HIV-1 evolution, disease progression and molecular epidemiology of HIV-1 single and HIV-1 and HIV-2 dual-infected individuals in Guinea-Bissau

Esbjörnsson, Joakim

2010

Document Version:
Publisher's PDF, also known as Version of record

HIV-1 evolution, disease progression and molecular epidemiology of HIV-1 single and HIV-2 dual-infected individuals in Guinea-Bissau

Joakim Esbjörnsson

Department of Experimental Medical Science, Section of Molecular Virology,
Lund University, Lund, Sweden

Hallands Nation

LUND UNIVERSITY
Faculty of Medicine

Fakultetsopponent

Professor Anne-Mieke Vandamme
Rega Institute for Medical Research,
Katholieke Universiteit Leuven, Leuven, Belgium

Handledare

Docent Patrik Medstrand
Department of Experimental Medical Science, Section of Molecular Virology,
Lund University, Lund, Sweden
Professor Eva Maria Fenyö
Department of Laboratory Medicine Lund, Division of Medical Microbiology,
Lund University, Lund, Sweden
Filosofie doktor Anders Kvist
Department of Clinical Sciences, Section of Oncology,
Lund University and Skåne University Hospital, Lund, Sweden

Akademisk avhandling

som med vederbörligt tillstånd av Medicinska Fakulteten vid Lunds universitet för avläggande av doktorsexamen i medicinsk vetenskap kommer att offentligen försvaras i Belfragesalen, BMC D15, Sölvegatan 19, Lund, fredagen den 17 december 2010, kl. 9.00.

DOKUMENTDATABLAD enl SIS 61 41 21

Organization
LUND UNIVERSITY

Document name
DOCTORAL DISSERTATION

Date of issue
December 17, 2010

Sponsoring organization

Author(s)
Joakim Esbjörnsson

Title and subtitle: HIV-1 evolution, disease progression and molecular epidemiology of HIV-1 single and HIV-1 and HIV-2 dual-infected individuals in Guinea-Bissau

Abstract

The two genetically related human lentiviruses known today, HIV-1 (which is pandemic) and HIV-2 (which mainly is confined to West Africa), are the causative agents of AIDS. Progressive immune dysfunction and AIDS develop in most cases of untreated HIV-1 infection, but only in approximately 25-30% of HIV-2 infected individuals. The V1-V3 region of the HIV-1 *env* gp120 is important for HIV-1 coreceptor use, and represents an informative region for both molecular epidemiology and intrapatient phylogenetic analyses due to high level of genetic variation. In this doctoral dissertation, HIV-1 V1-V3 sequences in combination with clinical disease markers were used to investigate HIV-1 evolution, disease progression, coreceptor tropism and molecular epidemiology of HIV-1. All sequences were derived from single (HIV-1 only) or dual-infected (HIV-1 and HIV-2) individuals from Guinea-Bissau, West Africa. The main findings was that CRF02_AG represents the most common form of HIV-1 in Guinea-Bissau, and that HIV-1 was introduced into the country on at least six different occasions between 1976 and 1981. Dual-infected individuals had a 46% lower mortality rate and a 53% longer progression-time to AIDS compared to single-infected individuals. $CD4^+$ T cell counts were higher at corresponding time-points after infection among dual-infected individuals, reflecting the slower disease progression rate at the cellular immune level. In addition, $CD8^+$ T cell counts were increasing at a faster rate in single than in dual-infected individuals. Stratified analyses showed that these observations were most prominent among the subgroup of dual-infected individuals that became HIV-1 infected after an established HIV-2 infection. Moreover, the HIV-1 genetic diversity was significantly lower in dual than in single-infected individuals at comparable time-points after infection. HIV-1 coreceptor tropism was investigated in late-stage disease by the use of a recombinant virus phenotypic assay that were confirmed to accurately predict the coreceptor tropism of HIV-1 subtype A and CRF02_AG. CXCR4 tropism has been coupled to an increased HIV-1 disease progression rate in late-stage disease. We found that HIV-1 CRF02_AG CXCR4 tropism was frequent (86%) and increased over time on the population level, indicating an evolving epidemic. In addition, a literature analysis showed a similar evolving epidemic for HIV-1 subtype C. Genotypic analysis suggested that the total number of charged amino acids could be important in predicting HIV-1 CRF02_AG coreceptor tropism. Finally, HIV-1 CXCR4-tropism was more common in single (79%) than in dual-infected individuals (35%). Understanding the underlying mechanisms responsible for the inhibitory effects exerted by HIV-2 against HIV-1 could be important for the development of future HIV-1 vaccines and therapeutics.

Key words: HIV-1, HIV-2, Guinea-Bissau, subtype, disease progression, evolution, coreceptor tropism, CXCR4, gp120, V1-V3, CD4%, CD8%, soluble immune markers

Classification system and/or index termes (if any):

Supplementary bibliographical information:

Language
English

ISSN and key title:
1652-8220

ISBN
978-91-86671-41-9

Recipient's notes

Number of pages
226

Price

Security classification

HIV-1 evolution, disease progression and molecular epidemiology of HIV-1 single and HIV-1 and HIV-2 dual-infected individuals in Guinea-Bissau

Joakim Esbjörnsson

Lund 2010

Department of Experimental Medical Science, Section of Molecular Virology, Lund University, Lund, Sweden

LUND UNIVERSITY
Faculty of Medicine

Joakim Esbjörnsson
HIV-1 evolution, disease progression and molecular epidemiology of HIV-1 single and HIV-1 and HIV-2 dual-infected individuals in Guinea-Bissau
Faculty of Medicine, Lund University, Lund, 2010

This work was supported by: The Faculty of Medicine, Lund, Lund University; The Swedish Research Council; The Swedish International Development Cooperation Agency/Department for Research Cooperation (SIDA/Sarec); The Crafoord Foundation; The Lars Hierta Memorial Foundation; The Royal Physiographic Society; The Konsul Thure Carlsson Fund for Medical Research; The Knut and Alice Wallenberg Foundation; The Anna Cederbergs Foundation for Medical Research; The Carl Swartz Memorial Foundation; and The Medical Society in Lund.

ISSN 1652-8220
ISBN 978-91-86671-41-9
Lund University, Faculty of Medicine Doctoral Dissertation Series 2010:125

Printed by Media-Tryck, Lund, 2010

To my family

Ever tried. Ever failed. No matter. Try again. Fail again. Fail better.

Samuel Beckett (1906-1989)

The significant problems we face cannot be solved at the same level of thinking we were at when we created them.

Albert Einstein (1879-1955)

TABLE OF CONTENTS

LIST OF PAPERS

I **Esbjörnsson J**, Mild M, Månsson F, Norrgren H and Medstrand P. HIV-1 molecular epidemiology in Guinea-Bissau, West Africa: Origin, demography and migrations. *Submitted.*

II **Esbjörnsson J***, Månsson F*, Kvist A, Isberg P-E, Biague A, da Silva Z, Fenyö EM, Norrgren H and Medstrand P. Natural inhibition of HIV-1 disease progression by contemporaneous HIV-2 infection. *Submitted.*
*These authors contributed equally to this study.

III Månsson F*, **Esbjörnsson J***, Norrgren H and Medstrand P. Previous HIV-2 infection results in better survival and disease-free follow-up time among HIV-1 seroincident individuals. *Manuscipt.*
*These authors contributed equally to this study.

IV **Esbjörnsson J**, Månsson F, Martínez-Arias W, Vincic E, Biague A, da Silva Z, Fenyö EM, Norrgren H, Medstrand P. 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, 22;7:23.*

V **Esbjörnsson J**, Månsson F, Fenyö EM, Norrgren H and Medstrand P. Natural alteration of HIV-1 coreceptor tropism by contemporaneous HIV-2 infection. *Manuscipt.*

The following papers are not included in this doctoral dissertation but are of relevance to the field:

- Repits J, Öberg M, **Esbjörnsson J**, Medstrand P, Karlsson A, Albert J, Fenyö EM, Jansson M. Selection of human immunodeficiency virus type 1 R5 variants with augmented replicative capacity and reduced sensitivity to entry inhibitors during severe immunodeficiency. *J Gen Virol. 2005, 86(Pt 10): 2859-2869.*

- Mild M, **Esbjörnsson J**, Fenyö EM, Medstrand P. Frequent intrapatient recombination between human immunodeficiency virus type 1 R5 and X4 envelopes: implications for coreceptor switch. *J Virol. 2007, 81(7): 3369-3376.*

- Mild M, Kvist A, **Esbjörnsson J**, Karlsson I, Fenyö EM, Medstrand P. Differences in molecular evolution between switch (R5 to R5X4/X4-tropic) and non-switch (R5-tropic only) HIV-1 populations during infection. *Infect Genet Evol. 2010, 10(3): 356-364.*

ABBREVIATIONS

aa	Amino acid
AIDS	Acquired immunodeficiency syndrome
AZT	Zidovudine
bp	Base pair
C1-C5	Constant region 1 to 5
CCR	CC chemokine receptor
CRF	Circulating recombinant form
C-terminus	Corboxy-terminus
CTL	Cytotoxic T lymphocyte
CXCR	CXC chemokine receptor
DC	Dendritic cell
DC-SIGN	DC-specific ICAM-3 grabbing non-integrins
DNA	Deoxyribonucleic acid
ds	Double stranded
ELISA	Enzyme-linked immunosorbent assay
env	Envelope gene
Env	Envelope protein
gag	Group antigen gene
GALT	Gut-associated lymphoid tissue
gp	Glycoprotein
HA	Hemagglutinin
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
HIV-1	Human immunodeficiency virus type-1
HIV-2	Human immunodeficiency virus type-2
HTLV-3	Human T-cell lymphotropic virus 3
IFN	Interferon
IL	Interleukin
IN	Integrase
kb	Kilobases
LAV	Lymphoadenopathy associated virus
LTNP	Long term non-progressor
LTR	Long terminal repeat
M-group	Major group
MHC	Major histocompatibility complex
MIP	Macrophage inflammatory protein

