Karolinska Institutet

Thesis for doctoral degree (Ph.D.) 2008

Genetic aspects of HIV-1 evolution and transmission

Irina Maljkovic Berry

Thesis for doctoral degree (Ph.D.) 2008

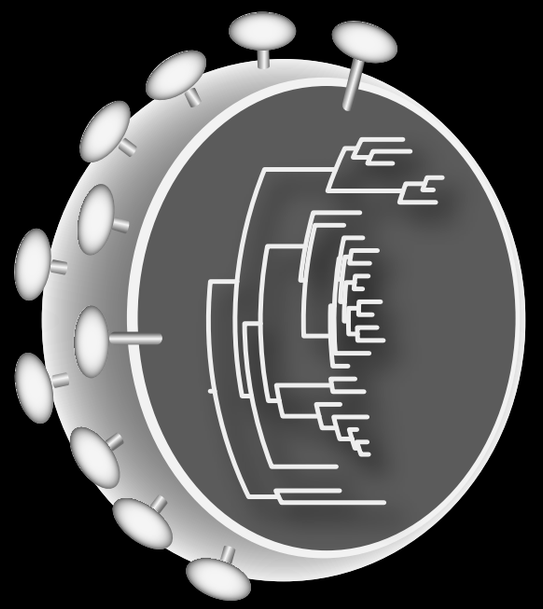
Genetic aspects of HIV-1 evolution and transmission

Irina Maljkovic Berry

Karolinska Institutet

Thesis for doctoral degree (Ph.D.)
2008

Genetic aspects of HIV-1 evolution and transmission



Irina Maljkovic Berry

Karolinska Institutet

From the Microbiology, Tumor and Cell Biology (MTC), Karolinska Institutet,
and the Swedish Institute for Infectious Disease Control, Stockholm, Sweden

Genetic aspects of HIV-1 evolution and transmission

Irina Maljkovic Berry

Karolinska Institutet

Stockholm 2008

ISBN: 978-91-7357-573-7

Printed by

REPROPRINT AB
Stockholm 2008
www.reproprint.se
Gårdsvägen 4, 169 70 Solna

ABSTRACT

HIV-1 is one of the fastest evolving organisms known to man. Its rate of evolution is approximately one million times faster than that of higher organisms such as ourselves, meaning that the amount of changes within the HIV-1 genome in just one year corresponds to the amount of changes within the human genome in one million years. The reason for this remarkable property of HIV-1 is its high amount of genetic variation, created by the rapid substitution introduction, fast generation time, vast number of viral particles produced per unit of time, and various selection forces. As a consequence, an HIV-1 population within a person consists of a large number of genetically related but non-identical viruses, a population structure that gives this pathogen an opportunity of rapid adaptation to changes in its environment. Viral escape variants quickly evolve as a response to the pressure of the human immune system or antiretroviral treatment assuring survival of the virus. In addition, the great genetic variability of HIV-1, both within a person and on the host population level, makes development of an effective vaccine a difficult and complicated task. These issues make studies on HIV-1 evolution and genetic variation highly relevant. This thesis examines different genetic aspects of HIV-1 evolution within a patient and in transmission events.

Prevalence of transmission of drug resistant HIV-1 in Sweden was investigated by analyzing *pol* gene sequences, derived from 100 newly infected and treatment naïve patients, for known resistance mutations. Mutations associated with high and intermediate level of resistance were found in 6 patients suggesting transmission of resistant viral variants. Mutations associated with low or unclear level of resistance were observed to occur at different frequencies in different subtypes. These subtype-specific patterns suggest the existence of different evolutionary paths that HIV-1 can take to develop drug resistance.

Phylogenetic analyses of viral clones and isolates from two HIV-1 infected mother-child pairs revealed the origin of X4 viruses in the children. Although the mothers carried X4 variants at the time of transmission, these were shown not to be the source of X4 variants in the children. Instead, child X4 viruses had evolved from child R5 viruses present early in infection. The initial R5 viruses in the children were correlated to maternal R5 variants that co-existed with maternal X4 at the time of transmission.

Viral phylogenies inferred from HIV-1 sequences derived from 10 patients belonging to a known HIV-1 transmission chain correctly reconstructed the epidemiological events from the chain, except for two of the transmissions and few of the sampling events. The few discrepancies were, however, explained by the existence of hidden viral lineages, that could make the epidemiological and virus trees completely compatible. In addition, the effect of hidden viral lineages could mislead the reconstruction of the root and the sequence evolutionary rate, indicating their importance in phylogenetic analyses of HIV-1 sequences.

We developed a fast and simple method for optimization of the root and evolutionary rate using samples from at least two different time points in a phylogenetic tree. The method had no bias and the estimation of an accurate evolutionary rate was possible even in cases where there was an error in the root and where the tree topologies were incorrectly reconstructed. Hence, the method is robust and thus suitable for rate estimations in real situations where the correct root and topology of a tree are usually unknown.

By analyzing HIV-1 sequences from different epidemics throughout the world we observed that the rate of evolution of HIV-1 on the population level depends on its rate of spread. The virus spreading rapidly in IDU standing social networks had significantly lower rate of evolution than the virus spreading more slowly through heterosexual contacts. In addition, viruses in mixed epidemics, spreading both slow and fast, showed an intermediate evolutionary rate. Epidemiological modeling predicted that the rate of evolution of HIV-1 spreading in a rapid manner will increase as the epidemic ages and the population gets saturated with infections.

This work is dedicated to all those affected by HIV

LIST OF PAPERS

This thesis is based on the following papers, referred to in the text by their Roman numerals:

I **Maljkovic I**, Wilbe K, Solver E, Alaeus A, Leitner T. 2003. Limited transmission of drug-resistant HIV type 1 in 100 Swedish newly detected and drug-naïve patients infected with subtypes A, B, C, D, G, U, and CRF01_AE. *AIDS Res Hum Retroviruses. 19(11):989-97.*

II Clevestig P, **Maljkovic I**, Casper C, Carlenor E, Lindgren S, Naver L, Bohlin A-B, Fenyo EM, Leitner T, Ehrnst A. 2005. The X4 phenotype of HIV type 1 evolves from R5 in two children of mothers, carrying X4, and is not linked to transmission. *AIDS Res Hum Retroviruses. 21(5):371-8.*

III **Maljkovic Berry I**, Franzen C, Albert J, Skar H, Aperia K, Leitner T. A known HIV-1 transmission history reveals limitations in the reconstruction of epidemiological events through analysis of viral phylogenies. *Submitted manuscript.*

IV **Maljkovic Berry I**, Ribeiro R, Kothari M, Athreya G, Daniels M, Lee HY, Bruno W, Leitner T. 2007. Unequal evolutionary rates in the human immunodeficiency type 1 (HIV-1) pandemic: the evolutionary rate of HIV-1 slows down when the epidemic rate increases. *J Virol. 81(19):10625-35.*

V Athreya G and **Maljkovic Berry I**, Kothari M, Daniels M, Korber B, Kuiken C, Leitner T. A simple method for optimizing the root and evolutionary rate in phylogenetic trees with taxa collected at a minimum of two different time points. *Submitted manuscript.*

ABBREVIATIONS

AIDS	acquired immune deficiency syndrome
APOBEC	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
AZT	zidovudine
CCR5	CC-chemokine receptor 5
CD4	cluster of differentiation 4
CRF	circulating recombinant form
CTL	cytotoxic T-lymphocyte
CXCR4	CXC-chemokine receptor 4
DNA	deoxyribonucleic acid
ER	endoplasmatic reticulum
env	envelope
FIV	feline immunodeficiency virus
FSU	former Soviet Union
gag	group specific antigen
gp	glycoprotein
HAART	highly active antiretroviral therapy
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HR	heptad repeat
HTLV-III	human T-cell lymphotropic virus type 3
IDU	intravenous drug user
IN	integrase
LAV	lymphadenopathy-associated virus
LTR	long terminal repeat
MHC	major histocompability complex
MRCA	most recent common ancestor
mRNA	messenger RNA
MSM	men who have sex with men
nef	negative factor
p	protein
PI	protease inhibitor
PIC	pre-integration complex
pol	polymerase
PR	protease
rev	regulator of virion
RNA	ribonucleic acid
RT	reverse transcriptase
SIV	simian immunodeficiency virus
tat	transactivator of transcription
tRNA	transfer RNA
URF	unique recombinant form
V3	third variable region
vif	virion infectivity factor
vpr	viral protein R
vpu	viral protein U

CONTENTS

INTRODUCTION

THE PANDEMIC

Over 33 million infected. More than 25 million dead. The HIV pandemic is one of the most challenging infectious diseases worldwide. Every day, approximately 6800 persons acquire the infection and 5700 die from AIDS. AIDS remains the leading cause of death in certain parts of the world and the HIV pandemic is considered one of the most destructive ones in recorded history.

The beginning

In 1981 an aggressive form of Kaposi's sarcoma (KS), which usually is a relatively benign cancer occurring in older people, was observed in young homosexual men in New York [1]. Simultaneously, there was an increase of pneumocystis carinii pneumonia (PCP), a rare lung disease, observed in homosexual men in both New York and California [2]. The increase in PCP was noticed at the Centers for Disease Control (CDC) in Atlanta, and a report was published about the occurrence of PCP without identifiable cause [3]. This report is sometimes referred to as the "beginning" of AIDS.

In 1982, it became clear that the new disease was not limited to the homosexual population, as the first cases of PCP were reported in injecting drug users (IDUs), hemophiliacs, and among individuals of Haitian origin [4-6]. A number of reports also described occurrence of the disease in Europe [7-11]. This year, the acronym AIDS (Acquired Immune Deficiency Syndrome) was suggested, because: i) the condition was acquired, ii) it caused a deficiency of the immune system, and iii) it was not a single disease but a syndrome manifested through a number of opportunistic infections [12, 13].

The first clear evidence that AIDS was caused by an infectious agent came at the end of 1982, when a child that received multiple transfusions of blood and blood products died of AIDS-related infections [14]. At the same time, the first cases of mother to child transmission were discovered [15]. In the beginning of 1983, it was suggested that the disease could also be transmitted heterosexually, as AIDS was discovered among women with no other risk factors [16, 17]. Meanwhile, new reports from Europe indicated the existence of two different AIDS epidemics: one in the European homosexual population correlated to the epidemic of North America, and one in individuals from, or with connections to, Central Africa [18-20].

In May of 1983, a new retrovirus named lymphadenopathy-associated virus (LAV), was isolated from a lymph node of a patient with lymphoadenopathy at the Pasteur Institute in Paris, and was suspected to be the cause of AIDS [21]. A year later, researchers at the National Cancer Institute in the United States isolated a virus called Human T-cell Lymphotropic Virus Type 3 (HTLV-III), which they suggested caused AIDS [22, 23]. In 1985 more detailed reports on the LAV and HTLV-III were published and it became clear that the viruses were the same [24, 25]. A new name for the causative agent of AIDS was suggested by the International Committee on the Taxonomy of Viruses [26]. The name was Human Immunodeficiency Virus (HIV).

Present time
The estimated number of people living with HIV in 2007 was 33.2 million, of which 2.1 million are children. The number of new infections in 2007 was 2.5 million and the number of deaths was estimated to 2.1 million. Recent global studies suggest that the HIV pandemic has now formed two broad trends, the one of local epidemics sustained in the general population of many Sub-Saharan African countries, and the epidemics in the rest of the world that are concentrated to the high-risk populations such as injecting drug users (IDUs), sex workers, and the male homosexual population [27].

Sub-Saharan Africa is still the most seriously affected region by the HIV pandemic (Figure 1). There, AIDS is the primary cause of death, taking approximately 1.6 million lives in 2007. Of the total global new HIV infections in 2007, 68% have occurred in this region. The estimated number of adults and children living with HIV in Sub-Saharan Africa is 22.5 million, and the adult prevalence varies from less than 2% in some countries to above 15% in most of southern Africa. An estimated 11.4 million children are orphaned due to AIDS in this region. However, due to prevention efforts aimed at reducing new HIV infections since 2000 and 2001, and scaling up of the antiretroviral treatment services, the prevalence of HIV infections and the number of deaths in the most of this region has reached a plateau or even started to decline [27].

Figure 1. Estimated number of adults and children living with HIV-1 in 2007. Reproduced by kind permission of UNAIDS.

In Asia, approximately 4.9 million people are infected with HIV and the epidemic trends vary widely between different countries. The epidemics in Cambodia, Myanmar and Thailand all show declines in HIV prevalence while those in Indonesia and Viet Nam are still growing. In China, all provinces and autonomous regions have reported the occurrence of HIV infections, and an estimated 0.7 million individuals are living with the virus. The rapid growth of the Chinese HIV epidemic has been predicted to result in 10 million or more infections by 2010 if no preventative measures are taken. In Eastern Europe and central Asia, the rapid increase in the number of HIV infected individuals has somewhat slowed, however, the HIV

prevalence is still increasing. The majority of new infections come from the Russian Federation and Ukraine, where the total number of persons with HIV increased nearly 150% between 2001 and 2007. The estimated number of people living with HIV in this region in 2007 was 1.6 million. This is also the estimated number for Latin America. Approximately 2.1 million people are living with HIV in North America and Western and Central Europe. The number of infections in these regions is slowly increasing while the number of deaths, primarily in the North America and Western Europe, has been kept low, 32 000 in 2007, due to widespread access to antiretroviral treatment [27].

There are two types of HIV spreading in the human population. HIV type 1 (HIV-1) is the cause of the pandemic described above. This is the more virulent strain and is the original virus isolated in 1983. HIV type 2 (HIV-2) was discovered in 1985 [28, 29]. This strain is far less virulent and thus found in fewer people. HIV-2 is largely confined to few countries of West Africa, such as Guinea-Bissau, Gambia, Senegal and Guinea, with a prevalence of 1-10%. It is also found in Portugal and in countries with past socio-economical links to Portugal, such as France, India, Angola, Mozambique and Brazil [30, 31]. HIV-2 is responsible for less than 1 million infections worldwide [27, 31].

ORIGIN OF HIV

It is not known how many people became infected with HIV and developed AIDS prior to its discovery in the 1980s. It is believed, however, that before its discovery the virus silently spread to at least five continents of the world and it has been suggested that at least 100 000-300 000 persons were infected [32]. The earliest known HIV-1 infection is that of an adult male from what now is Democratic Republic of Congo, whose plasma was collected in 1959 and later was shown to be HIV seropositive [33]. Phylogenetic analyses of HIV-1 and HIV-2 sequences, including the 1959 sample of HIV-1, dated the origin of HIV-1 infection leading to the pandemic of today to around 1920 -1930 in Central West Africa, and the origin of HIV-2 infection in humans to around 1940 in West Africa [34-36].

Crossing species

HIV is believed to be a zoonotic infection, transferred to humans by several cross-species transmission events of the closely resembling Simian Immunodeficiency Viruses (SIVs) found in the African non-human primates. SIVs are species-specific and in most cases do not appear to cause disease in their natural hosts. HIV-2 is believed to have originated from the SIVsm, found in sooty mangabey (*Cercocebus atys*) [37, 38]. This monkey is indigenous to West Africa, where most of the HIV-2 infections are found. Until 1999, the closest counterpart to HIV-1 was a common chimpanzee virus (SIVcpz), however, HIV-1 and SIVcpz showed significant differences [39]. In 1999 a SIVcpz virus almost identical to HIV-1 was discovered in a sample from a certain subspecies of chimpanzee, *P.t.t* (*Pan troglodytes troglodytes*), shown to reside in southern

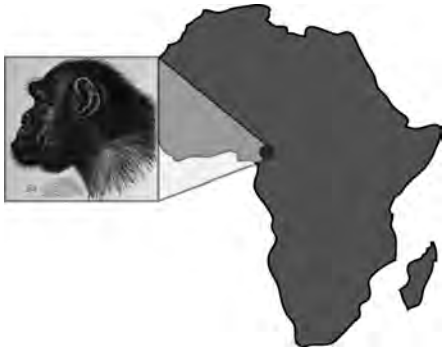


Figure 2.*Pan troglodytes* chimpanzee from Cameroon, West Africa.

Cameroon [40, 41] (Figure 2). It has also been suggested that SIVcpz*Ptt* is result of a recombination event in the chimpanzees, and that the two recombining viruses probably originated from two other monkey species on which chimpanzees prey, the red-capped mangabeys (*Cercocebus torquatus*) and the great spot-nosed monkeys (*Cercopithecus nictitans*)[42, 43].

How the virus crossed from monkeys to humans is debated, but the most accepted theory is the so called "hunter theory". Here, it is believed that the transmission took place when the chimpanzees were killed and eaten by humans, possibly by the hunter being bitten or cutting himself while dealing with an infected animal. In fact, transfer of retrovirus from apes to hunters is still occurring today. This transfer has, interestingly enough, been detected in Cameroon, which is believed to be the very same region where the cross-species transmission of SIVcpz to humans occurred [44]. Among other theories is the polio-vaccine theory, where it was suggested that the SIV crossed to humans by the use of an experimental polio vaccine prepared using chimpanzee kidneys. This theory was disproved, as analyses of original vaccine samples contained no traces of SIV or HIV. Furthermore, it was shown that only kidneys from Asian monkeys were used, in which no natural SIV infection has been found [45, 46].

TRANSMISSION

After the initial cross-species transmission event and introduction of HIV-1 into the human population, the virus continued spreading throughout the world via human-to-human transmission. The three major routes of human-to-human HIV-1 transmission are: blood or blood product route, sexual route, and mother to child transmission.

Blood and blood product route

Blood transfusions cause the greatest risk of HIV infection based on the risk from a single exposure. Transfusion of HIV-1 infected blood or blood products leads to infection in approximately 90 % of the cases [47]. Although screening of blood and blood products for HIV antibodies is performed in many countries, there are still people that do not have access to safe blood. This is estimated to cause 5-10% of HIV infections worldwide [48]. Injecting drug use proposes another risk of blood related infection, when unsterile needles and other equipment is used. Although the risk associated with a single known exposure is relatively low, less than 1%, repeated use of blood contaminated equipment increases the risk substantially [49, 50]. Approximately 10% of the world's HIV-1 infections are a result of injecting drug use. This number is substantially higher, 60%, in the areas with a high number of IDUs, such as in Eastern Europe and Central Asia [27, 51, 52]

Sexual route

Sexual transmission is estimated to account for 70-80% of all global HIV infections. The risk of acquiring HIV through sexual contact with an infected partner ranges from 0,1 to 1 % [53]. Although the probability that a sexual partner is infected with HIV is higher in some geographical areas then in others, the risk of exposure to HIV exists throughout the world. Several factors exist that increase individual susceptibility or infectivity such as genital ulcers and other sexually transmitted diseases, type of sexual act, stage of HIV-1 infection and viral load [53-57]. Homosexual transmission of HIV-1, i.e. transmission through sex between men (MSM), accounts for 5-10% of the world's HIV-1 infections, however, this number varies in

different geographical regions. The highest numbers of MSM infections are found in the developed world, such as North America and Western and Central Europe. Heterosexual intercourse remains the major route of infection in Sub-Saharan Africa, and the majority of HIV-positive individuals here (61%) are women. Biologically, women are twice as likely to be infected by HIV-1 through heterosexual intercourse then men. In addition, the majority of commercial sex workers are women. Transmission of HIV-1 through commercial sex work is highest in certain parts of Asia, and is increasing in Eastern Europe [27].

