COMPARISON OF THE GENETIC VARIABILITY OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1) IN BLOOD PLASMA AND CERVICOVAGINAL LAVAGE USING C2V3 REGION OF THE *ENV* GENE

By

MELAKU ADAL ESHETIE

A Thesis Submitted to the School of Graduate Studies,

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Biology (APPLIED GENETICS)

Addis Ababa University

School of Graduate Studies

Department of Biology

January 2003

Acknowledgements

I would like to express my deepest gratitude to my advisors Dr. Kifle Dagne and Dr. Wendelien Dorigo-Zetsma for their valuable comments, suggestions and encouragement throughout the course of this study. I am grateful to Dr. Dawit Wolday for his invaluable guidance and help in problem identification, proposal design, and for providing me with relevant literature and samples for the project. My deepest gratitude goes to mss. Worknesh Ayele for her interest, concern, support, valuable suggestions and consistent guidance from the start to the end of the study. I am also highly grateful to Dr. Georgios Pollakis and Dr. Michel de Baar for their useful guidance and suggestions in the course of the study and also the data analysis. Special thanks goes to ENARP PhD students, laboratory technicians, especially Ato Tesfaye Tilahun and the administrative staff for introducing me to all laboratory activities and set ups of ENARP and for their collaboration and support during my laboratory activities.

I would like to thank the Biology Department and School of Graduate Studies, Addis Ababa University for the training and the Ethio-Netherlands AIDS Research Project (ENARP), located at the Ethiopian Health and Nutrition Research Institute (EHNRI), for financial support of the research. My acknowledgement also goes to members of Biology Department and friends who in one way or another contributed for the success of this work. Finally, I would like to extend my deepest gratitude to Dr. Adamu Demeke for his unreserved friendly help and advice and to my family, especially to my mother Tiruye Teshale, for their consistent encouragement and support throughout the study period.

Table of Content

List of Tables

List of Figures

Abstract

To investigate the characteristics of HIV-1 in the female genital tract and the molecular epidemiology of subtypes, cell-free viral populations in the blood plasma and vaginal mucosa of women infected with HIV-1 using DNA sequencing of C2V3 region of the *env* gene were examined. Study subjects were those who had visited a sexually transmitted disease clinic for genital tract inflammation (n = 32 of which 29 had sexually transmitted infections) in Addis Ababa, Ethiopia. From the study subjects, 28 of the blood plasma and 22 of the cervicovaginal lavage viruses were successfully amplified and sequenced. 31 of the subjects carried subtype C and only one subject carried subtype A. Among the total individuals carrying subtype C, 12 and 19 were with subclusters C and C', respectively.

The sequences in cervicovaginal lavage differed significantly from those in plasma. The number of nonsynonymous substitutions significantly exceeded synonymous substitutions in both plasma and cervicovaginal lavage in almost all the study subjects. Sequences from both compartments within an individual were more closely related to each other than to those obtained from viruses of other individuals when evaluated through phylogenetic analysis. However, there were significant genotypic differences between viruses in the two compartments of an individual. Differences in nucleotide sequences, in net charge of amino acids at physiological pH and predicted amino acid sequences were also observed. These differences could be due to selective immune pressures operating in the two compartments, which could lead to differential evolution of the viruses in vaginal mucosa and plasma that can lead to differential evolution of the viruses.

These findings indicate that cellular replication of HIV-1 occurs in vaginal mucosa that results in a virus population with important differences from that in blood. This could result in different resistance patterns in the two compartments in the era of chemotherapies. Thus, resistance patterns in the viruses derived from the genital mucosa should be taken into account in order to prevent transmission of resistant viruses through exposure to the genital tract. For vaccine candidates, viruses derived from the genital compartment should be considered. So far, vaccine candidates are based on virus isolates from blood only. This indicates a need of considering all heterogeneity in all compartments to produce effective vaccines and therapies.

1. Introduction

1.1. Epidemiology of HIV/AIDS

According to WHO/UNAIDS report (2002), twenty years after the first clinical evidence of acquired immunodeficiency syndrome (AIDS) was reported, it has become the most devastating disease mankind has ever faced. HIV/AIDS is now the leading cause of death in Sub-Saharan Africa. Worldwide, it is the fourth biggest killer. The total number of people affected by HIV/AIDS from the start of the epidemic until the end of 2002 was estimated to be more than 60 million. Among 42 million people now living with HIV/AIDS, more than 70% are in the developing world. The highly affected region by the pandemic is Sub-Saharan Africa, where approximately 3.5 million new infections occurred in the year 2002, while the epidemic claimed the lives of an estimated 2.4 million Africans in the past year. Ten million young people (age 15-24) and almost 3 million children under 15 are living with HIV, bringing the total number of people living with HIV/AIDS in this region to 29.4 million.

In Ethiopia, there was no reported case of positive serum during 1982 and 1983. The first two positive sera were detected in 1984 collection (Tsega *et al*., 1988). The first AIDS cases were diagnosed in Yekatit 12 Hospital of the capital, Addis Ababa, in 1986 (Lester *et al*., 1988). The Ethiopian Disease Prevention and Control Department of the Ministry of Health (2002) reported that over 219,400 people may have already developed AIDS since the beginning of the epidemic although only a small portion of these (107,575) have been recorded in official statistics. In the year 2002, it was estimated that there were about 2.2 million people infected with HIV, out of which 200,000 were children under the age of five years. The estimated adult prevalence of 7.1 percent in 1997 increased to 7.3 percent in the year 2000, and decreased to 6.6 in 2002.

1.2. HIV and Its Life Cycle

Since the initial clinical reports in the summer of 1981 among patients with *Pneumocystis carinii* pneumonia in previously healthy homosexual men in the United States of America, a large number of people all over the world have been infected by the virus, and a considerable number of the patients have died as a result of the disease (Gottlieb *et al*., 1981). Two years after the first description of AIDS, its causative agent was identified to be a virus and was variously named as human T cell lymphotrophic virus type III (HTLV-III) (Gallo *et al*., 1983), lymphadenopathy associated virus (LAV) (Barre-Sinoussi *et al*., 1984), or AIDS associated retrovirus (ARV) (Levy *et al*., 1984). Thereafter, the International Committee on the Taxonomy of Viruses decided the virus to be called HIV (Coffin *et al*., 1986a). Two years later, HIV-2 that caused the same disease but was significantly different from HIV-1 serologically and molecularly was discovered in West Africa (Clavel *et al*., 1986).

Since the initial identification of HIV-1 in 1983 (Barre-Sinoussi *et al*., 1983; Gallo *et al*., 1984), much has been learned about the virus, its transmission and its pathogenic mechanisms. HIV is transmitted through sexual contact, contaminated injection needles, blood and blood product transfusion, and mother-to-child transmission. Although an acute mononucleosis-like illness may occur at the time of infection (acute retroviral infection), the infection is characteristically followed by a prolonged period of clinical latency (Moss and Bacchetti, 1989), with the median period of latency estimated to be approximately 10 years. It is now clear that during this clinically quiescent period, viral replication continues (Pantaleo *et al*., 1993) and is associated with the gradual erosion of immune competence. Subsequently, as a patient becomes progressively immunosuppressed, opportunistic infections and malignancies develop that are responsible for

most of the morbidity and, mortality of HIV/AIDS.

HIV is a member of the Lentivirinae subfamily of the Retroviridae. This family is known for being RNA viruses and use reverse transcriptase (RT) enzyme to become DNA and integrate into the human genome by the help of integrase (IN) enzyme. The lentiviruses are all A-rich and C-poor with nucleic acid compositions of 36%A, 18%C, 24%G and 22%T. The A richness and C poorness are roughly evenly distributed across the lentivirus genome with few exceptions (Kuiken *et al*., 2000).

HIV infects cells with CD4 cell-surface receptor molecules, using them to gain entry. In macrophages and in some other cells lacking CD4 receptors, such as fibroblasts, an Fc receptor site or complement receptor site may be used instead for entry of HIV. Cells of the mononuclear phagocyte system, principally blood monocytes and tissue macrophages, T lymphocytes, B lymphocytes, natural killer (NK) lymphocytes, dendritic cells (Langerhans cells of epithelia, and follicular dendritic cells in the lymph nodes), hematopoietic stem cells, endothelial cells, microglial cells in the brain, and gastrointestinal epithelial cells are the primary targets of the HIV infection (Vartanian and Wain-Hobson, 1994). After an asymptomatic period of variable length, HIV infections, accompanied by a dominant T helper 1 (TH1) response tend to proceed longer. The switch from a TH1 to TH2 response has been suggested as a factor in an irreversible destruction of the immune system, which finally leads to AIDS. The destruction of the immune system is caused mainly by the gradual depletion and loss of function of CD4 bearing helper/inducer T-lymphocytes in the peripheral blood from the normal count of about 1000 cells/mm^3 to less than 100 cells/mm^3. This is because $CD4^+$ T cell plays a central role in the immune system, which includes induction of the function of B-cells, NK cells, cytolytic T-

lymphocytes (CTL), macrophages through secretion of different cytokines. The precise mechanisms of CD4 T-cell depletion (or dysfunction) in HIV infection are not still well understood. The depletion may occur as a result of direct pathogenicity, syncytium formation, T-cells susceptibility to apoptosis by defective antigen presenting cells, suppression of the protective TH1 cells, induction of autoimmunity by cross reaction with HIV proteins, or coating of surface CD4 by gp120 immune complexes (Fauci, 1993; Murphy *et al*., 1995; Roitt, 1997; Popper *et al*., 1999).

Although the precise ordering of many of the steps has not been definitely determined and multiple steps are likely to occur in simultaneous fashion, the retroviral life cycle of HIV is summarized by Freed (1998) into the following series of events. The infection process begins when the surface (SU) *env* gp120 binds CD4 and interacts with coreceptors. A membrane fusion reaction induced by transmembrane (TM) *env* gp41 occurs between the lipid bilayer of the virion and the host cell plasma membrane, releasing the viral core into the cytoplasm. A series of ill-defined steps collectively referred to as uncoating takes place. During this process, capsid (CA, p24) protein is lost while at least some matrix (MA, p17) protein, as well as nucleocapsid (NC, p7/9), the *pol*-encoded enzymes integrase (IN, p32) and reverse transcriptase (RT, p66) and the viral protein R (*vpr*) are thought to be retained as part of a high molecular weight complex. During uncoating, reverse transcription of the viral RNA to generate a double-stranded DNA copy is largely completed in the presence of lysyl tRNA primer. The high molecular weight complex referred to as a preintegration complex is transported across the nuclear membrane with expense of ATP. In the nucleus, integration of only linear viral DNA into the host cell chromosome is catalyzed by IN. The integrated viral DNA, known as the provirus, serves as the

template for the synthesis of the viral RNAs, which are transported to the cytoplasm. The *env* glycoprotein is synthesized in the endoplasmic reticulum and is transported to the plasma membrane via the secretory pathway. The *gag* and the *gag-pol* polyproteins precursors are synthesized and transported by unknown mechanism to the plasma membrane. During or after transport, the *gag* precursor, recruits two copies of the single stranded viral RNA genomes, interacts with the *gag-pol* precursor, and assembles into structures visible by electron microscopy as dense patches lining the inner face of the plasma membrane. The assembled core proteins complex of *gag* gene (MA, CA, NC, and p6) induces membrane curvature, leading to the formation of a bud. During budding, the viral *env* glycoprotein is incorporated into the nascent particles. Budding is completed as the particle pinches off from the plasma membrane. During or immediately after budding, the viral protease (PR, p11) cleaves the *gag* and the *pol* proteins. PR cleavage leads to core condensation and the generation of a mature, infectious virion, which is now capable of initiating a new round of infection.

1.3. Genomic Organization and Gene Products of HIV

The genetic material of HIV is about 100,000 times smaller than the genetic information of human cell: a mere 9,749 nucleotides and consists of nine genes (Haseltine and Wong-Staal, 1988). Among these, three are typical proviral genes. All have the basic gene orders 5'-*gag-pol-env*-3' and are flanked by the characteristic long terminal repeat (LTR) sequences that contain transcription initiation (5' LTR) and termination (3' LTR). According to Patarca *et al*. (1987) and Coffin (1990), HIV LTR is a very active promoter *in vitro.* In addition, the genome of HIV encodes numerous regulatory proteins common to all retroviruses, and further complicated by containing three to five additional accessory regulatory genes than other retroviruses (*tat, rev,*

vpu, vif, nef and *vpr*) (Haseltine and Wong-Staal, 1988; Young, 1994; McClure and Dalgleish, 1998; Peeters *et al*., 2000).

The single-stranded viral RNA is converted to double-stranded DNA by the action of two *pol* enzymes packaged within the capsid, RT (RT, p51) and the RNase H (RNH, p15) after its entry to the cytoplasm. The RNH has the ability to degrade RNA in the context of a DNA-RNA hybrid, which takes place within a viral nucleoprotein complex. The minus single stranded DNA, which remains single after RNA degradation, is used as a template for the plus strand and makes the double stranded proviral DNA and form complexes with nucleoproteins. Retroviral DNA can integrate into the host genome at a large number of essentially random sites by the help of IN enzyme, although some target sites are preferred (Haseltine, 1988). Comparison of the cellular sequences flanking a modest number of independent integration events (12-20 in most cases) has revealed no common sequence, which might serve as a cellular target. The choice of host integration site appears random with regard to the DNA sequences but strongly depends on accessibility of the local chromatin. As increased accessibility of chromatin correlates with increased transcriptional activity, retroviruses may thus preferentially integrate in the vicinity of actively expressed genes (Coffin, 1990; Young, 1994).

The *env* protein is very heavily modified into glycoproteins by addition of complex carbohydrates via the action of cellular enzymes for a reason that such carbohydrates are not antigenic. The amino-terminal portion of the protein is located entirely exterior to the cell membrane. Gp120 is anchored noncovalently with its carboxyl terminus to the viral membrane. The conserved gp120 surfaces are involved in binding to its three minimally polymorphic ligands, gp41, CD4, and chemokine receptors, each exhibiting particular problems with respect to

the elicitation of or sensitivity to the neutralizing antibodies (Wyatt and Sodroski, 1998). Mulligan *et al.* (1990) indicated that *env* glycoprotein of HIV is central to its ability to specifically interact with and infect $CD4^+$ cells. These interactions include receptor binding, viral entry and syncytium formation, and perhaps other cytopathic effects. Interaction of gp120-CD4 allows for a change in the *env* protein structure to take place (The Global AIDS Policy Coalition, 1992). Once the gp120 interacts with the coreceptors, additional conformational changes expose fusion peptides at the N-termini of the gp41 moieties, which then mediate fusion of the viral envelope and cell membrane so that viral nucleoprotein complexes enter the target cell cytoplasm (Haseltine, 1988).

1.4. Genetic Diversity of HIV: Causes and Consequences

HIV-1 genome exhibits extensive nucleotide sequence variation in blood cells of infected individuals (Pang *et al*., 1991). One of the most important mechanisms for producing new types of RNA viruses is mutation, mostly through the accumulation of point mutations introduced by the error prone HIV-1 reverse transcriptase during replication (Haseltine, 1988; Strauss *et al*., 1990; Bonhoeffer *et al*., 1995; Mansky, 1998). This means that most DNA dependent RNA polymerases and RNA dependent DNA polymerase appear to lack 3'-5' exonuclease-proofreading activities, and their inherent error frequently is on the order of 1 in 10^4. 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 (Haseltine, 1988; Perelson *et al*., 1996).

Subsequent nucleotide sequence comparisons confirmed that sequence differences among different viral genomes are not evenly distributed throughout and is not the property of the entire

genome. In general, the extra-cellular envelope region and the *nef* gene are more variable than the LTR, *gag, pol, vif, vpr, tat,* or *rev* genes. Not only the degree of divergence is different among these genes, but also the type of nucleotide and amino acid changes are very different as well. It is known that in *gag* and *pol*, most nucleotide sequence changes are due to point mutations, whereas in the *env* there are clustered changes involving inframe deletions, insertions, and duplications. Furthermore, the majority of changes in the *gag* and *pol* are in the third base codon position, resulting in silent mutations. In contrast, more than half of the single nucleotide changes in *env* occur in the first or second positions. These findings suggest two possible mechanisms of selection of *env* mutations. First, there may be regions in the *env* gene that have little sequence-function constraints and therefore, would tolerate such changes; second, changes in parts of the *env* may confer a selective advantage for mutants. The rapid evolution of HIV isolates demonstrates the extreme plasticity of the HIV genome with regard to its genetic and biologic properties, including cell tropism (such as CD4 recognition, membrane anchorage and membrane to membrane fusion, virulence, and antigenicity) (Haseltine, 1988; Wang-Staal, 1990; McClure and Dalgleish, 1998).

The diversity of HIV-1 is mainly caused by mutations that affect genes part of genes encoding proteins. Amino acid sequences of gp120 from a number of different HIV-1 isolates have revealed five discontinuous regions (V1-V5) that contain highly variable amino acid residues interspersed with five relatively conserved regions designated C1-C5. The most antigenically dominant domain on the *env* of HIV-1 is the V3 region. The evolution of V3 sequences is apparently host dependent, rapid, and independent of the level of antigen expression (De Wolfs *et al.*, 1990). It is known that the number of charges and glycosylation in the V3-loop can affect

cellular and neutralization abilities of antibodies. The conformation of V3 domain and the antibody response directed against this epitope may vary from one variant to another, which results in the absence of high titer antibodies to the important functional domains such as the CD4 binding region and the fusion region. Because this region of the virus is highly variable, this variation leads to a further delay in the development of a universal prophylactic vaccine and therapeutics, and it ensures an endless source of antibody escape and drug-resistant variants which leads to disease progression in the presence of neutralizing antibodies (Young, 1994; Lenz *et al*., 2001).

Several genetic subtypes and circulating recombinant forms (CRFs) of HIV-1 have been identified. 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 and 13% in the total genome. Since its discovery, the HIV genome has been extensively studied, partly to determine the range of strains contributing to the global epidemic and partly in an attempt to understand the evolution of HIV-1 over time (Peeters *et al*., 2000).

Numerous strains of HIV-1 isolated from diverse geographical origins have revealed three distinct groups of viruses. These have been termed groups M (for main), N (non-M-non-O), and O (outlier). The vast majority of the strains found worldwide and are responsible for the pandemic belong to just one of the three lineages, group M. Group M includes the viruses that dominate the global epidemic which is further divided into subtypes, subsubtypes and circulating recombinant forms (CRFs), whereas group O contains a pool of highly divergent but genetically related viruses with its epicenter restricted to West Central Africa. Group N, the least spreading

of the viruses, is represented by only a handful of viruses identified in Cameroon (Peeters *et al*., 1997).