ML	Maximum-likelihood
nef	Negative regulatory factor gene
N-group	Non-M, non-O group
NK cell	Natural killer cell
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside reverse transcriptase inhibitor
NSI	Non-syncytium inducing
N-terminus	Amino-terminal
O-group	Outlier group
PBMC	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PIC	Pre-integration complex
PNGS	Potential N-linked glycosylation site
pol	Polymerase gene
qPCR	Quantitative polymerase chain reaction
PR	Protease
RANTES	Regulated on activation, normal T-cell expressed, and secreted
rev	Regulator of virion protein gene
RNA	Ribonucleic acid
RRE	Rev responsive element
RT	Reverse transcriptase
SDF-1	Stromal cell derived factor 1
SI	Syncytium inducing
SIV	Simian immunodeficiency virus
ss	Single stranded
TAR	Transactivation-responsive RNA
tat	Transactivator gene
TNF	Tumor-necrosis factor
V1-V5	Variable region 1 to 5
vif	Virion infectivity factor gene
vpr	Viral protein R gene
vpu	Viral protein U gene

AIMS OF THIS DOCTORAL DISSERTATION

The overall aim of this doctoral dissertation was to study the molecular epidemiology in Guinea-Bissau and different correlates in disease progresion rate between HIV-1 single and HIV-1 and HIV-2 dual-infected individuals.

Paper I: To investigate the molecular epidemiology in Guinea-Bissau by analysis of HIV-1 *env* gp120 V1-V3 sequences.

Paper II: To compare differences in disease progression rate between HIV-1 single and HIV-1 and HIV-2 dual-infected individuals.

Paper III: To investigate if HIV-1 and HIV-2 dual-infected individuals with previous HIV-2 infection differ in disease progression rate compared to dual-infected individuals with simultaneous recorded HIV-1 and HIV-2 infection.

Paper IV: To study the prevalence of CXCR4 tropism in late-stage disease for HIV-1 CRF02_AG over time on a population level.

Paper V: To compare differences in CXCR4 tropism in late-stage disease between HIV-1 single and HIV-1 and HIV-2 dual-infected individuals.

SUMMARY

The two genetically related human lentiviruses known today, HIV-1 (which is pandemic) and HIV-2 (which mainly is confined to West Africa), are the causative agents of AIDS. Progressive immune dysfunction and AIDS develop in most cases of untreated HIV-1 infection, but only in approximately 25-30% of HIV-2 infected individuals. The V1-V3 region of the HIV-1 *env* gp120 is important for HIV-1 coreceptor use, and represents an informative region for both molecular epidemiology and intra-patient phylogenetic analyses due to high level of genetic variation. In this doctoral dissertation, HIV-1 V1-V3 sequences in combination with clinical disease markers were used to investigate HIV-1 evolution, disease progression, coreceptor tropism and molecular epidemiology of HIV-1. All sequences were derived from single (HIV-1 only) or dual-infected (HIV-1 and HIV-2) individuals from Guinea-Bissau, West Africa. The main findings was that CRF02_AG represents the most common form of HIV-1 in Guinea-Bissau, and that HIV-1 was introduced into the country on at least six different occasions between 1976 and 1981. Dual-infected individuals had a 46% lower mortality rate and a 53% longer progression-time to AIDS compared to single-infected individuals. $CD4^+$ T cell counts were higher at corresponding time-points after infection among dual-infected individuals, reflecting the slower disease progression rate at the cellular immune level. In addition, $CD8^+$ T cell counts were increasing at a faster rate in single than in dual-infected individuals. Stratified analyses showed that these observations were most prominent among the subgroup of dual-infected individuals that became HIV-1 infected after an established HIV-2 infection. Moreover, the HIV-1 genetic diversity was significantly lower in dual than in single-infected individuals at comparable time-points after infection. HIV-1 coreceptor tropism was investigated in late-stage disease by the use of a recombinant virus phenotypic assay that were confirmed to accurately predict the coreceptor tropism of HIV-1 subtype A and CRF02_AG. CXCR4 tropism has been coupled to an increased HIV-1 disease progression rate in late-stage disease. We found that HIV-1 CRF02_AG CXCR4 tropism was frequent (86%) and increased over time on the population level, indicating an evolving epidemic. In addition, a literature analysis showed a similar evolving epidemic for HIV-1 subtype C. Genotypic analysis suggested that the total number of charged amino acids could be important in predicting HIV-1 CRF02_AG coreceptor tropism. Finally, HIV-1 CXCR4-tropism was more common in single (79%) than in dual-infected individuals (35%). Understanding the underlying mechanisms responsible for the inhibitory effects exerted by HIV-2 against HIV-1 could be important for the development of future HIV-1 vaccines and therapeutics.

SAMMANFATTNING PÅ SVENSKA

HIV = Humant immunbristvirus
AIDS = Samlingsnamn på de sjukdomar som kan vara dödliga för personer med svagt immunförsvar (t.ex. svamp- och herpesinfektioner)

HIV är en av de dödligaste farsoterna i världen och 30-40 miljoner människor beräknas ha dött AIDS. Genom att konstant förändra sig ligger viruset hela tiden ett steg före immunförsvaret, och immunförsvaret klarar därför inte av att bekämpa infektionen. HIV skiljer sig också från andra virus genom att det drabbar och långsamt utplånar immunförsvaret. Det gör att HIV kan finnas i kroppen under lång tid innan den smittade insjuknar i AIDS – i många fall över tio år. Det finns två olika typer av HIV – typ 1 och 2. HIV-1 finns i hela världen, medan den mildare HIV-2 framför allt finns i Västafrika. HIV-2 förökar sig långsammare än HIV-1, vilket leder till AIDS-utveckling i endast 25-30% av fallen (99-100 % för HIV-1).

Genom att undersöka hur HIV-1 har förändrat sig över tid i Guinea-Bissau och Västafrika har vi kunnat studera när viruset kom in i Guinea-Bissau och hur det sedan spred sig inom landet. Vi fann att HIV-1 introducerades i Guinea-Bissau vid minst sex olika tillfällen under slutet av 1970- och början av 1980-talet. Dessa introduktioner sammanföll med en tid av oroligheter i Guinea-Bissau, kort efter det 11 år långa självständighetskriget mot Portugal (1963-1974). Migrationsanalys visade att HIV-1 introducerades i huvudstaden Bissau, och sedan spreds därifrån ut till mindre landsortsbyar.

Vi har också upptäckt att Västafrikanska patienter som är infekterade med HIV-1 och HIV-2 samtidigt (dubbelinfekterade) har en i genomsnitt 1,5 gånger så lång överlevnad och sjukdomsfri tid jämfört med patienter som enbart är infekterade med HIV-1 (enkelinfekterade). Genom detaljerade studier av immunförsvaret och virusets förmåga att förändra sig och bilda nya varianter har vi kunnat ringa in olika delar i sjukdomsförloppet som särskiljer enkel- och dubbelinfekterade patienter. Flera av våra resultat pekar på att viruset hämmas redan tidigt i infektionsförloppet, vilket leder till att immunförsvaret sedan bättre kan kontrollera sjukdomsutvecklingen. Vi fann också att de patienter som blev infekterade av HIV-2 innan de blev infekterade av HIV-1 var de patienter som klarade sig allra bäst. Detta tyder på att HIV-2 på något sätt förbereder immunförsvaret för en kommande HIV-1-infektion. Vi försöker nu hitta de exakta mekanismer som leder till en bättre sjukdomsprognos för dubbelinfekterade patienter, något som direkt skulle kunna användas i utveckling av nya

behandlingsmetoder eller framtida vaccin.

Slutligen så har vi också studerat skillnader i receptoranvändning för HIV-1 i enkel- respektive dubbelinfekterade patienter. HIV använder olika cellreceptorer för att sig in i celler och föröka sig. Tidigare studier har visat att virus som kan använda sig av kemokinreceptorn CXCR4, så kallade X4 virus, ofta uppkommer sent i sjukdomsförloppet. Uppkomst av X4 virus har kopplats ihop med en snabbare utveckling av AIDS. När vi undersökte skillnader mellan enkel- och dubbelinfekterade patienter sent i sjukdom, fann vi att HIV-1 från dubbelinfekterade patienter var mycket mindre benägna att använda CXCR4 (endast 35% av patienterna hade X4 virus) jämfört med enkelinfekterade patienter (79%). De underliggande mekanismerna för denna skillnad kan ha stor betydelse av flera anledningar, inte minst med bakgrund av de nyligen introducerade HIV-behandlingsmetoder som har visat sig vara effektiva mot de virusvarianter som oftast finns tidigt i sjukdomsförloppet. Dock är de verkningslösa mot X4 varianter.

ORIGIN, EPIDEMIOLOGY, BIOLOGY AND PATHOGENESIS OF THE HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 AND TYPE 2

INTRODUCTION

The Pandemic

The human immunodeficiency virus (HIV) is one of the major threats to the global health, with approximately 33 million people infected today. Approximately 30 to 40 million people have died in aquired immunodeficiency syndrome (AIDS) since the first description of the disease in the early 1980's (**Fig. 1**). Even though the number of people living with HIV still increases every year, the increase has slowed down in recent years and the number of new infections is decreasing. The worst affected part is the sub-Saharan Africa, with two thirds of the world's HIV-infected individuals (**Fig. 2**). In some countries, for example in Botswana and Swaziland, the prevalence levels have raised above 25% in the adult population, with major consequences on both the individual and the regional level. The major route of transmission is via heterosexual intercourse, but there are also other common transmission routes, such as mother-to-child transmission, via intravenous drug use and between men who have sex with men (MSM).

