Phylogenetic Analysis and Diversity of HIV-1 Envelope V3 Sequences from Seroconverters of the two Ethiopian Cohorts: Akaki & Wonji

A thesis submitted to the School of Graduate studies of Addis Ababa University in partial fulfilment of the Degree of Master of Science in biology

By

Measho Hagos

Addis Ababa

June, 2001

Table of Contents

Acknowledgment i

List of figures and tables ii

List of abbreviations iii

Abstract iv

1. INTRODUCTION 1

1.1 HIV/AIDS 1

1.2 Transmission and epidemiology 1

1.3 HIV life cycle 2

1.4. Course of HIV Infection 2

1.5. The Virus 4
1.5.1. Morphology and Genetic Structure 5
1.5.2. The envelope glycoprotien, the third variable (V3) region of HIV-1 6

1.6. Genetic Diversity of HIV-1 8
1.6.1. Synonymous and non-synonymous mutation 10

1.7 HIV-1 subtypes 11
1.7.1. Global distribution of HIV-1 subtypes 13

1.8. Differences in biological properties of HIV-1 isolates 14

1.9. HIV-1/AIDS in Ethiopia 14

1.10. HIV-1 subtypes in Ethiopia 15
1.10.1. The HIV-1 subtype C genetic subcluster, C' 16

2. MATERIALS AND METHODS 19

2.1. Study participants 19

2.2. Blood collection and Plasma isolation 20

2.3. Enumeration of lymphocyte subsets 20

2.4. Viral load determination 20

2.5. RNA isolation 20

2.6. Reverse Transcription (RT) 21

2.7. Amplification by polymerase Chain Reaction (PCR) 21
2.7.1. First PCR 21
2.7.2. Second PCR 22

2.8. Agarose Gel Electrophoresis 22

2.9. HIV-1 biological Cloning 23

2.10. DNA sequencing 23

2.11. Phylogenetic analysis 24

3. RESULTS 25

3.1. Virological and Immunological Data 25

3.2. PCR and DNA sequencing 31

3.3. Phylogenetic analysis 31

3.4. HIV-1 subtype C genetic subcluster (C') 34

3.5. Predicted amino acid sequence 36

3.6. Amino acid sequence comparison of the subtype C and C' virus 36

3.7. Synonymous and nonsynonymous distance calculation 40

4. DISCUSSION 44

5. CONCLUSION 53

5. REFERENCES 54

Acknowledgment

I would like to express my sincere acknowledgment to my advisor Professor Endashaw Bekele, and my advisors at EHNRI/ENARP, Dr. Almaz Abebe and Dr Dawit Wolday, for their guidance and supervision. My special thanks goes to my advisors at EHNRI/ENARP, for their interest in the work, useful critics, unlimited support, and follow up. I am especially glad because I had unreserved access to both whenever I needed help.

I would like to thank the EHNRI/ENARP management for allowing me use the ENARP facilities for this study. The staff of Biology Department, the School of Graduate studies, and Alemaya University are acknowledged for facilitating this study.

I am so grateful to my former advisor Dr. Tobias Rinke de Wit for introducing me to ENARP and help at the beginning of the work and Dr. Georgios Pollakis for helping in the data analysis, Aletta Kliphuis for teaching me the techniques and Tesfaye Tilahun for helping me in the laboratory. All the ENARP staffs especially the laboratory technicians deserve my heartfelt thanks, all were friendly and helpful. I would like also to thank Ph.D. students at ENARP for their help and friendship especially Amha Kebede, for reading the manuscript and editorial comments.

My deepest gratitude goes to all my friends especially to Afework Kassu, Mamuye Hadis, Wondwossen Mekonnen, Wude Mihret, Medhanie Asmelash, and Adunga Woyesa for being great friends and for the wonderful moments we had.

My sincere thanks and love goes to my dear friend and husband, Tesfaye Kebede for all the help and encouragement. At last but not least, my sincere thanks and love goes to my sister, Mulualem Hagos and her family for the encouragement, love, and care.

List of Figures and Tables

Figure 1. HIV-1 Virion. 5

Figure 2. HIV-1 Genome 6

Figure 3. HIV-1 Envelope gp120 Region 8

Figure 4. A Phylogenetic Tree of all gp120 V3 Seroconverter Sequences 33

Figure 5. A Phylogenetic Tree of gp120 V3 Seoconverter Sequences Showing Subtype C and the Subcluster C' Sequences 35

Figure 6. Predicted Consensus gp120 C2-V3 Amino Acid Sequence 38

Figure 7. Vespa Supported Amino Acid Sequence Comparison 39

Figure 8. Synonymous and Nonsynonymous Distance Distribution Among and Between C And C' Sequences. 42

Figure 9. Intra-Person Synonymous Distance Variation 43

Figure 10. Intra-Person Nonsynonymous Distance Variation 43

Table 1. Follow up Data on 19 Cohort Participants who seroconverted to HIV-1.....26

Table 2. Mean Synonymous (ds) and Nonsynonymous (dn) Distance.....40

Table 3. Mean Intra Person Synonymous (ds) and Nonsynonymous (dn) Distance 41

Abbreviations

AIDS= Aquired Immunodeficiency Syndrome

AMV-RT= Avian myeloblastosis Virus Reverse transcriptase

HIV-1= Human Immunodeficiency virus type 1cDNA= complementary DNA

CSW= commercial sex workers

CTL= cytotoxic T-lymphocyte

DNA= Deoxynucleotide triphosphate

EDTA= Ethylene diamine tetra acetic acid

env= envelope

ENARP= EthioNetherlands AIDS Research Project

EHNRI= Ethiopian Health and Nutrition Research Institute

gag= group specific antigen

gp=glycoprotein

GUSCN= Guanidium Isothiocynate

kD= Kilodalton

kb= kilobase

LTR= Long terminal repeat

M-tropic= Macrophage tropic

NF-kB= nuclear factor kappa B

NSI= Non-syncytium Inducing

NASBA= Nucleic Acid Based Amplification

PCR= polymerase chain reaction

RT= Reverse transcription

PHA= phytohaemaglutinin

SI= Syncytium Inducing

Taq= *Thermus aquaticus*

TEMED= N-N-N-tetramethylethylenediamine

T-tropic= T lymphocyte tropic

TAE= Tris acetate

TBE= Tris borate

APS= Amonium persulphate

PBMC= Peripheral blood mononuclear cell

Abstract

HIV-1 is characterized by its genetic variability. Based on nucleotide sequence relatedness, the HIV-1 virus variants are grouped into three groups: M (major), O (outlier), and N (non-M or non-O). Group M viruses are further divided into 9 genetic subtypes or clades (A-D, F-H, J, & K). The Ethiopian HIV-1/AIDS epidemic is reported to be dominated by HIV-1 subtype C, which recently is reported to have a genetic subcluster designated as C'. In an effort to determine the diversity of gp120 V3 sequences and the frequency of subtype C and C' viruses among HIV-1 isolates from recently infected individuals, 19 Ethiopian factory workers who seroconverted for HIV-1 antibodies during the year 1997-2000 were included in the study. HIV-1 RNA was isolated from plasma samples of study subjects, the RNA was reverse transcribed into cDNA and the envelope gp120 V3 region was amplified. Phylogenetic analysis of the env V3 sequences showed all the 19 subjects to be infected by subtype C. 11 out of the 15 sequences were found to be subtype C' and 4 subtype C. Based on this observation we speculate that subtype C' viruses are spreading and might have a selective advantage for transmission. Amino acid sequence comparison of the two groups of viruses showed significant amino acid sequence differences. There was a 17% synonymous and 15% nonsynonymous distance among all the seroconverter sequences and a 24% synonymous and 17% nonsynonymous distance between the C and C' group of sequences. The synonymous distance among subtype C'(9%) is much less compared to the distance among the subtype C sequences (16%).

1. Introduction

1.1 HIV/AIDS

Human Immunodeficiency Virus (HIV) is the causative agent of acquired immunodeficiency syndrome (AIDS). The disease was first identified in USA in 1981 (Gallo *et al.*, 1984) among homosexual men who developed certain malignancies that appear when the immune system is compromised. The virus was then isolated in 1983 by Barre-Sinoussi *et al* (1983) and was initially named as Human T-lymphotropic Virus III (HTLV III), Lymphadenopathy Associated Virus (LAV) and AIDS associated Retrovirus (ARV) (Gallo *et al.*, 1984). But finally it was renamed as HIV (Coffin *et al.*, 1986). The virus infects and replicates in CD4+ cells of the immune system and brain leading to loss of CD4+ cell number and function thereby making the individual immunocompromized and prone to opportunistic infections and malignancies (Fauci, 1988).

1.2 Transmission and epidemiology

HIV can be transmitted via heterosexual and homosexual contact, through blood transfusion, needle sharing, and from mother to child. Since its identification in the early 1980's the virus has spread all over the world dramatically to reach epidemic levels and claimed the lives of many people. According to WHO report, currently 36.1 million people are estimated to be living with HIV/AIDS of which 25.3 million are in the Sub-Saharan Africa where the mode of transmission is mainly heterosexual (WHO/UNAIDS, 2000). In the same year, a total of 3.0 million people have died of HIV/AIDS of which 2.4 are in the sub-Saharan Africa and there were a total of 5.3 million new infections, the 3.8 million being in the sub-Saharan Africa (WHO/UNAIDS, 2000).

1.3 HIV life cycle

Entry of HIV into target cells requires virus binding to CD4 molecule, a major HIV receptor. CD4 is a cell surface glycoprotein mainly expressed on T lymphocytes and some other cell types of the macrophage lineage (Dalgleish *et al.*, 1984). In addition to the CD4 molecule, chemokine receptors have been identified as co-receptors for entry to target cells (D'Souza and Harden, 1996). Chemokine receptors are members of the seven transmembrane-spanning receptor family that have the α- and β- chemokines as their natural ligands (Feng *et al.*, 1996). CXCR4 and CCR5 are the two major coreceptor molecules used by HIV but there are also others that are alternatively/additionally used such as CCR3 and CCR2B (McNicholl *et al.*, 1997). CD4 and coreceptor binding leads to some conformational changes in the virus envelope which expose the fusion domain of the virus envelope and trigger fusion of virus and host cell membrane and subsequent internalisation of the virus (Doms and Peiper, 1997). After internalisation, the virus is uncoated and the viral RNA is reverse transcribed into double stranded DNA in the cytoplasm, which later is transported to the nucleus to be integrated into the host genome as a "provirus" by the virus enzyme integrase. After integration into the host genome, the provirus serves as a template for cellular DNA-dependent RNA polymerases to generate new viral RNA genomes as well as shorter subgenomic mRNAs. Then, the mRNAs are translated into structural and regulatory proteins, which later are assembled and bud from the cell as virions (Barre-Sinoussi, 1996). Most often the virus remains silent or with low level of replication until the cells are activated which will activate virus replication.

1.4. Course of HIV Infection

The course of HIV infection has three phases: primary, asymptomatic and AIDS.

Primary HIV-1 infection is the stage of the disease when the virus first disseminates throughout the body of newly infected individuals (Daar, 1998). This stage is characterized by acute retroviral syndrome (ARS), that appear in 30-90% of infected individuals. The common clinical manifestations being fever, myalgia, skin rash, sore throat, lymphadenopathy, headache, diarrhoea and some neurological problems (Cooper *et al.*, 1985). The symptoms are usually self-limited and disappear with in few weeks after infection. Primary HIV-1 infection is also characterized by high virus load in the blood and a very low CD4 count (Daar *et al*, 1991; Ho et al., 1995). But subsequently, the low CD4 count start to increase to subnormal levels and the high plasma virus load reduce primarily as a result of CD8+ cytotoxic T-cell activity (Niu *et al.*, 1993), which precedes the production of neutralizing antibodies. The individual could be viremic but HIV-1 seronegative for two to six weeks until antibodies against the virus are produced and this period is referred to as "window period". Seroconversion (antibody production) usually takes place within 2-6 weeks after infection, with antibodies to HIV-1 p24 antigen, followed by anti gp120 responses (Cooper *et al.*, 1985).

The asymptomatic phase of HIV-1 infection starts after the disappearance of symptoms of primary HIV infection and decline of viral load and increase in CD4 T cell count to subnormal levels. One year after infection, levels of HIV-1 RNA in plasma stabilise at a certain level called viral set point, which is predictive for clinical progression, the higher the viral set point the faster the disease progression (Mellors *et al*, 1996). The duration of the asymptomatic phase is highly variable and ranges between 8-15 years (Taylor, *et al.*, 1986). During this phase of HIV-1 infection the virus continuously replicate with gradual loss in

CD4 count and the person seems apparently healthy and remains free of other infections until the CD4 T cell number is highly reduced and the AIDS stage is reached.

Finally the last phase, known as the AIDS stage is reached when the infected individual has a very low CD4 count usually < 200/ml plasma and a functionally deteriorated immune system, which is no more capable of controlling infectious agents and is prone to a multiple of opportunistic infections as well as neoplastic diseases that lead to final death (Lifson *et al.*, 1988).

1.5. The Virus

The Human Immunodeficiency Viruses belong to the subfamily Lentivirinae in the family Retroviridae (White and Fenner, 1994). As the name lentiviruses indicates these viruses remain latent for a certain period of time before they induce diseases. The family includes enveloped RNA viruses that are characterized by having a diploid RNA genome and by their ability to make cDNA using the virus enzyme reverse transcriptase (Wong-Staal, 1990).

There are two types of Human Immunodeficiency viruses, HIV-1 and HIV-2 with 40-50% homology between their genomes (Korber et al., 1998). HIV-1 is distributed worldwide and accounts almost for all of the HIV infection in the world, while HIV-2 is restricted to West Africa, is less virulent and transmittable compared to HIV-1 (Marlink *et al.*, 1994). Since HIV-1 accounts for almost all of the infections in the world, it is clinically important and most of the studies and literature are on HIV-1.

1.5.1. Morphology and Genetic Structure

The HIV-1 particle has three parts: outer envelope, matrix protein, and virus core (Fig.1). The virus envelope is made up of a lipid bilayer membrane derived from the host cell and embedded in the lipid bilayer are virally encoded proteins, gp120 and gp41 (Kuritzkes, 2000). Electron microscopy has revealed the HIV-1 particle to have approximately 72 knobs containing trimers of the gp120 and gp41 complex. Beneath the envelope is the matrix protein (P^{17}), which is located surrounding the core. The virus core consists of the major capsid protein (P^{24}), the diploid single stranded RNA genome, the nucleocapsid proteins (P^7/P^9), and the three viral enzymes, protease, reverse transcriptase, and integrase (Robey *et al.*, 1985; Korber *et al.*, 1998).

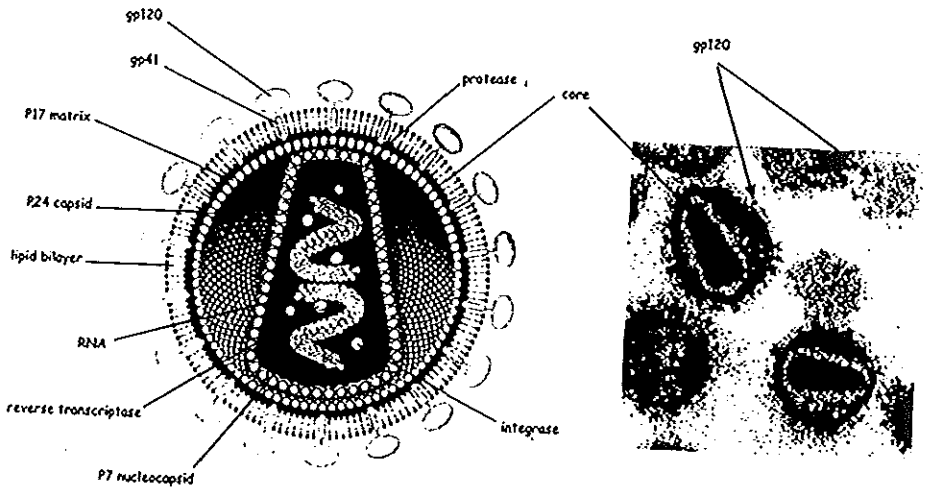


Figure 1. HIV-1 virion. (A). A schematic representation. (B). An electron micrograph. (Blaak, 1999).