Mother to child transmission

Transmission of HIV-1 from mother to child can occur during pregnancy, at delivery or through breastfeeding. Excluding breastfeeding, the major risk of maternal HIV-1 transmission appears to occur late during pregnancy or at delivery. Transmission rate of HIV-1 from mother to child varies throughout the world. In the absence of antiretroviral treatment, as the case is in many Sub-Saharan African countries, the transmission rate between mother and child is around 25% [58, 59]. However, where cesarean section and antiretroviral drug treatment are available, this risk can be reduced to as low as 1% [58, 60, 61]. The probability of HIV-1 transmission through breastfeeding has been shown to be similar to the probability of heterosexual transmission [62]. A number of other factors can affect transmission of HIV-1 from mother to child, such as maternal viral load, maternal neutralizing antibodies, and the stage of HIV infection [63-65]. Mother to child transmission is responsible for approximately 90% of HIV-1 infections in children worldwide. Furthermore, 90% of HIV-1 infected children live in Sub-Saharan Africa, followed by the Caribbean, Latin America and South and Southeast Asia [27].

PATHOGENESIS

Following infection with HIV-1, the average time to disease progression in the absence of treatment is 10 years [66, 67]. However, this time varies greatly between individuals and is generally shorter in children. A subset of HIV-1 infected persons, approximately 10-15%, are called rapid progressors and develop AIDS within four years after primary infection [68, 69]. On the other hand, approximately 5% show no signs of disease progression and remain asymptomatic for over 12 years [66, 70, 71]. These individuals are referred to as long-term non-progressors.

The course of HIV-1 infection can be divided into three stages (Figure 3). The first stage is the primary (acute) infection, which is defined as the time period between initial HIV infection and development of an antibody response, usually lasting seven to eight weeks. Symptoms of primary infection appear days to weeks following exposure to the virus, and usually resemble those of influenza. However, not all patients develop clinical symptoms during this stage of infection. Primary HIV infection is characterized by intense viral replication, resulting in high titre of virus particles in plasma, and a decrease in CD4+ T lymphocytes. Over the following weeks the number of CD4+ T-lymphocytes starts to recover, while the viremia declines several orders of magnitude reaching a setpoint. This setpoint has been suggested to be a good predictor of disease progression [72].

The second stage of infection is the clinically latent, chronic phase, initiated when the immune response to HIV-1 infection is fully developed. The characteristic of the chronic phase is persistent replication of the virus in the lymphoid tissue, accompanied by the

progressive depletion of CD4+ T-lymphocytes. Duration of the chronic infection is highly variable, lasting from few to ten years or more. The third stage occurs when the CD4+ T-cell count drops below a critical level, usually 200 T-cells/μl, resulting in the loss of cell-mediated immunity and collapse of the immune system. This leads to appearance of a number of opportunistic infections and development of AIDS. The final stage can last for a few years, longer with antiretroviral treatment, and eventually leads to death.

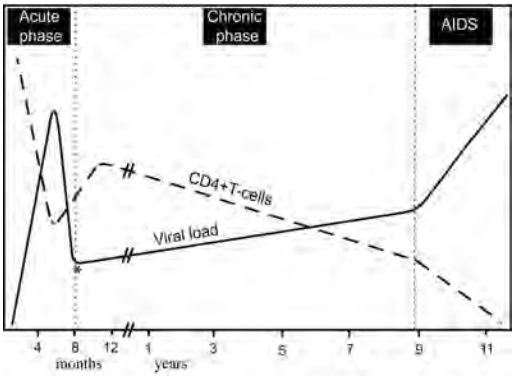


Figure 3. Schematic figure of the viral load and CD4 T-cell count within a treatment-naïve patient during the course of infection. * viral setpoint

THE VIRUS

HIV-1 is a lentivirus, which belongs to the Retroviridae family. Lentiviruses are found to infect many species, such as primates (SIV), cats (FIV) and sheep (Visna), and are characterized by causing prolonged sub clinical infections with persistent viremia, weak neutralizing antibody responses, and continuous virus mutation.

Structure

The virion is spherical with an average diameter of 100nm. Its external part consists of a lipid bilayer membrane, the envelope. The envelope is derived from the membrane of a host cell and thus contains host-specific proteins, such as the major histocompability complex (MHC). In addition, the envelope is equipped with virus-encoded glycoproteins, gp120 and gp41, non-covalently bound to each other. Three gp120-gp41 heterodimers form a “spike”, which is anchored to the envelope through its gp41 molecules, while the attaching gp120 molecules form a cap on the outside of the virion. Approximately 70 spikes are found on the surface of HIV-1. Beneath the envelope is a layer of matrix proteins, p17, surrounding the viral capsid. The capsid has a form of a truncated cone and is built up of capsid proteins, p24. Inside the capsid is the viral core, containing two positive-strand RNA molecules.

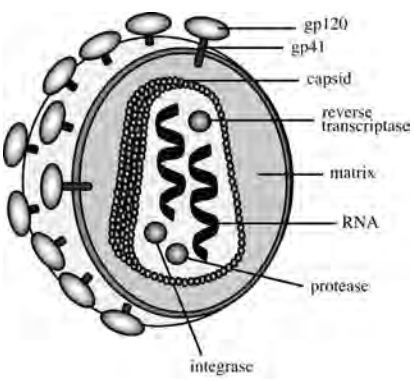


Figure 4. Structure of HIV-1.

The core also contains viral enzymes essential for viral replication: reverse transcriptase (RT), protease (PR), and integrase (IN) (Figure 4).

Genome

The HIV-1 genome consists of two copies of positive sense single-stranded RNA molecules, each approximately 10 000 bases long. Only nine genes are found in HIV-1, compared to 20 000-25 000 found in humans (Figure 5). The compact structure of HIV-1 genome is achieved by the use of all three reading frames and the use of differential splicing.

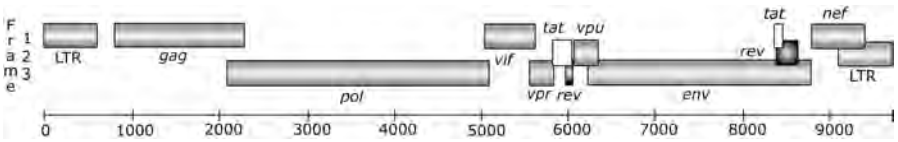


Figure 5. Organization of HIV-1 genome. Scale bar shows approximate nucleotide positions.

As in all retroviruses, the structural and enzymatic proteins of HIV-1 are encoded by *gag*, *pol* and *env* genes. The *gag* gene encodes polyprotein p55, which is cleaved into p24, p17, p7, p6, p2 and p1 proteins. The main purpose of p24, p17, p7 and p6 proteins is to build up viral matrix and capsid structures and to act in the process of viral assembly. The exact function of p1 and p2 is yet to be determined. The *pol* gene encodes viral enzymes RT, PR, and IN, which are necessary for viral replication. The *env* gene encodes viral polyprotein gp160, which is cleaved to form the gp120 and gp41 envelope glycoproteins. These are essential for viral attachment and entry into the host cell. In addition to the three structural genes, two regulatory and four accessory genes are found in the genome of HIV-1 (Table 1).

Gene	*Protein*	*Type*	*Function*
tat	Tat (p14/p16)	regulatory	viral transcriptional transactivator
rev	Rev (p19)	regulatory	upregulates expression of Gag, Pol and Env, downregulates itself and Tat
nef	Nef (p25/p27)	accessory	downregulates CD4 and MHC-I, prevents apoptosis
vif	Vif (p23)	accessory	promotes virion maturation and infectivity, inhibits APOBEC* function
vpr	Vpr (p12/p10)	accessory	involved in nuclear entry, prevents cell division
vpu	Vpu (p16)	accessory	downmodulates CD4 in ER, promotes virion release

Table 1. Regulatory and accessory genes of HIV-1. * APOBEC is a cellular cytidine deaminase that deaminates multiple cytosine (C) residues of viral RNA to uracil (U) during the reverse transcription. This results in guanine (G) to adenine (A) hypermutation of viral DNA leading to non-viable virus particles.

The coding regions of HIV-1 genome are flanked by long terminal repeats (LTRs). LTRs are non-coding regions containing promoters, enhancers, and other elements essential for viral replication and its interaction with cellular transcription factors.

REPLICATION CYCLE

HIV-1 infects cells of the human immune system that express its main entry receptor, CD4, on their surface (Figure 6). The primary targets of HIV-1 are T-lymphocytes and

macrophages, but the virus can infect many other cells, such as monocytes, dendritic cells, and microglial cells in the brain. Depletion of CD4+ T-lymphocytes due to HIV-1 infection leads to loss of cellular immunity and is the main cause of AIDS.

Binding and entry

The entry of HIV-1 into a host cell is initiated by a high-affinity binding of the viral gp120 glycoprotein to the CD4 receptor on the surface of the target cell. This interaction induces a conformational change in the gp120 molecule, so that previously hidden conserved regions and epitopes of its V3-loop are exposed [73, 74]. The newly exposed regions interact with the target cell's surface chemokine receptors, which are the secondary entry receptors of HIV-1. Binding of the virus to these co-receptors, mainly CCR5 and CXCR4, induces a subsequent conformational change, this time of the viral gp41 transmembrane molecule. The change of the gp41 conformation leads to exposure of its hydrophobic fusion domain and anchoring into the cell membrane, followed by interaction of its two helical regions, HR1 and HR2, to form a 6-helix bundle. This mechanism brings the viral and target cell membranes in close proximity to each other, allowing for their fusion and the release of the viral core into the target cell cytoplasm.

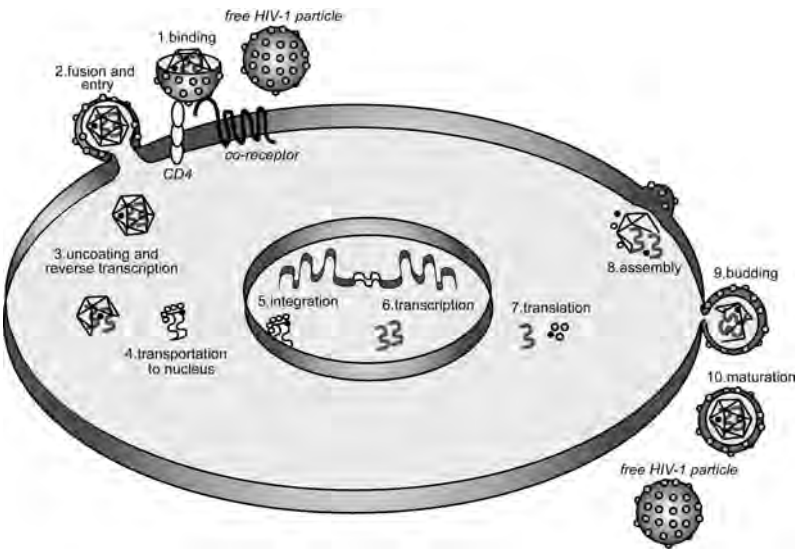


Figure 6. The lifecycle of HIV-1.

Reverse transcription and integration

Once the viral core is in the cytoplasm of the target cell, its genome and accompanying enzymes are freed by uncoating, and the reverse transcription of viral RNA begins. The reverse transcription is mediated by the RT enzyme, which initiates synthesis of DNA by adding nucleotides to a tRNA primer located on the viral RNA strand. As the first DNA strand is synthesized, the RNA is degraded by viral enzyme RNaseH, allowing synthesis of the complementary DNA. Since the rate of replication of HIV-1 is important, the RT enzyme lacks the time-consuming proofreading property found in cellular DNA-polymerases. Thus, the RT is highly error-prone, introducing substitutions, repetitions, insertions and deletions into the newly synthesized viral DNA. Another property of the RT is that it has the ability to jump over from one of the viral RNA copies to the other during the process of reverse

transcription. This becomes important in instances when the two RNA strands are not identical, and is an essential feature of the mechanism of HIV-1 recombination. The newly synthesized double stranded viral DNA forms, together with IN, vpr and viral matrix protein, a pre-integration complex (PIC). The reverse transcription complexes and PICs use cellular microtubule networks for their translocation from the cytoplasm to the nucleus of the host cell [75]. Once inside the nucleus, viral DNA is integrated into the host cell genome by the IN enzyme. The sites at which HIV genome is integrated are predominantly found in the active transcription units of the host cell DNA [76]. It has been suggested that IN is the main viral determinant of HIV integration specificity, however, a partial role of the cellular lens epithelium-derived growth factor (LEDGF/p75) for the favored integration of HIV has also been described [77, 78]. The integrated viral DNA is referred to as a provirus, and it can stay latent, or be transcribed by the cellular machinery upon the activation of the host cell.

Transcription and translation

Transcription of the provirus is mediated by the cellular RNA polymerase II, which generates sub-genomic mRNA translated into the early viral proteins Tat, Nef and Rev. Tat protein binds to the trans-activating region (TAR) in the LTR of the provirus, thus promoting phosphorylation of the RNA polymerase II, which enhances its activity. Nef protein mediates down-regulation of CD4, which prevents premature interaction of this molecule with viral particles produced in the cell [79, 80]. It also down-regulates MHC-I, thus avoiding detection by the immune system [80, 81]. In addition, Nef prevents apoptosis of the infected cell by interacting with its Apoptosis Signal-Regulating Kinase 1 (ASK1) [82]. The Rev protein helps with transportation of viral mRNA from nucleus to the cytoplasm for translation, and it regulates the switch from, and the balance between, early and late gene expression. The mRNA transcribed from the late genes is translated by the free polyribosomes in the cytoplasm into the gag and gag-pol polyprotein, and by the ER membrane-bound polyribosomes into the env polyprotein (gp160). gp160 is then transported to the Golgi apparatus, where it is extensively glycosylated and cleaved into the gp120 and gp41. From here, membrane bound trimeres of gp120 and gp41 are moved to the surface of the host cell.

Assembly and release

The assembly of viral polyproteins and its RNA genome takes place near the cellular membrane, where they form an immature capsid. The capsid buds from the cell, thereby acquiring its envelope already equipped with the gp120 and gp41 trimeres, as well as other cellular membrane-associated proteins (Figure 7).

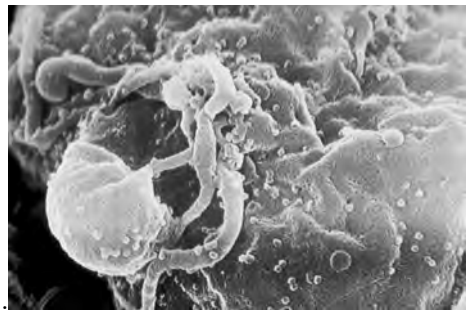


Figure 7. Scanning electron micrograph of HIV-1 particles budding from cultured lymphocyte. HIV-1 is seen as small dots on the surface of the cell.

The newly produced viral particle is still immature, and requires cleavage of its gag-pol polyprotein. This is done by the PR, which cleaves the pol portion into the viral enzymes, and the gag portion into the structural proteins. This process results in a mature virus particle, ready to infect new cells and produce progeny of its own.

TREATMENT

There are several different antiretrovirals currently used for treatment of HIV-1 infection. These antiretrovirals are constructed to interfere with the different parts of the viral replication cycle, thereby interrupting production of new virus particles. The antiretrovirals do not cure HIV-1 infection, they merely stabilize viremia and increase survival time of infected individuals.

Inhibitors today

AZT was the first approved drug against HIV-1 and it became available in 1987, approximately 4 years after the discovery of the virus. This drug belongs to the RT inhibitor class of antiretrovirals that target the reverse transcription phase of viral replication cycle. There are two types of reverse transcription inhibitors, nucleoside RT inhibitors and non-nucleoside RT inhibitors. Nucleoside RT inhibitors are nucleoside analogues that are inserted into the growing chain of viral DNA by the RT enzyme. There, they terminate the growth of the chain, thereby inhibiting HIV-1 transcription. Non-nucleoside RT inhibitors act in a different way, directly targeting the RT enzyme by binding to its active site and in that way inhibiting its activity. The introduction of RT inhibitors raised great hope for HIV-1 infected individuals, however, the effect of these drugs was soon shown to be limited. RT inhibitors only prolonged life for a few years, and their use as mono- or dual therapy quickly lead to the appearance of drug resistance [83].

In 1995 a new class of antiretrovirals was introduced that targeted the virion assembly phase of HIV-1 replication cycle. These were the protease inhibitors (PIs), interacting with the PR enzyme and inhibiting its cleavage of viral polyproteins in the budding virion, thus resulting in release of immature, non-infectious virus particles. Introduction of PIs lead to the use of highly active antiretroviral therapy (HAART), consisting of a combination of at least three drugs belonging to at least two different classes of antiretrovirals. The use of HAART lead to a decrease of HIV-1 associated morbidity and mortality, increasing the survival time with 4-12 years [84-86].

The two most recent classes of antiretrovirals are entry and integrase inhibitors. Entry inhibitors, also known as fusion inhibitors, consist of: co-receptor antagonists, which bind to the CCR5 co-receptor of the target cell thus preventing its interaction with viral gp120 glycoprotein; and fusion inhibitors, that bind to gp41 thereby preventing it from bringing viral and cell membranes together for their fusion [87, 88]. Integrase inhibitors block the action of IN enzyme in incorporating viral DNA into the genome of the host cell [89]. These classes of antiretrovirals are often used as salvage therapy for individuals whose virus is resistant to RT and PR inhibitors [90, 91].

Future treatment

Currently, a new class of antiretrovirals is under development, the maturation inhibitors. These are similar to the protease inhibitors as they prevent cleavage of the gag polyprotein

and result in the release of non-infectious particles. However, unlike PIs, maturation inhibitors bind to the gag polyprotein itself and not to the PR enzyme. When bound they prevent cleavage of this polyprotein, resulting in the release of structurally defective viral particles from the infected cell [92]. Additionally, recent studies reveal construction of a new enzyme that targets the integrated viral DNA. This recombinase enzyme recognizes an asymmetric sequence in the LTR region of the provirus and efficiently cuts out HIV-1 DNA from the genome of the infected cell [93]. The possibility of reversion of the viral integration is a promising mechanism for future antiretroviral use.