In Sweden, a total of 8,935 individuals had been reported HIV-positive by the end of 2009. Of these, 2,310 had been diagnosed with AIDS, and 2,045 deceased. Most registered HIV infected individuals were infected already at their arrival to Sweden. The major domestic transmission group during 2009 was MSM.

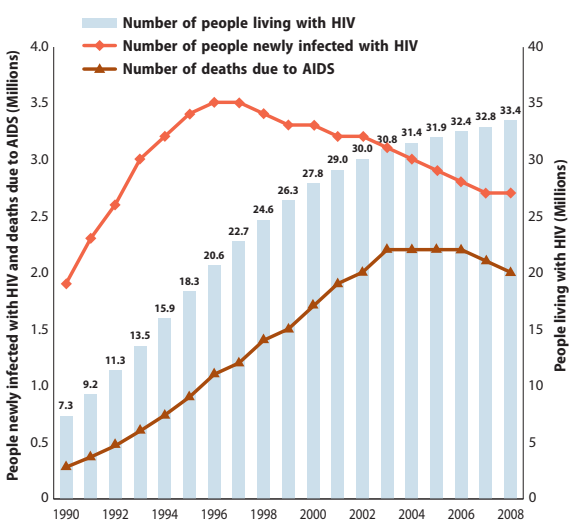


Figure 1. The HIV pandemic 1990-2008. Number of people living with HIV, number of people newly infected with HIV, and number of AIDS deaths worldwide, during 1990-2008. The figure was adopted with permission from the UN Millennium Development Report 2010[1].

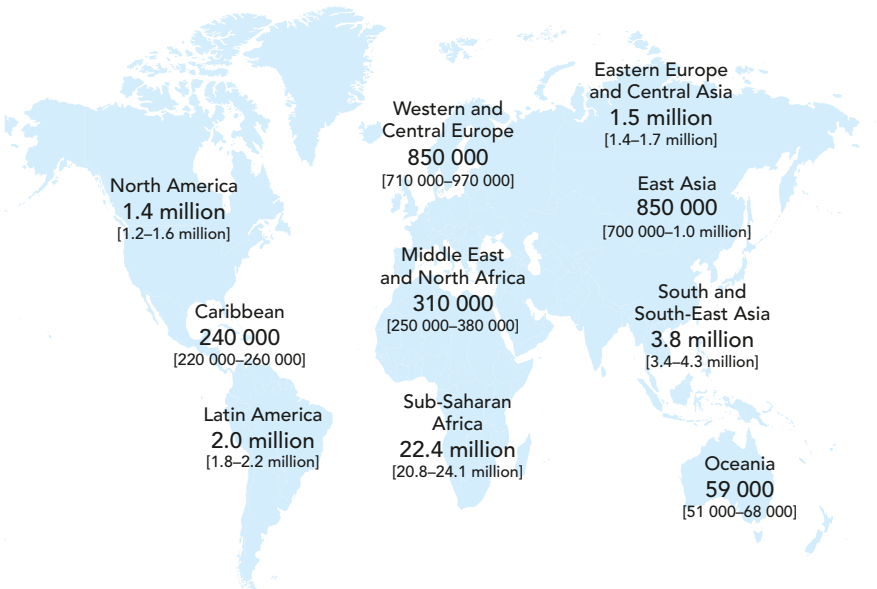


Figure 2. The HIV pandemic 2008 – regional estimates. Adults and children estimated to be living with HIV in 2008. The figure was adopted with permission from the Joint United Nations Programme on HIV/AIDS (UNAIDS) and World Health Organization (WHO) AIDS epidemic update 2009[2].

The discovery and history of HIV

In the late 1980 and early 1981, a number of American injection drug users and homosexual men without any known cause of immunodeficiency, sought health care due to symptoms of *Pneumocystis carinii* pneumonia (PCP)[3-5]. At the same time, or soon thereafter, reports of additional homosexual men with a rare skin cancer called Kaposi's sarcoma were presented[6-8]. Both these diseases represent rare opportunistic infections mainly linked to individuals with a severely compromised immune system. Shortly after these first reports, a number of similar cases were reported from several other parts of the world. The medical condition of these opportunistic symptoms was named acquired immune deficiency syndrome (AIDS). However, the cause of AIDS remained unknown until 1983, when a novel retrovirus was isolated from an AIDS patient with lymphadenopathy[9]. The virus was termed human immunodeficiency virus (HIV), after some initial suggestions such as lymphadenopathy-associated virus (LAV) and human T-lymphotropic virus type III (HTLV-III)[9-11]. In 2008, Luc Montagnier and Françoise Barré-Sinoussi were awarded the Nobel Prize in Physiology or Medicine for the discovery of HIV. The discovery of HIV was followed by a number

of important reports. In 1984, Robert C. Gallo and his collaborators published a series of significant articles that establishing HIV as the etiological agent of AIDS[12-15]. Shortly thereafter, in March 1985, the Food and Drug Administration (FDA) in the U.S. approved the first HIV antibody screening tests.

The increasing number of reports recognizing the medical condition of AIDS in humans also resulted in several other reports describing a resembling condition among captive monkeys[16,17]. In 1985, the etiological agent of this condition (termed simian AIDS) was isolated and termed simian immunodeficiency virus (SIV)[18]. The discovery of SIV also led to the discovery of a second type of human immunodeficiency virus in 1986, mainly circulating in West Africa[19]. Today, these two different human immunodeficiency viruses are termed HIV-1 (pandemic, discovered in 1983) and HIV-2 (endemic in West Africa, discovered in 1986).

Even though AIDS was not discovered as a clinical reality until late 1980, there have been several retrospective reports of earlier AIDS cases. The oldest case is thought to be the 25-year-old British painter David Carr. He served in the Royal Navy between 1955 and 1957, during which he reported to suffer from several rare symptoms, among them purplish skin lesions. After the military service his medical condition got worse, and he displayed classical AIDS symptoms such as shortness of breath, extreme fatigue, rapid weight loss, night sweats and high fever. He was unsuccessfully treated for tuberculosis and died in August 1959. The autopsy showed evidence of both PCP and cytomegalovirus infection[20,21]. Other early cases that have been described is (1) a documented HIV infection discovered in a preserved blood sample taken in 1959 from a man from Kinshasa, the Democratic Republic of the Congo; and (2) the identification of HIV in a preserved lymph node biopsy taken in 1960 from a women also from Kinshasa, the Democratic Republic of the Congo[22,23].

HIV genome and structure

HIV belongs to the genera *Lentiviridae* of the virus family *Retroviridae*. Lenti (Latin meaning slow) reflects the long incubation (disease-free) period seen among lentiviruses. One of the unique characteristics of lentiviruses is the ability to integrate their viral genome into the genome of non-dividing cells[24]. The family name *Retroviridae* comes from the ability of these viruses to perform a retrograde flow of information, meaning that they transcribes RNA into DNA instead of the regular direction of transcription of DNA into RNA (see the section *HIV replication cycle*).

HIV is a spherical enveloped virus of approximately 100 nm in size (**Fig. 3**)[24]. The

viral genome consists of two single-stranded RNA copies (ssRNA) that are embedded in a conical icosahedral capsid composed of the viral protein p24[25]. The viral RNAs are stabilized by the nucleocapsid protein p7. The capsid also encapsulates the viral enzymes reverse transcriptase, integrase, and protease, and some of the accessory proteins (**Table 1**). Surrounding the capsid is a matrix composed of the viral protein p17, ensuring the integrity of the viral particle[25]. The matrix structure is, in turn, enclosed by the viral envelope, which is composed of a phospholipidic bilayer derived from the infected host cell during budding (see the section *HIV replication cycle*). Early studies based on electron-microscopy pictures suggested that the outer membrane contains approximately 70-80 spiked knobs. However, recent biochemical studies and cryo-electron tomography reconstructions have estimated the number of spikes to be lower (range four to 35 spikes)[26-29]. The spikes are assembled as trimers of the viral glycoprotein (gp) 120, and bound to the transmembrane protein gp41[24,30-32]. In addition, several host cell proteins can be found within the viral envelope[24].

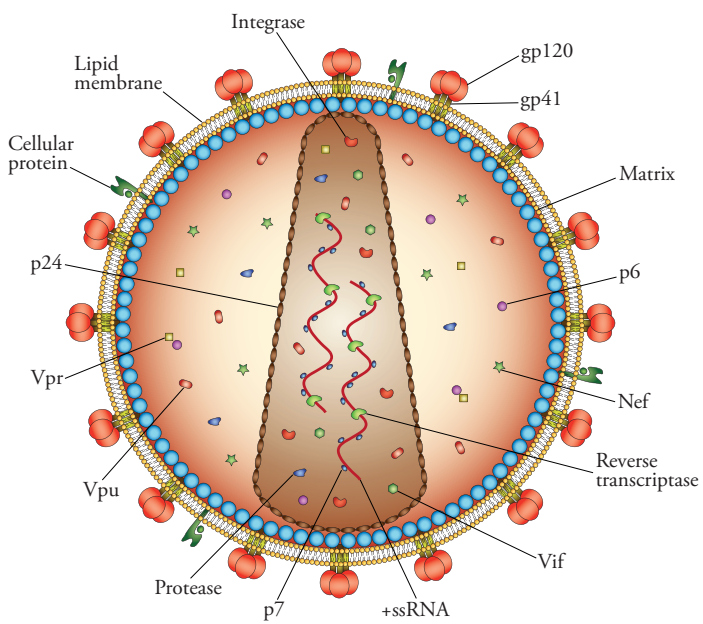


Figure 3. Structure of the HIV-1 virion. The viral envelope is formed from the host cell membrane, and the spherical form is maintained by the matrix structure. The core of the virion consists of the capsid and several different virus-specific proteins and enzymes.

Table 1. Functions of the different HIV-1 proteins.