The HIV-1 genome is about 9.7 Kb in length and is flanked by two long terminal repeats (LTRs) at the 5' and 3' ends. The three structural genes *gag*, *pol* and *env*, which are

common to all retroviruses code for the structural proteins. The *gag* gene product is a 55 kd precursor myristlated protein (P^{55}), which later is processed by the viral protease to give the matrix protein (p^{17}), the capsid protein (P^{24}) and the nucleocapsid proteins (P^{7}/P^{9}). The *pol* gene codes for the three viral enzymes, protease, reverse transcriptase and integrase. These enzymes are produced as a *gag-pol* precursor polyprotein, which is processed by the viral protease to give the three viral enzymes. The *env* gene codes for the envelope precursor glycoprotein, gp160, which is processed by a cellular protease into the two envelope proteins, gp120 (external glycoprotien) and gp41 (transmembrane glycoprotein) (Robey *et al.*, 1985; Thomas *et al.*, 1991). In addition to these structural genes the HIV-1 genome has six regulatory/ accessory genes, *tat, rev, nef, vif, vpr, and vpu or vpx* instead of *vpu* for HIV-2 (Fig.2) (Wong-Staal, 1990; Barre-Sinoussi, 1996; Korber *et al.*, 1998).

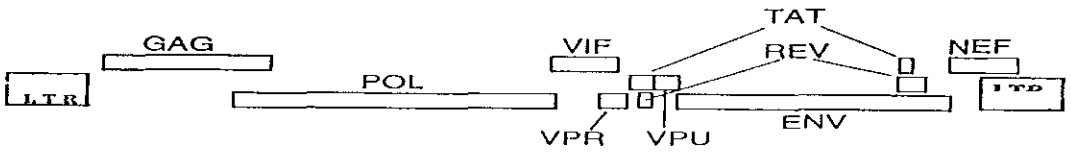


Figure 2. HIV-1 genome (Korber *et al.*, 1998).

1.5.2. The envelope glycoprotien, the third variable (V3) region of HIV-1

The viral envelope (*env*) gene product, gp160, is a 160kD glycosylated protein and is the most important part of the virion in its interaction with the host cell upon infection and it is a target for both humoral and cell mediated immune responses. This region has 24 aspargine linked glycosylation sites (N-linked glycosylation sites) (Fig. 3). Glycosylation or sugar coating of this protein occurs in the endoplasmic reticulum followed by oligomerization of

the glycosylated protein to a trimer for transport from the endoplasmic reticulum to the plasma membrane. (Willey *et al.*, 1988). The two glycoprotein subunits are non-covalently linked and the gp120/gp41 complex is anchored to the lipid bilayer of the virion via a single membrane-spanning region in the gp41 (Willey *et al.*, 1988).

Gp120 has sites for CD4 binding that involve several regions of gp120 that are brought together by the three-dimensional folding of gp120 (Wyatt *et al.*, 1998). The glycosylation of gp120 contributes to achieving the correct conformation of this protein, which is important for virus binding to CD4 (Li *et al.*, 1993). The sugar coating is also viewed as a protective mechanism, masking the virus from the host's immune system.

Based on amino acid sequence variability, gp120 is divided into five variable (V1-V5) and five conserved (C1-C5) regions. Four of the variable regions, V1-V4, form exposed loops formed as a result of disulfide bonds at their bases (Fig. 3) (Willey et al., 1988). The V3 region is the most studied region of gp120 as it is target to neutralizing antibodies, has epitopes for CTL response, cell tropism and cythopaticity (Kuiken, *et al.*, 1993; Zwart, 1994). The length of this region varies from 30-40 amino acids and it is flanked by two highly conserved cysteine (C) residues (Fig. 3).

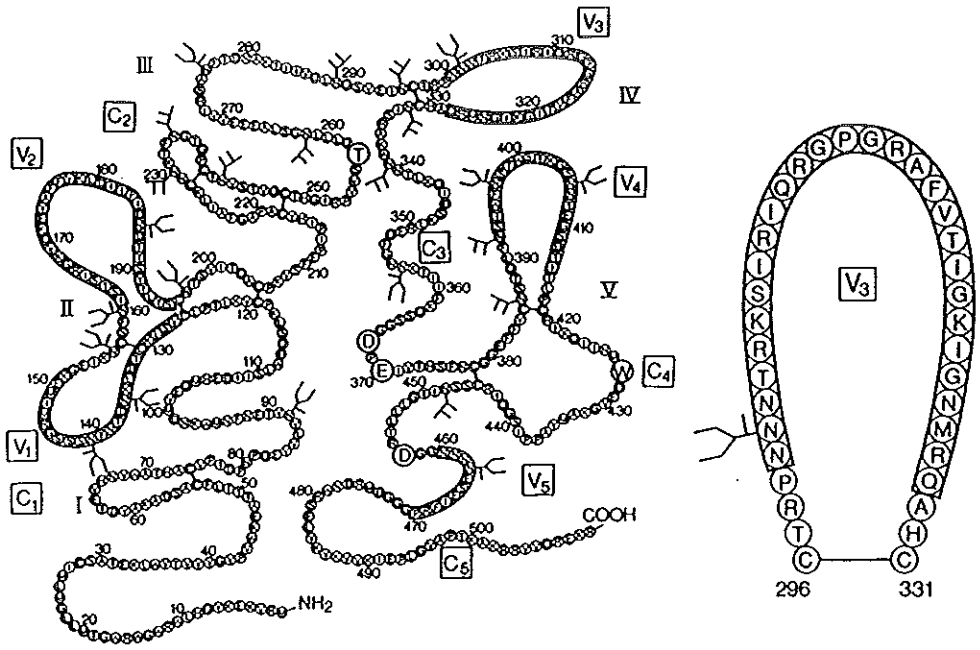


Figure 3. HIV-1 envelope gp120 region showing the loops formed as a result of disulfide bonds and the glycosylation sites (ζ). The V3 loop is shown on the right side.

1.6. Genetic Diversity of HIV-1

HIV-1 is characterized by its extensive genetic variation (Sharp *et al.*, 1995). At a certain time point, the HIV-1 population in infected individuals is generally heterogeneous and infected individuals have been shown to harbour a large number of genetically related viral genomes, called "quasi species" (Coffin, 1995). Genetic variation of HIV-1 is also observed between infected individuals within a population or among populations (Lukashove and Goudsmit, 1998). HIV-1 genes differ in their variability, the *gag* and *pol* genes are much more conserved than *env*. *tat* and *rev* likewise vary, but to a lesser extent than the *nef* and *env* genes (Kampinga, 1996).

The cause of HIV-1 genetic diversity is mainly accumulation of point mutations introduced by the error prone HIV-1 reverse transcriptase during replication (Mansky, 1998). Unlike many other DNA polymerases, HIV-1 reverse transcriptase lacks 3' exonuclease proof reading mechanisms. The rate of nucleotide substitutions introduced by reverse transcriptase is approximately 10^{-4} per nucleotide per cycle of replication or in other words a single point mutation per genome is introduced every time the virus replicates (Preston *et al.*, 1988; Nowak 1990). Besides, the virus has a high rate of replication, where about 10^{10} virions are produced each day thereby increasing the rate of error introduction (Prelson *et al.*, 1996).

Recombination between different virus strains also adds to the HIV-1 genetic diversity (Robertson *et al.*, 1995; Cornelissen *et al.*, 1996). Recombination between the two RNA strands that are packed in a virion occurs during replication because duplication of viral RNA involves copying alternatively from each of the strands forming the diploid RNA. Recombination could be either between strands of the same subtype (intrasubtype) or strands of different subtypes (intersubtype). Inter-subtype recombination occurs when individuals are co-infected with different subtypes and when the two RNA strands of the different subtypes are packaged in a virion (Robertson et al., 1995). Inter-subtype recombinants are common in regions where multiple subtypes co-circulate like in the sub Saharan Africa (Sherefa *et al.*, 1998; Carr *et al.*, 1998).

Besides point mutations due to the erroneous reverse transcriptase and recombination, insertions, deletions, and duplications also contribute to virus genetic diversity. (Lukashov and Goudsmit, 1998). This ongoing generation of variants within an individual over time

enables HIV-1 both to escape immune surveillance and to develop resistance to antiviral agents (Condra, *et al.*, 1995; Price *et al.*, 1997).

The high genetic variability of HIV-1 has led to rapid virus evolution (Mansky, 1998) under the selection pressure of the host environment mainly the immune system. This virus evolution follows the Darwinian evolution. A certain variant with a high relative fitness in a given intra host environment will predominate at some point, if the host conditions change, another variant become the best adapted and gives rise to a new line of dominating virus (Zwart *et al.*, 1994, McMichael et al., 1997). The evolution of HIV-1 *in vivo* is reflected by the emergence of virus isolates with increased replication potential and ability to infect cells as the clinical disease progresses. (Lukashov and Goudsmit, 1998) and the emergence of CTL escape and drug resistant strains (Condra *et al.*, 1995; Price *et al.*, 1997).

1.6.1. Synonymous and non-synonymous mutation

The nucleotide substitutions in the virus genome, due to the error prone reverse transcriptase, could result in either synonymous or non-synonymous substitutions. Nucleotide substitutions that do not result in change in amino acid are called synonymous substitutions while nucleotide substitutions that lead to change of the amino acid are called nonsynonymous substitutions. The frequency of these mutations is different along a given open reading frame. In the variable regions of gp 120, nonsynonymous nucleotide substitutions exceed synonymous substitutions (Kampinga, 1996) showing that this region of the virus gene is under greater selection pressure. Nonsynonymous nucleotide substitutions occur less frequently in progressors than in nonprogressors, reflecting a low

selective pressure due to a less efficient immune response of the progressors (Lukashove and Goudsmit, 1997).

Comparison of the numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions accumulated per site is used to predict the type of selection. Thus the ratio of $d_S/ d_N = 1$, indicates neutral selection, $d_S/ d_N < 1$, reflects positive selection, and $d_S/ d_N > 1$, reflects negative selection. (Holmes and de A. Zanotto, 1998; Mansky, 1998).

1.7 HIV-1 subtypes

The continuous variation of HIV-1 has led to the presence of different genetic clades or subtypes. Based on phylogenetic analysis of sequences of part of a genome or the whole genome, the HIV-1 genetic variants are divided into three groups: M (major), O (outlier) and N (non M or O) (Robertson *et al.*, 2000). These three groups are believed to result from three independent introductions of the ancestors of these groups into the human population (Simon *et al.*, 1998). Group M viruses are responsible for most of the infection in the world and have been further divided into nine genetic subtypes or clades (A-D, F-H, J and K) (Robertson *et al.*, 2000).

In phylogenetic analysis the genetic distance, which is an estimate of the number of nucleotide substitutions between sequences is calculated, the branch length in the tree representing the genetic distance from the common ancestor. Subtyping of a virus isolate is done by aligning nucleotide sequence of part or the whole virus genome with the already existing consensus reference sequences found in the Los Alamos HIV sequence database and by doing phylogenetic analysis. Subtyping is mainly done based on phylogenetic analysis of env gp120 and *gag* sequences (Simon *et al.*, 1998) where there is about 30%

nucleotide sequence divergence between the genetic subtypes in the *env* gene and 14% in the *gag* gene. However, in spite of the extensive variation of virus sequences within each HIV-1 subtype, a recent study suggested that intrasubtype genetic heterogeneity is limited by a set distance to the subtype consensus, indicating that HIV-1 genetic subtypes are entities which are stable in the course of the AIDS epidemic (Lukashov and Goudsmit, 1997).

According to recent HIV-1 nomenclature by Robertson *et al.* (2000), in order for a new virus isolate to be designated as a new subtype, the following criteria should be full filled:

1. At least three full-length sequences of the isolates should be available.
2. A new isolate should be roughly equidistant from all previously characterized subtypes in all regions of the genome with a distinct subtype branch similar to those of other subtypes.
3. The isolates should have been found in at least three epidemiologically unlinked individuals.

Currently besides the subtypes, inter subtype recombinant viruses are circulating in the population and are included in the HIV-1 nomenclature. At present, about more than 10% of the virus isolates are found to be recombinants (Cornelissen *et al.*, 1996; Jetzt *et al.*, 2000). Some of these mosaic strains are epidemic strains and are referred to as Circulating Recombinant Forms (CRFs). There are currently four defined CRFs, the CRF01_AE (CM240) in Southeast Asia (Carr *et al.*, 1996), the CRF02_AG (IbNg) of West central Africa (Carr *et al.*, 1998), the CRF03_AB (KAL153) strain in Russia (Bobkove *et al.*, 1998) and the CRF04-cpx in Cyprus (Gao *et al.*, 1998). The previously subtypes "E" and "I" are

problematic and subtype "E" virus is now designated as CRF01_AE and the previously subtype "I" as CRF04_cpx (Robertson *et al.*, 2000). Mosaic viruses with regions that resemble four or more subtypes will be called complex, and designated cpx.

1.7.1. Global distribution of HIV-1 subtypes

The HIV-1 subtypes are unevenly distributed in the world. Subtype B virus predominates in Europe and the United States, subtypes B and F in Brazil, subtypes B and E in south eastern Asia (Wang *et al*, 1998, UNAIDS, 1998), and subtype C in China and India (Lou *et al.*, 1995). In Africa the presence of all subtypes is confirmed with the preference of subtypes A, C, and D (Louwagie *et al.*, 1995; Takehisa *et al.*, 1998). Subtype A and its recombinant (A/G) predominate in West Africa (McCutchan *et al.*, 1999), subtype A and D predominate in East and central African countries such as Kenya, Uganda, and Tanzania (Janssens, *et al.*, 1997; Poss *et al.*, 1997; Hu *et al.*, 2000) while subtype C is dominating the AIDS epidemic in Southern and Eastern Africa such as South Africa, Botswana, Tanzania, and Ethiopia (Renjifo *et al.*, 1998; Abebe *et al.*, 1999; van Hermelen *et al.*, 1999).

HIV-1 subtype C is the most spreading and it is dominating the epidemic in Southern and Eastern Africa, India, and China and 50% of new HIV-1 infections in Africa are due to infections with this subtype. This rapid increase in the prevalence of subtype C virus in Africa and Asia may be due to behavioural or biological cofactors that facilitate transmission and the selective characteristics of the subtype C virus *env*elope (Janssen *et al.*, 1997; van Hermelen et al., 1999).

1.8. Differences in biological properties of HIV-1 isolates

Genetic variation of HIV-1 may be reflected in phenotypic differences that may determine transmissibility, pathogenesis and immunopathogenecity (Janssens *et al.*, 1997). HIV-1 variants have been shown to differ in biological properties such as replication rate, cell tropism and capacity to form multinucleated giant cells (syncytia), SI capacity (Schuitemaker *et al.*, 1992; Coffin *et al.*, 1995). Based on these phenotypic characteristics, HIV-1 variants are classified as slow/low Vs rapid/high, M-tropic Vs T-tropic, and SI Vs NSI, respectively. The M-tropic viruses replicate in macrophages, use CCR5 as coreceptor, and are of the NSI type. The NSI viruses are the transmitting viruses and they are mainly found at early or asymptomatic stage of HIV-1 infection (Zhu *et al.*, 1993). But as AIDS progresses, the SI strains appear in about 50-70% of infected individuals (Schuitemaker *et al.*, 1992). The T-tropic viruses infect T cell lines, are of the SI type, and use primarily CXCR4 as coreceptor, and CCR5 to a lesser extent. For subtype B viruses, the SI/NSI phenotype is associated with the presence of positively charged amino acids at positions 11 and 25 of the gp120 V3 region. The more the positively charged is the V3 loop, the higher the SI phenotype (Fouchier *et al.*, 1992). But for subtype C viruses, the emergence of SI viruses as disease progresses is rare (Tscherning *et al.*, 1998; Abebe *et al.*, 1999; Bjornal et al., 1999). A study on Ethiopian AIDS patients have shown the SI phenotype to be rare and when present was associated with positively charged amino acids at positions 8 and 29 (Abebe *et al.*, 1999).

1.9. HIV-1/AIDS in Ethiopia

In Ethiopia, 3.0 million people are estimated to be infected by the virus currently (MOH, 1999). The HIV-1/AIDS epidemic started relatively late in the country compared to other

African countries, the first HIV-1 positive sera being detected in 1984 and the first AIDS case reported in 1986 in Addis Ababa (Lester *et al.*, 1988; Tsega *et al.*, 1988). However, since then the virus has spread all over the country and has reached the following prevalence: 7-20% in the 15-20 year age groups in the urban areas by 1998, 18% in antenatal care attendants in Addis Ababa by 1996 (Fontanet *et al.*, 1998), and 74% among CSW of Addis Ababa visiting sexually transmitted disease (STD) clinics by 2000 (Aklilu *et al.*, 2001).