A virus isolated as a specific subtype should fulfill the following criteria: representative strains must be identified in at least three individuals with no direct epidemiological linkage, at least three near full-length genomic sequences should be preferably available, and the subtype should resemble each other but no other existing subtype throughout the genome (which may indicate recombination), a subtype should have consistent clustering and all parts of the genome should be approximately equidistant from one another (Robertson *et al*., 2000). In this light, there are nine subtypes of the HIV-1 group M: A, B, C, D, F, G, H, J and K. A putative new subtype, possibly to be named subtype L is described, and is found most commonly in Senegal (McCutchan *et al*., 200l). All known representatives of what was initially described as subtype E appear in fact to be recombinants of the subtypes A and E, and are now designated (CRF01-AE). CRF04-cpx viruses correspond to the previously described *env* subtype I virus, which are complex recombinant viruses involving at least four subtypes. Recombinant forms of the virus will continue to appear as long as the different subtypes of HIV-1 continue to circulate between continents and recombination continues to occur (Peeters and Delaporte, 1999; Robertson *et al*., 2000). The highest degree of genetic diversity in HIV-1 is observed in Africa where all subtypes and groups can be observed (McClure and Dalgleish, 1998).

McClure and Dalgleish (1998) indicated that recombination between isolates of different clades makes its own contribution to the range of the strains found in global pandemic of AIDS. Recombination in the *gag* gene is highly frequent among the major *env* subtypes and that selection of the recombinants is apparently based on particularly beneficial combinations of *gag*

and *env* products (Cornelissen *et al*., 1996). Inter-subtype recombination is an additional cause of HIV diversification and is relatively frequent (up to 10%) in regions where multiple subtypes co-circulate, such as Sub-Saharan Africa (Robertson *et al*., 1995; Bobkove *et al*., 1998; Carr *et al*., 1998; Sherefa *et al*., 1998). Phylogenetic analysis of HIV-1 isolates from UNAIDS plasma samples collected from four vaccine evaluation sites: Brazil, Rwanda, Thailand, and Uganda, previously subtyped by *env* sequence, showed a high frequency of *gag* recombinants (Cornelissen *et al*., 1996). Serum samples collected from immigrants in the Netherlands from 12 countries throughout Africa and phylogenetic analysis based on the *gag* sequences of 22 isolates showed the recovery of AG recombinants from African countries a thousand miles apart, indicating the active spread of new recombinants (Cornelissen *et al*., 2000).

Sequence data of the *env* gene on sera and plasma samples collected in Addis Ababa demonstrated that the majority of the circulating viruses belong to the subtype C. But sporadic infections with other subtypes, A and D, have also been reported (Abebe *et al*., 1997; Hussien *et al*., 2000). In neighboring countries, Kenya and Uganda, subtypes A and D are dominating. In Djibouti subtypes A and C are reported (Louwagie *et al*., 1995). Thus there is a possibility of influx of other subtypes to Addis Ababa from the neighboring countries. The first HIV-1 subtype C sequence in Ethiopia was reported in 1991 (Ayehunie *et al*., 1991) followed by partial *gag* and *env* sequences by Ayehunie *et al*. in 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, documented the first evidence of the subtype A/C recombinant from a 1991 Addis Ababa sample (Sherefa *et al*., 1998). In addition, from randomly selected HIV-1 subtype C infected sera collected from several Ethiopian towns, phylogenetic

analysis of the whole p17 and partial of p24 of the *gag* gene, and their respective C2V3 *env* regions showed intra-subtype recombinants between the two groups of viruses (C and C'), which co-circulate in the country (Pollakis *et al*., unpublished data). This shows that recombination can occur not only among the different subtypes of the HIV-1 but also within one-subtype strains. All-recombinant viruses carried the *env* of the C virus suggesting that this *env* may be more efficient for virus transmission than the envelope of subgroup C. The phenotypic expression of this genomic region involves how the virus attaches itself to the cell receptor and how it invades neutralizing antibodies produced by the immune system.

The HIV-1 subtypes are unevenly distributed in the world and the geographic distribution of subtypes is subject to constant change. Subtype B predominates in Europe and America, whereas subtype C is mainly found in South and East Africa, such as South Africa, Botswana, Burundi, Tanzania, and Ethiopia, and also in India and China (Lou *et al*., 1995; Engelstad, 1996; Renjifo *et al.*, 1998; Wang *et al*., 1998; Van Hermelen *et al*., 1999; Koch *et al*., 2001). HIV-1 subtype C is the most spreading, and 50% of the new 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 behavioral or biological factors that facilitate transmission and the selective characteristics of the subtype C virus envelope (Janssen *et al*., 1997; Van Hermelen *et al*., 1999). For example, subtype C has a genome organization similar to that of other subtypes of the group M that are fully sequenced HIV-1 isolates A, B, and D. However, it has an additional third binding site for the transcription factor NF-κB makes these strains distinctive among HIV-1 isolates, and this invites speculations about the possible role of this structural anomalies in the epidemiology of HIV-1 subtype C (Salminen *et al*., 1996).

Selection forces, besides a combination of mutations and recombinations, shape the diversity of viral quasispecies (Bonhoeffer *et al*., 1995). It is found that a greater systemic HIV-1 heterogeneity exists in HIV/TB patients than in HIV patients. Phylogenetic and sequence analysis of the C2V3 region of individual *env* pair clones of HIV and HIV/TB patients indicated more diverse clustering of the quasispecies and greater mutation frequencies in the TB patients. In addition, HIV/TB patients tended to have higher nonsynonymous (dn)/synonymous (ds) values than HIV patients, suggesting a stronger immune selection pressure in the former (Collins *et al*., 2001). This leads to a conclusion that an increase in HIV-1 heterogeneity with the advent of TB could have a profound effect on disease progression (Collins *et al*., 2001).

The accelerated rate of disease progression has been related to chronic immune stimulation due to pathogenic microbes, including parasites (Gilks, 1993; Bentwich *et al*., 1995). Evidence for an involvement of cytokine in the activation of HIV-1 has been documented. For example, tumor necrosis factor alpha (TNF-α) has been found to induce the production of NF-κB which binds to the HIV enhancer elements present in the LTR which initiates transcription of viral mRNA and augment viral replication (Gruber *et al*., 1992). STDs increase HIV shedding, thereby increasing the likelihood that genital secretions contain higher viral load of the HIV required for transmission (Hitchcock and Fransen, 1999). The local immune activation increased HIV-1 load in genital secretions, potentially increasing the risk of transmission. Furthermore, cytokine released in genital ulcers in people uninfected with HIV may render cells within the ulcers more susceptible to HIV-1 infection and promote systemic spread of HIV infection beyond the genital tract (Lawn *et al*., 2000). For example, sexually transmitted infections (STI) can be a risk for both HIV infection and/or transmission by altering the integrity of the mucosa, recruiting

susceptible activated lymphocytes and monocytes/macrophages to the genital mucosa, or increasing viral shedding of productively infected cells either in the genital mucosa or in the genital secretions (Cohen, 1998; Plummer, 1998).

1.5. Effect of the Genotype on the Biological Phenotype of HIV

The effect of genotype on viral phenotype is reported by different people who did researches in this field (Fouchier *et al*., 1992; Tersmette *et al*., 1989). In this context, isolates of the HIV-1 have been broadly divided into two groups; those which replicate in T-cell lines, grow very rapidly in cultures and induce the formation of syncytia in target cells, and those which are able to replicate in macrophages, grow relatively slowly in cultures and do not induce syncytia. These two groups are generally referred to as syncytium inducing (SI) and non-syncytium inducing (NSI), respectively (Bozzette *et al*., 1993). There is a strong correlation between the viral phenotype and progression to disease (Cheng-Mayer *et al*., 1988; Tersmette *et al*., 1989). NSI HIV-1 strains use primarily CCR5, while SI strains use CXCR4 receptors. SI strains use other chemokine receptors such as CCR3 and CCR2b in some instances (Abebe *et al*., 1999). CXCR4 usage was perfectly correlated to the biological phenotype for all subtypes. The role of SI viruses in transmission of HIV is still unclear. If highly pathogenic SI viruses are readily transmitted between individuals, one might expect a decreasing incubation period for AIDS as the epidemic progresses. However, this has not been observed (Lemp *et al*., 1990; Taylor *et al*., 1991).

SI variants appear in the later stage of the infection and for subtype B, this biological phenotype was proven to be a prognostic marker for disease progression with individuals harboring SI viruses progressing faster independent of the CD4 cell count (Chowdhury *et al*., 1995). HIV-1

strains differ in how they utilize chemokine receptors as co-factors for entry, and the ability of the chemokine receptor to facilitate entry depends on the cell in which it is expressed (Collman and Yi, 1999). Preferential transmissibility of certain NSI isolates compared with more pathogenic SI isolates may be one explanation for this finding. In this respect, Ethiopian patients with HIV-1 C-subtype harbor a remarkably low frequency of SI phenotype viruses (Abebe *et al*., 1999). Even if the reasons for this are not clear, it could be speculated that the expression of CCR5 chemokine receptor is higher on cells of subjects living in countries where HIV-1 subtype C epidemic is present, thus providing less selective advantage to the virus for mutation to CXCR4 tropic variants. CCR5 reportedly is preferentially expressed on activated and memory T-cells (Bleul *et al*., 1997). However, it is also possible that the primary immune response after HIV infection of an individual might be more efficient in eliminating SI viruses than in eliminating NSI viruses (Chesebro *et al*., 1992). There may also be subtype dependent differences in frequency of usage of certain co-receptors. This opens up the possibility that genetic subtypes may differ in important biological properties such as virulence, tissue tropism and transmissibility (Tscherning *et al*., 1998).

The V3 hypervariable region of HIV envelope protein influences cell tropism. Tropism is determined largely at entry and is encoded mainly by the HIV-1 *env* gene (Berger *et al*., 1997). The homogeneity of the macrophage tropic patient isolates appeared to be the result of selection based on the biological advantage *in vivo* in contrast to the heterogeneous V3 sequences of T cell tropic clones (Chesebro *et al*., 1992). *In vitro* mutagenesis of the V3 loop is able to convert the viral phenotype from NSI, low replicating into SI, high replicating. In short, the V3 region, and even specific regions herein are shown to be involved in the syncytium forming and replicating

capacity of HIV-1 (de Jong *et al*., 1992b).

1.6. Temporal and Spatial Variation of HIV in a Host

Virus in patients in early stage of infection are relatively homogenous and diverge with time, more consistently at its non-synonymous sites. Just prior to or coincident with the rapid decline in $CD4^+$ T cell number, sequences were found with basic amino acid substitutions clustered within downstream of the gp120 V3 domain (Shankarappa *et al*., 1998). Other data from small number of infected patients also show temporal changes in the number of genetically distinct strains of the virus throughout the incubation period, with slow but steady rise in the diversity during progression to disease (Nowak *et al*., 1991).

According to Ellerbrock *et al*. (2001), multiply spliced HIV-1 mRNA, which is found only in cells replicating virus, was detected in cervicovaginal lavage samples tested. There is a strong correlation between the amounts of multiply spliced HIV-1 RNA and viral RNA load in the lavage samples. In addition, significant genotypic differences were found in cell-free virus from matched blood plasma and vaginal secretions. These findings indicate that cellular replication of HIV-1 occurs in vagina mucosal cells and can result in a virus population with important differences from that in blood. All these lines of evidences suggest that HIV virus is actively produced in the vagina and not just a result of leakage into the vagina through the bloodstream.

In phylogenetic trees constructed from V3 of gp120, kidney derived sequences form tissue-specific subclusters within the radiation of blood mononuclear cells derived viral sequences from the patients. These data indicate the localized replication of HIV-1 in the kidney and the existence of a renal viral reservoir (Marras *et al*., 2002). Also, the ratio of infected to uninfected leucocytes in ejaculated semen specimens is highly discordant with paired blood samples, demonstrating that they derive from distinct populations of the infected cells (Kiessling *et al*., 1998).

Cervical viral shedding is more frequent in women with a higher plasma viral load and more advanced immunodeficiency (Gouldston *et al*., 1998). HIV-1 is present at high levels in genital tract secretions during the acute primary infection stage and is followed by reduced levels during the subsequent period of clinical latency and by increased levels during late-stage disease (Wahl *et al*., 1999). HIV-1 shedding was highest in the genital tract with more association of the combination of β-chemokines than virus load in blood indicating that local immune reaction strongly influence virus load in the cervical compartment (Iversen *et al*., 1998). In another study, Belec *et al*. (1995) indicated that IgG and IgA anti-HIV antibody titers were, respectively 30- and 12-fold higher in serum than in cervicovaginal secretions, but their mean specific activity were higher in cervicovaginal secretions than in serum, suggesting a local synthesis of both isotypes.

Although most researchers showed that most women had higher RNA concentrations in blood than in either swab or cervicovaginal lavage specimens, 4% of the women are found with higher RNA concentrations in the genital secretions. Even some have 10-fold higher RNA in genital tract specimens than in plasma (Kovacs *et al*., 2001). This in part may indicate that there is

compartmentalization between the peripheral blood and the vaginal mucosa. And, although there is an association between HIV-1 levels in plasma and genital secretions (semen and cervicovaginal fluid), there is variability in the ability to isolate HIV-1 RNA from genital secretions at different time points between individuals (Uvin and Caliendo, 1997; Mostad *et al*., 1998). For example, Mostad *et al*. (1998) evaluated cervical and vaginal secretions throughout the menstrual cycle in 17 women (21-31 sample times per subject) and found HIV-infected cells in 4% to 100% of the total endocervical swabs.

Vernazza *et al*. (1996) described 28 patients with detectable HIV-1 RNA in both semen and blood, with only 12 subjects showing levels within 1 log of each other. A similar proportion of subjects demonstrated a quantification discordance of greater than 10-fold in each direction. In the majority of the cases, discordant detection results in blood and genital compartments show plasma viremia without detectable genital secretion HIV-1 RNA. Cases of detectable genital secretion (cervicovaginal fluid and semen) HIV-1 RNA in individuals with undetectable plasma HIV-1 RNA levels have been described (Uvin and Caliendo, 1997; Kiessling *et al*., 1998; Zhu *al.*, 1996). Kiessling *et al*. (1998) have also described cases of culturable virus in semen from men who did not have culturable virus in blood.

Drug resistance-associated mutations appear in plasma virus several months before appearing in vaginal virus (Ellerbrock *et al*., 2001). In addition, infectious HIV is isolated from semen cells, but not from blood cells of an individual on triple antiretroviral therapy was found. The absence of major resistance-conferring mutations in the semen virus indicates that it was replicating in isolation from the antiviral agents (Kiessling *et al*., 1998). The compartmentalization of blood

and semen infection is further supported by genetic analysis of several infectious HIV clones isolated from semen cells and peripheral blood cells of another donor not on antiretroviral therapy using PR region of the *pol* gene. PR gene sequence analyses revealed significant divergence of the two viral populations. These findings confirm the distinct compartmentalization of HIV in the semen of this study, and support the concept that the semen HIV arises from an isolated reservoir of infection that may function independently in the patho-biology of HIV (Kiessling *et al*., 1998). Genetic and phenotypic compartmentalization of HIV-1 is found in different compartments. The presence of SI in blood and NSI phenotypic variants in genital fluid as a common property of subtype E (now AE recombinant) was observed by Kiessling *et al*. (1998).

Clinicians and patients must understand that effective antiretroviral therapy may reduce plasma HIV-1 RNA levels to undetectable levels even with ultra-sensitive assays, and still leave infectious virus in genital secretions. While sexual transmission on a population basis is likely to decrease with widespread use of effective antiretroviral therapy, individual infectiousness and transmissibility are highly variable across patients and even within patients over time. The evolution of resistance mutations for chemotherapies in blood and semen were frequently discordant, although over time similar patterns were seen. Sexual transmission of resistant variants may have a negative impact on treatment outcome in newly infected individuals and on the spread of diseases within a population (Eron *et al*., 1998). Another study by Dornadula *et al*. (1999) showed that resting $CD4^+$ T lymphocytes and cells in the other body compartments, including the genital tract, might contain replication competent proviral species in patients taking suppressive highly active antiretroviral therapy with undetectable viral RNA in peripheral blood.

This replicated virus may not only infect uninfected cells in the local microenvironment of viral producing cells, but may infect cells at a distance within the body.

Generally, the above review indicate that the development of safe and effective vaccine and therapeutics against HIV requires a good knowledge on genetic variability of HIV-1 virus in different body compartments. The existence of genetically varied group and detectable loads of viral RNA in nonblood body fluids stresses the importance of evaluating the impact of therapeutic treatment regimens and candidate vaccines in both blood and nonblood fluids (Shepard *et al*., 2000). But, still most researches for therapeutic and vaccine developments are based on blood viral genetic variability. Due to this reason, we intend to carry out comparisons of the genetic variation of HIV-1 in Ethiopian case, in the blood plasma and genital secretions of HIV-positive women using nested RT-PCR and automated sequencing of the C2V3 region of the *env* gene.

2. Objectives

2.1. General Objectives

The general objectives of this study are:

a) to compare the genetic variation of the HIV-1 in the blood plasma and genital secretions using nested RT-PCR and automated sequencing of the C2V3 of the *env* gene, and

b) to make phylogenetic analysis of the HIV-1 *env* sequences of circulating virus isolates in the STD study subjects and to determine the frequency of subtype C genetic subclusters C and C' viruses.

2.2. Specific Objectives

The specific objectives of this study are:

a) to determine the percentage of nucleotide and predicted amino acid sequence differences of HIV-1 in C2V3 of gp120 of the *env* gene of cell-free virus from blood plasma and vaginal secretions,

b) to investigate whether differences between amino acid sequences of cell-free HIV-1 in blood plasma and genital secretions resulted in differences of predicted cellular tropism by comparing the difference in the net charge and the signature amino acids in the V3 loop of HIV-1 infected women,

c) to construct the phylogenetic tree and observe the divergence between the HIV-1 sequences in genital secretions and in blood using the C2V3 region of the *env* gene, and

d) to relate the observed viral diversity with viral RNA load and CD4 count.

3. Subjects and Methods

3.1. Eligibility Criteria for Specimen Collection and Sample Selection

This analysis includes baseline data from examinations of HIV positive women enrolled in Teklehaimanot Clinic as part of the follow up of sexually transmitted diseases (STD) study under treatment. Blood plasma and cervicovaginal lavage samples were collected from HIV sero-positive women. Whole blood was collected in vacutainer tubes with ethylenediaminetetrachloroacetate (EDTA) (Becton and Dickinson, USA). Plasma was isolated by centrifugation at room temperature for ten minutes at 1180 revolution per minute (rpm) and stored at -80^0C. A cervicovaginal lavage was obtained by introducing 10 mL of phosphate buffered saline (PBS) into the vagina and by collecting the pooled fluid in the posterior vaginal fortex. Cervicovaginal lavage fluid samples were vortexed and centrifuged at room temperature, and cell-free supernatants of the cervicovaginal lavage were carefully removed and stored as aliquots at -80^0C with the DNA/RNA stabilizer. Of the study subjects, 32 sero-positive women were selected for the test of compartmentalization through the detection of the genetic variability based on C2V3 loop of the *env* gene. Only samples from the study participants with detectable viral loads in plasma and almost all cervicovaginal lavage of individuals were selected. The study subjects were not using antiretroviral therapy and there was no sexual intercourse for 72 hours preceding collection of cervicovaginal lavage and also no intra-vaginal medications. Clinical examinations for cervicovaginal lavage collection were not scheduled during menstruation to reduce the risk of blood contamination (Shepard *et al*., 2000; Ellerbrock *et al*., 2001).