MECHANISMS OF HIV-1 EVOLUTION

The main reason to why antiretroviral treatment does not protect against AIDS, and why there is no cure or vaccine against this virus, is its high level of genetic variation. In fact, HIV-1 is one of the fastest evolving organisms known to human, with an evolutionary rate estimated to be approximately one million times higher than the rate of higher organisms [94]. There are several factors influencing this extraordinary property of HIV-1, such as substitution introduction and selection.

Genetic variation

Introduction of substitutions into the viral genome is mediated by the RT enzyme during the reverse transcription phase of the viral replication cycle. The error frequency of the RT has been estimated very high, $3.4x10^{-5}$ substitutions per site per replication cycle [95]. Since the size of HIV-1 genome is approximately 10 000 bases, this means that around one substitution is introduced into every second to third newly synthesized viral genome. In addition to single substitution introduction, the RT also enters deletions, insertions and duplications. Furthermore, significant changes can be introduced into the viral genome by recombination (Figure 8). The recombination event has been estimated to occur two to three times per replication cycle [96]. Thus, the high genetic variability of HIV-1 is strongly influenced by the low fidelity of its RT enzyme.

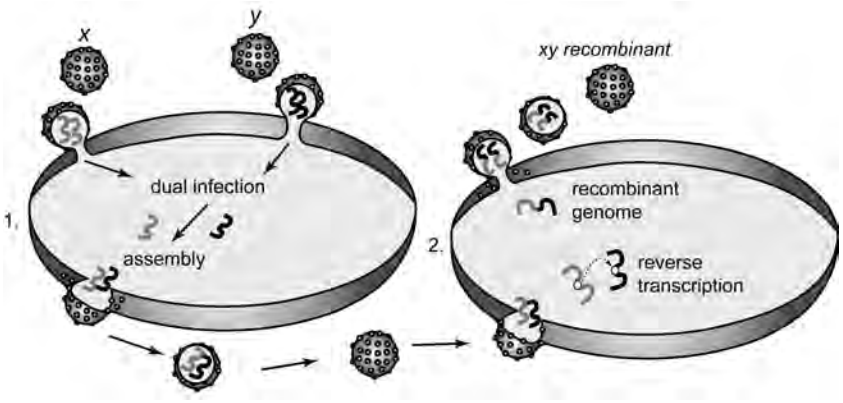


Figure 8. Recombination of HIV-1. 1. Infection of a cell with two different HIV-1 viruses, x and y, where the viral progeny gets one RNA copy from virus x and one from virus y. 2. The progeny infects a new cell where reverse transcriptase "jumps" from one RNA copy to the other creating a recombinant genome (xy). Packaging of the recombinant genome into a newly produced particle gives rise to a recombinant form of the virus.

Other important factors contributing to the high evolutionary rate of HIV-1 are its fast replication rate and the great number of viral particles produced per unit of time. The generation time of the virus, i.e. the time from the release of a virion until it infects a new cell and releases a progeny of its own, has been estimated to be approximately 2 days [97, 98]. In addition, the number of new viral particles produced per day has been approximated to 10^9 [97, 99]. This, together with the error introduced by the RT enzyme, results in a viral population within an individual consisting of a pool of genetically related but non-identical viruses, giving rise to the possibility of existence of every possible point substitution in that population. Such a population is often referred to as a quasispecies [100, 101]. The pre-existing variants of the HIV-1 quasispecies provide the viral population with an ability to rapidly adapt to changes in its environment.

Selection

A change in the environment imposes a selective pressure on the viral population, such that viral variants with genes best adapted to the new milieu are selected for. The impact of selection on a viral gene or gene segment can be determined by analyzing the ratio of synonymous to non-synonymous substitutions. Synonymous (silent) substitutions are those where the nucleotide change does not alter the encoded amino acid while non-synonymous substitutions are those that result in a change of the encoded amino acid. Generally, synonymous substitutions occur in the third position of a codon. When fixed in a population, they are believed to reflect the random genetic drift of the virus, occurring independently of environmental factors [102]. The occurrence and fixation of non-synonymous substitutions is, on the other hand, a reflection of the selection pressure imposed on the analyzed genetic region.

When the frequency of synonymous substitutions is greater than the frequency of non-synonymous substitutions, the selective pressure is assumed to be negative (purifying). In most cases negative selection predominates, as genes are striving to be conserved so that the proteins they code for maintain their structure and function. In view of negative selection, most of the substitutions are deleterious and are removed from the viral population. The HIV-1 *pol* gene is known to be under mostly negative selection pressure since it encodes important replication enzymes [103, 104]. Positive (diversifying) selection is observed when a change of an HIV-1 protein is required to ensure survival of the virus. Most of the substitutions here are also deleterious, however, some will result in an advantageous change in the protein and will, therefore, be fixed in the viral population. This scenario is reflected in the higher frequency of non-synonymous substitutions, and consequently, higher genetic variability. Certain regions of HIV-1 *env* gene, such as the V3 region, are highly variable and under strong positive selection pressure, since they are targeted by the immune system and in need of constant change [105, 106].

The selective pressure of the immune system: viral escape

The human immune system drives HIV-1 evolution by positively selecting variants with reduced sensitivity to cytotoxic T-lymphocytes (CTLs) and neutralizing antibodies. The best evidence for this is the emergence of viral escape mutants, which are, in the case of CTLs, associated with disease progression [107, 108]. CTL escape mutants can emerge shortly after the onset of symptoms of acute infection and become the dominant viral variants in the patient. Rapid emergence of CTL escape variants indicates their pre-existence in the viral population and points to the dominant role of CTLs as a selective force in the HIV-1 infection [107]. It has been suggested, however, that the emergence of CTL escape mutations

is correlated with loss of viral replication fitness [109]. Although the emergence of neutralizing antibody escape mutants has not been correlated to disease progression, their sole existence indicates that neutralizing antibodies, like CTLs, do exert selective pressure on HIV-1. The high variability of gp120 regions ensures escape from antibody recognition, while its heavy glycosylation reduces accessibility of the neutralizing antibody epitopes [110].

The selective pressure of antiretroviral treatment: drug resistance

Effective antiretroviral treatment has been shown to slow viral replication and to slow down and even totally abolish the evolution of HIV-1 [111, 112]. However, sub-optimal antiretroviral treatment allows continuation of viral replication. Given the quasispecies structure of an HIV-1 population, and the strong selective pressure of the treatment, the emergence of pre-existing viral variants with mutations conferring drug resistance is inevitable. Indeed, a considerable proportion of treated individuals develop resistance to one or more drugs in a relatively short period of time.

Drug resistance mutations selected early in the process of resistance accumulation are usually primary mutations. These are drug specific and tend to give rise to relatively high decrease in drug susceptibility [113]. For some antiretrovirals, only one primary mutation is enough to cause high-level resistance, while for others multiple mutations are required. However, the acquirement and maintenance of the primary resistance mutations is costly for the virus, as they usually occur in the active site of the RT and PR enzymes and thus directly affect viral replication fitness [114, 115]. Supporting this is the fact that the resistant variants tend to be outgrown by wild type viruses when treatment is discontinued. Secondary, or compensatory, mutations appear later in a viral population and confer little or no reduction in drug susceptibility. Their role is to alleviate some of the fitness cost caused by the primary mutations [116, 117]. Secondary mutations can arise in the PR and RT enzymes, as well as in other parts of the HIV-1 genome [116, 118].

EVOLUTION OF HIV-1 WITHIN A PATIENT

Consistent with the three stages of HIV-1 infection, the evolution of the virus within a host can be divided into three phases: the acute, the chronic, and the disease phase [119]. Furthermore, the chronic phase can be divided into three periods, in which HIV-1 divergence and diversity have been shown to follow distinct patterns [120]. Divergence of HIV-1 describes its evolution from a founder strain, while diversity is a measure of genetic variation within the virus population at a certain point in time.

Phase I: Diversity, bottlenecks and co-receptor phenotypes

Upon infection with HIV-1, the infected individual harbors a relatively homogenous population of the virus, as transmission is associated with a significant population bottleneck. The viral diversity during transmission of HIV-1 has been estimated very low, less than 1% in both *env* and *gag* genes. Especially the V3 loop of the *env* gene, which otherwise is highly variable, has been shown very homogenous [121-123]. It has also been suggested that the high level of homogeneity is a result of selection acting on the *env* gene, such that only certain variants can be transmitted or establish a successful infection. Supporting the selection theory is the preferential presence of R5 over X4 viruses in infected individuals during the first stage of HIV-1 infection [122, 123].

The R5 phenotype of HIV-1 is assigned to viruses using the CCR5 co-receptor as their secondary receptor for entry into a cell. Viruses using the CXCR4 co-receptor for entry are of X4 phenotype. R5X4 viruses are dual tropic and have the ability to use both co-receptor types. Primary targets for the R5 viruses are macrophages, while X4 mainly infect T-cells. There have been several suggestions as to why R5 is the more common viral phenotype found during the primary infection, such as an R5 transmission advantage over X4, and an advantage of R5 to establish infection. An important factor supporting the transmission advantage of the R5 viruses is the existence of the CCR5Δ32 allele, occurring in the Caucasian population at a frequency of 0.092. The allele has a 32bp deletion in the CCR5 gene resulting in a defective co-receptor, and individuals homozygous for CCR5Δ32 appear to be resistant from infection by R5 HIV-1 [124]. However, the defective allele does not protect from a more rarely occurring infection by X4 viruses [125]. Importantly, the emergence of X4 variants has been associated with an accelerated disease progression [126, 127].

Phase II: Diversity, divergence and different viral lineages

The chronic phase of HIV-1 evolution is characterized by a continuous pressure from the human immune system, resulting in a rapid turnover of viral genetic diversity. The V3 sequences from chronically infected patients have been shown to differ as much as 10-15%, and the rate of their divergence from the founding population has been estimated to roughly 1% per year [120]. During the early period of the chronic phase, viral diversity and divergence have been shown to linearly increase with a similar rate. The intermediate period is characterized by stabilization of viral diversity, while the divergence from the founder strain continues with the same pace. The emergence of X4 viruses in infected individuals is thought to start during the transition from the early to the intermediate period, and their peak in prevalence has been observed at the end of the intermediate period [120]. The X4 viruses have been suggested to evolve from the R5, as very few mutations within the V3 region are needed for R5 to gain the ability to use the CXCR4 co-receptor [128, 129 and **II**]. The increased prevalence of X4 viruses, and their contribution to rapid disease progression, may be correlated with their rapid replication rate resulting in greater fitness, as well as increased availability of activated T-cells during this stage of infection [130]. Finally, the last period of the chronic phase was shown to involve stabilization of viral divergence and a decline in viral diversity, as well as a decline in the prevalence of HIV-1 X4 phenotypes.

High genetic variation of HIV-1 within a patient during this phase can lead to the co-existence of several distinct viral lineages, or sub-populations, competing with each other and thereby increasing the complexity of HIV-1 dynamics [119 and **III**]. Observance of convergent evolution in the V3 region of viral sub-populations supports existence of positive selection in this phase of HIV-1 infection [131]. Indeed, several studies have found that positive selection plays a major role in driving the evolution of HIV-1 *env* gene [103, 106, 132, 133]. However, other studies suggest a greater importance of purifying and neutral selection [134, 135]. The rate of HIV-1 evolution has been shown to differ, not only in different genes of the virus, but also among patients.

Phase III: Diversity and disease progression

The last phase of HIV-1 evolution within a patient starts as the immune system collapses and the individual progresses to AIDS. As a consequence of the immune system collapse, the selection pressure on the virus decreases [136]. The less effective selection pressure results in a lower evolutionary rate of HIV-1, and the previously high level of the quasispecies

diversity now declines until the HIV-1 population is homogenous again [137, 138]. In addition to viral load and CD4+ T-cell count, several studies have suggested a correlation of the rate of disease progression to the evolutionary rate of HIV-1, and a correlation to viral fitness, which has been indicated to increase as infection within a patient progresses [120, 139-142 and **III**].

EVOLUTION OF HIV-1 ON THE POPULATION LEVEL

The effect of intense immune-mediated positive selection on HIV-1 within a patient is reflected in a phylogenetic tree as a clear temporal structure of the viral population showing constant adaptation and lineage extinction. On the host population level, however, the temporal pattern is missing and the structure of viral population shows multiple coexisting strains. The effect of immune-mediated positive selection here is weak and the phylogenetic structure is suggested to reflect demographic and spatial history of transmission [143, 144].

The lack of strong influence of positive selection on the evolution of HIV-1 on the host population level has been proposed to be due to several reasons. Transmission bottlenecks, which cause a significant reduction in viral diversity, in combination with behavioral aspects of the host, may result in limited transmission of viral strains with advantageous mutations across the host population. The presence of a selective component in the transmission bottlenecks might further reduce the amount of influence of positive selection on the HIV-1 evolution on the host population level. In addition, mutations advantageous for the virus in one individual might have too high fitness cost in another [108, 135, 145]. The latter point is an example of negative selection acting on the host population level, and is observed for instance in transmission of CTL escape mutants to individuals with HLA alleles that differ from HLA alleles in the donors. In such instances viral escape mutations frequently revert to ancestral forms quickly after transmission, as their benefit for immune evasion in the new environment is too low compared to their fitness cost [146-148]. This phenomenon is also observed for the pressure of antiretroviral treatment. Transmission of drug resistant viral variants has been shown to occur in 0 to over 20% of new infections in areas where antiretroviral treatment is common [149-153 and **I**]. However, in many instances mutations conferring drug resistance have been shown to revert to ancestral forms in newly infected treatment naïve individuals, suggesting lower replication fitness of these variants [150, 154]. In addition, drug-resistant variants have been suggested to have lower transmission fitness than drug-sensitive viruses [155]. Still, the impact of positive selection on the evolution of HIV-1 on the host population level is not non-existent. Its influence is observed in sustained transmission of some CTL escape variants among patients irrespectively of their HLA type [156, 157], and in transmitted drug resistance mutants reverting not to wild type, but to an intermediate variant with a better fitness and a greater likelihood of developing drug resistance [158, 159].

In addition to the impact of neutral genetic drift and positive and negative selection, the evolution of HIV-1 on the host population level is driven by the patterns of host behavior. Recently, several studies have indicated the importance of social network structures, transmission rates, and the size of epidemic on the HIV-1 inter-patient genetic variability and evolutionary rate [160-162 and **IV**]. The exact effects of these factors, however, are yet to be thoroughly examined.

Global variability of HIV-1: subtypes and recombinant forms

The high genetic variability of HIV-1 together with the forces driving its evolution has resulted in the emergence of several different viral lineages spreading throughout the world. The co-existence of these different evolutionary lineages is observed in a phylogenetic tree, where they cluster in various groups, or clades. The three main groups of HIV-1 are M (major), O (outlier) and N (non-M-non-O), genetically differing from each other by more than 30%.

The three groups appear to have entered the human population by three separate transmissions of SIV from apes to humans, which is evident in a phylogenetic tree where the HIV-1 groups intermix with different chimpanzee and gorilla SIV sequences. Phylogenetic relationships of these viruses indicate chimpanzees as original reservoir of SIVs found in chimpanzees and gorillas, and of HIV-1 in humans. It is suggested that groups M and N of HIV-1 were transmitted from chimpanzees to humans at two distinct transmission events, while the close relationship of group O to SIVgor suggests transmission of group O-like viruses from chimpanzees either to gorillas and humans independently, or first to gorillas that subsequently transmitted the virus to humans [163] (Figure 9).

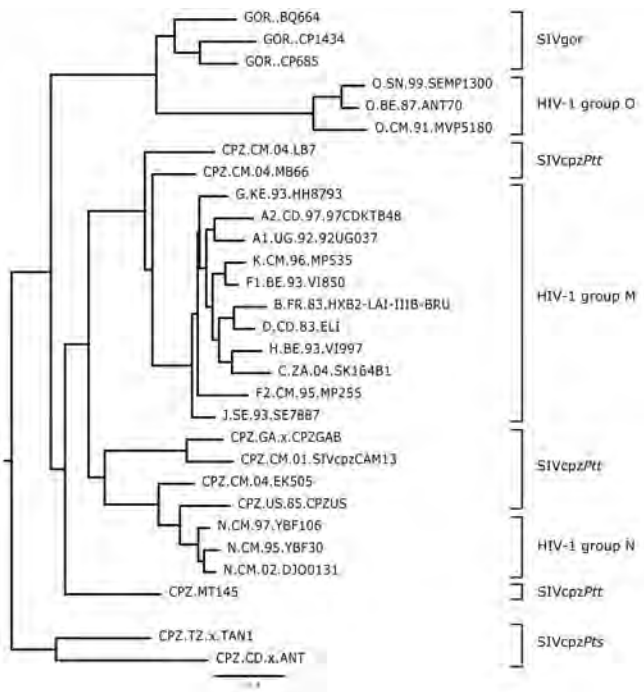


Figure 9. Phylogenetic tree showing relationships between HIV-1 groups and SIV sequences. The tree is based on partial *env* sequences and was inferred by maximum likelihood under a GTR model with 36 free site rates. The SIV sequences were kindly provided by dr. Brian Foley. *Ptt-Pan troglodytes troglodytes*. *Pts-Pan troglodytes schweinfurthii*.

Group O was first described in 1994 and named outlier because of its distinct clustering from group M [164]. Infections with group O viruses are more rare and mainly restricted to West-

central Africa. Group N lineage was identified in Cameroon in 1998 and is extremely rare [165]. The vast majority of HIV-1 infections in the world are caused by the group M HIV-1, which is by far the largest and most diverse of all the groups. It is composed of at least nine different subtypes: A, B, C, D, F, G, H, and J, and subtypes A and F are further divided into sub-subtypes, A1-A3 and F1-F2 (Figure 10). The subtypes of HIV-1 group M differ from each other 10-30%, with greatest diversity found in the *env* gene (30%), followed by *gag* (20%) and *pol* (15%) [166].