1.10. HIV-1 subtypes in Ethiopia

Several phylogenetic analysis done on the gp120 V3 and *gag* sequences of Ethiopian HIV-1 isolates have shown the Ethiopian HIV-1/AIDS epidemic to be dominated by subtype C viruses (Ayehunie *et al.*, 1991; Abebe *et al.*, 1997; Assefaw *et al.*, 1999; Abebe *et al.*, 2000a; Hussien *et al.*, 2000;). The first HIV-1 subtype C sequence was reported in 1991 (Ayehunie *et al.*, 1991) followed by partial *gag* and *env* sequences in 1993 (Ayehunie et al., 1993). The first full length Ethiopian subtype C sequence (ETH2220) was reported in 1996 from a 1986 Ethiopian sample (Salminen *et al.*, 1996) and the second full length Ethiopian sequence in 1998, documenting the first evidence of a subtype A/C recombinant from a 1991 Addis Ababa sample (Sherefa *et al.*, 1998). Sequence analysis done on samples collected from different population groups of Addis Ababa have shown 98.9% of the sequenced isolates to be subtype C and only one subtype A (Abebe *et al.*, 1997). Subsequent study on samples from six different towns of the country has further proved the dominance of subtype C out of Addis Ababa with only two non-C infections, subtype A from Addis Ababa and subtype D from Dessie. (Abebe *et al*, 2000). Related works by Hussien *et al* (2000) on CSW of Addis Ababa and Assefaw *et al*, (1999) on cohort participants of Akaki

and Wonji also confirmed the previous reports. In general, the Ethiopian epidemic is shown to be almost entirely by subtype C viruses with only few non-C viruses all of which were later found to be recombinants of subtype C in the different regions of the virus genome (Abebe, 2000). Ethiopia being surrounded by countries where non-subtype C viruses, mainly A and D, predominate, the fact that the epidemic is dominated by subtype C might be due to the rapid saturation this subtype among the CSWs which are important in heterosexual network (Abebe *et al.*, 2000).

The Ethiopian HIV-1 subtype C strains have been shown to have extra Nuclear factor kappa B (NF-kB) binding site at the promoter region. NF-kB is a DNA binding protein that binds to an enhancer motif in the promoter region of the IL2 gene to induce transcription of this gene. However, HIV has also homologues enhancer sequences at the LTR of the virus and binding of NF- kB initiates virus transcription. Although the biological significance of this extra site is not clearly known, it is suggested that it might contribute to the fast replication rate of the virus (Johansson *et al.*, 1995; Salminen *et al.*, 1996).

1.10.1. The HIV-1 subtype C genetic subcluster, C'

Studies by Abebe *et al.* (1997, 2000a) have reported the Ethiopian subtype C to have a genetic sub cluster designated C' (currently, the C and C' groups have been renamed as C1 and C2, respectively). In phylogenetic trees, the two groups formed separate clusters with significant bootstrap values in the *gag* and *env* regions. At the amino acid level the two groups showed significant differences when compared with the database consensus subtype C sequence and with each other, subtype C sequences being genetically closer to the database subtype C sequence than the C' sequence. The presence of lysine (K) at position

304 and valine (V) at position 294 of the V3 region in the C' group instead of glutamic acid (E) and aspargine (N), respectively found in subtype C sequences could be of interest because the former affects the number of positive charges in the V3 loop and the later leads to the loss of potential N-glycosylation site. The biological significance of this difference is not known although it is known that the number of charges and glycosylation in the V3 loop can affect cellular tropism and neutralization ability of antibodies (Abebe *et al.*, 2000a). The C and C' strains have been co-circulating in the population independent of geography, time of sample collection, and risk group (Abebe *et al*, 2000a). But the frequency of these two groups of viruses among recently infected individuals is not known. A related study by Pollakis *et al.* (2000), has reported intrasubtype recombination between the C and C' viruses with a recombination frequency of 20%. An interesting observation in their study was that all the recombinants carried a subtype C' envelope suggesting a more selective advantage of this envelope in terms of transmission than the subtype C envelope.

Objective

General:

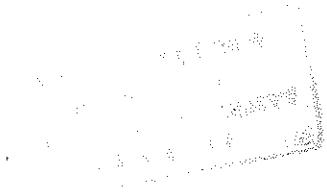
To make phylogenetic analysis of HIV-1 *env* sequences of currently circulating virus isolates in the two cohorts and to assess the diversity of gp120 V3 sequences in terms of synonymous and non-synonymous genetic distances.

To determine the frequency of subtype C and C' viruses.

Specific:

To identify the envelope V3 sequence characteristics of recently transmitted HIV-1 isolates in two cohorts of Ethiopia.

To relate the observed diversity with viral load and CD4 count.



2. Materials and Methods

2.1. Study participants

ENARP has two HIV-1 cohort sites: Akaki and Wonji that enrolled factory workers. Enrolment started on February 26, 1997 in Akaki, and October 7, 1997 in Wonji. In addition, information was available for some study subjects who also participated in pilot studies one to two years before at Akaki (between Dec 20, 1995 and April 2, 1996) and Wonji (between June 20, 1996 and October 21, 1996). HIV status of participants of the pilot study was linked to the actual cohort data, in order to document a possible seroconversion. At present, the cohort has enrolled 1641 factory workers in the two sites (801 Akaki and 840 Wonji). Cohort participants were enrolled only after signing an informed consent form. Withdrawal from the study can be at any time, without obligation to give reasons. As an incentive for their participation, study subjects and their immediate families were offered medical care to the standards of the country for the duration of the study. All participants attended an individual HIV-1 pre-test counselling session and knowledge of HIV status is voluntary. Blood samples were drawn every six months and other clinical and behavioural data are recorded during these visits. Ethical clearance for the study was given by both the National and Institutional, Ethiopian Health and Nutrition Research Institute (EHNRI) Ethical clearance committees.

This study included participants of the two ENARP HIV-1 cohorts who seroconverted between 1997 and 2000. The study population consisted of nineteen individuals, 15 from Akaki and 4 from Wonji, (10 males and 9 females). Data on haematology, immunology, viral load, parasitology, syphilis serology and TB status of participants are recorded in ENARP database. In this study, the nineteen seroconverters were designated by the alphabets A-O and Q-T.

2.2. Blood collection and Plasma isolation

Each participant visits the cohort clinics every six months and blood is withdrawn at each visit. Blood samples were collected in Vacutainer tubes with EDTA (Becton and Dickinson, USA) and transported to the ENARP central laboratory, Addis Ababa. Plasma was isolated by centrifugation at room temperature for 10 min at 1650 rpm and stored at -80°C. Serological, Immunological and virological tests are done on every sample and the data are recorded in ENARP database.

2.3. Enumeration of lymphocyte subsets

Lymphocyte subsets were enumerated on whole blood, using a FACSCAN flow cytometer (Becton and Dickinson, USA), according to the manufacturer's recommendations.

2.4. Viral load determination

Assessment of viral load in plasma was done by using a nucleic acid based amplification assay NASBA, (Organon Teknika, the Netherlands) on 100 ul plasma samples. In brief, RNA was extracted from plasma using an established guanidinium isothiocyanate based method and isothermally amplified in the presence of 3 RNA standards using NASBA kits (Organon Teknika, the Netherlands). Detection of amplified RNA was performed on a NucliSens reader (Organon Teknika, the Netherlands), using the RNA standards for calibration and calculation of viral loads. The lower detection limit of the assay is 80 RNA copies/ml of plasma and results are presented as $\log_{10}$ RNA copies/ml of plasma.

2.5. RNA isolation

Viral RNA was isolated from plasma samples according to Boom *et al.* (1990). This method uses GuSCN for lysing cells and inactivating nucleases and silica for binding nucleic acids.

For this purpose 100 µl of plasma was added to 900ul lyses buffer (L_6), containing GuSCN. 40 µl silica was added to the above mix, incubated for 10 min at room temperature, then centrifuged at 12000g for 30 sec. The pellet was washed with L2 buffer, ethanol and acetone. After washing the pellet was dried at 56°C and RNA was eluted with 50 µl of Low TE^{-4} buffer for 15 min.

2.6. Reverse Transcription (RT)

The viral RNA was reverse transcribed into cDNA by using AMV reverse transcriptase and ED33 primer (Promega kit). Each RT reaction tube contained 25 µl of the reaction mix, which contained: RT buffer (2.5 µl), $MgCl_2$ (5 µl), dNTP's (1.0 µl), RNAsin (0.5 µl), 3'ED33 primer (1.0 µl), AMV-RT (5.0 µl) and RNA (10 µl). The RT tubes were kept at 42°C for 45 min for cDNA synthesis and the reaction was stopped by keeping the reaction tubes at 95°C for 5 min.

2.7. Amplification by polymerase Chain Reaction (PCR)

The cDNA was subjected to two rounds of amplification by PCR, first and second PCR in a thermal cycler (Perkin Elmyer DNA thermal Cycler 480).

2.7.1. First PCR

In the first PCR, a 536 bp DNA fragment flanking the V3 region of the virus was amplified by using Taq polymerase and 5'ED31 primer. Each 1st PCR tube contained 100µl of the reaction mixture, which contained: $MgCl_2$ (10 µl), PCR buffer (8.0 µl), Taq polymerase (0.5 µl), 5'ED31 primer (1.0 µl), dH_20 (55.5 µl), and cDNA (25 µl).

Two drops of mineral oil was added to the PCR tubes to avoid evaporation in the thermal cycler and the cDNA was subjected to amplification under the following cycling conditions:

- 1 cycle of denaturation at 95 ^{0}C for 4 minutes.
- 35 cycles of denaturation, annealing, and elongation at 95^{0}C for 1 min, 55^{0}C for 1 min, and 72^{0}C for 2 min, respectively.
- 1 cycle of extension at 72^{0}C for 10 min.

2.7.2. Second PCR

In the second PCR, a 279 DNA fragment including the V3 region of the HIV-1 *env* gene was amplified in a nested manner by using a pair of primers, 5'V3 M13 and 3'V3 M13.

The 2nd PCR tubes contain 10 µl of the 1st PCR product and 90 µl of the 2nd PCR mix. The 2nd PCR mix contained: PCR buffer (10 µl), $MgCl_2$ (16.0 µl), dNTPs (0.4 µl), Taq polymerase (0.5 µl), 5'V3M13 (1.0 µl), 3'V3M13 (1.0 µl), and, dH_20 (61.1 µl). Finally, two drops of mineral oil was added to the PCR tubes and the following cycling conditions were used for amplification:

- 1 cycle of denaturation at 94°C for 4 min.
- 25 cycles of denaturation, annealing, and elongation at 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min, respectively.
- 1 cycle of final elongation at 72^{0}C for 10 min.

2.8. Agarose Gel Electrophoresis

The 1st and 2nd PCR products were run on 1.5% agarose gel, where 0.5µg/ml Ethidiumbromide (EtBr) was used as a DNA stain, to confirm amplification of the region of interest. 10µl of the PCR product, 5µl of dH_20 and 5µl of loading dye were mixed and loaded on the gel. λ 100 bp DNA fragment was used as a molecular weight marker. The gel was covered with 1500 ml 1X TAE buffer and it was run at 100mA (120V) for 1 hour and

then it was observed under UV light where PCR positive results fluoresce. Finally, Polaroid pictures were taken and the 536bp and 279bp bands were checked

2.9. HIV-1 biological Cloning

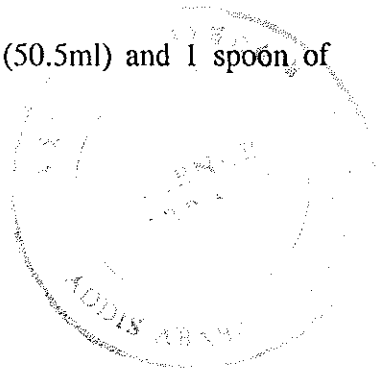
HIV-1 biological clones were generated from PBMC of subject G. Briefly, PBMC (1-2x 10^4 cells per well) from this subject was co-cultivated with PHA stimulated PBMC (10^5) from donor HIV negative subjects. Fresh medium with interluekin 2 and PHA stimulated donor PBMC were added every week. Virus replication was reflected by the presence of HIV-1 P^{24} antigen in the culture supernatant. The virus DNA was isolated from the culture supernatant according to Boom *et al* (1990) and *env* gp120 V3 region was amplified.

2.10. DNA sequencing

The nested PCR products were cycle sequenced with Thermo Sequenase fluorescent labelled primer cycle sequencing kit according to the manufacturer protocol (amersham pharmacia biotech) where the DNA polymerase, sequenase and the fluorescent-labelled primers, A.L.F. M13 17 Base Reverse primer, (5'-Fluorescein-d [CAGGAAACAGCTATGAC]-3') and A.L.F.M13-40 Primer (5'-Fluorescein-d [CGCCAGGGTTTTCCCAGTCACGAC]-3' were used in the cycle sequence reaction.

The cycle sequence reaction mix (Y-mix) containing the primer (2.1µl), dH_20 (14.7µl), DNA (5µl), was prepared and 5µl of this mix was liquated into 4 tubes labelled A, C, G, and T, followed by addition of 2µl of the A, C, G, and T mix in the respective tubes. The cycle sequence tubes were subjected to 1 cycle of denaturation at 95^0C for 4 min followed by 25 cycles of denaturation at 95^0C for 30 sec and annealing/extension at 55^0C for 30 sec.

Sequence gel (120ml) for A.L.F. sequencer was prepared as follows. Urea (50.5 g), accugel (40% (29:1) acrylamide:bisacrylamide solution) (19.4ml), dH_20 (50.5ml) and 1 spoon of



deionizer were mixed and the mix (acryl mix) was left for an hour being stirred. After 1 hour, the mix was filtered and 10X TBE (12ml) was added. Finally, 60 ml TEMED (catalyst) and 30% APS (polymerisation initiator) were added to 60 ml of the acryl mix and this solution was casted into the already prepared sequencing gel plate using a syringe within ten minutes of preparation and the gel plate was left for at least 3 hours for polymerisation. The cycle sequence products were mixed with loading dye and loaded on the gel. Electrophoresis of the cycle sequence products and analysis of the gp120 V3 and flanking regions sequences was performed by an automated DNA sequencer (Pharmacia biotech A.L.F.^{TM} DNA sequencer).

2.11. Phylogenetic analysis

Alignment of sequences was done manually where gaps were introduced for optimal alignment. The sequences were aligned with previously obtained HIV-1 *env* V3 sequences of Ethiopian strains and with consensus HIV-1 sequences for the different subtypes found in the Los Alamos HIV database. DNADIST, NEIGHBOR, DRAWTREE, SEQBOOT and CONSENSE programmes of the PHYLIP package were used for phylogenetic analysis of sequences with the Kimura two-parameter method (PHYLIP, 1996). MEGA (Molecular Evolutionary Genetic Analysis) software was used for synonymous and nonsynonymous genetic distance calculation with the modified Nei-Gojobori method and with the Jukes Cantor correction. (Kumar, 1993). The accuracy or stability of branches was determined by the bootstrap value and a value > 75% was considered as significant. DNA sequences were translated into amino acid sequences using the MEGA software and comparison of amino acid sequences was done by the VESPA software.

3. Results

3.1. Virological and Immunological Data

Data on age, sex, marital status, CD4 T-cell and CD8 T-cell count, and viral load data on the seroconverters during the follow up is depicted on Table 1. The follow up period of the participants ranged from 472 to 1690 days. Of the 19 subjects, 1 (subject H) has died of illness and 1 (subject E) has withdrawn from the study. The geometric mean CD4 T-cell and CD8-T-cell counts for the seroconverters three months after the estimated date of seroconversion were 431 and 1008, respectively. The median $\log_{10}$ RNA copies/ml of plasma at the same visit was 4.07. Four subjects (E, G, M, and S) had a detectable viral load in the pre-seroconversion sample. In particular, subject G had a very high $\log_{10}$ viral load in the two pre-seroconversion samples (4.26 and 7.23) making the number of pre-seoconversion samples with detectable viral load 5. In general, a total of 63 HIV-1 positive plasma samples and 5 pre-seroconversion samples with detectable viral load were collected from the 19 subjects.

Table 1. Follow up data on 19 cohort participants who seroconverted to HIV-1.

