag* (Myers *et al.*, 1992c). More than 10% of the HIV-1 strains that have been characterized by sequencing appear to be inter-subtype recombinants (Robertson *et al.*, 1995b). This figure is possibly an underestimate of the true frequency of recombinant viruses, both because of the excess representation of subtype B, and because the cross-over points of possible recombinant viruses do not fall within the gene being analysed (Robertson *et al.*, 1997). Recent studies indicated the frequency in Africa may be as high as 25% (Kampinga *et al.*, 1997).

Recombination occurs in geographic regions where multiple genetic subtypes co-circulate, such as Central Africa, South America, and Southeast Asia (Jain *et al.*, 1994, Louwagie *et al.*, 1995, Louwagie *et al.*, 1993, Lukashov *et al.*, 1995, McCutchan *et al.*, 1992a). A variety of recombinant viruses have been identified, including A/B, A/C, A/D, A/E, A/G, and B/F recombinants (Bobkov *et al.*, 1998, Carr *et al.*, 1998, Carr *et al.*, 1996, Leitner *et al.*, 1995b, Sabino *et al.*, 1994, Salminen *et al.*, 1997). It appears that certain subtypes recombine more often than others, as in the case of subtypes A and C while subtypes B and D show a more restricted pattern of genetic exchange (McCutchan *et al.*, 1996). Analysis of full-length genomes of subtypes A-H has revealed that all subtype E strains and the majority subtype G strains are

recombinants with subtype A (Leitner *et al.*, 1997). However, non-recombinant full-length genomes of HIV-1 subtype G have recently been described (Carr *et al.*, 1998). Recent analysis of a complete genome of a subtype I sequence, originally from Cyprus, revealed that it is multiply mosaic and composed of at least 3 subtypes (A, G, and I) (Gao *et al.*, 1998a). Other complex mosaic viruses comprising three different HIV-1 group M subtypes include the MAL strain which in addition to subtypes A and D also showed segments of subtype I (Robertson *et al.*, 1997). *Env* and *gag* fragments of subtype J genes have been isolated from Zairean immigrants in Sweden, but full-length sequences of this subtype are still unavailable (Leitner *et al.*, 1995a).

Recombination between HIV-1 and HIV-2 has not been described, although co-infection has been serological defined (Rubsamen-Waigmann *et al.*, 1994) and later confirmed by polymerase chain reaction (PCR) using nested primer pairs specific for either HIV-1 or HIV-2 (Grez *et al.*, 1994). The lack of recombination between these two lineages, which are only 40% homologous in th `e *env* genes (Guyader *et al.*, 1987) may be due to the high level of sequence divergence between them (Robertson *et al.*,). This suggests that in order to ensure a functional virus there may be a limit to genetic variation within HIV.

1.4.1.3 Selection by the immune system:

The frequent emergence of phenotypic variants in an individual over time is likely to be driven partly by the strong selective pressure imposed by the immune system (Coffin, 1992). HIV-1 mutates to escape neutralising antibodies (McKeating *et al.*, 1989) and cytotoxic T lymphocytes (CTL) (Phillips *et al.*, 1991). *In vitro* studies have shown that a single amino acid substitution in the sequence of the V3 epitope can lead to a highly significant reduction in the neutralizing activity of antibodies directed against this epitope (McKeating *et al.*, 1989). Similarly, variation in the amino acid sequence of known CTL epitopes have been shown to correlate significantly with reduced CTL specificity (Phillips *et al.*, 1991). This allows the virus to escape recognition by HIV-specific CTL prevalent at that time.

Rates of selection pressures vary for different parts of the retroviral genome. The viral envelope is continuously exposed to immune pressure which explains why *env* shows more genetic variation compared to the rest of the genome (Montagnier & Clavel, 1994). This suggests

that the observed levels of genetic sequence variation does not simply reflect intrinsic mutation rates (Holmes *et al.*, 1992), but rather a combination of viral and host factors.

1.4.1.4 Selective pressures of prolonged chemotherapy:

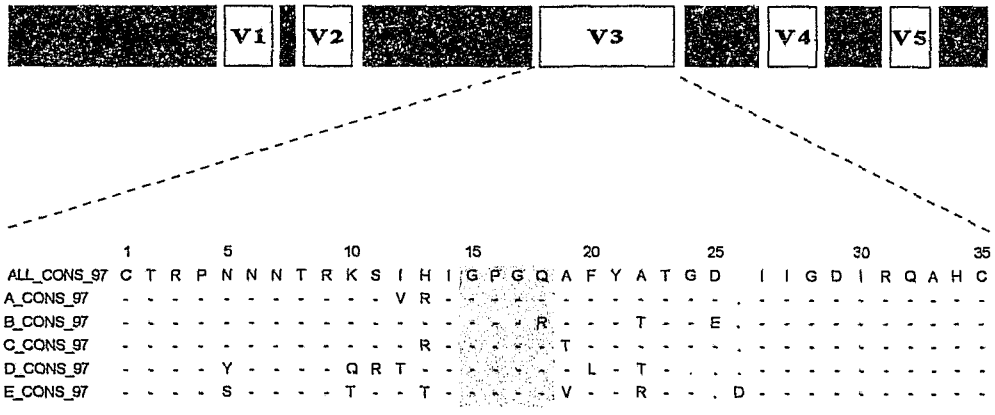
Drug-resistant mutants of HIV-1 have been shown to emerge under the selective pressure of prolonged chemotherapy. Resistance is characterized by specific mutations in the appropriate gene, for example mutations at codons 67, 70, 215, and 219 in the reverse transcriptase gene confer AZT resistance (Richman, 1992). There is strong evidence for the transmission of these drug-resistant strains (Imrie *et al.*, 1997), which not only contribute to the genetic pool of variants, but also complicate combination therapy in the recipient.

1.4.2 Biological effects of *env* gene diversity

The HIV-1 *env* gp120 protein contains five hyper variable domains (V1-V5) separated by six relatively conserved regions (Willey *et al.*, 1986) (Figure 1.3). *Env* is the most commonly studied gene for phylogenetic analysis because of its high degree of genetic variation. The third hyper variable region (V3 loop) has been mapped to a region of 32-35 amino acids formed by a disulphide bond between two conserved cysteines (Matsushita *et al.*, 1988). The tip of the loop has a highly conserved tetrapeptide motif, usually Gly-Pro-Gly-Gln (GPGQ (Figure 1.3), compared to the more variable immediate flanking regions (Foley & Korber, 1995). The V3 loop contains the principal neutralization domain and is the main target of antibody (Javaherian *et al.*, 1989) and T-cell-mediated responses (Takahashi *et al.*, 1992). It is therefore highly immunogenic and subjected to strong selective pressure. Diversity of the V3 loop sequence increases during HIV-1 infection from a relatively homogenous population during the acute phase, to a more heterogenous one at the onset of AIDS. This increased diversity is associated with an increase in antigenic variation which in turn has major implications for the gp120-$CD4^+$ interaction (Kuiken *et al.*, 1993).

The V3 loop is associated with several functions, such as infectivity (Freed *et al.*, 1992), viral fusion (Freed *et al.*, 1991), cytopathicity (Fouchier *et al.*, 1992) and cell tropism (Hwang *et al.*, 1991). It is these characteristics that help in defining the pathogenicity of the viral strain. *In vitro* studies indicate the predominance of macrophage tropic, slow replicating non-syncitium inducing (NSI) virus variants which use the CCR5 co-receptor (R5 viruses) during the early, asymptomatic stage of HIV-infection. The T-cell line tropic syncitium inducing (SI) variants,

which use the CXCR4 co-receptor (X4 viruses), emerge in about 50% of symptomatic patients (Tersmette *et al.*, 1989). The V3 domain functions during virus cell adsorption and fusion, after the initial binding between gp120 and the $CD4^+$ molecule on the surface of the host cell. Amino acid substitutions in the V3 loop cause changes in the confirmation of the epitope which has a direct impact on the tropism of HIV-1 (Freed *et al.*, 1991). Increased V3 loop charge, associated with the switch from macrophage to T cell tropism, indicates that a net positive charge is a determining factor in the SI capacity of a strain (Distler *et al.*, 1995). There is no significant association between biological phenotypic characteristics (slow/low and rapid/high) and the different HIV-1 genetic subtypes. This was evident from a study showing that HIV-1 isolates of diverse genetic subtypes derived from patients from different geographical regions, were quite heterogenous with respect to their biological phenotypes (Nkengasong *et al.*, 1995).



Redrawn from P.K. Dighe *et al.* (Dighe *et al.*, 1997) and R. Connor *et al.* (Connor *et al.*, 1997)

Figure 1.3 The HIV-1 *env* gp120 protein showing five hyper variable domains (V1-V5) separated by six relatively conserved regions (C1-C5). The V3 region alignment shows a consensus sequence generated for subtypes A-E. A dash (-) indicates concurrence with ALL consensus in the alignment, a period (.) indicates a deletion and cysteines (C) at the start and end of the V3 loop are shown. The tetrapeptide, GPGQ, is highly conserved in all subtypes except subtype B (GPGR).

1.4.3 Implications of genetic variability

1.4.3.1 Diagnostic tests:

The genetic variability of HIV-1 has major implications for the sensitivity and specificity of diagnostic tests. Previously used serological tests, based primarily on synthetic peptides and recombinant antigens of subtype B, failed to detect highly divergent strains such as HIV-1 subtype O (Schable *et al.*, 1994). This resulted in significant modifications to diagnostic tests to broaden their sensitivity towards current circulating subtypes. PCR-based diagnostic tests may also be compromised if primers are designed using regions of the genome that show genetic variability. This was the case with the first generation Roche Amplicor Monitor kit for quantitation of viral RNA. This test showed poor sensitivity for subtypes A, E and G, although, this has been corrected in a subsequent version by modifying the primers used. It is, therefore, important to take into account ongoing variation and the appearance of possible new subtypes, when designing diagnostic tests.

1.4.3.2 Drug resistance:

Genetic variation may affect the response of HIV-1 to chemotherapeutic agents. A low frequency of naturally occurring drug-resistant strains are present prior to therapy. The selection of these strains begins when drug therapy is initiated which then replace previously sensitive HIV strains. Resistance of non-B subtypes to azidothymidine (AZT) and didanosine (ddI) may have serious implications for effective chemotherapy in populations where non-B subtypes prevail. This may be relevant to the use of AZT in developing countries, such as Africa for preventing maternal-fetal transmission (Janssens *et al.*, 1997). Therefore, surveillance for these drug-resistant non-B mutants prior to the start of chemotherapy is essential.

1.4.3.3 Developing an HIV vaccine:

Candidate vaccines face the challenge of antigenic variation within the principle epitopes that elicit both humoral and cell-mediated immunity. Genetic variation results in the antigenic diversity of HIV-1 proteins, which in turn diminish the binding of neutralizing antibodies and the effectiveness of CTL. Some antibodies neutralize viruses, irrespective of HIV-1 subtypes, as a result of cross-reactive epitopes (Burton *et al.*, 1994), while others only show neutralization activity against viruses from the same subtype (Mascola *et al.*, 1994). Some CTL epitopes are conserved among several HIV-1 subtypes. Recent data has shown that subtype C-infected

individuals can recognize subtype B CTL epitopes indicating that cross-subtype CTL activity exists (Casement *et al.*, 1996) (Betts *et al.*, 1997). Thus, despite the extent of HIV-1 variation, the finding of cross-reactive antibody and CTL responses suggests that genetic variability may be less of a problem than what was initially thought. Knowing this, there is a recent trend towards the inclusion of conserved structural proteins, in addition to the more variant gp120/gp160, in vaccine development to elicit cross-clade immunity (Brodine *et al.*, 1997).

1.4.3.4 Transmissibility:

The hypothesis that HIV subtypes display differences in transmission efficiency as a result of functional differences, is controversial (Cohen, 1995). In Thailand the frequency of sexual transmission between men and women was found to be higher for HIV-1 subtype E than for B, the other major subtype present in intravenous drug users (Kunanusont *et al.*, 1995). It was found that subtype E viruses had an increased replication capacity in epithelial Langerhans cells (LC), present on the surface of the genital mucosa, possibly explaining the high rate of heterosexual transmission (Soto-Ramirez *et al.*, 1996). Studies counter-claiming this hypothesis have shown that HIV-1 subtypes do not have a preferred tropism for Langerhans cells *in vitro*, indicating that differences in transmissibility are probably subtype-independent (Dittmar *et al.*, 1997, Pope *et al.*, 1997). A further study in Brazil has shown that the incidence of subtype C is increasing together with higher rates of heterosexual transmission (Csillag, 1994). Other data lend very little support to this subtype-dependent transmission hypothesis as in the case of Haiti, which has experienced an explosive heterosexual epidemic of subtype B (Pape *et al.*, 1990). In vertical transmission there is selective transmission of certain maternal HIV-1 quasispecies (Wolinsky *et al.*, 1992), although no significant differences in transmission rates of the various HIV-1 subtypes were shown (Peckham & Gibb, 1995). Thus, there is evidence both for and against this hypothesis, but until further proof is available no conclusion can be drawn.

The variability of HIV-1 is very useful in establishing transmission patterns and epidemiological linkage between cases. Nucleotide sequence analysis showed strong similarities between the viral strains isolated from a dentist in Florida and those from five of his patients, proving HIV transmission occurred during dental procedures (Ou *et al.*, 1992).

1.4.3.5 Virulence and disease progression:

The possibility that HIV-1 subtypes cause different rates of disease progression was shown in a study which enrolled sex workers from Senegal. Women infected with a non-A subtype were 8 times more likely to develop AIDS than those infected with subtype A, the predominant subtype in Africa (Kanki *et al.*, 1999). However, a study conducted in London argues against genetic subtypes being an important determinant of HIV disease progression. Differences in progression could not be attributed to different viral subtypes, but rather to the lack of access to health care and exposure to opportunistic pathogens, particularly tuberculosis (TB), in Africa (Del Amo *et al.*, 1998). Differences in disease progression were initially apparent in the natural history of HIV-2. Although the clinical manifestation in HIV-2 infected individuals are quite similar to those of HIV-1, it is less pathogenic and disease progression appears to be slower (Pepin *et al.*, 1991). It still remains to be determined what, if any, impact different HIV-1 group M subtypes have on HIV-1 disease progression.

1.5 Classification of HIV

Two types of HIV have been characterized both genetically and antigenically, termed HIV-1 and HIV-2. Phylogenetic analysis of HIV-1 *env* sequences has revealed two groups within HIV-1: the major group (M), which is further divided into at least 10 genetically distinct subtypes (A-J), comprises the majority of HIV-1 isolates, and the outlier group (O), which is represented by a number of highly divergent HIV-1 strains. In addition, a new (N) subtype has recently been identified in Cameroon. Five genetic subtypes of HIV-2, designated A to E, have also been identified.

1.5.1 HIV-1

1.5.1.1 Group M:

Before 1992, HIV-1 strains were classified into two subgroups on the basis of their geographic origin, North American and African variants (Myers *et al.*, 1990). Since then, the V3 *env* gene has been used to classify globally prevalent viruses. The first five subtypes, identified as A, B, C, D and E (Myers *et al.*, 1992a), differed from each other by approximately 30% in their *env* gene and 14% in their *gag* gene (Janssens *et al.*, 1997). Another five subtypes were

described later and named F through to J (Janssens *et al.*, 1994, Kostrikis *et al.*, 1995, Leitner *et al.*, 1995a, Myers *et al.*, 1994). These ten subtypes belong to the main (M) group of HIV-1 and are approximately genetically equidistant from one another in the *env* gene (Foley & Korber, 1995). In a phylogenetic pattern these subtypes are referred to as a "star phylogeny", suggesting that the different HIV-1 subtypes might all have evolved and diverged synchronously from a common ancestor (Charneau *et al.*, 1994) (Figure 1.4). Group M subtypes differ from one another by an average nucleotide percentage distance of 27% (range 21- 31%) with an average intra-subtype genomic diversity of 11% within every subtype (Moore *et al.*, 1996). Phylogenetic analysis of HIV-1 nucleotide sequences (ZR59) recovered from a 1959 African plasma sample (Nahmias *et al.*, 1986), showed that this isolate is a representative of the common ancestor of HIV-1 subtypes B, D and F. This isolate may have existed in the 1940's or early 1950's, and perhaps all group M viruses may have evolved from this source (Zhu *et al.*, 1998). The oldest of the known hybrid viruses (Z321B), from a sample isolated in 1976, showed recombination between 3 (or possibly 4) subtypes. This supports that distinct subtypes must have been present earlier than the 1970s (Robertson *et al.*, 1997).

1.5.1.2 Group O:

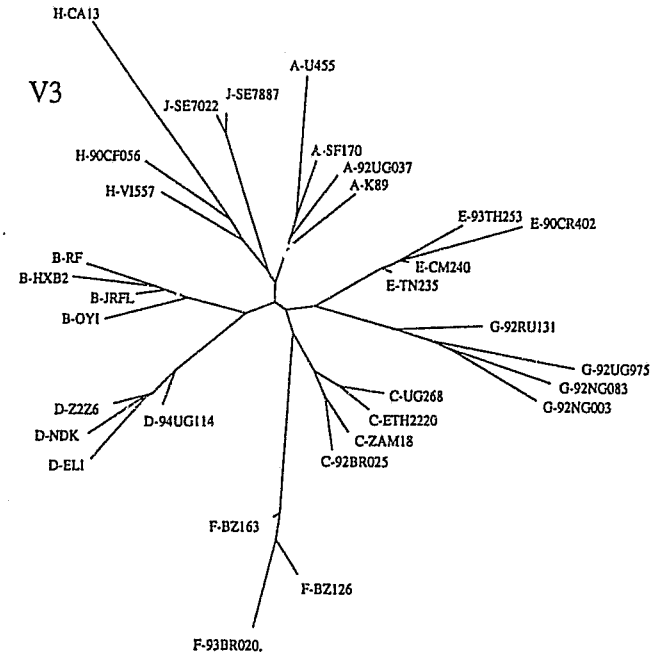
The discovery of highly divergent Cameroonian HIV-1 isolates in 1994, which share 50% amino-acid similarity in the *env* genes with group M viruses, led to the identification of group O (outlier) viruses (De Leys *et al.*, 1990, Gurtler *et al.*, 1994, Vanden Haesevelde *et al.*, 1994). These sequences are also highly divergent from HIV-2 isolates and are equidistant between HIV-2 and HIV-1 (Dondero *et al.*, 1994). Viruses within group O show as much variation as the 10 members of group M (Charneau *et al.*, 1994). However, they do not form subtype clusters in the same way as the group M sequences (Loussert-Ajaka *et al.*, 1995). Phylogenetically, group M and O viruses form a double-star-like configuration, i.e. two centres of evolutionary radiation, suggesting that HIV-1 arose from at least two separate zoonotic transmissions (Myers, 1994).

1.5.1.3 Group N:

In 1995, a highly divergent HIV-1 isolate, designated YBF30, was isolated from a Cameroonian woman. Phylogenetically it did not cluster with any of the known subtypes, including group O, and was as distinct from $SIV_{cpz\text{-}gab}$ as it was from HIV-1 group M. Therefore, YBF30 was considered as the prototype strain of a new HIV-1 group, designated "N (Simon *et*

al., 1998). Recently the N group was shown to be a mosaic of SIV_{cpz} and HIV-1 sequences. This highly divergent virus may represent an ancestral recombination event between HIV-1 and SIV possibly in a chimpanzee host (Gao *et al.*, 1999).

Evolutionary analysis of HIV-1 groups M, O and N isolates indicates that they probably originated as three separate transfers or zoonotic infections from chimpanzees or other primates into humans. All HIV-1 strains known to infect man, including all three groups, are closely related to just one SIV_{CPZ} strain, $SIV_{CPZ}US$, which is found in the chimpanzee subspecie *Pan troglodytes troglodyte* (Gao *et al.*, 1999).



Reprint from T. Leitner *et al.* (Leitner *et al.*, 1997)

Figure 1.4 Unrooted phylogenetic tree calculated with maximum likelihood methods using *env* V3 fragments of 9 HIV-1 subtypes (except I).

1.5.2 HIV-2

HIV-2 shares approximately 40% homology with HIV-1 and is sub-classified into five genetic subtypes, designated A - E. While HIV-1 has been widely disseminated throughout the world causing explosive epidemics in certain areas (UNAIDS/WHO, 1998), HIV-2 generally remains confined to western African countries (De Cock *et al.*, 1993). Lower quantitative HIV-2 viral load levels relative to HIV-1, at all stages of infection, may contribute to the decreased transmission efficiency of HIV-2 (De Cock *et al.*, 1993), as demonstrated in studies of perinatal and heterosexual transmission (Adjorlolo-Johnson *et al.*, 1994, Kanki *et al.*, 1994). Despite the differences in epidemiological spread and pathogenicity, HIV-1 and HIV-2 share modes of transmission as well as clinical manifestations. Recombinations within HIV-2 have also been documented; for example, the mosaic genome of 7312A involves HIV-2 subtypes A and B (Gao *et al.*, 1994).

1.5.3 Classification criteria for HIV-1 subtypes

In 1996, the proposed criteria for the identification of a new subtype were listed as: i) envelope sequences representing different subtypes must be approximately genetically equidistant from one another, ii) the phylogeny based on the *env* gene must be congruent with the *gag* gene, iii) two or more samples of the new sequence subtype are required, vi) clustering of at least two epidemiologically unrelated isolates which are separated from established subtypes, v) full sequence of at least a 1.5 kilobase region, and vi) the absence of any subsegment joining established subtypes. Less strict criteria may be used to classify a virus, and therefore, it is very important to state which region is used for classification. If only the *env* V3 region is used one might propose that the virus is *env* V3 subtype A, B, C, etc (Leitner, 1996). Subtype designation based on a gene or gene fragment may be correct, but recombination outside the region examined may have occurred (Foley & Korber, 1995). Since 1996, the criteria of a reference sequence has changed with a shift in emphasis to using full-length HIV genomic sequences (Leitner *et al.*, 1997).

Many sequences have been found not to group with known subtypes, either because they cluster differently using different phylogenetic analysis methods, or using different reference

subtype sequences. Insufficient data such as sequences that are too short or do not have enough representatives of a potential subtype, can also be the cause of sequences remaining unclassified. With the growing knowledge of inter-subtype recombinants, many of these unclassified sequences have been explained. The remaining unclassified sequences are included in the "U" or "unclassified" group (Foley & Korber, 1995).

1.6 Methods used for subtyping HIV

1.6.1 Genetic sequence analysis

The most accurate method of characterizing various HIV subtypes remains nucleic acid sequence analysis. The enzymatic chain-termination technique of Sanger involves synthesis, while the Maxam-Gilbert method involves chemical degradation of the original DNA (Sambrook *et al.*, 1989b). Although very different in principle, these two methods generate separate populations of radiolabeled (or digoxigenin-labelled) oligonucleotides that begin with a fixed point and terminate randomly at a fixed base or combination of bases. Because every base in the DNA template has an equal chance of being the terminating base, each population consists of a mixture of oliginucleotides with variable lengths determined by the location of a particular base along the length of the template. These populations of oligonucleotides are then resolved by loading them in adjacent lanes of a polyacrylamide gel under electrophoretic conditions that can discriminate between different lengths of individual DNA's by as little as one base. The order of nucleotides along the DNA template can be read directly from an autoradiographic image of the gel. The sequence template may be either proviral DNA, which requires the extraction of DNA from infected peripheral blood mononuclear cells (PBMC), or reverse transcribed RNA from HIV-1 virions in various body fluids (i.e. plasma, genital secretions). Indirect sequencing involves cloning of the PCR product prior to sequencing during which the sequencing population structure may be skewed by selection steps such as transformation, growing of clones, and clone picking. The fastest way of sequencing is by direct sequencing of the PCR product which suffers the least from methodological artifacts. DNA sequencing is usually performed on automated sequencers, enabling not only the simultaneous analysis of multiple HIV strains, but also hundreds of nucleotides per virus in a short period of time. It is important to remember that any HIV sample contains a population of related but genetically diverse variants, or quasispecies.

Therefore, it must be certain that a representative number of viral molecules are investigated to avoid erroneous conclusions, such as the detection of outlier sequences. While sequencing is the preferred method for subtyping, it is time consuming and requires expensive reagents and equipment.

1.6.2 PCR fingerprinting

A major challenge to larger-scale HIV characterization is the development of more efficient screening of isolates to determine which are strains of interest for subsequent sequencing. Several less cumbersome and less expensive techniques have been developed based on PCR. Sets of subtype-specific PCR primers have been used to amplify HIV isolates. PCR products are detected by hybridization with a ^{32}P-labelled probe which spans a fragment present in all products, and quantitated. Discrimination of HIV-1 genetic variants by PCR is based on template-primer mispairing. PCR reactions are scored "mispaired" when the product yield decreases 20-fold or more relative to reference isolates (McCutchan *et al.*, 1992a, McCutchan *et al.*, 1992b). This method of screening can only be useful in populations where no more than two subtypes are prevalent, as the various combinations of primer sets per subtype is very expensive.

1.6.3 Probe hybridization

Subtype-specific probe hybridization involves the binding of subtype-specific oligonucleotide probes, 3' or 5' end-labelled with digoxigenin (DIG)-dUTP or ^{32}P, to alkaline-denatured PCR *env* products of various HIV-1 isolates, already spotted onto a nylon membrane. In the case of DIG-labelled probes, detection is accomplished with the addition of alkaline phosphatase-conjugated anti-DIG antibody followed by phosphatase substrate and positive colour development (Luo *et al.*, 1998). Alternatively, an autoradiogragh is obtained after the membrane, hybridised with ^{32}P-labelled probe, is exposed to film (Ou *et al.*, 1993). This method shows adequate sensitivity and specificity in *env*-based studies where only a few subtypes circulate (Luo *et al.*, 1998, Subbarao *et al.*, 1996).

1.6.4 Restriction fragment length polymorphism

Restriction fragment length polymorphism (RFLP) is based on the observed correlation between the restriction maps of HIV-1 isolates with their phylogenetic classification. A ~300-bp

pol fragment (Pieniazek *et al.*, 1995) or 400-650-bp *gag* fragments (Van Harmelen *et al.*, 1999a) are amplified by nested PCR. Following endonuclease digestion, the products are loaded onto an acrylamide gel. Differences in electrophoretic mobility of the various digestion products result from the restriction site polymorphisms in the selected regions of the HIV-1 genome. Knowing the restriction patterns of various HIV-1 subtypes allows for quick recognition of the distinct subtypes. However, since HIV-1 is highly variable, single nucleotide mutations created either through substitution, deletions or insertions could destroy or generate a restriction site. Thus, RFLP can only be reliably applied for the rapid screening in an HIV-1 epidemic of limited diversity, such as in South Africa (Van Harmelen *et al.*, 1999a).

1.6.5 Single-strand conformational polymorphism

Single-strand conformational polymorphism (SSCP) has the advantage over RFLP analysis that it can detect DNA polymorphisms and point mutations at a variety of positions in short DNA fragments. It involves the amplification of a DNA fragment by PCR, melting of the PCR products, and analysis of the single strands by electrophoresis on non-denaturing polyacrylamide gels. The single strands adopt different conformations because of their nucleotide sequence differences and these can be distinguished by their different electrophoretic mobilities (Lin *et al.*, 1995, Orita *et al.*, 1989a). The technique has the potential to detect as little as one nucleotide change within sequences of 100-200 nucleotides (Orita *et al.*, 1989b). Therefore, SSCP is a useful tool for studying viral quasispecies (Lin *et al.*, 1998) rather than discriminating different subtypes. There are a number of possible applications of quasispecies analysis using SSCP: i) to monitor the emergence of drug-resistant variants in patients undergoing drug therapy, ii) to study the natural history of HIV-1 infection, and iii) to study the vertical transmission of variants in mother-infant pairs (Lin *et al.*, 1995).

