

From the DEPARTMENT OF MICROBIOLOGY, TUMOR AND
CELL BIOLOGY
Karolinska Institutet, Stockholm, Sweden

**PHYLOGENETIC AND PHENOTYPIC
PROPERTIES OF HIV-1 VARIANTS OF
DIFFERENT SUBTYPES,
IN MOTHER TO CHILD TRANSMISSION**

Rozina Caridha



**Karolinska
Institutet**

Stockholm 2013

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet. Printed by Larseries Digital Print AB.

© Rozina Caridha, 2013
rozinats@gmail.com
ISBN 978-91-7549-049-6

“On your journey to your dream, be ready to face oasis and deserts. In both cases, don't stop”
-- Paulo Coelho

In memory of my teacher Anneka Ehrnst

ABSTRACT

Transmission from mother to child is the most common way that children contract HIV-1 infection in developing countries; transmission through this route is prevented in most developed countries by antiretroviral treatment, elective Caesarean section and absence of breast-feeding. However, these measures are not fully available in developing countries.

By defining the properties of the virus that is transmitted from mother to child in great detail, we aimed at establishing a foundation of knowledge to improve preventive measures against HIV-1 transmission. Moreover, the understanding of the HIV-1 genetic variation and phenotype evolution is critical for unravelling correlates of disease progression. The third variable region (V3) of the HIV-1 envelope gp120 protein mediates co-receptor interaction and is an important marker of the viral phenotype. While the X4 phenotype is associated with disease progression, the R5 phenotype is associated with transmission and/or is the phenotype of the virus permitted to initially replicate following transmission. Thus, it is crucial to precisely identify properties of the transmitted virus, which may become the targets for new interventions.

In the present thesis, population and single genome sequences were obtained through nested PCR of *env* gp120 V3 and flanking regions. The co-receptor determination of HIV-1 isolates was previously performed in U87 cell lines. Geno2Pheno and an in-house method, the glycan-charge model, determined the genetic co-receptor predictions. Phylogenetic analysis was extensively used to map the origin and relation of virus isolates to other previously characterized HIV-1 strains.

In a prospective study conducted in northern Vietnam (paper I) a strategy combining post-test confidential counselling of HIV-1 infected mothers, formula feeding and antiretroviral prophylaxis of mothers and children resulted in low rates of delivery-associated and late HIV-1 transmissions. In 37 of the HIV-1 infected pregnant women from paper I, we further traced the origin of their HIV-1 *env* sequences (paper II). Their *env* sequences were classified as being CRF01_AE subtype and showed a relatively low evolutionary rate, which is compatible with a rapidly spreading, epidemic. In the third study (paper III), we investigated the HIV-1 predicted phenotypes of the CRF01_AE sequences, obtained from the mothers presented in paper I. Sequences from a separate group of vertically infected children from the same hospital in Hanoi were also added to paper III. In this study, we found a high overall proportion of the CXCR4-using phenotype; however, despite the dominant presence of CXCR4-using strains in mothers of infected children, it appears that CCR5-using strains would be favoured in transmission. The study IV was an attempt to identify and characterize the transmitted/founder virus in mother-to-child transmission. Over 700 single genome sequences were obtained. We preliminarily observed 11 matches in 8 cases, where infant sequences were identical to the maternal sequences. The earliest viruses obtained from the children harboured a virus with R5-like properties also in the expanded viral population. There was a tendency to an increased V3 charge over time and sequences of transmitted virus were often stable over time in the children samples from different subtypes.

In summary, this thesis may hopefully contribute to advance our knowledge on the viral characteristics, related to early transmission events from mother to child, which maybe helpful to take under considerations when improving or developing interventions against HIV.

LIST OF PAPERS

- I. Ha TT, Anh NM, Bao NH, Tuan PL, **Caridha R**, Gaseitsiwe S, Hien NT, Cam PD, Ehrnst A. HIV-1 mother-to-child transmission, post-test counselling, and antiretroviral prophylaxis in Northern Viet Nam: A prospective observational study. *Scand J Infect Dis*. 2012 Nov;44(11):866-73.
- II. **Caridha R***, Ha TT*, Gaseitsiwe S, Hung PV, Anh NM, Bao NH, Khang DD, Hien NT, Cam PD, Chiodi F, Ehrnst A. Short communication: phylogenetic characterization of HIV type 1 CRF01_AE V3 envelope sequences in pregnant women in Northern Vietnam. *AIDS Res Hum Retroviruses*. 2012 Aug;28(8):852-6. * Equal contribution
- III. **Caridha R**, Ha TT, Hung PV, Pramanik Sollerkvist L, Tuan PL, Cam PD, Ehrnst A. Co-receptor phenotypes in newly-delivered women and infants, infected with HIV-1 CRF01_AE in northern Viet Nam. *Submitted to Journal of General Virology*
- IV. **Caridha R**, Khoan G, Fried U, Pramanik Sollerkvist L, Lindgren S, Navér L, Clevestig P, Ehrnst A. Transmitted HIV-1 variants in HIV-1 infected mother-child pairs, carrying different subtypes. *Preliminary results*

TABLE OF CONTENTS

Introduction.....	1
HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).....	2
The HIV-1 genome.....	2
The Replication Cycle.....	3
The natural course of an HIV-1 Infection and Disease Progression	6
The HIV-1 origin and Genotypes.....	7
HIV-1 MOLECULAR EVOLUTION.....	9
Genetic Variation.....	9
Selection Pressure and Methods to study HIV-1 genetic variation.....	10
HIV-1 co-receptor phenotypes.....	11
Glycoprotein gp120 interaction with the HIV-1 co-receptors.....	12
Phenotype determination – Tropism testing.....	13
HIV-1 TRANSMISSION.....	15
Mother-To-Child Transmission of HIV and Intervention programs.....	15
Transmissible/Founder HIV-1 viruses.....	16
Co-Receptor Use, Subtypes, and Mother-to-Child Transmission.....	17
Aims of Thesis.....	19
Material and Methods.....	20
PATIENTS AND SAMPLES	20
METHODS.....	20
Virus Isolation	20
Phenotypic Determination of Co-receptor Use.....	20
Sample Preparation for PCR	21
Nested <i>pol</i> PCR – Testing for antiretroviral resistance.....	21
Nested <i>env</i> V3 PCR.....	21
Single Genome Sequencing	22
DNA Sequencing	22
Phylogenetic Analysis	22
<i>In silico</i> phenotype determination	23
Sequences submitted to the Los Alamos HIV Database	24
Statistical Analysis.....	24
Ethical Aspects.....	24
Results and Discussion.....	25
Paper I.....	25
Counseling of HIV-1 infected pregnant women in northern Vietnam.....	25
Mother to child transmission of HIV-1 in northern Vietnam.....	26
Resistance to antiretroviral drugs in the maternal samples.....	28
Paper II.....	28
The epidemic of HIV-1 CRF01_AE in Vietnam and neighboring countries.....	28
Synonymous and non-synonymous analysis.....	30
Paper III.....	30
HIV-1 co-receptor use predictions for CRF01_AE infected pregnant women.....	30
Predicted phenotype characteristics of the infant HIV-1 <i>env</i> sequences.....	32
Geno2Pheno vs glycan-charge model?	33
Study IV.....	34
Mother-child pairs infected different HIV-1 subtypes.....	34
Characteristics of the transmitted viruses.....	35
Conclusions & Perspectives.....	40
Acknowledgements.....	42
References.....	44

LIST OF ABBREVIATIONS

aa	Amino acid
AIDS	Acquired immune deficiency syndrome
ARV	Antiretroviral prophylaxis
AZT	Azidothymidine: a nucleoside analog reverse transcriptase inhibitor
bp	Base pairs
C1-C5	Constant regions 1 to 5 of the <i>env</i> gene
CCR2	Chemokine (C-C) receptor 2
CCR3	Chemokine (C-C) receptor 3
CCR5	Chemokine (C-C) receptor 5
CRFs	Circulating recombinant forms
CXCR4	Chemokine (C-X-C) receptor 4
ds	Synonymous amino acid changes
dn	Non-synonymous amino acid changes
ECS	Elective caesarean section
EDTA	EthyleneDiamineTetraacetic Acid: an anticoagulant for blood
Env	Envelope protein
FPR	False positive rate
gag	Group-specific antigen group
GALT	Gut-associated lymphoid tissue
G2P	Geno2Pheno
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
HTLV-III	Human T lymphotropic virus III
IN	integrase
LAV	Lymphadenopathy virus
LTR	Long terminal repeats
MHC-1	Major histocompatibility complex-1
MTCT	Mother-to-child transmission
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NNRTI	Nucleoside reverse transcriptase inhibitor
NSI	Non-syncytia inducing
NVP	Nevirapine: a non-nucleoside reverse transcriptase inhibitor
PBMC	Peripheral blood mononuclear cells
PCR	Polymerase Chain Reaction

PhyML	Phylogenetic Maximum Likelihood
PMTCT	Prevention of mother-to-child transmission
pol	Polymerase gene
PR	Protease
PSSM	Position-specific scoring matrix
p24	Capsid
RT	Reverse transcriptase
SGS	Single genome sequence
SI	Syncytia inducing
SIV	Simian Immunodeficiency Virus
SIVcpz	SIV chimpanzee
UNAIDS	Joint United Nations Program on HIV/AIDS
URFs	Unique recombinant forms
vpr	Viral proteins
V1-V5	Hyper-variable regions 1-5 of the <i>env</i> gene
WHO	World health organization
3TC	Lamivudine is a nucleoside analog reverse transcriptase inhibitor

Introduction

In the early 1980's a new disease was identified in the United States and Europe that caused immunological dysfunction in men who had sex with men. This was demonstrated by the rise of opportunistic infections in this risk group, such as *Pneumocystis carinii* pneumonia, *Toxoplasma gondii* encephalitis, and a number of unusual cancers, like Kaposi Sarcoma. The disease was given the name acquired immunodeficiency syndrome (AIDS). In 1983 a retrovirus was first isolated from a lymph node taken from an AIDS patient with lymphadenopathy at the Pasteur Institute in Paris and shortly thereafter the same virus was also identified in the United States (Barre-Sinoussi F *et al.*, 1983; Chermann JC *et al.*, 1983; Gallo RC *et al.*, 1984; Levy J.A *et al.*, 1984; Popovic M *et al.*, 1984). The virus was initially named human T lymphotropic virus III (HTLV-III) and lymphadenopathy virus (LAV). A group of experts decided later on to rename the virus to human immunodeficiency virus 1 (HIV-1) (Coffin J *et al.*, 1986), which today is in use. Since 1983, it is estimated that more than 60 million people have been infected with HIV-1. More than 25 million have died of AIDS since the beginning of the HIV/AIDS epidemic (UNAIDS, *AIDS epidemic update: 2011*; UNAIDS, *2012 Report on the global AIDS epidemic*).

According to The Joint United Nations Program on HIV/AIDS (UNAIDS), the global incidence of HIV-1 infection has stabilized and begun to decline in many countries (UNAIDS, *AIDS epidemic update: 2011*). One contributing factor could be that the number of people receiving antiretroviral therapy continues to increase, with 6.65 million people being treated at the end of 2010. A total of 2.7 million people acquired HIV infection in 2010 and 1.8 million deaths by AIDS were recorded, contributing to the total number of 34 million people living with HIV-1 in 2010 (UNAIDS, *AIDS epidemic update: 2011*). The number of children living with HIV-1 globally has levelled off in the past few years and in 2010 there were 390.000 new infections in children under 15. It is estimated that there are 3.4 million HIV-1 infected children worldwide, where more than 90% are living in sub-Saharan Africa (UNAIDS, *AIDS epidemic update: 2011*; UNAIDS, *2012 Report on the global AIDS epidemic*).

Today according to UNAIDS, HIV-1 has spread to all continents and among the 34 million infected people, 22.9 million are residents in Sub-Saharan Africa (UNAIDS, *AIDS epidemic update: 2011*). These figures indicate that much work remains to be done to control the spread of HIV, including vertical transmission, in different regions of the world.

Background

HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1)

HIV-1 is a lentivirus, belonging to the retroviridae family of viruses. Lentiviruses propagate slowly into their hosts, where symptoms of the disease usually develop after longer periods of time. HIV-1 and HIV-2 are two related types of HIV and they are similar at a genomic level but differ in their epidemiology, transmission and pathogenicity. From here on the focus of this thesis will be on HIV-1.

The HIV-1 genome

The HIV-1 genome is approximately 9.2 kbp, consists of two positive single stranded RNA and encodes three groups of genes common to all retroviruses: group-specific antigen (*gag*), polymerase (*pol*) and envelope (*env*), coding for structural proteins and enzymes pivotal for HIV-1 replication (See Figure 1). HIV-1 regulatory proteins are coded by the *tat* and *rev* genes. Moreover, *nef*, *vif*, *vpr* and *vpu* code for the accessory proteins of HIV-1. In addition to two regions flanking the genome, which are called long terminal repeats (LTRs), a total of nine genes have been identified in the HIV-1 genome. Based on their different degree of conservation, HIV-1 genes and proteins have been widely used for genotype characterization, drug resistance studies, and diagnosis. High priority is given to proteins from the HIV-1 *pol* gene for clinical purposes, since many mutations related to drug resistance are located in this gene. The *pol* gene codes for the three viral enzymes: reverse transcriptase (RT), integrase (IN) and protease (PR). It is quite stable throughout all HIV-1 genotypes, with an estimated intra- genotype diversity of about 15%. The HIV-1 *gag* gene encodes the polyprotein p55, which can have as much as 20% diversity. The p55 is processed by the viral protease into p24 (capsid), which is very useful for early diagnosis of HIV-1 infection, the conserved p17 protein, which is a matrix protein that has been considered as a potential target for immunotherapy of HIV-1 infected cells (Shang F *et al.*, 1991), p7 (nucleocapsid) and p6. The *env* gene encodes the envelope glycoproteins (Env), gp120 and gp41, that are responsible for binding and fusion of HIV-1 to the target cells. A region of particular relevance is the third variable region of gp120 *env* gene called V3 loop. The *env* gene is the most variable (30%) and is commonly used for the definition of the HIV-1 genotypes (Brun-Vezinet F *et al.*, 1999; Leitner T *et al.*, 2005; Berry IM 2008).

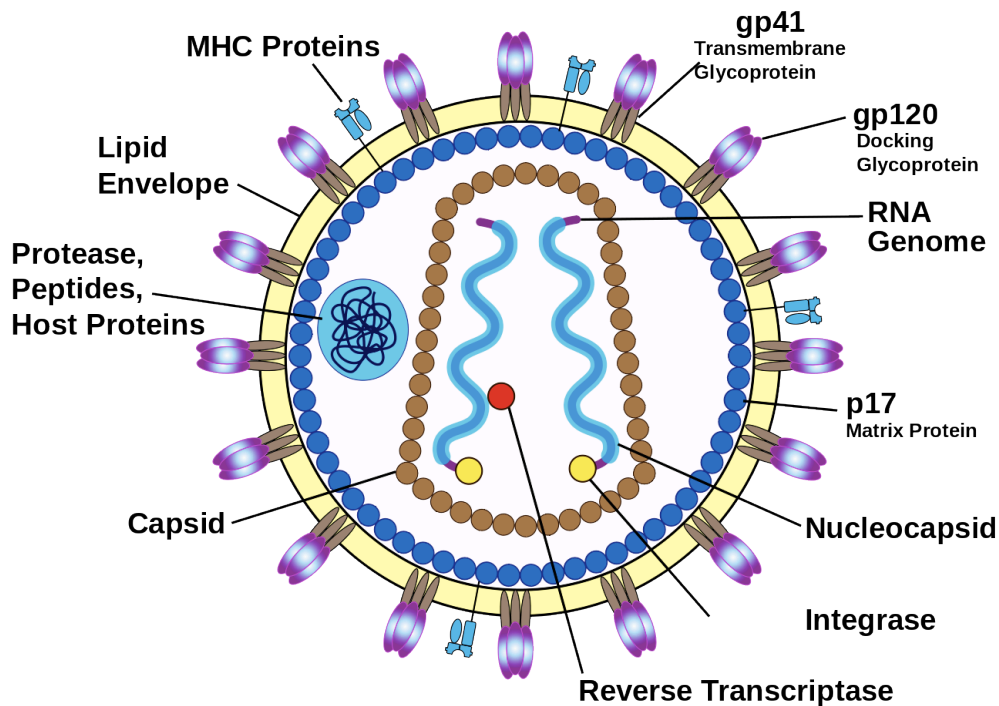


Figure 1. Illustration of an HIV-1 virion. (Adapted from NIH).

The Replication Cycle

Binding and entry

Through receptor-mediated membrane fusion, HIV-1 primarily infects T-lymphocytes, macrophages, dendritic and brain microglial cells, all of which express the CD4 receptor (Dalglish AG *et al.*, 1984; Klatzmann D *et al.*, 1984; Maddon PJ *et al.*, 1986; Habeshaw JA *et al.*, 1989) (Figure 2).

HIV-1 binding to the host cell occurs in a two-step process. It is initiated through high affinity binding of the trimeric gp120 to the primary host cell receptor, CD4. This interaction allows the gp120 trimer to undergo a conformational change and alters the positions of its V1/V2 and V3 loops (Wyatt R *et al.*, 1995; Trkola A *et al.*, 1996; Wu L *et al.*, 1996), forming the highly conserved co-receptor binding site (Rizzuto CD *et al.*, 1998). This site interacts with the CCR5 or CXR4 co-receptor, depending on the tropism of the virus. This interaction leads to an extensive conformational change in the trans-membrane gp41, and provides a mechanism for bringing the viral and host cell membranes closer together, allowing them to fuse and subsequently release the viral nucleocapsid into the host cell cytoplasm (Chan DC *et al.*, 1997; Weissenhorn W *et al.*, 1997).

Reverse transcription and integration

Once inside the cell, the viral core including viral genome, nucleocapsid core proteins, RT, and IN are released into the cytoplasm of the host cell. Thereafter DNA synthesis is initiated by reverse transcription by the RT through adding nucleotides to the 3' end

of the tRNA primer towards the end of the 5'LTR including the U5 and R regions. After it is completed, the minus strand DNA folds over into a circle and base pairs its 3' end primer binding site sequence with the complementary sequence of the shorter newly synthesized plus strand DNA strand. This allows the completion of the remaining plus-strand DNA, resulting in a complete double-stranded DNA genome including U3-R-U5 LTR sequences at both ends.

The double stranded DNA forms a pre-integration complex together with IN and other viral proteins (vpr and matrix proteins), and is actively trans-located into the nucleus. There it is integrated into the host cell genome in preferentially expressed genes by IN; in this form the viral DNA is named provirus (Brown PO, 1997). This provirus is the basis for retroviral infection, remains permanently associated with the host genome and is capable of establishing a persistent infection with constant production of new viral particles.

Transcription and translation

When integrated into the host cell chromosome, the provirus either remains latent or is transcribed by cellular RNA polymerase II when the cell is activated. The initial mRNA splice variants encode Tat, Rev and Nef. Tat binds a secondary RNA structure, enhancing RNA synthesis through the phosphorylation of RNA polymerase II. Nef appears to have several effects on host cell molecules and is responsible for down-regulation of the CD4 and CD28 molecules, which are internalized and degraded through the endolysosomal pathway (Aiken C *et al.*, 1994; Piguet V *et al.*, 1999), in contrast to *vpu*, which promotes proteolysis of CD4 in the endoplasmic reticulum. Nef also down-regulates molecules such as certain antigen presenting MHC-I and lipid presenting antigen CD1d (Scheppler JA *et al.*, 1989). This is likely a strategy developed by the virus to avoid detection by components of the immune system (Cohen GB *et al.*, 1999). It is also an important protein for the formation of mature infectious particles (Lama J, 2003).

The mRNA of late genes are trans-located out into the cytoplasm by the assistance of the Rev protein, either unspliced or partly spliced transcripts, where they are translated into structural proteins encompassing the viral core capsid, and the RT, PR and IN enzymes, producing the *gag* and *gag-pro-pol* precursor polyprotein (Swanstrom & Willis, 1997). The Env glycoprotein (gp160), also translated late, is properly folded and transported into the Golgi apparatus, where it is cleaved into gp120 and gp41, and the N-linked glycans are trimmed and modified into complex and hybrid types with sialic acid. The processed Env proteins are then transported to the host-cell surface and assembled as membrane bound trimers.