Subject ID	Sex	Age	Marital Status	Visit Date	Day	Lab ID	Serology	CD4	CD8	CD4/CD8	Viral load[aa]
A	F	30	Divorced F and married again	13-02-1996	1	0989	N	-	-	-	-
				20-05-1997	462	00352	N	-	-	-	<LDL
				20-11-1997	646	01104	**P**	399	670	0.6	4.02
				03-05-1999	1175	03024	P	207	561	0.4	4.63
				28-12-1999	1414	04196	P	184	406	0.5	4.54
				10-10-2000	1690	05665	P	156	510	0.3	4.93
B	F	29	Married	24-01-1996	1	0677	N	-	-	-	<LDL
				22-05-1997	484 [c]	00371	**P**	419	853	0.5	3.8
				08-12-1997	684	01178	P	563	998	0.6	4.7
				15-06-1998	874	01786	P	472	768	0.6	4.48
				15-12-1998	1057	02549	P	478	683	0.7	5.23
				08-07-1999	1262	03334	P	496	800	0.6	3.84
				03-01-2000	1441	04224	P	450	745	0.6	4.3
				25-07-2000	1644	05239	P	357	743	0.5	4.04
C	M	49	Married	26-05-1997	1	00384	N	-	-	-	-
				02-12-1997	190	C0149	N	-	-	-	<LDL
				10-06-1998	380	01774	**P**	365	469	0.8	4.87
				21-12-1998	574	02580	P	365	948	0.4	5.01
				08-07-1999	773	03335	P	353	806	0.4	4.77
				04-01-2000	953	04235	P	290	534	0.5	4.55
				14-08-2000	1175	05364	P	319	612	0.5	4.64
D	M	36	Single	17-06-1997	1	00474	N	-	-	-	-
				26-05-1998	343	C0394	N	1082	996	1.1	<LDL

				01-12-1998	532	02445	P	578	1033	0.6	2.75
				01-07-1999	744	03298	P	826	1530	0.5	2.40
				12-01-2000	939	04293	P	822	1613	0.5	2.51
				08-08-2000	1147	05334	P	720	1176	0.6	3.58
E*	F	42	Widowed	18-01-1996	1	AP6172/0566	N	-	-	-	3.15
				15-09-1997	606	00703	P	270	1475	0.2	5.45
F	M	40	Divorced A	16-10-1997	1	00788	N	900	551	1.6	<LDL
				26-08-1998	314	02018	P	95	539	0.2	4.52
				25-02-1999	442	02814	P	265	643	0.4	4.65
				30-08-1999	798	03554	P	352	690	0.5	5.58
				27-03-2000	1007	04656	P	413	971	0.4	5.46
G	F	41	Married to HIV+ spouse	17-11-1997	1	01061	N	-	-	-	4.26
				09-06-1998	204	C0420	N	-	-	-	7.23
				14-12-1998	392	02544	P	640	544	1.2	3.95
				21-06-1999	581	03226	P	935	821	1.1	3.96
				21-12-1999	764	04166	P	935	561	1.7	4.18
				12-09-2000	1019	05500	P	758	888	0.9	4.36
H=	M	45	Single	30-07-1996	1	07218	N	-	-	-	<LDL
				13-11-1997	472	00970	P	315	396	0.8	3.11
I	M	36	Married	18-07-1996	1	7159	N	-	-	-	<LDL
				03-12-1997	449	01164	P	350	1014	0.3	4.81
				25-06-1999	626	01828	P	256	948	0.3	4.99
				06-01-1999	989	02636	P	-	-	-	4.62
				20-07-1999	1184	03383	P	251	856	0.3	4.68
					1381	04380	P	342	1390	0.2	4.11
				24-08-2000	1584	05432	P	296	1128	0.3	4.43

J	M	51	DN	03-06-1997	1	00407	N	-	-	-	-
				11-12-1997	135	01210	N	-	-	-	-
				17-06-1998	295	01796	N	-	-	-	<LDL
				14-12-1998	475	02542	N	-	-	-	<LDL
				17-06-1999	660	03212	**P**	416	3024	0.1	4.87
				20-12-1999	819	04160	P	261	1416	0.2	4.97
				03-07-2000	1014	05142	P	181	650	O.3	4.68
K	M	34	Married	26-07-1996	1	7203	N	-	-	-	<LDL
				23-06-1998	559	01813	**P**	776	2236	0.3	5.31
				25-01-1999	775	02691	P	360	1000	0.4	5.04
				07-07-1999	938	03328	P	343	1152	0.3	4.81
				31-01-2000	1146	04344	P	256	971	0.3	5.58
				16-08-2000	1343	05390	P	239	1091	0.2	5.38
L	F	45	Married	03-09-1997	1	00685	N	-	-	-	-
				07-09-1998	369	C0611	N	-	-	-	<LDL
				29-03-1999	572	02902	**P**	824	1323	0.6	<LDL
				20-03-2000	928	04624	P	573	1504	0.4	<LDL
				11-10-2000	1133	05671	P	614	1376	0.4	<LDL
M	F	38	Married	24-09-1997	1	00726	N	-	-	-	-
				28-07-1998	307	C0527	N	-	-	-	<LDL
				10-03-1999	532	C1182	N	-	-	-	4.60
				21-10-1999	675	03837	**P**	688	669	1.0	3.53
				03-05-2000	869	04868	P	423	587	0.7	3.83
				23-11-2000	1073	05921	P	462	620	0.7	3.68
N	F	40	Married	11-12-1997	1	01206	N	-	-	-	-
				22-09-1998	285	02111	N	-	-	-	-
				05-04-1999	480	02917	N	-	-	-	<LDL

				11-10-1999	669	03725	**P**	668	954	0.7	2.48
				03-05-2000	873	04871	P	558	972	0.6	2.71
				28-11-2000	1082	05934	P	721	1416	0.5	<LDL
O	F	48	DN	08-04-1997	1	00227	N	-	-	-	-
				10-11-1997	216	00946	N	-	-	-	-
				21-04-1998	378	01663	N	707	521	1.4	-
				17-11-1998	560	02363	N	-	-	-	-
				26-05-1999	750	03127	N	-	-	-	<LDL
				01-02-2000	1001	04353	**P**	450	1349	0.3	4.20
				15-08-2000	1196	05376	P	396	1065	0.4	3.95
Q	M	33	DN	22-10-1997	1	00844	N	-	-	-	-
				06-05-1998	196	C0346	N	-	-	-	-
				07-12-1998	411	C0913	N	880	618	1.4	-
				30-06-1999	616	C1583	N	752	594	1.3	<LDL
				17-02-2000	848	04469	**P**	416	772	0.5	4.11
				06-12-2000	1140	05983	P	401	711	0.6	3.90
R	M	32	DN	25-03-1997	1	00182	N	447	927	0.5	-
				29-09-1997	188	C0038	N	-	-	-	-
				01-04-1998	372	C0300	N	-	-	-	-
				05-10-1998	559	C0708	N	-	-	-	-
				05-04-1999	741	C1269	N	-	-	-	-

				20-10-1999	939	C1976	N	-	-	-	<LDL
				10-05-2000	1141	04898	**P**	306	1072	0.3	4.00
				27-11-2000	1342	05922	P	414	1350	0.3	4.34
S	M	31	DN	02-06-1997	1	00400	N	-	-	-	-
				09-12-1997	190	C0158	N	-	-	-	-
				16-09-1998	471	C0636	N	-	-	-	-
				15-03-1999	651	C1193	N	-	-	-	-
				14-10-1999	864	C1949	N	-	-	-	4.11
				18-04-2000	1050	04782	**P**	348	2439	0.1	3.18
				23-10-2000	1238	05744	P	276	3588	0.1	3.34
T	F	30	DN	19-08-1997	1	00643	N	-	-	-	-
				12-03-1998	205	01519	N	-	-	-	-
				18-08-1998	364	01996	N	-	-	-	-
				09-03-1999	567	02847	N	-	-	-	-
				03-08-1999	714	03446	N	-	-	-	<LDL
				01-02-2000	896	04356	**P**	699	2429	0.3	4.52
				09-10-2000	1146	05657	P	294	1288	0.2	3.76

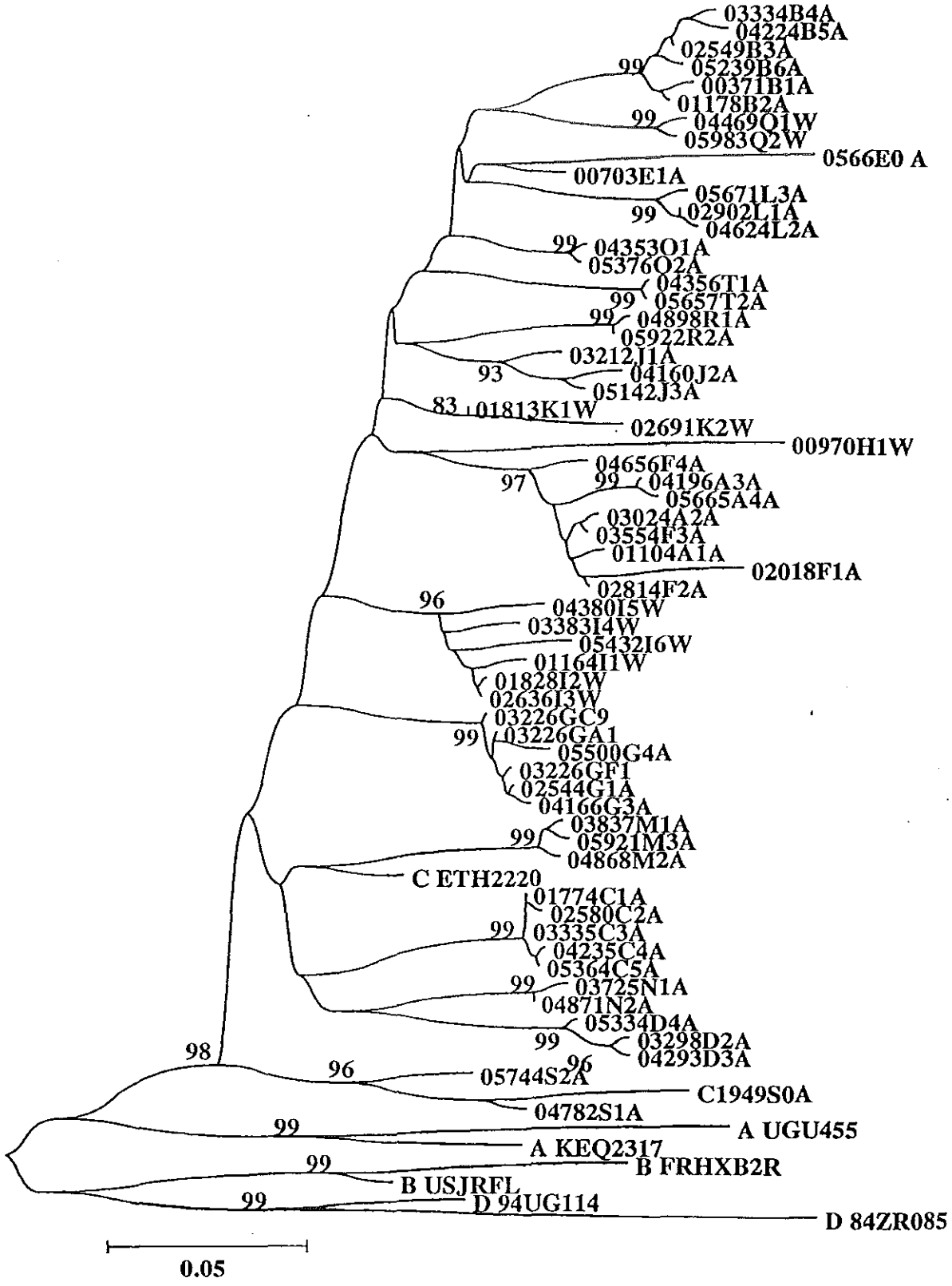
DN= Do Not Know
*Withdrawn
=Died
Subjects in bold have a pre-seroconversion viral load and the first HIV-1 positive visit is shadowed. [a] =log viral load

3.2. PCR and DNA sequencing

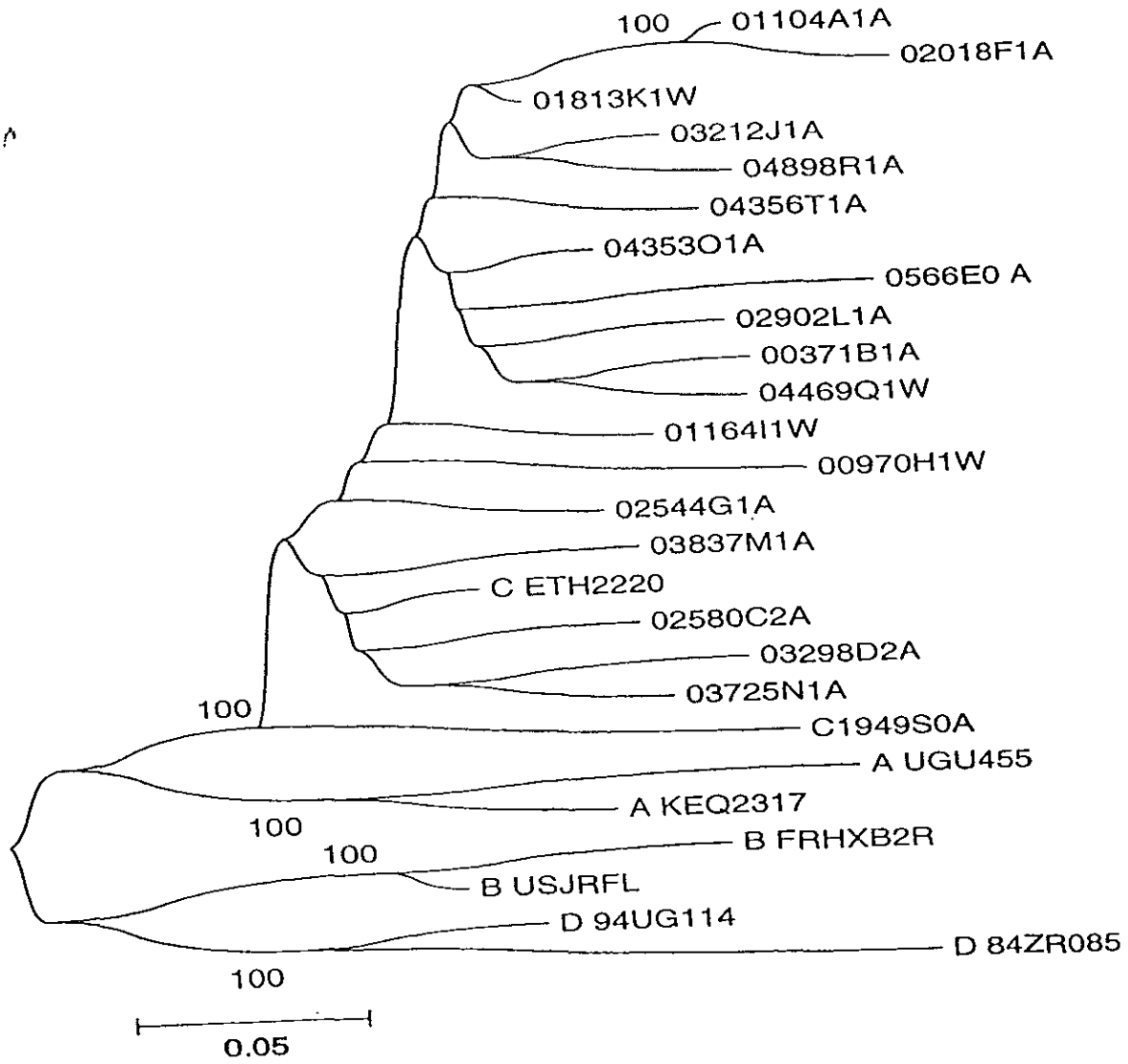
The cDNA was subjected to two rounds of amplification and a 536 bp DNA fragment in the gp120 region encompassing the V3 region was amplified in the first PCR and 279 bp DNA fragment including the C2V3 region was amplified in the nested PCR. 60 out of 63 (95%) HIV-1 positive samples and 2 out of 5 (40%) pre-seroconversion samples with viral load, gave PCR positive results. In total, 62 out of 68 (91%) of seroconverter samples gave PCR positive results. Samples from subject G and M were difficult to amplify with our standard RT-PCR procedure, so a low stringent PCR was used to amplify the gp120 V3 region of the isolates from these samples and a positive PCR result was obtained. Three clones (03226GF1, 03226GC9, and 03226GA1) were generated from PBMC of subject G (03226G2) and all were sequenced and included in the phylogenetic analysis. A total of 65 PCR positive products were subjected to sequencing of which 61 (94%) gave interpretable DNA sequences.

3.3. Phylogenetic analysis

All the gp120 V3 sequences were aligned with previously published subtype C sequence and reference sequences from other subtypes. A phylogenetic tree generated by a neighbour joining method is shown in Figure 4A demonstrating that all the seroconverter sequences belong to subtype C. As clearly seen in the tree, sequences of an individual were clustered together with high bootstrap values. It can also be seen that sequences from subject A and F who were married clustered together with a bootstrap value of 97%. Figure 4B was generated by taking the sequences from the first seropositive plasma samples.