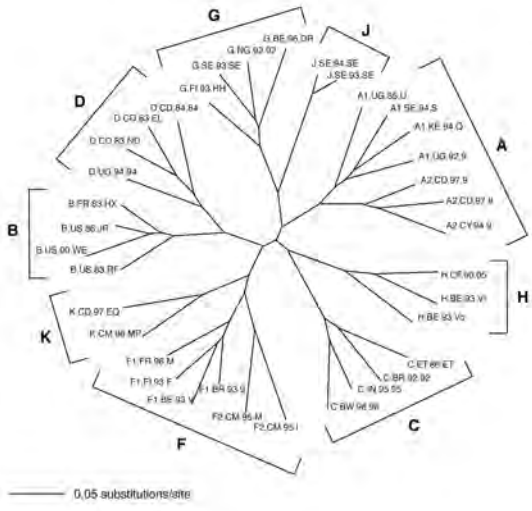


Figure 10. Phylogenetic tree of HIV-1 group M subtypes. The tree was made from *env* sequences using maximum likelihood tree building and a GTR+G+I evolutionary model.

The impact of recombination on the evolution of HIV-1 on the host population level became apparent when recombinant forms of viral subtypes were discovered in the mid-1990s [167-169]. Intersubtype recombinant viruses that successfully established infections in the human population were called circulating recombinant forms (CRFs), and to date, at least 37 of these have been described throughout the world [170]. Some of the CRFs have further recombined with pure HIV-1 subtypes and other CRFs giving rise to second-generation recombinants [171, 172]. Approximately 10-20% of all new HIV-1 infections in the world are caused by circulating recombinant forms. In addition to CRFs, many unique recombinant forms (URFs) exist in the regions where several subtypes co-circulate [170, 173-175]. The URFs are observed in single individuals and have limited transmission in the human population, however, they can account for as much as 30% of infections in the regions where they are found. Additionally, in 1999 it was discovered that the recombination event does not only occur between different subtypes of HIV-1, as a recombinant form between groups M and O was found [176, 177]. It has also been suggested that some pure HIV-1 subtypes are, in fact, recombinant forms [178, 179]. Thus, recombination is an essential factor that shapes the evolution of HIV-1 on the host population level. Its frequent occurrence combined with the high genetic variability of this virus makes the genetic classification, and thereby understanding of HIV-1 genetic variants, a difficult and complicated task.

Subtype distribution and differences

Group M HIV-1 is responsible for the great pandemic of today, however, its different subtypes and recombinant forms are unevenly distributed throughout the world. Several factors are responsible for this differential spread of HIV-1 strains, such as human genetic and social/behavioral factors, as well as founder effects.

The founder effect (single introduction followed by a rapid spread) is strongly supported as the cause of specific HIV-1 subtypes circulating in particular geographic regions, such as the original predominance of subtype B in China, Thailand and India, followed by additional founder events introducing other subtypes into these regions. Now, CRF01_AE predominates in Thailand and subtype C is expanding at a high pace in India [175, 180-183]. In China, the mixing of subtypes B and C has lead to their recombination, such that the CRF07_BC and CRF08_BC now dominate the epidemic in this region [184, 185]. Subtype C is, in fact, the most occurring subtype in the world, being responsible for 52% of all the HIV-1 infections. Its highest prevalence is in Southern and some regions of Eastern Africa [170].The founder effect, in combination with host behavior factors, is also evident in Eastern Europe, where subtype A1 rapidly spread after its introduction into the IDU population [186, 187]. Subtype B infections predominate in North America and Western Europe but the pattern is slowly changing as non-B subtypes and CRFs are being introduced due to immigration. An exception from the founder effect is West and Central Africa, particularly the Democratic Republic of Congo and Cameroon, where nearly every HIV type, group, subtype, and recombinant form can be found [188, 189]. The high diversity of HIV found in this region supports it as source of the HIV epidemic in humans.

Recent studies suggest an additional factor playing a role in the geographical distribution of HIV-1 subtypes: the viral fitness. It has been suggested that different types, groups, subtypes and recombinant forms of HIV-1 differ in their transmission efficiency (transmission fitness) and their replication capability (pathogenic, or replication fitness) [190]. For instance, the replication fitness of group M HIV-1 has been shown to be higher than that of HIV-2, followed by group O, which nearly perfectly matches their prevalence in the epidemic [191]. Subtype C has been shown to have lower replicative fitness than other subtypes, however, its transmission fitness has been suggested higher, which could explain its rapid expansion over other subtypes co-existing in the same regions [190]. It has also been suggested that the recombination event may result in viruses more fit than their parental strains. This has been proven correct for the CRF02_AG, which was shown to have significantly higher transmission and replication fitness than its parental strains A and G [192, 193]. It is possible that this explanation would also be valid for the CRF07_BC and CRF08_BC, which became the dominating strains in the HIV epidemic in China after their creation from previously predominating B and C strains.

Other biological differences between HIV-1 subtypes and CRFs have been suggested, such as difference in co-receptor usage, disease progression, effect of vaccines, and subtype-specific patterns of drug resistance mutation accumulation [194-200 and **I**]. Although important, the extent of these biological differences is still unclear, and needs to be studied in more detail. Nevertheless, the genetic differences between HIV-1 subtypes and CRFs add to the genetic variability of this pathogen, which is a major obstacle in the development of an effective HIV-1 vaccine. Furthermore, the continuous and rapid evolution of HIV-1 coupled with the human behavioral patterns will inevitably result in the rise and spread of new forms of this virus, forms with different and maybe even better adaptation and survival capabilities.

AIMS

The general aim of this thesis was to examine different genetic aspects of the HIV-1 evolution both within a patient and when the virus is transmitted *i.e.*, on the host population level. In more detail, the specific aims were to:

- Determine the prevalence and the characteristics of drug-resistance transmission in Sweden and investigate its impact on the evolution of HIV-1.
- Examine the evolution of HIV-1 phenotypes R5 and X4 when involved in transmission from mother to child and determine the origin of X4 variants in children.
- Investigate how different aspects of HIV-1 evolution influence phylogenetic reconstruction of epidemiological events, involving reconstruction of transmission events between patients, and sampling events within a patient.
- Examine how the evolution and the evolutionary rate of HIV-1 are affected by differential rate of spread in various types of epidemics found throughout the world.
- Construct a fast and easy method for estimation of evolutionary rate and a rooting point in a phylogenetic tree.

I Transmission of drug-resistant HIV-1 in Sweden: prevalence, evolution, and subtype specific patterns

The use of highly active antiretroviral therapy (HAART) has lead to reduction in disease progression and suppression of viral replication resulting in lower viral load in the majority of HIV-1 infected patients. The most common drugs used for HAART are the RT- and PR-inhibitors, targeting viral enzymes reverse transcriptase and protease, which are encoded by the *pol* gene. Mutations in the *pol* gene conferring drug resistance may arise as a consequence of neutral evolution or due to the selection pressure imposed on the viral population by the antiretroviral treatment. Loss of control of viral replication during therapy, mainly due to poor adherence, leads to outgrowth of viral variants carrying drug-resistance mutations to one or several antiretroviral drugs. It has been suggested that drug-resistant variants have lower transmissibility than the wild-type virus, however, transmission of drug-resistant HIV-1 has been shown to occur in 9-10% of new infections in Europe, and over 20% in some local and other regions where treatment is common [149-153, 155]. Although mutations associated with high level of resistance present an advantage for the viral population in an individual on treatment, these mutations confer significant loss of viral replication fitness when the pressure of antiretroviral drugs is absent. Because of this, HIV-1 drug-resistant variants are not expected to predominate in treatment-naïve individuals, and their presence in such environment indicates their transmission from individuals receiving therapy.

Transmission of drug resistant virus

In this paper, we examined the prevalence of transmission of drug-resistant HIV-1 by analyzing the RT and PR coding regions of the *pol* gene for drug resistance associated mutations. The sequences of the *pol* gene were derived from 100 newly infected and treatment-naïve patients in Sweden. The majority of the sequences, 91%, carried mutations of low or unclear resistance level. These mutations were, therefore, believed to be the result of neutral evolution of the HIV-1 population in treatment-naïve patients. In nine of the patients, on the other hand, the predominating HIV-1 populations were comprised of variants carrying mutations giving rise to high and intermediate level of resistance. In consistence with other studies, most of the transmitted resistance mutations were found to be directed to RT (8 patients) than to PR (1 patient) [149, 150] (Table 2). Primary mutations were found in five patients, and in two of those the impact of primary mutations was further increased by the presence of complementary secondary mutations. In the other four patients, only secondary resistance mutations were observed, however, at the time of the study these were scored as mutations giving rise to high or intermediate level of resistance. These nine patients were therefore believed to be infected by a virus that had experienced the pressure of antiretroviral treatment. Today, however, secondary mutations V75L and G190W are believed to occur naturally in a viral population and are usually not considered in genotypic assays. Thus, their presence in three of the nine patients is probably a result of neutral evolution. The estimated rate of transmission of drug-resistant HIV-1 variants in Sweden is therefore approximately 7%. This is consistent with the prevalence of transmitted drug-resistance in Sweden from a previous study, however, that study found that the most prevalent resistance mutations transmitted were those against the PR enzyme [201].

Risk group	*Subtype*	*Gene region*	*Resistance-associated mutation*
MSM	B	RT	K103N^{P}
MSM	B	RT	Y188H^{P}
MSM	B	RT	T69D^{P}, D67N, T215S
Unknown	CRF01_AE	PR	M46I^{P}, K20R, M36I, I93L
Heterosexual	B	RT	V118I^{P}
MSM	B	RT	V75L
MSM	B	RT	T69N
MSM	B	RT	M41L, T215SC
MSM	B	RT	G190W

Table 2. Nine patients with primary and secondary mutations associated with high and intermediate level of resistance. ^{P} primary mutations.

Evolution of resistant HIV-1 in the absence of treatment

The predominating HIV-1 populations in two of the above-described nine patients carried T215S and T215SC mutations in the RT coding region of the *pol* gene. These are believed to be a transition stage between the primary mutations 215Y/F and the wild-type amino acid T, and their presence suggested beginning of the reversion process in these individuals. In the absence of antiretroviral treatment, resistant HIV-1 strains may be outcompeted by better fit wild-type population available in the cells of immune system as a provirus. Because our patients were infected with resistant virus, however, no proviral wild-type variants existed in the cells. The only way that the HIV-1 populations in these two patients could recover from the fitness loss caused by the 215Y/F mutations, was to evolve back to wild-type themselves. However, it has been shown that viruses with transition stages in this position replicate effectively and are as susceptible to AZT as the wild-type, but require only one nucleotide change to become resistant while the wild-type requires two [159]. Therefore, it is possible that total reversion to the wild-type HIV-1 population in these patients may never occur, which could have devastating impact on the success of antiretroviral treatment. In addition, we cannot exclude the possibility that the reversion of resistant virus may have occurred in other sequences in our study, such as those with high and intermediate level-associated secondary mutations, which would lead to an underestimation of HIV-1 resistance transmission.

Subtypes and subtype specific patterns

Distribution of HIV-1 subtypes in our study was: A=3, B=55, C=29, D=2, G=1, CRF01_AE=9, and U=1, confirming previous finding that most of HIV-1 subtypes can be found in Sweden [202]. The majority of non-B subtypes originated from patients coming from, or being infected in, different regions of Sub-Saharan Africa and Thailand, while subtype B viruses originated from Europe and North America. Consistent with the findings of others, the majority of the transmitted resistant HIV-1 in our study was of subtype B [149, 150]. This was rather due to the fact that subtype B is mostly found in the industrialized world where antiretroviral treatment is most common, then to specific characteristics of non-B subtypes. Indeed, recent studies indicate that there is no significant difference in the rate of drug- resistance development, or the probability of transmission of drug-resistant HIV-1, between different subtypes of this virus [150, 203].

When the subtypes found in our study were compared, we found that some of the secondary naturally occurring mutations associated with low or unclear level of resistance were

occurring at different frequencies in different subtypes. This subtype-specific pattern was confirmed by an analysis of HIV-1 subtypes derived from the HIV sequence database. Although some of these subtype-specific resistance mutations have been described earlier, most studies show that there is no significant difference in virological response to therapy and primary and secondary resistance mutations used by B and non-B subtypes of HIV-1 [199, 200, 203-206]. Generally, all primary and secondary mutations can be found in both B and non-B subtypes. However, our study indicates that these mutations can appear at different frequencies in different subtypes, suggesting existence of several different evolutionary paths to drug resistance development.

II Transmission of HIV-1 from mother to child: evolution of X4 from R5

Although the use of antiretroviral treatment during pregnancy and at delivery substantially decreases the rate of transmission of HIV-1 from mother to child, the treatment is usually not available in developing countries where the rate of vertical transmission remains high. Early diagnosis and time of infection in children are usually difficult to estimate, complicated by the presence of maternal anti-HIV antibodies which cross the placenta to the fetus, and by temporary and often non-specific clinical symptoms of HIV infection in children. It has been suggested that a positive virus culture from infant blood sampled within the first 48 hours of life indicates transmission of HIV-1 *in utero*, associated with rapid progression of disease, while a negative culture at birth in children later proven to be HIV-1 positive indicates occurrence of infection late in pregnancy or at delivery, associated with a slower disease progression [207-209]. Similarly to infection in adults, HIV-1 populations in children show little variation in the beginning of an infection, and a preferential transmission of R5 variants is observed. In addition, a correlation of X4 emergence with disease progression has been confirmed in children [127]. Previous studies on the evolution of HIV-1 co-receptor usage and the switch from R5 to X4 in mother-child pairs have revealed a correlation between the emergence of X4 variants in children and the presence of X4 in their mothers. It was also observed that even in cases where the dominating viral population in mothers was of X4 phenotype, the virus establishing infection in children was R5. However, the R5 HIV-1 populations in children were generally outgrown by R5X4 and X4 variants at a later stage of infection [210]. In this study we further investigated the correlation between the X4 presence in mothers and their children by studying the origin of X4 variants found in the children.

Transmission of R5 phenotypes

HIV-1 sequences from single viral molecular clones and isolates covering the V3 region of the *env* gene were derived from two mother-child pairs (Ma-Ca and Mb-Cb). The samples were collected before, during, and after delivery for the mothers, and from 6 months of age to the age of 5 and 6.5 years, respectively, for the children. HIV-1 was not detected in the children at the time of birth, indicating that they became infected shortly before or at the time of delivery.

Phylogenetic analyses of maternal clones and isolates showed existence of several subpopulations of HIV-1 in the mothers, comprising of both R5 and X4 variants. Even though the dominating virus in the Mb showed no clear phenotype association in the tree, the sequence pattern of its clones indicated a closer relationship to R5 than to X4 phenotype. On the other hand, the dominating virus in Ma was clearly shown to be X4, although a minor R5 subpopulation existed. The transmission of R5 variants from the mothers to their children

was described in an earlier study, as the first viruses derived from the children were of R5 phenotype [211]. However, a clear relationship between maternal and child R5 variants became obvious when the earliest sampled child sequences were added to the maternal sequences in the phylogenetic trees. Sequences from Cb protruded from the Mb cluster of R5 variants and formed a homogenous cluster around an R5 isolate sequence. As expected, sequences from Ca showed closest relationship with the minor maternal R5 population. Interestingly, however, a maternal R5 isolate from the delivery time point clustered together with the sequences from the child, strongly indicating that the Ca was infected by a virus originating from the minor Ma R5 subpopulation around the time of delivery. The clearly preferential transmission of R5 viruses, together with homogenous HIV-1 populations in early infection in the infants, suggested a presence of a selective process during transmission and possibly the early phase of infection in children. Supporting this is the finding that HIV-1 infects CXCR4-expressing cells in placentas of non-transmitting mothers, while CCR5-expressing cells are almost exclusively infected in placentas from transmitting mothers [212]. Also, the macrophages derived from cord blood were shown to be more susceptible to R5 replication than the macrophages derived from adult peripheral blood [213].

Evolution of X4 from R5

The origin and evolution of X4 strains is not completely understood, however, two possible explanations exist, that do not necessarily exclude each other. One is *de novo* evolution of X4 variants from R5 in an infected individual, and the other is delayed emergence of transmitted X4 stored in cells at the time of infection. In our study, when child sequences sampled at later time points were added to the maternal sequences in the phylogenetic trees, they clearly clustered with maternal and child R5 subpopulations although they were of X4 phenotype. Thus, there was a strong indication that the X4 viruses in children had their origin in the R5 populations, and that they evolved independently of maternal X4 variants. In addition, comparison of V3-loop amino acid patterns in maternal and child sequences revealed similar patterns between the R5 viruses from the mothers and their children, while the X4 viruses showed striking differences. These results reinforced the possibility of unique evolution of child X4 variants, unrelated to transmission of maternal X4. The transmission could have still occurred giving rise to minor undetected X4 subpopulations in the children, however, this did not affect the conclusion that the major X4 populations in the children had evolved from their R5 variants. Similar results in the emergence and evolution of X4 in children have been shown by others, however, evidence of appearance of stored variants has also been presented [214, 215]. Several reasons to the co-receptor phenotype switch in adults have been proposed, and these were discussed earlier. In addition, it has been suggested that unlike in adults, the majority of CD4 T-lymphocytes in children express CXCR4 co-receptor on their surface. Predominance of these, and thus X4 viral variants in thymus, was proposed to play an important role in the evolution and emergence of CXCR4-tropic HIV-1 in children [216]. Since the mothers in our study carried X4 viruses, the co-receptor switch in the children could have also been predisposed by the common genetic factors of the viruses from the mothers and their children.

III Using viral phylogeny to reconstruct epidemiological events: the impact of hidden viral lineages on the reconstruction

Phylogenetic analyses have been used extensively in the studies of HIV-1, and have proven to be an invaluable tool, revealing important aspects of evolution, origins, disease

progression, and transmission of this virus. For instance, it was the use of phylogenetic analyses that unveiled SIV as a source of HIV infection in humans, and detected the existence and establishment of HIV-1 subtypes and recombinant forms throughout the world [37, 38, 40, 41, 166, 168]. In addition, phylogenetic analyses have been used to evaluate the evolution and evolutionary rate of this virus both within and among patients, to study the relationship of HIV-1 genetic divergence and diversity with disease progression, and to infer HIV transmission on several different levels, from the global spread to individual transmissions [34, 120, 139, 217-223]. As HIV-1 is a fast evolving organism, it will accumulate mutations within reasonable amount of time, and simultaneously preserve some common signature patterns with its ancestors, such that genetic relationships between viruses within and among patients can be reconstructed in phylogenetic trees. Previous studies on known HIV-1 transmission histories have shown that phylogenetic trees based on viral sequences correctly reflect known epidemiological events, indicating good reliability of these analyses [224-226]. However, it has been suggested that the complex nature of HIV-1 evolution within and between individuals may influence viral phylogenies in different ways, and therefore also influence a correct reconstruction and estimates of epidemiological events and viral evolutionary rate [119, 227, 228]. In this study, we used a well known and previously studied HIV-1 transmission chain [217, 224, 229], with added 10 years of sampling and thus a total of 50 years of viral separation time, to further investigate the agreement between viral phylogenies and known epidemiological events, and to examine possible features of HIV-1 evolution giving rise to discrepancies between virus and epidemiological trees (Figure 11).