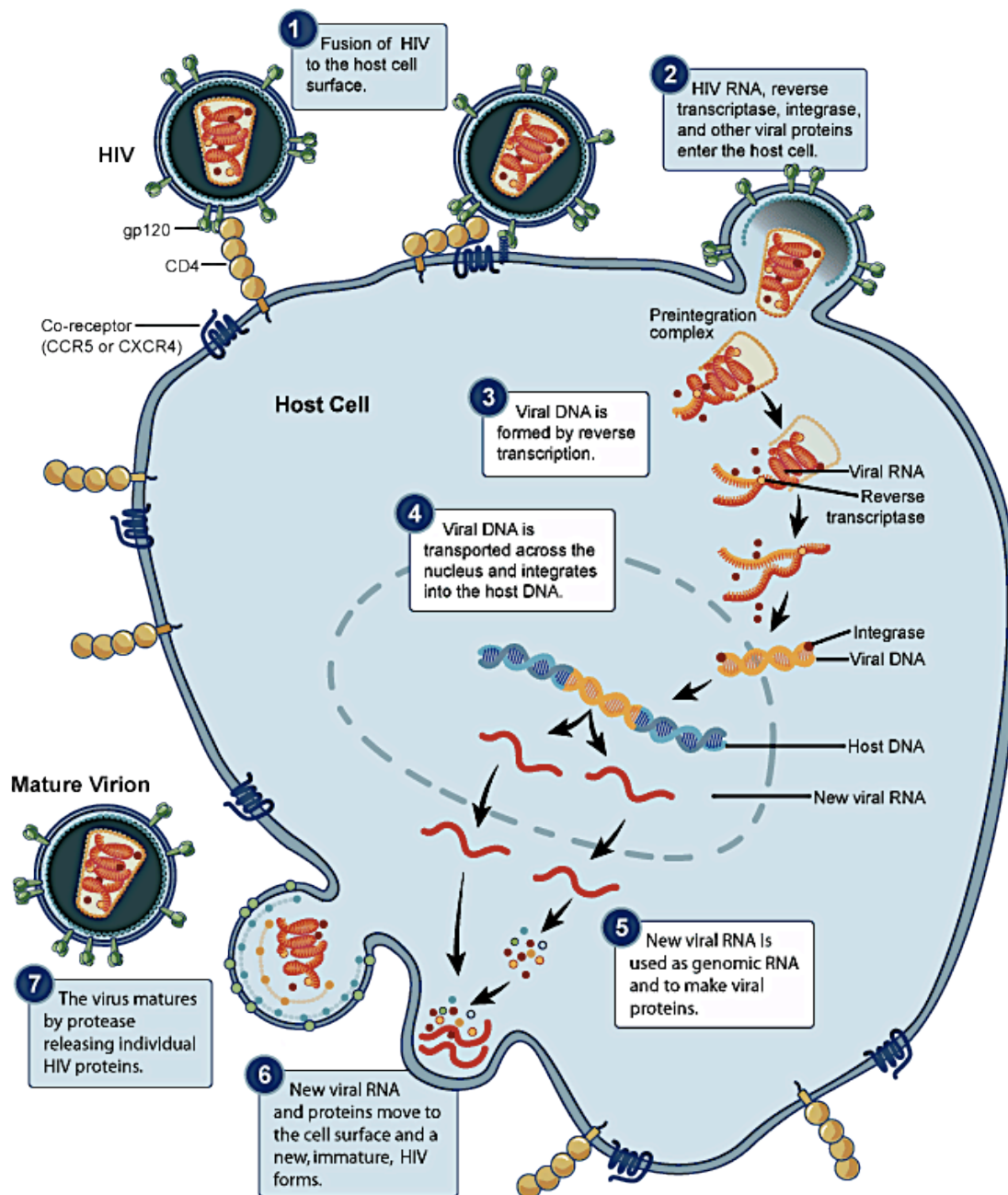


Figure 2. Steps in the HIV Replication Cycle: 1) Fusion of the HIV cell to the host cell surface, 2) HIV RNA, RT, IN, and other viral proteins enter the host cell, 3) viral DNA is coded by reverse transcription, 4) viral DNA is transported to the nucleus and integrates into the host DNA, 5) new viral RNA is used as genomic RNA and to produce viral proteins, 6) new viral RNA and proteins are assembled at the cell membrane and a new, immature virus buds out from the target cell, and 7) the virus matures and becomes infectious by viral protease cleavage of HIV precursor proteins. (Adapted from NIAID)

Assembly and release

Two copies of the *gag* and *gag-pro-pol* proteins assemble together at the cell membrane forming an immature capsid together with the viral RNA genome and cellular cyclophilin-A. The *env* proteins, already present on the cell membrane surface, are associated to the capsid during the budding process. Significant amounts of MHC and other membrane-associated proteins are gathered in the particle during the budding

process. As the particle finally buds from the host cell membrane, it is still an immature virus particle. The *pol* portion of the *gag-pol* polyprotein is cleaved into PR, IN and RT and the *gag* polyprotein is subsequently cleaved into p17, p24, p7 and p6, creating a mature and infectious virion.

The virus has developed three mechanisms to avoid premature interactions between viral glycoproteins and CD4 during the replication cycle; 1) through *vpu*, which removes newly synthesized CD4 from the ER, 2) through *nef* removing both surface CD4 and 3) MHC-1 (Garcia & Miller, 1992).

The natural course of an HIV-1 Infection and Disease Progression

The natural course of an HIV-1 infection in untreated individuals includes three main stages: the acute infection, clinical latency (or chronic phase) and progression to AIDS (Figure 3). During the acute stage of the infection, the viral load reaches a peak level and CD4⁺ T cells are greatly depleted (Schacker TW *et al.*, 1998; Guadalupe M *et al.*, 2003). In the gut-associated lymphoid tissue (GALT), around 80% of the CD4⁺ T cells are depleted (Alimonti JB *et al.*, 2003; Brenchley JM *et al.*, 2004). During the acute phase many individuals experience “flu-like” symptoms in parallel with peak viremia (Fauci A *et al.*, 1996). It takes a few weeks for the adaptive immune responses to mature and then the peak viremia drops to a steady state (Lyles RH *et al.*, 2000) also called set point, while there is partial recovery of the CD4⁺ T cells, especially in peripheral blood. During the chronic phase there is a slow but constant depletion of CD4⁺ T cells, which can last for years. The immune system is continuously activated due to on-going viral replication, however, this chronic immune activation is not only specific for the virus but translates into generally elevated levels of activated cells and cytokine levels, and this chronic immune activation eventually lead to exhaustion and general defects in immune responses (Grossman Z *et al.*, 2006). Several opportunistic infections might potentially arise. Current guidelines indicate that the AIDS phase begins when the CD4⁺ T cell count drops below 200 cells/mm³.

In the absence of antiretroviral therapy the median time from infection to development of AIDS-related symptoms has been estimated be approximately 10 to 12 years. However, a wide variation in rate of disease progression has been observed. Approximately 10 % of HIV-infected people in these studies have progressed to AIDS within the first 2 to 3 years following infection, while up to 5 % of individuals maintain stable CD4⁺ T cell counts and no symptoms even after 12 or more years of infection. A small percentage (1 %) of HIV-infected individuals spontaneously control viral replication in the absence of antiretroviral therapy. These individuals control viremia below the detection limit of standard viral load assays (50-75 copies/mL) for one year or more and are called elite controllers, elite suppressors or elite non-progressors (Deeks&Walker, 2007). Several known host genetic factors have been associated with HIV-1 control, such as different HLA alleles. B*35 HLA allele is associated with an

increased rate of progression, where patients progress to AIDS within 2-3 years (O'Brien SJ and Nelson GW, 2004). On the other hand, the HLA allele B*57, and to a less B*27, are associated with slower rate of HIV-1 disease progression (Kiepiela P *et al.*, 2004). Additionally, it has also been observed that plasma viral load and CD4+ T cell counts can be prognostic markers of HIV- 1 infection (Mellors JW *et al.*, 1997).

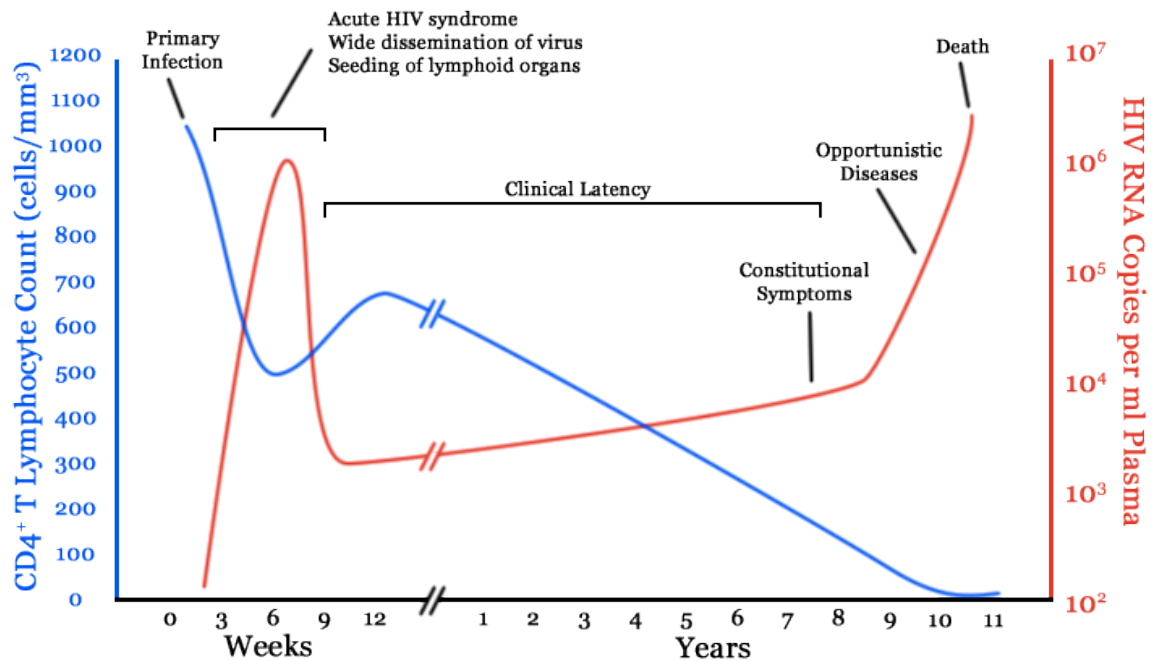


Figure 3. Representation of the clinical course of HIV-1 infection throughout the disease progression. CD4+ T cell counts are shown in blue and viral load levels in red. (Adapted from Fauci A. *et al.*, 1996)

Beginning in 1996, the initiation of highly active antiretroviral therapy (HAART) revolutionized HIV care. The introduction of HAART has dramatically improved the short-term survival of patients with HIV infection. This treatment includes potent combinations of three or more anti-HIV drugs and can reduce an infected individuals viral load to undetectable levels and in many cases delay the progression of HIV disease for prolonged period (Moore & Chaisson, 1999).

The HIV-1 origin and Genotypes

The origin of HIV group viruses has been traced to simian immunodeficiency viruses (SIVs), which were found in African apes (Gao F *et al.*, 1992; Gao F *et al.*, 1999, Keele BF *et al.*, 2006). It is known that more than 30 different species of non-human primates are natural hosts of species-specific SIV (Takehisa J *et al.*, 2009). Some of these SIVs have, through independent and multiple zoonotic transmission events, resulted in different HIV types and groups. There are two main types of HIV: Transmissions from West Central African chimpanzees, *Pan troglodytes troglodytes*) established the epidemic of HIV type 1 (HIV-1), and transmissions from sooty mangabeys (*Cercocebus atys atys*) established HIV type 2 (HIV-2) (Gao F *et al.*, 1999).

Phylogenetic analysis, using sequence data with known sampling times, have shown that the time to the most recent common ancestor for HIV-1 dates back in 1910s for HIV-1 and in 1940s for HIV-2 (Salemi M *et al*, 2009; Lemey P *et al*, 2001; Worebey M *et al*, 2008).

HIV-1 can be divided into three common groups: M (Major), O (Outlier) and N (Non M/Non O). Phylogenetic analysis revealed that groups M and N have their ancestor in SIVcpz, which was isolated from the Chimpanzee *Pan Troglodytes Troglodytes* in South Cameroon (Bartolo I *et al.*, 2008; Peeters M *et al.*, 2008; Taylor BS *et al.*, 2008). Group O instead originates from SIVgor of the gorillas in Western Africa (Van Heuverswyn F *et al.*, 2006; Taylor BS *et al.*, 2008).

Group M entered the human population along the Congo River (Taylor BS *et al.*, 2008), and traces of these viruses date back to 1940s (Zhu T *et al.*, 1998; Lemey P *et al.*, 2006). Groups N and O are endemic in West Equatorial African countries. The infection globally is dominated by group M and is divided in different subtypes: A, B, C, D, F, G, H, J and K, and at least 37 circulating recombinant forms (CRFs) of the virus have been found so far (Robertson DL *et al.*, 2000; Bartolo I *et al.*, 2008; Taylor BS *et al.*, 2008). Inter-subtype recombinant form is the denomination of a virus strain that is a hybrid from more than one CRF. In order for a subtype to be designated as a new subtype it needs to be found in at least two unrelated individuals in two geographically unrelated regions. The two sequences have to be similar to each other, but display enough variation from previously identified sequences throughout their whole genome.

There is a widespread geographical distribution of the different subtypes throughout the globe (Figure 4). HIV-1 subtype A and HIV-1 CRF02_AG dominate in Eastern Africa (Janssens W *et al.*, 2000), in addition to the HIV-1 subtype C and D. HIV-1 subtype B is most common in Europe, North America and Australia. HIV-1 subtype C is the most dominant subtype globally and is reported to be spread in China, India and Southern Africa (Neilson JR *et al.*, 1999; Taylor BS *et al.*, 2008). HIV-1 CRF07_BC and HIV-1 CRF08_BC are mixtures between subtypes B and C and were reported to be common in China (Su L *et al.*, 2000; McClutchan FE *et al.*, 2002; Qiu Z *et al.*, 2005). HIV-1 CRF01_AE is most common in Asian countries and is originated from HIV-1 subtype E that was first detected in Thailand in late 1980s. HIV-1 CRF01_AE is a hybrid between *gag* gene of HIV-1 subtype A and *env* gene of HIV-1 subtype E (McCutchan, FE *et al.*, 1992; Gao F *et al.*, 1996).

Most of the circulating recombinant forms are hybrids from two HIV-1 subtypes and are called HIV-1 CRF01_AE, 02_AG, 03_AC, AD, 07_BC, 08_BC, BF, BG, CD etc (Gao F *et al.*, 1996; Bartolo I *et al.*, 2008; Taylor BS *et al.*, 2008). There are also unique recombinant forms (URFs), which designate an HIV-1 form detected in one or many individuals in the same region. There is limited transmission of the URFs to the general population (Taylor BS *et al.*, 2008). This URF mosaic virus is labelled with

“U” until enough criteria are collected to designate a nomenclature for the virus (Robertson DL *et al.*, 2000).

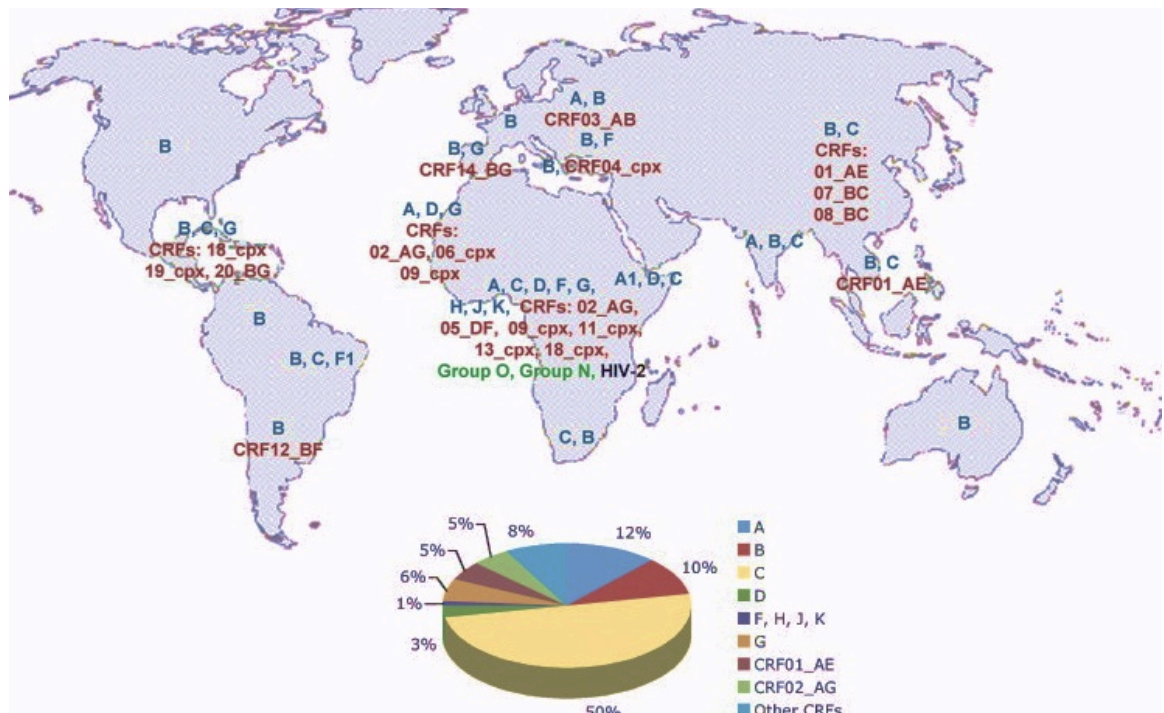


Figure 4. Geographical distribution of HIV genetic forms in 2007. The approximate location of the different HIV forms is indicated on the world map. HIV-1 group M subtypes are indicated in blue, while CRFs are given in red. The other HIV-1 groups and HIV-2 are indicated in green and black, respectively. The pie chart gives the prevalence of HIV-1 group M genetic forms; the global prevalence of each subtype and circulating recombinant form (CRF) is expressed as the percentage of the total number of group M HIV-1 isolates identified worldwide (Adapted from Ramirez BC *et al.*, 2008).

HIV-1 MOLECULAR EVOLUTION

Genetic Variation

HIV-1 is characterized by high genetic variability as well as rapid evolution and diversification, similarly to other RNA viruses (Seillier-Moisewitsch F *et al.*, 1994). It is through the elevated error rate of the reverse transcriptase, recombination and rapid HIV-1 turnover in infected individuals, that makes HIV-1 such a challenge to the development of vaccines and efficient antiretroviral drugs (Peeters M *et al.*, 2000). It has been estimated that the HIV-1 evolutionary rate is approximately one million times faster than the rate of cellular genes in higher organisms (Li WH *et al.*, 1988). The RT enzyme lacks proof reading and thereby introduces point mutations into the viral genome at a rate of 3.4×10^{-5} substitutions per site per replication cycle (Mansky LM *et al.*, 1995). Such an error rate results in the introduction of one nucleotide substitution to almost every newly synthesized viral genome, considering that the size of the HIV-1 genome is approximately 10,000 bases. Moreover HIV-1 genetic diversity is further escalated by recombination, which is the result of strand switching during reverse

transcription in infected cells. This enables exchange of genetic material and repair of the genome when breaks occur, which also involve completely different strains during super infection. This is also the basis for the existence of CRFs and URFs.

In addition to the general variability of HIV-1 within and between hosts, variation also occurs within different tissues, such as lymphoid organs and peripheral blood mononuclear cells (PBMC), and has been described as compartmentalization, and can probably be due to extreme immunological pressure and constant cell activation (Delwart EL *et al.*, 1997). The difference in variation within the HIV-1 genome is mainly depended on the function of the encoded proteins. The *env* is the most variable gene; in fact, in order to escape immune response the encoded envelope glycoproteins need a large variability. However, there are constant regions (C1-C5) interleaving with hyper-variable regions (V1-V5), within the *env* gene coding for gp120. The *pol* gene varies the least, mainly because the encoded enzymes have to maintain their function to enable replication, while the *gag* gene varies intermediately.

Another important factor is the high turnover rate of the virus *in vivo*, estimated to be up to 10^{10} virus particles per day at the peak viremia, contributing to the rapid evolution of the virus. The consequence of the factors mentioned above is that each HIV-1 infected individual harbours a uniquely diverse virus population, consisting of a pool of genetically distinct yet related viruses called quasispecies (Goodenow M *et al.*, 1989; Holland JJ *et al.*, 1992). The genetic variability of the HIV-1 quasispecies, can reach up to 10% nucleotide diversity within an infected subject. It also provides the viral population with the ability to adapt rapidly to environmental changes, such as immune response or antiretroviral drugs, and is the main challenge for the development of a vaccine and a treatment able to counter fight the infection.