1.6.6 Heteroduplex mobility assay

The heteroduplex mobility assay (HMA) provides a rapid and convenient method to subtype large numbers of HIV-1 isolates. This method is based on the migration differences of DNA heteroduplexes and homoduplexes in polyacrylamide gels (Delwart *et al.*, 1993). Heteroduplexes are formed by simply denaturing and reannealing partially complementary *env* or other PCR-amplified fragments from each of the reference subtypes with an unknown virus envelope gene fragment (Figure 1.5). Sufficient quantities of product, either 1.2 kb, 0.7 kb or 0.5

kb *env* fragments, are generated using nested PCR reactions. Structural distortions of the DNA double strand are caused by mismatched and unpaired nucleotides, resulting from base substitutions, mutations, insertions or deletions in aligned regions. The reduced mobility of DNA heteroduplexes in the gel is proportional to the degree of mismatches between the two strands. Thus, the most related sequences are the most rapidly migrating heteroduplex species. In addition, homoduplexes are formed between fragments from the same origin which migrate fastest. HMA is very useful in the initial screening of HIV-1 subtypes prevalent in countries where vaccine trials are in preparation (Buonaguro *et al.*, 1995, WHO, 1994). HMA is relatively simple and quick to perform, highly accurate and inexpensive relative to sequencing.

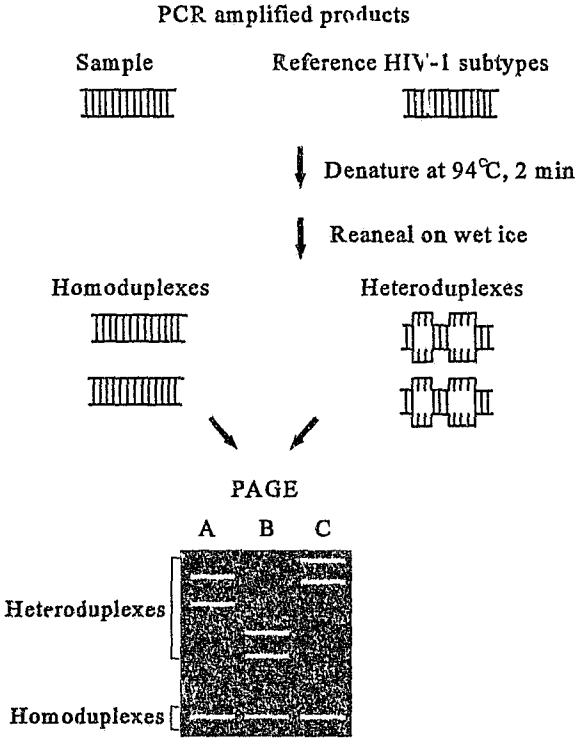


Figure 1.5 Schematic illustration of the heteroduplex mobility assay. A, B and C denote different HIV-1 reference subtypes. Heteroduplexes formed between the unknown sample and reference B migrate the fastest relative to other heteroduplexes thereby identifying this sample as a subtype B.

1.6.7 Combinatorial melting array

The combinatorial melting array (CoMA) is based on the principle that there is a direct correlation between the melting behaviour of long DNA heteroduplexes and the divergence among the single-stranded DNA strands. The higher the divergence the lower the T_m (T_m=temperature at which half of a double-stranded DNA fragment is denatured). Anti-sense, single-stranded DNAs were generated for each of the reference subtypes by asymmetrical PCR, with biotinylation of one of the primers facilitating capture to streptavidin-coated plates. Complementary sense-stranded DNAs are generated in the same manner for patients, but with DIG labelling, and added to the microtiter wells. After heat denaturation, heteroduplexes are allowed to form followed by a colorimetric assay using alkaline phosphatase-conjugated antibody against DIG. The subtype is determined by the comparative absorbance measurement with a particular reference subtype being higher than any other subtype by a factor of at least 2 (Kostrikis *et al.*, 1998).

1.6.8 Serotyping

All the above methods rely on PCR amplification of genetic material, and are therefore not suitable for large-scale surveillance of HIV subtypes. Phylogenetic analysis of the gp120 envelope gene of HIV-1 subtypes A-E has allowed the design of subtype-specific consensus sequences of the third variable region (V3). Synthetic V3 peptides derived from this region are employed in peptide-binding enzyme immunoassays (PEIA). These peptides are coated on to an ELISA microtiter plate after which the plate is incubated with seropositive HIV-1 sera. Colour development can be measured following the addition of peroxidase conjugate and substrate. V3 serotyping classifies HIV-1 based on the affinity of HIV-1 antibodies from infected individuals to bind V3 peptides derived from different HIV-1 *env* genetic subtypes (Cheingsong-Popov *et al.*, 1992). HIV-1 serotyping has shown both high sensitivity and specificity for large-scale identification of HIV-1 isolates in countries where the numbers of subtypes are limited, such as Thailand, Argentina, Ethiopia and South Africa (Cheingsong-Popov *et al.*, 1998b, Devito *et al.*, 1998, Engelbrecht *et al.*, 1999, Pau *et al.*, 1993, Sherefa *et al.*, 1997, Wasi *et al.*, 1995). However, it does not perform well in countries where a large variety of subtypes are circulating, eg. Central Africa (Hoelscher *et al.*, 1998, Murphy *et al.*, 1993). Although V3 serotypes B, C, and E correlate the closest with their genetic subtypes (Cheingsong-Popov *et al.*, 1998a, Cheingsong-Popov *et al.*, 1998b) subtype E sera shows the best concordance (Cheingsong-Popov *et al.*, 1998b). A

number of sera have been found to be non-typable. This can be explained by (1) patients who fail to produce antibodies to the V3 loop, (2) antibodies directed against a subtype variant which do not recognize the V3 peptide derived from the consensus subtype sequence, (3) antibodies directed against viral genotypes not yet discovered, and (4) distortion of the binding site when the synthetic peptides are mounted onto the solid phase of the microtiter plate (Cheingsong-Popov *et al.*, 1998a, Pau *et al.*, 1993, Sherefa *et al.*, 1997). In the case of antigenically similar subtypes (such as subtypes A and C, or subtypes B and D) cross-reactivity of A sera with C peptides, and D sera with B peptides, is observed. Serotyping in cohorts with these mixtures of subtypes may therefore result in decreased sensitivity and specificity (Cheingsong-Popov *et al.*, 1998a).

An initial serological screening by PEIA and subsequent HMA analysis of non-typable or dually reactive specimens allows for rapid and accurate screening of HIV-1 subtypes (Wasi *et al.*, 1995). If there is still uncertainty concerning the subtype identity of a HIV-1 positive specimen, genetic sequencing remains the ultimate tool for viral characterization.

1.7 The global distribution of HIV subtypes

HIV-1 subtypes have distinct geographical distributions (Figure 1.6) and the rate of spread of these subtypes is not uniform. In sub-Saharan Africa, the epidemic has spread most rapidly and extensively in East Africa compared to the relatively slower and more stable spread in Central Africa. There appears to be an association between the relatively slow epidemics and a higher variety of subtypes. This is in contrast to areas with explosive epidemics where only one (or at most three) subtype is present. It has been suggested that the relatively slow spread of the virus in a particular area causes more opportunities for new viruses to emerge. This results in the presence of a large variety of circulating subtypes as in the case of Central Africa where all HIV-1 subtypes, except subtype I, have been found (Janssens *et al.*, 1997).

1.7.1 HIV-1

1.7.1.1 Group M:

Subtype A is the most genetically divergent subtype (Leitner, 1996). The epidemics in all parts of sub-Saharan Africa, except southern Africa, appear to be dominated by subtype A

(Janssens *et al.*, 1997), but it has also been found in other parts of the world, including Europe (Alaeus *et al.*, 1997), Russia (Lukashov *et al.*, 1995), East Asia (Tsuchie *et al.*, 1995), and America (Brodine *et al.*, 1995).

The most intensely studied subtype is B, mainly because viruses of this subtype represent the vast majority circulating in Europe and the United States, but also because many original reference isolates are of this subtype (Janssens *et al.*, 1997). However, subtype B has also established itself in parts of southeast Asia (Weniger *et al.*, 1994), such as Thailand (Weniger *et al.*, 1991) and South America, and is the dominant subtype in Australia (Leitner, 1996). In Thailand, a distinct form of subtype B has been referred to as B' (Kalish *et al.*, 1995), which has been found spreading into China (Luo *et al.*, 1995), Malaysia and Japan (Cassol *et al.*, 1996). Subtype B is also found in Brazil where it has diverged from the consensus North American/European strains which may have important implications for the vaccine programs in Brazil (Morgado *et al.*, 1994). Surprisingly, subtype B strains in India showed closer homology to subtype B isolates from North America and Europe than to those from Thailand (Baskar *et al.*, 1994).

It has been estimated that subtype C is now the most prevalent virus world-wide. Subtype C viruses are associated with new, rapidly growing epidemics in southern Africa, India and China, but have also been reported to occur in other parts of the world, including Europe, Russia and Brazil (Leitner, 1996). In both India and South Africa, subtype C is the predominant subtype with heterosexual transmission as the major route of transmission (Janssens *et al.*, 1997, Lalvani & Shastri, 1996). Unlike South Africa, limited intra-patient nucleotide sequence divergence among Indian sequences suggests recent spread of HIV-1 from a common ancestor (Grez *et al.*, 1994), referred to as a clonal epidemic. The presence of subtype C strains in China suggest that these viruses were introduced via the travelling of infected people between these three countries. The low nucleotide divergence among subtype C's in China also suggests a recent introduction of this subtype. In Africa, subtype C is mainly found in the eastern and southern region. Subtype C is also the predominant strain in Ethiopia (Sonnerborg & Sherefa, 1997) with sequences that form a distinct sub-cluster different from the consensus C sequences from the international database. This suggests, as in India and China, an independent introduction of this subtype into Ethiopia, possibly from a common ancestor (Abebe *et al.*, 1997).

Viruses of subtype D, together with subtype A, are responsible for the predominant heterosexual epidemic in Eastern Africa (Janssens *et al.*, 1997). Subtypes D, E, G, H and J have been found in Central Africa (Janssens *et al.*, 1994, Leitner *et al.*, 1995a, Louwagie *et al.*, 1995, Louwagie *et al.*, 1993, Murphy *et al.*, 1993). Subtype E viruses are the major cause for HIV-1 infections among heterosexuals in Thailand (Janssens *et al.*, 1997) and have also entered the United States (Brodine *et al.*, 1995). Isolates from recently acquired subtype E HIV-1 infections in Thailand show very low genetic diversity indicating more recent introduction of this subtype compared to the more heterogeneous subtype E isolates in Central Africa (Murphy *et al.*, 1993). Despite the two isolates from Central Africa (Louwagie *et al.*, 1993), the majority of subtype F isolates has been identified in Brazil, South America (Morgado *et al.*, 1994).

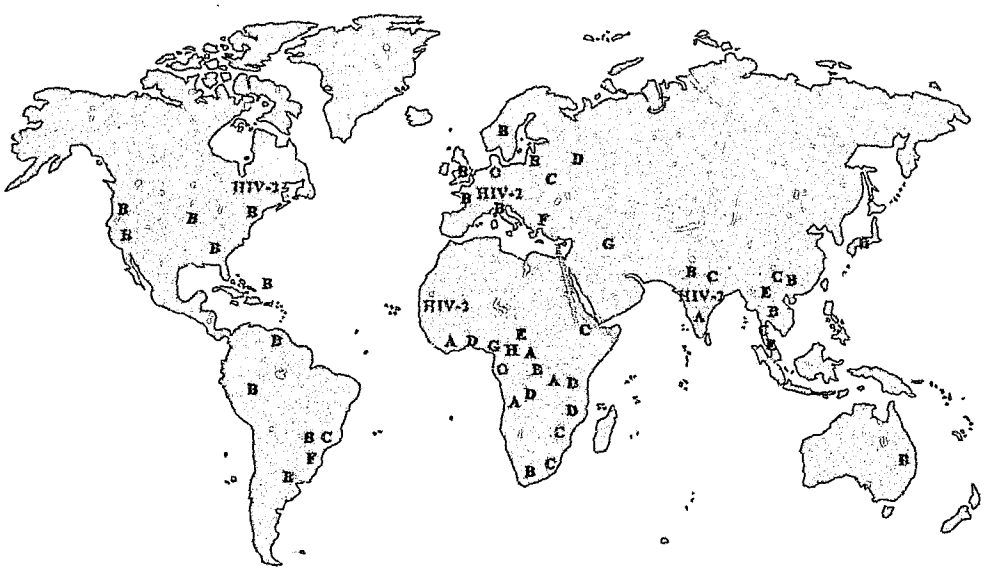
Thus far, subtype I viruses have only been found in patients from Cyprus (Kostrikis *et al.*, 1995). Although originally classified as I, this subtype was later identified as a triple recombinant with subtypes A, G and I (Gao *et al.*, 1998a).

1.7.1.2 Group O:

Initially the distribution of HIV-1 group O infections appeared to be endemic to Cameroon and neighbouring countries such as Gabon and Nigeria (Nkengasong *et al.*, 1993) accounting for less than 10% of HIV-1 cases (Zekeng *et al.*, 1994). Although HIV-1 group O infections have recently spread to West African countries such as Senegal, Niger and Togo, the prevalence remains very low (Peeters *et al.*, 1996). A survey conducted in France between December 1990 and March 1994 resulted in the identification of a group O HIV-1 strain in a French woman who had no links to Africa (Loussert-Ajaka *et al.*, 1995).

1.7.2 HIV-2

HIV-2 is endemic to West Africa and its appearance in other areas generally reflects epidemiological links to West Africa. Sporadic cases of HIV-2 infections have been identified throughout Europe, United States and South Africa (De Cock & Brun-Vezinet, 1989, Lyons *et al.*, 1988), with the greatest numbers of cases reported in Portugal, France, Germany and India (De Cock & Brun-Vezinet, 1989)(Horsburgh Jr & Holmberg, 1988)(Grez *et al.*, 1994, Kanki, 1991, Markovitz, 1993). India is the first country outside Africa where an HIV-2 epidemic occurred in the presence of an existing HIV-1 epidemic (Dietrich *et al.*, 1995).



Redrawn from D.J.Hu *et al.* (Hu *et al.*, 1996)

Figure 1.6 Geographic distribution of HIV-1 group M subtypes (A-I), group O and HIV-2.

1.8 South African HIV epidemic

In South Africa, HIV first spread during the early 1980s among the homosexual community after the virus was introduced possibly through contacts with gay communities in the United States and Europe (Ras *et al.*, 1983). The predominant subtype associated with homosexual transmission in both the United States and South Africa is subtype B, which supports the suggestion that this subtype was introduced to South Africa through the gay population (Van Harmelen *et al.*, 1997, Williamson *et al.*, 1995). During the late 1980s, the focus of the epidemic shifted to the heterosexual population after an alarming increase in the number of AIDS cases and HIV-1 positive people in the urban black community (Schoub *et al.*, 1990, Sher, 1989). Since the early 1990s the HIV-1 seroprevalence has increased exponentially in the heterosexual population (Department of Health, 1999). It is suggested that the virus has entered from countries north of South Africa through trucking routes which end in Durban. While there are many other factors

for this still rapidly rising epidemic in South Africa, among them migrancy and the high prevalence of sexually transmitted diseases (STD), are certainly the most important (Williams & Campbell, 1998).

The first South African strain to be sequenced, termed Nof (Myers *et al.*, 1992b), clustered with HIV-1 isolates from India (Dietrich *et al.*, 1993) indicating possible spread of HIV-1 from Central Africa to India via South Africa through the medium of merchant ships that call at South African ports (Becker *et al.*, 1995). Subtypes A, B, C and D have been found in South Africa with subtype C viruses as the predominant subtype, transmitted mainly heterosexually, followed by subtype B viruses mainly among homosexually-infected individuals (Williamson *et al.*, 1995, Van Harmelen *et al.*, 1997, Engelbrecht *et al.*, 1995, Van Harmelen *et al.*, 1999b). An average intra-subtype C divergence of 14% suggested this subtype has evolved mainly through multiple heterosexual introductions (Bredell *et al.*, 1998, Van Harmelen *et al.*, 1999b, Van Harmelen *et al.*, 1997). A few subtype D infections were also reported in this group during the mid-1980's (Becker *et al.*, 1995, Engelbrecht *et al.*, 1995, Williamson *et al.*, 1995, Van Harmelen *et al*, 1997) and more recently two subtype A infections have been documented (Bredell *et al.*, 1999, Van Harmelen *et al.*, 1999b).

1.9 Aims of this thesis:

Previous studies have indicated the prevalence of HIV-1 subtypes A, C and D in sub-Saharan African countries (Janssens *et al.*, 1997) with subtypes A, B, C and D being found in Cape Town, South Africa (Becker *et al.*, 1995, Engelbrecht *et al.*, 1995, Williamson *et al.*, 1995, Van Harmelen *et al.*, 1997, Van Harmelen *et al.*, 1999b). Similar studies to determine the circulating subtypes have not been done elsewhere in South Africa.

The objectives of this study are therefore:

1. to determine the HIV-1 genetic subtypes circulating in Gauteng, South Africa in different risk groups,
2. to determine whether inter-subtype recombinant viruses are present in South Africa,
3. to determine HIV-1 subtypes and patterns of transmission among migrant workers in South Africa.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Viral DNA lysates

Whole blood, collected in EDTA, was spun at 1,000 rpm for 10 min. Plasma was stored in volumes of 300 μl at -70°C. The cell fraction was resuspended in phosphate-buffered saline (PBS) and layered onto Ficoll-Hypaque, followed by density centrifugation at 2,000 rpm for 30 min. The white interface, representing PBMC, was resuspended in PBS and concentrated at 1,000 rpm for 10 min. Contaminating red blood cells were lysed using ammonium chloride (pH7), at room temperature for 5-7 min followed by 3 washes in PBS (1,000 rpm for 10 min) to remove contaminating platelets. PBMC were counted and stored as pellets of 10^6 cells at -70°C. To each pellet, 50 μl of lysis buffer containing 1.5 mg/ml proteinase K was added, and incubated at 56°C overnight. Proteinase K was heat inactivated at 98°C for 15 min prior to PCR.

2.2 Viral RNA extraction

In cases where only plasma samples were available, viral RNA was extracted and used as the template for reverse transcriptase-PCR (RT-PCR). Viral RNA was extracted from either fresh or frozen plasma using the QIAamp® Viral RNA isolation kit. Briefly, 560 μl lysis buffer was added to 140 μl plasma and incubated at room temperature for 10 min. This ensured isolation of intact viral RNA as well as the inactivation of RNases. After adding 560 μl 100% ethanol, the mixture was loaded onto the QIAamp spin column. Salt and pH conditions ensure that RNA binds to the silica-gel column. Proteins and other contaminants are washed through the column using two different wash buffers. RNA was eluted in a RNase-free buffer, ready to use for RT-PCR.

2.3 Reverse transcriptase-PCR

Single-stranded viral RNA was transcribed into complementary DNA (cDNA) using the

same primers as for the PCR. This was done by combining the RT-step and the first round PCR using primers ED5/ED12 or MSF12/SK431 (Table 2.1). The *env* RT-PCR mixture (50 μl) consisted of 10 pmol of each primer, 160 μM dNTPs, 1.4 mM $MgCl_2$, 10U RNase inhibitor, 5U Reverse Transcriptase, 0.625U Super-Therm DNA polymerase with recommended 10X buffer, and 5 μl isolated viral RNA. The *gag* RT-PCR mixture was identical, except for 200 μM dNTPs and 1.5 mM $MgCl_2$. Samples were placed in a GeneAmp® PCR System 2400/9600 thermocycler at 42°C for one hour, followed by identical cycling conditions to the first round *env* or *gag* PCR.

2.4 PCR amplification

2.4.1 *env* gene

A ~700-bp PCR product, spanning the gp120 V3 - V5 region of the HIV-1 envelope gene, was amplified from each sample by nested PCR. The PCR mixture (50 μl for first round, 100 μl for second round) consisted of 10 μl of each primer, 80 μM dNTPs, 1.4 mM $MgCl_2$, 0.625U Super-Therm DNA polymerase with recommended 10X buffer (without $MgCl_2$) and 5 μl lysed PBMC or 10 μl diluted (1:10) reference plasmid. The second round reactions were initiated with 5 μl of the first round reaction as template. The following cycling conditions for both first and second round reactions were used: 3 cycles of 94°C for 90 sec, 55°C for 45 sec, and 72°C for 90 sec; 31 cycles of 94°C for 1 min, 55°C for 45 sec, and 72°C for 90 sec; 1 cycle of 94°C for 1 min, 55°C for 45 sec, and 72°C for 7 min. Primers ED5/ED12 were used in the first round and ES7/ES8 in the second round PCR reaction (Delwart *et al.*, 1993) (Table 2.1).

The same fragment was directly amplified from reference plasmids A_2 (IC144 from the Ivory Coast), B_1 (BR20 from Brazil), B_2 (TH14 from Thailand), B_3 (SF162 from the USA), C_1 (MA959 from Malawi), C_2 (ZM18 from Zambia), C_3 (IN868 from India), C_4 (BR25 from Brazil), D_1 (UG21 from Uganda) and E_2 (TH06 from Thailand) using primers ES7/ES8 in a single PCR reaction.

4.2 *gag* gene

A ~540-bp product, spanning ~100 base pairs in the LTR region and the rest in the p17 *gag* gene, was amplified by nested PCR. The PCR mixture (50 μl for first round, 100 μl for

second round) consisted of 10 μl of each primer, 200 μM dNTPs, 1.5 mM $MgCl_2$, 0.625U Super-Therm DNA polymerase with recommended 10X buffer and 5 μl of lysed PBMC or 10 μl diluted (1:10) reference plasmid. The second round reactions were initiated with 5 μl of the first round reaction as template. The following cycling conditions for both first and second round reactions were used: 1 cycle of 94°C for 2 min, 45°C for 2 min, and 72°C for 4 min; 24 cycles of 94°C for 45 sec, 45°C for 60 sec, and 72°C for 90 sec; 1 cycle of 94°C for 45 sec, 45°C for 60 sec, and 72°C for 7 min. Outer primers MSF12 (Salminen *et al.*, 1995) and SK431 (Lynch *et al.*, 1992), were used in the first round, and inner primers LTR236 and gag778 (Van Harmelen *et al.*, 1997) (Table 2.1) in the second round PCR reaction. The same fragment was amplified and cloned from HIV-1 isolate 92UG029 (*gag* A/*env* A), 92BR021B (B/B), 92BR025 (C/C), 92RW009A (C/A), and 92UG001 (D/D) as references plasmids.

Table 2.1 Oligonucleotide primers used in PCR and sequencing reactions.

Primers	5'-Sequence-3'	Position*
HIV-1:		
ED5 forward	ATGGGATCAAAGCCTAAAGCCATGTG	6556-6581
ED12 reverse	AGTGCTTCCTGCTGCTCCCAAGAACCCAAG	7822-7792
ES7 forward	tgtaaaacgacggccagtCTGTTAAATGGCAGTCTAGC[#]	7001-7020
ES8 reverse	caggaaacagctatgaccCACTTCTCCAATTGTCCCTCA[#]	7667-7647
MSF12 forward	AAATCTCTAGCAGTGGCGCCCGAACAG	622-649
SK431 reverse	AGAGAACCAAGGGGAAGTGACATAGCAGG	1475-1501
LTR236 forward	CGCAGGACTCGGCTTGC	688-703
gag778 reverse	CACCTAGAACTTTAAATGCATGGG	1231-1254
PNDX forward	TGCACCATCCTCAGGAGG	7306-7323
ED33 reverse	TTACAGTAGAAAAATTCCCCTC	7359-7380
Human:		
PCO3 forward	CGATCACTTGTGTCAACACA	
PCO4 reverse	CCACTTGCACCTACTTCAAC	
Plasmid vectors:		
T7 Promoter Forward	TAATACGACTCACTATAGGGCGA	
Cy5-M13 Reverse	Cy5-d[CAGGAAACAGCTATGAC]	
21 M13 Forward	TGTAAAACGACGGCCAGT	
SP6 Promoter Reverse	AGTATTCTATAGTGTCACCTAAAT	

* = corresponding to various positions of HIV-1-HXBR-genome.
[#] = lower case letters are complementary to the universal M13-primer (ES7) or the reverse M13-primer (ES8) which allow direct sequencing of the PCR product.

For both *env* and *gag* PCR, PBMC lysates or plasma from HIV-seronegative individuals, and cells or supernatant fluid of an H9/IIIB-infected T-cell line were amplified under identical experimental conditions as negative and positive controls respectively. Amplification reactions were performed in either a GeneAmp® PCR System 2400 or 9600 or a Biometra TRIO-Thermoblock TB-1 cycler. Amplified PCR products, mixed with 5 μl bromophenol blue/xylenexyanol loading dye, were analysed on a 1% agarose gel (see Appendix A) containing 0.5 mg/ml ethidium bromide (EthBr) at 100V and visualized by ultraviolet (UV) transillumination.

When proviral DNA could not be amplified, the presence of intact DNA was assessed by PCR using primers PO3 and PO4 (Table 2.1) specific for human β-globin DNA and the following cycling conditions: 1 hold temperature of 94°C for 2 min; 1 cycle of 94°C for 90 sec, 50°C for 45 sec, and 72°C for 45 sec; 38 cycles of 94°C for 1 min, 55°C for 45 sec, and 72°C for 45 sec, and 1 cycle of 94°C for 1 min, 55°C for 45 sec, and 72°C for 10 min. The 50 μl PCR mixture consisted of 10 μl of each primer, 80 μM dNTPs, 1.4 mM $MgCl_2$, 0.625U Super-Therm DNA polymerase with recommended 10X buffer (without $MgCl_2$) and 5 μl lysed PBMC. Unsuccessful amplification of β-globin products resulted in viral RNA being used as an alternative template for PCR. If PCR product yields were too low an increased input of template (10 μl instead of 5 μl) was used.

2.5 Heteroduplex mobility assays (HMA)

2.5.1 *env* HMA

Env HMA was performed by mixing 7.5 μl ES7/ES8 PCR product from each specimen with 2.5 μl of the corresponding PCR fragment of each subtype reference plasmid and 1.1 μl of 10X heteroduplex annealing buffer (see Appendix A) in separate wells of a 96-well polycarbonate microtiter plate. Initially references A_2, B_1, C_2, D_1 and E_2 were used, but additional references B_2, B_3, C_1, C_3 and C_4 were added in cases of indeterminate HMA profiles and to confirm discrepant *env* and *gag* HMA results. DNA heteroduplexes were formed by thermal denaturation at 94°C for 2 min in a HYBAID Touch Down PCR system and reannealed by rapid cooling on wet ice. Heteroduplex reactions were mixed with 5 μl loading dye prior to loading on a 5% polyacrylamide gel (160 x 160 x 1.5 mm) (see Appendix A) followed by electrophoresis in 1X

TBE buffer (Delwart *et al.*, 1993) with an altered voltage of 200V for 3.5 hours. Hetero- and homoduplexes were visualized by UV transillumination after ~10 min EthBr staining (0.5 mg/ml in 1X TBE).