03334B4A
04224B5A
02549B3A
99
05239B6A
00371B1A
01178B2A
99
04469Q1W
05983Q2W
0566E0 A
00703E1A
05671L3A
99
02902L1A
04624L2A
99
04353O1A
05376O2A
04356T1A
99
05657T2A
99
04898R1A
05922R2A
03212J1A
93
04160J2A
05142J3A
83
01813K1W
02691K2W
00970H1W
04656F4A
97
99
04196A3A
05665A4A
03024A2A
03554F3A
01104A1A
02018F1A
02814F2A
96
04380I5W
03383I4W
05432I6W
01164I1W
01828I2W
02636I3W
03226GC9
99
03226GA1
05500G4A
03226GF1
02544G1A
04166G3A
03837M1A
99
05921M3A
04868M2A
C ETH2220
01774C1A
02580C2A
99
03335C3A
04235C4A
05364C5A
99
03725N1A
04871N2A
05334D4A
99
03298D2A
04293D3A
98
96
96
05744S2A
C1949S0A
04782S1A
99
A UGU455
A KEQ2317
99
B FRHXB2R
B USJRFL
99
D 94UG114
D 84ZR085
0.05

A.

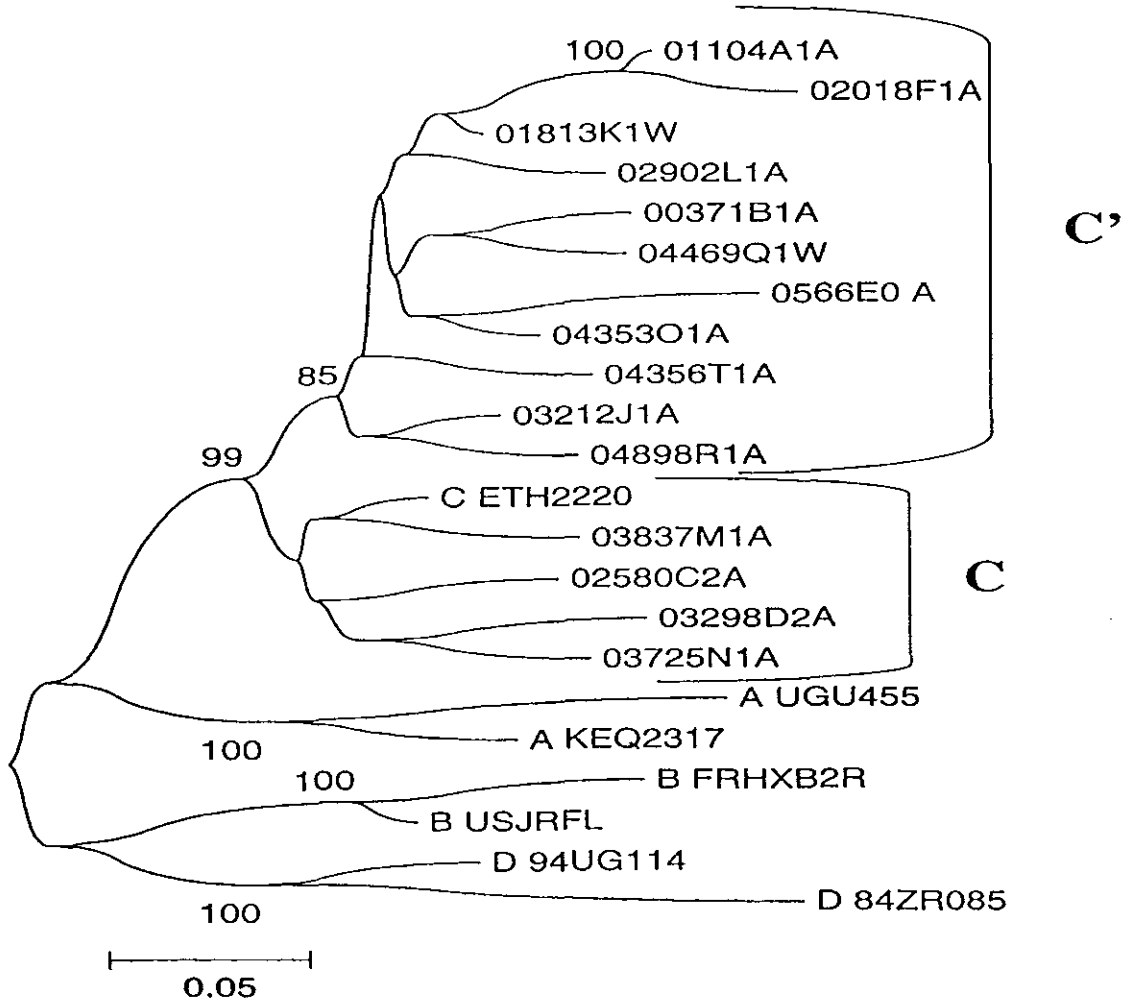


B.

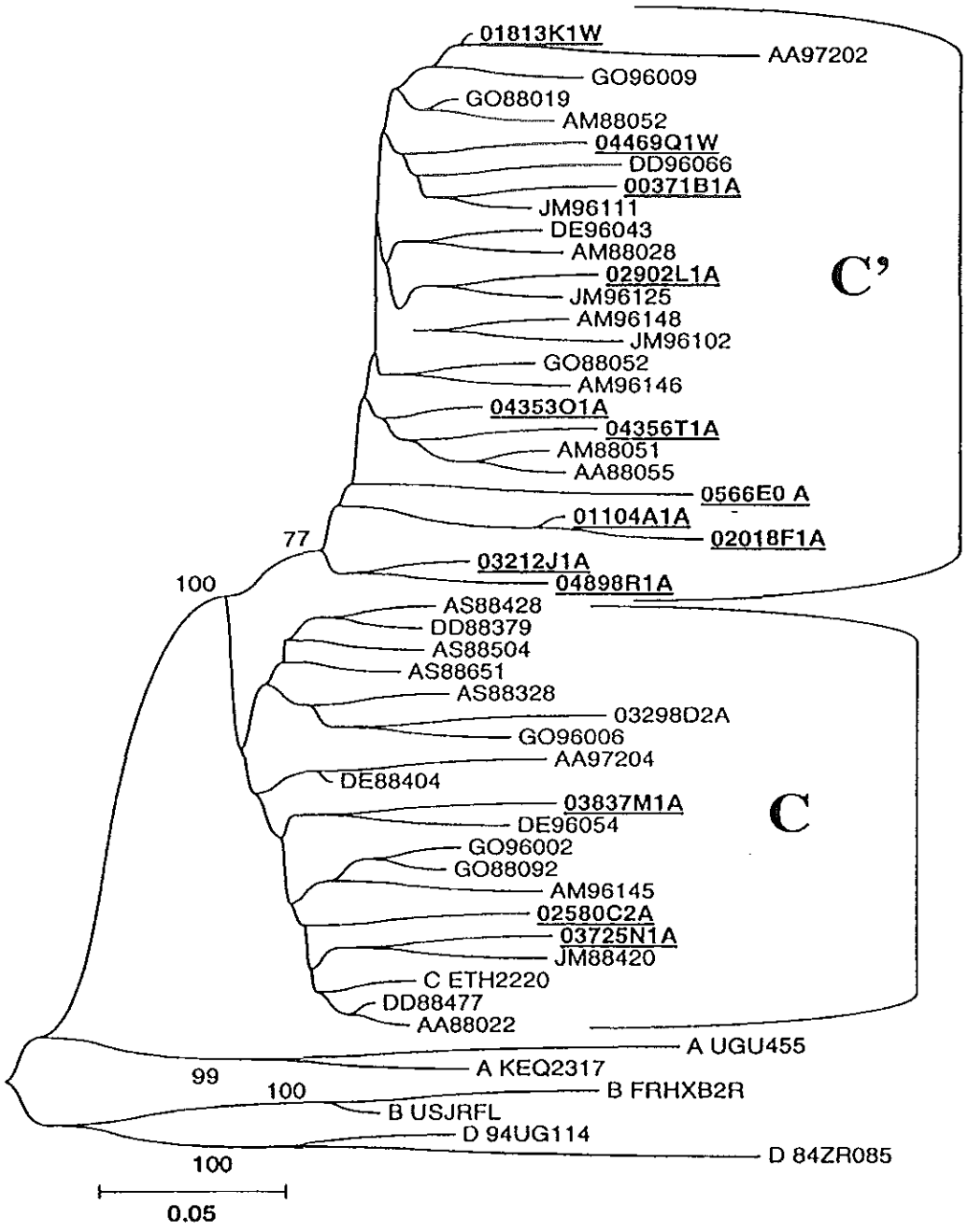
Figure 4. Phylogenetic analysis of gp120 V3 sequences of HIV-1 isolates from the 19 seroconverters.
The numbers in the seroconverter sequences represent the lab ID of the sample from which the virus is isolated, the alphabets (A-O, Q-T) refer to the subject ID, the numbers after the subject ID refer to the visit number (0= before seroconversion sample, 1=seroconversion sample etc). The letters at the end, A and W means Akaki and Wonji, respectively. The numbers at the branch node refer to the bootstrap value out 100 replications. The scale bar represents a genetic distance of 5%. (A) A neighbour-joining tree of all the 61 seroconverter sequences and seven reference sequences of subtypes C (ETH2220), A (UGU455 and KEQ2317), B (FRHXB2Rand USJRFL), and D (94UG114 and 84ZR085). (B) A neighbour-joining tree generated by taking a single gp120 V3 sequence from each individual. Sequence of HIV-1 isolates from the first HIV positive sample are taken (except for subject C and D).

3.4. HIV-1 subtype C genetic subcluster (C')

Neighbour joining trees of gp120 V3 seroconverter sequences are shown in Figure 5A and 5B where a single sequence from each individual is included. The trees clearly indicated two clusters, the C and C' sequences with high boot strap values (85% and 99%) in Fig. 5A and (100% and 77%) in Fig. 5B. As can be seen from the trees, 15 sequences out of 19 (79%) were classified as C and C'. Out of these 15 sequences, 11 were C' (73%) and only 4 were C (27%). Four of the sequences from subject S, G, H, and I, were difficult to be grouped as C or C' (data not shown). Besides, the Ethiopian reference sequence ETH2220 clustered with C group of sequences.



A.



B.

Figure 5.Phylogenetic analysis of gp120 V3 seoconverter sequences showing subtype C and the subcluster C' sequences. (A) Phylogenetic analysis of 15 HIV-1 gp120 V3 seroconverter sequences showing the main subtype C and the subcluster C'. The scale bar represents a genetic distance of 5%. (B) Phylogenetic analysis of 15 HIV-1 gp120 V3 seroconverter sequences and previous HIV-1 gp120 V3 sequences from samples of different towns showing the main subtype C and the subcluster C' (Abebe *et al.*, 2000a). The previous HIV-1 gp120 V3 sequences are indicated by the codes, AA= Addis Ababa, Am= Arbaminch, As= Assab, DD= Dire Dawa, DE, Dessie, GO= Gondar and JM= Jima. The seroconverter sequences are in boldface and underlined. Seven reference sequences of subtypes C (ETH2220), A (UGU455 and KEQ2317), B (FRHXB2Rand USJRFL), and D (94UG114 and 84ZR085) are included.

3.5. Predicted amino acid sequence

Figure 6 shows the deduced consensus gp120 C2V3 amino acid sequences of the HIV-1 isolates from 19 seroconverters. All the nucleotide sequences of an individual are translated into amino acid sequence and a single consensus sequence is generated for each subject. The length of the V3 loop, which is 35 amino acids long and is surrounded by the two cysteine residues was conserved in all the sequences. The GPGQT (glysine-proline-glysine-glutamine-threonine) motif, typical of subtype C viruses was conserved in all the 19 amino acid sequences. The first aspargine (N) at the three NNN residues, glycosylation site, was changed into glysine (G) in 7 of the 19 sequences while it was changed into serine (S) in two subjects, and completely changed into Glysine-Isoleucine-Isoleucine (GII) in subject H. The subtype C and C' group of viruses were also distinguished based on the amino acid sequences. Four sequences (G, H, I and S) were difficult to group as C and C'. Close look at the amino acid sequence showed subject S sequences to be different while the sequences from subject H, I, and G to have amino acid characteristics of both groups.

3.6. Amino acid sequence comparison of the subtype C and C' virus

Comparison between seroconverter C and seroconverter C' amino acid sequences (SER C and SER C') revealed significant amino acid changes at 12 amino acid positions most of which are outside of the V3 loop. Further closer look in the V3 loop region revealed two significant amino acid changes where the Glutamate (E) at position 10 and Isoleucine (I) at position 12 (start counting 1 at the first cysteine) in subtype C viruses were substituted by Lysine (K) and Methionine (M), respectively in subtype C' viruses (Fig. 7A). Comparison of Los Alamos consensus subtype C (LSA C) amino acid sequences with the two consensus seroconverter sequences, SER C and SER C' (Fig. 7B) showed a significant amino acid

change, the seroconverter C groups being genetically more closer to the Los Alamos consensus C. Comparison of previous Ethiopian C and C' sequences with the seroconverter C and C' showed a similar amino acid sequences except at position 290 where the proline (P) in Eth C and C' sequences is replaced by serine (S) in SER c and C'. Besides, the amino acid differences (N→K) at position 288 between Eth C and C' not observed between SER C and SER C' but a difference is observed at position 289 (E→K) between SER C and C' which is not observed between Eth C and C'. Finally as can be seen in figure 7, all groups showed a greater variability towards the right end, which is the most variable region.

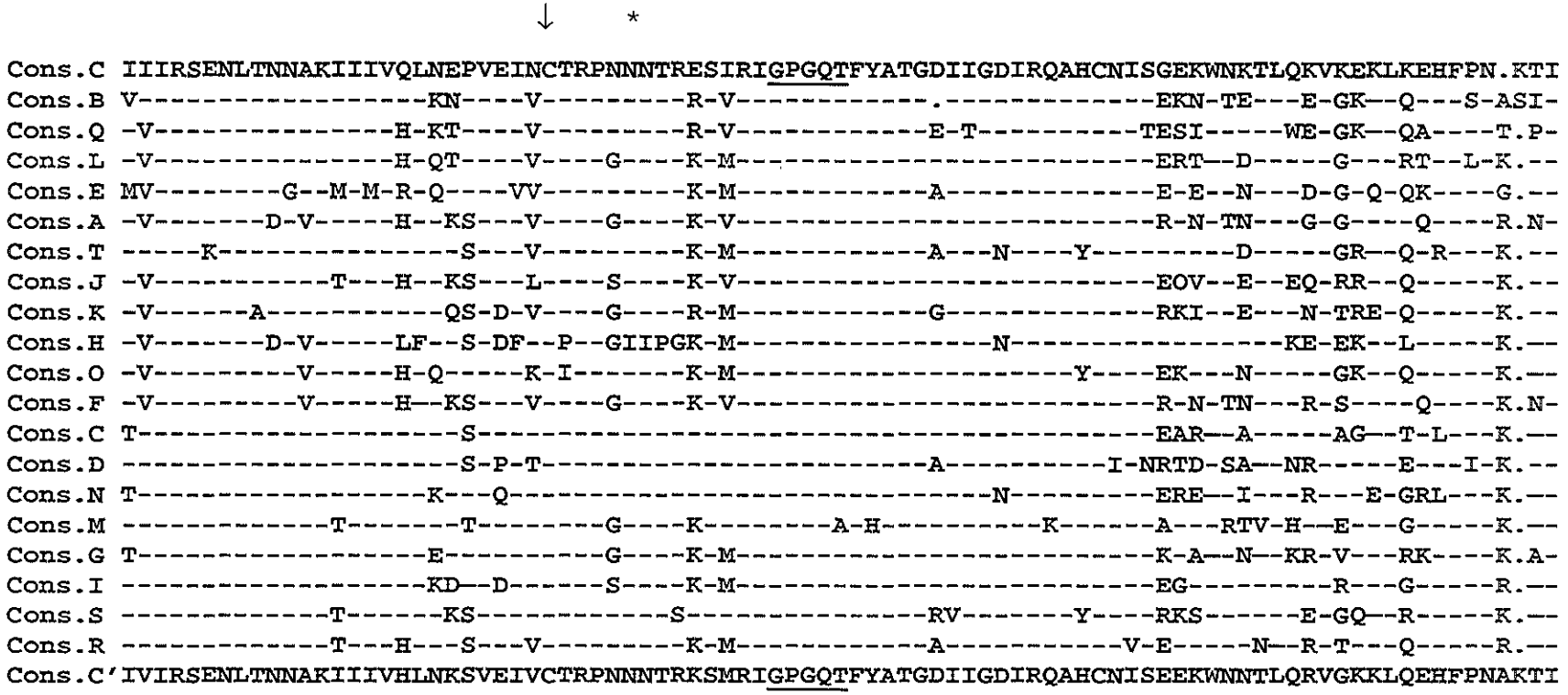


Figure 6. Predicted consensus gp120 C2-V3 amino acid sequence for the 19 seroconverters (↓ = Cysteine residues, * = N-linked glycosylation site. The GPGQT motif is underlined).