Name	Abbreviation	Size (kDa)	Function
Structural proteins			
Group-specific antigen	Gag		
Matrix	MA	p17	Membrane anchoring, Env interaction, nuclear transport of viral core
Capsid	CA	p24	Core capsid
Nucleocapsid	NC	p7	Binds to and stabilizes RNA
Core protein		p6	Binds to Vpr
Polymerase	Pol		
Protease	PR	p15	Proteolytic cleavage of Gag and Pol
Reverse Transcriptase	RT	p66	Reverse transcription and RNAse H activity
Integrase	IN	p31	DNA provirus integration
Envelope	Env		
Glycoprotein 120	Gp120	gp120	Viral entry to the cell
Glycoprotein 41	Gp41	gp41	Transmembrane protein, cell fusion
Regulatory proteins			
Trans-activator of transcription	Tat	p14	Viral transcriptional transactivator for the LTR promoter
Regulator of expression of viral protein	Rev	p19	Regulates viral mRNA prduction, binds to RRE and facilitates nuclear export of unspliced or singly spliced RNA
Accessory proteins			
Virion infectivity protein	Vif	p23	Promotes virion maturation and infectivity
Viral protein R	Vpr	p18	Enhances viral replication in primary cells, promotes nuclear localization of preintegration complex, inhibits cell division, causes G2 cell cycle arrest
Viral protein U	Vpu	p16	Enhances virion release from cells, downregulates CD4 and MHC class I expression
Negativity regulatory factor	Nef	p27	Alters viral replication, downregulates CD4 and MHC I

The HIV genome is approximately 10,000 base pairs long and contains three open reading frames (**Fig. 4**)[25]. The major genes are the structural genes *gag*, *pol*, and *env*, but the genome also comprises regulatory (*tat* and *rev*) and accessory genes (*nef*, *vif*, *vpr*, and *vpu* (HIV-1) or vpx (HIV-2)). All these genes encode different viral proteins important during the replication cycle or the assembly of the virus particle (**Table 1**). The protein-coding genes are flanked at the ends by the viral long terminal repeats (LTRs), containing transcriptional regulatory elements, RNA processing signals, packaging sites, and integration sites.

HIV replication cycle

Viral entry

The major target cells of HIV are $CD4^+$ T cells and macrophages, and the entry starts by interaction of the viral trimeric envelope complex and the cellular receptor CD4 and a coreceptor (most often CCR5 or CXCR4, see section *Viral receptors and tropism*) (**Fig. 5**)[33-37].

The HIV-1 gp120 is approximately 500 amino acids (aa) long and consists of five constant (C1-C5) and five variable regions (V1-V5) of different lengths (**Fig. 6**). The

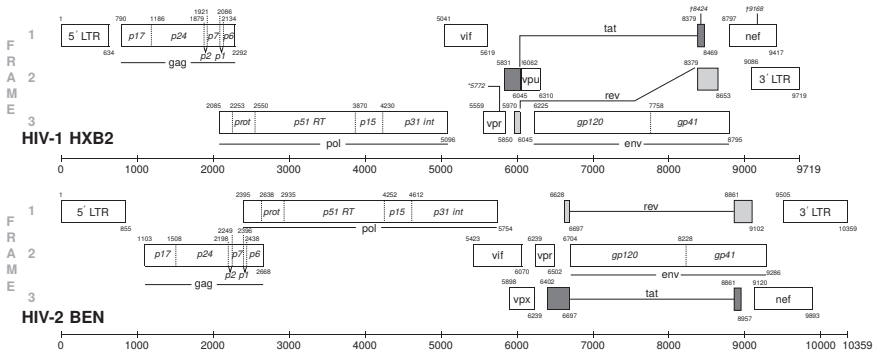


Figure 4. The HIV-1 and HIV-2 genomes in integrated state. Open reading frames are shown as rectangles. The gene start, indicated by the small number in the upper left corner of each rectangle normally records the position of the A in the ATG start codon for that gene, while the number in the lower right records the last position of the stop codon. The *tat* and *rev* spliced exons are shown as shaded rectangles. In HXB2, *5772 marks the position of a frameshift in the *vpr* gene caused by an "extra" T relative to most other subtype B viruses; !6062 indicates a defective ACG start codon in *vpu*; †8424 and †9168 mark premature stop codons in *tat* and *nef*. The figure was adopted with permission from[25].

CD4 binding domain of gp120 consists of several key epitopes that can be found in the V1-V2 stem, C2, C3, C4, and C5 regions[39]. The first interaction between gp120 and CD4 promotes a conformational change in gp120, which results in the exposure of the bridging sheet, composed of four anti-parallel β-strands from the V1-V2 and C4 regions (**Fig. 5 and 6**)[40]. The bridging sheet interacts with the coreceptor together with the V3 region, resulting in another conformational change. This second

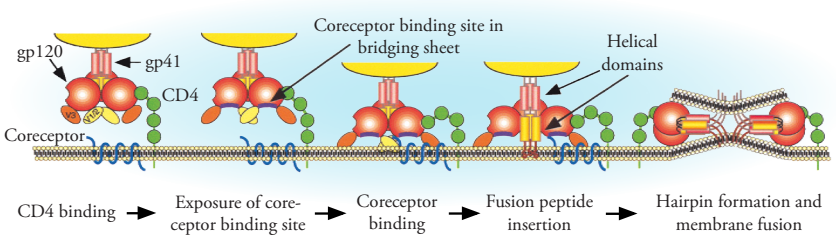


Figure 5. Schematic illustration of the HIV-1 entry. The entry process begins when the viral envelope and the target cell membrane get close enough for the viral gp120 to interact with the cellular receptor CD4. This induces a conformational change in gp120 resulting in exposure of the coreceptor binding site. Coreceptor binding induces a second conformational change, which allows insertion of the fusion peptide into the host cell membrane. The final step is a hairpin formation that brings the viral envelope and the cellular membrane close enough to fuse. The figure was adopted with permission from[38].

conformational change allows gp41 to insert a fusion peptide into the host cell membrane[41]. Finally, a six-helical bundle of gp41 is formed that brings the viral and cellular phospholipidic membranes close enough for fusion to occur, and the viral capsid is released into the cellular cytoplasm[42].

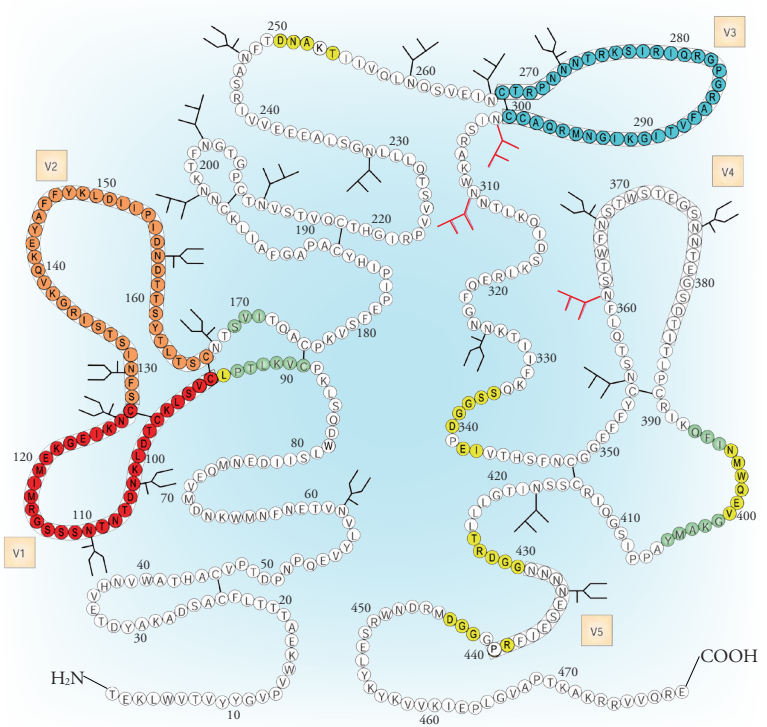


Figure 6. The HIV-1 gp120. Locations of the variable regions are marked with boxes (V1-V5). The variable regions studied in this doctoral dissertation are highlighted in colour (the V1 is red, the V2 is orange, and the V3 is blue). Key epitopes important for CD4 binding are highlighted in yellow, and the CD4-induced bridging sheet is marked in green. Glycosylation sites that contain high-mannose and/or hybrid-type oligosaccharide structures are indicated by branched structures, whereas complex-type oligosaccharide structures are indicated by U-shaped branches. Glycosylations that have shown the ability to induce neutralizing antibodies are highlighted in red. The figure was adopted with permission from[43].

Transcription and viral assembly

After fusion and delivery of the viral capsid into the cell cytoplasm, the viral enzyme reverse transcriptase starts to transcribe the viral single-stranded RNA into double-stranded DNA (dsDNA) (**Fig. 7**). This takes place simultaneously with the disassembly of the viral capsid. During the formation of dsDNA, a nucleoprotein complex, called the pre-integration complex (PIC), is formed. The PIC consists of the viral proteins Vpr, the matrix protein p17, the integrase, and the newly formed dsDNA. A nuclear localization signal within p17 directs the PIC into the cellular nucleus, where the provirus is integrated into the host cell genome[44]. The integration process is catalyzed by the integrase and it has been shown that the viral dsDNA preferentially integrates within active gene regions[45]. When the viral dsDNA is integrated into the host cell genome it is referred to as proviral DNA. HIV-1 lacks RNA polymerase II, and the virus is therefore dependent on the host cell transcription machinery after integration.