For each specimen a control lane containing 10 μl of PCR product alone, in the absence of reference plasmids, was run in parallel. Heteroduplexes formed in this lane were the result of intra-sample quasispecies diversity providing the background pattern to which deliberately formed heteroduplexes were compared.

2.5.2 *gag* HMA

For the *gag* HMA, 7.5 μl of LTR236/gag778 PCR product of each specimen was mixed with 2.5 μl of the corresponding PCR fragment from each of the *gag* reference plasmids A-D, and 1.1 μl of 10X annealing buffer. The rest of the *gag* HMA procedure is the same as for *env* HMA. The *gag* HMA was developed in collaboration with Gillian Hunt and Dr Caroline Tiemessen at the National Institute for Virology.

2.6 Cloning of PCR products

2.6.1 Inactivation of DNA polymerase activity

In order to inactivate DNA polymerase present in the PCR mixtures, the 700-bp ES7/ES8 PCR products were extracted with 1 volume chloroform:isoamylalcohol (24:1) (CIAA). Briefly, 1 volume CIAA was added to the PCR mixture, vortexed for 1 min and centrifuged at 14,000 rpm for 1 min. Two microliters of the aqueous (top) phase were added to the ligation reaction when using the pMOS*Blue* T-vector system.

2.6.2 Ligation and transformation

For each PCR product, a ligation reaction (10 μl) was set up according to the manufacturer's protocol. Briefly, 1 μl 10X ligase buffer, 0.5 μl dithiothreitol (DTT) (100 mM), 0.5 μl ATP (10 mM), 1 μl pMOS*Blue* T-vector (50 ng/μl), 0.5 μl T4 DNA ligase (2-3 weiss units), 4.5 μl nuclease free H_2O and 2 μl purified PCR product were mixed and incubated overnight at 16°C in a thermocycler. MOSBlue competent cells (20 μl/transformation reaction) were thawed on ice. One microliter of the ligation mix was added directly to the cells in a 1.5 ml

microcentrifuge tube, gently mixed and left on ice for 30 min. The mixture was heat shocked for 40 sec in a 42°C water bath, followed by another 2 min on ice. SOC medium (80 μl/transformation reaction) was added and incubated in a 37°C shaker (250 rpm) for 1 hour. 5-Bromo-4-chloro-3-indolyl-β-D galactoside (X-gal)/isopropylthio-β-D-galactoside (IPTG) Luria broth (LB) agar plates were prepared by spreading 35 μl 50 mg/ml X-gal and 20 μl 100 mM IPTG onto LB agar plates (see Appendix A) containing 50 μg/ml ampicillin and 15 μg/ml tetracycline. Fifty microliters of each transformation mixture was spread onto X-gal/IPTG LB agar plates and incubated overnight at 37°C (see Appendix A).

Due to discontinuance of the pMOS*Blue* vector kit during this work, the pGEM®-T vector kit was used to clone some of the PCR products. Briefly, 1 μl T4 ligase 10X buffer, 1 μl pGEM®-T vector (50 ng), 1 μl T4 DNA ligase, 5 μl nuclease free H_2O and 2 μl unpurified PCR product were incubated overnight at 4°C. The transformation protocol was almost identical to the pMOS*Blue* T-vector system, except for a few changes: 50 μl JM109 high efficiency competent cells were added to 2 μl ligation reaction followed by 20 min incubation on ice. Cells were heat shocked for 50 sec at 42°C and 950 μl SOC medium was added to each transformation mixture followed by 1.5 hr incubation at 37°C. One hundred microliters of transformation mixture was plated onto X-gal/IPTG LB agar plates.

Both cloning systems provided positive control inserts and plasmids to test the efficiency of ligation and transformation reactions, respectively. Successful cloning of an insert using both vectors interrupted the coding sequence of β-galactosidase, which contains the multiple cloning site. Insertional inactivation allowed recombinant clones to be directly identified by blue/white screening on X-gal/IPTG indicator plates.

2.6.3 Identification of recombinant clones

Ten white colonies, representing recombinant clones, were picked and each inoculated in 600 μl LB media (see Appendix A) containing 50 μg/ml ampicillin (and 15 μg/ml tetracycline in the case of pMOS*Blue* T-vector). After incubating overnight in a 37°C shaker (250 rpm), 5 μl of each mini-culture was subjected to PCR using the appropriate primers (ES7/ES8 for *env* clones; LTR236/gag778 for *gag* clones) to identify clones with *env* and *gag* inserts. PCR positive clones were expanded by adding 100 μl of the mini-cultures to 30 ml LB with the appropriate antibiotics and incubated overnight in a 37°C shaker (250 rpm).

2.7 Isolation of recombinant plasmids

Bacterial cells were harvested by centrifugation (6,000 rpm for 15 min) from 30 ml cultures, resuspended and alkaline lysed (200 mM NaOH, 1% SDS) using a QIAGEN Plasmid isolation kit. Genomic DNA, proteins and cell debris were precipitated using potassium acetate prior to centrifugation at 12,000 rpm for 30 min. The remaining supernatant was applied to QIAGEN anion-exchange resin, which allowed the plasmid DNA to bind under appropriate low-salt and pH conditions. Impurities were removed by a medium-salt wash followed by the elution of plasmid DNA using a high-salt buffer. Plasmid DNA was concentrated and desalted by isopropanol precipitation. Purified plasmids were quantitated as follows: Plasmid concentration [μg/ml] = $OD_{260}/50$; where 1 OD_{260} = 50 μg/ml double-stranded DNA. Plasmid purity was determined by: OD_{260}/OD_{280}. A ratio of 1.8 indicates pure plasmid DNA (Sambrook *et al.*, 1989a).

2.8 Sequencing of *env* and *gag* inserts

Purified plasmid DNA served as the template for automated sequencing, whereas PCR products from cloned DNA were used as templates for manual sequencing. For each isolate, one positive clone was sequenced in both directions. Early specimens were sequenced manually by the dideoxy-chain termination method using a Sequenase PCR Product Sequencing kit and later specimens by automated sequencing. An ALFexpress™ DNA Sequencer and ABI PRISM 377 DNA Sequencer was used with the Cy5 AutoRead Sequencing kit and dRhodamine Terminator Cycle Sequencing kit, respectively.

2.8.1 Sequenase PCR Product Sequencing kit

Prior to sequencing, 2 μl of PCR product was treated with a combination of 1 μl exonuclease I (to remove residual single-stranded primers) and 1 μl shrimp alkaline phosphatase (to remove remaining dNTPs) at 37°C for 15 min. The exonuclease and phosphatase were inactivated by heating to 80°C for 15 min. The treated PCR product (4 μl) was denatured at 100°C for 3 min in the presence of 10 pmol primer in a total volume of 10 μl followed by quick cooling in wet ice for 5 min. This allow the primers to anneal to the single-stranded DNA. PCR primers ES7 and ES8 were used as forward and reverse sequencing primers, respectively. In addition, ED33 (Delwart *et al.*, 1993) was used as a reverse primer, while PNDX was designed to

complement the forward primer (Table 2.1). The following labeling mixture was prepared and incubated at room temperature for 5 min: 10 μl ice-cold annealed DNA mixture, 2 μl 5X buffer, 1 μl DTT, 2 μl 1:5 diluted labeling mix (dGTP, dCTP, dTTP), 0.5 μl ^{35}S dATP (5μCi) and 2 μl DNA polymerase. In addition to the deoxynucleotides present, 1 μl each of dATP and dTTP was added to provide sufficient quantities in cases where A-T rich regions were sequenced, i.e. *env* fragments. While the labeling reaction took place, four 0.5 ml microcentrifuge tubes, labeled G, A, T and C were filled with 2.5 μl of the appropriate termination mixture (ddGTP, ddATP, ddTTP or ddCTP) and pre-warmed at 50°C for 1 min. One micro liter magnesium (Mn^{2+}) buffer was then added to the labeling mix, and 3.5 μl of the total volume was transferred to each of the termination tubes. Mn^{2+} allows for sequences less than 20 nucleotides from the primer to be read. The tubes were incubated for 5 min at 50°C followed by the addition of 4 μl stop solution. All reactions were done in a Biometra TRIO-Thermoblock.

Samples were heated to 80°C for 3 min before loading 3 μl onto an 8% acrylamide gel (see Appendix A). Gels were run at 60W (50°C), 2500V and 60mA in 1X Glycerol Tolerant Gel Buffer. For better resolution at the top of the gel, 3 M NaAc was added after 1 hr to the bottom buffer chamber to a final concentration of 1.5 M. Gels were fixed in 15% methanol:5% acetic acid for 30 min, dried for 1 hour at 80°C onto chromatography paper using a vacuum gel dryer and exposed overnight to Hyperfilm™-βmax autoradiography. Film was developed manually in the dark using developer (~3 min) and fixer (~1 min).

2.8.2 Cy5 AutoRead Sequencing kit

DNA denaturation. Purified plasmid DNA (10 μg), diluted in 32 μl H_2O, was denatured in the presence of 8 μl 2 M NaOH at room temperature for 10 min. DNA was precipitated with 7 μl 3 M NaAc, 4 μl H_2O and 120 μl 100% ethanol at -70°C for 15 min and pelleted for 15 min at 14,000 rpm in a microcentrifuge. The DNA pellet was then washed with 70% ethanol, spun for 10 min, vacuum dried and re-dissolved in 10 μl H_2O.

Annealing reaction. 2 μl 10 pmol/μl Cy5-M13 reverse primer (Table 2.1) was allowed to anneal to the denatured DNA (10 μl) in the presence of 2 μl annealing buffer at 65°C for 5 min, followed by two 10 min incubations at 37°C and room temperature, respectively.

Sequencing reaction. 1 μl Extension buffer containing DTT, 3 μl dimethyl sulfoxide (DMSO) and 2 μl diluted T7 DNA Polymerase (4U/μl) was added to the annealing reaction. Four tubes, labeled G, A, T and C containing 2.5 μl of the appropriate termination mixes (ddGTP,

ddATP, ddTTP or ddCTP in solution with dGTP, dCTP, dTTP and dATP) were pre-warmed at 37°C for at least 1 min. An aliquot (4.5 μl) of the annealing mixture was added to each of the 4 termination tubes and incubated for 5 in at 37°C. The reaction was stopped by adding 5 μl stop solution. All reactions were done in a Biometra TRIO-Thermo block.

Forward sequencing reactions were done using the Cy5-dATP labeling mix (Pharmacia Biotech, Uppsala, Sweden) in conjunction with the unlabeled T7 promoter forward primer (Table 2.1). The protocol is almost identical to the Cy5 AutoRead kit except for a few changes:

DNA denaturation. Template DNA was redissolved in 12 μl (not 10 μl).

Sequencing reaction. 1 μl Cy5-labeling mix (Cy5-dATP in solution with dGTP, dCTP, dTTP) and 2 μl diluted T7 DNA Polymerase (4U/μl) was added to the annealing reaction followed by 15°C incubation for 10 min. Four tubes (G, A, T and C) were each filled with 3 μl of the appropriate termination mix and pre-warmed at 37°C for 1 min. One micro liter extension buffer and 3.5 μl DMSO were added to the annealing reaction, and 5.4 μl was added to each of the 4 termination tubes, which were then incubated at 37°C for 5 min. The reaction was stopped by adding 6 μl stop solution.

Samples were heated to 85°C for 3 min before loading 6 μl onto an 5% Long Ranger acrylamide gel. Gels were run at 25W (55°C), 120V, 60 mA in 0.5X TBE buffer for 720 min using the ALFexpress™ DNA Sequencer.

2.8.3 dRhodamine Terminator Cycle Sequencing kit

This method is based on a cycle sequencing protocol. Briefly, 6 μl PCR product was mixed with 8 μl Terminator Ready Reaction Mix and 3.2 μl primer (3.2 pmol) in a total volume of 20 μl. Initially, primers ES7/ES8 and LTR236/gag778 were used, but in cases of poor primer binding, primers -21M13/SP6 were used (Table 2.1). The reaction was then subjected to the following cycling conditions in a GeneAmp® PCR System 2400 or 9600; 25 cycles of 96°C for 10 sec, 50°C for 5 sec and 60°C for 4 min. Hold at 4°C.

Excess dye terminators were removed from sequencing reactions by using CENTRI-SEP columns (Princeton Separations Inc., Adelphia, New Jersey). Columns were hydrated with RNase-free H_2O for at least 2 hours prior to using them. Excess H_2O was drained by gravity flow followed by a 3,000 rpm spin for 2 min in a microcenrifuge. All 20 μl of each completed

sequencing reaction was loaded onto the hydrated gel of a column and spun at 3,000 rpm for 2 min. Purified samples were collected and vacuum dried. Samples were resuspended in 6 μl (5 μl formamide and 1 μl Blue dextran with EDTA), heated at 90°C for 3.5 min prior to loading 2 μl onto a 4% acrylamide gel using the ABI PRISM 377 DNA Sequencer. Automated sequencing using this system was done by our Sequencing Core facility.

2.9 Nucleic acid analysis

Env and *gag* sequences were aligned with reference sequences representing various subtypes and $SIV_{cpz\text{-}Gab}$ as an outlier (Forum, 1999), using Clustal W (Thompson *et al.*, 1994). Refinement of the alignment was performed by hand editing and insertions and deletions within the alignment were excluded from all comparisons. Phylogenetic trees were constructed using both distance matrix and maximum parsimony approaches.

Pairwise distance matrixes were generated using the Kimura two-parameter model (Kimura, 1980) and Clustal W. Neighbour joining trees were generated using either TREECON for Windows 1.0 (Van de Peer, 1994) or PHYLIP 3.5c (Felsenstein, 1993). The reliability of the branches using TREECON was assessed by 500 repetitions of bootstrap analysis, which involved resampling the nucleic acid sequences whereby each column of the sequence alignment was selected using a (pseudo) random number generator. Bootstrap analysis done in PHYLIP used programs such as SEQBOOT (to generate 100 reshuffled sequences), DNADIST (maximum-likelihood model), NEIGHBOR, and CONSENSE from the PHYLIP package. Maximum parsimony trees were constructed using either PAUP 3.1.1 (Swofford, 1993) with gaps counted as "missing" and branch-swapping performed, or the program DNAPARS from PHYLIP.

CHAPTER THREE

HIV-1 SUBTYPES IN SOUTH AFRICA AND THE IDENTIFICATION OF INTER-SUBTYPE RECOMBINANTS

3.1 Introduction

In South Africa, HIV-1 was first recognized in the early 1980s in the homosexual population (Ras *et al.*, 1983), and by the late 1980s the epidemic expanded into the heterosexual population (Schoub *et al.*, 1990, Sher, 1989). Prevalence figures of HIV infection among women attending ante-natal clinics show that the epidemic in South Africa continues to rise resulting in one of the fastest growing epidemics in the world (Department of Health, 1999). Although subtypes A, B, C and D have been found in South Africa, the epidemic is dominated by subtype C viruses (Becker *et al.*, 1995, Bredell *et al.*, 1998, Engelbrecht *et al.*, 1995, Van Harmelen *et al.*, 1997, Van Harmelen *et al.*, 1999b, Williamson *et al.*, 1995).

The contribution of recombination to the genetic diversity of HIV-1 strains in the global epidemic has only recently been appreciated when it was estimated that more than 10% of the HIV-1 strains that have been sequenced appear to be inter-subtype recombinants (Robertson *et al.*, 1995b). A large number of HIV-1 recombinants have been identified from African countries, including Central African Republic, Zambia, Zaire, Nigeria, Rwanda, Uganda and Djibouti, which involve mainly subtype A with either subtypes C, D, E or G (Carr *et al.*, 1998, Choi *et al.*, 1997, Cornelissen *et al.*, 1996, Gao *et al.*, 1998b, Gao *et al.*, 1996, Kampinga *et al.*, 1997, Salminen *et al.*, 1997). All the inter-subtype recombinants have been found in geographic regions where multiple genetic subtypes are known to co-circulate, such as Central Africa, South America, and south-east Asia (Jain *et al.*, 1994, Janssens *et al.*, 1997, Louwagie *et al.*, 1995, Louwagie *et al.*, 1993, Lukashov *et al.*, 1995, McCutchan *et al.*, 1992a).

Previous studies have not revealed the existence of recombinant viruses in South Africa (Becker *et al.*, 1995, Van Harmelen *et al.*, 1999a, Van Harmelen *et al.*, 1997). The purpose of this study was to determine the HIV-1 subtypes present in different risk groups and if HIV-1

inter-subtype recombinant viruses circulate in South Africa. An HMA, based on the *gag* gene, was designed to complement the *env* HMA allowing a large number of samples to be screened.

3.2 Materials and Methods

3.2.1 Patients

Blood was collected in EDTA from 172 HIV-1 seropositive individuals between 1996 and 1998. Informed consent was obtained from all study subjects prior to collection of blood and the majority of patients were interviewed for demographic information including route and place of transmission (Appendix C). $CD4^+$ lymphocyte counts were determined in 117/172 (68%) of cases by flow cytometry. This study was approved by the Committee for Medical Research on Human Subjects of the University of the Witwatersrand (Appendix D).

3.2.2 PCR, HMA and phylogenetic analysis

Lysed PBMC were used as template for amplification of both the ~700-bp *env* and ~540-bp *gag* fragments, as described in Chapter 2. Amplified products were subtyped using a combination of *env* and *gag* HMA. In addition to *env* reference plasmids A_2, B_1, C_2, D_1, and E_2, reference plasmids B_2, B_3, C_1, C_3, and C_4 were also used in cases of indeterminate HMA profiles and to confirm discrepant *env* and *gag* HMA results (possible recombinants). The pMOS*Blue* T-vector was initially used and later the pGem®-T vector kit for cloning PCR products. Positive clones were identified by PCR using primers ES7/ES8 or LTR236/gag778 and sequenced in both directions by automated sequencing (ABI PRISM 377 DNA Sequencer) using the dRhodamine Terminator Cycle Sequencing kit. The V3-V5 *env* sequences were multiply aligned with 12 known *env* reference sequences belonging to subtypes A (U455 and SF1703), B (92BR020 and HXB2R), C (92BR025, D868, 93MW959 and ZAM18A), D (ELI and 92UG021) and E (92TH011 and 92TH006) (Forum, 1999), using Clustal W (Thompson *et al.*, 1994). Similarly, *gag* sequences were aligned with reference subtypes A (U455 and IBNG), B (HXB2R and SF2), C (ETH2220 and 92BR025) and D (ELI and Z2Z6). Two known HIV-1 recombinant strains, ZAM184 (A/C) and MAL (A/D) were included for comparison and $SIV_{cpz\text{-}Gab}$ was used as an outlier (Forum, 1999). Neighbour joining trees were constructed using PHYLIP (Felsenstein, 1993). Phylogenetic relationships were also analysed by the maximum-parsimony method with

repeated randomized input order, implemented by using the program DNAPARS from PHYLIP (Felsenstein, 1993).

The sequences were named by year of isolation, patient's identification number followed by the country code ZA for South Africa. Twenty-two HIV-1 sequences (16 *env* and 6 *gag*) have been submitted to GenBank.

3.3 Results

3.3.1 Patients

In order to monitor the spectrum of subtypes present in South Africa, 172 HIV-1 infected individuals representing different HIV risk groups were recruited for this study (Table 3.1). One hundred and sixty five patients were from hospitals and AIDS clinics in and around Johannesburg. Specifically, these groups included patients admitted with tuberculosis and other AIDS defining illnesses, individuals attending hemophiliac or ante-natal clinics, immigrant and mobile persons using urban clinics as well individuals under the care of private physicians. Seven commercial sex workers from KwaZulu/Natal were also included. There was a similar number of men (n=84) and women (n=88) in the cohort with ages ranging from 18 to 59 years. The majority of individuals were infected through heterosexual contact, followed by blood transfusion, homosexual transmission, intravenous drug use and occupational exposure. This cohort does not reflect the HIV infection rates across the different risk groups. Fifty two percent of individuals had $CD4^+$ T cell counts less than 200 cells/μl indicating that they had advanced disease and were thus likely to be have been infected for a long period of time.

3.3.2 *env* HMA subtypes

A 700-bp *env* PCR product (Figure 3.1) was amplified in 167 of 172 samples. PCR products were then used in an HMA with reference isolates from subtypes A-E (Figure 3.2). Samples from 3 commercial sex workers could not be subtyped due to the lack of clear heteroduplex formation (not shown). This resulted in 164 *env* subtypes, the majority of which were subtype C, followed by subtype B, with a few subtype A viruses (Table 3.2). Among heterosexually infected adults, 97% (104 of 107) were *env* subtype C. Subtype B isolates were

present in individuals infected through homosexual contact (93%) or through contaminated blood products (89%). In each of these 2 risk groups there was a single case of a subtype C transmission. Similarly subtype B was found among a heterosexually-infected individual. Two cases of occupational exposure were also examined. A subtype B virus was found in a laboratory worker and a subtype C virus in a physician. Intravenous drug use is still relatively rare in South Africa and of the 2 cases examined, one was subtype B and one was subtype C. Four *env* subtype A viruses were identified: 1 from a Zambian woman (95ZAM823ZA) who was later found to harbour an C/A recombinant (see below); 1 from a man from Mali whose route and place of transmission is unknown, 1 from a Polish man infected heterosexually in South Africa and 1 was a recipient of contaminated Factor IX.

Table 3.1 Patients' gender and $CD4^+$ range by different cohorts.

Cohorts	Patients			$CD4^+$ count
	n	Male	Female	
Hospitalized patients:				
Gold Fields West hospital	6	6		22 - 663
Sizwe hospital	48*	30	18	0 - 938
AIDS clinics - outpatients:				
Johannesburg hospital[$]	23	15	8	11 - 923
Hillbrow hopital	16	13	3	110 - 570
Private patients[$]	23*	19	4	29 - 509
Ante-natal clinic at Baragwanath hospital:	30*		30	ND
Haemophiliac clinic at Johannesburg hospital[#]:	19	1	18	5 - 974
Commercial Sex workers (Durban):	7		7	ND
Total	172	84	88	

ND = not done

* = all 30 mothers, 5 hospitalized patients and 1 private patient were omitted from *gag* PCR

[#] = patients infected through blood transfusion

[$] = some homosexually infected parients

3.3.3 Screening for recombinant viruses

A *gag* HMA was developed to complement the *env* HMA to facilitate HIV subtype determination in 2 regions of the genome and to identify inter-subtype recombinant viruses. Isolates representing subtypes A-D were used to generate 540-bp PCR *gag* gene fragments for use as references. The *gag* HMA was performed essentially the same as for the *env* HMA. There was, however, less discrimination between *gag* subtypes with heteroduplexes migrating closer to the homoduplexes compared to the *env* HMA. Nevertheless, subtypes A and C were easily identified using the *gag* HMA as these heteroduplexes migrated ahead of other subtypes. Subtype B was more difficult to distinguish as heteroduplexes migrated only slightly ahead or equivalently with subtype D and were classified as B or D (Figure 3.3 and Table 3.2).

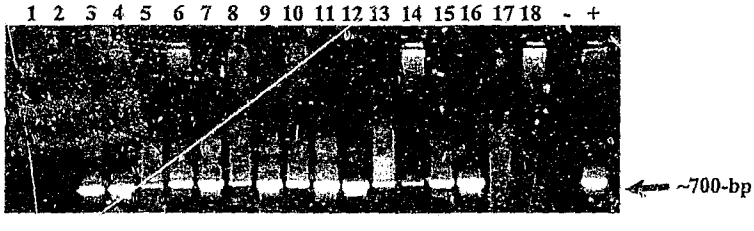


Figure 3.1 Agarose gel showing amplified ES7/8 *env* products. Lanes 1-18 represent PBMC from 18 HIV-1 infected patients, lane (-) an HIV-1 seronegative PBMC sample used as a negative control and lane (+) H9/IIIB-infected T-cells used as a positive control.

Of the 172 samples, 136 were used to amplify the *gag* gene (30 samples from the ante-natal clinic, 5 samples from patients infected heterosexually and 1 sample from an individual infected through occupational exposure, were not subjected to *gag* PCR). A 520-bp *gag* PCR product (Figure 3.4) was amplified in only 119 of the 136 samples. Two *gag* amplified fragments could not be subtyped by HMA, thus giving 117 *gag* subtypes (Table 3.2).

There was concordance between the subtypes determined by *env* and *gag* HMA in 113 samples. However, four samples showed discordant HMA results suggesting the presence of recombinant viruses (Table 3.3). Two of the patients, 96MOKZA and 97PP110ZA, were *env* subtype C and *gag* subtype A. 96MOKZA is a male from Congo who attended a private clinic in Johannesburg, while 97PP110ZA is a South African male who travelled extensively in Africa and Europe. The third individual (95ZAM823ZA) is a Zambian women infected heterosexually with a subtype C/A. Her husband harboured a subtype C/C virus (not shown). The last individual, CSW1, a commercial sex worker from KwaZulu/Natal, was infected with B/C mosaic virus.

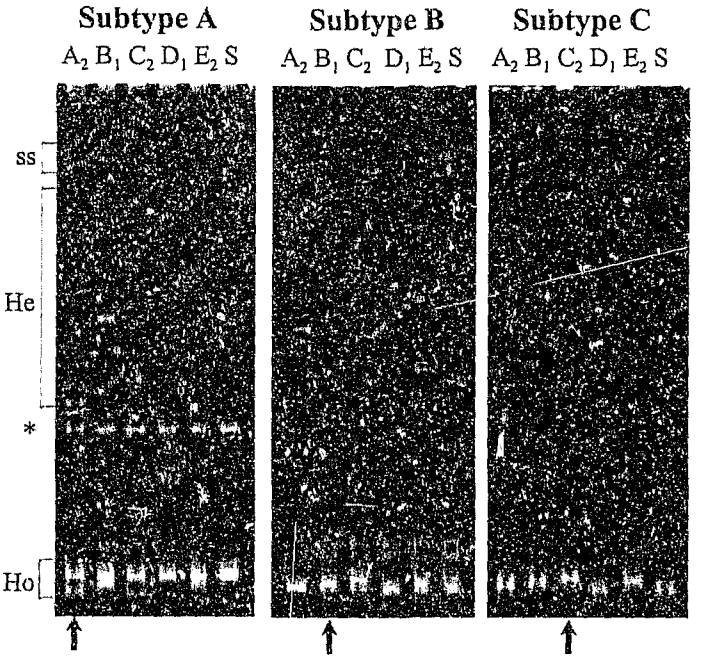


Figure 3.2 *env* HMA profiles of 3 South African samples representing subtypes A, B and C. Samples were run against reference plasmids from subtypes A_2 (IC144), B_1 (BR20), C_2 (ZM18), D_1 (UG21) and E_2 (TH06). Lane S is the sample in the absence of any reference plasmids. * This band, present in all lanes, represents non-specific bands or intra-patient quasispecies and is therefore ignored. The arrows indicate the lane with the fastest migrating heteroduplexes (He) which delineates the subtype. Homoduplexes (Ho) and single-stranded DNA (ss) are indicated.

Table 3.2 Genetic subtypes as determined by HMA.