Comment [s1]:

Comment [s2]:

Comment [s3]:

Comment [s4]:

3.2. HIV Screening

HIV screening was performed by HIV-SPOT (Genelabs Diagnostics, Singapore) and enzyme linked immuno-sorbent assay (ELISA) (Vironostika HIV Uni-Form II Organon, Teknika, the Netherlands). Discrepant results were confirmed by Western Blot Assay (HIV BLOT 2.2, Genelabs Diagnostics, Singapore).

3.3. Enumeration of Lymphocyte Subsets

Absolute number of leucocytes per microlitre of whole blood was obtained using coulter counter T540 (Coulter Electronics, Florida, USA). Lymphocyte subsets were enumerated based on three colors staining in whole blood, using a FACScan flow cytometer (Becton Dickinson, USA), according to the manufacturer's recommendations. The percentage of lymphocyte and the absolute count of leucocyte were used to calculate absolute count of the corresponding cell populations ($CD4^+$, $CD8^+$, B cells and NK cells) using the Multiset Analysis Software. However, only the CD4 count was used in this analysis.

3.4. Viral Load Determination

Assessment of the viral load in plasma and cervicovaginal lavage was done using the modified form of a nucleic acid sequence based amplification assay (NASBA) called Nuclisens NASBA Diagnostics (Organon Teknika, the Netherlands) on 200 μL plasma and cervicovaginal lavage samples according to the manufacturer's recommendations. In brief, RNA was extracted from blood plasma and cervicovaginal lavage using established guanidinium isothiocyanate silica bead RNA extraction method of Boom *et al*. (1990) and isothermally amplified in the presence of three synthetic RNA standards (Qa, Qb, Qc) of known high, medium and low concentrations,

respectively. These RNAs serve as internal caliberators each differing from the HIV-1 wild-type (WT) RNA by only a small sequence. 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 where the lower limit of detection of the viral load is 80 copies/mL. Results were also presented as $\log_{10}$ RNA copies/mL plasma or cervicovaginal lavage.

3.5. Nucleic Acid Isolation and Purification

HIV-1 RNA was isolated from the collected plasma and vaginal secretions according to guanidinium isothiocyanate (GuSCN) silica bead RNA extraction method of Boom *et al*. (1990). For this purpose, 200 µL of plasma/cervicovaginal lavage was added to 900 µL lysis buffer (prepared from 120 g GuSCN to 100 mL L1 buffer at a temperature of 60^0C with 8.8 mL of 0.5 M EDTA at pH = 8.0 and 2.4 mL (=2.6 g) Triton X-100 heated at 60^0C), containing GuSCN. L1 buffer is prepared by mixing 12.1 g Tris to 900 mL distilled H_2O with 42 mL 37% HCl pH = 8.0 up to the total volume of 1 L. 50 µL silica was added to the above mix, incubated for 10 minutes at room temperature, then centrifuged at 14000 rpm for 30 seconds. The pellet was washed after removing the supernatant with washing buffer (L2 prepared by mixing 120 g guandium isothiocyanate to 100 mL of L1 Buffer) and ethanol twice each and once with acetone with mixing and centrifugation. After washing, the pellet was dried at 56^0C for 10 minutes and RNA was eluted with 50 µL Low TE^{-4} (Louwagie Tris-EDTA) buffer for 15 minutes through shaking.

Comment [s5]:

3.6. Amplification of the C2V3 Region of the *Env* Gene

From the total RNA elute, 10 μL RNA elute was taken for the reverse transcription reaction and was reverse transcribed into cDNA by adding into 10 μL reverse transcription mix at 42^0C for 1 hour in the water bath with the total volume of 20 μL. Per sample, the total volume of mixes for reverse transcriptions were 10 μL with 0.5 μL of 1 units/μL avian myeoloblastosis virus reverse transcriptase (AMV), 2 μL of 10x RT buffer, 4 μL of 25 mM $MgCl_2$, 2 μL of 10 mM each dNTPs, 0.5 μL of 10 units/μL RNAsin, and 1 μL of 10 pmol/μL 3'-ED33, 5'-TTA CAG TAG AAA AAT TCC CCT C-3'. The product of the reverse transcription reaction was put at 95^0C for 5 minutes for the inactivation of the AMV. HIV-1 *env* C2V3 region was amplified using two rounds of amplification by PCR, in a thermal cycler (Perkin Elmer DNA Thermal Cycler 480). External PCR reactions were carried out in a 100 μL reaction mixture containing 10 μL of 25 mM $MgCl_2$, 8 μL of 10x PCR buffer, 0.5 μL of 5 units/μL of Taq polymerase, 1 μL of 10 pmol/μL 5'-ED31, 5'-CCT CAG CCA TTA CAC AGG CCT GTC CAA AG-3', and 60.5 μL distilled water with the 20 μL of the reverse transcription product. The dNTPs added for reverse transcription was enough for the first PCR. The first PCR condition was 1 cycle of denaturation at 95^0C for 4 minutes, 35 cycles for 1 min at 95^0C, 1 min at 55^0C, and 2 min at 72^0C for denaturation, annealing and elongation, respectively, followed by 1 cycle final incubation for 10 min at 72^0C for extension. This amplification resulted in a 536-bp DNA fragment of the C2V3 region. A nested PCR amplification of the C2V3 region of the gp120 was performed on the external PCR products using 5'-V3-M13 (17F), 5'-CAG GAA ACA GCT ATG ACA atg gia gic tag cag aa-3' and 3'-V3-M13 (-40R), 5'-CGC CAG GGT TTT CCC AGT CAC GAC aat ttc tig itc ccc tcc ig-3' primers. The nested PCR reactions were carried out in a 100 μL reaction mixture

Comment [D36]:

Comment [D37]:

containing 16 μL of 25 mM $MgCl_2$, 10 μL of 10x PCR buffer, 0.5 μL of 5 units/μL of Taq polymerase, 2 μL of 10 mM each dNTPs, 1 μL of 10 ρmol/μL 5'-V3-M13 (17F), 1 μL of 10 pmol/μL 3'-V3-M13 (-40R) and 59.5 μL distilled water with 10 μL of the first PCR product. The second PCR was for 25 cycles with the same condition as the first PCR. This second PCR resulted in, a 279-bp DNA fragment which was sequenced for this analysis. For those samples which were difficult to amplify and sequence, other nested primers 5INNgpCV3-dir, 5'-CGC CAG GGT TTT CCC AGT CAC GAC tta aat ggt agt mta gca gaa-3' and 3INNpgV3C-rev, 5'-CGA CGT TGT AAA ACG ACG GCC AGT gtt gta att tct agg tcc cct cct ga-3' were used for amplification starting from the first PCR products of primers ED33/ED31 for the second PCR. The reaction condition for second PCR by the new set of primer was carried out in a 50 μL reaction mixture containing 1.8 μL of 100 mM $MgCl_2$, 5 μL of 10x PCR buffer, 0.4 μL of 5 units/μL of Taq polymerase, 0.4 μL of 25 mM each dNTPs, 0.5 μL of 100 ng 5INNgpCV3-dir, 0.5 μL of 100 ng 3INNpgV3C-rev and 36.4 μL distilled water with 5 μL of the first PCR product for 25 cycles in the similar reaction conditions of the above.

3.7. Agarose Gel Electrophoresis

The presence of the amplified PCR products was verified on 1% agarose gels stained with ethidium bromide (EtBr) for both the first and second PCR products where 0.5 μg/mL EtBr was used as a DNA stain. To confirm amplification of the region of interest, 5 μL of the PCR product with 5 μL of distilled water and 5 μL of loading dye were mixed and loaded on the gel. 100 bp λ DNA fragment was used as a molecular weight marker. The gel was covered with 1500 mL 1x TAE (Tris-Acetate-EDTA) buffer and it was run at 100 milliampere (120 volt) for 1 hour and then it was observed under ultraviolet light where PCR positive results fluoresce. Finally,

Polaroid pictures were taken and the presences of the 536-bp and 279-bp bands were checked for the first and second PCR products, respectively.

3.8. Sequencing

The nested PCR products were cycle sequenced with Thermo-sequenase fluorescent-labeled primer cycle sequencing kit according to the manufacturer protocol (Amersham Pharmacia Biotech, Netherlands). The DNA polymerase, sequenase and the fluorescent-labelled primers, M1317 base reverse primer (5'-Fluorescein-d[CAG GAA ACA GCT ATG AC]-3') and M13-40 primer (5'-Fluorescein-d[CGC CAG GGT TTT CCC AGT CAC GAC]-3'), and ALF40FAM base direct primer (5'-Fluorescein-d[CGC CAG GGT TTT CCC AGT CAC GAC]-3') and ALFUNIFAM base reverse primer (5'-Fluorescein-d[CGA CGT TGT AAA ACG ACG GCC AGT]-3') for those difficult samples by the former sequencing primers were used using an A.L.F. DNA Sequencer Pharmacia Biotech following the manufacturer's instruction. The cycle sequence reaction mix containing the primer (2.1 µL, 3 pM/µL), distilled water (14.7 µL), DNA (5 µL, 50-100 ng) was prepared and 5 µL of this mix was added into four tubes labeled A, C, G, and T, followed by addition of 2 µL 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 minutes followed by 25 cycles of denaturation at 95^0C for 30 seconds and annealing/extension at 55^0C for 30 seconds. The cycle sequence products were mixed with loading dye, denatured to form single stranded DNA and loaded on the gel that was prepared in the following fashion. Sequence gel (120 mL) for A.L.F. sequencer sufficient for two runs was prepared consisting of urea (50.5 g), accugel [40% (29:1) acrylamide: bisacrylamide solution (19.4 mL)], distilled water (50.5 mL) and one spoon of deionizer. The mix (acryl mix) was left for an hour whilst being stirred. After one hour,

the mix was filtered and degassed, and 10x TBE (Tris-Borate-EDTA, 12 mL) was added. Finally, 60 µL N,N,N'N'-tetramethylethylenediamine (TEMED, catalyst) and 30% ammonium persulphate (APS, polymerization initiator) each 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. The gel plate was left for at least three hours for polymerization. The comb inserted before polymerization into the plates form holes following each lane. Lanes in the gel were washed after 3 hours, and is used for loading of the already prepared mix after the plate was arranged into the sequencing machine. The 0.6% 2000 mL TBE was put into the two buffer reservoirs fitted to the gel cassette and both connected to anode and cathode in order to allow flow of current. The sequencing machine was left on overnight with the appropriate current and voltage where the fragments produced following the Sanger's method move down in the polyacrylamide gel depending on their size and detected in laser beam following individual lanes. The electrogram was read from the computer for editing.

Comment [s8]:

3.9. Sequence Analysis

The sequenced products were edited using reverse and other direct sequences produced following the method in section 3.8. Samples difficult to sequence using the older sequencing primers were sequenced by the new set sequencing primers from the second PCR product of the new set of primers and edited with the help of sequenced results of the first and/or second set of sequencing primers. The sequences were aligned manually with previously obtained HIV-1 *env* C2V3 reference sequences of Ethiopian strains and with the consensus HIV-DNA sequence alignments with the Los Alamos database. Alignments and gaps were introduced for optimal alignment of the amino acid sequence; positions containing an alignment gap were excluded from pairwise

sequence comparisons. The pairwise nucleotide distances were generated by Kimura 2-parameter method included in the Molecular Evolutionary Genetics Analysis (MEGA) program. Phylogenetic analysis of the C2V3 nucleotide sequences was performed with the neighbor-joining (N-J) method of MEGA (Kumar, 1993) and the PHYLIP package, including the programs DNADIST, NEIGHBOUR (under the Kimura two-parameter model), DRAWTREE, SEQBOOT (100 sets) and CONSENSE (Felsenstein, 1996). Amino acid sequence comparisons were performed using the viral epidemiology signature pattern analysis (VESPA) program (Korber and Myers, 1992). MEGA was also used for synonymous and nonsynonymous genetic distance calculation with the modified Nei-Gojobori method with the Jukes-Cantor correction as described (Collins *et al*., 2000). A bootstrap value ≥ 70% based on one hundred replications were considered significant (Hills and Bull, 1993; Hills, 1997).

Comment [M9]:

3.10. Statistical Analysis

The statistical analyses were done using the statistical package SPSS for windows 10.0 version 4 for the comparison of clinical data, and synonymous and nonsynonymous nucleotide genetic distances. The t-test for means of continuous variables test, bivariate correlation analysis, the Mann-Whitney and Wilcoxon rank test for nonparametric methods in this package were used for the sake of the comparison of the above parameters.

3.11. Ethics

The study was performed as part of the Ethiopian Health and Nutrition Research Institute/Ethio-Netherlands AIDS Research Project (EHNRI/ENARP) STD-study, which was ethically cleared by the EHNRI and Science and Technology Commission. Informed consent was obtained from patients in this study. Institutional clearance from EHNRI and the Biology Department, Addis Ababa University was obtained.

4. Results

4.1. Study Subjects

The analysis includes baseline data from the initial examination of 32 HIV infected women among those enrolled in the STD/HIV co-infection follow-up study for both the two compartments. Matched samples of cervicovaginal lavage and blood plasma of HIV-1 infected women were examined. The median age of the women was 25 years (range 17-50 years) (Table 1). At the women's first study visits, their $CD4^+$ T lymphocyte counts varied from 25 to 670 cells/μL (median 190 cells/μL). Seventeen patients had CD4 cell counts below 200 cells/μL. HIV-1 RNA plasma virus loads ranged from 8,900 to 1,900,000 copies/mL (median 87,000 copies/mL), whereas cervicovaginal viral loads ranged from < 80 to 88,000 copies/mL (median, about 2,800 copies/mL only for those with HIV-1 shedding). As shown in Table-1, all the plasma samples had detectable viral load. Despite the detectable viral loads in plasma for all 32 women, for 2 (6.4%) of them there was no amplification in the cervicovaginal lavages and in 3 (9.4%), amplification was less than 80 copies/mL, which is below the lower detection limit of the Nuclisens assay. In cervicovaginal lavage of subject STD013, there was no amplification using the Nuclisens Assay. However, this sample was positive by PCR. Twenty-nine of the subjects (90.6%), with the exception of the three subjects (STD030, STD070 and STD075), were positive for STDs as shown in Table 1.

Table 1. Age, $CD4^+$ T lymphocyte cell count and HIV-1 viral load in plasma and cervicovaginal lavage of 32 HIV-1 infected women attending an STD Clinic in Addis Ababa.

Patient ID	Age	Absolute $CD4^+$ cell count/μL	Plasma viral load [a]	Cervicovaginal lavage viral load [b]
STD003	22	155	84000	1900
STD004	25	226	950000	3100
STD007	28	415	100000	7000
STD008	25	329	170000	1400
STD011	24	358	140000	220
STD012	18	212	35000	850
STD013	20	503	15000	No amplification [f]
STD015	27	101	9000	< 80 [c]
STD016	42	110	8900	< 80 [c]
STD017	43	122	57000	< 80 [c]
STD021	23	112	520000	2900
STD022	33	271	41000	2300
STD023	25	176	48000	1600
STD024	20	443	71000	1600
STD026	22	161	610000	55000
STD030 [d]	26	670	15000	No amplification
STD032	17	216	71000	9900
STD033	23	191	400000	83000
STD035	22	66	230000	2100
STD036	38	104	25000	2700
STD041	20	370	53000	1400
STD042	36	257	90000	9400
STD057	30	53	43000	2700
STD065	33	77	480000	88000
STD067	50	74	170000	26000
STD068 [e]	19	337	14000	23000
STD070 [d]	29	533	440000	59000
STD072	28	120	45000	2200
STD073	28	288	800000	69000
STD075 [d]	27	190	910000	29000
STD079	25	25	1900000	51000
STD088	24	81	710000	37000

Note:
[a] Cell-free HIV-1 RNA copies/mL in plasma
[b] Cell-free HIV-1 RNA copies/mL in cervicovaginal lavage
[c] Cell-free RNA below the level of quantification in cervicovaginal lavage (< 80 copies/mL).
[d] Subjects with no STD
[e] Subject with higher viral load in lavage than plasma
[f] Viral load negative in Nuclisens but amplification in the PCR

Although most women had higher RNA concentrations in blood plasma than cervicovaginal lavage specimens, 1 (3%) of the 32 women, subject STD068, had higher concentrations in the cervicovaginal lavage without considering the dilution made during washing for sample collection and RNA/DNA stabilizer for storage.

The actual cervicovaginal lavage viral load may be difficult to determine because the amount of the fluid in this compartment is unknown and may vary from person to person. But, for all individuals, the amount of PBS introduced for washing was equal, *i.e.*, 10 mL, similar volumes of nucleic acid stabilizer were added before freezing and 200 μL or cervicovaginal lavage was used as input in the Nuclisens assay.

Mean plasma HIV-1 RNA concentrations were 6-fold higher for women with genital tract shedding than for those with no shedding (Table-1). The amount of cell-free HIV-1 RNA in blood plasma was correlated with that in cervicovaginal lavages (Spearman's rank correlation coefficient (r) = 0.651, P = 0.0001 and Figure-1). No correlation was found between cell-free HIV-1 levels in either cervicovaginal lavages or blood plasma with $CD4^+$ T lymphocyte levels.

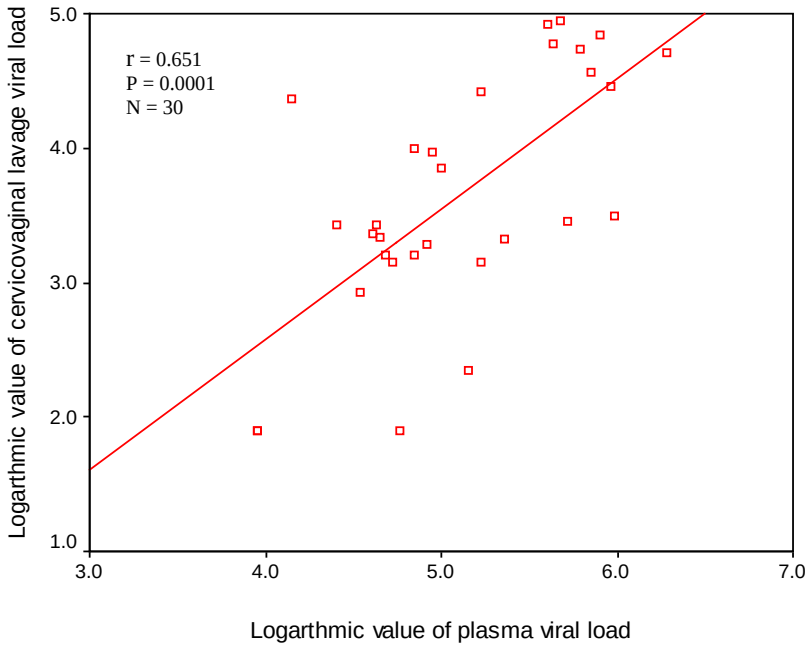

Figure 1. Comparison of the logarithmic value of viral load (copies/mL) of HIV-1 cell-free viruses in plasma and cervicovaginal lavage samples.