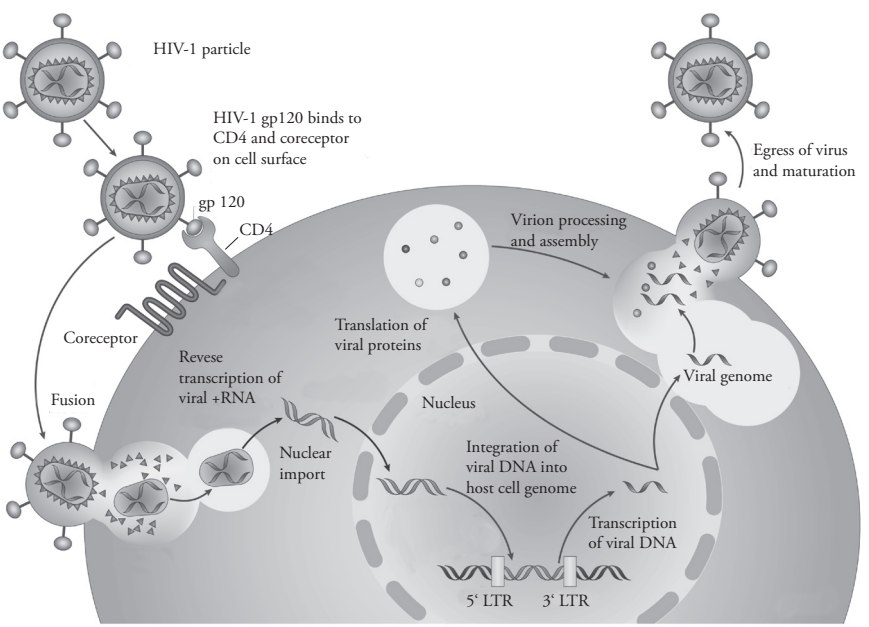


Figure 7. HIV-1 replication cycle. HIV-1 enters the target cell via interactions with the cellular receptor CD4 and a coreceptor. After fusion with the host cell membrane the viral RNA is reverse transcribed into proviral DNA that is transported into the cellular nucleus and integrated into the host cell genome. The integrated proviral DNA is then transcribed and translated into viral proteins by the cellular transcription and translation machinery. New viral particles are assembled at the host cell's outer membrane. After budding, the viral protease cleaves polyprotein precursors resulting in new mature and infectious viral particles. The figure was adopted with permission from[46].

The viral gene expression and particle production may take place either immediately or at some future time point. The timing presumably depends on the activity of the chromosomal locus hosting the provirus.

The HIV-1 gene expression is regulated by both viral and cellular factors. External factors, including coinfection with other agents, production of inflammatory cytokines, and cellular activation may trigger or enhance the viral replication[47]. Molecular mechanisms that can alter viral production include factors involved in different cellular pathways, such as the nuclear factor κB (NF-κB), that can bind to the enhancer region of the LTR[48,49]. A unique feature of HIV is that the expression of different viral RNAs is temporally regulated. Early after integration, activated cells produce double-spliced 2 kb mRNAs encoding for the viral proteins Tat, Rev and Nef[50,51]. The Tat protein induces an enhanced activity of the viral promoter by binding to the transactivation-responsive element (TAR) facilitating an increased RNA elongation[52]. The increased levels of these early and short transcripts are crucial for efficient production of late and longer transcripts. Accumulation of Rev results in a switch to enhanced expression of unspliced or singly spliced mRNAs encoding the different structural proteins of the viral particle. These proteins include Gag, Pol, Env, Vpu, Vpr and Vif, as well as genomic viral RNA. The delayed switch from early to late transcripts is thought to be a consequence of the requirement of a certain threshold concentration of Rev needed to interact with the Rev regulatory element (RRE) located in the incompletely spliced mRNAs[24].

The produced mRNAs are translated either cytosolic or membrane-bound by the common cellular translation pathways. The Env precursor protein gp160 is heavily glycosylated and cleaved by cellular proteinases into gp120 and gp41 in the endoplasmatic reticulum and Golgi apparatus. The different viral proteins are processed and transported to the host cell membrane for viral assembly, and the immature viral particle buds of from the cell. The final maturation step of proteolytic cleavage of the Gag-Pol polyprotein occurs after budding, and is performed by the viral protease. The mean generation time of HIV-1 has been estimated to 1.2-2.6 days[53-55].

HIV pathogenesis

One of the major characteristics of HIV is the continuous reduction of a functional host immune system, caused by an ongoing depletion of $CD4^+$ T cells. The natural course of an HIV infection is usually described by three stages: The acute stage, the asymptomatic stage, and the AIDS stage (**Fig. 8**).

The acute stage

The acute infection is characterized by viremia, rapid decrease in $CD4^+$ T cell counts (the main cellular target of HIV) and flu-like symptoms. This early stage of the HIV infection has been studied extensively during the last years, and the first weeks following HIV transmission have been shown to be extremely dynamic, both in

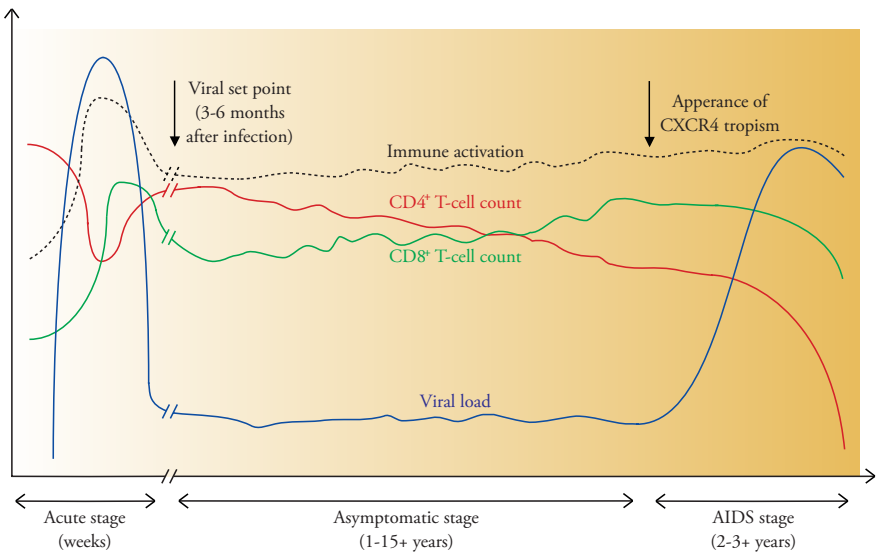


Figure 8. Schematic illustration of the general HIV disease course. The graph illustrates changes in levels of immune activation markers, $CD4^+$ T cells, $CD8^+$ T cells, and viral load over time during a typical HIV infection. Common fluctuations in duration of the different stages are given within brackets. Stabilization of the viral load after the acute stage is marked as the viral set point. Appearance of CXCR4 tropism is often seen in late-stage disease (see the section *"HIV receptors and tropism"*).

viral and immune levels (reviewed in[56]). Recently, the acute stage has been divided into several phases (**Fig. 9**). The time between infection and the first detection of viral RNA in blood plasma is referred to as the eclipse phase. Sequencing of the first detected virus variants during the eclipse phase suggests that approximately 80% of mucosally transmitted HIV-1 infections are initiated by a single virus variant[57-59]. It has been suggested that this founder virus establishes the infection at a single focus of infected mucosal $CD4^+CCR5^+$ T cells, and that early innate immune responses may support the recruitment of additional susceptible T cells to this focus[56,59,60]. A few days after this initial establishment virus particles and virus-infected cells reach the draining lymph nodes where further establishment of the infection is accomplished

by infection of activated $CD4^+CCR5^+$ T cells. The virus may also be transported to sites with high levels of $CD4^+$ T cells by dendritic cells (DC) and B cells[61,62]. HIV-1 gp120 can bind to C-type lectins, for example the DC-specific ICAM-3 grabbing non-integrin (DC-SIGN), on the surface of DCs. In addition, the virus often gets internalized by the DCs and can retain infectivity for up to four days[61]. The transfer of the virus from DCs to T-cells often occurs via the formation of virological synapses, which concentrate viruses and receptors at the site of contact between the two cells[63-65].

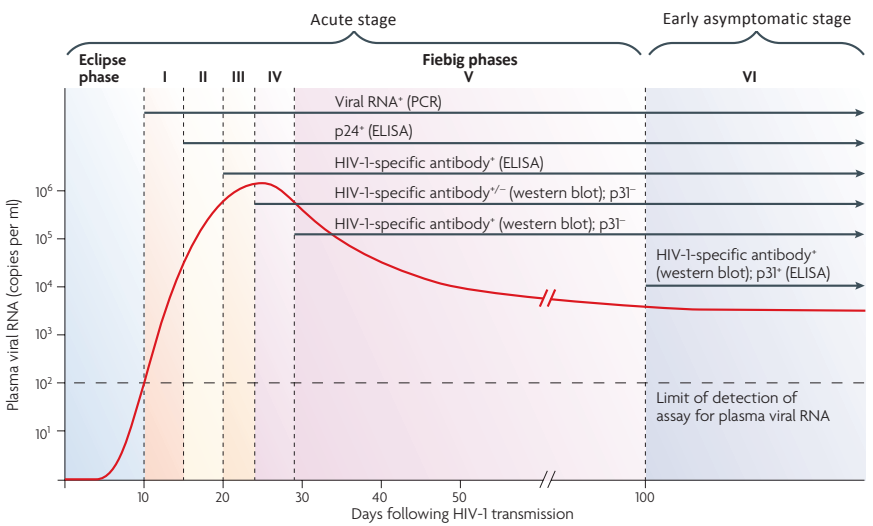


Figure 9. Schematic illustration of the different phases during the acute HIV-1 infection. Recently, different phases (Eclipse and I-VI) during the acute stage of infection were described based on clinical laboratory test results (different combinations of viral RNA measured by qPCR, p24 and p31 viral antigens measured by ELISA, and HIV-1-specific antibodies measured by ELISA or western blot)[66]. HIV-1 infected individuals progress through phase I to V during approximately three months after infection, and the plasma viral load usually stabilizes at a viral set point during phase VI three to six months after infection (see also **Fig. 8**). The figure was adopted with permission from[56].