Selection Pressure and Methods to study HIV genetic variation

Synonymous and non-synonymous substitutions

Changes in the environment can impose a selective pressure on evolving virus populations resulting in the fixation of genetic variants best adapted to the new milieu they were selected for. Such adaptation at the molecular level can be studied by analysing the relation between synonymous and non-synonymous substitutions. Synonymous substitutions, also called silent substitutions, do not alter the encoded amino acid and occur by random genetic drift. The synonymous substitution rate of any retrovirus is expected to be equal to the RT error rate, which is in turn proportional to the viral replication rate. Instead, non-synonymous substitutions result in a change of the encoded amino acid and may depend on selective pressure that can increase (positive selection) or decrease (negative selection) the fixation probability of specific amino acid changes. In this case non-synonymous substitution rates can be used as a measure of the adaptation rate (Salemi M *et al.*, 1999).

Phylogenetic tree inference

Phylogenetic trees can reflect the genetic variation and the relationship between sequences. In order to create a phylogenetic tree, first is needed to select a model of sequence evolution (substitution model). Several such models have been proposed to realistically describe sequence evolution by accounting for unbalanced base composition and mutation rates. One of the most complex substitution models is GTR (general time-reversible) model, in which each pair of nucleotide substitutions has different rates. For example, it assumes a time reversible symmetric substitution matrix in which A substitutes T with the same rate as T substitutes to A. However, mutations rates differ across sites of the genome. Several methods have been developed to account for these variation rates. The most commonly used adds a gamma-distributed rate parameter (G) to the substitutions model. In additions, information about invariant sites (I) can also be added to the model. This combination represents a complex model that often recapitulates HIV-1 evolution fairly realistically, known as GTR+G+I (Holder & Lewis, 2003). In general, the simplest model that adequately explains the data should be used. Several programs are available to obtain the best-fit model, such as FindModel (<http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html>).

Phylogenetic analysis can be used to study evolutionary relationship of different organisms or between strains of the same organism. It is possible to use phylogenetic trees for characterization of the fast evolution of HIV in evolutionary and epidemiological studies. The branching-pattern of the tree is called topology and the length of the branches describes their genetic distances. The four main methods to infer a phylogenetic tree are: Neighbor joining, Maximum likelihood, Parsimony and Bayesian inference (Holder & Lewis, 2003). Neighbor joining is broadly used, because it is a fast method and works well on closely related sequences. It creates a pair-wise joining distance matrix describing the evolutionary distance between sequences. In addition to the phylogenetic analysis, bootstrap analysis is a traditional method to assess confidence of the branches in the tree. During the bootstrap analysis the original alignment is randomly re-sampled with replacements to produce pseudo-replicate data sets. New trees are generated on these datasets and offer measurements of which part of the tree has higher or lower support. The main drawback of this method is the computational burden, since the analysis is repeated for each dataset (Holder & Lewis, 2003), i.e. at least 100 and often to 1000 times. It is important to note that a tree is the best attempt to explain the data given the model, which is not necessarily the same as the evolutionary history.

HIV-1 co-receptor phenotypes

The main receptor for attachment of HIV-1 to target cell is the CD4 molecule. HIV-1 isolates can also be designated according to the co-receptor that they use to enter into the target cell. Despite their great genetic variability, HIV-1 mainly uses two co-receptors. Viruses that use the CCR5 or CXCR4 co-receptor are called R5 or X4

viruses, respectively. Viruses that can use both co-receptors are called R5X4 or dual tropic viruses (Berger EA *et al.*, 1998; Huang W *et al.*, 2007; Taylor BS *et al.*, 2008). Some HIV-1 strains may also use other, alternative co-receptors, such as CCR2 or CCR3, as co-receptors (Matt C, 2001). These viral phenotypes are targets of new interventions and whether they differently affect disease progression and perhaps transmissibility is currently a question for research. Both CCR5 and CXCR4 are chemokine receptors, and more specifically seven trans-membrane G-protein coupled receptors. These chemokine receptors are found on a wide range of lymphoid cell types. T-lymphocytes, monocytes, macrophages and dendritic cells express CCR5, while CXCR4 is mainly expressed by T-cells, B-cells, myeloid, epithelial, endothelial and dendritic cells (Murdoch & Finn, 2000).

Current knowledge suggests that the R5 phenotype is favoured in the early infection events, such as in mother-to-child transmission (Scarlatti G *et al.*, 1997; Casper C *et al.*, 2002a). However in approximately 30-70% of chronically infected individuals who have progressed to AIDS, the presence of X4 viruses has been noted. A relation has been reported between the biological phenotype of HIV-1, virus transmission and diseases progression (Fenyo EM *et al.*, 1989; Berger EA *et al.*, 1998; Matt C, 2001; Scarlatti G, 2004; Ahmad N, 2005). In general, X4 viruses are associated with rapid disease progression, and appear to be transmitted at a low rate. R5 viruses occur early after infection, are more frequently transmitted and dominate until AIDS (Misrahi M *et al.*, 1998; Kiwanuka N *et al.*, 2008). In approximately 50% of AIDS patients, the X4 phenotype appears when the number of CD4⁺ T cells has been reduced, but this differs between the different subtypes. It is possible that X4 viruses may contribute directly to CD4⁺ T cell killing or emerge as the result of immunodeficiency.

Genetic studies of deficiencies in the CCR5 gene provide evidence that R5 viruses are more easily transmitted than X4 viruses. Individuals lacking the expression of CCR5 due to homozygous gene deletion (delta-32 deletion on the CCR5 gene) appear to be protected from HIV infection (Hutter G *et al.* 2009; Piacentini L *et al.*, 2009). Individuals with heterozygous CCR5 defects progress to AIDS slower than individuals with two normal copies (Matt C, 2001; Wasik TJ *et al.*, 2005; Verma R *et al.*, 2007).

Glycoprotein gp120 interaction with the HIV-1 co-receptors

The third variable region (V3 loop) of the gp120 has been shown to directly interact with the co-receptors (De Jong JJ *et al.*, 1992; Fouchier RA *et al.*, 1992; Shioda T *et al.*, 1992). The V3 loop is approximately 35 aa long with a loop structure held together by two cystein residues forming a disulfide bond and is found to have a net positive charge of between +2 and +10 (at pH 7.0). Substitutions of single amino acids within V3, for example the positive charged amino acids arginine and lysine, have been shown to influence co-receptor use in subtypes A, B, C, D and CRF01_AE (Fouchier RA *et al.*, 1992; Shioda T *et al.*, 1992; Verrier F *et al.*, 1999; Hu Q *et al.*, 2000a & 2000b; Briggs DR *et al.*, 2000). Such changes have been associated with changes towards

CXCR4 use, generating higher net charge for the entire V3 region (De Jong JJ *et al.*, 1992; de Wolf F *et al.*, 1994). In addition, it has been described that when X4 viruses evolve from R5 viruses, the sequon for the potential N- linked glycosylation site within V3 is lost (Pollakis G *et al.*, 2001; Polzer S *et al.*, 2001 and 2002). Others found this sequon to be necessary for CCR5, but not for CXCR4 use (Ogert RA *et al.*, 2001; Clevestig P *et al.*, 2006).

The V1/V2, C4 and V4 regions of gp120 are also of importance for co-receptor binding as they influence the accessibility of the V3 region (Cao J *et al.*, 1997; Wyatt R *et al.*, 1995). The mechanism for co-receptor binding to either CCR5 or CXCR4 is likely interplay between multiple regions on both gp120 and the chemokine receptor, with varying degrees of influence, which makes it difficult to identify the true nature of this interaction.

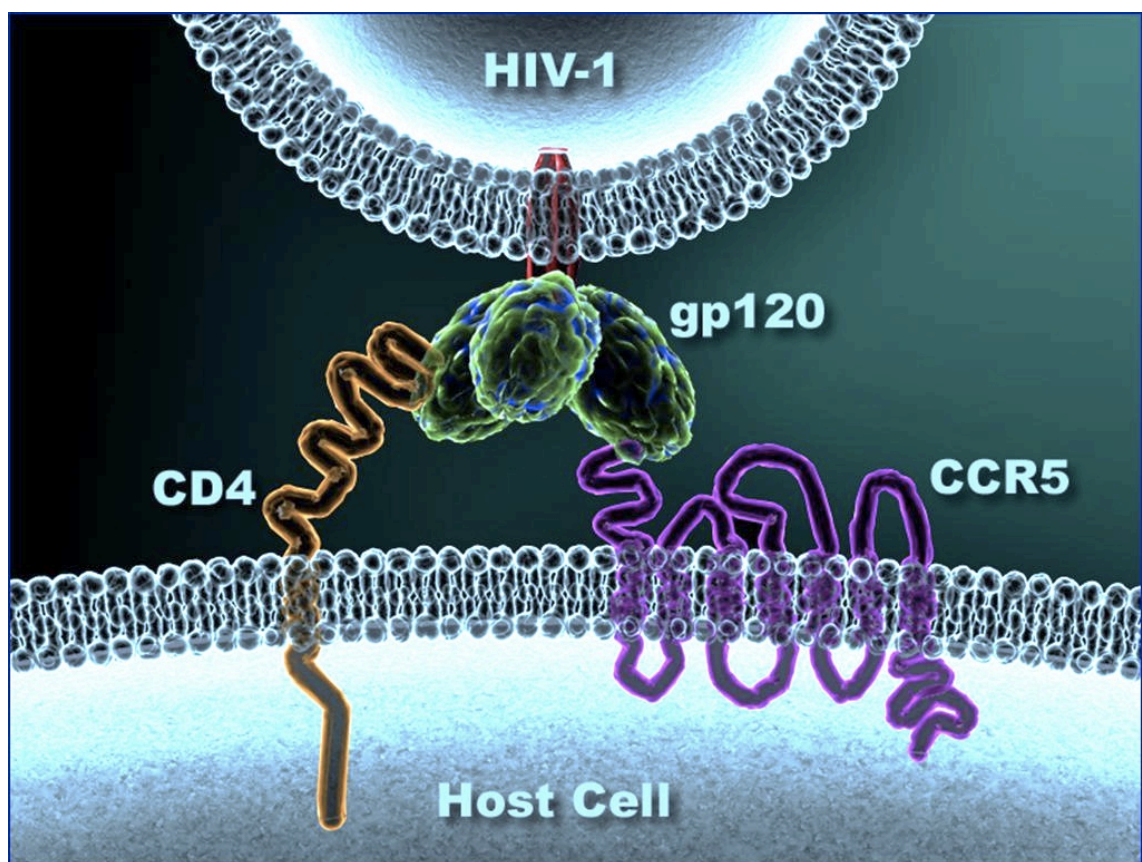


Figure 5. A simplified illustration of the binding between gp120, CD4 receptor and the co-receptor. Courtesy of Peter Clevestig.

Phenotype determination – Tropism testing

HIV phenotype refers to the ability of HIV-1 to enter CD4⁺ cells by the CCR5, CXCR4 or both co-receptors (tropism). Viral tropism can be predicted by genotypic studies, but the phenotypic determination is considered the golden standard.

Phenotypic assays

An early phenotypic classification of HIV-1 was based on the virus capacity to infect and replicate in established T-lymphoid and monocytoid cell lines (Åsjö B *et al.*, 1986). Viruses were characterized either as rapid/high, which grew rapidly to high titers in PBMC and established a productive infection in CD4 positive tumour cell-lines, or as slow/low, growing slowly with low titers and without the capacity to infect CD4 positive cell-lines. Additional methods to test the viral tropism were based on the capacity of primary isolates to form syncytia in PBMC and MT-2 T-cell lines, naming the viruses SI (syncytia inducing) or NSI (non-syncytia inducing) (Koot M *et al.*, 1993). These methods provided a system for categorizing HIV-1 of different tropisms and their results confirmed each other (Björndal A *et al.*, 1997).

Between 1995 and 1996 the role of the chemokine receptors as HIV-1 co-receptors were established (Cocchi F *et al.*, 1995; Feng Y *et al.*, 1996; Bleul CC *et al.*, 1996; Oberlin E *et al.*, 1996). Since then the viral tropism (phenotype) is described by the co-receptor use (Berger EA *et al.*, 1998). A modern phenotypic assay used today is the Trofile assay, which is included in the European guidelines on the clinical management of HIV-1 tropism testing (Whitcomb JM *et al.*, 2007). In this assay, the entire gp160 *env* gene is amplified directly from patient's plasma by PCR and cloned into an expression vector. In addition, this vector and a replication-defective proviral vector containing a luciferase reporter gene are co-transfected in a HEK293 cell line to produce a pseudo-virus stock. The pseudo-virus population is subsequently used to infect U87 cell lines expressing either CXCR4 or CCR5 co-receptor. Quantifiable light emission triggered by LTR-driven luciferase determines the infection. The reliability of this assay depends mainly on the sensitivity and accuracy of the cDNA synthesis and PCR and on the proportion of HIV-1 population amplified. The test can be done on both RNA and DNA but in Europe the commercial test is available only for plasma RNA.

Co-receptor prediction by genotypic testing

Genotypic testing is phenotype prediction based on sequences from the V3 region of the HIV-1 *env* gene of patients' samples (de Mendoza C *et al.*, 2008; Poveda E *et al.*, 2009; Seclen E *et al.*, 2010). Either population based sequencing or single genome sequencing approaches can be used for both viral RNA and DNA. There are three major bioinformatic interpretation techniques that can predict the phenotypes of the sequences, such as 11/25-charge rule, the position-specific scoring matrix (PSSM) and Geno2Pheno (G2P).

In brief, the 11/25-charge rule takes into account only the charge of the amino acids at key position 11 and 25 in the V3 loop (Resch W *et al.*, 2001). PSSM is a more advanced computer learning method, where the sequences' likelihood of being derived from an X4 virus for every possible amino acid at every individual position is calculated (Jensen MA & van 't Wout AB, 2003). The input sequences are compared

and aligned to subtype B sequences of known co-receptor use (e.g. X4). The better the fit, the higher PSSM score and the higher the score the higher likelihood that the sequence fragment has X4 properties. Sequences with values above -2.88 are considered X4, whereas sequences with scores below -6.96 are considered R5. PSSM can be accessed online at: <http://indra.mullins.microbiol.washington.edu/webpssm/>

The most broadly used prediction method is G2P[coreceptor] (Lengauer T *et al.*, 2007). This method is based on a statistical learning method, which is trained with a set of nucleotide sequences that corresponds to R5, dual/mixed tropism or X4 phenotypes. The result of interpretation is presented as a false positive rate (FPR), which defines the probability of false classification of an R5 virus as X4. There are many FPR cut offs optional to use, but the European guidelines on the clinical management of HIV-1 tropism testing recommend that a FPR of 5.75% should be used (Vandekerckhove LP *et al.*, 2011). This method can be accessed online at: <http://coreceptor.bioinf.mpi-inf.mpg.de/index.php>.

It is important to note that the overlapping between phenotypic and genotypic methods is not perfect (Garrido C *et al.*, 2008; Raymond S *et al.*, 2011; Chalmet K *et al.*, 2012) and it is common for the genotypic prediction tools to falsely predict R5 variants as X4 variants (Chalmet K *et al.*, 2012). Harrigan P *et al* previously compared PSSM and G2P with the Trofile assay. They found the sensitivities to be 56 and 63% and the specificities 90 and 91% for the two assays, respectively (Vandekerckhove LP *et al.*, 2011). Further attempts are required to improve and design better bioinformatic tools, as at the moment the predictions need to be interpreted with caution.

HIV-1 TRANSMISSION

The major route for HIV infection is through genital mucosal surfaces during heterosexual intercourse. Due to the protective environment of the genital tract being able to neutralize the majority of viral particles, the risk of being infected is less than 1:500, although women are twice more likely to be infected than men and more than 80% of infections are caused by only one single virus breaching through the mucosal defenses (Hladik F *et al.*, 2008 and 2009). For other routes of infection the risk is significantly higher. This includes intravenous blood-borne transmission, mother-to-child transmission, and transmission via anal intercourse. (Hladik F *et al.*, 2008 and 2009).

It has been suggested that in primary HIV-1 infection, a relatively homogeneous virus population resides in the beginning and diversifies into a heterogeneous population over time. As the individuals progress into AIDS, the population reverts into homogeneity (Goodenow M *et al.*, 1989; McNearney T *et al.*, 1992; Wolfs TF *et al.*, 1992; Shankarappa R *et al.*, 1999). In vertical transmission, minor subsets of maternal variants were also shown to be transmitted (Wolinsky SM *et al.*, 1992; Scarlatti G *et al.*, 1993). However, other studies have suggested that primary infection in children can occur with multiple variants (Nowak P *et al.*, 2002), with only the major variant

population detectable at the time of HIV-1 diagnosis (Dickover R *et al.*, 2001).

Mother-To-Child Transmission of HIV and Intervention programs

Mother-To-Child Transmission (MTCT) can occur in utero, during delivery and during breastfeeding and all these events have a cumulative effect over time (Dunn DT *et al.*, 1995; Coutoudis A *et al.*, 1999). In the absence of any intervention the cumulative HIV transmission is around 25-45% with 5-10% of infections occurring in utero, about 10-15% intrapartum and 10-20% through breastfeeding. Transmission rate increases with prolonged breastfeeding, between 0.51-1.57% per month of breastfeeding, depending on the woman's immune status and CD4 cell count levels (Bryson YJ *et al.*, 1992; Dunn DT *et al.*, 1995; Coutoudis A *et al.*, 1999; Kourtis AP *et al.*, 2006; Lehman&Farquhar, 2007; McDonald C *et al.*, 2007; Ahmad N 2008; WHO guidelines, 2010). With no breastfeeding, the overall transmission is 15-25%, which can be subdivided into 5-10% in utero and 10-15% intrapartum (Lehman&Farquhar, 2007; Ahmad N 2008). Approximately 40% of the infections occur in utero and 60% of the infected children acquire the infection at the time of delivery (De Cock KM *et al.*, 2000).

According to UNAIDS latest reports an estimated 390 000 children became infected with HIV in 2010 (UNAIDS, 2011). More than 90% of children living with HIV have acquired the infection through MTCT.

The rate of vertical HIV-1 transmission can be reduced by intervention with antiretroviral prophylaxis (ARV), elective caesarean section (ECS) and no breastfeeding (Grosch-Worner I *et al.*, 2000; Luzuriaga & Sullivan, 2005; Kourtis AP *et al.*, 2006; Newell ML 2006; Gray GE, 2008). By combining these interventions and by satisfying adherence of the mothers with the components of the prevention of mother-to-child transmission (PMTCT), the risk of MTCT in high-income countries has been reduced to about 1-2% (UNAIDS reports, 2011). The intervention package recommended by WHO contains: counselling, testing, provision of triple ARV therapy administered to HIV-infected women before, during and after delivery, Caesarean section, and provision of prophylactic ARVs to new-borns and replacement infant feeding (Newell ML 2006; UNAIDS 2011-2015). Many of these procedures are difficult to apply in developing countries because of the economic and social conditions, ethical factors and poor knowledge. In areas where clean water is hard to access, most women cannot provide safe replacement feeding to their infants and therefore are recommended to employ exclusive breastfeeding.

Despite the limited access in low-income settings to combination ARV regimens and the capacity to provide elective caesarean section and replacement feeding, the rate of MTCT can still be reduced to around 5%, if women are tested and enrolled in prevention programs (Kilewo C *et al.*, 2008; Namukwaya Z *et al.*, 2010). WHO and the UNAIDS have developed a plan towards the elimination of new HIV infections among children by 2015, in order to reduce the number of children being infected with HIV.

The aim is to reduce mother-to-child transmission of HIV globally to less than 5% (UNAIDS 2011-2015).

Transmissible/Founder HIV-1 viruses

A question of great importance is whether transmitted viruses have particular phenotypic properties that favour their transmission. If so, viruses with these properties should be targets of vaccination and microbicide efforts. The Env protein is a likely candidate for transmission related signatures. It has been shown that R5-viruses are transmitted far more frequently than those that utilize CXCR4 (Roos MT *et al.* 1992; Schuitemaker H *et al.*, 1992; Zhu T *et al.* 1993; Keele BF, *et al.*, 2008). Variations in *env* have also been linked to differences in the utilization of CD4 and co-receptor, the rate and efficiency of membrane fusion, and binding to C-type lectins such as DC-SIGN that are expressed on dendritic cells (DCs) and can function as virus attachment factors (Geijtenbeek TB *et al.*, 2000; Reeves JD *et al.*, 2002; Puffer B *et al.*, 2004).