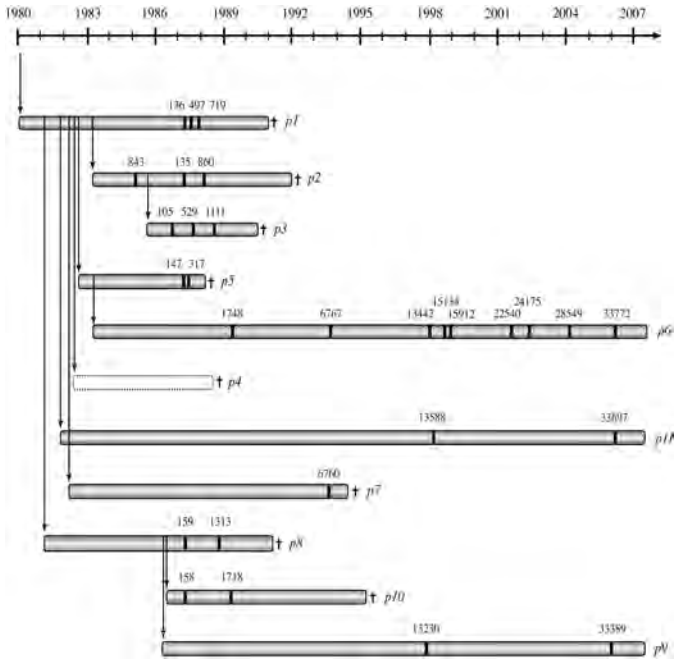


Figure 11. The Swedish transmission chain. Patients are shown as horizontal gray bars and marked with italic p and the corresponding patient number. Patient samples are shown as black vertical lines and their numbers are indicated in plain text above the corresponding line. Transmission and sampling events are shown according to the time-scale. Arrows represent transmissions. † represents patients that have died.

Hidden viral lineages affect reconstruction of epidemiological events
The Swedish transmission chain consists of 11 individuals where dates of infection and epidemiological links between the patients are known. Samples from 10 patients were available and the viral sequences covering the gp120 region of the *env* gene were used for analysis. A true epidemiological tree, reflecting the known epidemiological events, was compared to the maximum likelihood tree inferred from the virus sequences. The majority of epidemiological events were successfully reconstructed in spite of the fact that many transmission events occurred close in time to each other. However, the virus and the epidemiological tree topologies differed in two of the transmissions and in several sampling events.

During the course of infection, several viral lineages and subpopulations can exist within a patient. Some of these may be hidden at the time of sampling, because of the limitations in the sampling process, because they might have died out and been replaced by other lineages, or because they are in a latent stage. When taking the existence of hidden viral lineages into account, the observed topological discrepancies between the virus and the epidemiological trees were explained, and the trees were shown to be compatible. The limitations of inference of transmission events using viral sequences have been described previously, where it was shown that even though two sequences from two different patients cluster together in a tree, one cannot exclude the possibility of existence of a third epidemiological link, either infecting both of the patients or being an intermediate connection between them [230]. However, our study showed that even if two patients infect each other, they still might not be directly connected in a phylogenetic tree. This finding becomes especially significant for forensic analyses, which have been used extensively for investigations of HIV-1 transmission between individuals [222, 223, 231, 232]. In addition to their effect on reconstruction of transmission events, in our study we also showed that hidden viral lineages might influence the reconstruction of sampling events, as the order in which samples were taken within a patient was not matched by their order in our inferred tree. The divergence of these "misplaced" viral sequences *i.e.* their genetic distance from the infection event, was often, but not always, observed too low compared to the surrounding sequences from the same patient. However, too low divergence was also detected for some of the sequences placed in the correct order in the virus tree. These temporal inconsistencies in the patterns of HIV-1 evolution have been observed previously, where viruses obtained at a later time point showed lower divergence from the most recent common ancestor (MRCA) than viruses obtained earlier in an infected individual. The temporal inconsistencies were suggested to be due to activation of, or sampling from, the latent reservoir of HIV-1, and associated to genetic drift and purifying evolution acting on the viral population [233-235]. In our study, however, we also found few sequences with larger divergence than the surrounding sequences. This indicated on the existence of an additional factor influencing inference of genetic distances: the biological sampling error. The error might arise when different viral subpopulations within a patient have diverged different amounts from their MRCA, and when the frequencies of these subpopulations vary over time. While the biological sampling error can result in too high as well as too low than expected divergence, the amount of its contribution to the lower than expected divergence was masked by the impact of viral latency. The influence of hidden viral lineages on genetic distances in a phylogenetic tree, where the viral divergence does not correlate with the sampling time, can cause difficulties in estimations of a correct evolutionary rate.

Reconstruction of the correct epidemiological root by several different rooting methods proved to be difficult. None of the methods succeeded in finding the true root and they often suggested roots differing from each other as well. Once again, this could be explained by the existence of hidden viral lineages. Thus, hidden viral lineages may influence the inference of transmissions, sampling events, and cause temporal inconsistencies, as well as have impact on the estimation of the root in a phylogenetic tree. Further studies to uncover the exact impact and individual contribution of the different types of hidden lineages on the inference of phylogenetic trees are therefore necessary.

Evolutionary rate and disease progression

Comparison of the sequence genetic distances between normalized and weighted epidemiological and virus trees revealed that the distance of viral sequences measured from the infection events of three of the patients was consistently lower than the expected distance derived from the epidemiological tree. Interestingly, these three patients were the only survivors from the transmission chain. Calculation of the evolutionary rates confirmed that the amount of HIV-1 evolution was lower in these patients. One of the patients was heterozygous for the Δ32 CCR5 mutation, a mutation that has been associated with slower disease progression and, in some instances, poor and delayed virus kinetics [236, 237]. The other two patients had been receiving HAART, which has been shown to slow down and even totally abolish the replication and evolution of HIV-1 [234, 238, 239]. Although some studies suggest a correlation between viral evolutionary rate and disease progression, such correlation was not observed in our patients with exception of the three survivors. In addition, the significantly lower evolutionary rate in the survivors was possibly a consequence of the lower replication rate of the viral populations within these patients, due to the Δ32 mutation and HAART. Thus, the lower evolutionary rate of HIV-1 in these patients is probably correlated to, but not a direct cause of, their long-term survival.

IV Unequal evolutionary rates in HIV-1 epidemics: the rate of spread influences the rate of viral evolution

The rate of spread of HIV-1 is unequal throughout the world. Slow spread is mainly found in sexual transmissions of subtype B in Western Europe and North America and subtypes A and C in Africa. Fast spread, on the other hand, has been observed in IDU networks of the Former Soviet Union (FSU) and Asia involving several different subtypes of HIV-1 [220, 240-243]. Strikingly, virus sequences from the fast-spreading IDU HIV-1 epidemics have shown very high genetic similarity to each other, while this has not been observed for the sequences of slow-spreading viruses. When a virus is introduced into an already existing social network, such as those of IDUs, it will spread rapidly from person to person within the network. Here, HIV-1 will probably be transmitted while the host is in the first (primary) stage of infection in which the pressure of immune system is low or non-existent and the viral population is very homogenous. Consequently, highly similar viruses will be spread across the network, as observed in sequences collected from IDUs. Such a scenario is expected to result in a low evolutionary rate of HIV-1 on the host population level, reflecting the rate of a virus not subjected to the pressure of human immune system. In heterosexual transmission, on the other hand, the networks are mostly non-existent and the virus spreads as new contacts are formed. As this usually takes longer time, HIV-1 will most probably be transmitted when the host is in the second (chronic) stage of infection. Here, the immune response to HIV-1 infection is active, and the viral population shows high level of diversity, resulting in the

spread of genetically different viruses across the host population. The rate of evolution of HIV-1 on the host population level is thus expected to be higher. To examine whether the rate of spread influences the rate of evolution of HIV-1, we collected sequences from the Los Alamos HIV sequence database belonging to six different HIV-1 epidemics found throughout the world. The sequences covered the V3 region of the *env* gene and were of known sampling dates (Table 3).

A method for root and rate optimization

For this analysis we constructed a method for estimation of evolutionary rate and root in a phylogenetic tree. The method was based on a t-test, calculating genetic distance between two groups of taxa sampled at two different known time points. The genetic distance between the two groups was derived by subtracting the average genetic distance from the root to the first time point taxa from the average genetic distance from the root to the second time point taxa. In this way, errors in the tree prior to sequences from the first time point were essentially removed. Since the root of the tree needs to be known for the calculation, the method estimates all possible roots of the tree and chooses the best one based on maximized distance between the two groups of taxa with minimized variance within the groups. The advantages of this method include that it is fast and can be used for large datasets, it is user friendly and easily accessible as it will be available through the HIV sequence database website. In addition, it can use a phylogenetic tree calculated with or without molecular clock, and it gives the possibility of finding and comparing genetic distances and node heights for all possible rooting points of the tree. When tested in simulations, the method proved to estimate a correct genetic distance between the two groups of taxa. An exception was when this genetic distance was very small, in which case the method tended to somewhat overestimate the distance.

Evolutionary rate of HIV-1 from different epidemics

The evolutionary rate of HIV-1 subtype A1 spreading rapidly in IDU standing social networks in the FSU was estimated approximately 8.4 times lower than the evolutionary rate of the same subtype spreading slowly in the African heterosexual population (Table 3).

Region	*Epidemic*	*Subtype*	*Evolutionary rate (10^{-3} subst. site^{-1} year^{-1})*
FSU	IDU	A1	2.02
Southeast Asia	Mixed[A]	B'	10.8
Southeast Asia	Mixed[A]	CRF01_AE	8.32
Africa	Heterosexual	A1	16.9
Europe and North America	Heterosexual and MSM[B]	B	4.3
Africa	Heterosexual	C	9.65

Table 3. HIV-1 epidemics and evolutionary rates. [A] IDU and heterosexual, only IDU sequences were used for the analyses, [B] only heterosexual sequences were used for the analyses.

These results suggested that a slow spread of HIV-1 indeed correlated with most of the transmissions occurring in the chronic phase of infection, thus increasing viral diversity and rate of evolution on the host population level. On the other hand, the rapid spread of the virus in the FSU resulted in transmission of homogenous variants during primary infection, and decreased viral diversity and evolutionary rate within the epidemic. These findings were further confirmed when mixed epidemics, consisting of both rapid and slow spread, were

studied. An example of mixed epidemics is that of subtype B', or CRF01_AE, both found to circulate among IDUs, commercial sex workers, and their customers in Asia. Here, the virus was spreading through standing social networks and through forming of new contacts, and was thus transmitted during both primary and chronic stages of infection. Therefore, the evolutionary rate of HIV-1 in mixed epidemics was expected, and was found to be, an intermediate between the evolutionary rates of slow and fast spreading viruses.

Interestingly, the evolutionary rate of subtype C spreading in Africa was estimated 1.7 times lower than the evolutionary rate of subtype A from the same region. As both subtypes are spreading in a slow heterosexual manner, it would be expected that the diversity and evolutionary rates of these two viruses would be equal. However, several studies have indicated that subtype C in Africa is spreading more rapidly than the other subtypes in this region [244-246]. Consequently, a somewhat higher proportion of subtype C infections probably occur during the first stage of HIV-1 infection, resulting in an intermediate evolutionary rate of this virus on the host population level. An additional influence to the lower than expected rate of evolution of the subtype C might come from the conservation of its V3 loop, which is otherwise highly variable [246, 247]. However, if existent, this influence is minimal as the highly variable region downstream of the V3 loop was partially included in our sequences. The evolutionary rate of subtype B HIV-1 spreading in Western Europe and North America showed to be unusually low. However, subtype B differs substantially from the other subtypes by the fact that it is subjected to the effect of antiretroviral treatment most common in these parts of the world. It has been shown that effective antiretroviral treatment slows down the evolution of HIV-1 within a patient. This effect might thus be mirrored in the lower evolutionary rate of subtype B on the host population level, but due to very few sequences available from the period prior to the use of antiretroviral treatment this could not be further investigated. A possible factor adding to the unexpectedly low evolutionary rate of subtype B virus is a combination of its initial rapid spread within specific risk groups, followed by a slower spread later on. In addition, it has recently been shown that at least 25% of HIV-1 transmissions among MSM in London occurred within a few months of infection, suggesting a faster rate of spread of subtype B in this risk group [248].

Our results were further confirmed by an epidemiological model, showing that the evolutionary rate of a virus spreading rapidly, with most of the transmissions occurring soon after infection, was lower than the evolutionary rate of a virus spreading in a slow manner. The model also indicated that the rate of evolution of HIV-1 in a fast epidemic would increase as the epidemic ages and the host population reaches infection saturation. At this point, a higher proportion of newly infected individuals become infected by someone in the chronic stage of infection, increasing viral diversity and rate of evolution on the host population level. Moreover, re-sampling of individuals that now have been infected for a while becomes more frequent, and the evolutionary rate on the population level increases. Thus, we predict that the low rate of evolution of HIV-1 in the IDU population of the FSU will eventually increase as this epidemic ages and the population becomes more saturated.

Although our method tended to overestimate the evolutionary rate in cases where the genetic distance between the two groups of taxa was small, the incidence of such cases in our data was very low and they were not expected to affect the average estimate of the rate. This was confirmed by the Bayesian Markov Chain Monte Carlo analyses of the sequences from the fast and slow epidemics, which showed similar rate estimations. Our results indicated that the

rate of spread of HIV-1, correlated to the stage of infection of the host at the time of transmission and social behavior of the affected human population, influences the evolutionary rate of this virus on the population level. Furthermore, we showed that large differences exist in the rate of evolution of HIV-1 among different epidemics throughout the world, and that these rates can change depending on the phase and size of an individual epidemic. These results also become important to consider in the estimation of time of origin of HIV-1 infection in different parts of the world.

V A fast method for optimization of root and evolutionary rate in a phylogenetic tree

Fast evolving RNA viruses, such as HIV-1, are well suited for evolutionary studies as they accumulate enough mutations within reasonable amount of time. An important component of evolution is the evolutionary rate i.e. the rate at which substitutions occur over time. HIV-1 is one of the fastest evolving organisms known, with an evolutionary rate estimated to be approximately $1x10^{-3}$ substitutions $site^{-1}$ $year^{-1}$ in the *env* gene, and an even higher rate in the V3 region of this gene, estimated between 2 and $17x10^{-3}$ substitutions $site^{-1}$ $year^{-1}$ [34, 217, 249]. Initially, estimations of evolutionary rate were based on an assumption of a strict molecular clock, which is a constant accumulation of substitutions per unit of time [250, 251]. However, this assumption was suggested to be an oversimplification and more suitable methods that relaxed the molecular clock have been developed [252-257]. As these methods are more complicated and more time consuming, the need for a simple and fast method that simultaneously is able to manage large amount of data arises. In this study we describe development of a method for optimization of the root and estimation of the evolutionary rate in a phylogenetic tree with samples from at least two different time points. The rate can be estimated on a tree built with any existent method, including trees built with or without the assumption of a molecular clock. An earlier version of this method was used for estimation of evolutionary rate of HIV-1 on the population level (paper IV). In this paper, we describe an improved version of the method rigorously tested on simulated data.

Optimizing the root

Rooting of the tree is essential for the estimation of an evolutionary rate in our method, as the genetic distance between taxa from two time points ($\Delta\hat{d}$) is calculated as the average genetic distance from the root of the tree to time point 2 taxa minus the average genetic distance from the root to time point 1 taxa. Therefore several different statistics were tested in their ability to estimate the correct root. The best performing criteria for finding the correct root was minimizing the sum of the tip height variances of both groups of taxa (MSV), while the criterion used in the previous study, maximizing $\Delta\hat{d}$ while minimizing variances within the groups (MWP), performed less well at low $\Delta\hat{d}$s and smaller total tree heights. Finding the true root was shown to be dependent of $\Delta\hat{d}$, sequence length, number of taxa in groups 1 and 2, and total tree height. Although the correct root was always found in trees with high $\Delta\hat{d}$s and taxa consisting of very long sequences (100 000bp), in general, the method was not very efficient in finding the true root. However, this result is not surprising for trees containing little information, such as short sequences and low $\Delta\hat{d}$s, as the branches on such trees will be unresolved and the true topology and thus the correct root, will be impossible to find.

Estimating the evolutionary rate

In spite of the fact that the true root could not be estimated in instances where the amount of genetic information was limited, this did not affect the ability of our method to estimate a correct evolutionary rate. In such instances, branches closest to the true root are very short and even though the root gets misplaced to any of these branches, this will not contribute greatly to the total sum of distances within the groups, thus not markedly affecting the value of the evolutionary rate. In addition to being robust to errors in the estimation of the true root, the method proved to estimate a correct evolutionary rate even in cases where the topology of the tree was not correct. Although reconstruction of tree topology was less accurate in trees with low separation between the groups, the estimated rate of evolution was shown not to be affected, being within 10% of the true rate. In a real situation, when a phylogenetic tree is inferred based on sequence information, it is impossible to know whether the tree has a correct topology and a correct rooting point. Here we show that an accurate evolutionary rate can be estimated in such trees.

Our method, based on the MSV criterion, was unbiased in its average estimation of $\Delta\hat{d}$, and thus the rate of evolution. The variance around the average evolutionary rate was dependent on the separation between the two groups, number of taxa in the groups, sequence length and the height of the tree, where it decreased with increasing values of these parameters. The expected error in the evolutionary rate estimate from a single tree analysis was less than 10% when the separation between the groups was 0.01 substitutions site^{-1} or higher. As it was observed in paper IV, the MWP criterion tended to overestimate the evolutionary rate when the separation was less then 0.003 substitutions site^{-1}. Although some overestimation is expected due to the Skellam distribution, which arises when two Poisson distributed variables are subtracted from each other, this effect showed to be very low and drowned in the phylogenetic noise [258]. Thus it appears that, in instances where the separation between the two groups of taxa in a tree is very small, the MWP tends to be biased towards less defined distributions with greater $\Delta\hat{d}$s, and thereby overestimate the rate of evolution. In analyses of fast evolving organisms, such as the analysis performed on HIV-1 in paper IV, the incidence of such small separations was low, and thus did not markedly affect the estimations. The overestimation of evolutionary rate was not observed in the improved version of the method, where the average rate was estimated correctly even at low group separations. Thus the method is robust to errors in root placement and incorrect tree topologies, and accurately estimates the rate of evolution between two groups of taxa sampled at different time points. It is fast and user friendly, and it is able to handle large amounts of data. Although it requires prior computation of a phylogenetic tree, the taxons of the tree can be assigned and re-assigned differently to the two time point groups, and even chosen not to be included into the calculation. In this way several different estimations can be made without the need of re-constructing the tree.