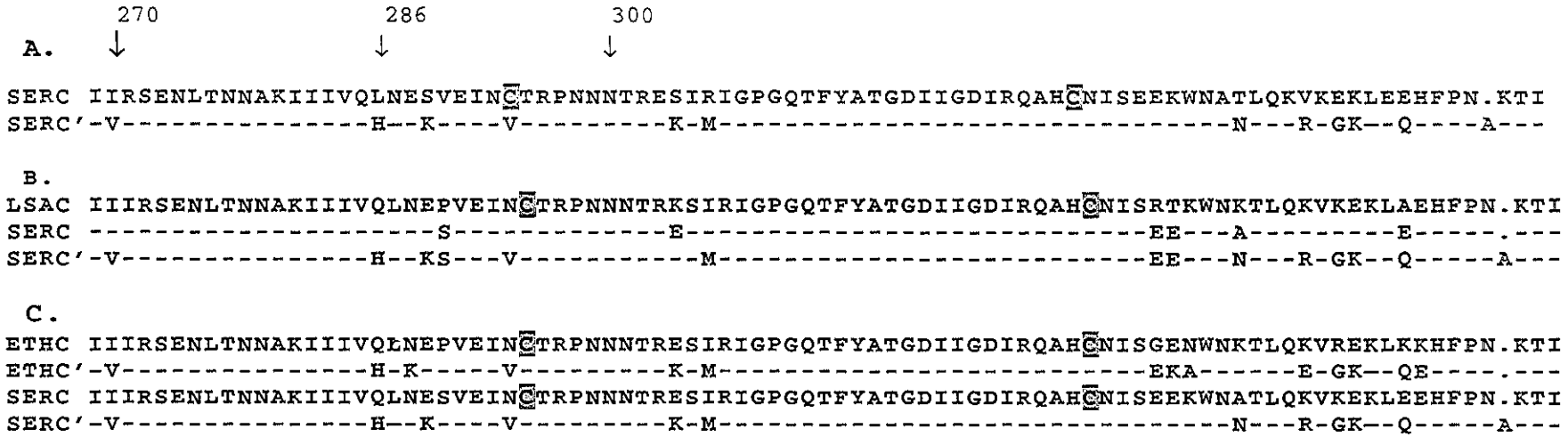


Figure 7.VESPA supported amino acid sequence comparison between different groups of consensus sequences. (A). Comparison between seroconverter C (SER C) and seroconverter C' (SER C'). (B). Comparison between Los Alamos (LSA C), SER C, and SER C'. (C). Comparison among consensus Ethiopian C (Eth C), consensus Ethiopian C' (Eth C'), SERC, and SER C'. Dashes represent amino acid identities and a dot represents gap or deletion. The two cysteine residues are shaded.

3.7. Synonymous and nonsynonymous distance calculation

The pair wise synonymous (ds) and nonsynonymous (dn) genetic distances among different sequence groups were calculated along the synonymous and nonsynonymous sites. Table 2 and Figure 8 summarise the mean synonymous and nonsynonymous distances with in different groups of sequences. There was a 17% ds and 15% dn among all the seroconverter sequences. The synonymous distance among subtype C' seroconverter sequences (8.8%) was less compared to subtype C seroconverter sequences (16%) while the nonsynonymous distance is comparable (15% and 16%). Calculation of genetic distance between subtype C and C' (C/C') viruses resulted in a 24% ds and 17% dn.

Table 3 and Figures 8 and 10 summarize the intra person ds and dn for 10 individuals with a minimum of three follow up sequences. Only 4 subjects (B, F, I, and S) out of ten had a significant ds while 8 subjects had a significant dn value. Sequences from subject S had the highest mean dn (10%) value while subject F had the highest ds value (6%).

The ds/dn ratio for 4 subjects (B, F, I and S) was also calculated.

Table 2. Mean synonymous (ds) and nonsynonymous (dn) distance.

Subtype	N	Mean ds (Standard deviation)	*p value	Mean dn (Standard deviation)	p value
C	10	0.1600 (±0.06379)	<0.0001	0.1572 (±0.02700)	<0.0001
C'	55	0.08847 (±0.02948)	<0.0001	0.1458 (±0.03490)	<0.0001
C/C'	44	0.2400 (±0.05697)	<0.0001	0.1666 (±0.03642)	<0.0001
Total	171	0.1749 (±0.07695)	<0.0001	0.1515 (±0.04194)	<0.0001

* significant at p<0.05

Table 3. Mean intra person synonymous (ds) and nonsynonymous (dn) distance

Subject	n	Mean ds (Standard deviation)	p value	Mean dn (Standard deviation)	p value	ds/dn
A	6	0.01993 (±0.01977)	0.0565*	0.02918 (±0.01639)	0.0073	
B	15	0.00584 (±0.008549)	0.0192	0.02511 (±0.008840)	<0.0001	0.23
C	6	0.00845 (±0.009257)	0.0756*	0.00575 (±0.008052)	0.0130	
D	0	0	0	0.0173 (±0.008052)	0.0652*	
F	6	0.05907 (±0.02818)	0.0037	0.0382 (±0.01679)	0.0026	1.55
Ga†	3	0.01197 (±0.01036)	0.1835*	0.01623 (±0.005658)	0.0382	
Gb†	3	0.01203 (±0.01042)	0.1835*	0.006467 (±0.005600)	0.1835*	
I	10	0.00296 (±0.003822)	0.0368	0.04234 (±0.02004)	<0.0001	0.07
J	3	0.01753 (±0.008852)	0.0755*	0.04507 (±0.01784)	0.0484	
L	3	0.0117 (±0.01013)	0.1835*	0.0116 (±0.005719)	0.0723*	
S	3	0.05563 (0.01987)	0.0400	0.1002 (±0.03266)	0.0336	0.55

* Not significant at $p<0.05$

† Ga and Gb refer to diversity among follow up sequences and diversity among the three clones from a single plasma sample of subject G, respectively.

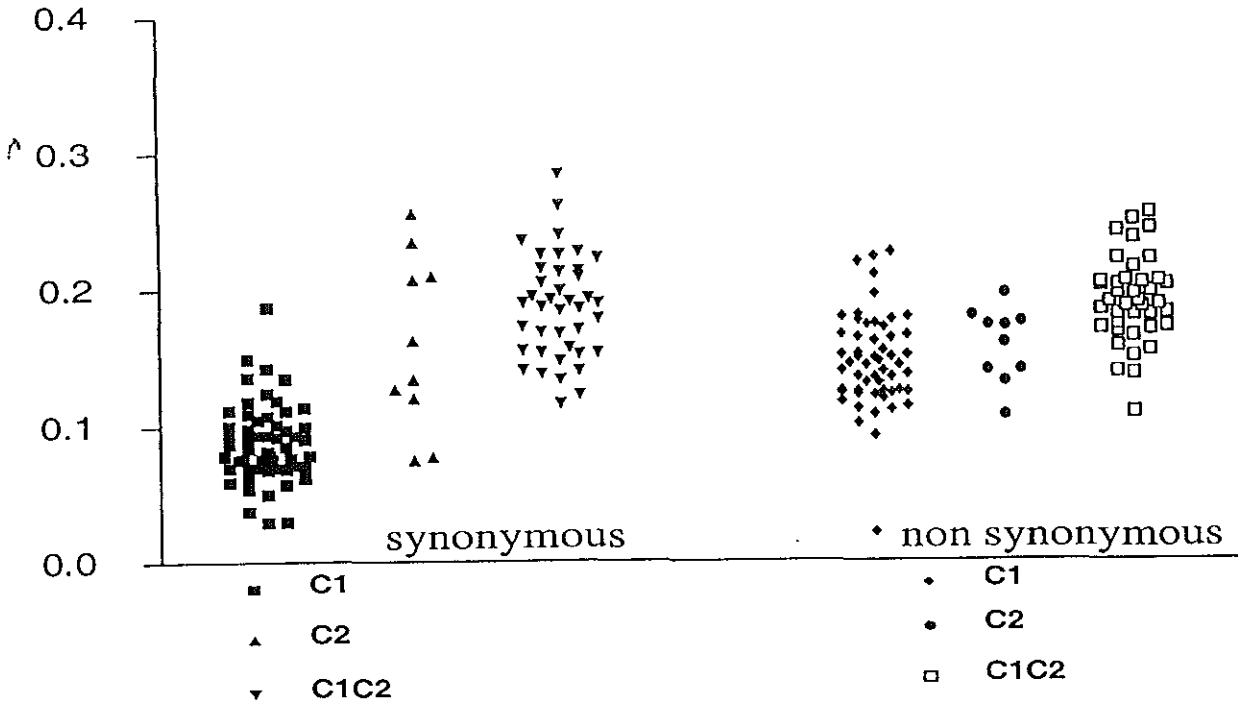


Figure 8 Synonymous and nonsynonymous distance distribution among subtype C (C1), C' (C2), and between C and C' (C1C2) sequences.

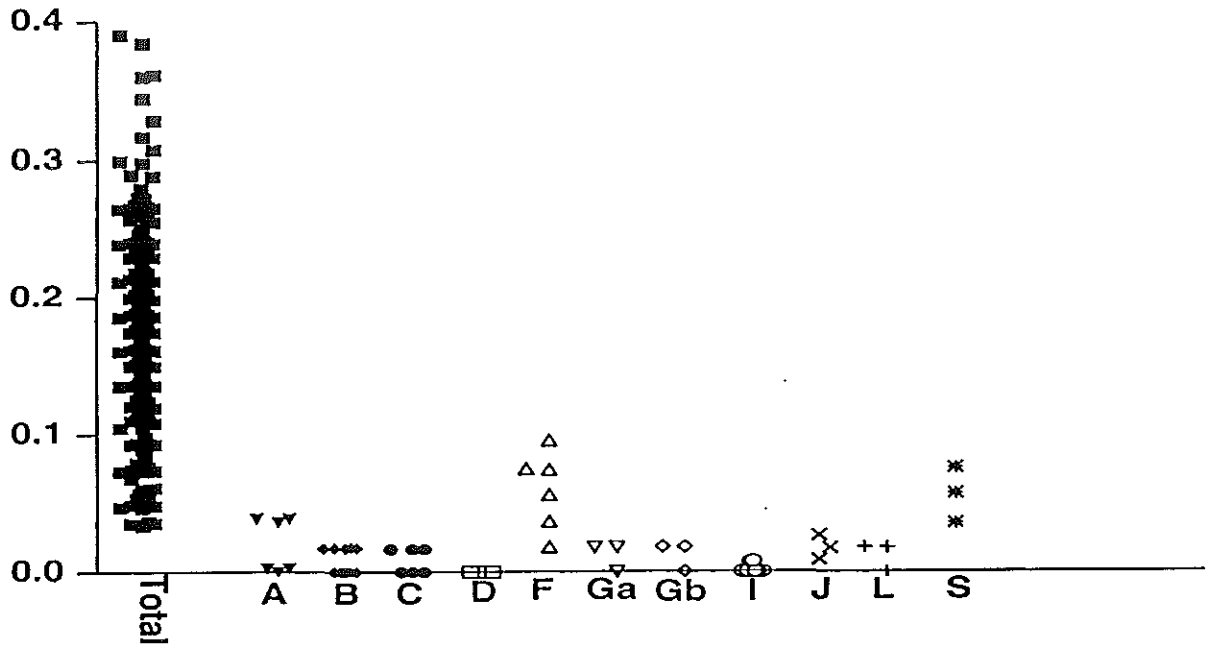


Figure 9 Synonymous distance variation among all seroconverter sequences (total) and among sequences of an individual.

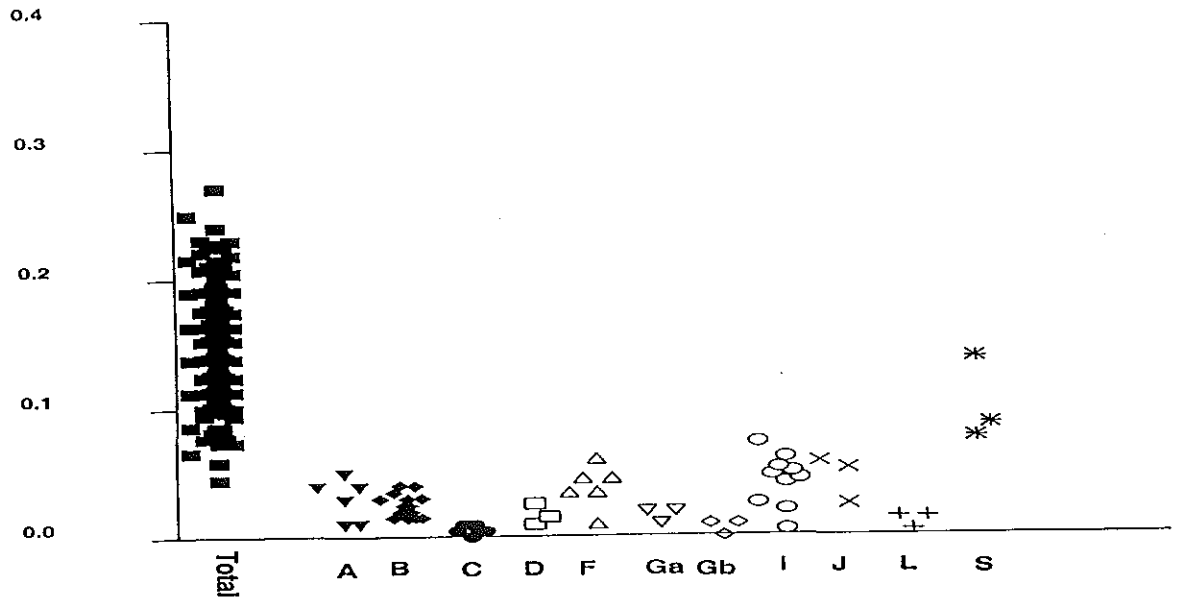


Figure 10 Nonsynonymous distance variation among all seroconverter sequences (total) and among sequences of an individual.

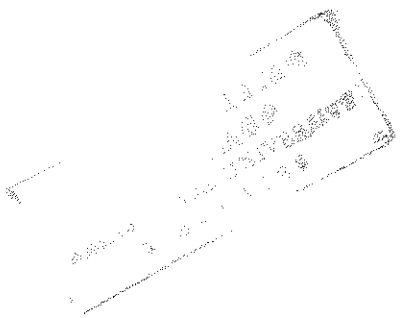
4. Discussion

HIV-1 is the most diversifying and evolving Retrovirus (Sharp *et al.*, 1995). The virus is increasing its diversity with time, and new virus isolates with better transmissibility are evolving (Kuiken *et al.*, 1993; Auwanit *et al.*, 2000, McCutchan et al., 2000). Because the ancestor of this virus has jumped into the human host, which is not the natural host, the virus is highly diversifying to explore the new environment, the human host (Menting, 2001).

The high variation of the virus and its capacity for rapid adaptation is challenging the effort for the development of successful vaccines and anti HIV-1 chemotherapeutic agents. Therefore, genetic variability of HIV needs to be taken into account when developing or adopting vaccines and diagnostic tests.

Vaccine strategies should target virus strains that are currently transmitting in the general population. Therefore, recently infected patients provide an opportunity to study the type of virus that has successfully transmitted from person to person (Zhu *et al.*, 1993; Wang *et al.*, 1998).

Since HIV-1 subtype C is highly spreading in Southern and Eastern Africa, China and India with 50% of new infections in the world being by this subtype, vaccines targeting subtype C viruses should be considered in the vaccine designs and information regarding this subtype will be of worth having (Janssens*et al.*, 1997; Renjifo *et al.*, 1998; UNAIDS, 1998).



In the present study, 19 individuals who seroconverted for HIV-1 antibodies in the last three or four years are included. The subjects had a median log_{10} viral load of 4.07 (n=19) three months after the estimated date of seroconversion, which is lower compared to Dutch homosexual and intravenous drug user seroconverters (4.72, n=157, p=0.03) (Table 1). When the viral load analysis was restricted to the comparison of Dutch intravenous drug users (n=52, 40% females) with Ethiopians (n=16, 50% females), the viral load at 3 months was still lower among Ethiopian compared to Dutch seroconverters (median of 4.07 and 4.78, respectively, p=0.04). The lowered viral load in Ethiopian seroconverters could be due to the low CD4 count available for the virus to replicate. Likewise, subtype C-infected seroconverters from Zimbabwe had some how similar plasma log viral load (n= 21, geometric mean= 4.39, range= 3.2-5.5, Tien et al., 1999) at the second seropositive visit (2 weeks to 6 months from seroconversion).