Route of transmission	*env* subtypes				*gag* subtypes			
	n	A	B	C	*n*	A	B or D	C
Heterosexual	*107*	2	1	104	*68*		2	*66*
Blood products	*19*	1	17	1	*19*	1	17	1
Homosexual	*15*		14	1	*13*		12	1
IVDU	*2*		1	1	*1*		1	
Occupational exposure	*2*		1	1	*1*			1
Unknown	*19*	1	5	13	*15*	2	3	10
TOTAL (%)	*164*	4 (2)	39 (24)	121 (74)	*117*	3 (3)	35 (30)	79 (68)

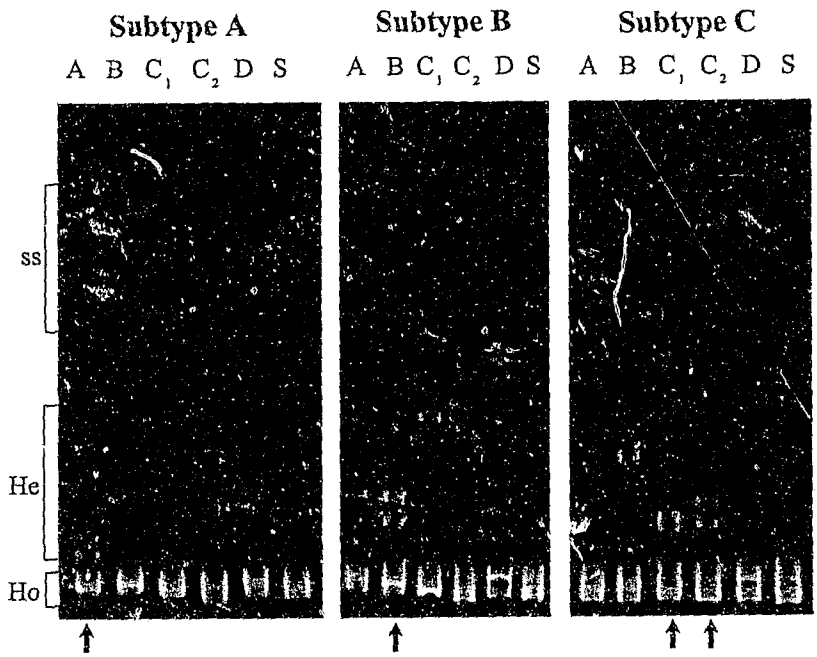


Figure 3.3 *gag* HMA profiles of 3 samples representing subtypes A (97PP110ZA), B (98HM12ZA) and C (95ZAM823ZA). Samples were run against reference plasmids from subtypes A (92UG029), B (92BR021B), C1 (92BR025), C2 (92RW009A) and D (92UG001). Lane S is the sample in the absence of any reference plasmids. The arrows indicate the lane with the fastest migrating heteroduplexes (He) which delineates the subtype. Homoduplexes (Ho) and single-stranded DNA (ss) molecular forms are indicated. Note that subtypes A and C are easily identified, whereas, subtype B heteroduplexes migrate only marginally faster than those in the D lane (middle gel).

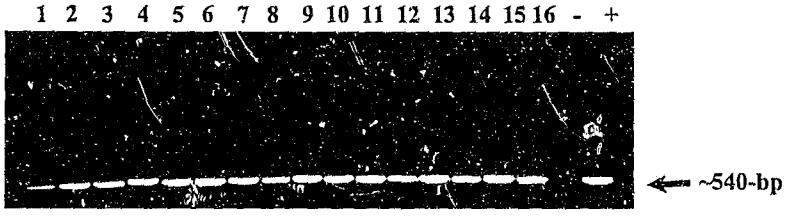


Figure 3.4 Agarose gel showing amplified LTR236/gag778 *gag* products. Lanes 1-16 represent PBMC from 16 HIV-1 infected patients, lane (-) an HIV-1 seronegative PBMC sample used as a negative control and lane (+) H9/IIIB-infected T-cells used as a positive control.

Table 3.3 Inter-subtype recombinants identified by *env* and *gag* HMA.

Patients	*gag* HMA	*env* HMA	Comments
PP110	A	C	South African male, place and route of transmission unknown
ZAM823	C	A	Zambian female, heterosexually* infected in Zambia
MOK	A	C	Congolese male, place and route of transmission unknown
CSW1	B	C	South African female commercial sex worker, infected in KwaZulu/Natal

*husband infected with C/C virus

3.3.4 *gag* and *env* sequence analysis

In order to validate the HMA results phylogenetic analysis was performed on 16 HIV-1 isolates identified by *env* HMA as subtypes A (n=4), B (n=8) and C (n=4). The 700-bp *env* sequence was compared with 12 references from subtypes A-E, 2 recombinant viruses ZAM184 (A/C) and MAL (A/D) and SIV_{cpzgab} (Forum, 1999) (Figure 3.5). HMA results were confirmed in 15 cases. Samples 96RS474ZA, 96SH523ZA, 98HM8ZA and 95ZAM823ZA clustered with subtype A; 96DS814ZA, 96DS815ZA, 96DS816ZA, 96DS833ZA, 96DS837ZA, 96DS897ZA, 96DS998ZA, 96DS1050ZA, 96DS814ZA and 96DS897ZA clustered with subtype B and 97PP110ZA, 97PP119ZA and 98CSW1ZA clustered with subtype C. However, one patient who

was subtyped as C (96MOKZA) by *env* HMA, did not cluster phylogenetically with a subtype group.

In order to confirm the discordant HMA profiles of the four putative recombinant isolates, the 520-bp *gag* fragment from patients with 95ZAM823ZA, 97PP110ZA, 96MOKZA and 98CSW1ZA was cloned and sequenced. One clone of each patient was phylogenetically analysed and compared with eight HIV-1 *gag* reference subtypes A-D and two known HIV-1 recombinant strains (Forum, 1999). Phylogenetic analysis of both *gag* and *env* gene fragments confirmed these 4 isolates to be recombinant viruses (Figure 3.5). Two A/C recombinants identified by HMA, 97PP110ZA and 96MOKZA, grouped with subtype A in the *gag* phylogenetic tree, but only the former grouped with *env* subtype C. The latter was shown to be an outlier as mentioned above. 95ZAM823ZA was confirmed as a C/A and 98CSW1ZA a B/C mosaic. In addition, *gag* fragments of 2 patients were sequenced; 97PP119ZA confirmed as C/C and 98HM8ZA confirmed as A/A. Both *env* and *gag* neighbour-joining trees showed the same topology as their parsimony trees (not shown), with clusters supported by bootstrap values of >75%.

3.3.5 Predicted protein sequence analysis

In order to reveal subtype specific patterns in the V3 loop, predicted amino acid sequences were aligned with an ALL consensus sequence (Dighe *et al.*, 1997). The characteristic tetrapeptide sequence GPGQ at the tip of the V3 loop was evident in all subtype A and C sequences, except for patient 95ZAM823ZA (GPGR) (Figure 3.6). For subtype B isolates GPGR was present in all isolates except for patient 96DS815ZA (GAGR).

3.4 Discussion

In this study, I describe the identification of HIV-1 subtypes A, B and C, and the relatively low prevalence of inter-subtype recombinants in Gauteng. Subtype C was the predominant subtype and was present mainly among individuals infected heterosexually. The majority of homosexually infected individuals and recipients of contaminated blood products were infected with subtype B viruses. Four recombinant viruses were found which were mosaics

between subtype C and regions belonging to either subtypes A or B.

The aim of this study was to document the HIV-1 subtypes in different risk groups in South Africa and to determine if new subtypes and inter-subtype recombinant viruses are emerging. Our results confirm those from other studies which showed that HIV-1 subtypes in South Africa segregate according to mode of transmission. Thus, the majority of heterosexual infections are due to *env* subtype C, while *env* subtype B predominates in the homosexual group (Engelbrecht *et al.*, 1995, Van Harmelen *et al.*, 1997, Williamson *et al.*, 1995). However, there is cross-over of subtypes between these groups as observed here and elsewhere (Van Harmelen *et al.*, 1997, Van Harmelen *et al.*, 1999b, Williamson *et al.*, 1995). Subtype B was also found in the majority of haemophiliacs infected through contaminated Factor VIII produced in the USA (Sher, 1986). A single case of subtype A infection (patient 98HM8ZA) was traced to a locally produced Factor IX. In this study, we found a total of 6 isolates with either or both *gag/env* subtype A fragments. This together with other cases of subtype A infection (Van Harmelen *et al.*, 1999b), probably signal the spread of this subtype from Central Africa where it is more common.

The prevalence of *gag/env* recombinant viruses in South Africa would appear to be relatively low (3%) compared to other countries in Africa, where in some areas prevalences of up to 25% have been recorded (Kampinga *et al.*, 1997). This is mainly due to the predominance of subtype C and low prevalence of other subtypes, but possibly also related to the more recent heterosexual epidemic in this region. Although there is a significant subtype B epidemic in South Africa, the apparent lack of recombination events between subtypes B and C probably reflects the limited overlap between the homosexual and heterosexual risk groups. However, the noted cross-over of these subtypes into different risk groups may alter this profile in the future. Indeed one B/C recombinant virus was identified in a South African commercial sex worker who may have been infected with more than one subtype as a result of multiple exposures. The other three recombinant viruses identified in this study involved subtypes A and C. Two of these were isolated from individuals from Zambia and Congo where subtype A circulates. The third recombinant virus, 96MOKZA, was from a local man whose place of transmission is unknown. This individual was classified as subtype A in *gag* and possibly had a mosaic *env* fragment as it was unclassified phylogenetically. It would be of interest to determine the possible existence of inter-subtype breakpoints within this *env* fragment. It is reported that subtypes A and C

recombine more frequently with other subtypes compared to subtypes B and D, which have a more restricted pattern of genetic exchange (McCutchan *et al.*, 1996). Thus, the apparent increase in subtype A viruses in South Africa may foster the emergence of more recombinant viruses in the future.

A combination of *env* and *gag* HMA provides a rapid and effective screening method for subtyping and identifying recombinant viruses or dual infections. This may underestimate the true frequency of recombinant viruses as recombination can also occur outside the two regions that were analysed. Although the *gag* HMA was able to identify subtype C viruses clearly, further work needs to be done to improve this assay for subtype B. It is likely that viruses classified as B or D are in fact subtype B because: i) they were subtype B in *env*; ii) subsequent sequencing analysis of 1 isolate confirmed it as subtype B; iii) subtype D has only rarely been reported in South Africa early on in the epidemic (Becker *et al.*, 1995, Engelbrecht *et al.*, 1995, Van Harmelen *et al.*, 1997, Williamson *et al.*, 1995); iv) subtype B viruses have previously been described among homosexual men in South Africa (Engelbrecht *et al.*, 1995, Van Harmelen *et al.*, 1997) and; v) individuals infected through contaminated blood products are known to have single source infections (Sher, 1986). Further refinement of the *gag* HMA may be complicated by the lower inter-subtype variation in the *gag* gene (14%) (Louwagie *et al.*, 1993) compared to the *env* gene (30%) (Janssens *et al.*, 1997), which probably accounts for less discrimination between subtypes. In particular, there is less genetic diversity between subtypes B and D, relative to other subtypes, which may explain why heteroduplexes formed with subtypes B and D migrated equivalently. It may be possible to improve on this by using representative local isolates as references strains or more variable regions of the *gag* gene (Becker *et al.*, 1995, Engelbrecht *et al.*, 1995, Van Harmelen *et al.*, 1997, Williamson *et al.*, 1995). Failing this it may be necessary to sequence all B or D indeterminates identified by HMA. In a country such as South Africa which has a predominantly subtype C epidemic, together with a lesser subtype B epidemic, the inability to distinguish between subtypes B and D may present only a minor problem. Nevertheless, the increase in the detection of other subtypes (such as A and possibly D) suggests that HMAs may need to be continually adapted to take into account new emerging viruses. Despite these limitations the HMA remains a powerful tool. It would have been considerably more difficult to identify the 4 recombinant viruses without the use of this quick screening method. We therefore, recommend the combination of *gag* and *env* HMAs to detect recombinant viruses in low frequency regions with a restricted number of subtypes.

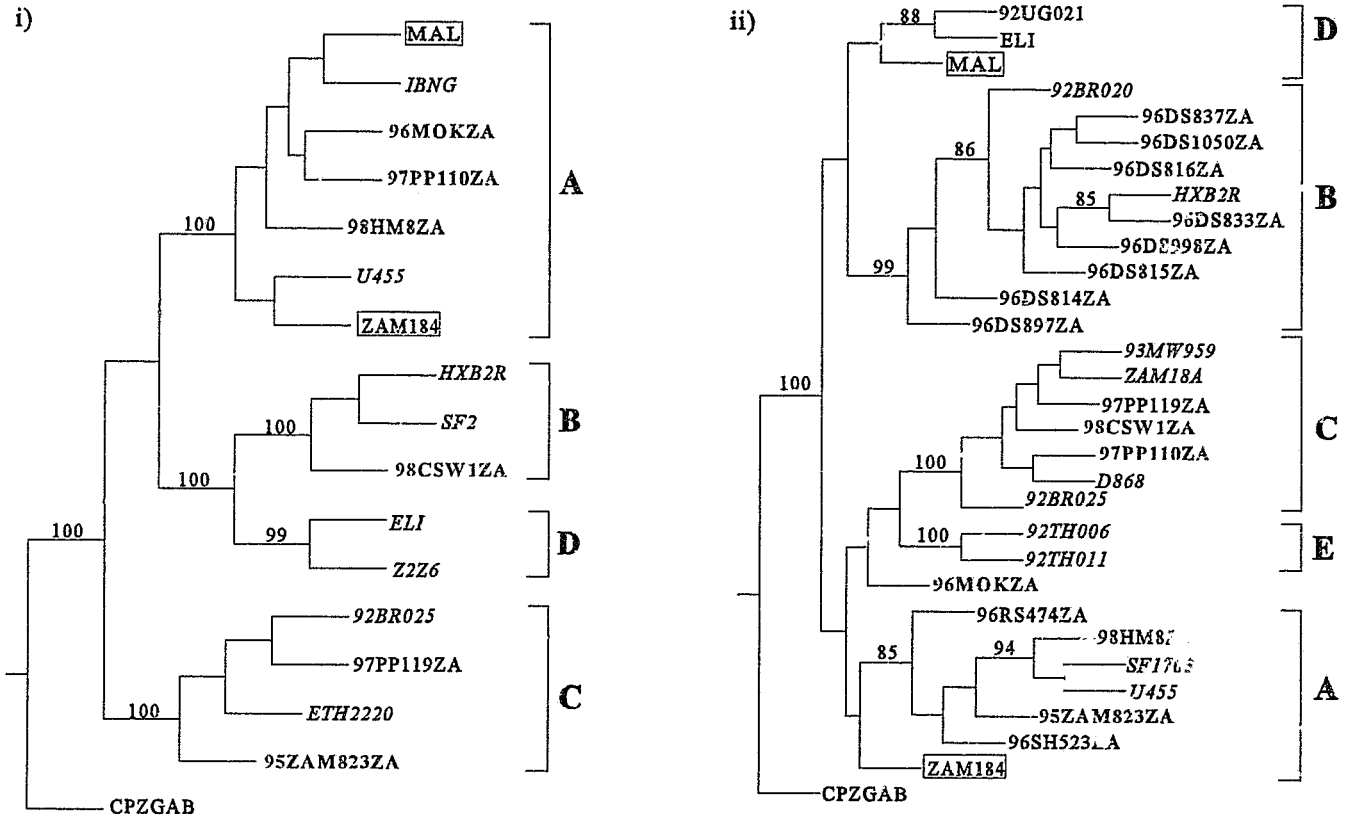


Figure 3.5 Phylogenetic analysis of 16 *env* (i) and 6 *gag* (ii) fragments of selected isolates including the 4 inter-subtype recombinants. Reference isolates from *gag* subtypes A-D (2 from each subtype) and *env* subtypes A-E (2 from each subtype except C where 4 were used) are used for comparison and are italicised. Two known HIV-1 recombinant viruses ZAM184 (A/C) and MAL (A/D) are shown in boxes. SIVcpz-Gab is used as an outlier. This neighbour-joining tree is supported by bootstrap values of >75%. Sequencing data confirmed the HMA results in all cases, except in patient 96MOKZA which is an outlier on the *env* tree despite being a clear subtype C by *env* HMA. The four inter-subtype recombinants are shown in colour: A/C (97PP110ZA and 96MOKZA), C/A (95ZAM823ZA) and B/C (98CSW1ZA). Additional isolates sequenced in this study are 98HM8ZA (A/A), 97PP119ZA (C/C) and 2 *env* subtype A isolates (96RS474ZA and 96SH523ZA) and 8 *env* subtype B isolates (all with the prefix 96DS) of which two are outliers to this group.

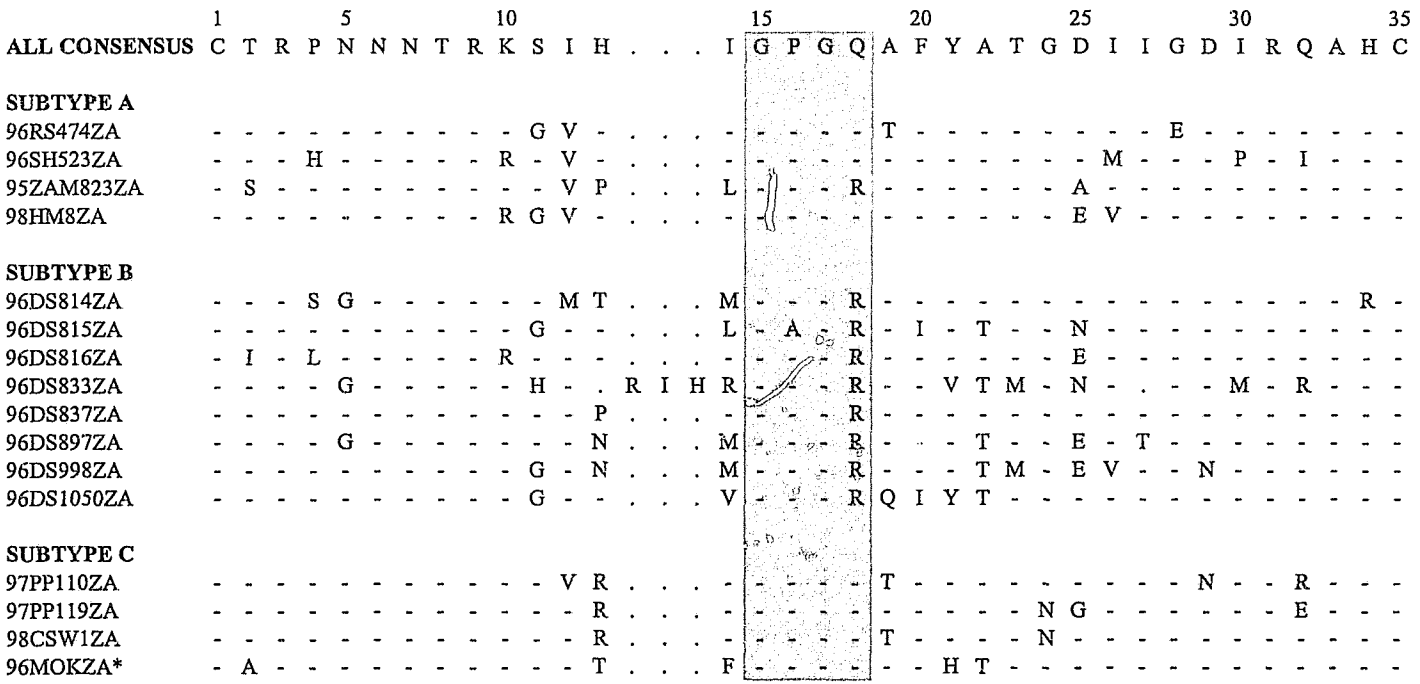


Figure 3.6 Predicted amino acid (aa) sequences of the *env* V3 loop of 16 isolates. Samples are grouped according to their subtype designation relative to the ALL consensus sequence (Dighe *et al.*, 199?). Dashes indicate identity and dots indicate gaps relative to the consensus. Changed aa are indicated. The tetrapeptide sequence at the tip of the V3 loop (position 15-18) is highlighted and shows that GPGQ is present in most A and C strains, while GPGR is in most B strains.

CHAPTER FOUR

GENETIC CHARACTERIZATION OF HIV-1 TYPE 1 FROM MIGRANT WORKERS IN THREE SOUTH AFRICAN GOLD MINES

4.1 Introduction

The spread of HIV-1 in South Africa is fueled by the migrant labor system. It not only creates situations of high risk behavior for the migrants, but also ensures the geographical spread of HIV-1 throughout the country. The mining industry in South Africa attracts job-seekers from many parts of the country as well as from other countries in southern Africa (Jochelson *et al.*, 1991). An HIV-1 seroprevalence survey conducted by the Chamber of Mines in 1986 showed that 4% of miners from Malawi, 0.3% from Botswana and less than 0.1% from Lesotho, Mozambique, Swaziland and South Africa were HIV-1 infected (Brink & Clausen, 1987, Chamber of Mines, 1993). Although no current data exist regarding current HIV prevalence among miners, mine doctors estimate it to be over 15% with an doubling time of 15 months (Williams & Campbell, 1996).

Subtypes A, C and D are most prevalent in sub-Saharan Africa (Janssens *et al.*, 1997) with subtypes A, B, C and D being found in South Africa (Becker *et al.*, 1995, Bredell *et al.*, 1998, Engelbrecht *et al.*, 1995, Van Harmelen *et al.*, 1997, Van Harmelen *et al.*, 1999b, Williamson *et al.*, 1995). Recent studies to determine the circulating subtypes showed the occurrence of subtype C in Maputo, Mozambique (Engelbrecht *et al.*, 1998), but similar studies in other southern African countries such as Botswana, Lesotho and Swaziland have not yet been done.

In order to investigate the genetic relationships among HIV-1 isolates from migrant workers employed by three adjacent gold mines in South Africa, 44 samples were phylogenetically analyzed. Based on published epidemiological data (Chamber of Mines, 1993) and phylogenetic analysis of the *env* gene, there is evidence of both epidemiological spread within the group of mines as well as multiple introductions of HIV-1 into this cohort.

4.2 Materials and Methods

4.2.1 Study setting

Gold Fields West Hospital serves a population of 28 500 gold miners working on three mines in Westonaria, a district situated 50 km from Johannesburg in the Gauteng province. The men, 18 to 65 years of age, were migrant workers living in single-sex hostels on the mines. They came from the rural areas of South Africa as well as the neighbouring countries of Botswana, Lesotho, Swaziland, and Mozambique, where they returned on annual leave. Patients sampled were proportionate to their representation on the mine (except for the Mozambicans, who were under-represented), but did not reflect the HIV prevalence rates in the different countries. It is company policy not to conduct pre-employment HIV testing and to test only when clinically indicated. Therefore, serostatus at the time of employment, date of seroconversion, and geographical place of infection were unknown.

4.2.2 Patients

Specimens were collected over a 4 month period, from October 1995 to January 1996, from 43 HIV-1 seropositive patients admitted with pulmonary tuberculosis. Except for samples 95ZA309, 95ZA327, 95ZA368 and 95ZA369 which were collagenase-digested lymph node cells (Gray *et al.*, 1996), all samples were uncultured PBMC isolated from a 5 ml EDTA-blood sample. HIV-1 seropositivity was confirmed using 3 ELISA tests (Genscreen® HIV1/2 version 2, Vironostika® HIV Uni-Form II plus O, Wellcozyme* HIV Recombinant). $CD4^+$ lymphocyte counts were determined by flow cytometry using Simultest monoclonal antibodies and a FACSort equipped with SimulSet software (all from Becton Dickinson, San Jose, California, USA). Specimens were obtained with the informed consent of all participants and ethical approval was obtained from the Committee for Research on Human Subjects of the University of the Witwatersrand (see Appendix C and D).

4.2.3 PCR, HMA and phylogenetic analysis

Lysed PBMC were used as templates for the amplification of the ~700-bp *env* fragment, as described in Chapter 2. PCR products from all 44 samples were subtyped by *env* HMA using reference plasmids A_2, B_1, C_2, D_1 and E_2. For patient 96ZA119BWA additional references B_2, B_3, C_1, C_3 and C_4 were also used. Products were cloned into the pMOS*Blue* T-vector and positive clones were identified by PCR using primers ES7/ES8. For each isolate, one positive PCR clone

was sequenced in both directions: 35 samples were sequenced by the dideoxy-chain termination method using the Sequenase PCR Product Sequencing kit, and the rest by automated sequencing (ALFexpress™ DNA sequencer) using the Cy5 AutoRead sequencing kit and Cy5-dATP labeling mix. The 44 ES7/ES8 *env* sequences were aligned with 38 reference sequences representative of subtypes A (2), B (2), C (25), D (2), E (2), F (2), G (2), and $SIV_{cpz\text{-}Gab}$ as an outlier (Forum, 1999), using Clustal W (Thompson *et al.*, 1994). Neighbor-joining trees were generated using TREECON (Van de Peer, 1994) and a limited set of sequences (n = 64, including all 44 sequences, 19 reference A-D subtype sequences and $SIV_{cpz\text{-}Gab}$) were also analysed using PAUP 3.1.1 (Swofford, 1993). All of the 200 equally parsimonious shortest trees were used to generate a strict 50% majority rule consensus tree.

The sequences were named by year of isolation, country of origin of the patient and his identification number. For the purpose of distinguishing between isolates, the last two or three characters denote geographical origin of the patient: ZA (South Africa), LSO (Lesotho), MOZ (Mozambique), SWZ (Swaziland) and BWA (Botswana). *Env* sequences are listed in GenBank, identified by accession numbers AF053181-AF053224.

4.3 Results

4.3.1 Study subjects

The 44 HIV-1 samples were obtained from 43 individuals originating from five different countries in southern Africa (Figure 4.1). Samples 95ZA853BWA and 96ZA119BWA were from the same patient, two months apart. The largest group (19 out of 43) of patients were from South Africa, 13 from Lesotho, 5 from Botswana and 3 each from Swaziland and Mozambique. The spectrum of $CD4^+$ lymphocyte cell counts ranged from 30 to 930 cells/μl. Forty percent of patients had counts less than 200 $CD4^+$ T-cells/μl indicating that many patients had advanced disease and likely infected for a long period of time (Table 4.1). One patient (95ZA961ZA) had recently seroconverted.

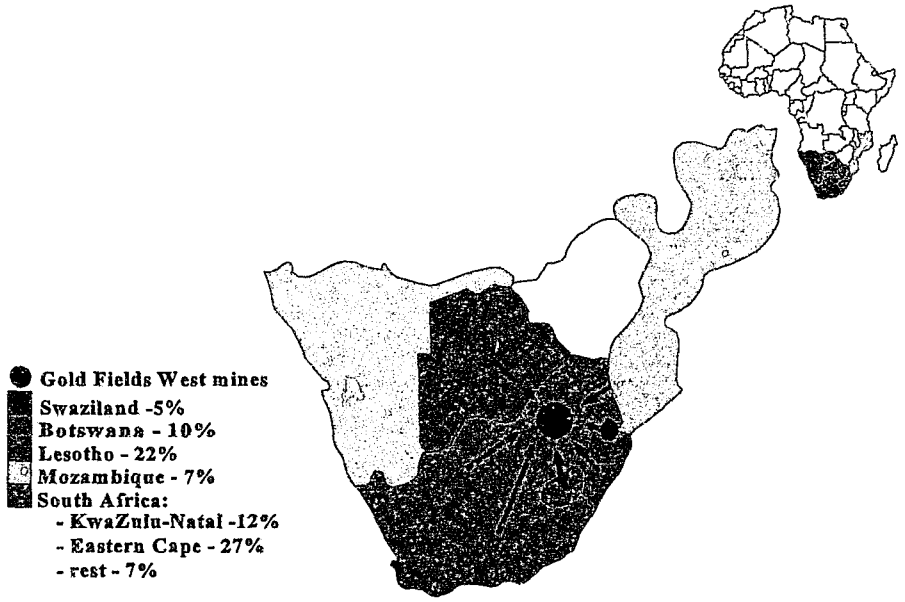


Figure 4.1 Origin of the workforce at the mines where this study was conducted. Less than half (46%) of the workforce are from South Africa. Mostly they migrate from the eastern part of the country from the provinces of Kwazulu/Natal and the Eastern Cape (formerly Ciskei and Transkei). The study site is situated in the province of Gauteng where only a small fraction of the workforce is drawn. Over half the workforce is recruited from countries bordering on South Africa such as Mozambique and Botswana or within the borders of South Africa such as Lesotho and Swaziland.