The major site of viral infection during the acute stage is the gut-associated lymphoid tissue (GALT), where activated $CD4^+CCR5^+$ memory T cells are present in high numbers. Approximately 80% of all $CD4^+$ T cells in the GALT can be depleted during the first three weeks of infection (either by direct infection (20%) or by apoptosis (60%))[67-69]. The massive mucosal depletion of $CD4^+$ T cells is closely followed by the onset of a generalized immune activation, resulting in increased production of proinflammatory cytokines, increased frequency of T cells expressing activation

markers, and increased turnover of B and T lymphocytes, and NK cells[70-73]. In addition, immune activation has been associated with both T and B cell apoptosis, resulting in the release of apoptotic microparticles[74]. However, the causes of HIV-associated immune activation during the acute stage are still to be defined. The massive infection seen during the first part of the acute stage generally results in an exponential increase in plasma viremia, peaking 21-28 days after the infection at virus levels of up to several million copies per ml of blood[56,75]. As a consequence of the induced immune response against the virus, the viral load decreases towards a stabilized level, known as the viral set point[76-78] (**Fig. 8**). The time to reach this set point can vary between three to six months. The viral set point is maintained by the balance between the virus turnover and the host immune response and has been shown to be correlated with the risk of disease progression[79]. The level of decrease in $CD4^+$ T cell numbers during the acute stage has been correlated to the viral load in blood plasma, reaching the lowest levels at the time of peak viremia. When the viral load then starts to decline, $CD4^+$ T cell levels in blood return to near normal, whereas they remains low in the GALT[67-69].

The asymptomatic stage

After the initial and often very dynamic acute stage, the different levels of viral load, $CD4^+$ and $CD8^+$ T cell counts, and immune activation seem to stabilize. The pathogenic process enters a more steady stage, most often referred to as the asymptomatic or chronic stage. The length of this stage can vary considerably between different individuals (from less than one year to more than 20 years), and is marked by low plasma viral loads and a slow but gradual decrease in the number of $CD4^+$ T cells (**Fig. 8**). Early studies, investigating differences in T cell population levels over time, have shown an inverse correlation between $CD4^+$ and $CD8^+$ T cells, suggesting maintenance of a constant number of T lymphocytes throughout the HIV disease course[80]. The general immune activation seems to persist on a high level displaying a slow gradual increase during the asymptomatic stage.

The AIDS stage

The gradual loss in $CD4^+$ T cell levels ultimately results in exhaustion of the immune system, which therefore fails to control the HIV infection. When the patient enters this final stage, an increase in viral load is often seen, accompanied by a continuous decrease in $CD4^+$ T cell count. As a consequence of the severely impaired immune system, opportunistic infections such as candidiasis (*Candida albicans*), cytomegalovirus, severe herpes simplex outbreaks, Kaposi's sarcoma (human herpes virus type 8), pneumocystis pneumonia (*Pneumocystis jirovecii)*, and tuberculosis become increasingly common. The common name for these opportunistic infections

is AIDS, and ultimately also the cause of the patient's death.

Differences in disease progression rate

Even though the general characteristics of HIV pathogenesis are similar between most patients, the time frames of the different stages can differ a lot. Suggestions of how to group individuals with different rates of disease progression have been made (**Table 2**). Unfortunately, and somewhat confusingly, different research teams have used different criteria and different names for their classifications, and to date there is no common consensus. However, the identification of different progressor groups is of high relevance, and thorough studies dissecting the question of why some people progress faster to AIDS than others could give insights into mechanisms important for the development of novel therapeutic and vaccine approaches.

Immune responses to HIV

One of the hallmarks of HIV is the infection of $CD4^+$ T cells, a cell type that has a central role and is essential in a functional human immune system. The first line of

Table 2. HIV-1 progressor groups. HIV-infected individuals can be classified into different groups, based on their disease progression rate. The table was made as an effort of combining data and classifications from previous studies to form a set of different progressor groups[81-88]. Previously used terms are shown in the corresponding progressor group. If untreated, the general HIV-infected individual would progress to AIDS in approximately 3-10 years, corresponding to progressor group number II.

Progressor group	Characteristics
I	
Fast progressors	Viral load: >50,000 copies/ml
Rapid progressors	CD4 decline: Fast, >100 cells/ul per year
	Time to AIDS: <3 years
II	
Progressors	Viral load: >2,000 copies/ml
Chronic progressors	CD4 decline: Moderate, 30-99 cells/ul per year
	Time to AIDS: 3-10 years
III	
Slow progressors	Viral load: 50-2,000 copies/ml
Viremic controllers	CD4 decline: Slow, <29 cells/ul per year
	Time to AIDS: >10 years
IV	
Long-term non-progressors	Viral load: <50 copies/ml
Long-term survivors	CD4 decline: None
Elite controllers	Time to AIDS: >15-20 years without disease progression

defence against HIV infection is the innate immune response. This works in a non-specific manner and is followed by the more specific adaptive immune response that consists of a cellular and a humoral part.

The innate immune response

The innate immune system comprises several different types of cells and mechanisms, and is activated by non-specific binding of pathogens to sentinel cells via Toll-like and pattern recognition receptors[89,90]. The innate defence comprises several different cell types such as macrophages, dendritic cells (DC), natural killer cells (NK), and γδ-T cells. When these cells encounter a pathogen they immediately start to secrete potent molecules, and effector cells targeting the invader are activated. NK-cells and γδ-T cells kill infected cells through cytolytic mechanisms[91,92]. Macrophages and DCs activate the adaptive immune response (see below) by presenting antigens on major histocompatibility complex (MHC). Interferon (IFN) α and β are produced by HIV infected cells and induce an antiviral state resulting in apoptosis of these cells upon infection. In addition, IFNs stimulate the up-regulation of MHC I and consequently increased antigen presentation and activation of cytotoxic $CD8^+$ T lymphocytes (CTL) (see *The cellular immune response* below). An important part of the innate immune system is the complement system. The complement system is a biochemical cascade that helps, or "complements", the ability of antibodies to clear pathogens or mark them for destruction by other cells. Infected cells can also be rapidly killed through the formation of membrane attack complexes (MACs), causing osmotic lysis of the target cell. Moreover, the virus can get inactivated by complement-induced opsonization. Finally, cells with phagocyting capacity like neutrophils and macrophages can engulf HIV and clear foreign material.

The cellular immune response

As indicated by the name, the cellular immunity is associated with the cellular part of the immune system. Apart from the cells involved in the innate immune response, a more specific cellular immune response involves the activation of antigen-specific cytotoxic $CD8^+$ T lymphocytes (CTLs). These cells can to induce apoptosis in cells that displays epitopes of foreign antigen on their surface (indication of infection). CTLs are activated by $CD4^+$ T helper cells type 1 (Th1) and have shown to be crucial in the control of viral infection. Studies of SIV in rhesus macaques have reported that depletion of CTLs during acute infection resulted in loss of initial control of viral replication and marked increase in viremia during the asymptomatic stage of infection[93]. Viral load has also been inversely correlated with the level of CTL response[94]. In addition, rhesus macaques vaccinated to elicit virus-specific CTL

responses, and subsequently infected with SIV or SHIV (SIV with an HIV-1 envelope) had lower viral loads, higher amounts of $CD4^+$ T cell counts, and survived longer than unvaccinated monkeys[95-99]. $CD8^+$ T cells are not only important in direct killing of infected cells; they can also produce cytokines and chemokines that can affect viral replication, such as IFN-γ, TNF-α, MIP1-α, MIP-1β, and RANTES[100-105].

The humoral immune response

The term humoral comes from the fact that it involves substances or antigens that can be found in the humors (parts of the human body), or body fluid. The humoral immune response in mediated by secreted immunoglobulins, or antibodies that is produced by B lymphocytes. Activation of these cells is aided by an antigen-presenting cell and $CD4^+$ T helper cells type 2 (Th2). The secreted antibodies bind to viral antigens either directly or on the surface of infected cells, flagging them for destruction. Antibody responses have a central role in clearing most viral infections, and have been the major mechanism employed in the development of effective and protective vaccines. However, accumulating data have suggested that this might not be true for HIV-1 infection[106-109]. In contrast, studies of limited numbers of non-human primates have pointed towards a direction where antibody-mediated protection could be feasible[110-112]. Depletion of B lymphocytes in Rhesus macaques before exposure to SIV resulted in lower neutralizing antibody titers, higher viral load, and faster disease progression, demonstrating that neutralizing antibodies are important in controlling viral infection[113,114].

In HIV-1 infection, neutralizing antibodies are mainly directed towards gp120, preventing viral interaction with host cell receptors[115-118]. In addition, examples of non-neutralizing antibodies have also been presented, e.g. viral-cell fusion by interfering with gp41, activation of the complement system, and antibody-dependent cell mediated cytotoxicity[119-124]. However, the rapid evolution of HIV makes the development of persisting and effective neutralizing antibodies difficult. Further complication is the structure of gp120, where the conserved epitopes are hidden beneath variable regions and a dense glycan shield (**Fig. 6**). Neutralizing antibodies have also been suggested to be one of the driving forces of viral evolution[125,126].