Despite continuous efforts, few studies have been able to clone and compare transmissible Envs from acutely infected individuals in different subtypes. Derdeyn *et al.* studied subtype C Envs from eight heterosexual transmission pairs and found that there were fewer putative N-linked glycosylation sites (PNGs), more compact variable regions, and enhanced neutralization sensitivity to donor plasma (Derdeyn CA *et al.*, 2004), but no functional differences have been revealed. Recent comparisons of thousands of subtype B transmitted and chronic *env* sequences confirmed significantly fewer total PNGs and a trend toward fewer PNGs in the V1/V2 loops of these transmitted Envs (Gnanakaran S *et al.*, 2011). In addition, a study of subtype A and D transmission pairs also identified shorter recipient Envs with a lower V3 charge, although no differences in the number of PNGs were noted (Sagar M *et al.*, 2009). These discrepancies might have occurred due to differences in sample size, demographic characteristics of acutely infected individuals, cloning strategy, and whether the Envs under investigation represented the actual transmitted viruses.

Studies to characterize the properties of transmitted HIV-1 strains face several challenges. Since it is difficult to identify adult individuals during the acute phase of HIV-1 infection, MTCT provides an ideal setting. In addition, individual viruses cloned from the PBMCs or plasma of acutely infected individuals within weeks from transmission may have already evolved away from the actual transmitted virus and have acquired phenotypic changes (Borrow P *et al.*, 1997). Furthermore in the absence of extensive sampling of the early viral quasiespecies by single-genome amplification, it is impossible to know if one or more virus strains established the clinical infection (Keele BF *et al.*, 2008).

Co-Receptor Use, Subtypes, and Mother-to-Child Transmission

Many research projects are aimed at finding the association between co-receptor use and subtype in order to explain why some subtypes are more easily transmitted or lead

to more rapid disease progression than others. Difference in co-receptor use and disease progression between individuals infected with different subtypes of HIV-1 has been suggested (Esbjörnsson J *et al.*, 2010). In approximately 20-50% of the individuals infected with HIV-1 subtype B CXCR4-using viruses may emerge; this percentage is lower in HIV-1 subtypes A and C whereas subtype D uses CXCR4 more frequently (Tonie Cilliers JN *et al.*, 2003; Church JD *et al.*, 2008; Taylor BS *et al.*, 2008). Studies in Uganda showed that HIV-1 subtype D is more often associated with faster progression and higher mortality rate than other subtypes, and HIV-1 subtype B had a lower transmission rate and less progression (Kiwanuka N *et al.*, 2008). Studies conducted in South Africa concluded that HIV-1 subtype C spreads more rapidly than other subtypes. In contrast, no significant differences in disease progression between HIV-1 patients infected with A, B, C and D were observed in a study in an older Swedish study (Alaeus AL *et al.*, 1999).

Studies with large sets of samples, where subtypes and co-receptor use were determined, have shown that CXCR4-using viruses less often emerge in individuals that are infected with subtype C, while there is a tendency towards higher frequencies of CXCR4-using viruses in those individuals infected with subtype D (De Wolf *et al.*, 1994; Tscherning C *et al.*, 1998; Björndal A *et al.*, 1999; Tscherning-Casper C *et al.*, 2000). Other studies have suggested that viruses of different subtypes have preferences in their transmission pathways. One such study suggested that that subtypes C and CRF01_AE are better adapted to sexual transmission than subtype B (Mastro TD *et al.*, 1997). A second study suggested that subtype B is associated with homosexual transmission while subtype C is associated with heterosexual transmission (van Harmelen J *et al.*, 1997).

Differences in MTCT rates for subtypes have also been reported. In one study performed on a cohort in Western Kenya, a higher rate of transmission in mothers infected with subtype D compared to subtype A was shown (Yang C *et al.*, 2003). A more recent analysis of the same cohort showed a higher *in utero* transmission rate for subtype C than for subtypes A, D and A-env or D-env recombinants (Renjuji B *et al.*, 2004). It has also been suggested that subtype C virus displays a higher vertical transmission rate than other subtypes in a study on a Kenyan cohort, showing that women infected with subtype C virus were more prone to mucosal (vaginal) shedding of virus than women infected with subtypes A, and D (John-Stewart GC *et al.*, 2005).

Other studies have shown little clinical relevance for subtypes and for the distribution of phenotypes (Alaeus A *et al.*, 2000) or when comparing subtypes with MTCT rates (Morgado MG *et al.*, 1998; Tapia N *et al.*, 2003). There are many studies showing that R5 is found early in infection in mother-to-child transmission (Scarlati G *et al.*, 1997; Casper C *et al.*, 2002a; Clevestig P *et al.*, 2005). However, it has been difficult to identify sufficient numbers of pregnant women infected with X4 strains, and hence it is a challenge to investigate the role of R5 and X4 viruses in MTCT (Arroyo MA *et al.*, 2002; Casper C *et al.*, 2002a; Clevestig P *et al.*, 2005; Church JD *et al.*, 2008;

Huang W *et al.*, 2009; Kittinunvorakoon C *et al.*, 2009; Matala E *et al.*, 2001; Salvatori F *et al.*, 2001; Sato H *et al.*, 1999; Tcherning-Casper C *et al.*, 2000; Zhang H *et al.*, 2002).

Whether subtypes have in fact implications for the emergence of specific phenotypes, differences in transmission rates, and/or in the progression to AIDS, still remains to be determined. However, the varying amounts of CXCR4-virus observed among different subtypes and the association between X4 and a more rapid disease progression be taken into consideration.

Aims of the thesis

Aims of this thesis were:

- I.** To study HIV-1 transmission from mother-to-child in the north of Vietnam and to test drug resistance of the viruses.
- II.** To follow the evolution of the HIV-1 epidemic within northern Vietnam and in relation to its neighboring countries using to HIV-1 sequences.
- III.** To investigate the predicted co-receptor use phenotype, by bioinformatic methods, employing a newly developed model.
- IV.** To identify the transmitted/founder virus in mother to child transmission and seek characteristics, which may be subtype specific or common across subtypes.

Materials and methods

The sections below provide a brief overview of the main methods in **Paper I- IV**. More detailed information about the specific methods can be found in Materials and Methods in the respective papers

PATIENTS AND SAMPLES

Papers I-III: In northern Vietnam a cohort of 135 pregnant women/mothers participated in a prospective follow up of their children up to the age of 12 to 18 month, from 2005-2007. Samples were collected to identify the women's HIV-1 status. This material was also used to define the HIV-1 genotype in this population. Venous blood was collected in EDTA. Nevirapine (NVP) was given to most women at delivery and to the newborn child. The women did not breastfeed. Most of the transmissions had occurred *in utero*, which was established by a positive PCR test at birth.

Paper III: In a period just preceding the study of the pregnant women, 13 samples from 12 untreated HIV infected infants, were collected. Venous blood was collected at birth, 1 month, 3 months, and up to six months of age in Hanoi. This is a separate cohort.

Study IV: Eight HIV-1 infected mother-child pairs from Sweden were included in this study. Two of these pairs were infected with subtype A, three with subtype C and three with subtype CRF01_AE. Samples were prospectively collected from different time points during pregnancy, delivery and 6 months afterwards. Their children were followed from birth and at regular intervals up to 18 months of age (Lindgren et al., 1993).

METHODS

Virus Isolation (Study IV)

Viruses in PBMC were isolated by co-culturing with phytohemagglutinin stimulated PBMC from two healthy blood donors (Ehrnst A *et al.*, 1988). Virus stocks were obtained through passage through donor PBMC and infection was tested by env V3 PCR.

Phenotypic Determination of Co-receptor Use (Study IV)

Co-receptor use was determined by infecting U87 astrogloma cell lines expressing CD4 and chemokine receptors CCR5 or CXCR4 (Study IV) and other co-receptors CCR1, CCR2b, CCR3 (Deng H *et al.*, 1996; Deng H *et al.*, 1997; Berger EA *et al.*, 1998). Additional co-receptor use determination was performed on GHOSTcells expressing CCR5, CXCR4 or orphan receptors BOB or BONZO (Casper C *et al.*, 2002a).

Sample Preparation for PCR (Papers I to III; Study IV)

Prior to PCR assays, two million infected PBMCs or U87 cells were treated with a lysis buffer containing proteinase-K at different temperatures. First, this procedure enables the proteinase-K to break down the cellular structures and expose the cell DNA, and later inactivates both the virus and the active enzyme, rendering the specimen safe for regular laboratory work and of good quality for PCR. For some of the samples from Vietnam, plasma was separated from PBMC by centrifugation in Lymphoprep (Axis-Shield PoC AS, Oslo, Norway). RNA was extracted from 140-200 µL plasma by QIAamp (QIAGEN GmbH, Hilden Germany). Viral RNA was used as a template to run cDNA synthesis using the outer primer JA170 by Fermentas kit.

Nested *pol* PCR – Testing for antiretroviral resistance (Paper I)

Mutations in the HIV *pol* gene mediating resistance to antiretrovirals were investigated by using PBMC DNA in samples from 23 women. The primers had a broad specificity for different HIV-1 subtypes and the protocols were adopted from Steegen K *et al.*, 2006. An outer *gag-pol* nested PCR product was obtained using the outer primers GAG2, PR1, RT137, and RT3303. The inner primers RT1 and RT4 spanned the amino acids 30–227 of the *pol* gene and the length was about 646 bp for the reverse transcriptase gene.

Nested env V3 PCR (Papers II-III; Study IV)

A nested PCR of the V3 region of gp120 was used as a basis for classifying HIV-1 M group subtypes, for DNA sequencing of population sequences, to create single genome sequences, and lately for surveillance of HIV-1 infection in PBMC cultures and in co-receptor use determinations.

Primers

JA167 Outer 5'-TATCTTTTGAGCCAATTCCTATACA-3'

JA 170 Outer 5'-GTGATGTATTTCARTAGAAAAATTC-3'

JA 168 Inner 5'-CAATG(C/T)ACACATGGAATTA(A/G)GCCA-3'

JA 169 Inner 5'-AGAAAAATTC(C/T)CCTC(C/T)ACAATTAAA-3'

PCR1: Ten µL DNA were amplified in a final volume of 50 µL containing 5 µL MgCl₂ 25mmol/l, 5 µL PCR buffer, 1 µL dNTP 2.5mmol/l, 0.2 µL Taq and 0.5 µL of each primer (10mol/l) and H₂O. PCR was run for 40 cycles of 92°C/30sec, 50°C/30sec and 72°C/30sec with the denaturation at 92°C/1min and incubation at 72°C/1min.

PCR 2: Five µL of DNA from PCR 1 were amplified in a final volume of 50 µL containing 5 µL MgCl₂ 25mmol/l, 5 µL PCR buffer, 1 µL dNTP 2.5mmol/l, 0.2 µL Taq and 0.5 µL of each primer 10mol/l and H₂O. PCR was run for 40 cycles at 92°C/30sec; 55°C/30sec, 72°C/30sec with the denaturation at 92°C/1min and

incubation at 72°C/1min. Over time different polymerase enzymes were used and the protocols were adapted to the respective annealing temperatures.

In studies I-IV, the PCR product was visualized in 1.5 % agarose gel after adding GelRed (Bio-Nuclear AB, Bromma, Sweden). The amplified DNA from microcentrifuge tubes was purified with Qiaquick PCR purification kit (Qiagen GmbH, Hilden, Germany) and the procedure followed the Qiaquick standard protocol, except for the elution step. Elution buffer was exchanged to 30 µl Gibco water (Life Technologies, Carlsbad, CA, USA) instead. Samples that were performed with 96-well reaction plate were purified in a 96 well Multiscreen® PCRµ96 (Millipore corporation Billerica MA, USA).

Single Genome Sequencing (Paper III and Study IV)

To obtain single genome V3 sequences from PBMC lysates, we performed a limiting dilution PCR with a quadruplet set of four-fold dilutions. The nested PCR described above was used and an agarose gel electrophoresis conducted on all samples. A dilution factor was determined, which would yield about one-quarter PCR positive results, of which the majority presumably would represent single DNA molecules, according to the Poisson distribution. About 50 replicas or more of this dilution were run in PCR to yield a sufficient number of single genome sequences. To be defined as such, the V3 section of the sequences must lack ambiguous nucleotides. Sequences with more than two ambiguity options in V3, were discarded completely.

DNA Sequencing (Papers I and Study IV)

Sequencing PCR was performed on all positive PCR reactions. This cycle sequencing PCR was conducted using both inner primers JA168 or JA169 in separate tubes or plates to provide two complementary sequences from each sample. Cycle sequencing PCR was performed, using the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, CA). The products were submitted to MWG Operon Eurofins Company, Ebersberg, Germany. Editing of the sequences was carried out with Sequencher 4.1 software programme (Gene Codes, Ann Arbor, MI, USA). Assembly and alignment were done with ClustalX 2.0.10, developed by the European Bioinformatics Institute and available at [<http://mac.softpedia.com/get/Math-Scientific/ClustalX.shtml>].

Phylogenetic Analysis (Papers II-III; Study IV)

Phylogenetic analysis of genetic material (DNA, RNA) is a method to study epidemiologically important relations and differences. It has proven to be valuable in the characterization of the HIV epidemic both geographically and with regard to changes over time and of differences in different risk groups (Albert J *et al.*, 1994; Leitner T *et al.*, 1996)

In paper II, phylogenetic tree analysis was used to determine the phylogenetic relationship between the new and present CRF01_AEenvV3 sequences from the pregnant women to sequences characterized in the past in Vietnam and neighboring countries. In paper IV, the phylogenetic relation of the virus population(s) in the mother and child was investigated after creating a large set of single genome sequences representing the HIV-1 present in PBMC and/or in plasma and relating them to the sequences of viral isolates of known co-receptor use (CCR5 and/or CXCR4). The sequences analyzed phylogenetically, included not only the V3 region, but also its flanking parts, 360bp long total. A maximum likelihood phylogenetic tree was inferred using PhyML v3.0 (PHYlogenetic inferences using maximum likelihood), using a GTR + G + I model with BioNJ program initial tree and an NNI search. Tree diagrams were plotted with FigTree v1.1.2, available at:

[<http://mac.softpedia.com/get/Graphics/FigTree.shtml>].

Moreover in paper II, another way to study selection pressure on the virus population would be to describe the synonymous (ds) and non-synonymous (dn) amino acid changes, where the synonymous changes would represent a background of mutations and non-synonymous would increase in relation to different pressures and both would accumulate over time. Ds and Dn values were calculated for all pairwise sequence comparisons within each geographically determined group using the SNAP program (Los Alamos HIV database), <http://www.hiv.lanl.gov26> according to the Nei and Gojobori method (Korber B, 2000; Nei M and Gojobori T, 1986).

***In silico* phenotype determination (Papers III and Study IV)**

Sequence based models have been developed to predict the HIV tropism (phenotype / co-receptor use). In papers III and study IV we compared the predicted phenotypes defined by the glycan-charge model with those obtained by Geno2Pheno. The glycan-charge model was deduced from Clevestig P *et al.*, 2006 as presented in Table 2. In brief, this model is based on the critical presence of the N-linked glycosylation motif within V3 for CCR5 use, while CXCR4 use was characterized by an increased positive charge of V3 amino acids. The net charge was obtained by counting all charged amino acid residues in V3, where R and K were each counted as +1, H as +0.1 and D and E each as -1. The V3 charge pattern was defined for each of the major subtypes of HIV-1 by Pramanik L *et al.*, 2011, based on sequences from HIV-1 strains, where the co-receptor use had been determined biologically. We chose to adopt the new cut-off values for CRF01_AE (Pramanik L *et al.*, 2011). Since we lack a biological marker for co-receptor use, we have described the phenotypes as putative.

Table 2. Glycan-charge model deducted from Clevestig P *et al.*, 2006

		V3 sequence charge		
		≤ 4.9	5-6.1	≥ 6.2
V3 glycan	NXT+	R5	Mixed	X4
	NXT-	Defective	X4	X4

Our interpretation of the data for CRF01_AE (Pramanik L *et al.*, 2011) using the glycan-charge model was as follows. Putative monotropic R5 strains would be characterized by the presence of the V3 glycan motif and V3 charge ≤ 3.9 . The putative monotropic X4 strains would express charge $\geq +5.1$, irrespective of the presence or absence of the glycan, and the dual-tropic sequences would possess both the N-linked glycosylation motif and charge in the range of 4.0 - 5.0 (Clevestig P *et al.*, 2006, Pramanik L *et al.*, 2011). However, the latter pattern could also represent R5 or X4 strains, as the glycan-charge model cannot discriminate between the phenotypes in sequences with both the N-linked glycosylation motif and charge in the range of 4.0 - 5.0. We have used the term “mixed phenotype” for them. When the criteria for neither CCR5 nor CXCR4 use were met, we designated the strains as putative defective/noninfectious.

The Geno2Pheno method is an often-used bioinformatic model for prediction of co-receptor use phenotypes and has been recommended for clinical use (Vandekerckhove LP *et al.*, 2011). It is recommended that you choose a cut-off, according to whether you want to maximize the sensitivity in detecting X4 sequences or minimize the false positive rate in the determination of X4 (Vandekerckhove LP *et al.*, 2011). We chose a cut-off of 5% false positive for population and single genome sequences to minimize the false positive results.

Submitted sequences to the Los Alamos HIV Database (Papers II-III)

For the published and submitted papers the sequences were submitted to the Los Alamos HIV Database. In paper II the maternal sequences received the accession numbers FJ009330 – FJ009366 and in paper III: JN987664 – JN987856, and JX021505 – JX021524. The infant sequences had the accession numbers JQ023668 – JQ023727.

Statistical Analysis (Papers I-III; Study IV)

Statistical methods were used in all papers. The non-parametric double-sided Fisher's exact test, McNemar's test, and student t-test were used to assess statistical probability in our data.

Ethical Aspects

Ethical permission for the studies presented in this thesis was given by the Ethics Committee of the Karolinska Institutet. All subjects were treated respectfully in accordance with the Declaration of Helsinki 1975, revised in 2000. Signed informed consent was obtained from each study participant (the mothers signed a personal permission regarding the children samples).

RESULTS & DISCUSSION

Paper I

Counselling of HIV-1 infected pregnant women in northern Vietnam

Ten years ago, the Ministry of Health in Vietnam developed a Volunteer Counselling and Testing program in each district, city and province. As a result of this program all Vietnamese citizens can get counselling, testing for HIV and condoms free of charge without reporting. This policy was expected to result in the early detection of HIV-1 infection in the population. In the hospitals, counselling and ARVs were provided free of charge for every HIV-infected person. Following the WHO treatment guidelines, the counselling and testing of HIV-infected pregnant women were set-up and run in the Obstetrics and Gynaecology hospitals in Hanoi and Haiphong (Oosterhoff P *et al.*, 2008). The HIV screening test was offered to all pregnant women at 7-8 months of gestation in Vietnam.

The study in paper I was initiated two years after the start of the national program with the purpose of investigating the uptake and effect of confidential counselling and antiretroviral prophylaxis in HIV-1 infected women in northern Vietnam to prevent mother-to-child HIV-1 transmission.

Counselling is usually the first step in approaching the HIV-infected patient and has a key role in the PMTCT. The gaining of the HIV patients' trust and confidence is an important step when advocating for strict compliance with prophylactic measures and follow-up. In this study we collaborated with Dr. Ha TT Tran who was trained in Sweden by her co-supervisor Anneka Ehrnst in the procedure of counselling HIV-1 infected pregnant women and also on the relevance of follow-up of the children at birth, 1-2 months after birth, and at the later time point, 12-18 months after birth, for a final determination of the HIV-1 status in the child. The consecutive sampling also allowed the determination of the time of transmission in infected children.

The study included HIV-1 infected pregnant women who attended the delivery hospitals in Hanoi and Haiphong. In the years 2005-2006, 234 women were given counselling, and 182 mothers agreed to enrol in our study; 135 of 182 (74.2%) mothers from Hanoi ($n=101$) and Haiphong ($n=34$) participated in the study and gave birth to 135 children that were prospectively followed from birth up to 12-18 months of age. Each woman was given ≈ 6 hours of counselling during 3-7 visits. Questions on how they contracted HIV-1 infection were not asked to the mothers, in order to focus on creating a situation of confidence and trust with the women included in the program. Counselling was offered to all HIV-1 pregnant women regardless of the length of pregnancy at enrolment in the study. However, counselling usually started after

delivery, since most women came directly to delivery without a known HIV diagnosis. In addition, an 18 month follow up of their children was also offered, not only concerning the HIV status but also the general health situation of the child. They were also informed about the opportunity to try to avoid transmission of HIV-1 to the child by NVP treatment of themselves and their baby, by formula feeding, and the importance of following up the child for early diagnosis of the virus infection. If there was no time for a meeting before delivery, the pregnant women were treated and a meeting was arranged with the mother as soon as possible to inform her about the HIV-1 status.