4.2. Polymerase Chain Reaction and DNA Sequencing

To analyze the C2V3 region of the *env* gene in both compartments, polymerase chain reaction (PCR) and DNA sequencing were performed using PCR and sequencing primers. A total of 32 plasma samples with the corresponding 30-cervicovaginal lavage samples (for subjects STD079 and STD088 no cervicovaginal lavage was available), were available for RNA extraction and amplification. All the blood plasma (100%) samples and 23 (77%) of the cervicovaginal lavage, with the exception of STD016, STD021, STD022, STD023, STD030, STD035 and STD068, were positive on gel for the PCR products of required molecular weight. From the amplified PCR products of blood plasma and cervicovaginal lavage viruses, 28 (87.5%), with the exception of STD026, STD036, STD057, and STD072, of the blood plasma and 22 (96%), with the exception of STD017, of the cervicovaginal lavages were sequenced and gave interpretable DNA sequences. This indicates that five amplicons (4 from plasma samples and 1 from vaginal samples) were difficult to sequence due to unknown reason.

4.3. Phylogenetic Analysis

All the sequenced products of the plasma and lavage gp120 C2V3 sequences were aligned with previously published Ethiopian subtype C sequences and reference sequences of other subtypes encompassing the same region from the HIV database. Based on the C2V3 sequences, the phylogenetic tree generated by a neighbor-joining (N-J) method indicated that all viruses belong to subtype C, except for subject STD003 who carried a subtype A virus as shown in Figure-2.

The phylogenetic tree reveals the two subclusters, C and C', showing the Ethiopian phylogeny with a bootstrap value of 50%. This value is not significant. Of the 31 subtype C samples, 19

(61.3%) were C' and 12 (38.7%) were C. In the overall phylogenetic analysis, the cervicovaginal lavage sequence of subject STD036 did not clearly cluster with C or C'. However, it clustered with C in the C/C' tree (data not shown). This may indicate presence of co-infection with the two subtype C subclusters (C/C') and this may affect the clustering because the sequences refer to populations of viruses in the individual subject.

The phylogenetic analysis also showed that the HIV-1 sequences from each patient clustered together, compared with those from other patients with a large bootstrap value which indicates that the two sequences come really from the same source but they are not identical. This also indicates the intra-patient variations between the viral species in plasma and cervicovaginal lavage of patients were not an artifact or contamination of samples. The branch length in phylogenetic tree between sequences of the corresponding compartments is different as could be seen from the tree (Figure-2). This implies that, even if the two sequences from the two compartments cluster together, there is a divergence, differ in rate of change or evolution and HIV strains with a longer branch are more diverse than those with a shorter branch.

As shown in the phylogenetic tree, the sequence from cervicovaginal lavage of STD007 was found to be subtype B and was found to be closer to the HIV-1 subtype B sequences from Brazil. The presence of subtype B, for the first time, in these sequences requires verification, if possible, by tracing the subject. If the result is confirmed true, it will be candidate for the investigation of double infections and should be the subject of future investigation.

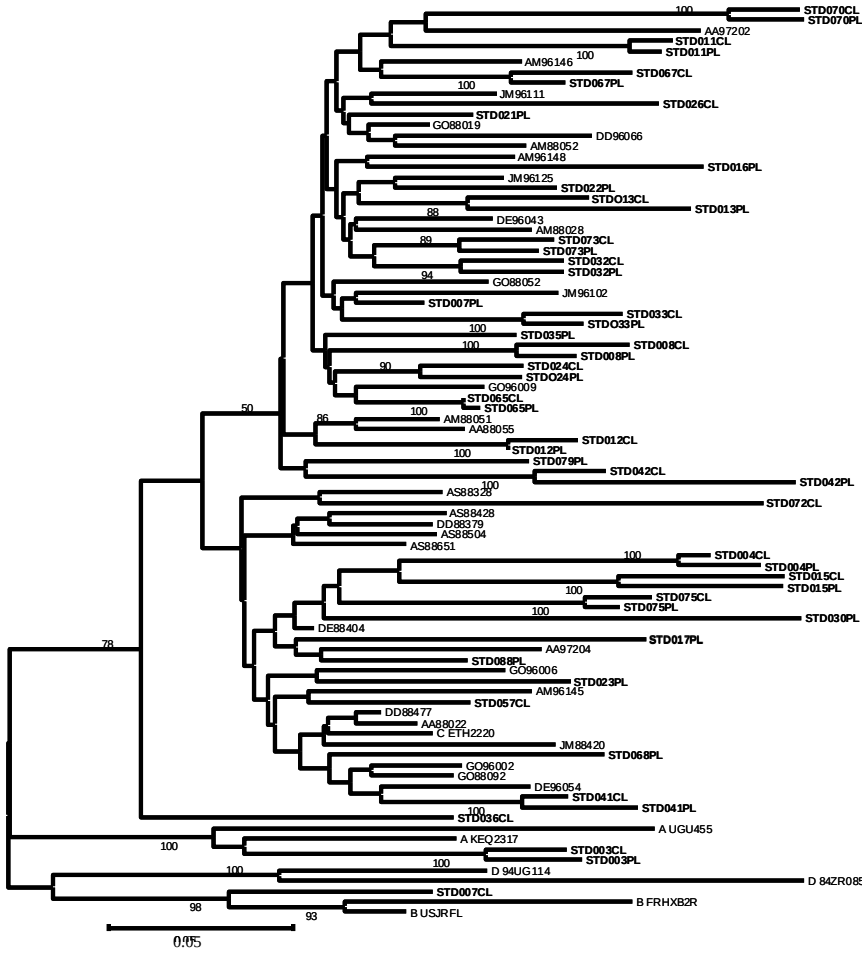

Figure 2. Unrooted phylogenetic tree of HIV-1 subtypes. Neighbor-joining tree of 264 bases showed evolutionary relationship of the C2V3 region of the gp120 of HIV-1 RNA from cervicovaginal lavage (CL) and plasma (PL) of 50 sequences from the STD study of which 28 were from plasma and 22 from cervicovaginal lavage with other published subtype C reference sequences. Other reference sequences include Ethiopian subtype C (ETH2220), and other subtypes which include A (UGU455 and KEQ2317), B (FRHXB2R and USJRFL), and D (94UG114 and 84ZR085) from Los Alamos database were used. The scale bar represents a genetic distance of 5%.

4.4. Comparison of CD4 and Viral Load in Ethiopian Subtype C Subclusters C and C'

Even if viral loads should be considered for groups matched for the same counts of CD4 cells, enough samples to split them into subgroups were not available. So it was tried to see whether the CD4 and age distribution data are comparable or not for the median and mean values. The comparison of the age and the CD4 counts in the two subclusters C and C' did not show any difference by the Mann-Whitney U-test ($P = 0.580$ and $P = 0.490$, respectively), making the viral load data of the two groups comparable. The median and mean values for the CD4 counts were found comparable, 149 cells/μL subcluster C and 216 cells/μL subcluster C', and 212 cells/μL subcluster C and 244 cells/μL subcluster C', respectively, although the standard deviation is quite high as shown in Table-2.

Table-2. The median and mean values with their standard deviations and errors of the age, absolute CD4 counts, plasma and cervicovaginal viral load in each of the groups C and C' subclusters from 32 HIV-1-infected women.

		Median	Mean	Std. Deviation	Std. Error
Age (yrs)	C	26.50	27.67	6.827	1.971
	C'	25.00	27.26	8.465	1.942
Absolute CD4 count (cells/μL)	C	149.0	212.5	174.6	50.400
	C'	216.0	244.3	153.8	35.290
Log_{10} plasma viral load	C	4.665	4.826	0.713	0.206
	C'	5.230	5.185	0.606	0.139
Log_{10} cervicovaginal lavage viral load	C	3.385	3.261	0.956	0.276
	C'	3.715	3.716	1.031	0.243

The statistical analysis of the plasma viral load showed that people infected with C' have higher viral loads than those infected with C. The values in plasma viral load were found significantly different by the independent t-test that gives weight to the mean value (P = 0.0009) and by the Wilcoxon signed rank test that gives weight to the median value (P = 0.002). However, although these data are exciting the groups compared are not big enough to be conclusive and even the standard deviation is bigger than the differences. The difference in cervicovaginal lavage viral load for the two subclusters was not done by any statistical test for the reason that no actual number of RNA was calculated for cervicovaginal lavage specimens due to the absence of dilution factor information at least in this case.

4.5. Comparison of the Amino Acid Consensus Sequences of the Subclusters C and C'

Comment [s10]:

The deduced consensus gp120 C2V3 amino acid sequences of the HIV-1 isolates from the 28 plasma (ConPLC/C') and 22 lavage (ConCLC/C') samples are shown in Figure-3 aligned with the consensus C (ConC), consensus of 1988 Ethiopian C/C' (Con88C/C'), consensus of Ethiopian 1996 C/C' (Con96C/C'). The consensus amino acid sequences comparison pointed out whether there are amino acid positions that characterize one particular group. This comparison between consensus sequences of the C and C' amino acid sequences of both plasma and lavage samples revealed significant changes at nine amino acid positions, most of which are outside of the V3 loop. In these case, the amino acids at positions Isoleucine (I) 270 Valine (V), Glutamine (Q) 286 Histidine (H), Asparagine (N) 288 Lysine (K), Asparagine (N) 294 Valine (V), Glutamic acid (E) 304 Lysine (K), Isoleucine (I) 306 Methionine (M), Glycine (G) 333 Glutamic acid (E), Lysine (K)/Arginine (R) 344 Glycine (G) and Lysine (K) 348 Glutamine (Q) stably distinguish C from C'.

As described above and shown in Figure-3, the V3 loop region revealed two significant amino acid changes where the Glutamine (E) at position 304 and Isoleucine (I) at position 306 in subcluster C virus were substituted by Lysine (K) and Methionine (M), respectively, in subcluster C'. In addition to the 2 amino acid differences in the V3 loop, there were seven amino acid differences in the flanking region of the V3 loop although the biological significance of these differences in amino acids is not known. So, further investigation is needed to see the biological differences of the two subclusters of subtype C.

The differences in the amino acid consensus sequences of the plasma and cervicovaginal lavage within subcluster C were at seven positions and within subcluster C' at four positions. However, the amino acid sequence differences within each of the two subclusters were outside of the V3 loop region (Figure-3). This is one important indication that the plasma sequences were different from the cervicovaginal lavage sequences with the potential that the consensus may avoid sequencing/PCR mistakes/artifacts.

The length of the V3 loop, which is 35 amino acids long and is surrounded by the two-Cysteine residues and GPGQT (Glysine-Proline-Glysine-Glutamine-Threonine) motif, typical of subtype C viruses, and Asparagines (N) at the three NNN residue glycosylation site positions 5-7 were conserved in all the consensus sequences. As shown in Figure-3, all groups had a greater variability towards the right end downstream of the V3 loop of gp120 in the *env* gene, which is the most variable region. However, the V3 loop region was relatively conserved in all consensus sequences in contrast to the region immediately downstream of it.

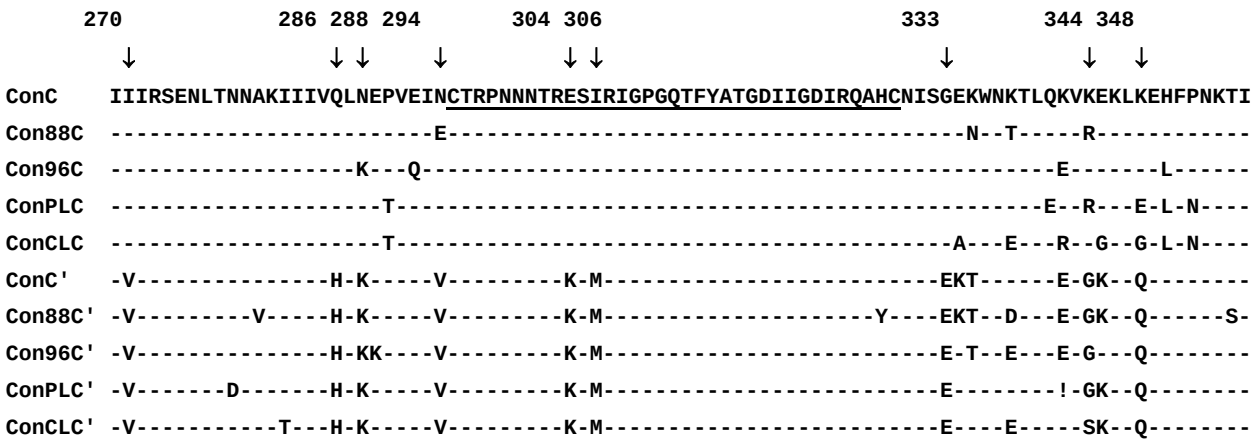

Figure 3. VESPA supported amino acid sequence comparison among different groups of predicted consensus amino acid sequences. The envelope gp120 V3 region for three periods of the two subclusters C and C' of HIV-1 subtype C of Ethiopian epidemic of which ConC/C', Con88C/C', Con96C/C', ConPLC/C' and ConCLC/C' refer to total Ethiopian consensus C, consensus of 1988, consensus of 1996 of Ethiopian sequences and consensus plasma and consensus cervicovaginal lavage 2001 sequences of the STD study of the respective Ethiopian subclusters C and C', respectively. The sequences are shown in comparison with the consensus C subcluster of Ethiopian subtype C. The underlined amino acids sequence refers to V3 loop of Gp120. Dashes indicate amino acid identities. ! indicates unreadable codon.

4.6. Comparison of the Predicted Phenotypes of the Cell-Free Virus

Figure-4 represents the amino acid sequence alignment of the 18 pairs of samples of the STD study for the comparison of genetic compartmentalization of viruses in the cervicovaginal lavage and plasma. All V3 loops including subtype A (STD003) were 35 amino acids long. Most of the amino acid sequences showed the characteristic C-subtype pattern of GPGQT at the apex. Some of the sequences (STD012, STD033, STD070 and STD075) showed a consensus subtype A apex (GPGQA). The viruses from cervicovaginal lavage of subject STD007 show GPGRA motif of typical subtype B and it was classified as subtype B in the phylogenetic tree (Figure-2). The first Asparagine (N) at the three NNN residue glycosylation site positions 5-7 was changed into Glycine (G) in 6 of the individuals from the total 18 pairs of sequences (33.33%) while it was changed into Serine (S) in one pair of samples (5.55%) from the two compartments (STD070). In subject STD041 Serine (S) in the cervicovaginal lavage substitutes Asparagine (N). At position 13 of subject STD007, Arginine (R) in plasma is replaced by Histidine (H) in the cervicovaginal lavage and in subject STD073 Serine (S) in plasma is replaced by Lysine (R) in the cervicovaginal lavage; at position 29 of subjects STD008 in plasma, STD032 cervicovaginal lavage, STD065 both in plasma and cervicovaginal lavage and STD070 in plasma, and at position 25 in subject STD075 of the cervicovaginal lavage, the negatively charged amino acids, Aspartic acid (D) and Glutamic acid (E) downstream of the starting Cysteine amino acid of the V3 was substituted by Asparagine (N) which may affect the phenotype of the virus by affecting the charge and may have a role in resistance towards the neutralizing antibodies due to the effect in glycosylation site (De Wolfs *et al*., 1990)

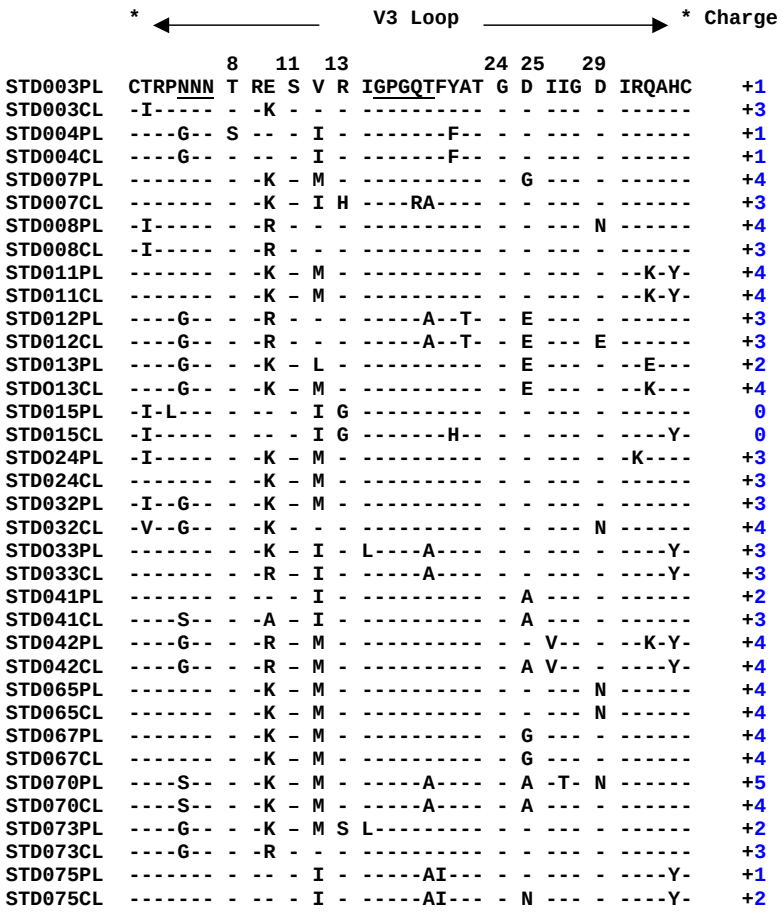

Figure 4. Alignment of the V3 loop derived amino acid sequence and net charge.The sequences of each of the 18 patient isolates were shown both for plasma (PL) and cervicovaginal lavage (CL). Dashes (-) indicate concurrence with the STD003PL with amino acid homology. Net charge is indicated to the right of the amino acid sequence. The GPGQT crown and N-linked glycosylation sites are underlined (* = Cysteine residue and 8, 11, 24, 25, and 29 = signature positively charged amino acids which determine the phenotype of the HIV-1 viruses and 13, 25 antigenic binding specificity).

To investigate whether differences between amino acid sequences of cell-free HIV-1 in blood plasma and cervicovaginal lavage resulted in differences of predicted cellular tropism, the net charge and signature amino acids in the V3 loop of the major HIV-1 species in plasma and cervicovaginal lavage of 18 HIV-1 infected women. For each sample tested in both compartments in this study, the net charge in the V3 loop predicted the same cellular tropism, *i.e.*, macrophage tropic in both compartments except the virus in the plasma of patient STD070 probably being SI because it has a charge +5. This individual was not in clinical AIDS stage and its absolute CD4 count was 533 during sample collection. Similarly, the distribution of charged amino acids in the V3 loop of viral envelope protein gp120 indicates that no positively charged amino acids were found at positions 8, 11, 24, 25 and 29 as shown in Figure-4, which is an indication that almost all subjects carry nonsyncytium inducing (NSI) viruses.