HIV receptors and tropism

HIV receptors

The first receptor for HIV-1, the cluster of differentiation 4 (CD4), was identified as critical for HIV entry already in 1984[33,36,37]. Soon after this discovery, it was shown that CD4 was not enough to confer HIV-1 entry by itself. The first steps towards

identifying the second receptor important for HIV entry was the discovery of three chemokines of the CC-family, RANTES (CC-chemokine ligand 5 (CCL5)), MIP-1α (CCL3) and MIP-1β (CCL4), that showed a potent and specific natural inhibition of HIV-1 infection *in vitro*[105]. A few months later, in the spring of 1996, the chemokine receptor CXCR4 was identified as a critical coreceptor for HIV-1 entry[127]. These two findings triggered a chain reaction resulting in several important discoveries. Within a year, a second coreceptor for HIV-1 was identified, the chemokine receptor CCR5[128-131]. The specific chemokine ligand for CXCR4, the SDF-1 (CXCL12), was also identified[132,133]. A third important discovery during this short time-period was the association of a specific genetic variant, a 32-base pair deletion in the coding region of the CCR5-gene (CCR5-Δ32), with resistance to HIV-1 infection[134-137]. Since the discovery of CCR5 and CXCR4 as the major coreceptors for HIV-1, several other, minor coreceptors have been described (**Table 3**)[128,138-150]. When these minor coreceptors are used by the virus *in vitro*, it is almost exclusively in combination with one or both of the major coreceptors[151]. However, the *in vivo* role of the minor receptors remains to be clarified.

Chemokine receptors belong to a class of seven-membrane G-protein-coupled receptors involved in intracellular signalling, cell migration, and inflammation[152,153]. Major hallmarks of these receptors are (1) they measure approximately 350 amino acids (aa) in length, (2) the overall acidic N-terminal is located outside the cell, whereas the C-terminal that is coupled to the signal transducing G-protein is located on the inside, (3) seven α-helical transmembrane domains, with three extra- and three intracellular connecting loops composed of hydrophilic amino acids, are oriented perpendicularly to the plasma membrane, and (4) a disulfide bond links highly conserved cysteins in extracellular loops one and two. The chemokine receptor family can be further subdivided into four different groups (C, CC, CXC, and CX_3C) according to the positioning of the first two closely paired and conserved cysteins[154]. Furthermore, there are other examples of minor HIV-1 chemokine receptors, such as orphan chemokine receptors (whose endogenous ligands have yet to be identified or were identified much later than the receptor (adopted orphans)), and the viral chemokine receptor US28 (encoded by the human cytomegalovirus (CMV) and expressed on the surface of CMV-infected cells). There have also been reports of non-chemokine receptors with the ability to function as HIV-1 coreceptors *in vivo* (**Table 3**).

The small peptides (92-125 aa) that interact with chemokine receptors are called chemokines. The chemokines that bind to the HIV-1 major chemokine coreceptors are RANTES (Regulated on activation, normal T-cell expressed, and secreted), MIP-

Table 3. Details of the HIV-1 coreceptor repertoire.

Coreceptor	Alternative names	Ligand(s)	Cellular distribution[1]
Human chemokine receptors			
CCR2b	CD192	MCP-1, -2, -3, -4	M, T, B
CCR3	CD193	Eotaxin-1, -2, -3, RANTES, MIP-5, MCP-2, -3, -4, -5	T, DC, As
CCR5	ChemR13, CD195	RANTES, MIP-1α, -1β, MCP-2	M, T, DC
CCR8	TER1, CKR-L1, ChemR1, CDw198	I-309	M
CCR9	GPR-9-6, D6, CDw199	TECK	T
CXCR4	LESTR, Fusin, CD184	SDF-1α, -1β	M, T, B, DC
CX_3CR1	V28	Fractalkine	M, T, NK
CXCR6	BONZO, STRL33, CD186	CXCL16	T, NK
Human orphan chemokine receptors			
APJ		Apelin	T, As
ChemR23	CMKLR1	Chemerin	M, DC
gpr15	BOB	Not known	M, T
Other human receptors			
BLTR		Leukotriene B4	Ne
α4β7 integrin		MadCam-1	M, T, B, DC, NK
Viral chemokine receptors			
US28		RANTES, MIP-1α, -1β, MCP-1	T

[1]M, monocyte/macrophage; T, T lymphocyte; B, B-lymphocyte; DC, dendritic cell; NK, natural killer cell; As, Astrocytes; Ne, neutrophil

1α (Macrophage inflammatory protein 1α), and MIP-1β for CCR5, and SDF-1α (Stromal cell-derived factor 1α), and SDF-1β for CXCR4. These chemokines have been shown to have a suppressive effect on HIV-1 infection in different experimental systems[104,133,155].

Detailed studies of how HIV-1 interacts with CCR5 and CXCR4 have shown that different parts of the receptors are involved in binding to the virus. The N-terminal and the first extracellular loop have been suggested to be the most important domains involved in CCR5 binding[130,156]. However, it has been shown that HIV-1 can display an evolving pattern of CCR5 use, resulting in a broader use of the receptor during late-stage disease, including extracellular loops two and three[157]. The key binding

structures in CXCR4 are the first and second extracellular loops[158-160]. The binding of HIV-1 to any of the coreceptors results in a signal cascade affecting the gene expression of the infected cell. The major group of genes that is affected is involved in the cell cycle, and the cell is often forced into an active state[161]. This promotes transcription of HIV-1 genes and results in increased virus production. In addition, it has been demonstrated that different viruses with different phenotypes, binding to either CCR5 or CXCR4, can induce distinct gene expression patterns in peripheral blood mononuclear cells (PBMCs)[161].

HIV tropism and molecular determinants

In the late 1980's it was discovered that HIV-1 from different patients could display different kinetics in terms of replication in PBMCs. Consequently, the virus was determined to be either slow/low (S/L) for slow replicating viruses, or rapid/high (R/H) for fast replicating viruses[162,163]. Another difference was the ability of some virus isolates to induce syncytia (multinucleated giant cells) in PBMCs or in the human T-cell line MT-2[164-167]. It was soon suggested that viruses having the ability to induce syncytia (syncytium-inducing (SI)) almost exclusively were of R/H phenotype, whereas S/L viruses lacked that ability (non-syncytium-inducing (NSI)). In addition, HIV-1 of R/H or SI phenotype was associated with a faster decline in $CD4^+$ T cells, and a faster progression to AIDS[168-171]. When the coreceptors were discovered during the mid 1990's, it became evident that those viruses that were able to use CXCR4 as a coreceptor also were the viruses that displayed a R/H and SI phenotype connected to a faster disease progression. This resulted in a new nomenclature for HIV-1, based on coreceptor tropism, and viruses were designated to be either monotropic, dualtropic, or multitropic (**Table 4**)[172]. Viruses using CCR5 were termed R5 viruses, viruses using CXCR4 were termed X4 viruses and viruses able to utilize both receptors were termed R5X4.

HIV-1 coreceptor use is mainly determined by the V3 region of gp120[173-175]. The V3 region is a 35 aa long loop structure that is held together by a disulphide bond between the cysteins in V3 position 1 and 35 (**Fig. 6**). Only a few aa changes in the V3 region has been shown to be enough to alter the coreceptor tropism. The presence of a positively charged aa in position 11 and/or 25 has been particularly associated with the CXCR4-using phenotype[176,177]. Other properties related to charge or specific potential N-linked glycosylations have also been suggested[178,179]. In addition, predictive algorithms and methods based on different molecular characteristics in the V3 region important for coreceptor tropism have been presented[180-184]. However, these methods seem to be limited to specific HIV-1 subtypes, and a large amount of data is needed

Table 3. Details of the HIV-1 coreceptor repertoire.

Class	Phenotype	Tropism	Syncytia properties (MT-2, PBMC)	Replicative properties
Monotropic	R3	CCR3	NSI	S/L
	R5	CCR5	NSI	S/L
	X4	CXCR4	SI	R/H
Dualtropic	R3X4	CCR3 and CXCR4	SI	R/H
	R5X4	CCR5 and CXCR4	SI	R/H
Multitropic	R3R5X4	CCR3, CCR5 and CXCR4	SI	R/H

to get accurate predictive tools with both high specificity and sensitivity (**Paper IV** and[185-188]). Although the V3 region has been shown to be the principal determinant, other regions such as the V1-V2, V4 and C4 regions, could also influence the HIV-1 coreceptor tropism[189-194].

HIV coreceptor switch

Changes in HIV-1 coreceptor tropism over the disease course have been studied extensively, and it is well established that most individuals have CCR5 tropic viruses early in infection[162,165,166,168-171]. During the disease course, the virus population may switch to or broaden its coreceptor use to include CXCR4. In most cases this happens in late-stage disease, close to the onset of AIDS. The prevalence of HIV-1 CXCR4 tropism has been debated, and a common suggestion is that it appears in approximately 50% of the infected individuals[195,196]. However, this is a very rough estimate, based on the first reports investigating the dynamics of coreceptor tropism, and the approximate of 50% might be biased by either low numbers of study subjects or by a combined analysis of subjects from different disease stages[197-199]. In addition, the first reports were mainly performed on individuals infected with HIV-1 of the genetic subtype B (the dominating form in Europe and North America, see section *HIV variation and subtypes*). More recent studies have shown that the genetic subtype is important for the coreceptor evolution throughout the disease course[200-204]. Two examples are subtype C and subtype D. While subtype C has shown low numbers of CXCR4 tropism (0-30%) in patients in late-stage disease, subtype D has been reported to have the ability of using CXCR4 to a high extent already early in infection[200,202,203,205]. The underlying mechanism for these subtype-specific differences remains to be investigated.

Despite intensive efforts, the exact cellular and molecular mechanisms responsible for the coreceptor switch remain unclear, although several mechanisms have been presented (reviewed in[206]). The main hypotheses are considered to be (1) the transmission-mutation hypothesis, (2) the target-cell-based hypothesis, and (3) the immune-based hypothesis. Each of these hypotheses proposes a different view on the selection pressures acting on the viral population and the interaction of the virus with the immune system of the host.