CONCLUSIONS AND FUTURE PERSPECTIVES

HIV-1 is a highly variable and fast evolving organism with great adaptation capabilities. This thesis summarizes several studies used to investigate the evolutionary features of this virus both within a patient and in context of transmission.

Transmission of drug resistant HIV-1 was shown to occur in approximately 7% of new infections in Sweden and the predominating subtype in such infections was subtype B. Its predominance is believed to be due to the higher occurrence of antiretroviral treatment in regions where this subtype is most common, rather then to the existence of specific characteristics of different subtypes of HIV-1. Generally, all resistance-associated mutations can be found in both B and non-B subtypes, and no difference in drug susceptibility between the subtypes exists. However, our study revealed subtype-specific patterns of the frequencies at which the resistance-associated mutations occur. The existence of such patterns indicated to the presence of several different evolutionary paths that this virus can use for the development of drug resistance. The exact relevance of such paths and subtype specific patterns remains to be examined. In addition, the discovery of transition mutations at sites associated with high-level drug resistance in two of the patients suggested initiation of the reversion process within their viral populations. However, whether a total reversion occurs or the populations stay in the more favorable transition stage, leading to a possible faster development of resistance during antiretroviral treatment, remains to be further evaluated.

The presence of X4 HIV-1 phenotypes in children initially infected with R5 viruses from their mothers was shown to be due to the independent evolution of these X4 variants from the child R5 populations, and was not correlated to transmission. Although the dominating HIV-1 population in one of the mothers was of X4 phenotype, a minor R5 variant was transmitted to the child, from which the X4 population in the child evolved. Furthermore, the preferential transmission of R5 viruses indicated presence of a selection mechanism favoring transmission or outgrowth of CCR5-tropic viruses in children. Several reasons to the favored R5 transmission in both children and adults have been suggested, however, the exact explanation to this phenomenon remains to be determined. Another unknown factor of HIV-1 phenotype evolution is the switch of co-receptor use associated with disease progression. Although we showed that the X4 populations in the children were not derived from their mother's X4 viruses, the common genetic background of the virus populations in the mothers and their children could have predisposed for the phenotype switch. Among other factors that have been suggested to contribute to the change in co-receptor use are activation of resting T-cells and virus replication fitness, however, the exact cause remains unknown and is in need of further investigation.

Reconstruction of epidemiological events by inferring phylogenetic trees based on viral sequences was shown to be reliable, however, the few discrepancies between the epidemiological and virus trees revealed the importance of hidden viral lineages on the reconstruction. The definition of hidden viral lineages included lineages not sampled due to latency, extinction, or due to limitations in the sampling process. When the effect of hidden viral lineages was taken into account the virus and the epidemiological tree were shown to be completely compatible. The hidden lineages had impact on the reconstruction of transmission and sampling events, as well as on the reconstruction of the correct root. For instance, we showed that two patients not directly connected in an inferred virus tree still might have

infected each other, a finding highly relevant for forensic purposes. In addition, hidden viral lineages may affect divergence of a sample, such that latent or re-activated lineages have lower amount of accumulated substitutions than expected from their sampling time, directly affecting correct estimations of viral evolutionary rate. Thus, more detailed studies are important to unveil how much and in what way different types of hidden viral lineages can influence an inferred phylogenetic tree.

On the host population level, the evolutionary rate of HIV-1 was shown to depend on the rate of spread of the virus in an epidemic. If the virus spreads rapidly, such as in IDU standing social networks, it will transfer from person to person in the early stage of infection where the pressure of the immune system is low and the viral population is homogenous. This was reflected on the host population level, where the viral variants spreading were very similar to each other and the rate of evolution was low. In a heterosexual epidemic, on the other hand, HIV-1 spreads as new contacts are formed, mostly in the chronic phase of infection. The viral variants here have experienced the pressure of human immune system and started to diverge, which was observed as a much higher rate of evolution of HIV-1. As expected, variants spreading in mixed epidemics i.e. epidemics composed of both rapid and slow HIV-1 spread, had an intermediate rate of evolution. Epidemiological modeling confirmed our results, and indicated that a slow evolutionary rate in a rapid spread will increase as an epidemic ages and the population in which it spreads gets saturated with infections. These findings are important to consider when calculating the origin of HIV-1 epidemic.

A fast and simple method for optimization of a root and estimation of an evolutionary rate between samples taken at two different time points in a phylogenetic tree was developed. Simulations showed that the method had no bias in estimations of the evolutionary rate and that the variance around the mean depended on the sequence length, the genetic distance between the two groups of taxa, the number of taxa in the two groups, and the total tree height. Finding a correct root proved to be difficult, although the true root could be found in 100% of the cases where the sequence length and the genetic distance between the two groups were large. However, even in cases where the correct root was not found the rate of evolution could be accurately estimated. Furthermore, simulations on uncertain tree topologies revealed that even though the topology of a tree is not reconstructed entirely correctly, the estimated rate of evolution is still accurate. These results showed that our method, besides being fast and able to handle large amounts of data, is robust and reliable when used in real situations where the root and the topology of a phylogenetic tree are uncertain.

The studies in this thesis all contribute with insights into the complex nature of HIV-1 evolution, both within a patient and on the population level. Direct implications include the finding of transmission of resistant variants in Sweden and their reversion to transition stages, strengthening the importance of routine genetic testing for resistance in newly infected patients. The finding of diverse effects of hidden viral lineages on the reconstruction of transmission events between individuals is very relevant for forensic purposes, and the variations of evolutionary rate of HIV-1 on the population level in different epidemics might be proven to be relevant in estimations of the time of the origin of HIV-1, as well as have impact on vaccine strategies. In general, these studies improve our knowledge about this very diverse and highly adaptable virus, and hopefully take us a few steps closer to overcoming the challenges that stand in the way of construction of an effective vaccine or cure against HIV-1.

ACKNOWLEDGMENTS

Thomas Leitner, words cannot express how grateful I am for everything you have done for me. Thank you for showing me what it means to be a good scientist and for sharing your knowledge, thank you for being such a good teacher and supervisor, for your patience and for believing in me. I hope to some day become as good scientist as you are, and even maybe as good supervisor to someone as you have been to me. I am looking forward to future collaboration and mutual projects, working with you is always rewarding both in terms of interesting discussions and new ideas, and in gaining knowledge about HIV in particular and science in general.

Jan Albert, thank you for your support and guidance, for interesting discussions and good advice, for taking me in as your student and for being such a good co-supervisor.

Åke Lundkvist, thank you for being my co-supervisor during the first part of my PhD studies, for your support and for giving me the opportunity to diversify my work on viruses.

Anneka Ehrnst, thank you for giving me the opportunity to do my exam work with you and introducing me to work in the P3-laboratory.

I would also like to thank all the *co-authors* of the papers in this thesis, and all *the patients* that contributed with blood samples and gave their consent for these studies. Without you, none of this would have been possible to do.

To all the present and former *staff* at the virology department at SMI, especially *Kajsa, Ellen, Afsaneh, Maj, Gunnel* and *Sirkka*, thank you for your help and for making working at SMI enjoyable.

All the present and former *PhD students*, especially *Jonas K, Peter C, Karin W, Dace B, Jonas H, Andreas B, Sara Å and Helena S,* it has been a pleasure working with you.

All the *members* of the T-10 group at the Los Alamos National Laboratory, especially *Bette Korber*, *Brian Foley* and *Ruy Ribeiro*, thank you for helping me and giving me the opportunity to work with you all.

Britta Wahren, thank you for examining my knowledge in virology and for your constructive criticism driving me to better myself in this field.

The *staff* at MTC, especially *Anna Logdberg* and *Lena Norenius*, thank you for your help and for making seemingly impossible things possible.

I would also like to thank all the *organizations* providing funding for my research and making completion of this thesis possible.

To the examination board, previous and current, including: *Anna-Lena Spetz, Helene Norder, Sven Britton, Matti Sällberg, Stefan Schwartz* and *Andy Leigh Brown*, thank you for your time and expertise for my licentiate and PhD degrees.

To all my friends, especially *Sait K, Ulrika J, Sofia W* and *Jennie H*, thank you for your support, generosity, and for all the fun that we have had.

Special gratitude goes to my family:

My parents *Olga* and *Vladimir*, my sister *Jovana*, my grandmother *Kova*, aunt *Liljana*, uncle *Mladen*, cousins *Bojan* and *Boris*, and other aunts and cousins, thank you for your endless love and support and for always being there for me. I miss you all and I wish we didn't live so far away from each other. My late aunt *Mira,* your words of encouragement during difficult times will never be forgotten. My grandfather *Milovan*, I will love and miss you forever.

My extended family, *Kathy* and *Paul Berry*, *Elizabeth, Heather* and *Krista,* with husbands and children, thank you for welcoming me into your family, for all your help and all the fun times.

Mateo, my favorite little guy, thank you for coming to us and for making our lives so much more beautiful and full of joy. Thank you for being such a good little baby, for all the adorable smiles and baby-hugs, for all the cute laughs and funny things you do. Mommy loves you very much.

Michael, my love, you have been absolutely incredibly amazing and there are not enough words in the world to express my love, appreciation and gratitude for you. Thank you for sharing your life with me, for your patience and understanding, for your support and advice, for all the fun that we have had, and for always being there for me when I need you. I love you so much and I look forward to our life together.

REFERENCES

1. Hymes, K.B. *et al*. 1981. *Kaposi's sarcoma in homosexual men-a report of eight cases*. Lancet 2:598-600.

2. Centers for Disease Control (CDC). 1981. *Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California*. MMWR Morb. Mortal. Wkly Rep. 30:305-8.

3. Centers for Disease Control (CDC). 1981. *Pneumocystis pneumonia--Los Angeles*. MMWR Morb. Mortal. Wkly Rep. 30:250-2.

4. Masur, H. *et al*. 1981. *An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction*. N. Engl. J. Med. 305:1431-8.

5. Centers for Disease Control (CDC). 1982. *Opportunistic infections and Kaposi's sarcoma among Haitians in the United States*. MMWR Morb. Mortal. Wkly Rep. 31:353-4, 360-1.

6. Centers for Disease Control (CDC). 1982. *Pneumocystis carinii pneumonia among persons with hemophilia A*. MMWR Morb. Mortal. Wkly Rep. 31:365-7.

7. du Bois, R.M., Branthwaite, M.A., Mikhail, J.R., and Batten, J.C. 1981. *Primary Pneumocystis carnii and cytomegalovirus infections*. Lancet 2:1339.

8. Gerstoft, J. *et al*. 1982. *Severe acquired immunodeficiency in European homosexual men*. Br. Med. J. (Clin. Res. Ed.) 285:17-9.

9. Vilaseca, J., Arnau, J.M., Bacardi, R., Mieras, C., Serrano, A., and Navarro, C. 1982. *Kaposi's sarcoma and toxoplasma gondii brain abscess in a Spanish homosexual*. Lancet 1:572.

10. Rozenbaum, W., Coulaud, J.P., Saimot, A.G., Klatzmann, D., Mayaud, C., and Carette, M.F. 1982. *Multiple opportunistic infection in a male homosexual in France*. Lancet 1:572-3.

11. Francioli, P. *et al*. 1982. *Acquired immunologic deficiency syndrome, opportunistic infections and homosexuality. Presentation of 3 cases studied in Switzerland*. Schweiz Med. Wochenschr. 112:1682-7.

12. Marx, J.L. 1982. *New disease baffles medical community*. Science 217:618-21.

13. Centers for Disease Control (CDC). 1982. *Update on acquired immune deficiency syndrome (AIDS)--United States*. MMWR Morb. Mortal. Wkly Rep. 31:507-8, 513-4.

14. Centers for Disease Control (CDC). 1982. *Possible transfusion-associated acquired immune deficiency syndrome (AIDS) - California*. MMWR Morb. Mortal. Wkly Rep. 31:652-4.

15. Centers for Disease Control (CDC). 1982. *Unexplained immunodeficiency and opportunistic infections in infants--New York, New Jersey, California*. MMWR Morb. Mortal. Wkly Rep. 31:665-7.

16. Centers for Disease Control (CDC). 1983. *Immunodeficiency among female sexual partners of males with acquired immune deficiency syndrome (AIDS) - New York*. MMWR Morb. Mortal. Wkly Rep. 31:697-8.

17. Harris, C. *et al*. 1983. *Immunodeficiency in female sexual partners of men with the acquired immunodeficiency syndrome*. N. Engl. J. Med. 308:1181-4.

18. Weller, I. *et al*. 1984. *Homosexual men in London: lymphadenopathy, immune status, and Epstein-Barr virus infection*. Ann. N. Y. Acad. Sci. 437:238-53.

19. Clumeck, N. *et al*. 1984. *Acquired immunodeficiency syndrome in African patients*. N. Engl. J. Med. 310:492-7.

20. Glauser, M.P. and Francioli, P. 1984. *Clinical and epidemiological survey of acquired immune deficiency syndrome in Europe*. Eur. J. Clin. Microbiol. 3:55-8.

21. Barre-Sinoussi, F. *et al*. 1983. *Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)*. Science 220:868-71.

22. Gallo, R.C. *et al*. 1984. *Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS*. Science 224:500-3.

23. Popovic, M., Sarngadharan, M.G., Read, E., and Gallo, R.C. 1984. *Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS*. Science 224:497-500.

24. Ratner, L., Gallo, R.C., and Wong-Staal, F. 1985. *HTLV-III, LAV, ARV are variants of same AIDS virus*. Nature 313:636-7.

25. Marx, J.L. 1985. *A virus by other name*. Science 227:1449-51.

26. Coffin, J. *et al*. 1986. *What to call the AIDS virus?* Nature 321:10.

27. UNAIDS 2007. *2007 AIDS epidemic update*. http://www.unaids.org/en/.

28. Barin, F. *et al*. 1985. *Serological evidence for virus related to simian T-lymphotropic retrovirus III in residents of west Africa*. Lancet 2:1387-9.

29. Clavel, F. *et al*. 1986. *Isolation of a new human retrovirus from West African patients with AIDS*. Science 233:343-6.

30. Reeves, J.D. and Doms, R.W. 2002. *Human immunodeficiency virus type 2*. J. Gen. Virol. 83:1253-65.

31. Schim van der Loeff, M.F. and Aaby, P. 1999. *Towards a better understanding of the epidemiology of HIV-2*. AIDS 13:S69-84.

32. Mann, J.M., *AIDS: A worldwide pandemic*., in *Current topics in AIDS, volume 2*, M.S. Gottlieb, et al., Editors. 1989, John Wiley & Sons: Chichester, UK.

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

34. Korber, B. *et al*. 2000. *Timing the ancestor of the HIV-1 pandemic strains*. Science 288:1789-96.

35. Lemey, P. *et al*. 2004. *The molecular population genetics of HIV-1 group O*. Genetics 167:1059-68.

36. Lemey, P., Pybus, O.G., Wang, B., Saksena, N.K., Salemi, M., and Vandamme, A.M. 2003. *Tracing the origin and history of the HIV-2 epidemic*. Proc. Natl. Acad. Sci. USA 100:6588-92.

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

38. Gao, F. *et al*. 1992. *Human infection by genetically diverse SIVSM-related HIV-2 in west Africa*. Nature 358:495-9.

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

40. Gao, F. *et al*. 1999. *Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes*. Nature 397:436-41.

41. Keele, B.F. *et al*. 2006. *Chimpanzee reservoirs of pandemic and nonpandemic HIV-1*. Science 313:523-6.

42. Bailes, E. *et al*. 2003. *Hybrid origin of SIV in chimpanzees*. Science 300:1713.

43. Leitner, T., Dazza, M.C., Ekwalanga, M., Apetrei, C., and Saragosti, S. 2007. *Sequence diversity among chimpanzee simian immunodeficiency viruses (SIVcpz) suggests that SIVcpzPts was derived from SIVcpzPtt through additional recombination events*. AIDS Res Hum Retroviruses 23:1114-8.

44. Wolfe, N.D. *et al*. 2004. *Naturally acquired simian retrovirus infections in central African hunters*. Lancet 363:932-7.

45. Blancou, P. *et al*. 2001. *Polio vaccine samples not linked to AIDS*. Nature 410:1045-6.

46. Berry, N. *et al*. 2001. *Vaccine safety: Analysis of oral polio vaccine CHAT stocks*. Nature 410:1046-7.

47. Donegan, E. *et al*. 1990. *Infection with human immunodeficiency virus type 1 (HIV-1) among recipients of antibody-positive blood donations*. Ann. Intern. Med. 113:733-9.

48. WHO 2000. *Blood safety.for too few*. http://www.who.int/inf-pr-2000/en/pr2000-25.html.

49. Kaplan, E.H. and Heimer, R. 1992. *A model-based estimate of HIV infectivity via needle sharing*. J. Acquir. Immune. Defic. Syndr. 5:1116-8.

50. Kaplan, E.H. and Heimer, R. 1995. *HIV incidence among New Haven needle exchange participants: updated estimates from syringe tracking and testing data*. J. Acquir. Immune. Defic. Syndr. Hum. Retrovir. 10:175-6.

51. Trace, M., Riley, D., and Stimson, G. 2005. *UNAIDS & the prevention of HIV infection through injecting drug use*. The Beckley Foundation Drug Policy Programme.

52. Aceijas, C., Stimson, G.V., Hickman, M., Rhodes, T., and United Nations Reference Group on HIV/AIDS Prevention and Care among IDU in Developing Transitional Countries 2004. *Global overview of injecting drug use and HIV among injecting drug users*. AIDS 18:2295-303.

53. Varghese, B., Maher, J.E., Peterman, T.A., Branson, B.M., and Steketee, R.W. 2002. *Reducing the risk of sexual HIV transmission: quantifying the per-act risk for HIV on the basis of choice of partner, sex act, and condom use*. Sex. Transm. Dis. 29:38-43.

54. 1992. *Comparison of female to male and male to female transmission of HIV in 563 stable couples. European Study Group on Heterosexual Transmission of HIV*. BMJ 304:809-13.

55. Quinn, T.C. *et al*. 2000. *Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group*. N. Engl. J. Med. 342:921-9.

56. Laga, M., Taelman, H., Van der Stuyft, P., Bonneux, L., Vercauteren, G., and Piot, P. 1989. *Advanced Immunodeficiency as a risk factor for heterosexual transmission of HIV*. AIDS 3:361-6.