4.3.2 PCR and HMA

Uncultured PBMC and collagenase-digested lymph node cells were all found to be HIV-1 positive by nested PCR using primers ED5/ED12 and ES7/ES8. The ~700-bp nested PCR products, spanning the V3 - V5 region, were subtyped by HMA. Each sample was compared against one representative of each reference subtype: A_2, B_1, C_2, D_1 and E_2 (Figure 4.2). Forty-three of the 44 samples could be clearly assigned to HIV-1 subtype C on the basis of markedly faster mobilities of heteroduplexes formed with C_2 (defined as migrating two-thirds of the distance compared to one-third or less with other subtypes, relative to the homoduplexes). One sample, however, 95ZA853BWA, showed equivalent migration with B_1 and C_2 (Figure 4.3)

Table 4.1 Patients' geographic origin and $CD4^+$ count.

Sample	CD4 count	Geographic origin
95ZA309	390	Lesotho
95ZA327	620	KwaZulu/Natal, SA
95ZA368	80	KwaZulu/Natal, SA
95ZA369	370	Lesotho
95ZA650	190	KwaZulu/Natal, SA
95ZA65.?	50	Mozambique
95ZA653	ND	Lesotho
95ZA654	30	Lesotho
95ZA656	120	Mozambique
95ZA658	730	Lesotho
95ZA659	270	KwaZulu/Natal, SA
95ZA711	1176	Eastern Cape, SA
95ZA712	125	Botswana
95ZA716	120	Mpumalanga, SA
95ZA718	277	Eastern Cape, SA
95ZA722	589	Swaziland
95ZA724	269	Botswana
95ZA726	630	Botswana
95ZA728	285	Mozambique
95ZA731	ND	Northern Province, SA
95ZA735	456	KwaZulu/Natal, SA
95ZA736	797	KwaZulu/Natal, SA
95ZA739	635	Eastern Cape, SA
95ZA742	ND	Eastern Cape, SA
95ZA770	90	Eastern Cape, SA
95ZA784	356	Lesotho
95ZA787	700	Lesotho
95ZA799	442	Eastern Cape, SA
95ZA801	395	Eastern Cape, SA
95ZA803	379	Lesotho
95ZA804	930	Eastern Cape, SA
95ZA828	385	Lesotho
95ZA830	412	Swaziland
95ZA834	153	Lesotho
95ZA836	325	Botswana
95ZA845	274	Lesotho
95ZA853*	93	Botswana
95ZA854	107	Lesotho
95ZA857	ND	Eastern Cape, SA
95ZA859	168	Lesotho
95ZA863	96	Eastern Cape, SA
95ZA868	122	Swaziland
95ZA961#	ND	Eastern Cape, SA
96ZA119*	108	Botswana

* Sample 96ZA119 is a two month follow-up of sample 95ZA853

recent seroconverter

suggesting that this patient was either co-infected with subtype B and C, or infected with a virus showing equivalent homology to B_1 and C_2. Twenty separate clones from this sample also showed equivalent migration with B_1 and C_2 indicating that this patient was not co-infected. Minor variation in heteroduplex formation represent intra-patient quasispecies (Figure 4.4). Further analysis using two additional reference B plasmids (B_2 from Thailand and B_3 from USA) and three additional C plasmids (C_1 from Malawi, C_3 from India, and C_4 from Brazil) confirmed that 95ZA853BWA was indeed a subtype C virus: heteroduplexes in C_1, C_2, C_3, and C_4 all migrated ahead of heteroduplexes formed with B_2 and B_3, but equivalent to B_1 (Figure 4.5). A follow-up sample from this patient, 96ZA119BWA, was clearly subtype C (Figure 4.3). The subtype designation was later confirmed by sequence analysis, which showed that both samples from this patient were HIV-1 subtype C.

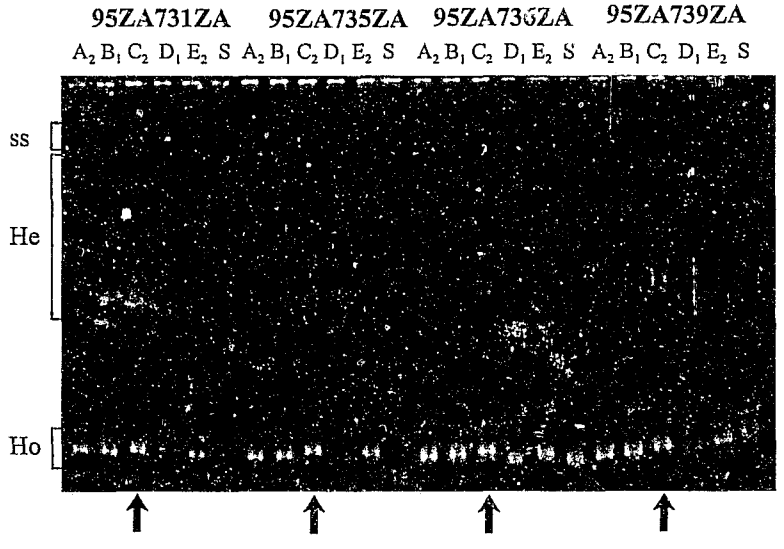


Figure 4.2 *env* HMA profiles of 4 samples (95ZA731ZA, 95ZA735ZA, 95ZA736ZA and 95ZA739ZA) representing subtype C. Samples were run against reference plasmids from subtypes A_2 (IC144), B_1 (BR20), C_2 (ZM18), D_1 (UG21) and E_2 (TH06). Lane S is the sample in the absence of any reference plasmids. The arrows indicate the lane with the fastest migrating heteroduplexes (He) which delineates the subtype. Homoduplexes (Ho) and single-stranded DNA (ss) are indicated.

4.3.3 DNA sequencing and phylogenetic analysis

The ~700-bp PCR products of the 44 isolates were cloned and one representative clone of

each isolate was sequenced in both directions. Sequences were phylogenetically compared with 25 subtype C sequences (Forum, 1999), 6 from South Africa (90ZA514, 92ZA517, NOF, DLU, U0GOM and BOOYD) (Becker *et al.*, 1995, Engelbrecht *et al.*, 1995), 12 from other regions in Africa (ZAM18A, UG268A2, 93MW959.18, SH750, 93MW960.3, 93MW965.26, SM145A, ZAM20A, Z8A1, DJ259A, DJ373A, and SE364A), 6 from India (D744, D747, D757, D760, D808 and D1044), and 1 from Brazil (92BR025.9). Two reference sequences each of subtype A (SF1703 and U455), B (SF2 and HXB2R), D (ELI and Z2Z6), E (CAR4017 and 93TH966.8), F (CA20 and BZ163A) and G (92UG975.10 and RU511B) were also included. All 44 miner sequences grouped with known subtype C sequences, confirming the HMA results. The topology of the neighbor joining trees (Figure 4.6) was essentially the same as that of the PAUP tree except in some of the terminal branches (not shown), and clusters supported by bootstrap values >80% were also observed in the PAUP tree.

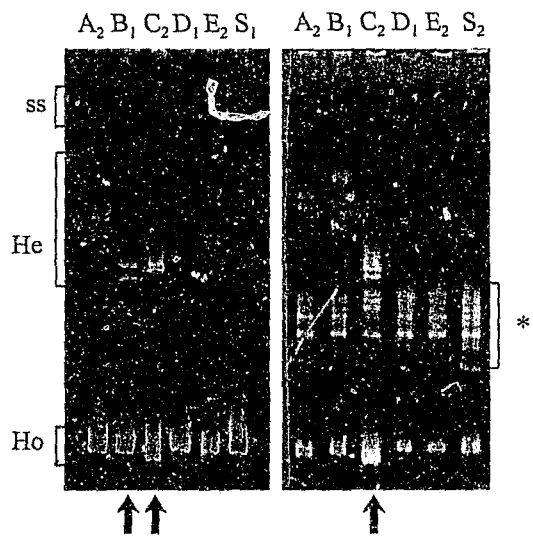


Figure 4.3 *env* HMA profiles of sample 95ZA853BWA (S_1) and a follow-up sample 96ZA119BWA (S_2) from the same patient. Samples were run against reference plasmids from subtypes A_2 (IC144), B_1 (BR20), C_2 (ZM18), D_1 (UG21) and E_2 (TH06). Lane S is the sample in the absence of any reference plasmids. * These bands, present in all lanes, represent non-specific bands or intra-patient quasispecies and are therefore ignored. The arrows indicate the lane with the fastest migrating heteroduplexes (He) which delineates the subtype. Note that sample 95ZA853BWA shows equivalent migration with B_1 and C_2 whereas the follow-up sample 96ZA119BWA was clearly subtype C. Homoduplexes (Ho) and single-stranded DNA (ss) are indicated.

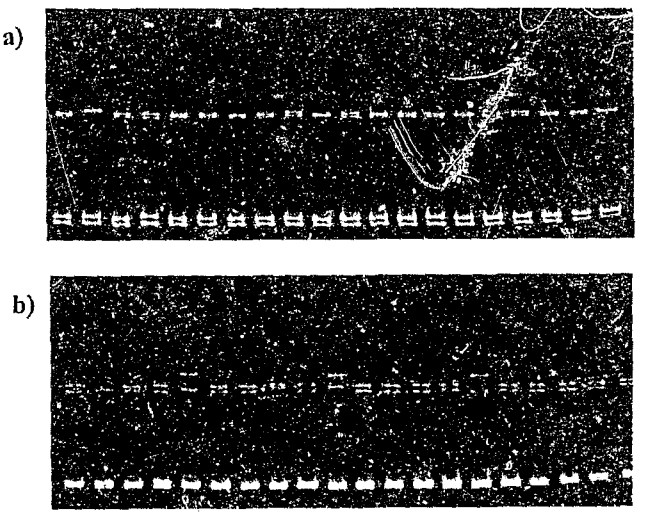


Figure 4.4 *env* HMA profiles of twenty clones from sample 95ZA853BWA showing equivalent migration with B_1 (a) and C_2 (b).

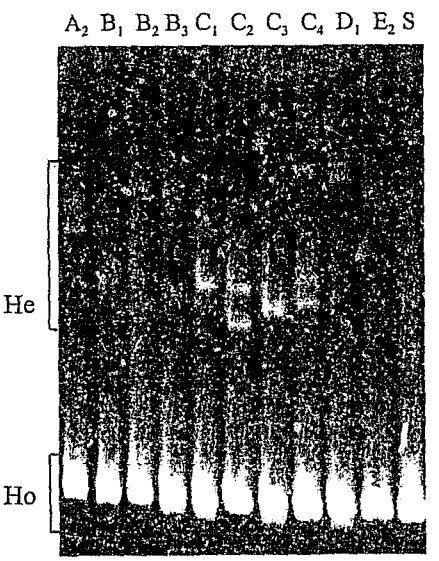


Figure 4.5 *Env* HMA profile of sample 95ZA853BWA using additional reference plasmids. Samples were run against reference plasmids from subtypes A_2 (IC144), B_1 (BR20), B_2 (TH14), B_3 (SF162), C_1 (MA959), C_2 (ZM18), C_3 (IN868), C_4 (BR25), D_1 (UG21) and E_2 (TH06). Lane S is the sample in the absence of any reference plasmids. Homoduplexes (Ho) and heteroduplexes (He) are indicated. Note that heteroduplexes in C_1, C_2, C_3, and C_4 all migrated ahead of heteroduplexes formed with B_2 and B_3 but equivalent to B_1.

To determine whether HIV-1 isolates grouped according to the geographic origin of the migrant workers, each isolate was compared to isolates from patients from the same geographic region (intra-group) and to isolates from patients from other geographic regions (inter-group). In addition, isolates were compared to two reference subtype C isolates and two reference subtype B isolates. The mean nucleotide diversity within and between each group and subtype is shown in Table 4.2.

Table 4.2 Intra- and inter-subtype nucleotide diversity among 54 HIV-1 isolates[a,b].

	Percentage mean nucleotide diversity							
Origin(n)	ZA(19)	LSO(13)	BWA(6)	MOZ(3)	SWZ(3)	REF B(2)	REF C(2)	INDIAN(6)
ZA(19)	**10.4**	11.2	10.5	11.8	10.6	22.5	9.1	9.7
LSO(13)	11.2	**10.6**	10.8	12.2	10.8	22.6	9.1	10
BWA(6)	10.5	10.8	**7.3**	10.9	10	22	8.2	8.5
MOZ(3)	11.8	12.1	10.8	**8.3**	11.8	22.9	9.7	11.3
SWZ(3)	10.6	10.8	9.5	11.7	**7.6**	22.7	9.3	9.5
REF B(2)	22.5	22.6	22	22.9	22.7	**3.3**	20.2	22.1
REF C(2)	9.1	9.1	8.2	9.6	9.3	21.2	**3.3**	7.6
INDIAN(6)	9.7	10	8.6	11.3	9.5	22.1	7.6	**4.2**

[a]Including migrant workers infected with subtype C virus, reference B isolates, reference C isolates and subtype C isolates from India.

[b]Nucleotide distances were calculated from 54 HIV-1 sequences, using the Kimura method. The 44 migrant workers all belong to subtype C and are grouped according to their geographic origin (ZA, South Africa; LSO, Lesotho; BWA, Botswana; MOZ, Mozambique; SWZ, Swaziland), which may not necessarily represent the source of their infection. Values for the intra-group and intra-subtype distances are shown in boldface. Comparison between and within groups and subtypes was done by calculating the genetic differences between each individual isolate and all the other isolates in that group (intra-group) or within the same subtype (intra-subtype; e.g. reference C and IndiaC) or inter-subtype (reference B). The two subtype B isolates in this analysis (SF2 and HXB2R) show close homology and are not representative of the divergence seen within this group.

There was at least a 20% difference between isolates from subtypes B and C, clearly delineating these two subtypes. However, except for a few clusters (see below), the genetic diversity among the groups of migrant workers presented as a spectrum with no apparent phylogenetic relatedness within groups. This was reflected in the overlap between the mean intra-group nucleotide diversity of 7.3%-10.6% and the mean inter-group diversity of 9.5%-12.2%. Thus, subtype C viruses from a cohort of migrant workers could not be separated on the basis of the geographic origins of the infected individual.

HIV-1 subtype C sequences from the miners were not more closely related to each other than to other subtype C viruses from South Africa. The average DNA difference between the 44 miners' sequences was 11%, compared with the intra-subtype C virus DNA difference of 8.9% for the 6 published South African isolates and 10.6% for the 19 subtype C viruses of non-South African origin. This is in contrast to the relationship seen between sequences from India, which have previously been shown to have resulted from recent spread of a single ancestor (Grez *et al.*, 1994) and, in our analysis, showed a mean divergence of only 4% (Table 4.2). This small degree of diversity among isolates within the Indian group of viruses compared to isolates from other groups was also confirmed by the single cluster of Indian sequences in the phylogenetic tree (Figure 4.6). None of the miners' sequences showed significant homology with the Indian sequences.

Five clusters of closely related sequences were identified within the mining cohort supported by bootstrap values of 75% or more (Figure 4.6, indicated by number): (1) 95ZA830SWZ, 95ZA653LSO, together with 95ZA784LSO and 95ZA368ZA; (2) 95ZA854LSO and 95ZA787LSO; (3) 95ZA739ZA and 95ZA654LSO, and (4) 95ZA868SWZ, 95ZA859LSO, 95ZA718ZA, 95ZA857ZA and 95ZA799ZA. These clusters linked isolates from miners with the same geographic origins (cluster 2 and three patients in cluster 4 from the Eastern Cape region of South Africa) and miners who worked on the same mine (cluster 3 and four of the miners in cluster 4). There was no similar obvious linkage between the isolates in cluster 1. The average DNA distances between the isolates in these four clusters were 7%, 8%, 4% and 4% respectively. Samples 95ZA853BWA and 96ZA119BWA, the fifth cluster, showed 97% homology, as would be expected given that these two samples were from the same patient taken 2 months apart. No apparent differences in genetic diversity between samples from lymph nodes and PBMC were evident, although a more extensive analysis of the quasispecies in the lymph nodes and PBMC from a single patient would more fully address this issue.

4.3.4 Predicted protein sequence analysis

The characteristic tetrapeptide sequence at the tip of the V3 loop of subtype C viruses, GPGQ, was present in the predicted amino acid sequences of all the isolates except for 95ZA742ZA (RPGQ), 95ZA652MOZ and 95ZA961ZA (both GPGR) (Figure 4.7). The latter sample was from a patient who had recently seroconverted. Amino acid positions 12, 24 and 25 of isolates from patients from all 5 southern African countries showed extensive variation from

the published subtype C consensus sequence (Dighe *et al.*, 1997). However, analysis of the isolates according to the geographic origins of the patients did not reveal any characteristic sequence patterns for geographic groups (Figure 4.7). The median V3 loop charge for this group was +4. All of the 44 subtype C viruses had neutral amino acids at position 11, 25 and 32 associated with a non-syncitium-inducing phenotype (NSI). There were two amino acid changes in sample 96ZA119 compared with the previous sample, 95ZA853, from the same patient taken two months earlier.

4.4 Discussion

Genetic analysis of HIV-1 from 43 infected individuals from a mining population recruited from the rural areas of South Africa and from the neighboring countries of Botswana, Lesotho, Swaziland and Mozambique confirms the predominance of subtype C in this region. The lack of significant genetic relatedness amongst isolates in this cohort suggests that HIV-1 subtype C has evolved through multiple introductions of subtype C into this population.

In this study we investigated the extent of HIV-1 genetic diversity within the mining cohort to identify possible epidemiological linkage between HIV-1 isolates. Phylogenetic analysis showed a lack of genetic relatedness between isolates from individuals from the same geographic region. Compared to the Indian isolates, which showed close homology, the isolates from the miners were very heterogenous in the V3-V5 *env* region. Thus, there was considerable intra-group (miners from the same geographic region) and inter-group (miners from different geographical regions) genetic diversity that did not delineate separate clusters. This indicates that most of the HIV-1 infections on the mine did not originate from the recent spread of a common progenitor virus. This supports previous studies done elsewhere in South Africa, where the intra-subtype DNA distance for subtype C viruses indicated multiple introductions of the same subtype into the heterosexual community rather than a clonal epidemic (Van Harmelen *et al.*, 1999b). The 5 clusters identified in the phylogenetic tree showed between 3-8% diversity. Since it has been estimated that HIV-1 evolves at a rate of approximately 1% per year in this region (Myers *et al.*, 1993), the relatively low diversity seen among these clusters indicates recent spread from common sources and possible epidemiological linkage. In most cases these clusters involved individuals of disparate geographic origins, some of whom worked on the same mine,

suggesting in these cases possible transmission within the mine or within communities around the mine.

Recent studies report that the SI-phenotype, i.e. CXCR4 dependence, appears to be rare among subtype C isolates (Morris *et al.*, 1999, Peeters *et al.*, 1999). SI variants evolve in approximately half of subtype B infected individuals associated with an increased positive charge in the V3 loop and the use of the CXCR4 co-receptor (Richman & Bozzette, 1994, Tersmette *et al.*, 1989). The overall positive charge of the V3 loop in this study was moderately low especially considering that 40% of the miners had $CD4^+$ counts <200 cell/μl. In addition, the amino acids at positions 11, 25 and 32 were all neutral, which is consistent with a NSI phenotype. Although we were not able to examine the *in vitro* phenotypic properties of these viruses, our data suggest that they are likely to be NSI.

The movement of migrant work-seekers from rural areas to the urban and mining workplace and then back home is a common feature of southern African society. While on the mines, the majority of miners live in single-sex hostels without wives and families, but in contact with surrounding communities including a mobile female sex worker population. Although transmission of HIV-1 is predominantly through heterosexual contact with sex workers, homosexual activities also occur in the hostels, but mainly involves contact rather than anal penetration (Jochelson *et al.*, 1991). Furthermore, the rate of other sexually transmitted diseases is high in the mineworker population which may facilitate transmission (Jochelson *et al.*, 1991). The extent of the migrant labor system in South Africa means that once HIV-1 is present in the urban areas it would inevitably spread to rural areas (Campbell & Williams, 1996), and vice-vera. Movement of migrant workers between South Africa and its neighboring countries could facilitate spread of HIV-1 in this sub-region of Africa. HIV-1 infection is therefore unlikely to remain confined to a specific geographic region for any length of time. The HIV-1 heterogeneity seen in this mining cohort is likely the result of many factors including external sources of infection as a result of migrancy, together with exposure of mine workers to the local population. While these data suggest multiple introductions of HIV-1 into the mining population it gives no clues as to the origin of the viruses. Further studies into the molecular epidemiology of HIV-1 in neighboring countries and among local sex workers may shed some light on this question.

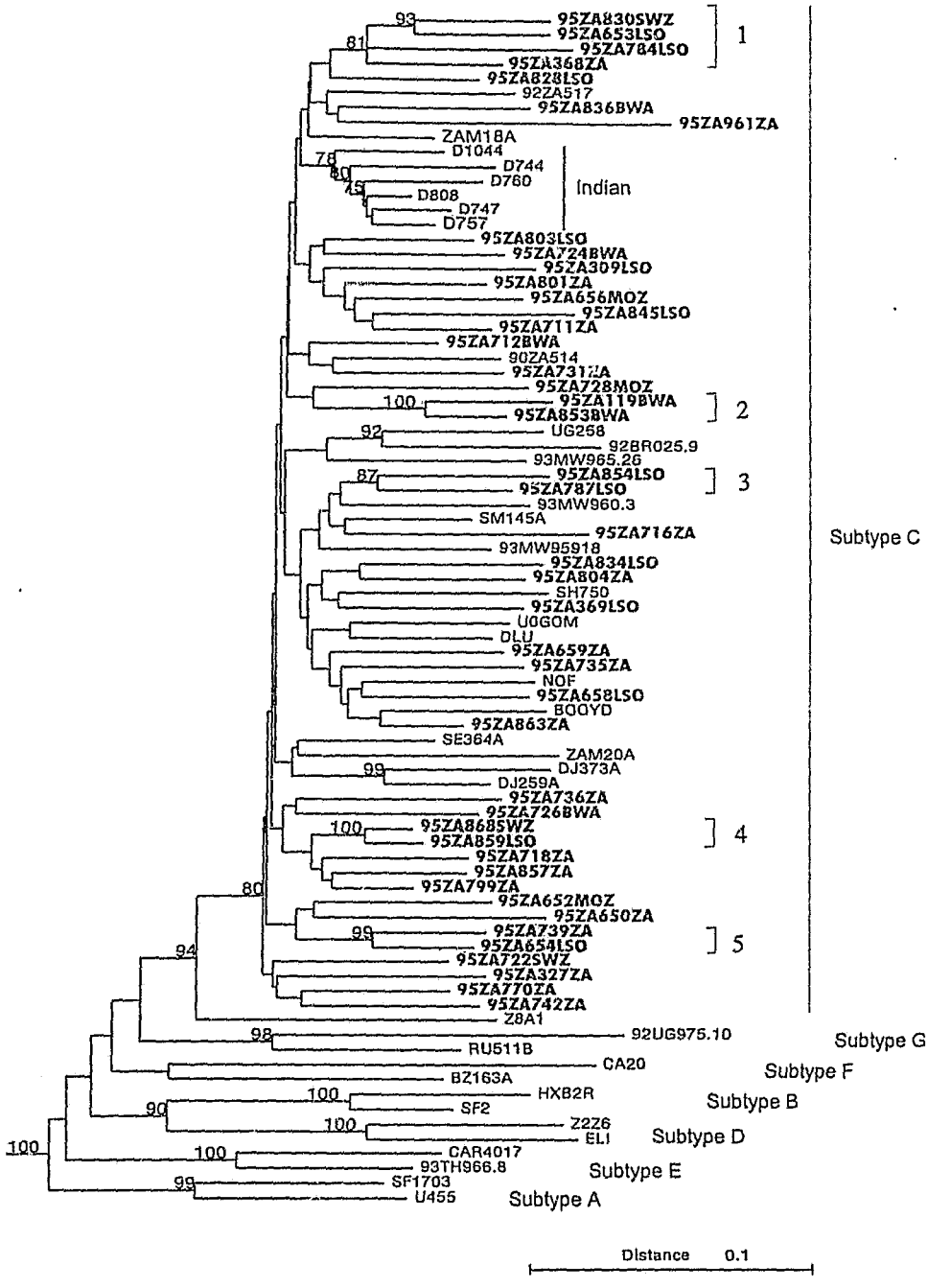


Figure 4.6 Phylogenetic analysis of the gp120 V3-V5 region (~700-bp) of 44 HIV-1 isolates from 43 miners from South Africa (ZA, *n*=19), Lesotho (LSO, *n*=13), Botswana (BWA, *n*=5), Mozambique (MOZ, *n*=3) and Swaziland (SWZ, *n*=3). Phylogenetic trees were constructed using both distance matrix and maximum parsimony approaches. The phylogenetic tree with bootstrap values of >75% are shown. Samples from this analysis are in boldface and clusters of interest are indicated. For comparison, other subtype C isolates from South Africa (*n*=6) and India (*n*=6), and reference C isolates (*n*=25), were included. The Indian cluster is labelled. In addition, two reference isolates from subtypes A, B, D, E, F and G were included for comparison (see text for details). $SIV_{cpz\text{-}Gab}$ was used as an outlier (not shown).