The mothers were recommended not to start breastfeeding and were advised to bring up their children by formula feeding after ensuring that they could follow this recommendation and apply it safely. One more difficult aspect was for the mothers how to explain why they would not use breast-feeding without informing about their HIV status. To resolve this aspect they were provided with plausible explanations, which were not directly related to HIV/AIDS and could be offered to their relatives and friends.

Mother to child transmission of HIV-1 in northern Vietnam

The ARV was given free of charge to every HIV pregnant woman under the supervision of the National program for PMTCT. The following regimen was provided: if the HIV-1-infection was detected during delivery, the woman was provided with one dose of 200 mg NVP at delivery or at least two hours before caesarean section. If the HIV-1-infection was detected during gestation, the women were provided with AZT (azidothymidine) + 3TC (lamivudine) + nelfinavir (300 mg + 150 mg + 1250 mg twice per day, respectively). NVP was provided in case the maternal level of CD4+ T cells was less than 200 cells/ μ l, while nelfinavir was preferred if the level of CD4+ T cells was more than 200 cells/ μ l. Their children were given 2 mg of NVP liquid/kg weight within the first 48 hours of life. AZT liquid was administered for a week to those children whose mothers had received ARV before delivery.

In this study, 182 children were followed up to 18 months of life; for 135 children there were stringent criteria to identify their HIV-1 status and among them, 9 contracted infection from their mothers ($9/35 = 6.7\%$; see Figure 6). Seven infected children of 167 had a positive PCR test at birth, as evidence of intrauterine transmission in 4.2% of the children. Two children were negative at birth but positive at one month, suggesting that these 2 children were exposed to intra-partum transmission, which had then occurred in 1.5% of cases. There was no sign of late transmission through breast feeding, as all 135 children were tested 1-3 months after birth and later, with no new infection detected later than at 1-3 months. The frequency of intra-uterine and intra-partum transmission in our group of mothers can be compared with the frequency of 4.4% and 3.8% in Thailand in 2005 (Jourdain G *et al.*, 2007).

Two of the 15 children, who came for diagnosis at or after 6 months of age, were HIV-1-infected and the time point of infection could not be determined in these 2 children, since there was no birth sample. All included, 11/150 (7.3%) children became infected.

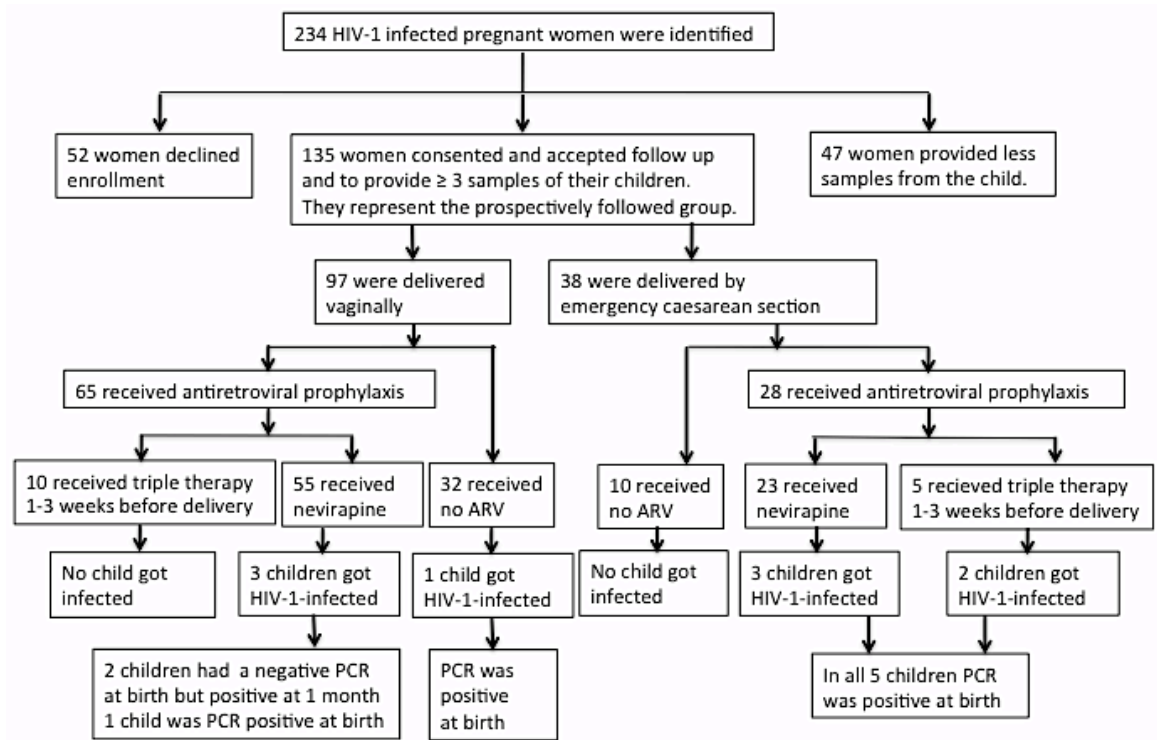


Figure 6. Flow chart of the outcome of maternal choice to participate in the study, antiretroviral prophylaxis and infection in the children

For the majority of the pregnant women the HIV-1 status became known late in pregnancy and therefore the full course of ART could not be provided. This problem was also reported in a previous study (Le CT *et al.*, 2008). ARV was given to 69% women.

Nearly 60% of the women in this study received one dose of NVP at delivery, which could not have affected *in utero* transmission. This probably explains the higher rate of intrauterine transmission in our study. Fifteen of the mothers (11%) got triple ARV treatment (AZT + 3TC + nelfinavir) for a few weeks before delivery and their children got more AZT liquid for one week after delivery. Two of these children were infected with HIV-1 at birth, which provides evidence of intrauterine infection. Since ARV was provided to the mothers only 4 weeks before delivery, these children could have been infected before ARV treatment was started. On the other hand, 13/15 of the mothers, who got triple ARV treatment, gave birth to a child who was not infected, demonstrating the usefulness of triple ARV treatment also in a shorter perspective.

Did the prevention strategy influence transmission?

In summary, our study indicates that the frequency of intrauterine MTCT was unaffected since most women started prophylaxis at delivery. In contrast, the frequency of delivery-associated and post-delivery MTCT seems to have been significantly reduced. In the absence of breastfeeding the reported accumulated transmission rate was estimated to be approximately 15% (ECS, 2001), which can be compared to 7% in our study, almost half; this figure can be interpreted as a clear indication of the reduction of the delivery-associated transmission.

According to recent WHO recommendations for PMTCT, triple combination therapy should be started as early as possible. Such programs are available in Vietnam, but only few HIV-1 pregnant women are treated, because the identification of their HIV-1 status occurs late in pregnancy. Here, our strategy of confidential counselling, formula feeding, and antiretroviral prophylaxis appears to have been effective. However, it is essential that the HIV status be defined early, in order for the pregnant women to fully benefit from the PMTCT programs.

Resistance to antiretroviral drugs in the maternal viruses

Since NVP is especially prone to induce mutations linked to resistance to anti-viral drugs (Grob PM *et al.*, 1992), it became important to analyse the frequency of these changes and to evaluate the possible interference with the effect of single-dose nevirapine to prevent mother-to-child transmission.

Mutations were detected in about 30% of the tested maternal virus *pol* sequences in positions commonly associated with RT resistance. The detected NVP-related NNRTI mutations were common polymorphisms that did not reduce susceptibility to the drug. Although we detected nucleoside RT inhibitor (NRTI) resistance mutations in 5 cases, none of these mutations was of clinical relevance; thus, we saw no pre-prophylaxis resistance to NVP. Our estimate is that the prevalence of circulating resistance is still low, supporting the value of the present PMTCT strategy. Lastly, resistance testing of HIV-1 infected women is not necessary unless they already have been exposed to therapy or single dose NVP before.

Paper II

The epidemic of HIV-1 CRF01_AE in Vietnam and neighbouring countries

The samples for this study were obtained from an important and sensitive group, from which it is difficult to receive samples and information. The study involved 37 HIV-1 infected pregnant women from northern Vietnam who were a subset of the patients in paper I. Characterization of HIV-1 in pregnant women is important to assist in the surveillance of mother-to-child transmission of HIV-1. In order to establish a material for comparison of the sequences derived from our study, a total of 464 V3 *env* HIV-1

sequences from Asia (Thailand, China, Vietnam, Japan, Indonesia, Singapore and Korea) were retrieved from the Los Alamos HIV database; these sequences were obtained until 2008. The sequences were analyzed and presented in a phylogenetic tree (See Figure 7).

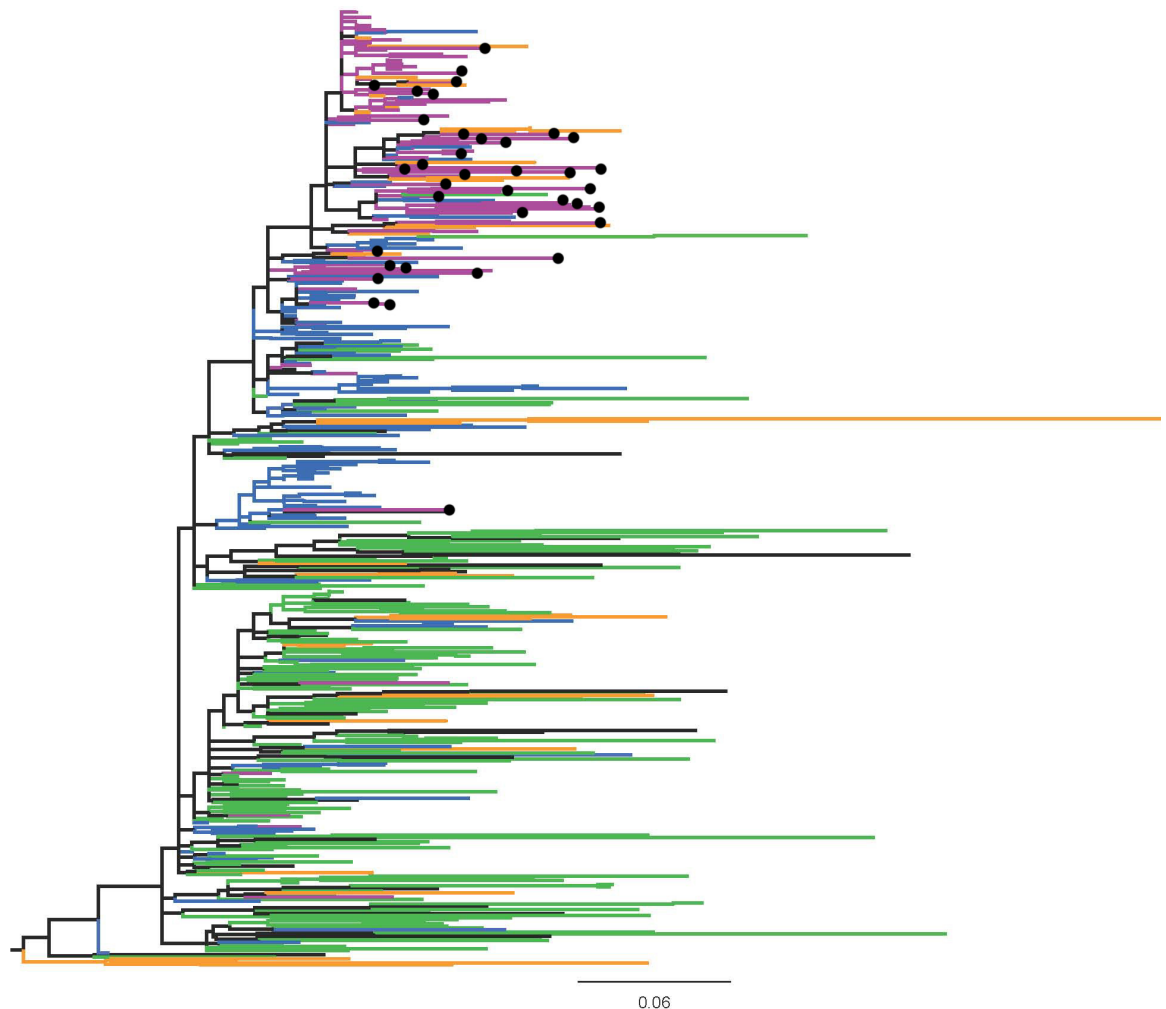


Figure 7. Phylogenetic tree of the CRF01_AE envV3 molecular epidemic in Vietnam and neighboring countries in Southeast Asia. The phylogenetic tree shows the sequences from Thailand in green, from China in orange, from the south of Vietnam in blue, and from the north of Vietnam in purple. Sequences from other countries are shown in black. The indicated scale bar is in units of substitutions/site. Filled black circles indicate the maternal sequences.

The upper part of the tree was dominated by sequences from northern Vietnam (blue for northern and purple for southern Vietnam), including sequences from the pregnant women (black circles) and sequences from southern China (orange). The bottom part was dominated by sequences from Thailand (in green), containing also sequences from China, Vietnam and from other countries (black). The visual impression of the maternal sequences obtained from our cohort was that they were scattered throughout the other sequences collected from southern Vietnam and southern China, apart from one sequence that appeared together with sequences from southern Vietnam. The bootstrap values were here too low to be a basis for statistical analysis and determine the exact branching pattern, indicating that the phylogenetic relations of the sequences were

complex and without clear phylogenetic paths. The bootstrap value support of the branching was not sufficiently high to allow the tree to be a basis for statistical analysis. However, the topology of the tree was similar to a recently reported phylogenetic analysis of the spread of HIV-1 CRF01_AE from Thailand to southern Vietnam and then to the north (Liao H *et al.*, 2009). According to the latter report, the HIV-1 epidemic in Vietnam seemed to arrive first from Thailand by heterosexual spread into the south of Vietnam, followed by the spread to intravenous drug users in the late 1980s and subsequently to intravenous drug users in the north during the mid-1990s. Interestingly enough, the sequences in our tree displayed a similar pattern.

It is important to note that here we are missing data from more neighbouring countries and areas such as, Burma, Laos and the Golden triangle, which have had major impact on the epidemic through drug trafficking. Thus, it would be interesting to retrieve newer submissions of more HIV-1 CRF01_AE reported sequences and map the virus evolution up to the current date.

Synonymous and non-synonymous analysis

Another way to study the relation to other groups of HIV-infected people is to describe the synonymous (ds) and non-synonymous (dn) amino acid changes, where the synonymous changes represent a background of mutations and non-synonymous increase in relation to different pressures. Studying the individual nucleotide changes intrapatiently, showed that the maternal sequences had significantly larger numbers of ds changes than those of the intravenous (IV) drug users in Hanoi in 2002 (closely related in time and area) and other sequences from northern Vietnam, respectively ($p < 0.001$). Altogether, the comparisons of the respective dn and ds average means may reflect the time-related evolution of the maternal sequences versus the older ones from Vietnam.

In addition we compared the maternal sequences to geographically more distant groups such as IV drug users from Hanoi and other sequences from northern Vietnam, southern Vietnam and Thailand. The sequences from Thailand were significantly more evolved than the maternal sequences and all other groups with regard to both ds and dn changes ($p < 0.001$). This is probably a sign of the longstanding epidemic in Thailand, broader range of sources and a wider range in time of collection.

The maternal sequences had less ds, but more dn changes than the group from southern Vietnam. The analysis also indicated that they had evolved when compared to the older sequences from northern Vietnam. It was not possible to discern whether HIV-1 in the pregnant women has evolved as a result of spreading into the general population, or is still closely related to the epidemic among intravenous drug users, a group that dominates as a source of HIV sequences in this part of the country.

Paper III

The CRF01_AE subtype appears to be associated with a higher proportion of CXCR4-using viruses (Yu XF *et al.*, 1995, Pramanik L *et al.*, 2011), but this is not widely recognized. Using two bioinformatic models described above, we frequently detected the CXCR4-using phenotype in CRF01_AE HIV-1 from pregnant women, while in the population sequences of unrelated infants at birth we detected only the R5 phenotype.

HIV-1 co-receptor use predictions for CRF01_AE infected pregnant women

A greater proportion of CXCR4-using strains was found in mothers of infected children by the glycan-charge model when compared to mothers of uninfected children. This could also be linked to disease progression and the high viral loads in the mothers of infected children, consistently with previous work showing a lower CD4 +T cell count in mothers transmitting HIV-1 (Ha TT *et al.*, 2012). The relative proportion of X4 to R5 was similar to recent findings (Pramanik L *et al.*, 2011).

In order to increase the resolution of the relative proportion of putative R5 and X4 strains we examined single genome sequences from PBMC samples. The main outcome was that sequences representing CXCR4-using strains were significantly more frequent in mothers of infected children using the glycan-charge model. (See Figure 8). A similar comparison, using Geno2Pheno, was not statistically significant but visualized that the mixed population sequences included all phenotypes. In about 93% of the cases, the single genome sequences from each woman reflected a similar or identical composition of the phenotypes, as for the population sequences obtained with Geno2Pheno and glycan-charge model.

In 4 women the viral phenotypes were classified as defective or non-infectious. We could only retrieve a few single genome sequences from these mothers, in accordance with the finding that they had low HIV-1 DNA levels. This supports the interpretation of the model that, if the sequences lack characteristics that match with either CCR5 or CXCR4 use, the corresponding viruses would be defective/noninfectious. Population sequences with the sequence pattern resembling defective/noninfectious virus were found in mothers of uninfected children, but not in mothers of infected children. A lower proportion of R5 viruses, either due to a higher proportion of X4 viruses or to the presence of more defective/noninfectious strains, could perhaps be associated with lower levels of transmission.

Figure 8 represents the phylogenetic tree of the 200 maternal single genome sequences. For the majority of the mothers, their single genome sequences clustered into distinct groups, as supported by bootstrap values.

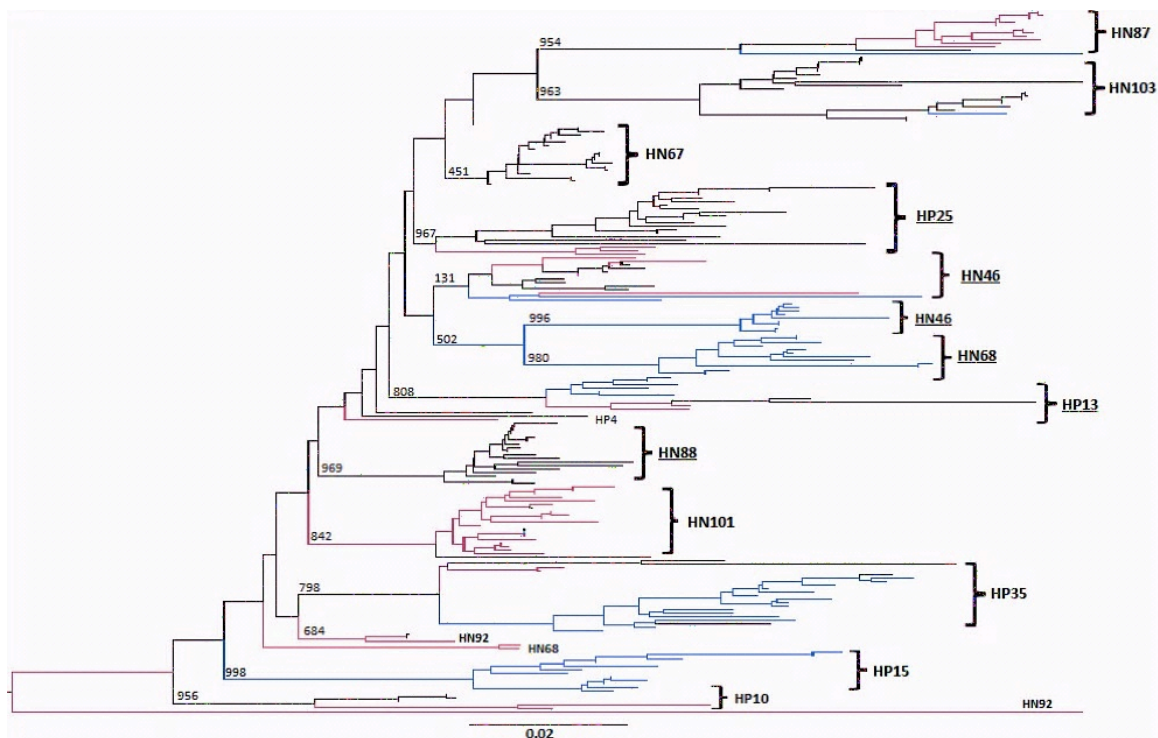


Figure 8. Phylogenetic tree of the women's *env* single genome sequences. The putative R5 sequences are displayed in red, putative X4 sequences in blue, and putative mixed sequences in black color. The scale bar indicates the degree of divergence in the branch lengths, equivalent to 0.02 substitutions per site.