57. Cohen, M.S. 2004. *HIV and sexually transmitted diseases: lethal synergy*. Top. HIV Med. 12:104-7.

58. Coovadia, H. 2004. *Antiretroviral agents--how best to protect infants from HIV and save their mothers from AIDS*. N. Engl. J. Med. 351:289-92.

59. Ryder, R.W. *et al*. 1989. *Perinatal transmission of the human immunodeficiency virus type 1 to infants of seropositive women in Zaire*. N. Engl. J. Med. 320:1637-42.

60. Herzmann, C. and Karcher, H. 2004. *Nevirapine plus zidovudine to prevent mother-to-child transmission of HIV*. N. Engl. J. Med. 351:2013-5.

61. Kind, C. *et al*. 1998. *Prevention of vertical HIV transmission: additive protective effect of elective Cesarean section and zidovudine prophylaxis. Swiss Neonatal HIV Study Group*. AIDS 12:205-10.

62. Richardson, B.A. *et al*. 2003. *Breast-milk infectivity in human immunodeficiency virus type 1-infected mothers*. J. Infect. Dis. 187:736-40.

63. Fang, G. *et al*. 1995. *Maternal plasma human immunodeficiency virus type 1 RNA level: a determinant and projected threshold for mother-to-child transmission*. Proc. Natl. Acad. Sci. USA 92:12100-4.

64. Scarlatti, G. *et al*. 1993. *Mother-to-child transmission of human immunodeficiency virus type 1: correlation with neutralizing antibodies against primary isolates*. J. Infect. Dis. 168:207-10.

65. Bryson, Y.J. 1996. *Perinatal HIV-1 transmission: recent advances and therapeutic interventions*. AIDS 10:S33-42.

66. Buchbinder, S.P., Katz, M.H., Hessol, N.A., O'Malley, P.M., and Holmberg, S.D. 1994. *Long-term HIV-1 infection without immunologic progression*. AIDS 8:1123-8.

67. Lifson, A.R. *et al*. 1991. *Long-term human immunodeficiency virus infection in asymptomatic homosexual and bisexual men with normal CD4+ lymphocyte counts: immunologic and virologic characteristics*. J. Infect. Dis. 163:959-65.

68. Anzala, O.A. *et al*. 1995. *Rapid progression to disease in African sex workers with human immunodeficiency virus type 1 infection*. J. Infect. Dis. 171:686-9.

69. Pantaleo, G. and Fauci, A.S. 1996. *Immunopathogenesis of HIV infection*. Annu. Rev. Microbiol. 50:825-54.

70. Cao, Y., Qin, L., Zhang, L., Safrit, J., and Ho, D.D. 1995. *Virologic and immunologic characterization of long-term survivors of human immunodeficiency virus type 1 infection*. N. Engl. J. Med. 332:201-8.

71. Pantaleo, G. *et al*. 1995. *Studies in subjects with long-term nonprogressive human immunodeficiency virus infection*. N. Engl. J. Med. 332:209-16.

72. Mellors, J.W. *et al*. 1995. *Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion*. Ann. Intern. Med. 122:573-9.

73. Wyatt, R., Moore, J., Accola, M., Desjardin, E., Robinson, J., and Sodorski, J. 1995. *Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding*. J. Virol. 69:5723-33.

74. Rizzuto, C.D. *et al*. 1998. *A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding*. Science 280:1949-53.

75. McDonald, D. *et al*. 2002. *Visualization of the intracellular behavior of HIV in living cells*. J. Cell Biol. 159:441-52.

76. Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. 2002. *HIV-1 integration in the human genome favors active genes and local hotspots*. Cell 110:521-9.

77. Lewinski, M.K. *et al*. 2006. *Retroviral DNA integration: viral and cellular determinants of target-site selection*. PLoS Pathog. 2:e60.

78. Ciuffi, A. *et al*. 2005. *A role for LEDGF/p75 in targeting HIV DNA integration*. Nat. Med. 11:1287-9.

79. Garcia, J.V. and Miller, A.D. 1992. *Downregulation of cell surface CD4 by nef*. Res. Virol. 143:52-5.

80. Piguet, V., Schwartz, O., Le Gall, S., and Trono, D. 1999. *The downregulation of CD4 and MHC-I by primate lentiviruses: a paradigm for the modulation of cell surface receptors*. Immunol. Rev. 168:51-63.

81. Cohen, G.B. *et al*. 1999. *The selective downregulation of class I major histocompability complex proteins by HIV-1 protects HIV-infected cells from NK cells*. Immunity 10:661-71.

82. Xu, X.N. and Screaton, G. 2001. *HIV-1 Nef: negative effector of Fas?* Nat. Immunol. 2:384-5.

83. Larder, B.A., Darby, G., and Richman, D.D. 1989. *HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy*. Science 243:1731-4.

84. Palella, F.J.J. *et al*. 1998. *Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators*. N. Engl. J. Med. 338:853-60.

85. Tassie, J.M. *et al*. 2002. *Time to AIDS from 1992 to 1999 in HIV-1 infected subjects with known date of infection*. J. Acquir. Immune. Defic. Syndr. 30:81-7.

86. King, J.T.J., A.C., J., Roberts, M.S., Chang, C.C., Fusco, J.S., and Team., C.i.H.O.R.-U.P. 2003. *Long-term HIV/AIDS survival estimation in the highly active antiretroviral therapy era*. Med. Decis. Making 23:9-20.

87. Biswas, P., Tambussi, G., and Lazzarin, A. 2007. *Access denied? The status of co-receptor inhibition to counter HIV entry*. Expert Opin. Pharmacother. 8:923-33.

88. Cervia, J.S. and Smith, M.A. 2003. *Enfuvirtide (T-20): a novel human immunodeficiency virus type 1 fusion inhibitor*. Clin. Infect. Dis. 37:1102-6.

89. Markowitz, M. *et al*. 2007. *Rapid and durable antiretroviral effect of the HIV-1 inhibitor raltegravir as part of combination therapy in treatment-naive patients with HIV-1 infection: results of a 48-week controlled study*. J. Acquir. Immune. Defic. Syndr. 46:125-33.

90. Lazzarin, A. *et al*. 2003. *Efficacy of enfuvirtide in patients with drug-resistant HIV-1 in Europe and Australia*. N. Engl. J. Med. 348:2186-95.

91. Grinsztejn, B. *et al*. 2007. *Safety and efficacy of the HIV-1 integrase inhibitor raltegravir (MK-0518) in treatment-experienced patients with multidrug-resistant virus: a phase II randomized controlled trial*. Lancet 369:1261-9.

92. Salzwedel, K., Martin, D.E., and Sakalian, M. 2007. *Maturation inhibitors: a new therapeutic class targets the virus structure*. AIDS Rev. 9:162-72.

93. Sarkar, I., Hauber, I., Hauber, J., and Buchholz, F. 2007. *HIV-1 proviral DNA excision using an evolved recombinase*. Science 316:1912-5.

94. Li, W.H., Tanimura, M., and Sharp, P.M. 1988. *Rates and dates of divergence between AIDS virus nucleotide sequences*. Mol. Biol. Evol. 5:313-30.

95. Mansky, L.M. and Temin, H.M. 1995. *Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase*. J. Virol. 69:5087-94.

96. Jetzt, A.E., Yu, H., Klarmann, G.J., Ron, Y., Preston, B.D., and Dougherty, J.P. 2000. *High rate of recombination throughout the human immunodeficiency virus type 1 genome*. J. Virol. 74:1234-40.

97. Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M., and Ho, D.D. 1996. *HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time*. Science 271:1582-6.

98. Markowitz, M. *et al*. 2003. *A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and T-cell decay in vivo*. J. Virol. 77:5037-8.

99. Wei, X. *et al*. 1995. *Viral dynamics in human immunodeficiency virus type 1 infection*. Nature 373:117-22.

100. Holland, J.J., De La Torre, J.C., and Steinhauer, D.A. 1992. *RNA virus populations as quasispecies*. Curr. Top. Microbiol. Immunol. 176:1-20.

101. Goodenow, M., Huet, T., Saurin, W., Kwok, S., Sninsky, J., and Wain-Hobson, S. 1989. *HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions*. J. Acquir. Immune. Defic. Syndr. 2:344-52.

102. Sala, M. and Wain-Hobson, S. 2000. *Are RNA viruses adapting or merely changing?* J. Mol. Evol. 51:12-20.

103. Seibert, S.A., Howell, C.Y., Hughes, M.K., and Hughes, A.L. 1995. *Natural selection on the gag, pol, and env genes of human immunodeficiency virus 1 (HIV-1)*. Mol. Biol. Evol. 12:803-13.

104. Cornelissen, M., van den Burg, R., Zorgdrager, F., Lukashov, V., and Goudsmit, J. 1997. *pol gene diversity of five human immunodeficiency virus type 1 subtypes: evidence for naturally occurring mutations that contribute to drug resistance, limited recombination patterns, and common ancestry for subtypes B and D*. J. Virol. 71:6348-58.

105. Wolfs, T.F., de Jong, J.J., Van den Berg, H., Tijnagel, J.M., Krone, W.J., and Goudsmit, J. 1990. *Evolution of sequences encoding the principal neutralization epitope of human immunodeficiency virus 1 is host dependent, rapid, and continuous*. Proc. Natl. Acad. Sci. USA 87:9938-42.

106. Yamaguchi, Y. and Gojobori, T. 1997. *Evolutionary mechanisms and population dynamics of the third variable envelope region of HIV within single hosts*. Proc. Natl. Acad. Sci. USA 94:1264-9.

107. Borrow, P. *et al*. 1997. *Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus*. Nat. Med. 3:205-11.

108. Goulder, P.J. *et al*. 1997. *Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS*. Nat. Med. 3:212-7.

109. Liu, Y. *et al*. 2007. *Evolution of human immunodeficiency virus type 1 cytotoxic T-lymphocyte epitopes: fitness-balanced escape*. J. Virol. 81:12179-88.

110. Back, N.K. *et al*. 1994. *An N-glycan within the human immunodeficiency virus type 1 gp120 V3 loop affects virus neutralization*. Virology 199:431-8.

111. Parera, M., Ibanez, A., Clotet, B., and Martinez, M.A. 2004. *Lack of evidence for protease evolution in HIV-1-infected patients after 2 years of successful highly active antiretroviral therapy*. J. Infect. Dis. 189:1444-51.

112. Rodrigo, A.G., Goode, M., Forsberg, R., Ross, H.A., and Drummond, A.J. 2003. *Inferring evolutionary rates using serially sampled sequences from several populations*. Mol. Biol. Evol. 20:2010-8.

113. Hirsch, M.S. *et al*. 2000. *Antiretroviral drug resistance testing in adult HIV-1 infection: recommendations of an International AIDS Society-USA Panel*. JAMA 283:2417-26.

114. Croteau, G., Doyon, L., Thibeault, D., McKercher, G., Pilote, L., and Lamarre, D. 1997. *Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors*. J. Virol. 71:1089-96.

115. Mammano, F., Petit, C., and Clavel, F. 1998. *Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients*. J. Virol. 72:7632-7.

116. Boyer, P.L., Gao, H.Q., and Hughes, S.H. 1998. *A mutation at position 190 of human immunodeficiency virus type 1 reverse transcriptase interacts with mutations at positions 74 and 75 via the template primer*. Antimicrob. Agents Chemother. 42:447-52.

117. Nijhuis, M. *et al*. 1999. *Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy*. AIDS 13:2349-59.

118. Zhang, Y.M., Imamichi, H., Lane, H.C., Falloon, J., Vasudevachari, M.B., and Salzman, N.P. 1997. *Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites*. J. Virol. 71:6662-70.

119. Leigh Brown, A.J. and Holmes, E.C. 1994. *Evolutionary biology of human immunodeficiency virus*. Annu. Rev. Ecol. Syst. 25:127-65.

120. Shankarappa, R. *et al*. 1999. *Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection*. J. Virol. 73:10489-502.

121. Edwards, C.T. *et al*. 2006. *Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1*. BMC Evol. Biol. 6:28.

122. Zhang, L.Q., MacKenzie, P., Cleland, A., Holmes, E.C., Brown, A.J., and Simmonds, P. 1993. *Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection*. J. Virol. 67:3345-56.

123. Zhu, T. *et al*. 1993. *Genotypic and phenotypic characterization of HIV-1 patients with primary infection*. Science 261:1179-81.

124. Huang, Y. *et al*. 1996. *The role of a mutant CCR5 allele in HIV-1 transmission and disease progression*. Nat. Med. 2:1240-3.

125. Michael, N.L. *et al*. 1998. *Exclusive and persistent use of the entry coreceptor CXCR4 by human immunodeficiency virus type 1 from a subject homozygous for CCR5 delta32*. J. Virol. 72:6040-7.

126. Connor, R.I., Sheridan, K.E., Ceradini, D., Choe, S., and Landau, N.R. 1997. *Change in coreceptor use correlates with disease progression in HIV-1-infected individuals*. J. Exp. Med. 185:621-8.

127. Scarlatti, G. *et al*. 1997. *In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression*. Nat. Med. 3:1259-65.

128. Shioda, T., Levy, J.A., and Cheng-Mayer, C. 1992. *Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell-line and macrophage tropism of human immunodeficiency virus type 1*. Proc. Natl. Acad. Sci. USA 89:9434-8.

129. De Jong, J.J., De Ronde, A., Keulen, W., Tersmette, M., and Goudsmit, J. 1992. *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. 66:6777-80.

130. Koning, F.A., van Rij, R.P., and Schuitemaker, H., *Biological and molecular aspects of HIV-1 coreceptor usage*., in *HIV sequence compendium*, C.L. Kuiken, et al., Editors. 2002, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory: Los Alamos, NM, LA-UR 03-3564.

131. Holmes, E.C., Zhang, L.Q., Simmonds, P., Ludlam, C.A., and Brown, A.J. 1992. *Convergent and divergent evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient*. Proc. Natl. Acad. Sci. USA 89:4835-9.

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

133. Ross, H.A. and Rodrigo, A.G. 2002. *Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration*. J. Virol. 76:11715-20.

134. Shriner, D. *et al*. 2004. *Influence of random genetic drift on human immunodeficiency virus type 1 env evolution during chronic infection*. Genetics 166:1155-64.

135. Edwards, C.T. *et al*. 2006. *Evolution of the human immunodeficiency virus envelope gene is dominated by purifying selection*. Genetics 174:1441-53.

136. Lemey, P. *et al*. 2007. *Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics*. PLoS Comput. Biol. 3:e29.

137. Delwart, E.L. *et al*. 1997. *Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS*. J. Virol. 71:7498-508.

138. McNearney, T. *et al*. 1992. *Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease*. Proc. Natl. Acad. Sci. USA 89:10247-51.

139. Wolinsky, S.M. *et al*. 1996. *Adaptive evolution of human immunodeficiency virus type 1 during the natural course of infection*. Science 272:537-42.

140. Quinones-Mateu, M.E. *et al*. 2000. *A dual infection/competition assay shows a correlation between ex vivo human immunodeficiency virus type 1 fitness and disease progression*. J. Virol. 74:9222-33.

141. Troyer, R.M. *et al*. 2005. *Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression*. J. Virol. 79:9006-18.

142. Halapi, E. *et al*. 1997. *Correlation between HIV sequence evolution, specific immune response and clinical outcome in vertically infected infants*. AIDS 11:1709-17.

143. Pond, S.L., Frost, S.D., Grossman, Z., Gravenor, M.B., Richman, D.D., and Brown, A.J. 2006. *Adaptation to different human populations by HIV-1 revealed by codon-based analyses*. PLoS Comput. Biol. 2:e62.

144. Grenfell, B.T. *et al*. 2004. *Unifying the epidemiological and evolutionary dynamics of pathogens*. Science 303:327-32.

145. Rambaut, A., Posada, D., Crandall, K.A., and Holmes, E.C. 2004. *The causes and consequences of HIV evolution*. Nat. Rev. Genet. 5:52-61.

146. Leslie, A.J. *et al*. 2004. *HIV evolution: CTL escape mutation and reversion after transmission*. Nat. Med. 10:229-30.

147. Fernandez, C.S. *et al*. 2005. *Rapid viral escape at an immunodominant simian-human immunodeficiency virus cytotoxic T-lymphocyte epitope exacts a dramatic fitness cost*. J. Virol. 79:5721-31.

148. Li, B. *et al*. 2007. *Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution*. J. Virol. 81:193-201.

149. Wensing, A.M. *et al*. 2005. *Prevalence of drug-resistant HIV-1 variants in untreated individuals in Europe: implications for clinical management*. 192:958-66.

150. SPREAD programme 2008. *Transmission of drug-resistant HIV-1 in Europe remains limited to single classes*. AIDS 22:625-35.

151. Little, S.J. *et al*. 2002. *Antiretroviral drug resistance among patients recently infected with HIV*. N. Engl. J. Med. 347:385-94.

152. Shet, A. *et al*. 2006. *Tracking the prevalence of transmitted antiretroviral drug-resistant HIV-1: a decade of experience*. J. Acquir. Immune. Defic. Syndr. 41:439-46.

153. Cane, P. *et al*. 2005. *Time trends in primary resistance to HIV drugs in the United Kingdom: multicentre observational study*. BMJ 331:1368.

154. Yerly, S. *et al*. 1998. *Switch to unusual amino acids at codon 215 of the human immunodeficiency virus type 1 reverse transcriptase gene in seroconverters infected with zidovudine-resistant variants*. J. Virol. 72:3520-3.

155. Leigh Brown, A.J. *et al*. 2003. *Transmission fitness of drug-resistant human immunodeficiency virus and the prevalence of resistance in the antiretroviral-treated population*. J. Infect. Dis. 187:683-6.

156. Edwards, C.T., Pfafferott, K.J., Goulder, P.J., Phillips, R.E., and Holmes, E.C. 2005. *Intrapatient escape in the A*0201-restricted epitope SLYNTVATL drives evolution of human immunodeficiency virus type 1 at the population level*. J. Virol. 79:9363-6.

157. Allen, T.M. *et al*. 2005. *Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution*. J. Virol. 79:13239-49.

158. De Ronde, A. *et al*. 2001. *Establishment of new transmissible and drug-sensitive human immunodeficiency virus type 1 wild types due to transmission of nucleoside analogue-resistant virus*. J. Virol. 75:595-602.

159. Garcia-Lerma, J.G., Nidtha, S., Blumoff, K., Weinstock, H., and Heneine, W. 2001. *Increased ability for selection of zidovudine resistance in a distinct class of wild-type HIV-1 from drug-naive persons*. Proc. Natl. Acad. Sci. USA 98:13907-12.