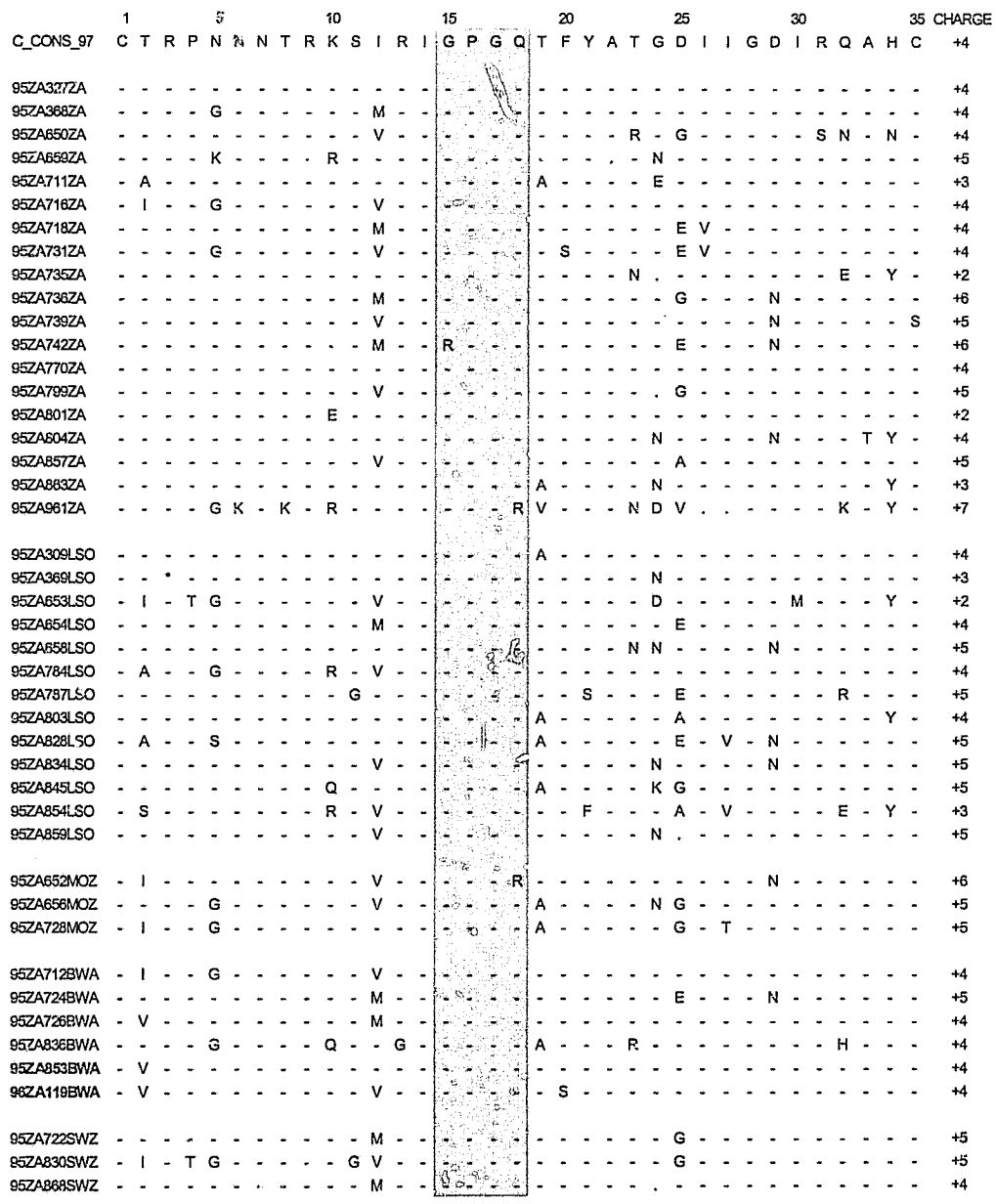


Figure 4.7 Predicted amino acid sequences of V3 loop from 43 migrant workers grouped according to country of origin relative to the consensus subtype C sequences (Dighe *et al.*, 1997). Amino acids (aa) that are identical to the published sequences are shown as a dash and changed aa are indicated. Dots indicate gaps and the asterisk in 95ZA369LSO indicates a stop codon. Two samples from the same patient collected 2 months apart are in bold. The net amino acid charge of the V3 loop was calculated by subtracting the number of negatively charged amino acids (D and E) from the number of positively charged ones (H, K and R).

The HIV-1 epidemic in South Africa is increasing exponentially. Based on the current trend it is predicted that between 18% and 27% of the South African population will be HIV-1 infected by the year 2010 (Doyle, 1991). The vast majority of isolates in South Africa are subtype C (Van Harmelen *et al.*, 1999b), although A, B and D have also been identified (Becker *et al.*, 1995, Bredell *et al.*, 1998, Engelbrecht *et al.*, 1995, Van Harmelen *et al.*, 1997, Van Harmelen *et al.*, 1999b, Williamson *et al.*, 1995). Given the extensive movement of people in this sub-region and the presence of multiple HIV subtypes in countries to the north and north-east (Janssens *et al.*, 1997), it is likely that other subtypes will appear in South Africa further contributing to the HIV-1 genetic pool. Globally HIV-1 subtype C is the most prevalent virus and it is likely that future vaccine strategies will be directed towards this subtype. The high level of genetic diversity that are seen among subtype C viruses (Bredell *et al*, 1998, Van Harmelen *et al.*, 1997, Van Harmelen *et al.*, 1999b) is likely to complicate the identification of a South African prototype subtype C strain as well as the design of vaccines cross-reactive with other HIV-1 subtypes. Thus, a comprehensive analysis of subtype C viruses and continual monitoring of the spread and genetic evolution of this subtype is an important aspect in the global control of HIV-1.

CHAPTER FIVE

CONCLUDING REMARKS

Genetic analysis of HIV-1 from 215 infected individuals has confirmed the predominance of subtype C in Gauteng. This subtype was found mainly within the heterosexually infected group currently estimated to affect over 3 million individuals. A smaller subtype B epidemic was present among the majority of homosexually infected individuals and those infected through contaminated blood products. It is unlikely that subtype B will expand significantly in the future as education programs among homosexual groups and screening of blood have been effective interventions. The four *env* subtype A sequences that were identified probably represents a recent introduction of this subtype from Central Africa. It remains to be seen if subtype A will expand significantly in South Africa given that subtype C is already so well established. Nevertheless, the presence of other subtypes contributes to the genetic pool and may also facilitate the emergence of recombinant viruses. While at present there appears to be a low frequency of inter-subtype recombinant viruses in Gauteng, it is likely that the true frequency of these events may be underestimated as only 2 gene regions were analysed. Full-length sequencing or the analysis of additional gene fragments will become a necessity for a full appraisal of the extend of recombination events.

HIV-1 diversity increases as a consequence of the movement of infected individuals between different countries and within a particular country. Migrancy is a major feature of the southern African region and has contributed significantly to the spread of HIV. The majority of migrant workers are driven by rural poverty to seek job opportunities in the urban areas of South Africa, such as the mines. This has created a market for prostitution in mining towns allowing for inter-mixing of different HIV-1 subtypes within and between urban and rural communities. Despite this, all miners were infected with HIV-1 subtype C. Although the source of these infections cannot be determined, attending doctors believe miners to be infected locally on the mines and this may not be a significant source of new viral strains. In contrast, non-C subtypes were found among individuals attending urban clinics in central Johannesburg. These clinics service areas with a large number of immigrants, particularly from other countries in Africa. Thus, it is likely that this represents introductions of other viral subtypes into South Africa.

The genetic characterization of HIV-1 isolates, not just highly diverse strains, but also strains that are globally prevalent, together with the frequency and distribution of different subtypes within a population, provides the basis for vaccine development. Sampling of isolates for genetic analysis should be as diverse as possible, both geographically and in terms of transmission modes, but should also include recent infections so as to reflect the strains currently being transmitted. Vaccine development is focussing on HIV-1 subtype C which now accounts for almost 50% of all infections world-wide. South Africa is well poised to play an important role in developing a vaccine against subtype C. Thus, continual monitoring and development of new methodologies are needed to stay abreast of the evolution of HIV in high incidence areas such as South Africa.

REFERENCES

Abebe, A., Kuiken, C. L., Goudsmit, J., Valk, M., Messele, T., Sahlu, T., Yeneneh, H., Fontanet, A., De Wolf, F. & Rinke De Wit, T. F. (1997). HIV type 1 subtype C in Addis Ababa, Ethiopia. *AIDS Res. and Hum. Retroviruses* **13**, 1071-1075.

Adjorlolo-Johnson, G., De Cock, K. M., Ekpini, E., Vetter, K. M., Sibailly, T., Brattegaard, K., Yavo, D. D., R. , Whitaker, J. P., Kestens, L., Ou, C. Y., George, J. R. & Gayle, H. D. (1994). Prospective comparison of mother-to-child transmission of HIV-1 and HIV-2 in Abidjan, Ivory Coast. *JAMA* **272**, 462-466.

Ahmad, N., Baroudy, B. M., Baker, R. C. & Chappey, C. (1995). Genetic analysis of human immunodeficiency virus type 1 envelope V3 region isolates from mothers and infants after perinatal transmission. *J. Virol.* **69**, 1001-1012.

Alaeus, A., Leitner, T., Lidman, K. & Albert, J. (1997). Most HIV-1 genetic subtypes have entered Sweden. *AIDS* **11**, 199-202.

Alkhatib, G., Combadiere, C., Broder, C. C., Feng, Y., Kennedy, P. E. & Murphy, P. M. (1996). CC CKR5: a RANTES, MIP-1a, MIP-1b receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* **272**, 1955-1958.

Anderson, R. M. & May, R. M. (1988). Epidemiological parameters of HIV transmission. *Nature* **333**, 514-519.

Artenstein, A. W., VanCott, T. C., Mascola, J. R., Carr, J. K., Hegerich, P. A., Gaywee, J., Sanders-Buell, E., Robb, M. L., Dayhoff, D. E., Thitivichianlert, S., Nitayaphan, S., McNeil, J. G., Birx, D. L., Michael, R. A., Burke, D. S. & McCutchan, F. E. (1995). Dual infection with human immunodeficiency virus type 1 of distinct envelope subtypes in humans. *J. Infect. Dis.* **171**, 805-810.

Barre-Sinoussi, F. (1996). HIV as the cause of AIDS. *Lancet* **348**, 31-35.

Baskar, P. V., Ray, S. C., Rao, R., Quinn, T. C., Hildreth, J. E. & Bollinger, R. C. (1994). Presence in India of HIV type 1 similar to North American strains. *AIDS Res. Hum. Retroviruses* **10**, 1039-1041.

Becker, M. L. B., De Jager, G. & Becker, W. B. (1995). Analysis of partial *gag* and *env* gene sequences of HIV-1 type 1 from South Africa. *AIDS Res. Hum. Retroviruses* **11**, 1265-1267.

Bedinger, P., Moriarty, A., von Borstel, R. C. I., Donovan, N. J., Steimer, K. S. & Littman, D. R. (1988). Internalization of the human immunodeficiency virus does not require the cytoplasmic domain of CD4. *Nature* **334**, 162-165.

Betts, M. R., Krowka, J., Santamaria, C., Balsamo, K., Gao, F., Mulundu, G., Luo, C., N'Gandu, N., Sheppard, H., Hahn, B. H., Allen, S. & Frelinger, J. A. (1997). Cross-clade human immunodeficiency virus (HIV)-specific cytotoxic T-lymphocyte responses in HIV-infected Zambians. *J. Virol.* **71**, 8908-8911.

Bobkov, A., Kazennova, E., Selimova, L., Bobkov, M., Khanina, T., Ladnaya, N., Kravchenko, A., Pokrovsky, V., Cheingsong-Popov, R. & Weber, J. (1998). A sudden epidemic of HIV type 1 among injecting drug users in the Former Soviet Union: Identification of subtype A, subtype B, and novel *gag*A/*env*B recombinants. *AIDS Res. Hum. Retroviruses* **14**, 669-676.

Bredell, H., Hunt, G., Morgan, B., Tiemessen, C. T., Martin, D. J. & Morris, L. (1999). Identification of HIV-1 inter-subtype recombinants in South Africa using *env* and *gag* heteroduplex mobility assays. *AIDS Res. Hum. Retroviruses* (in press).

Bredell, H., Williamson, C., Sonnenberg, P., Martin, D. J. & Morris, L. (1998). Genetic characterization of HIV type 1 from migrant workers in three South African gold mines. *AIDS Res. Hum. Retroviruses* **14**, 677-684.

Brink, B. A. & Clausen, L. (1987). The acquired immune deficiency syndrome. *J. Mine Med. Officers Assoc. SA* **63**, 10-17.

Brodine, S. K., Mascola, J. R. & McCutchan, F. E. (1997). Genotypic variation and molecular epidemiology of HIV. *Infect. Med.* **14**, 739-748.

Brodine, S. K., Mascola, J. R., Weiss, P. J., Ito, S. I., Porter, K. R., Artenstein, A. W., Garland, F. C., McCutchan, F. E. & Burke, D. S. (1995). Detection of diverse HIV-1 genetic subtypes in the USA. *Lancet* **346**, 1198-1199.

Buonaguro, L., Del Gaudio, E., Monaco, M., Greco, D., Corti, P., Beth-Giraldo, E., Buonaguro, F. M., Giraldo, G. & Program, T. I.-U. C. A. (1995). Heteroduplex mobility assay and phylogenetic analysis of V3 region sequences of human immunodeficiency virus type 1 isolates from Gulu, Northern Uganda. *J. Virol.* **69**, 7971-7981.

Burton, D. R., Pyati, J., Koduri, R., Sharp, S. J., Thornton, G. B., Parren, P. W. H. I., Sawyer, L. S. W., Michael Hendry, R., Dunlop, N., Nara, P. L., Lamacchia, M., Garratty, E., Stiehm, E. R., Bryson, Y. J., Cao, Y., Moore, J. P., Ho, D. D. & Barbas III, C. F. (1994). Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science* **266**, 1024-1027.

Campbell, C. M. & Williams, B. G. (1996). Academic research and HIV/AIDS in South Africa. *S. Afr. Med. J.* **86**, 55-60.

Carr, J. K., Salminen, M. O., Albert, J., Sanders-Buell, E., Gotte, D., Birx, D. L. & McCutchan, F. E. (1998). Full genome sequences of human immunudeficiency virus type 1 subtypes G and A/G intersubtype recombinants. *Virology* **247**, 22-31.

Carr, J. K., Salminen, M. O., Koch, C., Gotte, D., Artenstein, A. W., Hegerich, P. A., St. Louis, D., Burke, D. S. & McCutchan, F. E. (1996). Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol.* **70**, 5935-5943.

Casement, K. S., Arlinghaus, R. B. & Sastry, K. J. (1996). Cytotoxic T-lymphocyte response induced by a V3 loop synthetic peptide from an African HIV-1 isolate is cross-reactive against HIV-1 strains from North America/Europe region. *AIDS* **10**, 1440-1441.

Cassol, S., Weniger, B. G., Babu, P. G., Salminen, M. O., Zheng, X., Htoon, M. T., Delaney, A., O'Shaughnessy, M. & Ou, C.-Y. (1996). Detection of HIV type 1 env subtypes A, B, C, and E in Asia using dried blood spots: a new surveillance tool for molecular epidemiology. *AIDS Res. Hum. Retroviruses* **12**, 1435-1441.

CDC. (1981a). Kaposi's sarcoma and *Pneumocystis* pneumonia among homosexual men-New York City and California. *MMWR* **30**, 305-308.

CDC. (1981b). *Pneumocystis* pneumonia-Los Angeles. *MMWR* **30**, 250-252.

CDC. (1996). Update: Provisional Public Health Services Recommendations for Chemoprophylaxis After Occupational Exposure to HIV. *MMWR* **45**, 468-472.

Chamber of Mines. (1993). Statistical Tables. Johannesburg.

Charneau, P., Borman, A. M., Quillent, C., Guetard, D., Chamaret, S., Cohen, J., Remy, G., Montagnier, L. & Clavel, F. (1994). Isolation and envelope sequence of a highly divergent HIV-1 isolate: Definition of a new HIV-1 group. *Virology* **205**, 247-253.

Cheingsong-Popov, R., Callow, D., Beddows, S., Shaunak, S., Wasi, C., Kaleebu, P., Gilks, C., Petrascu, I.V., Garaev, M.M, Watts, D.M., Constantine, N.T. & Weber, J.N. (1992). Geographic Diversity of Human Immunodeficiency Virus Type 1: Serologic Reactivity to *env* Epitopes and Relationship to Neutralization. *J. Infect. Dis.* **165**, 256-261.

Cheingsong-Popov, R., Osmanov, S., Pau, C. P., Schochetman, G., Barin, F., Holmes, H., Francis, G., Ruppach, H., Dietrich, U., Lister, S. & Weber, J. (1998a). Serotyping of HIV type 1 infections: definition, relationship to viral genetic subtypes, and assay evaluation. UNAIDS Network for HIV-1 Isolation and Characterization. *AIDS Res. Hum. Retroviruses* **14**, 311-318.

Cheingsong-Popov, R., Williamson, C., Lister, S., Morris, L., van Harmelen, J., Bredell, H., Wood, R., Sonnenberg, P., van der Ryst, E., Martin, D. & Weber, J. (1998b). Usefulness of HIV-1 V3 serotyping in studying the HIV-1 epidemic in South Africa [letter]. *AIDS* **12**, 949-950.

Choi, D. J., Dube, S., Spicer, T. P., Slade, H. B., Jensen, F. C. & Poiesz, B. J. (1997). HIV type 1 isolate Z321, the strain used to make a therapeutic HIV type 1 immunogen, is intersubtype recombinant. *AIDS Res. Hum. Retroviruses* **13**, 357-361.

Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M.-A., Santos-Ferreira, M. O., Laurent, A. G., Dauguet, C., Katlama, C., Rouzioux, C., Klatzmann, D., Champalimaud, J. L. & Montagnier, L. (1986). Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233**, 343-346.

Coffin, J., Haase, A., Levy, J. A., Montagnier, S., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Vardus, H., Vogt, P. & Weiss, R. (1986). Human immunodeficiency viruses [letter]. *Science* **232**, 697.

Coffin, J. M. (1992). Genetic diversity and evolution of retroviruses. In *Current Topics in Microbiology and Immunology*, pp. 143-164. Edited by J. J. Holland. Berlin, Heidelberg: Springer-Verlag.

Cohen, J. (1995). Differences in HIV strains may underlie disease patterns. *Science* **270**, 30-31.

Connor, R., Ho, D., Kuritzkes, D. & Richman, D. (1997). Human immunodeficiency viruses. In *Clinical Virology*, 1st edn, pp. 707-754. Edited by D. Richman, R. J. Whitley & F. G. Hayden. New York: Churchill Livingstone.

Cornelissen, M., Kampinga, G., Zorgdrager, F., Goudsmit, J. & Characterization, U. N. f. H. I. a. (1996). Human immunodeficiency virus type 1 subtypes defined by *env* show high frequency of recombinant *gag* genes. *J. Virol.* **70**, 8209-8212.

Csillag, C. (1994). HIV-1 subtype C in Brazil [letter]. *Lancet* **344**, 1354.

De Cock, K. M., Adjorlolo, G., Ekpini, E., Sibailly, T., Kouadio, J., Maran, M., Brattegaard, K., Vetter, K. M., Doorly, R. & Gayle, H. D. (1993). Epidemiology and transmission of HIV-2. Why there is no HIV-2 pandemic. *JAMA* **270**, 2083-2086.

De Cock, K. M. & Brun-Vezinet, F. (1989). Epidemiology of HIV-2. *AIDS* **3**, S89-S95.

De Leys, R., Vanderborght, B., Vanden Haesevelde, M., Heyndrickx, L., Van Geel, A., Wauters, C., Bernaerts, R., Saman, E., Nijs, P., Willems, B., Taelman, H., Van der Groen, G., Piot, P., Tersmette, T., Huisman, J. G. & Van Heuverswyn, H. (1990). Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *J. Virol.* **64**, 1207-1216.

Del Amo, J., Petruckevitch, A., Phillips, A., Johnson, A. M., Stephenson, J., Desmond, N., Hanscheid, T., Low, N., Newell, A., Obasi, A., Paine, K., Pym, A., Theodore, C. M. & De Kock, K. M. (1998). Disease progression and survival in HIV-1-infected Africans in London. *AIDS* **12**, 1203-1209.

Delwart, E. L., Shpaer, E. G., Louwagie, J., McCutchan, F. E., Grez, M., Rubsamen-Waigmann, H. & Mullins, J. I. (1993). Genetic relationships determined by a DNA heteroduplex mobility assay: Analysis of HIV-1 *env* genes. *Science* **262**, 1257-1261.

Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C., Schall, T. J., Littman, D. R. & Landau, N. R. (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**, 661-666.

Department of Health. (1999). 1998 National HIV sero-prevalence survey of women attending public antenatal clinics in South Africa.

Devito, C., Levi, M., Hinkula, J., Fernandez Medina, R. D., Libonatti, O. & Wigzell, H. (1998). Seroreactivity to HIV-1 V3 subtypes A to H peptides of Argentinian HIV-positive sera. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **17**, 156-159.

Dietrich, U., Grez, M., von Briesen, H., Panhans, B., Geissendorfer, M., Kuhnel, H., Maniar, J., Mahambre, G., Becker, W. B., Becker, M. L. B. & Rubsamen-Waigmann, H. (1993). HIV-1 strains from India are highly divergent from prototypic African and US/European strains, but are linked to a South African isolate. *AIDS* **7**, 23-27.

Dietrich, U., Maniar, J. K. & Rubsamen-Waigmann, H. (1995). The epidemiology of HIV in India. *Trends Microbiol.* **3**, 17-21.

Dighe, P. K., Korber, B. T. & Foley, B. T. (1997). Global variation in the HIV-1 V3 region. In *Human Retroviruses and AIDS: A compilation and analysis of nucleic acid and amino acid sequences*, pp. III74-III206. Edited by B. Korber, B. Hahn, B. Foley, J. W. Mellors, T. Leitner, G. Myers, F. McCutchan & C. Kuiken. New Mexico: Theoretical Biology and Biophysics.

Distler, O., McQueen, P. W., Tsang, M. L., Evans, L. A., Hurren, L., Byrne, C., Penny, R., Cooper, D. A. & Delaney, S. F. (1995). Primary structure of the V3 region of gp120 from sequential human immunodeficiency virus type 1 isolates obtained from patients from the time of seroconversion. *J. Infect. Dis.* **172**, 1384-1387.

Dittmar, M. T., Simmons, G., Hibbitts, S., O'Hare, M., Louisirirotchanakul, S., Beddows, S., Weber, J., Clapham, P. R. & Weiss, R. A. (1997). Langerhans cell tropism of human immunodeficiency virus type 1 subtype A through F isolates derived from different transmission groups. *J. Virol.* **71**, 8008-8013.

Dondero, T. J., Hu, D. J. & George, J. R. (1994). HIV-1 variants: yet another challenge to public health. *Lancet* **343**, 1376.

Dougherty, J. P. & Temin, H. M. (1988). Determination of the rate of base-pair substitution and insertion mutations in retrovirus replication. *J. Virol.* **62**, 2817-2822.

Doyle, P. R. (1991). The impact of AIDS on the South African population. In *AIDS in South Africa: The Demographic and Economic Implications*. Center for Health Policy, University of the Witwatersrand, Johannesburg, South Africa.

Dragic, T., Litwin, V., Allaway, G. P., Martin, S. R., Huang, Y., Nagashima, K. A., Cayanan, C., Maddon, P. J., Koup, R. A., Moore, J. P. & Paxton, W. A. (1996). HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381**, 667-673.

Engelbrecht, S., Koulinska, I., Smith, T.-L., Barreto, J. & Janse van Rensburg, E. (1998). Variation in HIV Type 1 V3 Region *env* Sequences from Mozambique. *AIDS Res. Hum. Retroviruses* **14**, 803-805.

Engelbrecht, S., Laten, J. D., Smith, T.-L. & Janse van Rensburg, E. (1995). Identification of *env* subtypes in fourteen HIV type 1 isolates from South Africa. *AIDS Res. Hum. Retroviruses* **11**, 1269-1271.

Engelbrecht, S., Smith, T.-L., Kasper, P., Faatz, E., Zeier, M., Moodley, D., Clay, C. G. & Van Rensburg, E. J. (1999). HIV Type 1 V3 Domain Serotyping and Genotyping in Gauteng, Mpumalanga, KwaZulu-Natal, and Western Cape Provinces of South Africa. *AIDS Res. Hum. Retroviruses* **15**, 325-328.

Fauci, A. S. (1988). The human immunodeficiency virus: infectivity and mechanisms of pathogenesis. *Science* **239**, 617-622.

Felsenstein, J. (1993). PHYLIP -Phylogeny Inference Package (Version 3.5c). Washington: University of Washington.

Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. (1996). HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872-877.

Foley, B. & Korber, B. (1995). Global variation in the HIV-1 V3 region. In *Human Retroviruses and AIDS: A compilation and analysis of nucleic acid and amino acid sequences*, pp. III70-III145. Edited by G. Myers, B. H. Hahn, J. W. Mellors, L. E. Henderson, B. Korber, K.-T. Jeang, F. E. McCutchan & G. N. Pavlakis. New Mexico: Theoretical Biology and Biophysics.

Forum, S. S. (1999). HIV Sequence Database. Edited by http://hiv-web.lanl.gov.

Fouchier, R. A. M., Groenink, M., Kootstra, N. A., Tersmette, M., Huisman, H. G., Miedema, F. & Schuitemaker, H. (1992). Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp 120 molecule. *J. Virol.* **66**, 3183-3187.

Freed, E. O., Delwart, E. L., Buchschacher Jr., G. L. & Panganiban, A. T. (1992). A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity. *Proc. Natl. Acad. Sci. USA* **89**, 70-74.

Freed, E. O., Myers, D. J. & Risser, R. (1991). Identification of the principal neutralizing determinant of human immunodeficiency virus type 1 as a fusion domain. *J. Virol.* **65**, 190-194.

Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M. & Hahn, B. H. (1999). Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* **397**, 436-441.

Gao, F., Robertson, D. L., Carruthers, C. D., Li, Y., Bailes, E., Kostrikis, L. G., Salminen, M. O., Bibollet-Ruche, F., Peeters, M., Ho, D. D., Shaw, G. M., Sharp, P. M. & Hahn, B. H. (1998a). An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). *J. Virol.* **72**, 10234-10241.

Gao, F., Robertson, D. L., Carruthers, C. D., Morrison, S. G., Jian, B., Chen, Y., Barre-Sinoussi, F., Girard, M., Srinivasan, A., Abimiku, A. G., Shaw, G. M., Sharp, P. M. & Hahn, B. H. (1998b). A Comprehensive Panel of Near-Full-Length Clones and Reference Sequences for Non-Subtype B Isolates of Human Immunodeficiency Virus Type 1. *J. Virol.* **72**, 5680-5698.

Gao, F., Robertson, D. L., Morrison, S. G., Hui, H., Craig, S., Decker, J., Fultz, P. N., Girard, M., Shaw, G. M., Hahn, B. H. & Sharp, P. M. (1996). The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virol.* **70**, 7013-7029.

Gao, F., Yue, L., Robertson, D. L., Hill, S. C., Hui, H., Biggar, R. J., Neequaye, A. E., Whelan, T. M., Ho, D. D., Shaw, G. M., Sharp, P. M. & Hahn, B. H. (1994). Genetic diversity of human immunodeficiency virus type 2: Evidence for distinct sequence subtypes with differences in virus biology. *J. Virol.* **68**, 7433-7447.

Gelderblom, H. R., Hausmann, E. H. S., M., O., Pauli, G. & Koch, M. A. (1987). Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. *Virology* **157**, 171-176.

Goedert, J. J., Duliege, A.-M., Amos, C. I., Felton, S. & Biggar, R. J. (1991). High risk of HIV-1 infection for first-born twins. *Lancet* **338**, 1471-1475.

Goff, S. P. (1990). Retroviral reverse transcriptase: synthesis, structure, and function. *J. Acquir. Immune Defic. Syndr.* **3**, 817-831.

Goodenow, M., Huet, T., Saurin, W., Kwok, S., Sninsky, J. & Wain-Hobson, S. (1989). HIV-1 isolates are rapidly evolving quasispecies: Evidence for viral mixtures and preferred nucleotide substitutions. *J. Acquir. Immune Defic. Syndr.* **2**, 344-352.