There were differences in the amino acid sequences between plasma and cervicovaginal lavage in all subjects, except for STD011, STD065 and STD067. Furthermore, the overall V3 positive charge was different between the plasma and cervicovaginal lavage sequences for 50% of the analyzed sample pairs. There was also a difference in the position of the negatively charged Aspartic acid (D) and Glutamic acid (E) and positively charged Lysine (K), Arginine (R) and Histidine (H) amino acids. For subjects STD003, STD013, STD032, STD041, STD073 and STD075, the cervicovaginal lavage sequences had higher V3 positive charge than the plasma amino acid sequences in contrast to the samples of STD007, STD008 and STD070, where the cervicovaginal lavage amino acid sequences had lower V3 positive charge than the plasma amino acid sequences. Thus, this result showed relatively higher net positive charge in cervicovaginal lavage in 6 of the 18 pairs. However, in 3 of 18 women the net positive charge in plasma was higher than in cervicovaginal lavage, and in 9 of the 18 women the net positive charge was equal

in the two compartments. Generally paired sample t-test showed that there is no significant net charge difference (P = 0.205) between amino acids of the respective compartments.

4.7. Comparison of Viral RNA Genotypes

The major genotypes in the populations of viruses based on the C2V3 region of HIV-1 RNA to see whether there is a divergence or not in matched blood plasma and vaginal secretions were sequenced. Comparison of the percentage differences of the predicted amino acid sequences of C2V3 region of cell-free virus in the two compartments was performed in order to investigate possible differences in HIV phenotypes of 17 HIV infected women (Table-3). Ellerbrock and his colleagues (2001) showed that≥ 5% predicted amino acid difference is assumed to be the minimum divergence value. In this study, the predicted amino acids differences between plasma and cervicovaginal lavage for these women of the V3 loop region were 1.0%, 2.1%, 2.6%, 2.9%, 3.2%, 4.2%, 4.5%, 4.5%, 4.8%, 5.3%, 5.4%, 5.9%, 6.4%, 8.3%, 10.4%, 11.4% and 14.7%, indicating that in eight (47.1%) of the women the virus in the plasma and cervicovaginal lavage constitute different populations.

It is possible that HIV-1-infected cells from blood populate in the genital tissues and are the major source of cell-free HIV-1 in genital secretions as cited by Ellerbrock *et al*. (2001). To address this possibility, the major species of HIV-1 in blood plasma and cervicovaginal lavage RNA based on the nucleotide difference as indicated in Table-3 was compared. Previous studies have shown that samples with a < 2% nucleotide difference in a variable region of the HIV-1 genome originate from the same source (Learn *et al*., 1996). All (100%) of the 17 women had a C2V3 region nucleotide divergence of > 2% in the two compartments. The median nucleotide difference between the C2V3 sequences of viral RNA in cervicovaginal lavage samples and

plasma was 8.3%. In none of the 17 women tested, the nucleotide sequences of the C2V3 region in the gp120 envelope in the two compartments were identical or had < 2% nucleotide difference.

Table 3. Percentage of predicted amino acid and nucleotide differences in the C2V3 region of the human immunodeficiency virus type 1 (HIV-1) gp120 envelope between cell-free virus population in peripheral blood and cervicovaginal lavage samples from 17 HIV-1-infected women.

Subject ID	Subtype	% amino acid sequence differences	% nucleotide sequence differences
STD003	A	8.3	11.0
STD004	C	2.9	3.7
STD008	C’	4.5	9.7
STD011	C’	2.1	3.6
STD012	C’	2.6	2.6
STD013	C’	11.4	15.0
STD015	C	14.7	24.0
STD024	C’	5.9	9.0
STD032	C’	6.4	9.4
STD033	C’	4.2	8.6
STD041	C	4.8	7.8
STD042	C’	10.4	15.2
STD065	C’	1.0	2.4
STD067	C’	4.5	8.3
STD070	C’	5.4	6.8
STD073	C’	5.3	6.8
STD075	C	3.2	4.7
Median		4.8	8.3

Note: % amino acid differences ≥ 5% and % nucleotide differences ≥ 2% are assumed to be indicators of divergence of the cell-free RNA populations in peripheral blood and cervicovaginal fluid as indicated by Ellerbrock *et al*. (2001). The % amino acid difference is assumed to be equal to the nonsynonymous distances because this distance is the result of 95%, 100% and 28% the first, second, and third nucleotide changes, respectively, in the codons. And % nucleotide difference is a result of the sum of synonymous and nonsynonymous nucleotide distances according to Nei and Gojoberi (1986).

4.8. Synonymous and Nonsynonymous Genetic Distances

The calculation of average synonymous and nonsynonymous genetic distances of overall codons in the C2V3 region of HIV-1 was done by comparing matched plasma and CVL sequences. The mean synonymous and nonsynonymous genetic distances were significantly different by the paired t-test ($P = 0.0003$) when all subjects were treated together. The nonsynonymous mean distances were higher than the synonymous distances (0.053 versus 0.029; $P = 0.021$) for C' and still higher but not statistically significant for C (0.064 versus 0.037; $P = 0.248$) by the Mann-Whitney test (Table 2). This resulted in a ratio of $ds/dn < 1$ which is an indication of positive selection. The mean of synonymous and nonsynonymous genetic distances among the sequences of the C group seems higher than those of the C' group in the study population described here (Table 2). However, these differences were not statistically significant by one-way ANOVA for the two subclusters of subtype C ($P = 0.537$) and ($P = 0.622$) for both synonymous and nonsynonymous genetic distances, respectively.

The absolute CD4 count in blood was negatively correlated but not significant with the ds/dn ratios (Spearman's rank correlation coefficient (r) = -0.347, $P = 0.172$ and Table 2) which may indicate that the magnitude of the nonsynonymous substitution increases when the individual infected has strong immune pressure and the fittest virus population is positively selected and proliferate to dominate the environment in the host. The viral load in blood plasma was positively correlated with ds/dn ratio but not significant (Spearman's rank correlation coefficient (r) = 0.036, $P = 0.890$), which may indicate that the magnitude of the nonsynonymous substitution decreases when the individual infected has developed higher viral load, in other words, as the immune pressure decreases, the viral load increases.

Table 4. The mean cervicovaginal lavage versus plasma synonymous (ds) and nonsynonymous (dn) nucleotide distance for each couple of sequences with their standard deviation.

Subject ID	Subtype	Mean ds (standard deviation)	Mean dn (standard deviation)	ds/dn
STD003	A	0.027 (±0.019)	0.083 (±0.024)	0.325
STD008	C’	0.052 (±0.028)	0.045 (±0.018)	1.156
STD011	C’	0.015 (±0.015)	0.021 (±0.011)	0.714
STD012	C’	0.000 (±0.000)	0.026 (±0.011)	0.000
STD013	C’	0.036 (±0.022)	0.114 (0.028)	0.316
STD024	C’	0.031 (±0.021)	0.059 (±0.018)	0.525
STD032	C’	0.030 (±0.020)	0.064 (±0.018)	0.469
STD033	C’	0.044 (±0.026)	0.042 (±0.015)	1.048
STD042	C’	0.048 (±0.025)	0.104 (±0.028)	0.462
STD065	C’	0.014 (±0.014)	0.010 (±0.007)	1.400
STD067	C’	0.038 (±0.022)	0.045 (±0.020)	0.844
STD070	C’	0.014 (±0.015)	0.054 (±0.019)	0.259
STD073	C’	0.015 (±0.015)	0.053 (±0.016)	0.283
Mean value	**C’**	**0.029 (±0.015)**	**0.053 (±0.030)**	**0.547**
STD004	C	0.008 (±0.008)	0.029 (±0.017)	0.276
STD015	C	0.093 (±0.027)	0.147 (±0.048)	0.633
STD041	C	0.030 (±0.021)	0.048 (±0.015)	0.625
STD075	C	0.015 (±0.015)	0.032 (±0.013)	0.469
Mean value	**C**	**0.037 (±0.039)**	**0.064 (±0.056)**	**0.578**

Note:
ds/dn < 1 refers to positive selection
ds/dn >1 refers to negative selection
ds/dn = 1 refers to neutral selection

The pairwise mean synonymous (ds) and nonsynonymous (dn) genetic distances among different consensus sequence groups were calculated along the synonymous and nonsynonymous sites as shown in Table-5. This is a different representation of a mean value since the consensus does not include individual differences but only group differences. The distances between the groups are all comparable and close to 10% both for synonymous and nonsynonymous distances. The nearly 10% synonymous genetic distances between the two groups indicate the diversity at the protein level and the value for nonsynonymous distances between the two groups indicate the differential evolution in the two subclusters.

The distance between the consensus sequences of the cervicovaginal lavage and plasma within the group are much lower, which is characteristic of intra-patient variation. This is not an unexpected result in blood plasma or cervicovaginal lavage sequences with high degree of sequence identity within such individuals, because any two virions may have shared a common ancestor within the patients. The synonymous genetic distance within the subcluster C' is the lowest which is an indication of similarity in evolutionary stage but the presence of relatively high (2%) nonsynonymous genetic distances is an indication of the presence of factors which lead the virus to evolve.

Table 5. The mean synonymous (ds) and nonsynonymous (dn) nucleotide distances of pooled sequences. The genetic distances of consensus sequences of the Ethiopian pooled subtype C with subclusters C/C' of two different years and these clusters within cervicovaginal lavage and plasma of the present study with the standard deviation

	Mean ds (standard deviation)	Mean dn (standard deviation)
EthC/EthC'88	0.103 (±0.040)	0.107 (±0.025)
EthC/EthC'96	0.108 (±0.041)	0.089 (±0.027)
CLC/CLC'	0.100 (±0.043)	0.119 (±0.032)
PLC/PLC'	0.110 (±0.043)	0.106 (±0.028)
PLC'/CLC'	0.000 (±0.000)	0.021 (±0.010)
PLC/CLC	0.030 (±0.021)	0.042 (±0.015)

As shown below in Figure-5 and Table-6, the distance differences for both synonymous and nonsynonymous nucleotide distances generally seems increasing over the years from 1988 to 1996 and seems more increasing for both plasma and cervicovaginal lavage sequences obtained in the present study. It indicates that the virus in the Ethiopian HIV-1 epidemic is in continuous evolution over time. This is an indication of the presence of strong positive selection due to the presence of immune pressure may be due to continuous immune activation in this population.

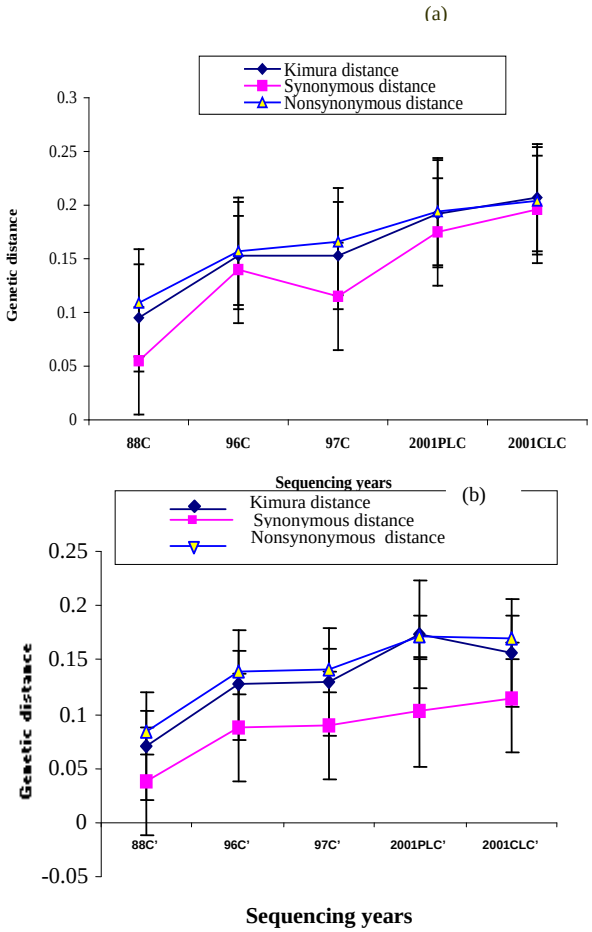

Figure 5. Kimura, mean synonymous and nonsynonymous genetic distances for both plasma (past + present sequences) and cervicovaginal lavage sequences of the present study. The ratios are derived from C2V3 sequences of the two Ethiopian subclusters C and C' (a and b, respectively) obtained from HIV-1 infected patients with the corresponding years. The line graph indicates the parallel increment of all parameters within the years.

Comparison of the plasma sequences of the former and the present studies showed that the synonymous distance for subtype C subcluster C' sequences (7.7%) was less compared to subtype C subcluster C (9.7%) while the nonsynonymous distance is relatively comparable (13.0% and 14.3%). For the cervicovaginal lavage sequences of the present study, the synonymous distance among subcluster C' sequences (11.5%) was less compared to subcluster C (19.6%) while the nonsynonymous distance is relatively comparable (17% and 20.4%). For the plasma sequences of the present study, the synonymous distance among subcluster C' sequences (10.2%) was less compared to subcluster C (17.5%) while the nonsynonymous distance is relatively comparable (17.1% and 19.4%). The calculation of the genetic distance between C and C' (C/C') subcluster viruses resulted in 12.5% at synonymous and 16.4% at nonsynonymous sites.

Synonymous and nonsynonymous genetic distances increased with time due to the parallel accumulation of synonymous and nonsynonymous substitutions (with $P < 0.001$ for both of the two variables) and nonsynonymous substitutions overweigh within the given time. The paired t-test for synonymous versus nonsynonymous distances of all sequences with no partition by clustering was found highly significant ($P = 0.0001$ for both variables). Even though the C' was shown to be dominating in the present study, the mean of synonymous and nonsynonymous nucleotide variation among the sequences of the C group seems higher than in the C' group but not statistically significant ($P = 0.113$ and $P = 0.307$, respectively) by one-way ANOVA. When the analysis was performed independently between the two subtype C subclusters C and C', the synonymous and the nonsynonymous genetic distances were not statistically different ($P = 0.465$ and $P = 0.076$) by the Mann-Whitney test for subcluster C and subcluster C', respectively.

Table 6. The mean synonymous (ds) and nonsynonymous (dn) nucleotide distance variation within a group of consensus sequences of pooled Ethiopian subtype C from C/C' subclusters of C2V3 region of the *env* gene. Pooled C and C' of different years and pooled results of the cervicovaginal lavage and plasma sequences of the present study (2001) with standard deviation of each group are indicated.

Subtype	Number of Sequences	Sequence group	Kimura Distance (standard deviation**)**	Mean ds (standard deviation)	Mean dn (standard Deviation)	ds/dn
C/C'		ETH C/C'	0.156 (±0.013)	0.125 (±0.020)	0.164 (±0.022)	0.762
C	70	ETHC	0.132 (±0.012)	0.097 (±0.012)	0.143 (±0.020)	0.678
	28	88C	0.095 (±0.010)	0.055 (±0.011)	0.109 (±0.017)	0.505
	17	96C	0.153 (±0.014)	0.140 (±0.021)	0.157 (±0.023)	0.892
	21	97C	0.153 (±0.013)	0.115 (±0.016)	0.166 (±0.023)	0.693
	9	PLC	0.192 (±0.016)	0.175 (±0.025)	0.194 (±0.024)	0.902
	7	CLC	0.207 (±0.020)	0.196 (±0.031)	0.204 (±0.030)	0.961
C'	77	ETHC'	0.117 (±0.010)	0.077 (±0.009)	0.130 (±0.017)	0.592
	18	88C'	0.071 (±0.009)	0.038 (±0.008)	0.083 (±0.014)	0.458
	20	96C'	0.127 (±0.012)	0.088 (±0.015)	0.139 (±0.019)	0.633
	40	97C'	0.129 (±0.012)	0.090 (±0.010)	0.141 (±0.020)	0.638
	16	PLC'	0.173 (±0.018)	0.102 (±0.016)	0.171 (±0.023)	0.596
	13	CLC'	0.157 (±0.015)	0.115 (±0.018)	0.170 (±0.026)	0.676

Note:

Significant value of sequence differences in each of the 13 groups for both synonymous and nonsynonymous genetic distances is $p < 0.001$ by two tailed test.

Ds/dn < 1 refers to positive selection

Ds/dn >1 refers to negative selection

Ds/dn = 1 refers to neutral selection

5. Discussions

In this study, sequencing of the C2V3 region of the gp120 of the *env* gene was performed to see if there exists compartmentalization of HIV in peripheral blood and vaginal mucosa in samples collected from HIV-1 positive women attending an STD clinic in Addis Ababa, Ethiopia. Both cervicovaginal lavage and blood plasma samples were used for this compartmentalization testing through phylogenetic and genotypic analysis.

There were discordance in amplification and sequencing of the viral PCR product in plasma and cervicovaginal lavage samples. The most probable reason for problem in sequencing of these samples could be the presence of some fragmented products and protein materials in the sequencing mix, which come from the nested PCR product due to the absence of purification steps and not updated sequencing machine. The reasons contributing to the absence of amplification could be low viral load, presence of amplification inhibitors, the aberrant nature of the RNA and possibly repeated freezing and thawing of the samples during laboratory activities which may result in RNA degradation.

All the present 31 sequences belong to subtype C except for subject STD003 who carries a subtype A virus. This is an indication of the dominance of subtype C in the Ethiopian epidemic of HIV/AIDS, with sporadic subtype A and D infections as reported by Abebe *et al*. (1997, 2000) and Hussein *et al*. (2000). The present subtype A is the fifth documented HIV-1 subtype A infection in Ethiopia. Four earlier reported subtype A viruses are: AFAR11-A detected from a 25-years old man in Afar Regional State from police recruits (Zewde *et al*., 2001), TP95001-A from a pregnant woman and AA97206-A from a commercial sex worker in Addis Ababa (Abebe *et al*.

1997, 2000) and A3099-AC recombinant, where C at its V3 loop region, with an almost full length sequence (Sherefa *et al*., 1996).

The phylogenetic tree showed the two subclusters, C and C' of the Ethiopian subtype C subclusters, with no significant bootstrap value. Of the 31 subtype C samples, 19 (61.3%) were C' and 12 (38.7%) were C. The percentage of C' was lower than the previous report by Rinke de Wit *et al*. (2002), where 75% of the sequences belong to subtype C' while 25% to subtype C. These results are in contrast to Abebe *et al*.'s report (2000) that showed that from the analyzed samples 43% of 1988, and 52% of 1996 and 1997 contain a C' virus. These studies confirm the spreading of C' virus among the study population with time.

Patients infected with subtype C subcluster C' had significantly higher plasma viral loads than those infected with subcluster C. The comparison of the age and the CD4 counts in the two subclusters C and C' did not show any differences. This is in agreement with the finding by Rinke de Wit *et al*. (2002). The dominance of subcluster C' in comparison with C and the presence of higher viral load in subcluster C' than C might be due to the selective advantage in terms of transmission of C' viruses. In fact, this suggestion is in agreement with Pollakis *et al*. (unpublished data) on the presence of C/C' recombinant viruses, where all the recombinants carried the C' virus in the *env* region. It indicates that C' viruses could be selectively advantaged possibly due to their higher replication efficiency.