Predicted phenotype characteristics of the infant HIV-1 *env* sequences

In samples from 13 children, unrelated to the women's group, we found only R5 sequences in the population sequences at birth. This complies with other studies showing that R5 is found early in the infected child following HIV-1 mother-to-child transmission (Scarlati G *et al.*, 1997; Casper C *et al.*, 2002a; Clevestig P *et al.*, 2005). The presence of only one putative mixed strain at birth in all infant sequences suggests that R5 strains are favored during transmission or possibly in the outgrowth. However we also observed many sequences of a defective/non-infectious nature. The possible explanations for these findings could be that there is a high degree of lymphocyte proliferation early after birth and that the replication of lymphocytes containing the integrated HIV-1 DNA of a defective virus, would lead to amplification of the defective virus as well. Furthermore, complementation could influence replication of non-infectious virus by the simultaneous presence of non-infectious and infectious virus (Salzwedel & Berger, 2009).

The gp120 V3 N-linked glycosylation motif NNT was present in almost all sequences. The charge appeared to shift in the six months period after birth. Thus, while all population sequences had the charge +3 up to 3 months, the charge increased to +4 in the two population sequences from 6 months ($p=0.018$, Fisher's exact test). Altogether this suggests a tendency of a rapid induction of genetic

changes, or selection of new variants from a minor population of HIV-1 in these infants.

Geno2Pheno vs glycan-charge model?

Comparisons of the two in silico methods

The agreement between these two different *in silico* methods was surprisingly high, except for the sequences that were designated as mixed or defective by the glycan-charge model. The Geno2pheno method also predicted a relatively high proportion of the CXCR4-using phenotype.

It was not clear how to compare the outcomes of the two methods used in our study (Table 3). For all population sequences there was a complete agreement regarding the monotropic R5 and X4 between the Geno2Pheno versus the glycan-charge model.

In Table 3b the outcome of the two methods in single genome sequences is compared. If only the single genome sequences, classified by the glycan-charge model as either R5 or X4, were included the agreement between the two methods was 87 %.

Table 3. Comparison of the HIV-1 phenotype predictions made by the glycan-charge model and Geno2Pheno

a) HIV-1 Population Sequences of women and children									
Glycan – Charge model									
Geno2Pheno		R5		X4		Mixed		Defective	
	R5	31		0		12		6	
	X4	0		7		4		1	
	Total	31		7		16		7	
b) HIV-1 Single genome sequences of women (W) and children (I)									
Glycan – Charge model									
Geno2Pheno		R5		X4		Mixed		Defective	
		W	I	W	I	W	I	W	I
	R5	44	10	2	0	35	10	4	23
	X4	3	0	57	0	53	0	2	1
	Total	47	10	59	0	88	10	6	24

It appears as if Geno2Pheno interprets the mixed and defective sequences in the children as predominantly R5, while the corresponding single genome sequences of the women were interpreted as predominantly X4. The discrepancy may be related to that the glycan-charge model is a more analytical method and simply based on the sequence characteristics, it is comprehensive and could be suitable for use in routine clinical practice.

On the other hand, phenotype predictions using the Geno2Pheno were previously found to have relatively poor sensitivity, but high specificity for predicting non-R5 virus, as compared to biological testing (Harrigan PR, 2011). Geno2pheno relies on a number of algorithms based on bioinformatic theory, which have become increasingly sophisticated over time. Such recent additions allow clinical data to be included in the model, which are considered to improve the predictions.

Furthermore, Geno2Pheno provides a reliable discrimination between R5 and X4 sequences when the false-positive rate (FPR), is set between 5 and 10%. Triplicate replicas are recommended for the testing, which means that samples have to undergo three separate PCR amplifications followed by separate sequencing of the three PCR products (Swenson L *et al.*, 2010). Therefore, three separate results have to be obtained for each sample, and if any sequence is identified as X4, presence of CXCR4-using variants is reported. In contrast, a single V3 sequence is enough for the glycan-charge model predictions. However, further improvements are also required for this model, in order to discriminate between the undetermined groups of mixed and defective viruses. An analysis is needed of sequences of true dual-tropic isolates, so that the presence of a mixture of R5 and X4 virus strains can be eliminated. This would probably increase the predictions of most *in silico* methods.

Study IV (preliminary results)

Mother-child pairs infected with different HIV-1 subtypes

The basis for the maternal-infant transmission remains poorly understood. It has previously been associated with specific viral selection (Wolinsky SM *et al.*, 1992), neutralization resistance (Scarlatti G *et al.*, 1993; Kliks SC *et al.*, 1994; Dickover R *et al.*, 2006; Wu X *et al.*, 2006; Zhang H *et al.*, 2009) and enhanced replicative capacity (Kong X *et al.*, 2008) of the transmitted viruses. Moreover, other factors such as selection for reduced glycosylation of gp120 and shorter *env* variable regions have also been proposed (Wu X *et al.*, 2006). Three different patterns of vertical transmission of HIV-1 have been suggested: transmission of the dominating maternal strain, of a minor strain and of multiple strains (Nowak P *et al.*, 2002). Often, the transmitted variant represented a minor population of the maternal strains, suggesting that selection plays a role during this process (Briant L *et al.*, 1995; Pasquier C *et al.*, 1998; Wade CM *et al.*, 1998).

In the present study, we have generated and analysed over 700 single genome sequences spanning the V3 and flanking regions from 8 HIV-1 infected mother-child pairs to trace the origin of the maternal virus transmitted to the child. Three children were classified as contracting the infection in utero and five during delivery. The majority of the women (5 pairs) was from Africa, carried subtypes A and C and was naïve to treatment. There were also three pairs from Asia, carrying the circulating recombinant form CRF01_AE, one of which was treated with NVP at delivery (M1-

C1_AE). Twenty-eight isolates from PBMCs and 13 from plasma were available for this study (Table 4). We have included several samples from each pair; collected during and after pregnancy, including the very first positive samples from the children and, in some cases, follow up samples from up to 4 years of life. Thereby, we could obtain a relatively large population of single genome sequences, which were used for comparison of the amino acid composition of the V3 region and for the phylogenetic tree analysis.

Table 4. Overview of the cases included in the Study IV

	Case ID	Time of transmission	No. Isolates		No. SGS	
			PBMC	Plasma	PBMC	Plasma
Subtype A	M1-C1_A	In utero	8	4	36	9
	M2-C2_A	Delivery	6	5	101	0
Subtype C	M1-C1_C	Delivery	4	2	83	17
	M2-C2_C	Delivery	2	0	114	20
	M3-C3_C	In utero	3	1	85	41
Subtype CRF01_AE	M1-C1_AE	In utero	0	0	0	14
	M2-C2_AE	Delivery	0	0	55	0
	M3-C3_AE	Delivery	5	1	105	50
Total	8		28	13	579	151

Besides the single genome sequences (SGS) generated in this study, we also used sequences from PBMC and plasma isolates, which have been previously characterized biologically as being of the R5 phenotype (Casper C *et al.*, 2002a, 2002b). In addition to this, we included some additional sequences from Contag CH *et al.*, 1997 for case M1-C1_A and from Clevestig P *et al.*, 2005 for case M3-C3_AE. An advantage with the isolate sequences is that they represent infectious virus and the disadvantage is that they may have undergone selection during the primary isolation step and also mutated *in vitro* during passage. On the other hand, single genome sequences represent a more detailed picture of the viral population in the patient, although there is no certainty of whether the represented viral strains are viable and infectious.

Characteristics of the transmitted viruses

Phylogenetic analysis of the *env* sequences revealed that all eight mother-child pairs were epidemiologically linked subtype wise, the pairs are distinctly separated and that within each pair there was a clear link between the maternal and infant sequences (see Figure 9). The transmitted virus could be identified by the combined use of the phylogenetic tree and by comparing the amino acid pattern of the V3 sequence. More detailed scrutiny of the phylogenetic trees displayed a pattern suggesting a possible single variant transmission for 5 of the 8 pairs. It was remarkable than in many cases, the children showed a more diverse and hence evolved HIV-1 population than their

chronically infected mothers. This could also be due to the fact that the children were followed up for longer and more samples from different time points were available.

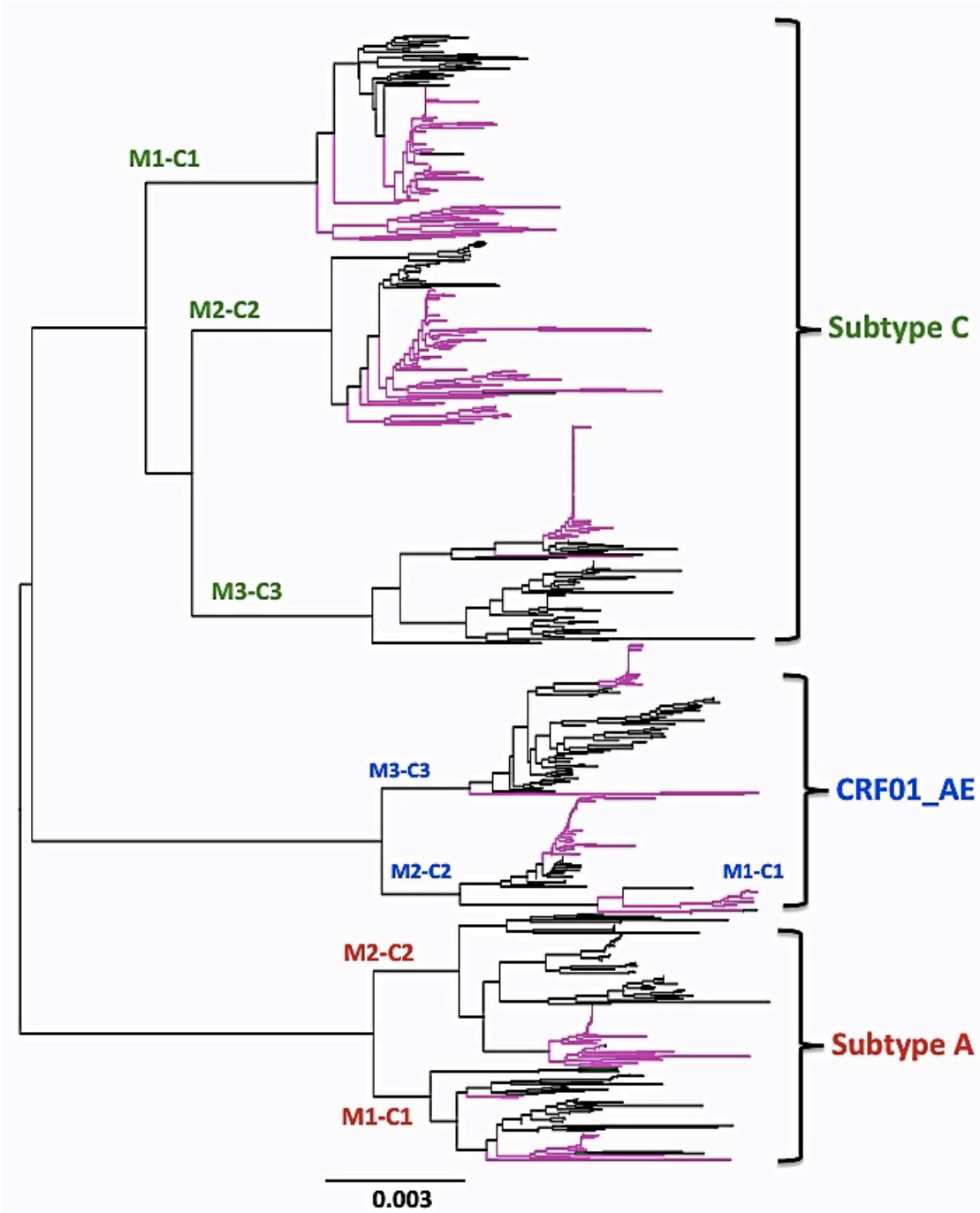


Figure 9. Phylogenetic tree displaying the layout of the HIV-1 *env* V3 sequences and flanking regions for the 8 mother-child pairs. The sequences cluster separately into their different subtype groups and group accordingly to the respective pair. Infant sequences are displayed in purple and maternal sequences in black. The brackets indicate the subtype; the number after the letters M and C indicates the mother-child pair. The scale bar indicates the degree of divergence in the branch lengths, equivalent to 0.03 substitutions per site.

Overall, we identified 11 likely transmitted viral sequences in 7/8 mother-child pairs, as matching sequences, including identical V3 amino acid composition in both maternal and infant sequences; these sequences are presented in Table 5 A-C. A consensus

sequences was chosen for each subtype from the HIV Los Alamos Database and was used for the comparisons. No reduction in the V3 region size was observed and the V3 glycosylation site was present in all of the sequences as NNT (Asn-Asn-Thr). The phenotype determinations included in the tables were based on the outcomes from the glycan-charge *in silico* model. There was a complete agreement between the glycan-charge model predictions and the biologically obtained phenotypes, regarding the sequences from the isolates, which were the only sequences that had a known biological phenotype in this study. A transmission selection of few viruses was observed among all the pairs and the transmitted viruses showed a consistent CCR5 dependence. This is in agreement with the majority of the reported MTCT cases (Scarlati G *et al.*, 1993; Ahmad N *et al.*, 1995; Dickover RE *et al.*, 2001).

Subtype A mother-child pairs

Only one viral sequence was detected as being transmitted in each mother/child pair for subtype A. Both V3 sequences had a charge of +3 and the same pattern was seen *in utero* and at delivery transmission. For the mother-child pair 1 the matching V3 sequence was present in the child in a birth isolate and in the mother in 8 SGS from the 2nd trimester and 6 months after delivery. However, for the mother-child pair 2 the transmitted virus seemed to be present in all samples from 3, 6 and 24 months, as well as in plasma and PBMC isolates from the child, but only at 3 months after delivery in the mother. These sequences were also found to cluster together in the phylogenetic tree.

Table 5A. Transmitted HIV-1 subtype A sequences in mother-child pairs*

Subtype A	CTRPNNNTRKSVRIGPGQAFYATGDIIGDIRQAH	Charge	NNT	Phenotype
M1.2tr.D(8)Y.	3	+	R5
M1.6mpp.PiY.	3	+	R5
C1.b.DiY.	3	+	R5
M2.3mpp.PiL.....Y.	3	+	R5
M2.3mpp.DiL.....Y.	3	+	R5
C2.3m.PiL.....Y.	3	+	R5
C2.3m.D(22)L.....Y.	3	+	R5
C2.6m.DiL.....Y.	3	+	R5
C2.24m.D(3)L.....Y.	3	+	R5

*In the sequence designation the first letter stands for maternal (M) or child (C) sequence, the second character indicates the pair number. The next number indicates time collection of the sample: tr stands for trimester, mpp stand for months post-partum, p stands for partum, b for birth and m for months. The next letter indicates if it was a PBMC (D) or plasma (P) sample and i stands for isolate. The numbers in brackets indicate the total number of single genome sequences that had the same V3 sequence.

Table 5B. Transmitted HIV-1 subtype C V3 sequences in mother child-pairs*

Subtype C	CTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQAHC	Charge	NNT	Phenotype
M1.3tr.P1V.....	3.1	+	R5
M1.3tr.D1V.....	3.1	+	R5
M1.p.D(5)V.....	3.1	+	R5
M1.6mp.D(7)V.....	3.1	+	R5
M1.11mp.DiV.....	3.1	+	R5
C1.7w.D(18)V.....	3.1	+	R5
C1.6m.DiV.....	3.1	+	R5
C1.6m.PiV.....	3.1	+	R5
C1.12m.PiV.....	3.1	+	R5
C1.12m.DiV.....	3.1	+	R5
C1.12m.P(9)V.....	3.1	+	R5
C1.12m.D(9)V.....	3.1	+	R5
C1.36m.D(7)V.....	3.1	+	R5
C1.24m.DiV.....	3.1	+	R5
M1.6mp.D2M.....	3.1	+	R5
C1.12m.D5M.....	3.1	+	R5
C1.12m.P(6)M.....	3.1	+	R5
M2.p.D(4)E.....	1.1	+	R5
C2.30m.D15E.....	1.1	+	R5
M2.p.D(4)	.A.....E...L.....A.....	2.1	+	R5
C2.5m.D3	.A.....E...L.....A.....	2.1	+	R5
M2.p.D(11)E.....A.....	2.1	+	R5
C2.5m.D(38)E.....A.....	2.1	+	R5
M3.3tr.D5M.....E...N...Y.	4	+	R5
M3.p.PiM.....E...N...Y.	4	+	R5
M3.p.P(6)M.....E...N...Y.	4	+	R5
M3.p.D(4)M.....E...N...Y.	4	+	R5
C3.b.DiM.....E...N...Y.	4	+	R5
C3.b.D(19)M.....E...N...Y.	4	+	R5
C3.1m.DiM.....E...N...Y.	4	+	R5
C3.1m.D(10)M.....E...N...Y.	4	+	R5
C3.1m.P(30)M.....E...N...Y.	4	+	R5

*The first letter stands for maternal (M) or child (C) sequence and the second character indicates the pair number. The next number indicates time collection of the sample: tr stands for trimester, mp stand for months post-partum, p stands for partum, b for birth and m for months. The next letter indicates if it was a PBMC (D) or plasma (P) sample and i stands for isolate. The numbers in brackets indicate the total number of single genome sequences that had the same V3 sequence.

Subtype C mother-child pairs

Interestingly, there was a greater divergence within the sequences obtained from the subtype C group. This could also be due to the fact that a larger number of SGS was obtained for subtype C. There were 1-3 viral sequences detected as transmitted in each case and the V3 charge range was between +1.1 and +4. For mother-child pair 1, the

transmitted virus seemed to be present in all infant samples from 7 weeks up to 36 months of age. Furthermore, in mother-child pair 2, three different viral strains from the mother at delivery appeared to be present in the child at 5 and 30 months of age. For the third mother-child pair of this subtype, we had samples available from birth and up to one month of life, which displayed a very homogenous viral population. There were many maternal SGS intermingling within the child's cluster, containing the same exact sequence.

Subtype CRF01_AE mother-child pairs

From the mother-child pairs infected with the subtype CRF01_AE we only obtained sequence matches for the delivery-associated transmissions. For these mother-child pairs, the SGS from the infants with an identical V3 sequences as the maternal SGS and isolates, did not group together in the phylogenetic tree. This suggests that the flanking regions had further evolved in the infants, or maybe there were multiple variants transmitted containing the same V3 sequence.

Moreover for mother child-pair 3, the maternal sequences were derived from three time points and grouped in different clusters. There was no clear separation between plasma and PBMC derived sequences. The sequences from the child were homogeneously distributed in an early time point. When using the CRF01_AE cut offs for the glycan-charge model in this case, the putative phenotypes of matched sequences were determined as mixed, while Geno2Pheno characterized them as X4.

Table 5C. Transmitted HIV-1 CRF01_AE V3 sequences in mother child-pairs*

CRF01_AE	CTRPSNNTRTSITIGPGQVFYRTGDIIGDIRKAYC	Charge	NNT	Phenotype
M2.p.D(13)T.....	3	+	R5
C2.6m.D(34)T.....	3	+	R5
M3.p.P(3)Y.....VR.....A..K.....R...	4	+	mixed
C3.12m.DiY.....VR.....A..K.....R...	4	+	mixed
M3.2tr.1PY.....VR.....A..K....V....R...	4	+	mixed
M3.p.D(5)Y.....VR.....A..K....V....R...	4	+	mixed
M3.6mpp.D(3)Y.....VR.....A..K....V....R...	4	+	mixed
M3.6mpp.5PY.....VR.....A..K....V....R...	4	+	mixed
C3.6m.D(25)Y.....VR.....A..K....V....R...	4	+	mixed

*The first letter stands for maternal (M) or child (C) sequence and the second character indicates the pair number. The next number indicates time collection of the sample: tr stands for trimester, mpp stand for months post-partum, p stands for partum, b for birth and m for months. The next letter indicates if it was a PBMC (D) or plasma (P) sample and i stands for isolate. The numbers in brackets indicate the total number of single genome sequences that had the same V3 sequence.

Summary of Study IV

In summary, preliminary results obtained in Study IV suggest that the transmitted HIV-1 variants in MTCT represent a transmission selection of few viruses. No reduction in the length of the variable region size, no loss of the V3 glycosylation site, and CCR5 dependence was observed among all the pairs.