Gray, C. M., Morris, L., Murray, J., Keeton, J., Shalekoff, S., Lyons, S. F., Sonnenberg, P. & Martin, D. J. (1996). Identification of cell subsets expressing intracytoplasmic cytokines within HIV-1-infected lymph nodes. *AIDS* **10**, 1467-1475.

Green, W. C. (1991). The molecular biology of human immunodeficiency virus type 1 infection (Review). *N. Engl. J. Med.* **324**, 308-317.

Grez, M., Dietrich, U., Balfe, P., Von Briesen, H., Maniar, J. K., Mahambre, G., Delwart, E. L., Mullins, J. I. & Rubsamen-Waigmann, H. (1994). Genetic analysis of human immunodeficiency virus type 1 and 2 (HIV-1 and HIV-2) mixed infections in India reveals a recent spread of HIV-1 and HIV-2 from a single ancestor for each of these viruses. *J. Virol.* **68**, 2161-2168.

Gurtler, L., Hauser, P. H., Eberle, J., Von Brunn, A., Knapp, S., Zenkeng, L., Tsague, J. M. & Kaptue, L. (1994). A new subtype of human immunodeficiency virus type 1 (MVP-5180) fro Cameroon. *J. Virol.* **68**, 1581-1585.

Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. & Alizon, M. (1987). Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature* **326**, 662-669.

Hahn, B. H., Shaw, G. M., Taylor, M. E., Redfield, R. R., Markham, P. D., Salahuddin, S. Z., Wong-Staal, F., Gallo, R. C., Parks, E. S. & Parks, W. P. (1986). Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science* **232**, 1548-1553.

Hehlmann, R. (1991). Human Retroviruses. In *Textbook of Human Virology*, 2nd edn, pp. 274-306. Edited by R. B. Belshe. St. Louis, Missouri: Mosby Year Book.

Heyndrickx, L., Alary, M., Janssens, W., Davo, N. & Van der Groen, G. (1996). HIV-1 group O and group M dual infection in Benin [letter]. *Lancet* **347**, 902-903.

Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M. & Markowitz, M. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123-126.

Hoelscher, M., Hanker, S., Barin, F., Cheingsong-Popov, R., Dietrich, U., Jordan-Harder, B., Olaleye, D., Nagele, E., Markuzzi, A., Mwakagile, D., Minja, F., Weber, J., Gurtler, L. & Von Sonnenburg, F. (1998). HIV type 1 V3 serotyping of Tanzanian samples: probable reasons for mismatching with genetic subtyping. *AIDS Res. Hum. Retroviruses* **14**, 139-149.

Holmes, E. C., Zhang, L. Q., Simmonds, P., Ludlam, C. A. & Brown, A. J. L. (1992). Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc. Natl. Acad. Sci. USA* **89**, 4835-4839.

Horsburgh Jr., C.R. & Holmberg, S.D. (1988). The global distribution of human immunodeficiency virus type 2 (HIV-2) infection. *Transfusion* **28**, 192-195.

Htun, Y., Bredell, H., Martin, D., Dangor, Y., Radebe, F., Tshabala, V. & Ballard, R. (1999). Influence of reproductive tract infections and their treatment on vaginal shedding of HIV. *13th Meeting of the International Society for Sexually Transmitted Diseases Research.* Denver, July 11-14.

Hu, D.J., Dondero, T.J., Rayfield, M.A., George, R., Schochetman, G., Jaffe, H.W., Luo, C.-C., Kalish, M.L., Weniger, B.G., Pau, C.-P., Schable, C.A. 7 Curran, J.W. (1996). The emerging genetic diversity of HIV. *JAMA* **275**, 210-216.

Hu, W.-S. & Temin, H. M. (1990). Genetic consequences of packaging two RNA genomes in one retroviral particle: psuedodiploidy and high rate of genetic recombination. *Proc. Natl. Acad. Sci. USA* **87**, 1556-1560.

Husson, R. N., Lan, Y., Kojima, E., Venzon, D., Mitsuya, H. & McIntosh, K. (1995). Vertical transmission of human immunodeficiency virus type 1: Autologous neutralizing antibody, virus load, and virus phenotype. *J. Pediatr.* **126**, 865-871.

Hwang, S. S., Boyle, T. J., Lyerly, H. K. & Cullen, B. R. (1991). Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* **253**, 71-74.

Imrie, A., Beveridge, A., Genn, W., Vizzard, J. & Cooper, D. A. (1997). Transmission of human immunodeficiency virus type 1 resistant to nevirapine and zidovudine: Sydney Primary HIV Infection Group. *J. Infect. Dis.* **175**, 1502-1506.

Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J. & Varmus, H. E. (1988). Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. *Nature* **331**, 280.

Jain, M. K., John, T. J. & Keusch, G. T. (1994). A review of human immunodeficiency virus infection in India. *J. Acquir. Immune Defic. Syndr.* **7**, 1185-1194.

Janssens, W., Buve, A. & Nkengasong, J. N. (1997). The puzzle of HIV-1 subtypes in Africa [editorial]. *AIDS* **11**, 705-712.

Janssens, W., Heyndrickx, L., Fransen, K., Motte, J., Peeters, M., Nkengasong, J. N., Ndumbe, P. M., Delaporte, E., Perret, J.-L., Atende, C., Piot, P. & Van der Groen, G. (1994). Genetic and phylogenetic analysis of *env* subtypes G and H in Central Africa. *AIDS Res. Hum. Retroviruses* **10**, 877-879.

Javaherian, K., Langlois, A. J., McDanal, C., Ross, K. L., Eckler, L. I., Jellis, C. L., Profy, A. T., Rusche, J. R., Bolognesi, D. P., Putney, S. D. & Matthews, T. J. (1989). Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. *Proc. Natl. Acad. Sci. USA* **86**, 6768-6772.

Jochelson, K., Mothibeli, M. & Leger, J.-P. (1991). Human immunodeficiency virus and migrant labor in South Africa. *Int. J. Health Serv.* **21**, 157-173.

Kalish, M. L., Baldwin, A., Raktham, S., Wasi, C., Luo, C.-C., Schochetman, G., Mastro, T. D., Young, N. L., Vanichseni, S., Rubsamen-Waigmann, H., Von Briesen, H., Mullins, J. I., Delwart, E., Herring, B., Esparza, J., Heyward, W. L. & Osmanov, S. (1995). The evolving molecular epidemiology of HIV-1 envelope subtypes in injecting drug users in Bangkok, Thailand: implications for HIV vaccine trials. *AIDS* **9**, 851-857.

Kampinga, G. A., Simonon, A., Van de Perre, P., Karita, E., Msellati, P. & Goudsmit, J. (1997). Primary infection with HIV-1 and their offspring in Rwanda: Findings of heterogeneity at seroconversion, coinfection, and recombinants of HIV-1 subtypes A and C. *Virology* **227**, 63-76.

Kanki, P. (1991). Biologic features of HIV-2. An update. In *AIDS Clinical Review*, pp. 17-38. Edited by P. Vilberding & M. A. Jacobson. New York: Marcel Dekker.

Kanki, P. J., Hamel, D. J., Sankale, J. L., Hsieh, C., Thior, I., Barin, F., Woodcock, S. A., Gueye-Ndiaye, A., Zhang, E., Montano, M., Siby, T., Marlink, R., NDoye, I., Essex, M. E. & MBoup, S. (1999). Human immunodeficiency virus type 1 subtypes differ in disease progression. *J. Infect. Dis.* **179**, 68-73.

Kanki, P. J., Travers, K. U., MBoup, S., Hsieh, C. C., Marlink, R. G., Gueye-NDiaye, A., Siby, T., Thior, I., Hernandez-Avila, M., Sankale, J.-L., NDoye, I. & Essex, M. E. (1994). Slower heterosexual spread of HIV-2 than HIV-1. *Lancet* **343**, 943-946.

Kaplan, A. H., Manchester, M. & Swanstrom, R. (1994). The activity of the protease of human immunodeficiency virus type 1 is initiated at the membrane of infected cells before the release of viral proteins and is required for release to occur with maximum efficiency. *J. Virol.* **68**, 6782-6786.

Kimura, M. (1980). A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **19**, 111-120.

Kostrikis, L. G., Bagdades, E., Cao, Y., Zhang, L., Dimitriou, D. & Ho, D. D. (1995). Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: Identification of a new subtype designated subtype I. *J. Virol.* **69**, 6122-6130.

Kostrikis, L. G., Shin, S. & Ho, D. D. (1998). Genotyping HIV-1 and HCV strains by a combinatorial DNA melting assay (COMA). *Mol. Med.* **4**, 443-453.

Kuiken, C. L., Zwart, G., Baan, E., Coutinho, R. A., Van den Hoek, J. A. R. & Goudsmit, J. (1993). Increasing antigenic and genetic diversity of the V3 variable domain of the human immunodeficiency virus envelope protein in the course of the AIDS epidemic. *Proc. Natl. Acad. Sci. USA* **90**, 9061-9065.

Kunanusont, C., Foy, H. M., Kreiss, J. K., Rerks-Ngarm, S., Phanuphak, P., Raktham, S., Pau, C. P. & Young, N. L. (1995). HIV-1 subtypes and male-to-female transmission in Thailand. *Lancet* **345**, 1078-1083.

Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648-659.

Lalvani, A. & Shastri, J. S. (1996). HIV epidemic in India: opportunity to learn from the past. *Lancet* **347**, 1349-1350.

Leitner, T. (1996). Genetic subtypes of HIV-1. In *Human Retroviruses and AIDS 1996: A compilation and analysis of nucleic acid and amino acid sequences*, pp. III28-III40. Edited by G. Myers, B. Foley, J. W. Mellors, B. Korber, K.-T. Jeang & S. Wain-Hobson. Los Alamos, NM: Theoretical Biology and Biophysics.

Leitner, T., Alaeus, A., Marquina, S., Lilja, E., Lidman, K. & Albert, J. (1995a). Yet another subtype of HIV type 1. *AIDS Res. Hum. Retroviruses* **11**, 995-997.

Leitner, T., Escanilla, D., Marquina, S., Wahlberg, J., Brostrom, C., Hansson, H. B., Uhlen, M. & Albert, J. (1995b). Biological and molecular characterization of subtype D, G, and A/D recombinant HIV-1 transmissions in Sweden. *Virology* **209**, 136-146.

Leitner, T., Korber, B., Robertson, D., Gao, F. & Hahn, B. (1997). Updated proposal of reference sequences of HIV-1 genetic subtypes. In *Human Retroviruses and AIDS: A compilation and analysis of nucleic acid and amino acid sequences*, pp. III19-III24. Edited by B. Korber, B. Foley, T. Leitner, F. McCutchan, B. Hahn, J. W. Mellors, G. Myers & C. Kuiken. New Mexico: Theoretical Biology and Biophysics.

Levy, J. A. (1994). Features of HIV transmission and acute infection. In *HIV and the pathogenesis of AIDS*, 1st edn, pp. 11-33. Edited by J. A. Levy. Washington, D.C.: American Society for Microbiology Press.

Levy, J. A. (1996). Infection by human immunodeficiency virus - CD4 is not enough. *N. Engl. J. Med.* **335**, 1528-1530.

Levy, J. A., Mitra, G. A., Wong, M. F. & Mozen, M. M. (1985). Inactivation by wet and dry heat procedures of AIDS- associated retrovirus (ARV) during factor VIII purification from plasma. *Lancet* **i**, 1456-1457.

Lin, H. J., Siwak, E. B., Lauder, I. J. & Blaine Hollinger, F. (1995). Single-strand conformation polymorphism study of human immunodeficiency virus type 1 RNA and DNA in plasma, peripheral blood mononuclear cells, and their virologic cultures. *J. Infect. Dis.* **171**, 1619-1622.

Lin, H.J., Siwak, E.B., Lauder, I.J. & Hollinger, F.B. (1998). Selection of appropriate HIV-1 genomic regions for single-strand conformation polymorphism analysis of the diversity, modification, and transmission of HIV-1 quasispecies. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **18**, 409-416.

Loussert-Ajaka, I., Chaix, M.-L., Korber, B., Letourneur, F., Gomas, E., Allen, E., Ly, T.-D., Brun-Vezinet, F., Simon, F. & Saragosti, S. (1995). Variability of human immunodeficiency virus type 1 group O strains isolated from Cameroonian patients living in France. *J. Virol.* **69**, 5640-5649.

Louwagie, J., Janssens, W., Mascola, J., Heyndrickx, L., Hegerich, P., van der Groen, G., McCutchan, F. E. & Burke, D. S. (1995). Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. *J. Virol.* **69**, 263-271.

Louwagie, J., McCutchan, F. E., Peeters, M., Brennan, T. P., Sanders-Buell, E., Eddy, G. A., Van der Groen, G., Fransen, K., Gershy-Damet, G.-M., Deleys, R. & Burke, D. S. (1993). Phylogenetic analysis of *gag* genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS* **7**, 769-780.

Lukashov, V. V., Cornelissen, M. T. E., Goudsmit, J., Papuashvilli, M. N., Rytik, P. G., Khaitov, R. M., Karamov, E. V. & De Wolf, F. (1995). Simultaneous introduction of distinct HIV-1 subtypes into different risk groups in Russia, Byelorussia and Lithuania. *AIDS* **9**, 435-439.

Luo, C. C., Tian, C., Hu, D. J., Kai, M., Dondero, T. & Zheng, X. (1995). HIV-1 subtype C in China [letter]. *Lancet* **345**, 1051-1052.

Luo, C.-C., Downing, R. G., Dela Torre, N., Baggs, J., Hu, D. J., Respess, R. A., Candal, D., Carr, L., George, J. R., Dondero, T. J., Biryahwaho, B. & Rayfield, M. A. (1998). The Development and Evaluation of a Probe Hybridization Method for Subtyping HIV Type 1 Infection in Uganda. *AIDS Res. Hum. Retroviruses* **14**, 691-694.

Lynch, C. E., Madej, R., Louie, P. H. & Rodgers, G. (1992). Detection of HIV-1 DNA by PCR: evaluation of primer pair concordance and sensitivity of a single primer pair. *J. Acquir. Immune Defic. Syndr.* **5**, 433-440.

Lyons, S. F., Smith, A. N., McGillivray, G. M. & Schoub, B. D. (1988). HIV-2 infection in South Africa. *Trans. R. Soc. Trop. Med. Hyg.* **82**, 757.

Maddon, P. J., Dalgleish, A. G., McDougal, J. S., Clapham, P. R., Weiss, R. A. & Axel, R. (1986). The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47**, 333-348.

Markovitz, D. M. (1993). Infection with the human immunodeficiency virus type 2 [Review]. *Ann. Intern. Med.* **118**, 211-218.

Mascola, J. R., Louwagie, J., McCutchan, F. E., Fischer, C. L., Hegerich, P. A., Wagner, K. F., Fowler, A. K., McNeil, J. G. & Burke, D. S. (1994). Two antigenically distinct subtypes of human immunodeficiency virus type 1: viral genotype predicts neutralization serotype. *J. Infect. Dis.* **169**, 48-54.

Masquelier, B., Burgard, M., Doussin, A., Simon, F., Cottalorda, J., Izopet, J., Tamalet, C., Fluery, H., Mayaux, M. J., Blanche, S. & Rouzioux, C. (1999). Zidovudine Genotypic Resistance in HIV-1 Infected Newborns in the French HIV-1 Perinatal Cohort Study. *6th Conference on Retroviruses and Opportunistic Infections*. Chicago, Jan 31-Feb 4.

Matsushita, S., Robert-Guroff, M., Rusche, J., Koito, A., Hattori, T., Hoshino, H., Javaherian, K., Takatsuki, K. & Putney, S. (1988). Characterization of a human immunodeficiency virus neutralizing monoclonal antibody and mapping of the neutralizing epitope. *J. Virol.* **62**, 2107-2114.

McCune, J. M., Rabin, L. B., Feinberg, M. B., Lieberman, M., Kosek, J. C., Reyes, G. R. & Weissman, I. L. (1988). Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. *Cell* **53**, 55-67.

McCutchan, F. E., Hegerich, P. A., Brennan, T. P., Phanuphak, P., Singharaj, P., Jugsudee, A., Berman, P. W., Gray, A. M., Fowler, A. K. & Burke, D. S. (1992a). Genetic variants of HIV-1 in Thailand. *AIDS Res. Hum. Retroviruses* **8**, 1887-1895.

McCutchan, F. E., Salminen, M. O., Carr, J. K. & Burke, D. S. (1996). HIV-1 genetic diversity. *AIDS* **10**, S13-S20.

McCutchan, F. E., Ungar, B. L. P., Hegerich, P., Roberts, C. R., Fowler, A. K., Hira, S. K., Perine, P. L. & Burke, D. S. (1992b). Genetic analysis of HIV-1 isolates from Zambia and an expanded phylogenetic tree for HIV-1. *J. Acquir. Immune Defic. Syndr.* **5**, 441-449.

McKeating, J. A., Gow, J., Goudsmit, J., Pearl, L. H., Mulder, C. & Weiss, R. A. (1989). Characterization of HIV-1 neutralization escape mutants. *AIDS* **3**, 777-784.

Montagnier, L. & Clavel, F. (1994). Human immunodeficiency viruses. In *Encyclopedia of Virology*, 1st edn, pp. 674-681. Edited by R. G. Webster & A. Granoff. London: Academic Press.

Montagnier, L., Gruest, J., Chamaret, S., Dauguet, C., Axler, C., Guetard, D., Nugeyre, M. T., Barre-Sinoussi, F., Chermann, J. C. & Brunet, J. B. (1984). Adaption of lymphadenopathy associated virus (LAV) to replication in EBV-transformed B lymphoblastoid cell lines. *Science* **225**, 63-66.

Moore, J. P., Cao, Y., Leu, J., Qin, L., Korber, B. & Ho, D. D. (1996). Inter- and intraclade neutralization of human immunodeficiency virus type 1: genetic clades do not correspond to neutralization serotypes but partially correspond to gp120 antigenic serotypes. *J. Virol.* **70**, 427-444.

Morgado, M. G., Sabino, E. C., Shpaer, E. G., Bongertz, V., Brigido, L., Guimaraes, M. D. C., Castilho, E. A., Galvao-Castro, B., Mullins, J. I., Hendry, R. M. & Mayer, A. (1994). Polymorphism in the V3 region of the envelope protein of HIV-1 in Brazil: divergence from prevalent North American/European subtype B strains and identification of newly described subtype F. *AIDS Res. Hum. Retroviruses* **10**, 569-576.

Morris, L., Cilliers, T., Phoswa, M. & Martin, D. J. (1999). No evidence for change in co-receptor use in progressing patients infected with HIV-1 subtype C (submitted).

Murphy, E., Korber, B., Georges-Courbot, M. C., You, B., Pinter, A., Cook, D., Kieny, M. P., Georges, A., Mathiot, C., Barre-Sinoussi, F. & Girard, M. (1993). Diversity of V3 region sequences of human immunodeficiency viruses type 1 from the central African Republic. *AIDS Res. Hum. Retroviruses* **9**, 997-1006.

Myers, G. (1994). Tenth anniversary perspectives on AIDS. HIV: Between past and future. *AIDS Res. Hum. Retroviruses* **10**, 1317-1324.

Myers, G., Berzofsky, J. A., Korber, B., Smith, R. F. & Pavlakis, G. N. (1992a). Analysis. In *Human Retroviruses and AIDS 1992: A compilation and analysis of nucleic acid and amino acid sequences*, pp. III2-III3. Los Alamos, NM: Theoretical Biology and Biophysics.

Myers, G., Berzofsky, J. A., Korber, B., Smith, R. F. & Pavlakis, G. N. (1992c). HIV and SIV Nucleic Acid Sequences. In *Human Retroviruses and AIDS 1992: A compilation and analysis of nucleic acid and amino acid sequences*, pp. IA9-IA117. Los Alamos, NM: Theoretical Biology and Biophysics.

Myers, G., Berzofsky, J. A., Korber, B., Smith, R. F. & Pavlakis, G. N. (1992b). HIV and SIV Nucleic Acid Sequences. In *Human Retroviruses and AIDS 1992: A compilation and analysis of nucleic acid and amino acid sequences*, pp. IA76-IA99. Los Alamos, NM: Theoretical Biology and Biophysics.

Myers, G., Berzofsky, J. A., Rabson, A. B. & Smith, T. F. (1990). Analysis. In *Human Retroviruses and AIDS 1990: A compilation and analysis of nucleic acid and amino acid sequences*, pp. III-8. Los Alamos, NM: Theoretical Biology and Biophysics.

Myers, G., MacInnes, K. & Myers, L. (1993). Phylogenetic moments in the AIDS epidemic. In *Emerging viruses*, pp. 120-137. Edited by S. S. Morse. Oxford: Oxford University Press.

Myers, G., Wain-Hobson, S., Henderson, L.-E., Korber, B., Jeang, K.-T. & Pavlakis, G. N. (1994). Analysis. In *Human Retroviruses and AIDS 1994: A compilation and analysis of nucleic acid and amino acid sequences*, pp. III26-III27. Los Alamos, NM: Theoretical Biology and Biophysics.

Nahmias, A. J., Weiss, J., Yao, X., Lee, F., Kodsi, R., Schanfield, M., Matthews, T., Bolognesi, D., Durack, D., Motulsky, A., Kanki, P. & Essex, M. (1986). Evidence for human infection with an HTLV III/LAV-like virus in Central Africa, 1959 [letter]. *Lancet* **1**, 1279-1280.

Newell, M. L. & Peckham, C. (1993). Risk factors for vertical transmission of HIV-1 and early markers of HIV-1 infection in children. *AIDS* 7, S91-S97.

Nkengasong, J. N., Peeters, M., vanden Haesevelde, M., Musi, S. S., Willems, B., Ndumbe, P. M., Delaporte, E., Perret, J. L., Piot, P. & van den Groen, G. (1993). Antigenic evidence of the presence of the aberrant HIV-1 ant70 virus in Cameroon and Gabon [letter]. *AIDS* 7, 1536-1538.

Nkengasong, J. N., Peeters, M., Zhong, P., Willems, B., Janssens, W., Heyndrickx, L., Fransen, K., Ndumbe, P. M., Gershy-Damet, G.-M., Nys, P., Kestens, L., Piot, P. & Van der Groen, G. (1995). Biological phenotypes of HIV-1 subtypes A and B strains of diverse origins. *J. Med. Virol.* 47, 278-284.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sekiya, T. (1989a). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* 86, 2766-2770.

Orita, M., Suzuki, Y., Sekiya, T. & Hayashi, K. (1989b). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* 5, 874-879.

Ou, C. Y., Ciesielski, C. A., Myers, G., Bandea, C. I., Luo, C. C., Korber, B. T., Mullins, J. I., Schochetman, G., Berkelman, R. L., Economou, A. N., Witte, J. J., Furman, L. J., Satten, G. A., MacInnes, K. A., Curran, J. W., Jaffe, H. W., Group, L. I. & Group, E. I. (1992). Molecular epidemiology of HIV transmission in a dental practice. *Science* 256, 1165-1171.

Ou, C.-Y., Takebe, Y., Weniger, B., Luo, C.-C., Kalish, M. L., Auwanit, W., Yamazaki, S., Gayle, H. D., Young, N. L. & Schochetman, G. (1993). Independent introduction of two major HV-1 genotypes into distinct high-risk populations in Thailand. *Lancet* 341, 1171-1174.

Ozel, M., Pauli, G. & Gelderblom, H. R. (1988). The organization of the envelope projections on the surface of HIV. *Arch. Virol.* 100, 255.

Padian, N., Shiboski, ·., Glass, S. & Vittinghoff, E. (1997). Heterosexual transmission of human immunodeficiency virus (HIV) in Northern California: results from a ten-year study. *Am. J. Epidemiol.* 146, 350-357.

Palella Jr , F.J., Delaney, K.M., Moorman, A.C., Loveless, M.O., Fuhrer, J., Satten, G.A., Aschman, D.J. & Holmberg, S.D. (1998). Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *N. Engl. J. Med.* 338, 853-860.

Pape, J. W., Stanback, M. E., Pamphile, M., Boncy, M., Deschamps, M. M., Verdier, R. I., Beaulieu, M. E., Blattner, W., Liautaud, B. & Johnson Jr., W. D. (1990). Prevalence of HIV infection and high-risk activities in Haiti. *J. Acquir. Immune Defic. Syndr.* 3, 995-1001.

Pathak, V. K. & Temin, H. (1990a). Broad spectrum of *in vitro* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc. Natl. Acad. Sci. USA* 87, 6019-6023.

Pathak, V. K. & Temin, H. M. (1990b). Broad spectrum of *in vitro* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. *Proc. Natl. Acad. Sci. USA* 87, 6024-6028.

Pau, C.-P., Lee-Thomas, S., Auwanit, W., George, J. R., Ou, C.-Y., Parekh, B. S., Granade, T. C., Hoiloman, D. L., Phillips, S., Schochetman, G., Young, N. L., Takebe, Y., Gayle, H. D. & Weniger, B. G. (1993). Highly specific V3 peptide enzyme immunoassay for serotyping HIV-1 specimens from Thailand. *AIDS* 7, 337-340.

Peckham, C. & Gibb, D. (1995). Mother-to-child transmission of the human immunodeficiency virus. *N. Engl. J. Med.* **333**, 298-302.

Peeters, M., Gaye, A., Mboup, S., Badombena, W., Bassabi, K., Prince-David, M., Develoux, M., Liegeois, F., van der Groen, G., Saman, E. & Delaporte, E. (1996). Presence of HIV-1 group O infection in West Africa. *AIDS* **10**, 343-344.

Peeters, M., Vincent, R., Perret, J.-L., Latsky, M., Patrel, D., Liegeois, F., Courgnaud, V., Seng, R., Matton, T., Molinier, S. & Delaporte, E. (1999). Evidence for differences in MT2 cell tropism according to genetic subtypes of HIV-1: syncitium-inducing variants seem rare among subtype C HIV-1 viruses. *J. Acquir. Immune Defic. Syndr.* **20**, 115-121.

Pepin, J., Morgan, G., Dunn, D., Gevao, S., Mendy, M., Gaye, I., Scollen, N., Tedder, R. & Whittle, H. (1991). HIV-2 induced immunosuppression among asymptomatic West African prostitutes: evidence that HIV-2 is pathogenic, but less so than HIV-1. *AIDS* **5**, 1165-1172.

Phillips, R. E., Rowland-Jones, S., Nixon, D. F., Gotch, F. M., Edwards, J. P., Ogunlesi, A. O., Elvin, J. G., Rothbard, J. A., Bangham, C. R. M., Rizza, C. R. & McMichael, A. J. (1991). Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354**, 453-459.

Piatak Jr., M.,Saag, M.S.,Yang, L.C.,Clark, S.J.,Kappes, J.C.,Luk, K.C.,Hahn, B.H.,Shaw, G.M. & Lifson, J.D. (1993). High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* **259**, 1749-1754.

Pieniazek, D., Janini, L. M., Ramos, A., Tanuri, A., Schechter, M., Peralta, J. M., Vicente, A. C. P., Pieniazel, N. J., Schochetman, G. & Rayfield, M. A. (1995). HIV-1 patients may harbour viruses of different phylogenetic subtypes: Implications for the evolution of the HIV/AIDS pandemic. *Emerg. Infect. Dis.* **1**, 86-88.

Pope, M., Frankel, S. S., Mascola, J. R., Trkola, A., Isdell, F., Birx, D. L., Burke, D. S., Ho, D. H. & Moore, J. P. (1997). Human immunodeficiency virus type 1 strains of subtypes B and E replicate in cutaneous dendritic cell--T-cell mixtures without displaying subtype-specific tropism. *J. Virol.* **71**, 8001-8007.