160. Gordo, I. and Campos, P.R. 2007. *Patterns of genetic variation in populations of infectious agents*. BMC Evol. Biol. 7:116.

161. Buckee, C.O., Koelle, K., Mustard, M.J., and Gupta, S. 2004. *The effects of host contact network structure on pathogen diversity and strain structure*. Proc. Natl. Acad. Sci. USA 101:10839-44.

162. Goodreau, S.M. 2006. *Assessing the effects of human mixing patterns on human immunodeficiency virus-1 interhost phylogenetics through social network simulation*. Genetics 172:2033-45.

163. Van Heuverswyn, F. *et al*. 2006. *SIV infection in wild gorillas*. Nature 444:164.

164. Gurtler, L.G. *et al*. 1994. *A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon*. J. Virol. 68:1581-5.

165. Simon, F. *et al*. 1998. *Identification of a new human immunodeficiency virus type 1 distinct from group M and group O*. Nat. Med. 4:1032-7.

166. Leitner, T. *et al*., eds. *HIV Sequence Compendium*. 2005, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, LA-UR 06-0680.

167. Sabino, E.C. *et al*. 1994. *Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil*. J. Virol. 68:6340-6.

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

169. Leitner, T. *et al*. 1995. *Biological and molecular characterization of subtype D, G, and A/D recombinant HIV-1 transmissions in Sweden*. Virology 209:136-46.

170. *HIV sequence database*. http://www.hiv.lanl.gov.

171. Janssens, W. *et al*. 2000. *Near full-length genome analysis of HIV type 1 CRF02.AG subtype C and CRF02.AG subtype G recombinants*. AIDS Res. Hum. Retroviruses 16:1183-9.

172. Mamadou, S. *et al*. 2003. *Emergence of complex and diverse CRF02-AG/CRF06-cpx recombinant HIV type 1 strains in Niger, West Africa*. AIDS Res. Hum. Retroviruses 19:77-82.

173. Tebit, D.M., Ganame, J., Sathiandee, K., Nagabila, Y., Coulibaly, B., and Krausslich, H.G. 2006. *Diversity of HIV in rural Burkina Faso*. J. Acquir. Immune. Defic. Syndr. 43:144-52.

174. Yang, R., Xia, X., Kusagawa, S., Zhang, C., Ben, K., and Takebe, Y. 2002. *On-going generation of multiple forms of HIV-1 intersubtype recombinants in the Yunnan Province of China*. AIDS 16:1401-7.

175. Tran, T.T., Maljkovic, I., Swartling, S., Phung, D.C., Chiodi, F., and Leitner, T. 2004. *HIV-1 CRF01_AE in intravenous drug users in Hanoi, Vietnam*. AIDS Res. Hum. Retroviruses 20:341-5.

176. Takehisa, J. *et al*. 1999. *Human immunodeficiency virus type 1 intergroup (M/O) recombination in Cameroon*. J. Virol. 73:6810-20.

177. Peeters, M. *et al*. 1999. *Characterization of a highly replicative intergroup M/O human immunodeficiency virus type 1 recombinant isolated from a Cameroonian patient*. J. Virol. 73:7368-75.

178. Robertson, D.L. *et al*. 2000. *HIV-1 nomenclature proposal*. Science 288:55-6.

179. Abecasis, A.B. *et al*. 2007. *Recombination confounds the early evolutionary history of human immunodeficiency virus type 1: subtype G is a circulating recombinant form*. J. Virol. 81:8543-51.

180. Kalish, M.L. *et al*. 1994. *Early HIV type 1 strains in Thailand were not responsible for the current epidemic*. AIDS Res. Hum. Retroviruses 10:1573-5.

181. Subbarao, S. *et al*. 1998. *HIV type 1 in Thailand, 1994-1995: Persistence of two subtypes with low genetic diversity*. AIDS Res Hum Retroviruses 14:319-327.

182. Shankarappa, R. *et al*. 2001. *Human immunodeficiency virus type 1 env sequences from Calcutta in eastern India: identification of features that distinguish subtype C sequences in India from other subtype C sequences*. J. Virol. 75:10479-87.

183. Siddappa, N.B. *et al*. 2004. *Identification of subtype C human immunodeficiency virus type 1 by subtype-specific PCR and its use in the characterization of viruses circulating in the southern parts of India*. J. Clin. Microbiol. 42:2742-51.

184. Qiu, Z. *et al*. 2005. *Characterization of five nearly full-length genomes of early HIV type 1 strains in Ruili city: implications for the genesis of CRF07_BC and CRF08_BC circulating in China*. AIDS Res. Hum. Retroviruses 21:1051-6.

185. McCutchan, F.E. *et al*. 2002. *Precise mapping of recombination breakpoints suggests a common parent of two BC recombinant HIV type 1 strains circulating in China*. AIDS Res. Hum. Retroviruses 18:1135-40.

186. Lukashov, V.V., Karamov, E.V., Eremin, V.F., Titov, L.P., and Goudsmit, J. 1998. *Extreme founder effect in an HIV type 1 subtype A epidemic among drug users in Svetlogorsk, Belarus*. AIDS Res. Hum. Retroviruses 14:1299-303.

187. Bobkov, A.F. *et al*. 2004. *Temporal trends in the HIV-1 epidemic in Russia: predominance of subtype A*. J. Med. Virol. 74:191-6.

188. Vidal, N. *et al*. 2000. *Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa*. J. Virol. 74:10498-507.

189. Takehisa, J. *et al*. 1998. *Various types of HIV mixed infections in Cameroon*. Virology 245:1-10.

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

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

192. Njai, H.F. *et al*. 2006. *The predominance of human immunodeficiency virus type 1 (HIV-1) circulating recombinant form 02 (CRF02_AG) in West Central Africa may be related to its replicative fitness*. Retrovirology 3:40.

193. Konings, F.A., Burda, S.T., Urbanski, M.M., Zhong, P., Nadas, A., and Nyambi, P.N. 2006. *Human immunodeficiency virus type 1 (HIV-1) circulating recombinant form 02_AG (CRF02_AG) has a higher in vitro replicative capacity than its parental subtypes A and G*. J. Med. Virol. 78:523-34.

194. Zhang, L., Huang, Y., He, T., Cao, Y., and Ho, D.D. 1996. *HIV-1 subtype and second-receptor use*. Nature 383:768.

195. Tscherning, C. *et al*. 1998. *Differences in chemokine coreceptor usage between genetic subtypes of HIV-1*. Virology 241:181-8.

196. Kanki, P.J. *et al*. 1999. *Human immunodeficiency virus type 1 subtypes differ in disease progression*. J. Infect. Dis. 179:68-73.

197. Baeten, J.M. *et al*. 2007. *HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads*. J. Infect. Dis. 195:1177-80.

198. Wahren, B. *et al*. 2002. *HIV subtypes and recombination strains--strategies for induction of immune responses in man*. Vaccine 20:1988-93.

199. Pieniazek, D. *et al*. 2000. *Protease sequences from HIV-1 group M subtypes A-H reveal distinct amino acid mutation patterns associated with protease resistance in protease inhibitor-naive individuals worldwide*. AIDS 14:1489-95.

200. Caride, E. *et al*. 2001. *Genotypic and phenotypic evidence of different drug-resistance mutation patterns between B and non-B subtype isolates of human immunodeficiency virus type 1 found in Brazilian patients failing HAART*. Virus Genes 23:193-202.

201. Birk, M. and Sonnerborg, A. 1998. *Variations in HIV-1 pol gene associated with reduced sensitivity to antiretroviral drugs in treatment-naive patients*. AIDS 12:2369-75.

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

203. van de Vijver, D.A. *et al*. 2006. *The calculated genetic barrier for antiretroviral drug resistance substitutions is largely similar for different HIV-1 subtypes*. J. Acquir. Immune. Defic. Syndr. 41:352-60.

204. Bannister, W.P. *et al*. 2006. *HIV-1 subtypes and response to combination antiretroviral therapy in Europe*. Antivir. Ther. 11:707-15.

205. Alexander, C.S. *et al*. 2002. *Prevalence and response to antiretroviral therapy of non-B subtypes of HIV in antiretroviral-naive individuals in British Columbia*. Antivir. Ther. 7:31-5.

206. Atlas, A., Granath, F., Lindstrom, A., Lidman, K., Lindback, S., and Alaeus, A. 2005. *Impact of HIV type 1 genetic subtype on the outcome of antiretroviral therapy*. AIDS Res. Hum. Retroviruses 21:221-7.

207. Bryson, Y.J., Luzuriaga, K., Sullivan, J.L., and Wara, D.W. 1992. *Proposed definitions for in utero versus intrapartum transmission of HIV-1*. N. Engl. J. Med. 327:1246-7.

208. Dickover, R.E. *et al*. 1994. *Rapid increases in load of human immunodeficiency virus correlate with early disease progression and loss of CD4 cells in vertically infected infants*. J. Infect. Dis. 170:1279-84.

209. Naver, L. *et al*. 2001. *Broad spectrum of time of detection, primary symptoms and disease progression in infants with HIV-1 infection*. Eur. J. Clin. Microbiol. Infect. Dis. 20:159-66.

210. Casper, C.H. *et al*. 2002. *Link between the X4 phenotype in human immunodeficiency virus type 1-infected mothers and their children, despite the early presence of R5 in the child*. J. Infect. Dis. 186:914-21.

211. Casper, C. *et al*. 2002. *Coreceptor change appears after immune deficiency is established in children infected with different HIV-1 subtypes*. AIDS Res. Hum. Retroviruses 18:343-52.

212. Behbahani, H. *et al*. 2000. *Up-regulation of CCR5 expression in the placenta is associated with human immunodeficiency virus-1 vertical transmission*. Am. J. Pathol. 157:1811-8.

213. Reinhardt, P.P., Reinhardt, B., Lathey, J.L., and Spector, S.A. 1995. *Human cord blood mononuclear cells are preferentially infected by non-syncytium-inducing, macrophage-tropic human immunodeficiency virus type 1 isolates*. J. Clin. Microbiol. 33:292-7.

214. Ghaffari, G. *et al*. 2005. *Complex determinants in human immunodeficiency virus type 1 envelope gp120 mediate CXCR4-dependent infection of macrophages*. J. Virol. 79:13250-61.

215. Westby, M. *et al*. 2006. *Emergence of CXCR4-using human immunodeficiency virus type 1 (HIV-1) variants in an minority of HIV-1-infected patients following treatment with the CCR5 antagonist maraviroc is from a pretreatment CXCR4-using virus reservoir*. J. Virol. 80:4909-20.

216. Salemi, M., Burkhardt, B.R., Gray, R.R., Ghaffari, G., Sleasman, J.W., and Goodenow, M. 2007. *Phylodynamics of HIV-1 in lymphoid and non-lymphoid tissues reveals a central role for the thymus in emergence of CXCR4-using quasispecies*. PLoS ONE. 2:e950.

217. Leitner, T. and Albert, J. 1999. *The molecular clock of HIV-1 unveiled through analysis of a known transmission history*. Proc. Natl. Acad. Sci. USA 96:10752-7.

218. Mikhail, M. *et al*. 2005. *Role of viral evolutionary rate in HIV-1 disease progression in a linked cohort*. Retrovirology 2.

219. Ripamonti, C. *et al*. 2007. *Biological and genetic evolution of HIV type 1 in two siblings with different patterns of disease progression*. AIDS Res. Hum. Retroviruses 23:1531-40.

220. Balode, D. *et al*. 2004. *Rapid epidemic spread of HIV type 1 subtype A1 among intravenous drug users in Latvia and slower spread of subtype B among other risk groups*. AIDS Res Hum Retroviruses 20:245-9.

221. Holmes, E.C. *et al*. 1995. *The molecular epidemiology of human immunodeficiency virus type 1 in Edinburgh*. J. Infect. Dis. 171:45-53.

222. Ou, C.Y. *et al*. 1992. *Molecular epidemiology of HIV transmission in a dental practice*. Science 256:1165-71.

223. Albert, J., Wahlberg, J., Leitner, T., Escanilla, D., and Uhlen, M. 1994. *Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 pol and gag genes*. J. Virol. 68:5918-24.

224. Leitner, T., Escanilla, D., Franzen, C., Uhlen, M., and Albert, J. 1996. *Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis*. Proc. Natl. Acad. Sci. USA 93:10864-9.

225. Paraskevis, D. *et al*. 2004. *Phylogenetic reconstruction of a known HIV-1 CRF04_cpx transmission network using maximum likelihood and Bayesian methods*. J. Mol. Evol. 59:709-17.

226. Lemey, P. *et al*. 2005. *Molecular footprint of drug-selective pressure in a human immunodeficiency virus transmission chain*. J. Virol. 79:11981-9.

227. Leitner, T. and Fitch, W.M., *The phylogenies of known transmission histories*., in *The evolution of HIV*., K.A. Crandall, Editor. 1999, The Johns Hopkins University Press: Baltimore, Maryland.

228. Leitner, T., *The use of molecular epidemiology*., in *The molecular epidemiology of human viruses*., T. Leitner, Editor. 2002, Kluwer Academic Publishers: Norwell, Massachusetts.

229. Leitner, T., Kumar, S., and Albert, J. 1997. *Tempo and mode of nucleotide substitutions in gag and env gene fragments in human immunodeficiency virus type 1 populations with a known transmission history*. J. Virol. 71:4761-70.

230. Leitner, T. and Albert, J. 2000. *Reconstruction of HIV-1 transmission chains for forensic purposes*. AIDS Rev. 2:241-51.

231. Birch, C.J. *et al*. 2000. *Molecular analysis of human immunodeficiency virus strains associated with a case of criminal transmission of the virus*. J. Infect. Dis. 182:941-4.

232. Metzker, M.L., Mindell, D.P., Liu, X.M., Ptak, R.G., Gibbs, R.A., and Hillis, D.M. 2002. *Molecular evidence of HIV-1 transmission in a criminal case*. Proc. Natl. Acad. Sci. USA 99:14292-7.

233. Gunthard, H.F. *et al*. 1999. *Evolution of envelope sequences of human immunodeficiency virus type 1 in cellular reservoirs in the setting of potent antiviral therapy*. J. Virol. 73:9404-12.

234. Frenkel, L.M. *et al*. 2003. *Multiple viral genetic analyses detect low-level human immunodeficiency virus type 1 replication during effective highly active antiretroviral therapy*. J. Virol. 77:5721-30.

235. Bello, G. *et al*. 2007. *Lack of temporal structure in the short term HIV-1 evolution within asymptomatic naive patients*. Virology 362:294-303.

236. Rodes, B. *et al*. 2004. *Differences in disease progression in a cohort of long-term non-progressors after more than 16 years of HIV-1 infection*. AIDS 18:1109-16.

237. Picchio, G.R., Gulizia, R.J., and Mosier, D.E. 1997. *Chemokine receptor CCR5 genotype influences the kinetics of human immunodeficiency virus type 1 infection in human PBL-SCID mice*. J. Virol. 71:7124-7.

238. Drummond, A., Forsberg, R., and Rodrigo, A.G. 2001. *The inference of stepwise changes in substitution rates using serial sequence samples*. Mol. Biol. Evol. 18:1365-71.

239. Finzi, D. *et al*. 1997. *Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy*. Science 278:1295-300.

240. Bobkov, A. *et al*. 1997. *An HIV type 1 epidemic among injecting drug users in the Former Soviet Union caused by a homogeneous subtype A strain*. AIDS Res Hum Retroviruses 13:1195-1201.

241. Bobkov, A.F. *et al*. 2004. *An HIV type 1 subtype A outbreak among injecting drug users in Kazakhstan*. AIDS Res Hum Retroviruses 20:1134-1136.

242. Kalish, M.L. *et al*. 1995. *The evolving molecular epidemiology of HIV-1 envelope subtypes in injecting drug users in Bangkok, Thailand: Implications for HIV vaccine trials*. AIDS 9:851-857.

243. Zheng, X. *et al*. 1994. *Injecting drug use and HIV infection in southwest China*. AIDS 8:1141-1147.

244. Vidal, N. *et al*. 2005. *Distribution of HIV-1 variants in the Democratic Republic of Congo suggests increase of subtype C in Kinshasa between 1997 and 2002*. J. Acquir. Immune Defic. Syndr. 40:456-62.

245. Gordon, M. *et al*. 2003. *Molecular characteristics of human immunodeficiency virus type 1 subtype C viruses from KwaZulu-Natal, South Africa: Implications for vaccine and antiretroviral control strategies*. J. Virol. 77:2587-99.

246. Engelbrecht, S., de Villiers, T., Sampson, C.C., zur Megede, J., W., B.S., and van Rensburg, E.J. 2001. *Genetic analysis of the complete gag and env genes of HIV subtype C primary isolates from South Africa*. AIDS Res Hum Retroviruses 17:1533-47.

247. Korber, B.T., MacInnes, K., F., S.R., and Myers, G. 1994. *Mutational trends in V3 loop protein sequences observed in different genetic lineages of human immunodeficiency virus type 1*. J. Virol. 68:6730-44.

248. Lewis, F., Hughes, G.J., Rambaut, A., Pozniak, A., and Leigh Brown, A.J. 2008. *Episodic sexual transmission of HIV revealed by molecular phylodynamics*. PLoS Med. 5:e50.

249. Salemi, M. *et al*. 2001. *Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution*. FASEB J. 15:276-8.

250. Zuckerkandl, E. and Pauling, L., *Evolutionary divergence and convergence in proteins*., in *Evolving genes and proteins*., V. Bryson and H.J. Vogel, Editors. 1965, Academic Press: New York.

251. Kimura, M. 1980. *A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences*. J. Mol. Evol. 16:111-20.

252. Gillespie, J.H. 1984. *The molecular clock may be an episodic clock*. Proc. Natl. Acad. Sci. USA 81:8009-13.

253. Takahata, N. 1987. *On the overdispersed molecular clock*. Genetics 116:169-79.

254. Thorne, J.L., Kishino, H., and Painter, I.S. 1998. *Estimating the rate of evolution of the rate of molecular evolution*. Mol. Biol. Evol. 15:1647-57.

255. Huelsenbeck, J.P., Larget, B., and Swofford, D. 2000. *A compound poisson process for relaxing the molecular clock*. Genetics 154:1879-92.

256. Drummond, A.J., Ho, S.Y., Phillips, M.J., and Rambaut, A. 2006. *Relaxed phylogenetics and dating with confidence*. PLoS Biol. 4:e88.

257. Yang, Z. and Rannala, B. 2006. *Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds*. Mol. Biol. Evol. 23:212-26.

258. Skellam, J. 1946. *The frequency distribution of the difference between two Poisson variates belonging to different populations*. Journal of the Royal Statistical Society 109:296.