The geometric mean CD4 T-cell count three months after the estimated date of seroconversion for these subjects was 431/ul (n=19), which is remarkably low (p=0.008) when compared with CD4 count for Dutch homosexual seroconverters (631/ul, n=132, Dawit Wolday, personal communication). This difference could be seen in light of the already decreased CD4 count in adult HIV-1 negative individuals compared to their Dutch counter parts (Messele *et al.*, 1999). However, the CD4 count for Ethiopian seroconverters is some how comparable with the subtype C-infected seroconverters from Zimbabwe (n=21, mean CD4 count 506, range= 148-1450, Tien et al., 1999).

In general, the median $\log_{10}$ viral load for the 19 seroconverters showed an increasing tendency where as, the CD4 count showed a decreasing tendency during the follow up period (Table 1).

Subject L had plasma viral load less than the lower detection limit (<LDL) in all the three seropositive visits while all these samples gave PCR positive results. The above discrepancy could be explained by the differences in the region of the virus genome amplified by the two techniques(NASBA and PCR) where the gag gene is amplified in NASBA while the env gene is amplified by PCR. Subject N had a decreased viral load (<LDL) in the third visit that was associated with an increase in CD4 count.

The cDNA was amplified by PCR with a PCR success rate of 91% which is good as compared to previous sequencing works at ENARP with 81% PCR success rate (Assefaw, 1999). Only 6/68 (9%) of the samples failed to give PCR positive products which could be due to several reasons such as failure of the specific primers to anneal to target sequences, lack of enough target RNA, and technical errors. However, the samples that failed to be amplified had a detectable viral load ruling out the possibility of lack of target RNA and leaving technical errors and primer failure as possible causes.

Isolates from subject G and M were difficult to amplify in the gp120 V3 region using our standard RT-PCR procedure and three clones were generated from PBMC of subject G (03226G2). Amplification of the gp120V3 region of the HIV-1 isolates from the rest of the samples from these subjects was possible through low stringent PCR. All the sequences of subject M clustered together. Sequences of the three clones from subject G clustered with sequences of the low stringent PCR product with high bootstrap value (99%) confirming the

correctness of our sequences. Our data showed subject G, a 41-year-old woman, to have some striking features. This subject had a very high $\log_{10}$ viral load (4.26) at intake while HIV-1 seronegative and remained seronegative even in the next visit after 204 days with high $\log_{10}$ viral load (7.23) (Table 1). This individual had the highest CD4 T-cell count (640/ul) in the first HIV-1 seropositive visit and a very high and stable CD4 count in the next visits. The fact that there is no decline of CD4 T-cell count in this subject and no inversion of CD4/CD8 T-cell ratio in the first seropositive visit could lead to the conclusion that the individual might be infected with an aberrant virus strain or the individual might have a unique strong immune background (Rinke de Wit *et al.*, 2000). Besides, virus isolates from this subject were difficult to amplify and subsequent amino acid sequence characterization of the gp120 V3 region from this subject revealed unique amino acid residues, which supports the above conclusion.

The clustering of sequences of an individual in the phylogenetic tree gives evidence for the validity of our sequences. This is further confirmed by the clustering of sequences of subject A and F who were married and who probably infected one another (Fig. 4A). Another remarkable observation in the tree was the long branch length of the pre-seroconversion sample isolates from subject S and E (C1949S0 and 05660E0). Based on this observation we might speculate that the HIV-1 strains found before seroconversion are more diverse than the later sequences.

Phylogenetic analysis of the gp120 V3 seroconverter sequences further confirmed the dominance of the Ethiopian epidemic by subtype C and the presence of a genetic subcluster (C'), among seroconverters. In the present study, the C and C' sequences cluster separately

with high bootstrap values (99 % and 85%) (Fig.5A). However, not all sequences were grouped as C and C', only 15 sequences out of 19 were grouped as C and C'. Interestingly, 73% of these sequences were subtype C' while the rest 27% were subtype C. Therefore, our data showed the spreading of C' viruses among the study population which might be due to the selective advantage in terms of transmission of C' viruses. In fact, this suggestion is in agreement with previous report by Pollakis *et al.* (2000) on the presence of C/C' recombinant viruses, where all the recombinants carried the C' virus in the *env* region. Our data together with reports by Pollakis *et al.*, (2000) call for the need to consider subtype C' viruses in future vaccine designs. In the present study, the Ethiopian subtype C reference sequence, ETH2220, cluster with group C seroconverter sequences which is in agreement with previous work by Abebe *et al.* (2000a).

Predicted consensus amino acid sequences of the V3-C2 region showed the C and C' group of seroconverter sequences to differ at 12 significant amino acid positions although only two, are in the V3 loop. Especially the presence of lysine (K) at position 304 and valine (V) at position 294 in subtype C' viruses instead of the glutamate (E) and aspargine (N), respectively, in subtype C viruses were of interest because the substitution of Glutamate (E) with Lysine (K) in C' viruses affects the positive charge of the V3 loop and the substitution of aspargine (N) with valine (V) leads to loss of possible glycosylation site. Because the charge of an amino acid affects the property of the protein and since the glycosylation affects the three -dimensional structure of the envelope protein and the accessibility of this protein to neutralizing antibodies, these two substitutions might affect the biological property of the virus (Abebe *et al.*, 2000a). Comparison of amino acid sequences of the SER C and C' with the LSA C database C showed significant differences the SER C being more

closer to the database C than the C'. This observation is in agreement with previous report (Abebe *et al.*., 2000a)

Sequences from subjects, S, G, H, and I were difficult to group as C or C'. Closer look at the amino acid sequences revealed these sequences to have amino acids characteristics of both groups. This could be either because these viruses are undergoing evolution or there is recombination between the two groups of viruses (C and C'). The high nonsynonymous distance observed among sequences of subject S might suggest the evolution of the virus in this individual which could be supported by the high intra-person non synonymous distance (10%) in this subject. In addition two of the four sequences that failed to be grouped as C/C' belong to subjects (G and S) that had a high pre-seroconversion plasma viral load.

The consensus amino acid sequence for all the 19 seroconverters was predicted based on nucleotide sequence of the C2V3 region and showed the conservation of the GPGQT motif typical of subtype C sequences in all the sequences.

The presence of positively charged amino acids at positions 11 and 25 was associated with SI phenotype for subtype B, A, D, and E viruses (Foucher et al., 1992), but for subtype C viruses, the positively charged amino acids associated with the SI phenotype (even though it was only observed for 3 individual) were found at positions 8 and 29 (Abebe *et al.*, 1999). None of the sequences in our study had the positively charged amino acid at the above positions except the sequence from subject S that had a positively charged residue at position 25. This observation indicates the presence of NSI phenotype sequences in almost all the sequences in this study.

The first aspargine at N-linked glycosylation site at position 5-7 was replaced by G in 7 of the 19 sequences while all the three aspargines were substituted by glysin-isoleucine-isolucine (GII) in subject H. The effect of this substitution is not known but the loss of one of the potential glycosylation site can have a possible effect on virus biological property.

The 17% synonymous distance (ds) among Ethiopian seroconverter subtype C sequences (Table 2) is less compared to the 26% found among seroconveter subtype C sequences of the Zimbabwean cohort of male factory workers who participated in the study between 1995-1997. However, the nonsynonymous distances (15%) among all Ethiopian seroconverter sequences is comparable with the Zimbabwean (16%) (Batra *et al.*, 2000). The high genetic diversity (ds) observed among the Zimbabwean seroconverter sequences could either be due to the circulation of subtype C viruses for a long time in Zimbabwe or the multiple introductions of subtype C viruses in the country (Batra *et al.*, 2000).

The less synonymous distance (8.8%) among subtype C' sequences compared to subtype C sequences (16%) showed the low genetic diversity among subtype C' viruses which could be explained as either this group of viruses are introduced late or the C group of viruses have a greater replication rate. However, both of the above explanations are ruled out because a study on the years of introduction of these viruses in Ethiopia estimated the year 1982 and 1983 to be the year of introduction of subtype C' and C viruses, respectively. The greater replication rate of group C viruses cannot be an explanation too because if these group of viruses had a higher replication rate they would have replaced the C' viruses which is not the case at present. Rather it seems the C' viruses are spreading according to the

present study and previous study by Pollakis *et al* (2000) on C/C' recombinants where all the recombinants had a C' envelope. Besides, there was no significant increase viral load among individuals infected with subtype C compared to those infected with C' which might indicate increased virus replication in the former groups. Therefore, a more detailed comparison of the biological properties of the two virus envelopes might provide an answer.

There was a 24% synonymous and 17% nonsynonymous distance between the C and C' sequences. The 17% non synonymous distance between the two groups indicates the high diversity at the protein level, which could have an implication in terms of differences in biological properties between the two groups. In relation with this, we have attempted to see if there is a difference in CD4 count and viral load between the two groups of viruses. This could be an indicator for differences in virus property, but there was no significant difference between the two (data not shown). Besides, sample size of the two groups of viruses (C'=11, C=4) reduced the statistical power to make appropriate comparisons.

Finally, we calculated the intra-person synonymous and non synonymous distances genetic distances (Table 3 and Fig. 3). Only four subjects out of ten (B, F, I, and S) had significant ds while all subjects except two (D and G) had significant dn values. This result indicated the majority of subjects to have intra-person nonsynonymous genetic distance, which in turn indicates the presence of immune pressure. This is actually to be expected because the subjects are recently infected and their immune system has not yet been deteriorated.

According to Holmes and Zanotto (1998), the ratio ds/dn shows the type of selection, in which a ratio ds/dn = 1 is indicative of neutral selection, ds/dn <1 is indicative of positive

selection, and ds/dn > 1 is indicative of negative selection. The ratio can also be interpreted in terms of the immune states of the individual because the ds/dn < 1 in individuals with strong immune response and ds/dn> 1 in individuals with weak immune response. Accordingly, we have attempted to relate the ds/dn ratio with immune status such as CD4 T-cell count and viral load. Subject F who had a ds/dn >1 (1.55) showed a very low CD4 T-cell count which is in agreement with the Holmes and Zantoo (1998) proposal where as subject B (0.23), S (0.55), and I (0.07) who had ds/dn <1, had a higher CD4 count compared to subject F.

5. Conclusion

In conclusion, our data confirmed the presence of a genetic subcluster of subtype C, C' among recently infected individuals. Based on our data the C' virus seems to spread and dominate among currently infected individuals leaving in Addis Ababa and the surrounding and it should be considered in future vaccine designs. Of course, further study with larger sample size is recommended to confirm the spreading of the C' viruses. Besides, there is a need to make full length sequence comparison between the C and C' group. The two groups of viruses have shown significant differences both at the nucleotide and amino acid level and further investigation is needed to explain the biological implications of these differences. The diversity among the C' viruses is shown to be less compared to the C viruses and because of our small sample size, the statistical power to make comparisons between the two groups in terms of differences in viral loads and CD4 count is reduced. And thus, further study on comparisons between the two groups in terms of CD4 count and viral load is recommended.

5. References

Abebe, A. (2000). HIV-1 subtype C in Ethiopia Genotypic and Phenotypic Variation. Ph.D. Thesis. Department of Human Retrovirology, Academic Medical Centre, University of Amsterdam. The Netherlands.

Abebe, A. Dereje, D., Goudsmit, J., Kuiken, C., *Pol*larkis, G., Schuitemaker, H., Fontanet, A. and Rinke de Wit, T. (1999). HIV-1 subtype C Syncytium and non-syncytium inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS. *AIDS*. **13**: 1305-1311.

Abebe, A., Kuiken, C. Goudsmit, J., *et al.* (1997). HIV-1 Subtype C in Addis Ababa, Ethiopia. *AIDS*. **13**: 1071-1075.

Abebe, A., Lukashov, V., *Pol*lakis, G., Kliphuis, A., Fontanet, A., Goudsmit, Y., and Rinke de Wit, T. (2000b). Year of HIV-1 subtype C introduction into Ethiopia: estimated on the basis of genetic analysis of early samples and subsequent virus diversification. Submitted for publication.

Abebe, A., Lukashov, V., Rinke de Wit, T., Fisseha, B., Belete, T., Kliphuis, A., Tesfay, G., Negassa, H., Fontanet, A., Goudsmit, J., *Pol*lakis, G. (2000c). Timing of the introduction into Ethiopia of the subcluster C' of the HIV-1 subtype C. Submitted for publication.

Abebe, A., *Pol*lakis, G., Fontanet, A., Fisseha, B., Tegbaru, B., Kliphuis, A., Tesfaye, G., Negassa, H., Cornelissen, M., Goudsmit, J., and Rinke de Wit, T. (2000a). Identification of a genetic subluster of HIV-1 subtype C (C') widespread in Ethiopia. *AIDS Res Hum Retrovirol.* **16**: 1909-1914.

Aklilu, M., Messele, T., Tsegaye, A., Biru, T., Mariam, H., van Benthem, B., Coutinho, R., Rinke de Wit, T., and Fontanet, A. (2001). Factors associated with HIV-1 infection among sex workers of Addis Ababa, Ethiopia. *AIDS*. **15**: 87-96.

Assefaw, B. (1999). M.Sc. thesis. Department of Biology, Addis Ababa University.

Auwanit, W., Ayuthaya, P., Duangchanda, S., Mukai, T., Kurata, T., and Ikuta, K. (2000). Highly variable sequences in the *env* V3 region of HIV type 1 distributing among Thai carriers from 1995 to 1997. *AIDS Res Hum Retroviruses*. **16**: 283-289.

Ayehunie, S., Johansson, B., Salminen, M., Leinikki, P., Sonnergorg, A., Zewdie, DW., Britton, S., and Strannegard, O. (1991). HIV-1 in Ethiopia: phylogenetic divergence from other HIV-1 strains. *Virus Gen.* **5**: 359-366.

Ayehunie, S., Johansson, B., Sonnerborg, A., Zewdie, D., Britton, S., Strannegard, O. (1993). Sequence analysis of selected regions of the *env* (V3 loop and gp 41/ and *gag* (P17) reading frames of Ethiopian Human Immunodeficiency Virus Type-1 strains. *Arch Virol.* **128**: 229-239.

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

Barre-Sinoussi, F., Chermann, J., Rey R. (1983). Isolation of T-lymphotropic retrovirus from a patient at risk of acquired immune deficiency syndrome. Science. 220: 868-871.

Batra, M., Tien, P., Shafer, R., Contag, C., and Katzenstein, D. (2000). HIV Type 1 *Env*elope Subtype C Sequences from Recent Seroconverters in Zimbabwe. *AIDS Res. Hum. Retro.* **16**: 973-979.

Bjornal, A., Sonnerborg, A., Tscherning, C., Albert, J., and Fenyo, EM. (1999). Phenotypic characteristic of human immunodeficiency virus type 1 subtype C isolates of Ethipian AIDS patients. AIDS Res. Hum. Retroviruses. **15**: 647-653.

Blaak, H. (1999). Virus and host determinants of HIV-1 infection and AIDS pathogenesis. PhD. Thesis. University of Amsterdam, Amsterdam, The Netherlands.

Bobkove, A., Kazennova, E., Selimova, l., Bobkova, M., Khanina, T., Ladnaya, N., Kravchenko, A., Pokrovsky, v., Cheinsong-Popov, R., and 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 Retrovirol.* **14**: 669-676.

Boom, R., Sol, JA., Salimans, M., Jansen, CL., Wertheim-van Dillen, M., and van der Noordaa, J. (1990). Rapid and simple method for purification of Nucleic acids. *Journal of clinical microbiology.* **28**: 495-503.

Carr, JK., Salminen, MO., Albert, J. (1998). Full genome sequences of human immunodeficiency virus type 1 subtype G and A/G intersubtype recombinants. *Virology.* **247**: 22-31.

Carr, JK., Salminen, MO., Koch, C. (1996). Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J Virol.* **70**: 5935-5943.

Coffin JM (1995). HIV-1 population dynamics in *vivo*: implications for genetic variation, pathogenesis, and therapy. *Science.* **267**: 483-489.

Condra, Jh., Schleif, WA., Blahy, OM. (1995). In vivo emergence of HIV-1 variants to multiple protease inhibitors. *Nature.* **374**: 569-571.

Cooper DA, Gold J, MacLean P, Donovan, B., Finlayson, R., Barnes, TG., Michelmore, HM., Brooke, P., and Penny R. (1985). Acute AIDS retrovirus infection: Definition of a clinical illness associated with seroconversion. *Lancet.*1: 537-40.

Cornelissen, M., Kampinga, G., Zorgdrager, F., Goudsmit, J. and The UNAIDS Network for HIV isolation and characterization. 1996. HIV-1 subtypes defined by *env* show high frequency of recombinant *gag* genes. J. Virology. 70: 8209-8212.