Ras, G. J., Simson, I. W., Anderson, R., Prozesky, O. W. & Hamersma, T. (1983). Acquired immunodeficiency syndrome. A report of 2 South African cases. *S. Afr. Med. J.* **64**, 140-142.

Registro de casos de SIDA. Actualizacion a Junio de 1993. (1993). Madrid: Ministerio der Sanidad y Consumo.

Ricchetti, M. & Buc, H. (1990). Reverse transcriptases and genomic variability: the accuracy of DNA replication is enzyme specific and sequence dependent. *EMBO J.* **9**, 1583-1593.

Richman, D. & Bozzette, S. (1994). The impact of the syncitium-inducing phenotype of human immunodeficiency virus on disease progression. *J. Infect. Dis.* **169**, 968-974.

Richman, D. D. (1992). HIV drug resistance. *AIDS Res. Hum. Retroviruses* **8**, 1065-1071.

Robertson, D. L., Gao, F., Hahn, B. H. & Sharp, P. M. (1997). Intersubtype recombinant HIV-1 sequences. In *Human Retroviruses and AIDS: A compilation and analysis of nucleic acid and amino acid sequences*, pp. III25-III30. Edited by B. Korber, B. Foley, T. Leitner, F. McCutchan, B. Hahn, J. W. Mellors, G. Myers & C. Kuiken. Los Alamos, NM: Theoretical Biology and Biophysics.

Robertson, D. L., Hahn, B. H. & Sharp, P. M. (1995a). Recombination in AIDS viruses. *J. Mol. Evol.* **40**, 249-259.

Robertson, D. L., Sharp, P. M., McCutchan, F. E. & Hahn, B. H. (1995b). Recombination in HIV-1. *Nature* **374**, 124-126.

Royce, R. A., Sena, A., Cates, W. & Cohen, M. S. (1997). Sexual Transmission of HIV. *N. Engl. J. Med.* **336**, 1072-1078.

Rubsamen-Waigmann, H., Maniar, J., Gerte, S., Brede, H. D., Dietrich, U., Mahambre, G. & Pfutzner, A. (1994). High proportion of HIV-2 and HIV-1/2 double-reactive sera in two Indian states, Maharashtra and Goa: first appearance of an HIV-2 epidemic along with an HIV-1 epidemic outside of Africa. *Int. J. Med. Microbiol. Virol. Parasitol. Infect. Dis.* **280**, 398-402.

Sabino, E. C., Shpaer, E. G., Morgado, M. G., Korber, B. T. M., Diaz, R. S., Bongertz, V., Cavalcante, S., Galvao-Castro, B., Mullins, J. I. & Mayer, A. (1994). Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiological linked individuals from Brazil. *J. Virol.* **68**, 6340-6346.

Salminen, M. O., Carr, J. K., Robertson, D. L., Hegerich, P., Gotte, D., Koch, C., Sanders-Buell, E., Gao, F., Sharp, P. M., Hahn, B. H., Burke, D. S. & McCutchan, F. E. (1997). Evolution and probable transmission of intersubtype recombinant human immunodeficiency virus type 1 in a Zambian couple. *J. Virol.* **71**, 2647-2655.

Salminen, M. O., Koch, C., Sanders-Buell, E., Ehrenberg, P. K., Michael, N. L., Carr, J. K., Burke, D. S. & McCutchan, F. E. (1995). Recovery of virtually full-length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase chain reaction. *Virology* **213**, 80-86.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989a). Appendix E: Commonly used techniques in molecular cloning. In *Molecular cloning: A laboratory manual*, 2nd edn, pp. E.1-E.39. Edited by N. Ford, C. Nolan & M. Ferguson. New York: Cold Spring Harbor Laboratory Press.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989b). DNA sequencing. In *Molecular cloning: A laboratory manual*, 2nd edn, pp. 13.1-13.104. Edited by N. Ford, C. Nolan & M. Ferguson. New York: Cold Spring Harbour Laboratory Press.

Schable, C., Zekeng, L., Pau, C. P., Hu, D., Kaptue, L., Gurtler, L., Dondero, T., Tsague, J. M., Schochetman, G., Jaffe, H. & George, J. R. (1994). Sensitivity of United States HIV antibody tests for detection of HIV-1 group O infections. *Lancet* **344**, 1333-1334.

Schoub, B. D. (1994). The transmission of HIV infection. In *AIDS & HIV in perspective: A guide to understanding the virus and its consequences*, 1st edn, pp. 91-124. Edited by B. D. Schoub. Cambridge: Cambridge University Press.

Schoub, B. D., Smith, A. N., Johnson, S., Martin, D. J., Lyons, S. F., Padayachee, G. N. & Hurwitz, H. S. (1990). Considerations on the further expansion of the AIDS epidemic in South Africa--1990. *S. Afr. Med. J.* **77**, 613-618.

Sher, R. (1986). Acquired immune deficiency syndrome (AIDS) in the RSA. *S. Afr. Med. J.* **70**, 23-26.

Sher, R. (1989). HIV infection in South Africa, 1982-1988 - a review. *S. Afr. Med. J.* **76**, 314-318.

Sherefa, K., Sallberg, M., Johansson, B., Salminen, M. & Sonnerborg, A. (1997). Subtyping of human immunodeficiency virus type 1 strains by using antibodies specific for the third variable domain (V3) of gp120: Results may be affected by divergent V3 sequences. *J. Clin. Microbiol.* **35**, 2419-2421.

Simmons, G., Wilkinson, D., Reeves, J. D., Dittmar, M. T., Beddows, S., Weber, J., Carnegie, G., Desselberger, U., Gray, P. W., Weiss, R. A. & Clapham, P. R. (1996). Primary, syncytium-inducing human immunodeficiency virus type 1 isolates are dual-tropic and most can use either Lestr or CCR5 as coreceptors for virus entry. *J. Virol.* **70**, 8355-8360.

Simon, F., Mauclere, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M.C., Saragosti, S., Georges-Courbot, M.C., Barre-Sinoussi, F. & Brun-Vezinet, F. (1998) Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* **4**, 1032-1037.

Songok, E. M., Libondo, D. K., Rotich, M. C., Oogo, S. A. & Tukei, P. M. (1996). Surveillance for HIV-1 subtypes O and M in Kenya [letter]. *Lancet* **347**, 1700.

Sonnerborg, A. & Sherefa, K. (1997). HIV-1 subtype C epidemic in Ethiopia. *AIDS* **11**, 1897.

Soto-Ramirez, L., Renjifo, B., McLane, M. F., Marlink, R., O'Hara, C., Sutthent, R., Wasi, C., Vithayasai, P., Vithayasai, V., Apichartpiyakul, C., Auewarakul, P., Cruz, V. P., Chui, D.-S., Osathanondh, R., Mayer, K., Lee, T.-H. & Essex, M. (1996). HIV-1 Langerhans' cell tropism associated with heterosexual transmission of HIV. *Science* **271**, 1291-1293.

Stein, B. S., Gowda, S. D., Lifson, J. D., Penhallow, R. C., Bensch, K. G. & Engleman, E. G. (1987). pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. *Cell* **49**, 659-668.

Subbarao, S., Luo, C.-C., Limpakarnjanarat, K., Bhumisawasdi, J., Young, N. L., Mastro, T. D., Schochetman, G. & Kalish, M. L. (1996). Evaluation of oligonucleotide probes for the determination of the two major HIV-1 *env* subtypes in Thailand. *AIDS* **10**, 350-351.

Subbarao, S. & Schochetman, G. (1996). Genetic variability of HIV-1. *AIDS* **10**, S13-S23.

Swofford, D. L. (1993). PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1.1. Champaign, IL: Illinois Natural History Survey.

Takahashi, H., Nakagawa, Y., Pendleton, C. D., Houghten, R. A., Yokomuro, K., Germain, R. N. & Berzofsky, J. A. (1992). Induction of broadly cross-reactive cytotoxic T cells recognizing an HIV-1 envelope determinant. *Science* **255**, 333-336.

Tersmette, M., Gruters, R. A., de Wolf, F., de Goede, R. E., Lange, J. M., Schellekens, P. T., Goudsmit, J., Huisman, H. G. & Miedema, F. (1989). Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the pathogenesis of acquired immunodeficiency syndrome: studies on sequential HIV isolates. *J. Virol.* **63**, 2118-2125.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.

Tsuchie, H., Saraswathy, T. S., Sinniah, M., Vijayamalar, B., Maniar, J. K., Monzon, O. T., Santana, R. T., Paladin, F. J. E., Wasi, C., Thongcharoen, P., Hossain, M. M., Yamada, O., Kageyama, S., Kitamura, T. & Kurimura, T. (1995). HIV-1 variants in South and South-East Asia. *Int. J. STD AIDS* **6**, 117-120.

UNAIDS/WHO (1998). Report on the global HIV/AIDS epidemic, pp. 1-75. Geneva: SERVICES CONCEPT.

Van de Peer, Y. (1994). TREECON for Windows, Version 1.0. Antwerp: Department of Biochemistry, University of Antwerp.

Van der Ryst, E. (1996). Eight annual meeting of the National Cooperative Vaccine Development group for AIDS. *TB & HIV* **9**, 18-21.

Van Harmelen, J., Wood, R., Lambrick, M., Rybicki, E. P., Williamson, A. L. & Williamson, C. (1997). An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. *AIDS* **11**, 81-87.

Van Harmelen, J., Van der Ryst, E., Wood, R., Lyons, S. F. & Williamson, C. (1999a). Restriction fragment length polymorphism analysis for rapid gag subtype determination of human immunodeficiency virus type 1 in South Africa. J. Virol. Methods 78, 51-59.

Van Harmelen, J. H., Van der Ryst, E., Loubser, A. S., York, D., Madurai, S., Lyons, S., Wood, R. & Williamson, C. (1999b). A predominantly HIV type 1 subtype C-restricted epidemic in South African urban populations. *AIDS Res. Hum. Retroviruses* **15**, 395-398.

Vanden Haesevelde, M., Decourt, J.-L., De Leys, R. J., Vanderborght, B., Van der Groen, G., Van Heuverswijn, H. & Saman, E. (1994). Genomic cloning and complete sequence analysis of a highly divergent African human immunodeficiency virus isolate. *J. Virol.* **68**, 1586-1596.

Varmus, H. (1988). Retroviruses. *Science* **240**, 1427-1435.

Vartanian, J.-P., Meyerhans, A., Asjo, B. & Wain-Hobson, S. (1991). Selection, recombination, G-A hypermutation of human immunodeficiency virus type 1 genomes. *J. Virol.* **65**, 1779-1788.

Villari, P., Spino, C., Chalmers, T. C., Lau, J. & Sacks, H. S. (1993). Cesarean section to reduce perinatal transmission of human immunodeficiency virus: a metaananalysis. *Online J. Curr. Clin. Trials* **74**, 1993 (Doc No 74).

Wasi, C., Herring, B., Raktham, S., Vanichseni, S., Mastro, T. D., Young, N. L., Rubsamen-Waigmann, H., von Briesen, H., Kalish, M. L., Luo, C.-C., Pau, C. P., Baldwin, A., Mullins, J. I., Delwart, E. L., Esparza, J., Heyward, W. L. & Osmanov, S. (1995). Determination of HIV-1 subtypes in injecting drug users in Bangkok, Thailand, using peptide-binding enzyme immunoassay and heteroduplex mobility assay: evidence of increasing infection with HIV-1 subtype E. *AIDS* **9**, 843-849.

Weniger, B. G., Limpakarnjanarat, K., Ungchusak, K., Thanprasertsuk, S., Choopanya, K., Vanichseni, S., Uneklabh, T., Thongcharoen, P. & Wasi, C. (1991). The epidemiology of HIV infection and AIDS in Thailand. *AIDS* **5**, S71-S85.

Weniger, B. G., Takebe, Y., Ou, C.-Y. & Yamazaki, S. (1994). The molecular epidemiology of HIV in Asia. *AIDS* **8**, S13-S28.

WHO. (1995). Global Programme on AIDS. The current global situation of the HIV-AIDS pandemic. Geneva.

WHO network for HIV isolation and characterization. (1994). HIV type 1 variation in World Health Organization-sponsored vaccine evaluation sites: Genetic screening, sequence analysis, and preliminary biological characterization of selected viral strains. *AIDS Res. Hum. Retroviruses* **10**, 1327-1343.

Willey, R. L., Bonifacino, J. S., Potts, B. J., Martin, M. A. & Klausner, R. D. (1988). Biosynthesis, cleavage, and degradation of the human immunodeficiency virus 1 envelope glycoprotein gp160. *Proc. Natl. Acad. Sci. USA* **85**, 9580-9584.

Willey, R. L., Rutledge, R. A., Dias, S., Folks, T., Theodore, T., Buckler, C. E. & Martin, M. A. (1986). Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. *Proc. Natl. Acad. Sci. U S A* **83**, 5038-5042.

Williams, B. & Campbell, C. (1996). Mines, migrancy and HIV in South Africa-managing the epidemic. *S. Afr. Med. J.* **86**, 1249-1251.

Williams, R. & Campbell, C. (1998). How to control the Aids epidemic. In *FOCUS TEN* **April**, pp. 12-13.

Williamson, C., Engelbrecht, S., Lambrick, M., van Rensburg, E. J., Wood, R., Bredell, W. & Williamson, A. L. (1995). HIV-1 subtypes in different risk groups in South Africa [letter]. *Lancet* **346**, 782.

Wolinsky, S. M., Wike, C. M., Korber, B. T., Hutto, C., Parks, W. P., Rosenblum, L. L., Kunstman, K. J., Furtado, M. R. & Munoz, J. L. (1992). Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. *Science* **255**, 1134-1137.

Xin, K.-Q., Ma, X.-H., K.A., C., Bukawa, H., Ishigatsubo, Y., Kawamoto, S. & Okuda, K. (1995). Dual infectionwith HIV-1 Thai subtype B and E [letter]. *Lancet* **346**, 1372-1373.

Zekeng, L., Gurtler, L., Afane Ze, E., Sam-Abbenyi, A., Mbouni-Essomba, G., Mpoudi-Ngolle, E., Monny-Lobe, M., Tapka, J. B. & Kaptue, L. (1994). Prevalence of HIV-1 subtype O infection in Cameroon: preliminary results [letter]. *AIDS* **8**, 1626-1628.

Zhu, T., Korber, B. T., Nahmias, A. J., Hooper, E., Sharp, P. M. & Ho, D. D. (1998). An African HIV-1 sequence from 1959 and implications for the origin of the epidemic [letter]. *Nature* **391**, 594-597.

Zhu, T., Wang, N., Carr, A., Wolinsky, S. & Ho, D. D. (1995). Evidence for coinfection by multiple strains of human immunodeficiency virus type 1 subtype B in an African seroconvertor. *J. Virol.* **69**, 1324-1327.

APPENDIX A: Composition of buffers, media and gels

1. Reagents for making viral DNA lysates

Lysis buffer:

1 M Tris-HCL (pH 8.3)	1 ml
0.5 M EDTA	200 μl
10% Triton X-100	5 ml
10% SDS	10 μl

Make up to 100 ml with sterile, deionized, distilled water.

2. Reagents for agarose electrophoresis

1% agarose gel:
1 g D-1 LE agarose (molecular grade) in 100 ml 1X TBE. Microwave until dissolved. Allow it to cool to room temperature and add 5 μl 10 mg/ml EthBr before pouring the gel.

10X TBE stock buffer:

Tris-HCL	108 g
Boric acid	55 g
0.5 M EDTA	40 ml

Make up to 1000 ml with distilled water. Dilute 1:10 for 1X TBE and 1:20 for 0.5X TBE running buffer.

3. Reagents for HMA

10X Heteroduplex annealing buffer:

1 M NaCl	5.84 g
1 M Tris-base (pH 7.8)	10 ml
0.5 M EDTA	4 ml

Make up to 100 ml with sterile, deionized, distilled water.

4. Reagents for polyacrylamide gel electrophoresis

5% PAGE gel:

30% acrylamide:bis-acrylamide (30:0.8)	8.3 ml
10X TBE	5 ml
sterile, deionized, distilled water	36.7 ml
10% ammonium persulphate	500 μl
TEMED	33 μl

5. Reagents for transformation

SOC medium:

bacto-tryptone	20 g
bacto-yeast extract	5 g
1 M NaCl	1 ml
1 M KCL	250 μl

Add 97 ml distilled water and autoclave for 15 min at 121 °C. Allow it to cool to room temperature and then add 1 ml 2 M Mg^{2+} stock (1 M $MgCl_2.6H_2O$/1 M $MgSO_4.7H_2O$) and 1 ml 2 M sterile glucose (both filter-sterilized). Make up to 100 ml with distilled water. Filter the complete medium through a 0.2 μm filter unit. Final pH should be 7.0.

LB agar plates:

LB agar (Lennox L agar)	32 g

Make up to 1 L with distilled water and autoclave for 15 min at 121 °C. Allow it to cool to 60 °C and then add 1 ml 50 mg/ml ampicillin (and 3 ml of 5 mg/ml tetracycline in the case of pMOS*Blue* T-vector) before pouring the plates.

LB growth medium:

LB broth base (Lennox L broth base)	20 g

Make up to 1000 ml with distilled water and autoclave for 15 min at 121 °C. Allow it to cool to room temperature and then add 1 ml 50 mg/ml ampicillin (and 3 ml 5 mg/ml in the case of pMOS*Blue* T-vector).

6. Reagents for sequencing

8% acrylamide gel:

acrylamide	7.6 g
bis-acrylamide	0.4 g
urea	48 g
10X TBE	10 ml
sterile, deionized, distilled water	40 ml
10% ammonium persulphate	500 μl
TEMED	50 μl

20X Glycerol Tolerant running buffer:

Tris-base	216 g
Taurine	72 g
$Na_2EDTA.2H_2O$	4 g

Make up to 1 L with distilled water and autoclave for 15 min at 121 °C. Dilute 1:20 for 1X Glycerol Tolerant running buffer

APPENDIX B: List of suppliers

acrylamide	Boehringer Mannheim GmbH, USA
ABI PRISM 377 DNA Sequencer	Perkin Elmer Applied Biosystems, USA
ALFexpress™ DNA Sequencer	Pharmacia Biotech AB, Sweden
ammonium persulphate	Pharmacia Biotech, Sweden
Ampicillin	Boehringer Mannheim GmbH, Germany
bacto-tryptone	Unipath Ltd., England
bacto-yeast extract	DIFCO Laboratories, USA
Biometra TRIO-Thermoblock TB-1 cycler	Gottingen, Germany
bis-acrylmaide	Boehringer Mannheim GmbH, Germany
boric acid	Merck, Germany
bromophenol blue	Merck, Germany
CENTRI-SEP columns	Princeton Sperations Inc., USA
chloroform	Merck, Germany
chromatography paper	Whatman Scientific Limited, England
Cy5 AutoRead Sequencing kit	Pharmacia Biotech, Sweden
Cy5-dATP labeling mix	Pharmacia Biotech, Sweden
D-1 LE agarose	Whitehead Scientific, South Africa
dNTP's	Boehringer Mannheim GmbH, Germany
dRhodamine Terminator Cycle Sequencing kit	Perkin Elmer Applied Biosystems, Great Britain
EDTA	Boehringer Mannheim GmbH, Germany
ethidium bromide	Promega, USA
Ficoll-Hypaque	Amersham Pharmacia Biotech AB, Sweden
GeneAmp® PCR System 2400/9600 thermocycler	Perkin Elmer Applied Biosystems, Great Britain
Genscreen® HIV1/2 version 2	Sanofi Diagnostics Pasteur, USA
HCL	Sigma Chemical Co., USA
HMA kit	NIH AIDS Research and Reference Program, US
HYBAID Touch Down PCR system	HYBAID Limited, UK
Hyperfilm™-βmax autoradiograpgy film	Amersham International plc., UK
IPTG	Boehringer Mannheim GmbH, Germany
isoamylalcohol	Merck, Germany
KCl	Sigma Chemical Co., USA
LB growth medium	GIBCOBRL, Life Technologies, Scotland
LB agar	GIBCOBRL, Life Technologies, Scotland
$MgCl_2$	Merck, Germany
$MgSO_4$	Merck, Germany
NaCl	Merck, Germany
pGEM® -T vector kit	Promega, USA
pMOS*Blue* T-vector kit	Amersham LIFE SCIENCE Inc., England
polycarbonate microtiter plate	HYBAID Limited, UK

Proteinase K	Boehringer Mannheim GmbH, Germany
QIAGEN Plasmid isolation kit	QIAGEN, Germany
QIAmp® Viral RNA isolation kit	QIAGEN GmbH, Germany
Reverse transcriptase	Boehringer Mannheim GmbH, Germany
RNase inhibitor	Boehringer Mannheim GmbH, Germany
Sabax water (sterile, deionized, distilled)	Adcock Ingram Limited, South Africa
SDS	Merck, Germany
Sequenase PCR Product Sequencing	Amersham LIFE SCIENCE Inc., USA
Super-Therm DNA polymerase (including 10X buffer and $MgCl_2$)	Advanced Biotechnologies Limited, UK
Taurine	Pharmacia Biotech, Sweden
TEMED	Stratagene Cloning System, California
tetracycline	Boehringer Mannheim GmbH, Germany
Tris-base	Pharmacia Biotech, Sweden
Tris-HCL	Boehringer Mannheim GmbH, Germany
Triton X-100	Boehringer Mannheim GmbH, Germany
urea	Boehringer Mannheim GmbH, Germany
Vironostika® HIV Uni-form II plus O	Organon Teknika, Belgium
Wellcozyme* HIV Recombinant	Murex, UK
X-gal	Boehringer Mannheim GmbH, Germany
xylenexyanol	Fluka AG, Switzerland

APPENDIX C: Patient Information, Informed Consent and Questionnaire

PATIENT INFORMATION LEAFLET

GENETIC CHARACTERIZATION OF HIV-1 IN SOUTH AFRICA

Your doctor has asked you if you would be willing to take part in this study. This information sheet is to help you decide if you would like to participate. Please read it carefully and do not hesitate to ask your doctor any questions about it. You should not agree to take part unless you are completely happy with it.

The aim of this study is to characterize the strains (or subtypes) of the human immunodeficiency virus (HIV) that are currently circulating among people living in South Africa. The reason for doing this is to allow us to follow the spread of the epidemic and to determine the rate at which the virus is evolving. This is a research project, the results of which will be used by Epidemiologists and Public Health Officials, and this information will also be essential when HIV vaccine trials are undertaken in South Africa.

If you decide to take part in this study you will be asked to provide a 5 ml blood sample which will be taken at the same time as your routine checks. The HIV in this specimen will be subjected to molecular analyses and this information will be passed on to your doctor. In addition, a CD4 cell count will also be done on your blood specimen which will b given to your doctor which he/she will use in your clinical management. You will also be asked a number of questions by a doctor or nurse involved in this study relating to your geographic and sexual history. Some of these questions may be too personal and you may wish to refrain from answering them. This will not jeopardize the study and will not affect your access to other medical care at the clinic.

By taking part in this study you are contributing to a great understanding of HIV in South Africa which is a first step towards attempting to curtail this epidemic. This study will not affect your own situation directly, that is, it will not provide a cure or alter the course of your infection. All information obtained during the course of this study is strictly confidential and any data that maybe reported will not include information which identifies you as a patient.

Thank you for taking the time to read this.

INFORMED CONSENT FORM (Appendix C)

I hereby confirm that I have been informed by the investigator, Dr.....

of the nature, purpose and risks of this clinical study.

I have also received and understand the Patient Information Leaflet regarding this study.

I am also aware that the results of this study, including personal details regarding my age, sex, diagnosis and risk factors will be anonymously processed into a publishable report.

I may at any stage, without prejudice withdrawn my consent and participation in the trial.

I have had sufficient opportunity to ask questions and (of my own free will) declare myself prepared to participate in this study.

Patient:

Signature: .....

Name (please print): .....

Witness:

Signature: ..... Date: ...../.. ...../.....

Dr who obtained informed consent:

I hereby confirm that the above patient has been informed fully about the nature, purpose and risks of this study.

Signature: ..... Date: ...../...../.....

Name (please print): .....

HIV SUBTYPING STUDY AT NATIONAL INSTITUTE FOR VIROLOGY (NIV)

(Appendix C)

Contacts at NIV: Dr Barak Morgan, Dr Lynn Morris, Ms Ba Bredell
at 321 4200 (phone), 882 0596 (fax)

Please fill this form an send it with a 5 ml EDTA (purple top) blood specimen within 48 hrs.

Referring doctor + phone # ..... Hospital .....

Patient I.D. (Initials/Code) ..... Hosp. No. .....

Age ..... Sex ..... Date of specimen ...../...../.....

Race ..... Nationality ..... Ethnicity .....

Occupation .....

Geographic/Residential history .....

.....

Date of diagnosis .....

Risk factor or likely route of transmission (see below) .....

Place and time of transmission .....

Clinical events including CDC category .....

.....

.....

Drug history .....

.....

Latest CD4 count ..... Date of specimen ...../...../.....

Other .....

.....

Patient selection

White patients especially women
Patients infected via blood transfusions or factor VIII preparations
Patients with Kaposi's sarcoma
Non-South African patients who you suspect were infected outside South Africa
Homosexuals
Intravenous drug users
HIV negative patients with a clinical diagnosis of "AIDS"

Your co-operation is much appreciated. Thank you.

APPENDIX D: Ethical clearance

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

COMMITTEE FOR RESEARCH ON HUMAN SUBJECTS (MEDICAL)
Ref: R14/49 Morris

CLEARANCE CERTIFICATE PROTOCOL NUMBER M 960610

PROJECT	Genetic subtyping of HIV in Southern Africa
INVESTIGATORS	Dr L Morris
DEPARTMENT	Virology, Nat Inst for Virology
DATE CONSIDERED	960628
DECISION OF THE COMMITTEE *	Unconditionally approved
DATE	960710

CHAIRMAN. .....(Professor P E Cleaton-Jones)

c c Supervisor: Dr L Morris
Dept of Virology, Nat Inst for Virology

==========

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and ONE COPY returned to the Secretary at Room 10001, 10th Floor, Senate House, University.

I/~~we~~ fully understand the conditions under which I am/~~we are~~ authorized to carry out the abovementioned research and I/~~we~~ guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee.

DATE. 31.7.96...SIGNATURE .....

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

APPENDIX E: Publications from this thesis

1. Cheingsong-Popov R, Williamson C, Lister S, Morris L., van Harmelen J, Bredell H, Wood R, Sonnenberg P, van der Ryst E, Martin D, Weber J. (1998). Usefulness of HIV-1 V3 serotyping in studying the HIV-1 epidemic in South Africa. *AIDS* **12**, 949-50.

2. Helba Bredell, Carolyn Williamson, Pamela Sonnenberg, Desmond J. Martin and Lynn Morris. (1998). Genetic Characterization of HIV Type 1 from Migrant Workers in Three South African Gold Mines. *AIDS Res. Hum. Retroviruses* **14**, 677-684.

3. Helba Bredell, Gillian Hunt, Barak Morgan, Caroline T. Tiemessen, Desmond J. Martin and Lynn Morris. (1999). Identification of HIV-1 Inter-subtype Recombinants in South Africa using *env* and *gag* Heteroduplex Mobility Assays. *AIDS Res. Hum. Retroviruses* (in press).