Diversity in the viral genome is due in part to the high error rate and lack of proof reading activity of the retroviral reverse transcriptase (Coffin, 1986b; Preston *et al*., 1988). It is also influenced by the cellular (Epstein *et al*., 1991; Itescu *et al*., 1994) and immunological (Borrow *et al*., 1997) environment of the host. Provirus studies have shown that HIV-1 proviral sequences

Comment [s11]:

can be non-uniformly distributed throughout the body. Differences in the composition of virus population in plasma, PBMC and other compartments have been described for a number of viral genes that support the concept of anatomically distinct, independently evolving quasispecies. Distinct proviral sequence variants have been reported in some patients in PBMC compared to brain tissue (Epstein *et al*., 1991; Hughes *et al*., 1997; Korber *et al*., 1994; Pang *et al*., 1991; Power *et al*., 1994; Wong *et al*., 1997), cerebrospinal fluid (Kuiken *et al*., 1995; Steuler *et al*., 1992), spleen (Epstein *et al*., 1991), lymph node (Ball *et al*., 1994, Sheehy *et al*., 1996), lung (Itescu *et al*., 1994) and semen (Kiessling *et al*., 1998).

Detection of HIV-1 RNA in cervicovaginal secretions is primarily associated with increased plasma HIV-1 RNA load (Gouldston *et al*., 1998). In this study, the subjects were not on anti-retroviral treatment and the amount of cell-free HIV-1 RNA in blood plasma was correlated with that in vaginal secretions. This correlation should be evaluated in subjects on anti-retroviral therapies. However, statistical test of significance of viral loads between the two compartments is difficult due to the dilution difference in different individuals and there is a need of developing a method to know the actual dilution factor in the cervicovaginal compartment. Although most women in this study had higher concentrations in blood plasma than cervicovaginal lavage specimens, one woman had higher RNA concentrations in the cervicovaginal lavage. This result is in agreement with the work by Kovacs *et al*. (2001) that 4% of the subjects in their study had a higher cervicovaginal viral load than the plasma viral load. There is an indication that these two compartments are spatially different replication reservoirs; otherwise, in the absence of this compartmentalization with the presence of the dilution of the cervicovaginal lavage samples, it is expected to see less amount of RNA viral load in the genital compartment. Effective treatment of STDs results in significant reduction of genital shedding while failure of treatment is associated

with persistently elevated HIV shedding (Wolday *et al*. AIDS Conference, Barcelona, 2002), and the later may be associated with increasing viral diversity due to selection pressure of the persistent immune activation.

Phylogenetic analysis of these sequences showed that the HIV-1 sequences from both plasma and cervicovaginal lavage of each patient clustered together, compared with those from other patients which indicated that the two sequences come really from the same source but they are not identical. This implies that even though the two sequences from the two compartments cluster together, there is a divergence and HIV strains with longer branch are more diverse than the shorter branch. These data provide phylogenetic evidence for HIV compartmentalization between peripheral blood and vaginal mucosa. The data provide evidence for replication of the HIV-1 in vaginal mucosa and indicate that HIV strains residing in the vaginal mucosa microenvironment are not in a state of unrestricted bi-directional movements with viruses circulating in the blood. This in turn suggests that the vaginal mucosa may serve as a viral reservoir potentially harboring HIV-1 strains that have evolved under tissue specific selection pressures. The work by Birdsall *et al*. (2002) showed the above fact that reverse migration of HIV-1 infected leucocytes out of the perivascular reservoirs may provide a way to disseminate HIV-1 and expand the body burden of the virus in some patients receiving highly active antiretroviral therapy.

All these results are in agreement with results from other studies done in matched HIV-1 sequences derived from cervicovaginal fluid and from PBMC (Poss *et al*., 1995; Overbaugh *et al*., 1996; Ellerbrock *et al*., 2001) even though the RNA populations in the vaginal mucosa and blood plasma were compared. Genetic differences were found between proviral variants in the peripheral blood and those in the genital secretions in women infected with HIV-1 envelope

sequence subtypes A and D (Poss *et al*., 1995). Evidences support this independent evolution of mucosal lineages in individuals chronically infected with HIV-1 (Overbaugh *et al*., 1999).

The prediction of cellular tropism, in the 35 amino acids V3 loop predicted positive charges < 5.0 which suggests that the virus is macrophage tropic, whereas positive charges ≥ 5.0 suggest that the virus is T cell tropic (Fouchier *et al*., 1992). Amino acids at positions 11, 24 and 25 in the V3 loop are defined and used as signature amino acids for the determination of SI and NSI phenotypes (Chesebro *et al*., 1992; De Wolf *et al*., 1994) of which these are proved to be true in A, B, D and E (now AE recombinant) subtypes. In addition, it is indicated that amino acid positions 8, and 29 were important for the syncytium-inducing, high replication rate conferring property of the V3 Region (de Jong *et al*., 1992a; Abebe *et al*., 1999) and amino acid positions 13 and 25 are thought to be used in antigen binding specificity (Wolfs *et al*., 1991; Shioda *et al*., 1994).

In the present study, differences in the predicted amino acid sequences of plasma and cervicovaginal lavage at seven positions for subcluster C and at four positions for subcluster C' were found. This is one important indication that the plasma sequences were different from the cervicovaginal lavage sequences. A more positive charge of the V3 loops were found in the blood strains compared with genital fluid strains (Sutthent *et al*., 2001). However, relatively higher net positive charge in cervicovaginal lavage in 6 of the 18 pairs was found although in 3 of 18 women net positive charge in plasma was higher than cervicovaginal lavage viruses. The finding that there are not only amino acid differences but also V3 positive charge differences might reflect the difference in quasispecies viruses in the two compartments. There may be also a co-receptor usage difference in the two environments following the difference in the charge because

it is already proven that charge has a role in resulting different phenotypes of viruses (Fouicher *et al*., 1992).

The analysis of nucleotide and the predicted amino acid sequences of major quasispecies of HIV-1 RNA in blood plasma in comparison with cervicovaginal lavage, indicates that local HIV replication in the female genital tract provides the opportunity for development of virus phenotypes that may be different from those in blood. In this study, the predicted phenotype in almost all subjects was nonsyncytium inducing except for subject STD070 who had a predicted syncytium inducing phenotype in plasma sample. This subject has no clinical symptoms of AIDS and its absolute CD4 count was 533 cells per cubic millimeter. In support of this result, Katzentein *et al*. (1996) reported that CD4 cell counts were not significant predictors of clinical outcome even though the syncytium-inducing phenotype were significantly associated with an increasing risk of progression to AIDS or death. The dominance of syncytium inducing phenotype in this result is in agreement with the work by Bjorndal *et al*. (1999) and Abebe *et al*. (1999) on biological phenotype and co-receptor usage of HIV-1 subtype C isolates obtained from Ethiopian AIDS patients, revealing scarcity of CXCR4 usage of subtype C isolates that distinguishes subtype C from other HIV-1 subtypes.

The premise that phylogenetic structure arose by independent migration events and subsequent viral proliferation is consistent with the biology of memory cells homing to specific tissue compartments (Mackay *et al*., 1992a; Mackay *et al*., 1992b) which means that there is repeated migration of viruses into the different compartments followed by localized expansion of these populations. Tissue specific variant populations and distributional differences of variants in the genital mucosa and PBMCs suggest that selective pressures on the two virus pools are different or nonsynchronous. Local chemokines and HIV-1 specific immunoglobulins production in

female genital tract secretions are reported to be different to those in blood (Belec *et al*., 1995; Iversen *et al*., 1998). This indicates that there is also immunological compartmentalization in peripheral blood and vaginal mucosa.

Elucidation of the mechanism for viral evolution within a human body is of a particular importance in understanding the evolutionary mechanism of viruses and for developing effective strategies for antiviral therapy and vaccination. The phylogeny through genetic distance of the sequences in the present study were compared with sequences of previous years and were able to show the viral evolutionary pattern over time. The type of selective pressure operating on the virus population dynamics in the two compartments at a given point of time in the present study were also shown.

Studies of viral evolution frequently involve portion of the envelope gene, which encodes the envelope glycoprotein of the HIV-1, gp120 because it contains determinants of cell tropism (Shioda *et al*., 1991). HIV is characterized by enormous genetic flexibility, which gives rise to drug resistance, escape from immune responses and failures of the immune responses. The diversity of viral quasispecies, however, is shaped by a combination of mutations and selection forces. The main selective forces that have been proposed to derive HIV diversity are the immune response, cell tropism and random activation of infected cells (Bonheoeffer *et al*., 1995). According to Holmes and Zanotto (1998), Bonhoeffer *et al*. (1995) and Collins *et al*. (2000), the ratio of 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 status of an individual because the ds/dn < 1 in an individual signifies strong immune response and ds/dn > 1 in individuals signifies weak immune response. In this context, the absolute CD4 count in blood plasma was

negatively correlated but not significant with that of the genetic distance rate ratio ds/dn of the two compartments, which may indicate that the magnitude of the nonsynonymous substitution increases when the individual infected has strong immune pressure and the fittest virus population is positively selected and proliferates to dominate the environment in the host.

Synonymous and nonsynonymous genetic distances at synonymous and nonsynonymous amino acid sites support the clade C HIV-1 evolutionary divergence study between cervicovaginal lavage and blood plasma. Nucleotide substitution analysis in matched plasma and cervicovaginal lavage samples of an individual indicated that there was positive selection pressure in almost all women. The mean synonymous and nonsynonymous distances between the two compartments is significantly different, regardless of the two subclusters of subtype C, which means there is a drive that pushes the virus in the two compartments to differentiate. This indicates that there is also the compartmentalization of the immune pressure as one of the selective forces for the HIV evolution in patients.

In the settings of HIV-1 infection, cellular activation in the mucosa is enhanced by the stimulation of proinflammatory cytokines (Kotler *et al*., 1993). The majority of the mucosal CD4 lymphocytes express high levels of the CCR5 and CXCR4 co-receptors in the presence of proinflammatory cytokines and chemoattractant, resulting in infiltration of immunocompetent cells. This increases the chance of infection of cells by the HIV-1 virus (Olsson *et al*., 2000). The presence of certain genital tract infections is associated with substantial increases in the level of HIV-1 in genital secretions (Wahl *et al*., 1999). Nearly all subjects in this study presented with sexually transmitted infections. According to Wright *et al*. (2001), cervical inflammation and ulceration are associated with local HIV-1 expression increasing up to 10,000-fold, thereby increasing the amount of HIV-1 shedding into genital secretions. This may explain why sexually

transmitted diseases are important risk factors for HIV transmission. Increased viral replication due to cellular activation in these compartments may result in higher mucosal HIV-1 load and, in the face of selective pressure through immune response or antiretroviral medications to differential evolution.

The synonymous distance in subcluster C' of Ethiopian virus sequences (including the present ones) was less compared with the synonymous distance in subcluster C, while the nonsynonymous distances are relatively comparable. Even though C' was shown to be dominating (Rinke de Wit *et al*., 2002) and for present the study the samples were collected in 2001, the mean of synonymous and nonsynonymous nucleotide variation among the sequences of the C group is higher than in the C' group even if it is not significant. This could be explained as either this subcluster C' viruses are the results of intra-patient variation so that C' evolves from C, or they are introduced late. However, both of the above explanations are ruled out because a study on the years of introduction of these viruses in Ethiopia estimated the years 1982 and 1983 to be the year of introduction of subtype C' and C viruses respectively (Abebe *et al*., 2001a; Abebe *et al*., 2001b). The greater replication rate of group C viruses cannot be an explanation too because if this group of viruses had a higher replication rate, they would have replaced the C' virus which is not the case at present. In contrast, the viral load of subcluster C' was found to be higher than C which may lead to the conclusion that it is selectively advantaged for transmission and then replication. So, the reason for the presence of larger synonymous genetic distances in subcluster C than C' should be verified.

Virus in patients diverged with time, more consistently at its non-synonymous sites (Shankarappa *et al*., 1998). Synonymous and nonsynonymous genetic distances increased over years from 1988 to 2001, may be due to the parallel accumulation of synonymous and nonsynonymous

substitutions even though nonsynonymous substitutions overweigh within the given time. It indicates that the virus in the Ethiopian HIV-1 epidemic is in continuous evolution. This evolution can be caused by a strong positive selection as a consequence of continuous immune activation in the study population due to intestinal parasitic infections (Kassu *et al*., 2001). It is found that a greater systemic HIV-1 heterogeneity exists in HIV/TB patients than in HIV patients. Phylogenetic and sequence analysis of the C2V3 region of individual *env* pair clones of HIV and HIV/TB patients indicated more diverse clustering of the quasispecies and greater mutation frequencies in the TB patients. In addition, HIV/TB patients tended to have higher nonsynonymous (dn)/synonymous (ds) values than HIV patients, suggesting a stronger immune selection pressure in the former (Collins *et al*., 2001).

The limitations of this study were absence of purification of the PCR products for sequencing and not having performed sperm test for those subjects under the study. This is because protein materials and fragmented PCR products in the PCR may affect the migration pattern of the actual fragmented products for sequencing in unpurified PCR product. The sperm test was important because the study subjects might have mated in the prohibited period to meet the criterion of the study. If the subject(s) probably had sex with HIV positive individuals, there may be washing of the sperm and the virus in it. So there may be amplification of the virus in the sperm, which may affect the comparison of the virus genetic variability in the two compartments. For example, in subject STD007, the subtype found in cervicovaginal lavage is B; in contrast, it is found to be subtype C in plasma. This discordance may be due to the above limitation or there may be multiple infections in this individual where subtype B population is dominant in cervicovaginal lavage and subtype C population in blood. This means that the two subtypes may be found in the two compartments. However, this subtype difference may be from sampling of the viral

populations during amplification for sequencing. But, the absence of contamination in other pairs other than STD007 is confirmed through the phylogenetic analysis as shown in Figure-2. However, the problem of viral contamination by viruses from the sperm is not occurred as it is proven in the phylogenetic analysis where viral sequences from blood plasma and cervicovaginal lavage of an individual cluster together with large bootstrap value.

6. Conclusions and Recommendations

This study showed that the Ethiopian subtype C subcluster C' is dominating the subcluster C. Both subclusters are in continuous evolution may be due to the presence of continuous immune activation in our population. The increasing of HIV-1 subtype C genetic heterogeneity within these isolates may be an indication for the role of STDs in immune activation and further selection of fittest isolates in vaginal mucosa and contributing to the systemic heterogeneity because there is no complete barrier between different compartments within a body (partial compartmentalization). Effective treatment of STDs are found to reduce the viral shedding and further helps in reduction of viral transmission through heterosexual intercourse. This result also showed that subjects infected with subcluster C' virus had higher viral load compared to those infected with the subcluster C virus even though the sample size was small. This finding should prompt further investigation with a sample size large enough for conclusive results.

The virus populations in blood plasma and vaginal secretions were found significantly different. The presence of divergence in peripheral blood and vaginal mucosa must be considered in the view of the application of chemotherapies and future vaccines. This result also implies that consideration of both subtype C subclusters C and C' in the Ethiopian case is necessary for development of vaccine(s) or therapies. Of course, there is need of whole genome sequence analysis to see whether there is difference between the two subclusters and to come with protective vaccine. Adding of fifty sequences to Ethiopian sequences is another achievement and these sequences will be submitted to the genebank with the specified accession numbers as far as the publication is submitted and published in an international journal.

Basic knowledge on genetic variability of the HIV-1 virus in different body compartments is indispensable as it can affect infants exposed to the HIV-1 variants during birth and men during heterosexual intercourse. Therefore, the existence of genetically varied HIV-1 variants in cervicovaginal fluid stresses the importance of considering viruses in both blood and cervicovaginal fluids in evaluating the impact of therapeutic treatment regimens and candidate vaccines.

7. References

Abebe, A., Demissie, D., Goudsmit, J., Brouwer, M., Kuiken, C. L., Pollakis, G., Schuitemake, H., Fontanet, A. L. and Rinke de Wit, T. F. (1999). HIV Subtype C Syncytium and Non-Syncytium Ethiopian Patients with AIDS. *AIDS*. **13**: 1305-1311.

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

Abebe, A., Lukashov, V.V., Pollakis, G., Kliphuis, A., Fontanet, A. L., Goudsmit, J. and Rinke de Wit, T. F. (2001a). Year of HIV-1 Subtype C Introduction in Ethiopia: Estimated on the Basis of Genetic Analysis of Early Samples and Subsequent Virus Diversification. *AIDS*. **15**: 1-7.

Abebe, A., Lukashov, V.V., Rinke de Wit, T. F., Fisseha, B., Tegbaru, B., Kliphuis, A., Tesfaye, G., Negassa, H., Fontanet, A. L., Goudsmit, J. and Pollakis, G. (2001b). Timing of the Introduction into Ethiopia of the subcluster C' of the HIV-1 Subtype C. *AIDS Res. Hum. Retroviruses*. **17**: 657-661.

Abebe, A., pollakis, G., Fontanet, A. L., Fisseha, B., Tegbaru, B., Kliphuis, A., Tesfaye, G., Negassa, H., Cornelissen, M., Goudsmit, J. and Rinke de Wit, T. F. (2000). Identification of a Genetic Sub-Cluster of the HIV-1 Subtype C (C') Widespread in Ethiopia. *AIDS Res. Hum. Retrovir*. **16** : 1909-1914.

Ayehunie, S., Johansson, B., Salminen, M., Leinikki, P., Sonnerborg, A., Zewdie, D. W., 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. W., Britton, S. and Strannegard, O.

(1993). Sequence Analysis of the Selected Regions of the *Env* (V3 Loop and Gp41) and Gag (p17) Reading Frames of Ethiopian Human Immunodeficiency Virus Type-1 Strains. *Arch Virol*. **128**: 229-239.

Ball, J. K., Holmes, E. C., Whitwell, H. and Desselberger, U. (1994).Genomic Variation of Human Immunodeficiency Virus Type 1 (HIV-1): Molecular Analyses of HIV-1 in Sequential Blood Samples and Various Organs Obtained at Autopsy. *J. Gen. Virol*. **75**: 67-69.

Barre-Sinoussi, F., Chemann, J. C., Rey, F. (1983). Isolation of a T-lymphotrophic Retrovirus from a Patient at Risk for Acquired Immunodeficiency Syndrome (AIDS). *Science*. **220**: 868-871.

Belec, L., Dupre, T., Prazuck, T., Tevi-Benissan, C., Kanga, J., Pathey, O., Lu, X. and Pillot, J. (1995). Cervicovaginal Overproduction of the Specific IgG to Human Immunodeficiency Virus (HIV) Contrasts with Normal or Impaired IgG Local Response in HIV infection. *J. Infect. Dis*. **172**: 691-697.

Bentwich, Z., Kalinkovich, A. and Weisman, Z. (1995). Immune Activation is a Dominant Factor In the Pathogenesis of African AIDS. *Immunol. Today*. **11**: 187-191.

Berger, E. A. (1997). HIV Entry and Tropism: the Chemokine Receptor Connection. *AIDS*. **11**(Suppl. A): S3-16.

Birdsall, H.H., Siwak, E. B., Trial, J., Rodriguez-Barradas, M., White Jr. A. C., Wietgrefe, S. and Rossen, R. D. (2002). Transendothelial Migration of Leucocytes Carrying Infectious HIV-1: an Indicator of Adverse Prognosis. *AIDS*. **16**: 5-12.