In this study it was interesting that many of the isolate sequences appeared to be stable over time. A possible explanation could be that there is a core of infectious virus *in vivo*, transmissible and stable over time. The fact that most of our isolate sequences were obtained from PBMCs from the very primary isolates means that there has been no extra passage before the sequencing PCR was conducted. This may have prevented the expected mutations induced *in vitro*. However, further investigations are needed to complete the analysis of the large material included.

In addition, the transmitted viral sequences found in the children were detected in different time points from 3-36 months, in different subtypes. It is important to note that, the spread of sequences within one time point may illustrate the great variety of HIV-1 sequence populations present at any given time. When mutations occur these may affect only a portion of the population and in the relatively short time span of these cases, the diversity may cause a lack of time linearity.

One of the limitations of this study was that only the V3 and flanking regions were sequenced. Since we have in our hands isolates as well as proviral/cDNA and not only these sequences, these viruses could be used for a more detailed genetic analysis of other parts of the genome. In addition, it would be interesting to test the neutralization sensitivity of these isolates against homologous and heterologous sera, and broadly neutralizing antibodies for the purpose of investigating potential selection forces during mother-child-transmission.

Conclusions & Perspectives

This thesis contributes to outline the picture of the HIV-1 epidemic in northern Vietnam in respect to viral evolution and outcomes of PMTCT strategies, as well as further increases the understanding of the genotypic prediction of viral phenotypes involved in mother-to-child transmission.

Programs for PMTCT have been improved in Vietnam and other developing countries, but the national PMTCT programs need to be further developed and shaped to fit the specific culture of a country. ART is of great benefit for HIV-1 infected patients but it is also accompanied by certain side effects. The identification of the HIV-1 status late during pregnancy makes it hard to provide ART to the infected mother in time. Extensive use of NVP at delivery has been efficient and is still safe to be used, as there has been no major resistance to therapy detected in the viral sequences. However, this drug is out-dated and not anymore included in the general recommendations, due to its ability to easily select for resistance mutations. Moreover, the counselling strategy should be further implemented and it could possibly be assisted by social workers or other experienced HIV patients, so that the doctors could be become available for other type of assistance. Financial support for formula feeding can increase significantly the effectiveness of PMTCT and should be added to PMTCT programs, while the impact of elective caesarean section in Vietnam should be clarified in future studies.

The dominating HIV-1 subtype in northern Vietnam is CRF01_AE and this feature of the epidemic in Vietnam has not been changing significantly over time. In order to get the full picture of this rapidly spreading subtype, it would be important to extend this study to additional countries and to other HIV-1 risk groups. Moreover, including extensive parts of the viral genome would make it possible to detect potential new recombinants.

One more issue raised in this thesis is the co-receptor profiles of CRF01_AE viruses from the group of pregnant women. Basic knowledge of the HIV-1 evolution of co-receptor use has become increasingly important due to the recent introduction of CCR5 antagonists as part of antiretroviral therapy. Since these drugs have no effect on the virus X4 populations, there is an increased risk that the HIV-1 populations may shift from CCR5 to CXCR4-use upon treatment with CCR5 antagonists, which could lead to treatment failure and faster disease progression. Since the R5 phenotype appears to be critical in HIV-1 mother-to-child transmission, targeting the R5 phenotype with an increase armamentarium of drugs seems pivotal.

Today, subtype-specific differences regarding how frequently CXCR4-using populations appear in late-stage disease have been noted. However, there is no clear evidence that the V3 region has the same impact on co-receptor interaction among different subtypes. In the first sample from the infected infants we detected no CXCR4-

using strains, suggesting selective transmission of R5 viruses when the mothers carried CXCR4-using strains. Indeed, we found a generally high proportion of viral sequences with predicted CXCR4-use from the pregnant women, which may reflect more advanced disease conditions or an intrinsic feature of viruses of the CRF01_AE subtype. Subtyping of the infected mothers might not yet be necessary for this particular risk group, since it probably does not affect prophylaxis or other clinical decisions. However, understanding HIV-1 subtype-specific differences regarding the ability to develop CXCR4-using populations may be of great importance for defining differences in HIV-1 pathogenesis and for future treatment guidelines using co-receptor antagonists.

The field is becoming increasingly more confident in predicting the viral phenotype, i.e. co-receptor use, with genotypic methods, but there is still need for improvements. Looking to the future, additional genotypic methods such as ultra-deep sequencing may be developed for improved sensitivity, which may help to further define the clinically relevant characteristics of CXCR4-using variants and improve our understanding of X4 evolution. Moreover, the possibility exists that alterations of Env functions could provide a sufficiently selective advantage during the HIV-1 transmission events and result in preferential transmission of viruses with specific properties, as we identified the selective transmission of a minor population of viruses with R5-like properties, which seemed to be stable over time. The growing application of SGS technology coupled with the sophisticated cell-to-cell and *ex vivo* tissue systems may provide means for identifying immunologic and phenotypic traits associated with HIV-1 transmission.

Aiming towards a generation free of AIDS and HIV, we still need to obtain substantial knowledge that may underpin the design of improved vaccine and treatment approaches.

ACKNOWLEDGEMENTS

The past four years and the efforts put into this thesis have been somehow agonizing. Many people have played a crucial role into making this work happen and I would like to express my most sincere gratitude to the following:

Anneka Ehrnst, my supervisor for her great courage to take me in her group despite not knowing me at all. Anneka was a great mentor and did everything to help me develop as an independent scientist and to think critically. I was very lucky to have her as my supervisor, not only because of her excellent scientific guidance, but also for showing me how to always fight towards your goals. I hope I made you proud!

Ingemar Ernberg, my official supervisor, who took me under his wing in a very vulnerable time and lead the end of my PhD. Thank you for your support and encouragement. Thank you for all the interesting morning meeting discussions and for being such an inspiring leader. Your group has been a real family to us.

Francesca Chiodi, words cannot express how grateful I am for everything you have done for me. Thank you for being such a wonderful support, for your patience and for believing in me.

Marianne Jansson, my third saving angel! Thank you for your great scientific input and for all your support on these difficult times.

Gunilla Karlsson-Hedestam, for your help and support over the years and always being there for us. You are truly an inspiration! **Martin Rottenberg**, for being a fantastic mentor and always cheering for me. **Jan Albert**, for your great scientific advice.

My fantastic group members! **Sven Grutzmeier**, thank you for kindness, hospitality and your infinite knowledge. **Lotta Pramanik**, my first and real friend in Sweden. I have been more than lucky to have you next to me on this journey. Thank you for all your help and for always being there whenever I needed you. I will always need you!

This thesis would have never been official if it was not for the great support of the MTC administrators. Thank you **Marie Arsenian-Henriksson**, **Helene Stambeck** and **Marlene Brink-Sinervo** for making those magic numbers appear in our account. Thank you **Gesan Arulampalam**, for being so understanding, supportive and helped out in the most critical moments.

My favorite MTC people, who made work a better place. Thank you: **Mariam**, **Frank**, **Berit**, **Bence**, **Marc**, **Pegah**, **Marijke**, **Arnnika**, **Rossana**, **Sammer**, **Jess**, **Marilena**, **Speranta** and **Suman**. Extra thanks to my super sweet **Chaniya** and **Qin Li**, who never say no to anything.

To my amazing partners in crime at the legendary MTC Pubs! Thank you **Simone**, **Antonio**, **Cyprian** and **Mario** for all those unforgettable nights. Special thanks to my

favorite Polish persons: **Lech Ignatowicz**, you have always had my back and made sure that I would land smoothly. Life has some extra fun around you! **Jola Mazurek**, thank you for your support throughout these years and for being such a good listener. I love you both!

The lovely super-star **Maria-Lisa Knudsen**. I have been very lucky to have you in my life. Thank you for all the fun times, your advice and your support on both life and science.

There have been a great number of people who helped me tremendously with everything, outside the Karolinska world.

What would I do without my crazy Greeks?! Thank you **Souzana** and **Vasw**, for bringing joy and fun in my life. Thank you for all you have done and for trying to keep me connected to the roots.

The super sweet and fun Swedish gang, who welcomed me into their lives and added so much more flavor into mine. Thank you **Alex, Anna E, Christoffer, Elin, Andreas, Camilla, Philip, Sarah, Christian** and **Anne**. The amazing **JohAnna**, for being such good and caring friends; thank you **Johan** and **Anna**!

The best friend in the world, **Laura Levi**, who always has something nice to say and so much love to give. Thank you **Claudette, Alessandra**, and **Joao** for being yourselves and for being around. You guys bring joy in my life!

Special thanks go to:

My Swedish family, who literally adopted me these last years and gave me more than just affection. Thank you **Kurt** for taking care of me, **Monica** and **Thomas** for your incredible kindness, **Fredrik** and **Therese** for being so lovely and supportive.

My family! My lovely parents **Anna** and **Gavrill** and my sister **Despoina**, who would cheer and support me in any moment. Their love and encouragement have been a strong driving force to anything I have done in my life. I promise to become better at coming home more often!

Patrik, the most amazing person who deals with all sort of Rozina-problems and can still love me everyday! My life has had such a wonderful turnaround since I met you. I could not thank you enough for everything you do for me. This is just one milestone; we have many more to achieve together. I love you more than anything!!!

REFERENCES

- Aiken C, et al., Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell* 1994;76:853-64.
- Ahmad N. The vertical transmission of human immunodeficiency virus type 1: molecular and biological properties of the virus. *Crit Rev Clin Lab Sci*, 2005. 42(1): p. 1-34.
- Alaeus A. Significance of HIV-1 genetic subtypes. *Scand J Infect Dis*. 2000;32:455-63.
- Albert J, et al., Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 pol and gag genes. *J Virol*. 1994;68:5918-24.
- Ahmad N., Molecular mechanisms of HIV-1 vertical transmission and pathogenesis in infants. *Adv Pharmacol*, 2008. 56: p. 453-508.
- Alaeus AL, et al., Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A-D. *AIDS*, 1999. 13(8): p. 901-7.
- Alimonti JB, et al., Mechanisms of CD4+ T lymphocyte cell death in human immunodeficiency virus infection and AIDS. *The Journal of general virology* 2003;84:1649-1661.
- Arroyo MA, et al., Virologic risk factors for vertical transmission of HIV type 1 in Puerto Rico. 2002, *AIDS Res Hum Retroviruses* 18, 447-460.
- Barre-Sinoussi F et al., Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, 1983. 220(4599): p. 868-71.
- Bartolo I, et al., Highly divergent subtypes and new recombinant forms prevail in the HIV/AIDS epidemic in Angola: New insights into the origins of the AIDS pandemic. *Infect Genet Evol*, 2008.
- Berger EA, et al., A new classification for HIV-1. *Nature*, 1998. 391(6664): p. 240.
- Berry IM. Genetic aspects of HIV-1 evolution and transmission, in *Microbiology, Tumor and Cell Biology (MTC)*. 2008, Karolinska Institutet: Stockholm. p. 48.
- Björndal A, et al., Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype. *J Virol*. 1997;71:7478- 87.
- Björndal A, et al., Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients. *AIDS Res Hum Retroviruses* 1999;15:647-53.
- Bleul CC, et al., The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature* 1996;382:829-33.
- Brenchley JM, et al., CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *The Journal of experimental medicine* 2004;200:749- 759.
- Briant L, et al., Analysis of envelope sequence variants suggests multiple mechanisms of mother to child transmission of HIV-1. *J Virol* 1995; TW: 3778-3788.
- Briggs DR, et al., Envelope V3 amino acid sequence predicts HIV-1 phenotype (co-receptor usage and tropism for macrophages). *AIDS* 2000;14:2937-9.
- Brown PO. Integration. In J. M. Coffin et al. (ed.), *Retroviruses*. 1997, Cold Spring

Harbor Laboratory Press, Plainview, N.Y.

Brun-Vezinet F., Damond F and Simon F, Variability of human immunodeficiency virus type 1. *Bull Soc Pathol Exot*, 1999. 92(4): p. 261-3.

Bryson YJ, et al., Proposed definitions for in utero versus intrapartum transmission of HIV-1. *N Engl J Med*, 1992. 327(17): p. 1246-7.

Cao J, et al., Replication and neutralization of human immunodeficiency virus type 1 lacking the V1 and V2 variable loops of the gp120 envelope glycoprotein. *J Virol*. 1997;71:9808-12.

Casper C, et al., Coreceptor change appears after immune deficiency is established in children infected with different HIV-1 subtypes. *AIDS Res Hum Retroviruses*, 2002a, 18;343-352.

Casper C, et al., 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*, 2002b, 186;914-921.

Chalmet K, et al., Presence of CXCR4-using HIV-1 in patients with recently diagnosed infection: correlates and evidence for transmission. *J Infect Dis*, 2012. 205(2): p. 174-84.

Chan DC, et al., Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 1997;89:263-73.

Chermann JC, et al., Isolation of a new retrovirus in a patient at risk for acquired immunodeficiency syndrome. *Antibiot Chemother*, 1983. 32: p. 48- 53.

Church JD, et al., HIV-1 tropism and survival in vertically infected Ugandan infants. *J Infect Dis*, 2008. 197(10): p. 1382-8.

Clevestig P, et al., The X4 phenotype of HIV type 1 evolves from R5 in two children of mothers, carrying X4, and is not linked to transmission. 2005, *AIDS Res Hum Retroviruses* 21, 371-378.

Clevestig P, et al., CCR5 use by human immunodeficiency virus type 1 is associated closely with the gp120 V3 loop N-linked glycosylation site. 2006 *J Gen Virol* 87, 607-612.

Clevestig P. (2006). In *HIV-1 phenotype and genetic variation in vertical transmission*. Stockholm, Karolinska University Press, ISBN 19-1740-631-X, p. 29.

Cocchi F, et al., Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science* 1995;270:1811-5.

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

Coffin J, et al., Human immunodeficiency viruses. *Science*, 1986. 232(4751): p. 697.

Contag CH, et al., Mother-to-Infant Transmission of Human Immunodeficiency Virus Type 1 Involving Five Envelope Sequence Subtypes. *J. Virol*. 1996;71:1292-1300

Coutsoudis A, et al., Influence of infant-feeding patterns on early mother-to- child transmission of HIV-1 in Durban, South Africa: a prospective cohort study. *South African Vitamin A Study Group. Lancet*, 1999. 354(9177): p. 471-6.

Dalglish AG, et al., The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* 1984;312:763-7.

De Cock KM, et al., Prevention of mother-to-child HIV transmission in resource-poor countries: translating research into policy and practice. *Jama*, 2000. 283(9): p. 1175-82.

Deeks SG, Walker BD. Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity* 2007;27:406-416.

de Mendoza C, et al., Performance of a population-based HIV-1 tropism phenotypic assay and correlation with V3 genotypic prediction tools in recent HIV-1 seroconverters. *J Acquir Immune Defic Syndr*, 2008. 48(3): p. 241-4.

Deng H, et al., Identification of a major co-receptor for primary isolates of HIV-1. *Nature* 1996;381:661-6.

Deng HK, et al., Expression cloning of new receptors used by simian and human immunodeficiency viruses. *Nature* 1997;388:296- 300.

De Jong JJ, et al., 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*. 1992;66:6777- 80.

De Wolf F, et al., Syncytium-inducing and non-syncytium-inducing capacity of human immunodeficiency virus type 1 subtypes other than B: phenotypic and genotypic characteristics. WHO Network for HIV Isolation and Characterization. *AIDS Res Hum Retroviruses* 1994;10:1387-400.

Delwart EL, et al., Slower evolution of human immunodeficiency virus type 1 quasiespecies during progression to AIDS. *J Virol*. 1997;71:7498-508.

Derdeyn CA, et al.. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science*, 2004, 303:2019–2022.

Dickover RE, et al., Perinatal transmission of major, minor, and multiple maternal human immunodeficiency virus type 1 variants in utero and intrapartum. *J Virol*. 2001;75:2194-203.

Dickover R, et al., Role of maternal autologous neutralizing antibody in selective perinatal transmission of human immunodeficiency virus type 1 escape variants. *J Virol*, 2006, 80: 6525–6533.

Dong XN, et al., Short communication: HIV type 1 phenotype, tropism, and sequence patterns: association and preference. *AIDS Res Hum Retroviruses* 2005;21:234-8.

Dunn DT, et al., The sensitivity of HIV-1 DNA polymerase chain reaction in the neonatal period and the relative contributions of intra-uterine and intra- partum transmission. *Aids*, 1995. 9(9): p. F7-11.

Ehrnst A, et al., Efficient isolation of HIV from plasma during different stages of HIV infection. *J Med Virol*. 1988;26:23-32.

Esbjörnsson J, et al., Frequent CXCR4 tropism of HIV-1 subtype A and CRF02_AG during late-stage disease--indication of an evolving epidemic in West Africa. *Retrovirology*, 2010, 7;23

European Collaborative Study. HIV-infected pregnant women and vertical transmission in Europe since 1986. *AIDS* 2001;15:761–70.

Fauci AS, et al., Immunopathogenic mechanisms of HIV infection. *Annals of internal medicine* 1996,124:654-663.

Feng Y, et al., HIV-1 entry cofactor: functional cDNA cloning of a seven-

transmembrane, G protein-coupled receptor. *Science* 1996;272:872- 7.

Fenyo, E.M., J. Albert, and B. Åsjö, Replicative capacity, cytopathic effect and cell tropism of HIV. *Aids*, 1989. 3 Suppl 1: p. S5-12.

Fouchier RA, et al., Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol.* 1992;66:3183-7.

Gallo RC, et al., Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science*, 1984. 224(4648): p. 500-3.

Gao F, et al. Origin of HIV-1 in the chimpanzee *Pan troglodytes*. *Nature* 1999;397:436-441.

Gao F, et al. Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature* 1992;358:495-499.

Gao F, et al. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J Virol.*, 1996. Oct(70(10)): p. 7013-29.

Garcia JV, and Miller AD. Downregulation of cell surface CD4 by nef. *Res Virol.* 1992;143:52-5.

Garrido C, et al., Evaluation of eight different bioinformatics tools to predict viral tropism in different human immunodeficiency virus type 1 subtypes. *J Clin Microbiol*, 2008. 46(3): p. 887-91.

Geijtenbeek TB, et al., DC-SIGN, a dendritic cell-specific HIV-1- binding protein that enhances trans-infection of T cells. *Cell*, 2000, 100:587–597.

Gnanakaran S, et al., Recurrent signature patterns in HIV-1 B clade envelope glycoproteins associated with either early or chronic infections. *PLoS Pathog.* 2011 Sep;7(9):e1002209. doi: 10.1371/journal.ppat.1002209.

Goodenow M, et al., HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. *J Acquir Immune Defic Syndr* 1989;2:344- 352.

Gray GE. Walking the tightrope in prevention of mother-to-child transmission of HIV infection. *Clin Infect Dis*, 2008. 46(4): p. 622-4.

Grob PM, et al., Nonnucleoside inhibitors of HIV-1 reverse transcriptase: nevirapine as a prototype drug. *AIDS Res Hum Retroviruses* 1992;8:145–52.

Grosch-Worner I, et al., An effective and safe protocol involving zidovudine and caesarean section to reduce vertical transmission of HIV-1 infection. *Aids*, 2000. 14(18): p. 2903-11.

Grossman Z, et al., Pathogenesis of HIV infection: what the virus spares is as important as what it destroys. *Nat Med.* 2006 Mar;12(3):289-95.

Guadalupe M, et al., Severe CD4+ T-cell depletion in gut lymphoid tissue during primary human immunodeficiency virus type 1 infection and substantial delay in restoration following highly active antiretroviral therapy. *Journal of virology* 2003;77:11708-11717.

Ha TT, et al., HIV-1 mother to child transmission, post-test counselling, and antiretroviral prophylaxis in Northern Vietnam - a prospective observational study. *Scand J Infect Dis* 2012 Nov;44(11):866-73.

Harrigan PR. Screening for HIV tropism using population-based V3 genotypic analysis: a retrospective virological outcome analysis using stored plasma screening samples from MOTIVATE-1. 5th IAS Conference on HIV Pathogenesis and Treatment, 2009. Abstract no. WELBA101

Habeshaw JA, and Dalgleish AG. The relevance of HIV env/CD4 interactions to the pathogenesis of acquired immune deficiency syndrome. *J Acquir Immune Defic Syndr*. 1989;2:457-68.

Hladik F. and M. J. McElrath, "Setting the stage: host invasion by HIV.," *Nature reviews. Immunology*, vol. 8, no. 6, pp. 447–57, Jun. 2008.

Hladik F. and T. J. Hope, "HIV infection of the genital mucosa in women," *Current HIV/AIDS Reports*, vol. 6, no. 1, pp. 20–28, Jan. 2009.