D'Souza, P. and Harden, V. (1996). Chemokines and HIV-1 second receptors. Confluence of two fields generates optimism in AIDS research. *Nature Medicine.* **2**: 1293-1300.

Daar, E. (1998). Virology and Immunology of Acute HIV Type 1 Infection. *AIDS Res. Hum. Retro.***14**: S229-S234.

Daar, E., Tarsem Moudgil, M., Meyer, R., and Ho, D. (1991).Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection. *New Engl J Med.***324:**961-964.

Dalgleish, A., Beverley, P., Clapham, P., McDougal, JS., Cort, Sp. (1984). The CD4 antigen (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature.* **312**: 763-767.

Doms, R., and Peiper, S. (1997). Unwelcomed guests with master keys: how HIV uses Chemokine receptors for cellular entry. Virology. 235: 179-190.

Downing, R., Pieniazek, D., Hu, D., Biryahwaho, B., Fridlund, C., Rayfield, M., Sempala, S. and Lal, R. (2000). Genetic characterization and Phylogenetic analysis of HI-1 subtype C from Uganda. *AIDS Res. Hum. Retroviruses.* **16**: 815-819.

Fauci, AS. (1988). The human immunodeficiency virus: infectivity and mechanism of pathogenesis. Science. 239: 617-622.

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

Fontanet, AL., Messele, T., Dejene, A., Enqusellassie, F., Abebe, A., Cutts, FT., Rinke de Wit, T., Sahlu, T., Bindels, P., Yeneneh, H., Coutinho, RA., Nokes, DJ. (1998). Age and gender-specific HIV-1 seroprevalence in the urban community setting of Addis Ababa, Ethiopia. *AIDS*. **12**: 315-322.

Fouchier, RA., Groenink, M., Kootstra, NA. (1992). Phenotype associated sequence variation in the third variable domain of the HIV-1 gp120 molecule. J Virol. 66:3183-3187.

Gallo, R., salahuddin, S., Popovic, M., *et al.* (1984). Frequent detection and isolation of cytopathic retrovirus (HTLVIII) from patients with AIDS and at risk of AIDS. *Science*. **224**: 500-503.

Gao, F., Robertson, D., Carruthers, CD., Li, Y., Bailes, E., Kostrikis, LG., Salminen, OM., Ho, DD., Shaw, M., Sharp., PM.., and Hahn, BH. (1998). An isolate of HIV-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.

Ho, D., Neuman, A., Perelson, A., Chen, W., Leonard, J., Markowitz, M. (1995). Rapid Turnover of Plasma Virions and CD4 Lymphocytes in HIV-1 Infection. *Nature*. **373**:123-126.

Holmes, E. and de A. Zanotto, P. (1998). Genetic drift of Human Immunodeficiency Virus Type 1?. *J.Virol.* **72**:886-887.

Hu, DJ., Baggs, J., Downing, GR., Pieniazek, D., Dorn, J., Fridlund, C., Biryahwaho, B., Sempala, S., Rayfield, MA., Dondero, T., and Lal, R. (2000). Predominance of HIV-1 subtype A and D infections from Uganda. *Emerging Infectious Diseases*. **6**: 609-615.

Hussien, M., Abebe, A., *Pol*larkis, G., Brouwer, M., Petros, B., Fontanet, A., and Rinke de Wit, T. (2000). HIV-1 Subtype C in Commercial Sex Workers in Addis Ababa, Ethiopia. *JAIDS*. **23**: 120-127.

Janssens, W., Buve, A., Nkengasong, J. (1997). The puzzle of HIV-1 subtypes in Africa. *AIDS*. **11**:705-711.

Jetzt, A., Yu, H., Klarmann, G., Ron, Y., Preston, B. and Dougherty, J. (2000). High rate of recombination throughout the Human Immunodeficiency Virus type 1 genome. *J. Virol.* **74**: 1234-1240.

Johansson, B., Sherefa, K., Sonnerborg, A. (1995). Multiple enhancer motifs in HIV-1 strains from Ethioipia. *AIDS Res Hum. Retroviruses*. **11**: 761-764.

Kampinga, GA. (1996). HIV variation and transmission. PhD. Thesis. Academic Medical Center, Amsterdam, The Netherlands.

Kimura, M. (1981). Estimation of evolutionary distances between homologues nucleotide sequences. *Proc. Natl. Acad. Sci. USA.* **78**:454-458.

Korber, B., Kuiken, C., Foley, B., Hahn, B., McCutchan, F., Mellors, J., and Sodroski, J. (1998). Human Retroviruses and AIDS. A compilation and analysis of Nucleic acid and amino acid sequences. Theoretical Biology and Biophysics Group. Los Alamos National Laboratory, Los Alamos, New Mexico 87545 U.S.A.

Kuiken C., Zwart, G., Baan E. (1993). Increasing antigenic diversity and genetic diversity of the V3 variable domain of the human immunodeficiency virus *env*elope protein in the course of the AIDS epidemic. Proc Natl Acad Sci USA. 90: 9061-9065.

Kuiken, C. and Kober, B. (1994). Epidemiological significance of intra- and inter-person variation of HIV-1. *AIDS.* **8** (suppl1): S73-S83.

Kumar,S., Tamura, K.and Nei, M. (1993). Molecular Evolutionary Genetics Analysis (MEGA), version 1.01, Institute of Molecular Evolutionary Genetics, Pennsylvania State University, USA.

Kuritzkes, D. (2000). HIV pathogenesis and viral markers. Continuing Medical Education. http:/www.medscape.com

Lester, F., Ayehunie, S., Zewdie, D. (1988). Acquired Immunodeficiency Syndrome: seven cases in an Addis Ababa hospital. *Ethiop Med J.* **26**: 139-145.

Li, Y., Lou, L., Rasool, N., and Kang, Y. (1993). Glycosylation is necessary for the correct folding of human immunodeficiency virus gp120 in Cd4 binding. *JVirol.* **67**: 584-588.

Lifson AR, Rutherford TW, Jaffe HW. (1988). The natural history of human immunodeficiency virus infection. *J Infect Dis.* **158**:1360-1367.

Lou, CC., Tian, C., Hu, DJ. (1995). HIV subtype c in China [letter]. *Lancet.* **345**: 1051-1052.

Louwagie, J., Janssens, W., Mascola, J., Heyndrickx, L., Hederich, P., van der Groen, G., McCutchan, F., Burke, D. (1995). Genetic diversity of *env*elope glycoprotein from Human Immunodeficiency Virus type 1 isolates of African origin. *J. Virol.* **69**: 263-271.

Lukahov, V., Goudsmit, J. (1997). Evolution of the human immunodeficiency virus type 1 subtype -specific V3 domain is confined to a sequence space with a fixed distance to the subtype consensus. *J Virol.* **71**: 6332-6338.

Lukashov, V., and Goudsmit, J. (1997). Founder Virus Population related to Route of virus Transmission: a Determinant of Intrahost Human immunodeficiency Virus type I Evolution?, *J. Virol.* **71**:2023-2030.

Lukashov, V., Kuiken, C., and Goudsmit, J. (1995). Intrahost Human Immunodeficiency Virus Type 1 Evolution is related to the Immunocompetent Period. *J.virol.* **69**: 6911-6916.

Mansky, L. 1998. Retrovirus mutation rate and their role in genetic variation. J. General Virology. 79:1337-1345.

Marlink R, Kanki P, Thior I, *et al.* Reduced rate of disease development after HIV-2 infection as compared to HIV-1. Science. 1994;265: 1587-1590.

McCutchan, F., Carr, J., Bajani, M., Sanders-Buell, E., Harry, T., Stoeckkli, T., Robbins, K., Gashau, W., Nasidi, A., Janssens, W. and Kalish, M. (1999). Subtype G and multiple forms of A/G recombinants of HIV-1 in Nigeria. Virology. 54: 226-234.

McCutchan, F., Viputtigul, K., De Souza, M.S., Carr, JK., Markowitz, LE., Buapunth, P., McNeil, JG., Robb, ML., Nitayaphan, S., Brix, DL., and Brown, AE. (2000). Diversity of *Env*elope Glycoprotein from Human Immunodeficiency Virus type 1 of recent seroconverters in Thailand. *AIDS Res. Hum. Retoviruses.* **16**: 801-805.

McMichael, AJ., and Phillips, RE. (1997). Esacape of human immunodeficiency virus from immune control. *Ann Rev Immunol.* **15**: 271-296.

McNicholl, J., Smith, K., Qari, H., and Hodge, T. (1997). Host genes and HIV: The role of the chemokine receptor gene CCR5 and its allele (Δ32 CCR5). *Emerging Infectious Diseases.* **3**: 261-271.

Mellors, J., Rinaldo C., Gupta, P., *et al.* (1996). Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science.* **272**:1167-1170.

Menting, A. (2001). Responding to the changing conditions of life: the molecular evolution of HIV. In The evolving virus. Harvard AIDS review. Harvard AIDS Institute, Boston, USA.

Messele, T., Abdulkadir, M., Fontanet, A., Petros, B., Hamann, D., Koot, M., Roost, MTL., Schellenkens, PTA., Miedema, F., Rinke de Wit, T. (1999). Reduced Naïve and increased activated CD4 and CD8 cells in healthy adult Ethiopians compared with their Dutch counterparts. *Clin Exp Immunol.* **115**: 443-450.

Niu, M., Stein, D., and Schnittman, S. (1993). Primary human immunodeficiency virus type-1 infection: review of pathogenesis and early treatment intervention in human and animal retrovirus infections. *J Infect Dis.***168**: 1490-1501.

Nowak MA, Anderson RM, McLeaan AR, Wolfs TFW, Goudsmit J, May RM: Antigenic diversity thresholds and the development of AIDS. Science 1991, 254:963-969

Nowak, MA. (1990). HIV mutation rate [letter]. *Nature.* **347**:522.

PHYLIP: Phylogeny Inference Package, V.3.52. (1996). Seattle: University of Washington. Felsenstein, J. http://evolution.Genetics.Washington.edu/phylip.html

Pollakis, G., Abebe, A., Kliphuis, A., Fisseha, B., Tegbaru, B., Tesfaye, G., Negassa, H., Mengistu, Y., Fontanet, A., Rinke de Wit, T., Cornelissen, M., and Goudsmit, J. (2000). Intra-subtype recombination of Human Immunodeficiency virus type 1 subtype C: evidence for advantage of a gp 120 *env*elope subgroup. Submitted for publication.

Poss, M., Gosink, J., Thomas, E., Kreiss, J., Ndinya-Achola, J., Mandaliya, K., Bwayo, J., and Overbaugh, J. (1997). Phylogenetic evaluation of Kenyan HIV type 1 isolates. *AIDS Res Hum Retrovirol.* **13**: 493-499.

Prelson, AS., Neumann, AU., Markowitz, M. (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell life span, and viral generation time. *Science.* **271**: 1582-1586.

Preston, B., Poieesz, B., Loeb, L. (1988). Fidelity of HIV-1 reverse transcriptase. Science. 242: 1168-1171.

Price, D., Goulder PJR, Klenerman, P.,*et al.* (1997). Positive selection of HIV-1 cytotoxic T-lymphocyte escape variants during primary infection. Proc Natl Acad Sci USA. 94: 1890-1895.

Renjifo, B., Chaplin, B., Mwakagile, D., Shah, P., vanberg, F., Msamanga, G., Hunter, D., Fawzi, W. and Essex, M. (1998). Epidemic expansion of HIV type 1 subtype C and recombinant genotypes in Tanzania. *AIDS Res Hum Retroviruses.* **14**: 635-638.

Rinke de Wit, T., Tsegaye, A., Kliphuis, A., Hagos, M., *Pol*lakis, G., Hailu, B., Aklilu, M., and Fontanet, A. (2000). HIV-1 seroconversion in a prospective cohort of Ethiopian factory workers: a subject with a window period of at least 204 days. In Press.

Robertson, D., Sharp, P., McCutchan, F. and Hahn, B. (1995). Recombinatiion in HIV-1. Nature.374: 124-126.

Robertson, DL., Anderson, JP., Bradac, JA., Carr, JK., Foley, B., Gao, F., Hahn, BH., Kalish, ML., Kuiken, C., Learn, GH., McCutchan, F., Peeters, M., Salminen, M., Sharp, PM., Wolinsky, S., and Korber, B. (2000). HIV-1 nomenclature proposal. *Science*. **288**: 55-57.

Robey, WG., Safai, B., Oroszlan, S. (1985). Characterization of *env*elope and core structureal gene products of HTLV-III with sera from AIDS patients. *Science*. **228**:593-595.

Salminen, M., Johanssen, B., Sonnerborg, A., *et al.* (1996). Full-length sequence of an Ethiopian HIV-1 isolate of Genetic subtype C. *AIDS Res. Hum. Retroviruses*. **12**: 1329-1339.

Schuitemaker, H., Koot, M., Kootstra, NA. (1992). Bilological phenotype of HIV-1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus populations. *J Virol*. **66**: 1354-1360.

Sherefa, K., Johansson, B., Sonnerborg, A. (1998). Full length sequence of human immunodeficiency virus type 1 A, recombined with subtype C in the *env* domain. *AIDS Res. Hum. Retroviruses*. **14**: 289-292.

Simon, F., Mauclere, P., Roques, P. (1998). Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nature medicine*. **4**: 1032-1037.

Takehisa, J., Zekeng, L., Ido, E. (1998). Various types of HIV mixed infections in Cameroon. *Virology*. **245**: 1-10.

Taylor, Jm., Schwartz, K., Detels, R. (1986). The time from infection with human immunodeficiency virus (HIV) to onset of AIDS. *J Infect Dis*. **154**: 694-697.

Thomas, DJ., Wall, JS., Hainfeld, JF. (1991). Gp160, the *env*elope glycoprotein of human immunodeficiency virus type 1, is a dimmer of 125kd subunits stabilized through interactions between their gp 41 domains. *J Virol*. **65**: 3797-3803.

Tien, P., Chiu, T., Latif, A., Ray, S., Batra, M., Contag, C., Zejena, L., and Katzenstein, D. (1999). Primary subtype C HIV-1 infection in Harare, Zimbabwe. *J Acquired Immunodeficiency Syndromes and Human Retrovirology*. **20**: 147-153.

Tscherning, C., Alaeus, A., and Fredriksson, R. (1998). Differences in chemokine coreceptor usage between genetic subtypes of HIV-1. *Virology*. **241**:181-188.

Tsega, E., Mengesha, B., Nordenfelt, E., Hansson, B-G., Lindberg, J. (1988). Serological survey of Human Immunodeficiency Virus infection in Ethiopia. *Ethiop Med J*. **26**: 179-184.

van Hermelen, J., van der Ryst, E., Loubser, A., *et al.* (1999). A predominantly HIV-1 subtype C-restricted epidemic in South African urban population. *AIDS Res Hum Retroviruses.* **15**: 395-398.

Wang, Z., Lyles, C., Beyrer, C., Celentano, D., Vlahov, D., Natpratan, C., Nelson, K., and Yu, X. (1998). Diversification of subtype E human immunodeficiency virus type 1 *env* in heterosexual seroconverters from Northern Thailand. *JID.* **178**: 1507-1511.

White, DO and Fenner, FJ. (1994). Medical virology. Fourth edition. Academic Press. San Diego.

WHO/UNAIDS Report. (2000).

Willey, RL., Smith, DH., Lasky, LA., Theodore, TS., Earl, PL., Moss, B., and Martin, MA. (1988). Identification of conserved and divergent domains with in the *env*elope gene of the acquired immunodeficiency syndrome retrovirus. Proc. Natl. Acad. Sci. USA. 1986. 83: 5038-5042.

Wolfs, TFW., Zwart, G., Bakker, M., *et al.* (1991). Naturally occurring mutations within the HIV-1 V3 genomic RNA lead to antigenic variation dependent on a single amino acid substitution. Virology. 185: 195-205.

Wong-Staal, F. (1990). Human Immunonodeficiency viruses and their replication. Virology. 2nd ed. Raven Press. Ltd. New York.

Zhu, T., Mo, H., Wang, N., Nam, S.D., Cao, Y., Koup, A.R., and Ho, D. (1993). Genotypic and phenotypic characterization of Hiv-1 in patients with primary infection. *Science.* **261**:1179-1181.

Zwart, G., Back, N., Ramautarsing, C., Valk, M., van der Hoek, L., and Goudsmit, J. (1994). Frequent and early HIV-1_{MN} neutralizing capacity in sera from Dutch HIV-1 seroconverters is related to antibody reactivity to peptide from the GP120-V3 domain. *AIDS Res Hum Retrovirl.* **10**: 243-249.