Bjorndal, A., Sonnerborg, A., Tschrning, C., Albert, J. and Fenyo, E. M. (1999). Phenotypic Characterstics of Human Immunodeficiency Virus Type 1 Subtype C Isolates of Ethiopian AIDS Patients. *AIDS Res. Hum. Retroviruses*. **15**: 647-653.

Bleul, C. C., Wu, L., Hoxie, J. A., Springer, T. A. and Mackay, C. R. (1997). The HIV Co-Receptor Usage between T-Lymphocytes. *Proc. Natl Acad. Sci. USA*. **94**: 1925-1930.

Bobkove, A., Kazennova, E., Selimova, L., Ladnaya, N., Kravcchenko, A., Foley, B., Morriso, I., Pokrovsky, V., Weber, J. and Cheingsong-Popov, R. (1998). HIV-1 *Gag* D/G *Env* Recombinants in Russia. *AIDS Res. Hum. Retroviruses*. **14**: 1597-1599.

Bonhoeffer, S., Holmes, E. C. and Nowak, M. A. (1995). Causes of HIV Diversity. *Nature* (Letter). **376**: 125.

Boom, R., Sol, C. J. A., Salimans, M. M. M., Jansen, C. L., Wertheim-Van Dillen, P. M. E. and Van de Noordaa, J. (1990). A Rapid and Simple Method for Purification of Nucleic Acid. *J.Clin. Microbiol*. **28**: 495-503.

Borrow, P., Lewicki, H., Wei, X., Horwitz, M. S., Peffer, N., Meyers, H., Nelson, J. A., Gairin, J. E., Hahn, B. H., Oldstone, M. B. A. and Shaw G. M. (1997). Antiviral Pressure Exerted by HIV-1 exerted by HIV-1-Specific Cytotoxic T Lymphocytes (CTLs) during Primary Infection Demonstrated by Rapid Selection of CTL Escape Virus. *Nat. Med*. **3**: 205-217.

Bozzette, S. A., Spector, S.A. and Richman, S. (1993). A Cross-Sectional Analysis of Persons with Syncytium and Non-Syncytium Inducing Human Immunodeficiency Virus. *J. Infect. Disease*. **168**:1374-1379.

Carr, J., Salminen, M., Albert, J., Sanders-Buell, E., Gotte, D., Birx, D. and McCutchan, F. (1998). Full Genome Sequences of HIV-1 Subtype G and A/G Inter-Subtype Recombinants. *Virology*. **247**: 22-31.

Cheng-Mayer, C., Seto, D., Tateno, M. and Levy, J. A. (1988). Biological Features of HIV-1 that Correlate with Virulence in the Host. *Science*. **240**: 80-82.

Chesebro, B., Wehrly, K., Nishio, J. and Perryman, S. (1992). Macrophage Tropic Human

Immunodeficiency Virus Isolates from Different Patients Exhibit Unusual V3 Envelope Sequence Homogeneity in Comparison with T-cell-Trophic Isolates: Definition of Critical Amino Acids Involved in Cell Tropism. *J. Virol*. **66**: 6547-6554.

Chowdhury, I. H., Potash, M. J. and Volsky, D. J. (1995). Redefinition of Tropism of Common Macrophage-Tropic Human Immunodeficiency Virus Type 1. *AIDS Res. Hum. Retroviruses.* **11**: 1467-1471.

Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M-A., Santos-Ferreira, M. O., Laurent, A. G., Dauguet, C., Katlama, C., Rouzioux, C., Kalatzmann, D., Champalimaud, J. L. and Montagnier, L. (1986). Isolation of a New Human Retrovirus from West African Patients with AIDS. *Science*. **233**: 343-346.

Coffin, J. M. (1986b). Genetic Variation in AIDS Viruses. *Cell*. **46**: 1-4.

Coffin, J. M. (1990). Retroviridae and Their Replication. **In**: *Fields Virology*. Pp. 1435-1500. (Fields, B. N., Knipe, D. M., Chanock, R. M., Hirsch, M. S. Melinick, J. L., Monta, T. P., Roizman, B. eds.). Raven Press.

Coffin, J., Hause, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teichi, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P. and Weiss, R. (1986a). Human Immunodeficiency Viruses. *Science*. **232**: 697.

Cohen, M. S. (1998). Sexually Transmitted Disease Enhance HIV Transmission-No Longer Hypothesis. *Lancet*. **251**(Suppl. 13): S5-S7.

Collins, K. R., Mayanja-Kizza, H., Sullivan, B. A., Quinones-Mateu, M. E., Toossi, Z. and Arts, E. J. (2000). Greater Diversity of HIV-1 Quasispecies in HIV-Infected Individuals with Active Tuberculosis. *J. Acquir. Immune Defic. Syndromes*. **24**: 408-417.

Collman, R. G. and and Yi, Y. (1999). Cofactors for Human Immunodeficiency Virus Entry into Primary Macrophages. *J. Infect. Dis*. **179** (Suppl.3): S422-426.

Cornelissen, M., van Den Burg, R., Goudsmit, J. and Zorgdrager, F. (2000). Spread of Distinct Human Immunodeficiency Virus Type 1 AG Recombinant Lineages in Africa. *J Gen Virol*. **81**: 515-523.

Cornelissen, M., Kampinga, G., Zorgdrager, F., Goudsmit, J. and The UNAIDS Network for HIV Isolation and Characterization. (1996). Human Immunodeficiency Virus Type 1 Subtypes Defined by the *Env* Show High Frequency of Recombinant *Gag* Genes. *J. Virol*. **70**: 8209-8212.

De Jong, J., de Ronde, A., Keulen, W., Tersmette, M. and Goudsmit, J. (1992a). Minimal Requirements for the Human Immunodeficiency Virus Type 1 V3 Domain to Support the Syncytium-Inducing Phenotype: Analysis by Single Amino acid Substitution. *J. Virol*. **66**: 6777-6780.

De Jong, J., Goudsmit, J., Keulen, W., Klaver, B., Krone, W., Tersmette, M. and de Ronde, A. (1992b). Human Immunodeficiency Virus Type 1 Clones Chimeric for the Envelope V3 domain differ in Syncytium Formation and Replication Capacity. *J. Virol*. **66**: 757-765.

De Wolfs, T., de Jong, J., van den Berg, H., Tijnagel, J., Krone, W. and Goudsmit, J. (1990). Evolution of Sequences Encoding the Principal Neutralization Epitope of HIV-1 is Host Dependent, Rapid, and Continuous. *Proc. Natl. Acad. Sci. USA*. **87**: 9938-9942.

De wolf, F., Hogervorst, E., Goudsmit, J., Fenyo, E. M., Rubsamen-Waigmann, H., Holmes, H., Galvao-Castro, B., Karita, E., Wasi, C., Sempala, S. D., *et al*. (1994). Synsytium-Inducing and Nonsynsytium-Inducing Capacity of Human Immunodeficiency Virus Type 1 Subtypes Other Than B: Phenotypic and Genotypic Characteristics. WHO Network for HIV Isolation and Characterization. *AIDS Res. Hum. Retroviruses*. **10**: 1387-1400.

De Wolf, T. F., Zwart, G., Bakker, M., Kuiken, C. L., Goudsmit, J. (1991). Naturally Occurring Mutations within HIV-1 V3 Genomic RNA Lead to Antigenic Variation Dependent on a Single Amino Acid Substitution. *Virology*. **185**: 195-205.

Diaz, R., Sabino, C., Mayer, A., Mosely, J., Busch, M., and The Transfusion Safety Study Group. (1995). Dual HIV-1 Infection and Recombination in Dually Exposed Transfusion Recipient. *J. Virol*. **69**: 3273-3271.

Dornaddula, G., Zhang, H., VanUitert, B., Stern, J., Livornese, Jr. L., Ingerman, M. J., Witek, J., Kedanis, R. J., Natkin, J., DeSimone, J. and Pomerantz, R. J. (1999). Residual HIV-1 RNA in the Blood Plasma of Patients Taking Suppressive Highly Active Antiretroviral Therapy. *JAMA*. **282**: 1627-1632.

Ellerbrock, T. V., Lennox, J. L., Clancy, K. A., Schinazi, R. F., Wright, T. C., Pratt-Palmore, M., Evans-Strickfaden, T., Schnel, C., Pal, R., Corley, L. J., Parrish-Kohler, E. E., Bush, T. J., tatti. K. and Hart, C. E. (2000). Cellular Replication of the Human Immunodeficiency Virus Type 1 Occurs in Vaginal Secretions. *J. Infect. Dis.* **184**: 26-36.

Engelstad, M. (1996). Genetic Diversity of Env V3-Loop Sequences in HIV Type 1 Infected Homosexuals, Heterosexuals, Intravenous Drug Users, and Hemophilia Patients in Norway: High Prevalence of Subtype B and Detection of the Subtype C. *AIDS Res Hum Retroviruses*. **12**: 1733-1738.

Epstein, L. G., Kuiken, C., Blumberg, B. M., Hartman, S., Sharer, L. R., Clement, M. and Goudsmit, J. (1991). HIV-1 V3 domain Variation in Brain and Spleen of Children with AIDS: Tissue Specific Evolution within Host Determined Quasispecies. *Virology*. **180**: 583-590.

Erron, J.J., Vernazza, P.L., Johnston, D. M., Seiller-Moiseiwitch, F., Alco, T. M., Ficus, S. A. and Cohen, M. S. (1998). Resistance of HIV-1 to Antiretroviral Agents in Blood and Seminal Plasma: Implication of Transmission. *AIDS*. **12**: F181-189.

Fauci, A. S. (1993). Multifactorial Nature of Human Immunodeficiency Virus Disease: Implication for Therapy. *Science*. **262**: 1011-1018.

Felsenstein J. http://evolution.Genetics.Washington.edu/phylip.html. PHYLIP: phylogeny inference package. Version 3.52. Seatle: University of Washington, 1996.

Fouchier, R. A. M., Groenink, M. and Kootstra, N. A., Termette, M., Huisman, H. G., Miedema, F. and Schuitemaker, H. (1992). Phenotype Associated Sequence Variation in the Third Variable Domain of the Human Immunodeficiency Virus Type 1 Gp120 Molecule. *J. Virol.* **66**: 3183-3187.

Freed, E. O. (1998). HIV-1 Gag Protein: Diverse Functions in the Viral Life Cycle. *Virology*. **251**: 1-15.

Gallo, R. C., Salahuddin, S. Z, Popovic, M. (1984). Frequent Detection and Isolation of Cythopathic Retroviruses (HTLV-III) from Patients with AIDS and at Risk for AIDS. *Science*. **224**: 500-503.

Gallo, R. C., Sarin, P. S., Gelmann, E. P., Robert-Guroff, M., Richardson, E., Kalyanaramanam, V. S., Mannsidhu, G. D., Stahl, R. E., Zolla-Pazner, S., Leibowitch, J. and Popovic, M. (1983). Isolation of Human T Cell Leukemia Virus in Acquired Immunodeficiency Syndrome (AIDS). *Science*. **224**: 500-503.

Gilks, C. F. (1993). The Clinical Challenge of the HIV Epidemic in the Developing World. *Lancet*. **342**: 1037-1039.

Gouldston, C., McFarland, W. and Katzenstein, D. (1998). Human Immunodeficiency Virus Type 1 RNA Shedding in the Female Genital Tract. . *J. Infect. Dis.* **177**: 1100-1103.

Comment [M12]:

Gottlieb, M. S., Schroff, R., Schanker, H. M., Weissman, J. D., Fan, P. T., Wolf, R. A. and Saxon, A. (1981). Pneumonocystis carinni Pneumonia and Mucosal Candidiasis in the Previously Healthy Homosexual Male. Evidence of a New Acquired Cellular Immunodeficiency. *New Engl. J. Med.* **305**: 1425.

Gruber, M. F., Hewlett, I. K., Simms, T., Vujcic, L., Manischewitz, J. and Golding, H. (1992). Study of Viral Replication in HIV-1-Infected CEM T-cell Subclones Which are Reduced in Their Ability to Form Syncytia. *AIDS Res. Hum. Retroviruses*. **8**: 1139-1146.

Haseltine, W. A. (1988). Replication and Pathogenesis of the AIDS Virus. *J. Acquir. Immune Defic. Syndr*. **1**: 217-240.

Hills, D. M. (1997). Phylogenetic Analysis. *Current Biology*. **7**: R129-R131.

Hills, D.M. and Bull, J. J. (1993). An Empirical Test for Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Systematic Biology*. **42**: 182-192.

Hitchcock, P. and Fransen, L. (1999). Preventing HIV infection: Lesson from Mwanza and Rakai. *Lancet*. **353**: 513-515.

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

Hughes, E. S., Bell, J. E., and Simmonds, P. (1997). Investigations of the Dynamics of the Spread of Human Immunodeficiency Virus to Brain and Other Tissues by Evolutionary Analysis of Sequences from the p17 and *Env* Genes. *J. Virol.* **71**:1272-1280.

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

Itescu, S., Simonelli, P. F., Winchester, R. J. and Ginsberg, H. S. (1994). Human Immunodeficiency Virus Type 1 Strains in the Lungs of Infected Individuals Evolve

Independently from Those in Peripheral Blood and Are Highly Conserved in the C-Terminal Region of the Envelope V3 Loop. *Proc. Natl. Acad. Sci. USA*. **91** : 11378-11382.

Iversen, A. K. N., Fugger, L., Eugen-Olsen, J., Balslev, U., Jensen, T., Wahl, S., Gerstoft, J., Mullins, J. I. and Skinhoj, P. (1998). Cervical Human Immunodeficiency Virus Type 1 Shedding is Associated with Genital β-Chemokine Secretions. *J. Infect. Dis*. **178**: 1334-1342.

Kassu, A., Tsegaye, A., Petros, B., Wolday, D., Hailu, E., Tilahun, T., Hailu, B., Roos, M.T.L., Fontanet, A. L., Hamann, D. and Rinke de Wit, T. F. (2001). Distribution of Lymphocyte Subsets in Healthy Human Immunodeficiency Virus-Negative Adult Ethiopians from Two Geographic Locales. *Clin. Diagn. Lab. Immunol*. **8**: 1171-1176.

Katzenstein, D.A., Hammer, S. M., Hughes, M. D., Gundacker, H., Ackson, J. B., Iscus, S., Asheed, S., Lbeik, T., Eichman, R., Apour, A., Merigan, T., Hirschm, S. and For AIDS Clinical Rials Group Tudy 175 Virology Tudy EAM. (1996). The relation of virologic and immunologic markers to clinical outcomes after nucleoside therapy in HIV-infected adults with 200 to 500 CD4 cells per cubic millimeter. *N. Engl. J. Med.* **335**: 1091-1098.

Kiessling, A. A., Fitzgerald, L. M., Zhang, D., Chhay, H., Brettler, D., Eyre, R. C., Steinberg, J., McGowan, K. and Byrn R. A. (1998). Human Immunodeficiency Virus in Semen Arises from a Genetically Distinct Virus Reservoir. *AIDS Res. Hum. Retroviruses*. **14**: S33-S41.

Koch, N., Ndihokubwayo, Yahi, N., Tourres, C., Fantani, J. and Tamalet, C. (2001). Genetic Analysis of HIV type 1 Strains in Bujumbura (Burundi): Predominance of Subtype C Variant. *AIDS Res Hum Retroviruses*. **17**: 269-273.

Korber, B. T. M., Kunstman, K. J., Patterson, B. K., Furtado, M., McEvilly, M. M., Levy, R. and Wolinsky, S. M. (1994). Genetic Difference between Blood- and Brain-Derived Viral Sequences from Human Immunodeficiency Virus Type-1-Infected Patients: Evidence of Conserved Elements in the V3 Region of the Envelope Protein of the Brain Derived Sequences. *J. Virol*. **68**: 7467-7481.

Korber B and Myers G. (1992) Signature pattern analysis: a method of assessing viral sequence relatedness. *AIDS. Res. Hum. Retroviruses.* **8**: 1549-1560.

Kotler, D. P., Reka, S. and Clayton. F. (1993). Intestinal Mucosal Inflammation Associated with Human Immunodeficiency Virus Infection. *Dig. Dis. Sci*. **38**: 1119-1127.

Kovacs, A., Wasserman, S. S., Burns, D., Wright, D. J., Cohn, J., Landay, A., Weber, K., Cohen, M., Levine, A., Minkoff, H., Miotti, P., Palefsky, J., Young, M., Reichelderfer and the DATRI and WIHS Study Groups. (2001). Determinants of the HIV-1 Shedding in the Genital Tract of Women. *Lancet*. **358**: 1593-1601.

Kuiken, C. L., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J., Mullins., J., Sodroski, J., Wolinsky, S. and Korber, B. (2000). HIV Sequence Compendium 2000. Los Alamos National Laboratory, Los Alamos, New Mexico 87545 U.S.A. Pp. 35-43.

Kuiken, C. L., Goudsmit, J., Weiller, G. F., Armstrong, J. S., Hartman, S., Portegies, P., Dicker, J. and Cornelissen, M. (1995). Differences in Human Immunodeficiency Virus Type 1 V3 Sequences from Patients with and without AIDS Dementia Complex. *J. Gen. Virol*. **76**: 175-180.

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.

Lawn, S. D., Subbarao, S., Wright, Jr. T. C., Evans-Strickfaden, T., Ellerbrock, T. V., Lennox, J.

L., Butera, S. T. and Hart, C. E. (2000). Correlation between Human Immunodeficiency Virus Type 1 RNA Levels in the Female Genital Tract and Immune Activation Associated with Ulceration of the Cervix. *J. Infect. Dis*. **181**: 1950-1956.

Learn, G. H., Korber, B. T. M., Foley. B., Hahn, B. H., Wolinsky, S. M. and Mullins, J. I. (1996). Maintaining the Integrity of Human Immunodeficency Virus Sequence Databases. *J. Virol*. **70**: 5720-5730.

Lemp, G. F., Payne, S. F., Rutherford, G. W., Hessol, N. A., Winkelstein, W., Wley, J. A., Moss, A. R., Chaisson, R. E., Chen, D. T., Feigal, D. W., Thomas, P. A. and Werdegar, D. (1990). Projections of AIDS Morbidity and Mortality in San Francisco. *JAMA*. **263**: 1497-1501.

Lenz, J., Su, M., Mizrachi, Y., Burke, M. and Rubinotein, A. (2001). V3 Variation in HIV Seropositive Patients Receiving a V3 Targeted Vaccine. *AIDS*. **15**: 577-581.

Lester, F.T., Ayehunie, S. and Zewdie, D. (1988). Acquired Immunodeficiency Syndrome: Seven Cases in Addis Ababa Hospital. *Ethiop. Medical J.* **26**: 139-145.

Levy, J. A., Hoffman, A. D., Kramer, S. M., Landis, J. A., Shimabukuro, J. M. and Oschiro, L. S. (1984). Isolation of Lymphocytopathic Retroviruses fron San Francisco Patients with AIDS. Science. 225: 840.