Holder, M. and P.O. Lewis, Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet*, 2003. 4(4): p. 275-84.

Holland JJ, De La Torre JC, Steinhauer DA. RNA virus populations as quasispecies. *Curr Top Microbiol Immunol* 1992;176:1-20.

Hu Q, et al., Identification of ENV determinants in V3 that influence the molecular anatomy of CCR5 utilization. *J Mol Biol*. 2000a;302:359-75.

Hu QX, et al., Evolution of the human immunodeficiency virus type 1 envelope during infection reveals molecular corollaries of specificity for coreceptor utilization and AIDS pathogenesis. *J Virol*. 2000b;74:11858-72.

Huang W, et al., Coreceptor tropism in human immunodeficiency virus type 1 subtype D: high prevalence of CXCR4 tropism and heterogeneous composition of viral populations. *J Virol*, 2007. 81(15): p. 7885-93.

Huang, W., Eshleman, S.H. & Toma, J. Vertical transmission of X4-tropic and dual-tropic HIV-1 in five Ugandan mother-infant pairs. *AIDS*, 2009, 23;1903-1908.

Hutter G, et al., Long- term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *The New England journal of medicine* 2009;360:692-698.

Hwang SS, et al., Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* 1991;253:71-4.

Hymes KB, et al., Kaposi's sarcoma in homosexual men-a report of eight cases. *Lancet*, 1981. 2(8247): p. 598-600.

Janssens W, et al., Near full-length genome analysis of HIV type 1 CRF02_AG subtype C and CRF02_AG subtype G recombinants. *AIDS Res Hum Retroviruses*, 2000. 16(12): p. 1183-9.

Jensen, M.A. and A.B. van 't Wout, Predicting HIV-1 coreceptor usage with sequence analysis. *AIDS Rev*, 2003. 5(2): p. 104-12.

John-Stewart GC, et al., J Infect Dis. Subtype C Is associated with increased vaginal shedding of HIV-1. *J Infect Dis*. 2005;192:492-6.

Jourdain G, et al., Risk factors for in utero or intrapartum mother-to-child transmission of human immunodeficiency virus type 1 in Thailand. *J Infect Dis*, 2007. 196(11): p. 1629-36.

Keele BF, et al., Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*, 2006;313:523-526.

Keele BF, et al., Identification and characterization of transmitted and early founder

virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.*, 2008, 105:7552–7557.

Klatzmann D, et al., Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes. *Science*, 1984, 225:59-63.

Kiepiela P, et al., Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature*, 2004, 432:769-775.

Kilewo C, et al., Prevention of Mother-to-Child Transmission of HIV-1 Through Breast-Feeding by Treating Infants Prophylactically With Lamivudine in Dar es Salaam, Tanzania: The Mitra Study. *J Acquir Immune Defic Syndr*, 2008.

Kiwanuka N, et al., Effect of human immunodeficiency virus Type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. *J Infect Dis*, 2008. 197(5): p. 707-13.

Kittinunvorakoon, C, et al., Mother to child transmission of HIV-1 in a Thai population: role of virus characteristics and maternal humoral immune response. *J Med Virol*, 2009, 81, 768-778.

Kliks SC, et al., Features of HIV-1 that could influence maternal-child transmission. *JAMA*, 1994, 272: 467–474.

Kong X, et al. The human immunodeficiency virus type 1 envelope confers higher rates of replicative fitness to perinatally transmitted viruses than to nontransmitted viruses. *J Virol*, 2008, 82: 11609–11618.

Korber B: HIV signature and sequence variation analysis. In: *Computational Analysis of HIV Molecular Sequences* (Rodrigo AG, Learn GH, Eds.). Kluwer Academic Publishers, Dordrecht, Netherlands, 2000, pp. 55–72.

Koot M, et al., Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS. *Ann Intern Med*. 1993;118:681-8.

Kourtis AP, et al., Mother-to-child transmission of HIV-1: timing and implications for prevention. *Lancet Infect Dis*, 2006. 6(11): p. 726-32.

Lama J. The physiological relevance of CD4 receptor down-modulation during HIV infection. *Curr HIV Res*. 2003;1:167-84.

Le CT, et al., Preventing Mother-to-Child Transmission of HIV in Vietnam: An Assessment of Progress and Future Directions. *J Trop Pediatr*, 2008.

Lehman, D.A. and C. Farquhar, Biological mechanisms of vertical human immunodeficiency virus (HIV-1) transmission. *Rev Med Virol*, 2007. 17(6): p. 381-403.

Leitner T, et al., Molecular epidemiology and MT-2 cell tropism of Russian HIV type 1 variant. *AIDS Res Hum Retroviruses* 1996;12:1595-603.

Leitner T, HIV Sequence Compendium. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM., 2005.

Lemey P, et al., Tracing the origin and history of the HIV-2 epidemic. *Proc Natl Acad Sci U S A* 2003,100:6588-6592.

Lemey P, R.A., Pybus OG., HIV evolutionary dynamics within and among hosts. *AIDS Rev.*, 2006. 8(3): p. 125-40.

Lengauer T, et al., Bioinformatics prediction of HIV coreceptor usage. *Nat Biotechnol*, 2007. 25(12): p. 1407-10.

Levy JA, et al., Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science*, 1984. 225(4664): p. 840-2.

Li WH, Tanimura M, Sharp PM. Rates and dates of divergence between AIDS virus nucleotide sequences. *Mol Biol Evol* 1988;5:313-330.

Liao H, et al. Phylodynamic analysis of the dissemination of HIV-1 CRF01_AE in Vietnam. *Virology*, 2009, 391:51–56.

Lindgren S, et al., Screening for HIV-1 antibodies in pregnancy: results from the Swedish national programme. *BMJ* 1993;307:1447-51.

Lyles RH, et al. Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. Multicenter AIDS Cohort Study. *The Journal of infectious diseases* 2000;181:872-880.

Luzuriaga, K. and J.L. Sullivan, Prevention of mother-to-child transmission of HIV infection. *Clin Infect Dis*, 2005. 40(3): p. 466-7.

Maddon PJ, et al., The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* 1986;47:333-48.

Mansky LM, Temin HM. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* 1995;69:5087-5094.

Mastro TD, et al., Why do HIV-1 subtypes segregate among persons with different risk behaviors in South Africa and Thailand? *AIDS* 1997;11:113-6.

Matala, E, et al., Biological characterization of HIV type 1 envelope V3 regions from mothers and infants associated with perinatal transmission. 2001, *AIDS Res Hum Retroviruses* 17, 1725-1735.

Matt C, Genetic determinants of pediatric HIV-1 infection: vertical transmission and disease progression among children. *Mol Med*, 2001. 7(9): p. 583-9.

McDonald C, et al., Why are children still being infected with HIV? Experiences in the prevention of mother-to-child transmission of HIV in south London. *Sex Transm Infect*, 2007. 83(1): p. 59-63.

McDougal JS, et al., The T4 glycoprotein is a cell-surface receptor for the AIDS virus. *Cold Spring Harb Symp Quant Biol.* 1986;51 Pt 2:703-11.

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

McCutchan FE, et al., Genetic variants of HIV-1 in Thailand. *AIDS Res Hum Retroviruses*, 1992. 8(11): p. 1887-95.

McClutchan FE, et al., Precise mapping of recombination breakpoints suggests a common parent of two BC recombinant HIV type 1 strains circulating in China. *AIDS Res Hum Retroviruses.* 2002. 18(15): p. 1135-40.

Mellors JW, et al. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Annals of internal medicine* 1997;126:946-954.

Misrahi M, et al., CCR5 chemokine receptor variant in HIV-1 mother-to- child transmission and disease progression in children. French Pediatric HIV Infection Study Group. *Jama*, 1998. 279(4): p. 277-80.

Moore RD, Chaisson RE. Natural history of HIV infection in the era of combination antiretroviral therapy. *AIDS* 1999;13:1933-42.

Morgado MG, et al., Molecular epidemiology of HIV-1 in Brazil: high prevalence of HIV-1 subtype B and identification of an HIV-1 subtype D infection in the city of Rio de Janeiro, Brazil. Evandro Chagas Hospital AIDS Clinical Research Group. *J Acquir Immune Defic Syndr Hum Retrovirol.* 1998;18:488-94.

Namukwaya Z, et al., The Impact of Maternal Highly Active Antiretroviral Therapy and Short-Course Combination Antiretrovirals for Prevention of Mother- to-Child Transmission on Early Infant Infection Rates at the Mulago National Referral Hospital in Kampala, Uganda, January 2007 to May 2009. *J Acquir Immune Defic Syndr*, 2010.

Nei M and Gojobori T: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986;3:418–426.

Neilson JR, et al., Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. *J Virol.*, 1999. 73(5): p. 4393-403.

Newell ML. Current issues in the prevention of mother-to-child transmission of HIV-1 infection. *Trans R Soc Trop Med Hyg*, 2006. 100(1): p. 1-5.

Nowak P, et al., The selection and evolution of viral quasispecies in HIV-1 infected children. *HIV Med.* 2002;3:1-11.

O'Brien SJ, Nelson GW. Human genes that limit AIDS. *Nature genetics* 2004;36:565-574.

Oberlin E, et al., The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. *Nature* 1996;382:833-5.

Ogert RA, et al., N-linked glycosylation sites adjacent to and within the V1/V2 and the V3 loops of dualtropic human immunodeficiency virus type 1 isolate DH12 gp120 affect coreceptor usage and cellular tropism. 2001, *J Virol* 75, 5998-6006.

Oosterhoff P, et al., Dealing with a positive result: routine HIV testing of pregnant women in Vietnam. *AIDS Care*, 2008. 20(6): p. 654-9.

Palella FJ Jr. et al., Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *N. Engl. J. Med.* 338,853-860 (1998).

Piguet V, et al., The downregulation of CD4 and MHC-I by primate lentiviruses: a paradigm for the modulation of cell surface receptors. *Immunol Rev.* 1999;168:51-63

Qiu Z, et al., 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.*, 2005. 21(12): p. 1051-6.

Pasquier C, et al. Molecular evidence for mother to child transmission of multiple variants by analysis of RNA and DNA sequences of HIV-1. *J Virol* 1998; UP: 8493-8501.

Peeters M, Sharp PM. Genetic diversity of HIV-1: the moving target. *AIDS* 2000;14 Suppl 3:S129-140. Peeters M, C.M., Delaporte E, Genetic diversity and phylogeographic distribution of SIV: how to understand the origin of HIV. *Med Sci (Paris)*. 2008. 24(6-7): p. 621-8.

Piacentini L, et al., Genetic correlates of protection against HIV infection: the ally within. *Journal of internal medicine* 2009;265:110-124.

Pollakis G, et al., N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization. *J Biol Chem.* 2001;276:13433-41.

Polzer S, et al., Loss of N-linked glycans in the V3-loop region of gp120 is correlated to an enhanced infectivity of HIV-1. *Glycobiology* 2001;11:11-9.

Polzer S, et al., The N-linked glycan g15 within the V3 loop of the HIV-1 external glycoprotein gp120 affects coreceptor usage, cellular tropism, and neutralization. *Virology* 2002;304:70-80.

Popovic M, et al., Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science*, 1984. 224(4648): p. 497-500.

Poveda E, et al., Design and validation of new genotypic tools for easy and reliable estimation of HIV tropism before using CCR5 antagonists. *J Antimicrob Chemother*, 2009. 63(5): p. 1006-10.

Pramanik L, et al., Charged amino acid patterns of coreceptor use in the major subtypes of human immunodeficiency virus type 1. *J Gen Virol*, 2011, 92;1917-1922.

Puffer BA, et al., Determinants within gp120 and gp41 contribute to CD4 independence of SIV Envs. *Virology*, 2004, 327:16–25.

Raymond S, et al., Frequency of CXCR4-using viruses in primary HIV-1 infections using ultra-deep pyrosequencing. *AIDS*, 2011. 25(13): p. 1668-70.

Reeves JD, et al. Sensitivity of HIV-1 to entry inhibitors correlates with envelope/coreceptor affinity, receptor density, and fusion kinetics. *Proc. Natl. Acad. Sci. U. S. A.*, 2002, 99:16249–16254.

Resch, W., Hoffman, N. & Swanstrom, R. Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks. *Virology*, 2001, 288;51–62.

Renjufi B, et al., Tanzanian Vitamin and HIV Study Group. Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D. *AIDS* 2004;18:1629-36.

Rizzuto CD, et al., HIV-1 nomenclature proposal. *Science*, 2000. 288(5463): p. 55-6.

Roos MT, et al., Viral phenotype and immune response in primary human immunodeficiency virus type 1 infection. *J. Infect. Dis.* 1992, 165:427–432.

Sagar M, et al., Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. *J. Infect. Dis.* 2009, 199:580–589.

Salemi M, et al., Different population dynamics of human T cell lymphotropic virus type II in intravenous drug users compared with endemically infected tribes. *Proceedings of the National Academy of Sciences of the United States of America* 1999,96:13253-13258.

Salemi M, et al., 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 journal : official publication of the Federation of American Societies for Experimental Biology* 2001,15:276-278.

Salvatori, F. & Scarlatti, G. HIV type 1 chemokine receptor usage in mother-to-child transmission. *AIDS Res Hum Retroviruses*, 2001, 17:925-935.

Sato H, et al., Evolution and biological characterization of human immunodeficiency virus type 1 subtype E gp120 V3 sequences following horizontal and vertical virus transmission in a single family. *J Virol* 1999, 73;3551-3559.

- Scarlatti G, et al., Transmission of human immunodeficiency virus type 1 (HIV-1) from mother to child correlates with viral phenotype. *Virology*. 1993 Dec;197(2):624-9.
- Scarlatti G, et al., Neutralizing antibodies and viral characteristics in mother-to-child transmission of HIV-1. *AIDS*, 1993, 7 Suppl 2: S45–48.
- Scarlatti G, et al., In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression. *Nat Med*. 1997;3:1259-65.
- Scarlatti G. Mother-to-child transmission of HIV-1: advances and controversies of the twentieth centuries. *AIDS Rev*, 2004. 6(2): p. 67-78.
- Schacker TW, et al., Biological and virologic characteristics of primary HIV infection. *Annals of internal medicine* 1998;128:613-620.
- Scheppler JA, et al., Down- modulation of MHC-I in a CD4+ T cell line, CEM-E5, after HIV-1 infection. *J Immunol*. 1989;143:2858-66.
- Schuitmaker H, et al. Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocyctotropic to T-cell-tropic virus population. *J. Virol*. 1992, 66:1354–1360.
- Seclen E, et al., High sensitivity of specific genotypic tools for detection of X4 variants in antiretroviral- experienced patients suitable to be treated with CCR5 antagonists. *J Antimicrob Chemother*, 2010. 65(7): p. 1486-92.
- Seillier-Moiseiwitsch F, Margolin BH, Swanstrom R. Genetic variability of the human immunodeficiency virus: statistical and biological issues. *Annual review of genetics* 1994;28:559-596.
- Shang F, et al., Characterization of monoclonal antibodies against the human immunodeficiency virus matrix protein, p17gag: identification of epitopes exposed at the surfaces of infected cells. *J Virol*, 1991. 65(9): p. 4798-804.
- Shankarappa R, et al., Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol*. 1999;73:10489-502.
- Shioda T, Levy JA, Cheng-Mayer C. 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 U S A* 1992;89:9434-8.
- Sing, T., Low, A.J. & Beerenwinkel, N. Geno2Pheno: Predicting HIV co-receptor usage based on genetic and clinical covariates. *Antiviral Therapy*, 2007, 12,1097–1106.
- Sodroski J. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science* 1998;280:1949-53.
- Su L, et al., Characterization of a virtually full-length human immunodeficiency virus type 1 genome of a prevalent intersubtype (C/B') recombinant strain in China. *J Virol*, 2000. 74(23): p. 11367-76.
- Stegen K, et al., A sensitive in-house RT-PCR genotyping system for combined detection of plasma HIV-1 and assessment of drug resistance. *J Virol Methods* 2006; 133:137–45.
- Swanstrom R, and Wills JW. Synthesis, assembly and processing of viral proteins. In (ed. J. M. Coffin et al.), *Retroviruses*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 1997.

Swenson L, et al., Improved detection of CXCR4-using HIV by V3 genotyping: application of population-based and 'deep' sequencing to plasma RNA and proviral DNA. *J Acquir Immune Def Syndr* 2010; 54:506–510.

Takehisa J, et al., Origin and biology of simian immunodeficiency virus in wild-living western gorillas. *J Virol* 2009;83:1635-1648.

Tapia N, et al., Influence of human immunodeficiency virus type 1 subtype on mother-to-child transmission. *J Gen Virol*. 2003;84:607-13.

Taylor BS, et al., The challenge of HIV-1 subtype diversity. *N Engl J Med*, 2008. 358(15): p. 1590-602.

Tonie Cilliers JN, et al., The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C. *J Virol*, 2003. 77(7): p. 4449-56.

Trkola A, et al., CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature* 1996;384:184-7.

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

Tscherning-Casper C, et al. Coreceptor usage of HIV-1 isolates representing different genetic subtypes obtained from pregnant Cameroonian women. European Network for In Utero Transmission of HIV-1. *J Acquir Immune Defic Syndr*. 2000;24:1-9.

UNAIDS, AIDS epidemic update: December 2011 available at http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2011/20111130_ua_report_en.pdf

UNAIDS, 2012 Report on the global AIDS epidemic. Available at http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_en.pdf

UNAIDS. Global Plan towards the elimination of new HIV infections among children by 2015 and keeping their mothers alive. 2011-2015

Van Heuverswyn F, et al., Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature*, 2006. 444(7116): p. 164.

van Harmelen J, et al., An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. *AIDS* 1997;11:81-7.

Vandekerckhove LP, et al., European guidelines on the clinical management of HIV-1 tropism testing. *Lancet Infect Dis*, 2011. 11(5): p. 394-407.

Verma R, et al., Distribution of CCR5delta32, CCR2-64I and SDF1-3'A and plasma levels of SDF-1 in HIV-1 seronegative North Indians. *J Clin Virol*, 2007. 38(3): p. 198-203.

Verrier F, et al., Role of the HIV type 1 glycoprotein 120 V3 loop in determining coreceptor usage. *AIDS Res Hum Retroviruses* 1999;15:731-43.

Wade CM, Lobidel D, Brown AJ. Analysis of human immunodeficiency virus type 1 env and gag sequence variants derived from a mother and two vertically infected children provides evidence for the transmission of multiple sequence variants. *J General Virol* 1998; UW: 1055-1068.

Wasik TJ, et al., Effects of CCR5-delta32, CCR2-64I and SDF-1-3'A polymorphic alleles on human immunodeficiency virus 1 (HIV-1) infection in the Polish population.

Wiad Lek, 2005. 58(9-10): p. 500-7.

Weissenhorn W, et al., Atomic structure of the ectodomain from HIV-1 gp41. *Nature* 1997;387:426-30.

WHO, Antiretroviral drugs for treating pregnant women and preventing HIV infection in infants: towards universal access. Recommendation for a public health approach. Available at http://whqlibdoc.who.int/publications/2010/9789241599801_eng.pdf. 2011.

Whitcomb JM, et al., Development and characterization of a novel single-cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor tropism. *Antimicrob Agents Chemother*, 2007. 51(2): p. 566-75.

Wolfs TF, et al., HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. *Virology* 1992;189:103-10.

Wolinsky SM, et al., Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. *Science* 1992;255:1134-7.

Wu L, et al., CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature* 1996;384:179-83.

Wu X, et al., Neutralization escape variants of human immunodeficiency virus type 1 are transmitted from mother to infant. *J Virol*, 2006, 80: 835–844.

Wyatt R, et al., 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*. 1995;69:5723-33.

Worobey M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* 2008;455:661-664.

Yang C, et al., Genetic diversity of HIV-1 in western Kenya: subtype-specific differences in mother-to-child transmission. *AIDS* 2003;17:1667-74.

Yu XF, et al., Phenotypic and genotypic characteristics of human immunodeficiency virus type 1 from patients with AIDS in northern Thailand. *J Virol* 1995;69:4649-4655

Zhang H, et al., Phylogenetic and phenotypic analysis of HIV type 1 env gp120 in cases of subtype C mother-to-child transmission. *AIDS Res Hum Retroviruses* 2002, 18, 1415-1423.

Zhang H, et al., Functional properties of the HIV-1 subtype C envelope glycoprotein associated with mother-to-child transmission. *Virology*, 2009.

Zhu T, et al., Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science*, 1993, 261:1179–1181.

Zhu T, et al., An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature*, 1998. 391(6667): p. 594-7.

Åsjö B, et al., Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. *Lancet* 1986;2:660-2.