Lou, C. C., Tian, C., Hu, D. J. (1995). HIV Subtype C in China. *Lancet* (Letter). **345**: 1051-1052.

Louwagie, J., Jaansens, W., Mascola, J., Heyndrich, P., Van Der Groen, G., McCutchan, E., Burke, S. (1995). Genetic Diversity of the Envelope Glycoprotein from the HIV-1 Isolates of African Origin. *J. Virol*. **69**: 263-271.

Mackay, C. R., Marston, W. and Dudler, L. (1992a). Altered Patterns of T Cell Migration through Lymphnodes and Skin Following Antigen Challenge. *Eur. J. Immunol*. **22**: 2205-2210.

Mackay, C. R., Marston, W. L, Dudler, L., Spertini, O., Tedder, T. F. and Hein, W. R. (1992b). Tissue-Specific Migration Pathways by Phenotypically Distinct Subpopulations of Memory T Cells. *Eur. J. Immunol*. **22**: 2205-2210.

Mansky, L. (1998). Retrovirus Mutation Rate and Their Role in Genetic Variatian. *J. Gen. Virol*. **79**: 1337-1345.

Marras, D., Bruggeman, L. A., Gao, F., Tanji, N., Mansukhani, M. M., Cara, A., Ross, M. D., Gusella, G. L., Benson, G., D'Agati, V. D., Hahn, B. H. and Klotman, M. (2002). Replication and Compartmentalization of HIV-1 in Kidney Epithelium of Patients with HIV-Associated Nephropathy. *Nature Med*. **8**: 522-526.

McClure, M. and Dalgleish, A. (1998). Human Immunodeficiency Viruses. **In**: *Encyclopedia of Immunology*. Second Edition. (Delves, P. J. and Roitt, I.M. eds.) Pp. 1330-1339. Academic Press: San Diego, London, Boston, New York, Sydney, Tokyo, and Toronto.

McCutchan, F., Peeters, M., Takebe, Y., Kalish, M., Goudsmit, J., Mullins, J., Korber, B. and Burke, D. (2001). Genetic Variability of HIV-1. **In**: Approaches to the Development of Broadly Protective HIV-1/AIDS Vaccines: Challenges Posed by Genetic, Biological and Antigenic Variability of HIV-1. WHO-UNAIDS Report. *AIDS*. **15**: W1-W25.

McMichael, A.J. and Philips, R. E. (1997). Escape of Human Immunodeficiency Virus from Immune Control. *Ann. Rev. Immunol*. **15**: 271-296.

Ministry of Health (MOH), (2002). AIDS in Ethiopia, Disease Prevention and Control Department. Fourth Edition.

Moss, A. R. and Bacchetti, P. (1989). Natural History of HIV Infection. *AIDS*. **3**: 55-61.

Mostad, S. B., Jackson, S., Overbaugh, J., Reilly, M., Chohan, B., Mandaliya, K., Nyange, P., Ndinya-Achola, J., Bwayo, J. J. and Kreiss, J. K. (1998). Cervical and Vaginal Shedding

of Human Immunodeficiency Virus Type 1-Infected Cells Throughout the Menstrual Cycle. *J Infect Dis*. **178**: 983-991.

Mulligan, M. T., Kumar, H. H., Owens, R. J., Ritter, G. D., Hahn, B. H. and Compans, R. W. (1990). The *Env* Protein of an Infectious Noncytopathic HIV-2 is Deficient in Syncytium Formation. *AIDS Res. Hum. Retroviruses*. **6**: 707-720.

Murphy, S., Sylwester, A., Kennedy, R. and Soie, D. (1995). Phagocytosis of Individual $CD4^+$ T Cell by HIV-induced T Cell Syncytia. *AIDS Res. Hum. Retroviruses*. **11**: 432-442.

Nei, M. and Gojobori, T. (1996). Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. *Mol. Bio. Evol*. **3**: 418-426

Nowak, M. A., Anderson, R. M., McLean, A. R., Wolfs, T. F. W., Goudsmit, J. and May, R. M. (1991). Antigenic Diversity Thresholds and the Development of AIDS. *Science*. **254**: 963-969.

Olsson, J., Poles, M., Spetz, A. L., Elliott, J., Hultin, L., Giorgi, J., Andersson, J. and Anton, P. (2000). Human Immunodeficiency Virus Type 1 Infection Is Associated with Significant Mucosal inflammation Characterized by the Increased expression of CCR5, CXCR4, and Beta-Chemokines. *J. Infect. Dis*. **182**: 1625-1635.

Overbaugh, J., Kreiss, J., Poss, M., Lewis, P., Mostad, S., John, G., Nduati, R., Mbouri-Ngacha, D., Martin, H. Jr., Richardson, B., Jackson, S. Neilson, J., Long, E. M., Panteleeff, D., Welch, M., Rakwar, J., Jackson, D., Chohan, B., Laureys, L., Mandaliya, K., Ndinya-Achola, J. and Bwayo, J. (1999). Studies of Human Immunodeficiency Virus Type 1 Mucosal Shedding and Transmission in Kenya. *J. Infect. Dis*. **179** (Suppl. 3): S401-404.

Overbaugh,J., Anderson, R. J., Ndinya-Achola, J. O. and Kreiss, J. K. (1996). Distinct but Related Human Immunodeficiency Virus Type 1 Variant Populations in Genital Secretions and Blood. *AIDS Res. Hum. Retroviruses*. **12**: 107-115.

Pang, S., Vinters, H., Akashi, T., O, Briens, W. A. and Chen, I. S. Y. (1991). HIV-1 *Env* Sequence Variation in Brain Tissue of Patients with AIDS Related Neurologic Disease. *AIDS*. **4**: 1082-1292.

Pantaleo, G., Graziosi, C. and Demarest, J. (1993). HIV Infection is Active and Progressive in Lymphoid Tissue during the Clinically Latent Stage of Disease. *Nature*. **362**: 355-358.

Peeters, M. and Delaporte, E. (1999). Genetic Diversity of HIV Infection Worldwide and Its Consequences. *Med Trop*. **59**: 449-55.

Peeters, M., Esu-Williams, E., Vergne, L., Montavon, C., Mukanga-Kabeya, C., Harry, J., Ibironke, A., Lesage, D., Patrel, D. and Delaporte, E. (2000). Predominance of Subtype A and B HIV Type 1 in Nigeria, With Geographgical Differences in Their Distribution. *AIDS Res. Hum. Retroviruses*. **16**: 315-325.

Peeters, M., Gueye, A., Mboub, S., Bibollet-Ruche, F., Ekaza, E., Mulanga, C., Ouedrago, R., Gandji, R., Mpele, P., Dibanga, G., Koumare, B., Saidou, M., Esu-Williams, E., Lambart, J., Badombena, W., Luo, N., Haesevelde, M. V., and Delaporte, E. (1997). Geographical Distribution of the HIV-1 Group O Viruses in Africa. *AIDS*. **11**: 493-498.

Perelson, A. S., Neumann, A. U., Markowitz, M. (1996). HIV-1 Dynamics In Vivo: Virion Clearance Rate, Infected Cell Life Span, and Viral Generation Time. *Science*. **271**: 1582-1586.

Plummer, F. A. (1998). Heterosexual Transmission of Human Immunodeficiency Virus Type 1 (HIV): Interactions of Conventional Sexually Transmitted Diseases, Hormonal Contraception and HIV-1. *AIDS Res. Hum. Retroviruses*. **14** (Suppl.1): S5-S10.

Pollakis, G., Abebe, A., Kliphuis, A., Fisseha, B., Tegbaru, B., Tesfaye, G., Negassa, H., Mengistu, Y., Fontanet, A. L., Rinke de Wit, T. F., Cornelissen, M. and Goudsmit, J. (unpublished data). Intra-subtype Recombination of the Human Immunodeficiency Virus Type 1 Subtype C: Evidence for the Advantage of a Gp120 Envelope Subgroup.

Popper, S. J., Sarr, A. D., Travers, K. U., Gueye-Ndiaye, A., Mboup, S., Essex, M. E. and Kanki, P. (1999). Lower Human Immunodeficiency Virus (HIV) Type 2 Viral Load Reflects the Difference in Pathogenicity of HIV-1 and HIV-2. *J. Infect. Dis*. **180**: 1116-1121.

Poss, M., Martin, H., Kreiss, J., Nyange, P., Mandaliya, K., Chohan, B. and Overbaugh. (1995). Diversity in Virus Populations from Genital Mucosa and Peripheral Blood in Women Recently Infected with Human Immunodeficiency Virus. *J. Virol*. **69**: 8118-8122.

Power C., McArthur, J. C., Johnson, R. T., Griffin, D. E., Glass, J. D., Perryman, S. and Chesebro, B. (1994). Demented and Non-Demented Patients with AIDS Differ in Brain – Derived Human Immunodeficiency Virus Type 1 Envelope Sequences. *J. Virol*. **68**: 4643-4649.

Preston, B. D., Poiesz, B. J. and Loeb, L. A. (1988). Fidelity of HIV-1 Reverse Transcriptase. *Science*. **242**: 1168-1171.

Renjifo, B., Chaplin, B., Mwakagile, D., Shah, P., Vanberg, F., Msamanga, G., Hunter, D., Fawzi, W. and Essex, M. (1998). Epidemic Expansion of the HIV Type 1 Subtype C and Recombinant Genotypes in Tanzania. *AIDS Res. Hum. Retroviruses*. **14**: 635-638.

Rinke de Wit, R. F., Tsegaye, A., Wolday, D., Hailu, B., Aklilu, M., Sanders, E., Hagos, M., Kliphuis, A., Pollakis, G., Krol, A., Geskus, R., Miedema, F., Goudsmit, J., Coutinho, R. and Fontanet, A. L. (2002). Primary HIV-1 Subtype C Infection in Ethiopia. *J. Acquir. Immun. Defic. Symdromes.* **30**: 463-470.

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

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., McCutchan, F., Peeters, M., Salminen, M., Sharp, P. M., Wolinsky, S. and Korber, B. (2000). HIV-1 Nomenclature Proposal. *Science*. **298**: 55-56.

Roitt, I., Brostoff, J. and Male, D. (1996). *Immunology*. Fourth Edition. London, Barcelona, Boston, Chicago, Madrid, Sydney, Philadelphia, Tokyo, Toronto. Pp. 6-7.

Salminen, M. O., Johansson, B., Sonnerborg, A., Ayehunie, S., Gotte, D., Lenikki, P., Burke, D. S. and McCutchan, F. E. (1996). Full-Length Sequence of an Ethiopian Human Immunodeficiency Virus Type 1 (HIV-1) Isolates of the Genetic Subtype C. *AIDS Res. Hum. Retroviruses.* **12**: 1329-1339.

Shankarappa, R., Gupta, P., Learn, G. H., Rodrigo, A. G., Rinaldo, C. R., Gorry, M., Mullins, J. I., Nara, P. L. and Ehrlich, G. D. (19990. Evolution of Human Immunodeficiency Virus Type 1 Envelope Sequence in Infected Individuals with Differing Disease Progression Profiles. *Virology*. **241**: 251-259.

Sheehy, N., Desselberger, U., Whitwell, H. and Ball, J. K. (1996). Concurrent Evolution of Regions of the Envelope and Polymerase genes of Human Immunodeficency Virus Type 1 During Zidovudine (AZT) Therapy. *J. Gen. Virol*. **77**: 1071-1081.

Shepard, R. N., Schock, J., Robertson, K., Shugars, D. C., Dyer, J., Vernazza, P., Hall, C., Cohen, M. S. and Fiscus, S. A. (2000). Quantification of Human Immunodeficiency Virus Type 1 RNA in Different Biological Compartments. *J. Clin. Microbiol.* **38**: 1414-1418.

Sherefa, K., Johanssen, B., Salminen, M. and Sonnerborrg, A. (1998). Full Length Sequence of HIV-1 Subtype A Recombined with Subtype C in the *Env* Domain. *AIDS Res. Hum. Retroviruses*. **14**: 289-292.

Shioda, T., Levy, J. A., and Cheng-Mayer, C. (1991). Macrophage and T Cell-Line Tropisms of HIV-1 are Determined by Specific Regions of the Envelope Gp120 Gene. *Nature*. **349**: 167-169.

Shioda, T., Oka, S., Ida, S., Nokihara, K., Toriyoshi, H., Mori, S., Takebe, Y., Kimura, S., Shimada, K, Nagai, Y., et al. (1994). A Naturally Occurring Single Basic Amino Acid Substitutions in the V3 Region of the Human Immunodeficiency Virus Type 1 *Env* protein alters the cellular host Range and Antigenic Structure of the Virus. *J. Virol.* **68**: 7689-7696.

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.

Steuler, H., Sorech-Hagenlocher, B. and Widemann, B. (1992). Distinct Populations of Human Immunodeficiency Virus Type 1 in Blood and Cerebrospinal Fluid. *AIDS Res. Hum. Retroviruses.* **8**: 53-59.

Strauss, E. G., Strauss, J. H. and Levine, A. J. (1990). Virus Evolution. **In**: *Fields Virology*. V I. Second Edition. Pp. 167-190. (Fields, B. N., Knipe, D. M., Chanock, R. M., Hirsch, M. S., Melinick, J. L., Monta, T. P. and Roizman, B. eds.). Raven Press.

Sutthent, R., Sumragsurp, K., Wirachsilp, P., Chaisilwattana, P., Roongpisuthipong, A., Chaiyakul, P., Nooma, P., Honda, M. and Warachit, P. (2001). Diversity of HIV-1 Subtype E in Semen and Cervicovaginal Secretions. *J. Hum. Virol.* **20**: 260-268.

Taylor, J. M. G., Kuo, J. and Detels, R. (1991). Is the Incubation Period of AIDS Lengthening? *J. Acquir. Immune Defic. Syndr*. **4**: 69-75.

Tersmette, M., Gruters, R. A., de Wolf, F., REY de Goede, Lange, J. M. A., Schellekens P.A., Goudsmit, J., Huisman, J. G. and Miedema, F. (1989). Evidence for a Role of Virulent Human Immunodeficiency Virus (HIV) Variants in the Pathogenesis of Acquired Immunodeficiency Syndrome: Studies on Sequential HIV Isolates. *J. Virol*. **63**: 2118-2125.

The Global AIDS Policy Coalition. (1992). *AIDS in the World*. Pp. 267-277. (Mann, J. M., Tarantola, D. J. M. and Netter, T. W. eds.). Harvard University Press: Cambridge, Massachusetts, and London.

Tscherning, C., Alaeus, A., Fredriksson, R., Bjorndal, A., Deng, H., Littman, D. R., Fenyo, E. M., and Albert, J. (1998). Differences coreceptor Usage between Genetic Subtype of the HIV-1. *Virology*. **241**: 181-188.

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

Uvin, S. C. and Caliendo, A. M. (1997). Cervicovaginal Human Immunodeficiency Virus Secretion and Plasma Viral Load in Human Immunodeficiency Virus-Seropositive Women. *Obstet Gynecol*. **90**: 739-743.

Van Hermelen, J., van der Ryst, E. Loubser, A. S., York, D., Madurai, S., Lyons, S., Wood, R. and Williamson, C. (1999). A Predominantly HIV-1 Subtype C-Restricted Epidemic in South Africa Urban Population. *AIDS Res. Hum. Retroviruses*. **15**: 395-398.

Vartanian, J. and Wain-Hobson, S. (1994). The Molecular Biology of HIV-1. **In**: *AIDS In Africa*. Pp 21-46. (Essex, M., Mboup, S., Kanki, P. J. and Kalengayi, M. R. eds.). Raven Press Ltd., New York.

Vernazza, P.L., Dyer, J. R., Fiscus, S. A., Eron, J. J. and Cohen, M. S. (1997). HIV-1 Viral Load in Blood, Semen and Saliva (Letter; Comment). *AIDS*. **11**:1058-1059.

Wahl, S. M., Smith, P. D. and Janoff, E. N. (1999). Human Immunodeficiency Virus Type 1 Infection, Mucosal Immunity and Pathogenesis: Comments and Conference Summary. *J. Infect. Dis.* **179**(Suppl. 3): S397-400.

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. *J. infect. Dis.* **178**: 1507-1511.

Wang-Staal, F. (1990). Human Immunodeficiency Viruses and Their Replication. **In**: *Fields Virology*. V II. Second Edition. Pp. 1529-1543. (Fields, B. N., Knipe, D. M., Chanock, R. M., Hirsch, M. S., Melinick, J. L., Monta, T. P. and Roizman, B. eds.). Raven Press.

WHO/UNAIDS. (2002). AIDS Epidemic Update: *December 2002. http://www.unaids.org/*

Wolday, D., G/mariam, Z., Mohammed Z., Girma, M., Dorigo, W., Meles, H., Seme, W., Sanders, E., Geyid, A. and Mayaan, S. The Impact of Syndromic Treatment of Sexually Transmitted Diseases on Vertical Shedding of HIV-1. XIV International AIDS Conference 2002, Barcelona (Spain).

Wong, J. K., Igacio, C. C., Torriani, F., Havler, D., Fitch, N. J. S. and Richman, D. D. (1997). *In Vivo* Compartmentalization of Human Immunodeficiency Virus: Evidence from the Examination of *Pol* Sequences from Autopy Tissues. *J. Virol*. **71**: 2059-2071.

Wright, T. C. Jr., Subbarao, S., Ellerbrock, T. V., Lennox, J. L., Evans-Strickfaden, T., Smith, D. G. and Hart, C. E. (2001). Human Immunodeficiency Virus 1 Expression in Female Genital Tract in Association with Cervical Inflammation and Ulceration. *Am. J. Obstet. Gynecol.* **184**: 279-285.

Wyatt, R. and Sodroski, J. (1998). The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, and Immunogens. *Science*. **280**: 1884-1888.

Yang, C., Bash, B. Hanna, S. L., Frances, H. S., Nzilambi, N., Colebunders, R. C., Louis, M. S., Quinn, T. C., Folks, T. M. and Lal, R. B. (2000). Predominance of HIV Type 1 Subtype G Among Commercial Sex Workers from Kinishasa, Democratic Republic of Congo. *AIDS Res. Hum. Retroviruses*. **17**: 361-365.

Young, J. A. T. (1994). The Replication Cycle of HIV-1. **In**: *The AIDS Knowledge Base*. Pp. 1-24, (Cohen, P.T., Sande, M.A., Volberding, P., Feinberg, M.B., Gerberding, J. L., Kaplan, L. D., Osmond, D. H., Safrin, S. and Wofsy, C. B. eds.). Little, Brown and Company Boston, New York, Toronto, London.

Zewde, A., Bahiru, S., Sanders, E., Tilahun, T., Beyene, A., Alebachew, M., Schaap, A., Wolday, D., Rinke de Wit, T. F. (2002). HIV-1 Seroprevalence and Subtypes in Police Recruits from Afar Regional State, Ethiopia. *Ethiop. Med. J.* **40** (Suppl. 1): 1-10

DECLARATION

I, the undersigned, declare that this thesis is my own work and it has not been presented in other Universities, Colleges or Institutions, seeking for similar degree or other purpose.

Melaku Adal

Name	Date of Submission	Signature