The transmission-mutation hypothesis proposes that CCR5 tropism is fundamental for transmission, and that the emergence of CXCR4 tropism is a result of a gradual mutation of the transmitted CCR5-tropic founder strains. This hypothesis is supported by the fact that people that are homozygous for the Δ32-deletion in CCR5 are highly resistant to HIV-1 infection, and that infection with X4 viruses is extremely rare[135,207-209]. In addition, epithelial cells, important for mother-to-child-transmission and infections resulting from oral-genital contact, express mainly CCR5 and not CXCR4[210]. Moreover, dendritic cells that transport virus particles from mucosal tissues to lymph nodes have been shown to preferentially bind R5 viruses over X4 viruses[211,212]. Finally, SDF-1, the natural ligand for CXCR4, has been shown to be highly expressed in mucosal tissues, and it has been suggested that this could further limit the probability of X4 transmission[213]. It has also been shown that individuals infected via intravenous drug use or blood transmission also have R5 viruses early in infection[208]. Thus, any hypothesis trying to explain the predominance of R5 populations early in infection has to invoke mechanisms that are not dependent on the route of transmission. Furthermore, results of Cornelissen *et al.* and Pratt *et al.* showed that the virus populations in individuals infected with CXCR4-using viruses switched to R5 phenotype, suggesting a replicative advantage of R5 populations early in infection[214,215]. Further support for the transmission-mutation hypothesis came from a recent study by Clevestig *et al.* suggesting that the X4 subpopulations identified in two children had evolved from their own R5 subpopulations, and was not the cause of any transmission of X4 variants[216]. It might be argued that the transmission-mutation hypothesis is too simplistic due to the extremely high mutation-rate and viral turnover of HIV-1 in combination with the very few mutations needed for a coreceptor switch to occur (see section *HIV variation and subtypes*). In addition, mathematical modelling has suggested that X4 populations would emerge already three years after infection[75,76,206,217].

The immune-based hypothesis emphasizes that X4 viruses are more sensitive to a strong and well functioning immune system as compared to R5 viruses and, as a

consequence, are better recognized and suppressed early in infection. In late-stage disease, when the host immune system is severely dysfunctional, X4 viruses are allowed to emerge. Studies of neutralizing antibodies have shown that X4 populations are more sensitive to neutralization compared with R5 populations[218]. In addition, studies of the macaque model using SHIVs (a hybrid simian-human immunodeficiency virus) showed that dual-infections with R5 and X4 resulted in a dominance of R5 variants[99]. When monoclonal antibody depletion of the $CD8^+$ T cells was done in the macaques prior to infection, X4 variants became the dominating phenotype. Moreover, genetic analysis of the *env* gene has shown that CXCR4-using subpopulations are under higher selective pressure, perhaps by the immune system, than R5 subpopulations[219,220]. Another strong support for the immune-based hypothesis is the numerous reports connecting CXCR4 tropism with individuals in late-stage disease and with a severely impaired immune system.

Last, the target-cell-based hypothesis states that the viral phenotype is affected by the composition of the target cell pool. This hypothesis was presented by Davenport *et al.*, and was based on the differential expression of CCR5 and CXCR4 on naïve and memory $CD4^+$ T cells and differences in these cells population dynamics during the disease course[221]. The increasing proliferation rate of naïve $CD4^+$ T cells (expressing high levels of CXCR4) would result in a selection in favour of X4 viruses[222]. A mathematical model based on division rates of naïve and memory $CD4^+$ T cells and the ability of R5 and X4 viruses to bind to these cells support this hypothesis[223]. However, results contradicting the target-cell-based hypothesis have been presented. Memory $CD4^+$ T cells (mainly expressing CCR5) have been shown to increase in frequency relative to naïve $CD4^+$ T cells during the disease course[224-226]. In addition, increased immune activation over the disease course has been coupled to increased levels of CCR5 expression on $CD4^+$ T cells[227,228]. Moreover, van Rij *et al.* found that the relative frequency of CXCR4-expressing $CD4^+$ T cells was not associated with a switch in viral phenotype[229]. Instead, they found that early viral load and $CD4^+$ T cell count could be coupled to the emergence of CXCR4 tropism.

HIV variation and subtypes

One of the major characteristics of HIV-1 is the very high genetic variability and the ability to evolve at extremely high rates. This is primarily due to a combination of high viral turn-over, an error prone viral reverse transcriptase and frequent recombination. The composite half-life of HIV-1 in plasma and virus-producing cells are 1-2 days, and approximately 10^{10} new virions are generated every day, displaying the tremendous viral turnover[76,230]. The viral reverse transcriptase lacks proof-reading

capacity, and generates 1-4 substitutions per genome per round of replication[217]. Another feature of the viral reverse transcriptase is the ability of jumping across RNA strands, causing recombination (**Fig. 10**). Recombination is one of the major forces in HIV evolution, and it occurs at an estimated rate of 2-9 crossovers per genome per replication cycle[231-235].

To further illustrate the hallmark of HIV-1 genetic variability, it can be compared to another well-characterized and wide-spread virus, the influenza virus. An informative way of illustrating genetic variability is to construct phylogenetic trees based on

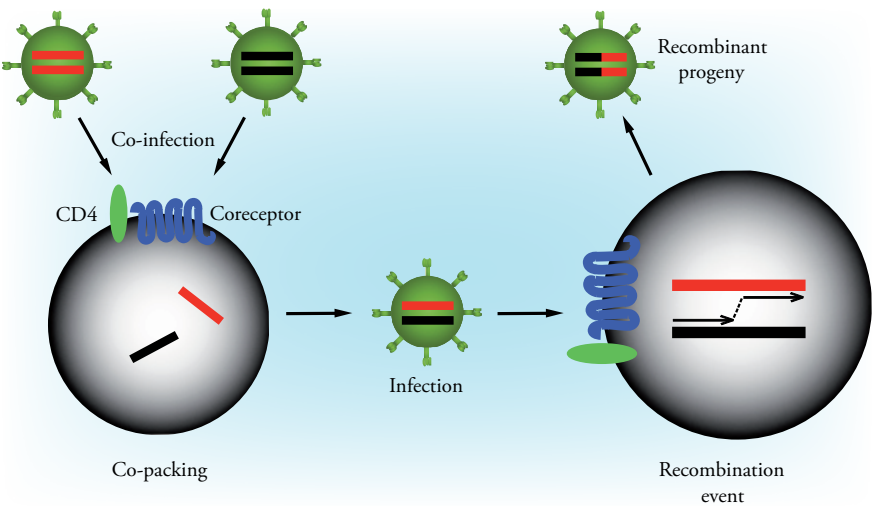


Figure 10. HIV-1 recombination. Two genetically distinct virus particles (red and black) infect the same target cell. Both viruses start to replicate within the cell, and during assembly one genome from each virus variant are co-packaged into the same budding particle. Second, this new particle infects a new target cell, and during reverse transcription the viral reverse transcriptase jumps between the RNA strands, resulting in recombination events. Consequently a recombinant progeny, with a mosaic genome from both the parental strains is generated.

sequence data. The influenza virus is thought of as a highly variable virus, and if we want to be protected against infection we need to take a new vaccine annually on every new season. In this perspective, it is interesting to note that the HIV population present in one single individual six years after infection can be as great as the global variation for an influenza outbreak (**Fig. 11**).

The high level of genetic variability has led to the diversification of HIV-1 into three

genetically distinct groups, the "major" group (M), the "non-M, non-O" group (N), and the "outlier" group (O) (**Fig. 12**). The far most prevalent group is group M, and both group N and O are tightly linked to Cameroon[238,239]. Recently, the existence of a fourth group, "pending the identification of further human cases" (P), was suggested, based on a virus isolated from a Cameroonian woman residing in France in 2009[240]. Group M has been further subdivided into different subtypes (A-D, F-H, J-K), subsubtypes (A1-A4, F1-F2) and 48 circulating recombinants forms (CRFs)

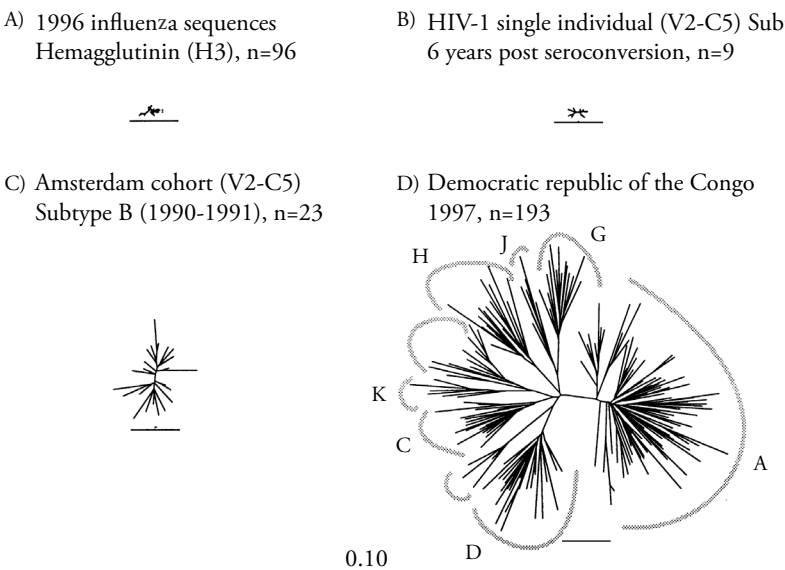


Figure 11. Genetic variation of influenza and HIV-1. Maximum likelihood trees using a general time-reversible model allowing for rate variation at different sites (see section *Phylogenetic inference*). A) Phylogenetic tree based on 96 hemagglutinin sequences encoding the HA1 domain of human influenza H3N2 viruses from 1996. B) Phylogenetic tree based on 9 HIV-1 subtype B C2-V5 *env* sequences from an infected individual 73 months after infection. C) Phylogenetic tree based on HIV-1 subtype B *env* C2-V5 sequences from 23 individuals residing in Amsterdam 1990-1991. D) Phylogenetic tree based on HIV-1 C2-V5 *env* sequences sampled from 193 individuals residing in the Democratic Republic of the Congo in 1977. The bar represents 0.10 nucleotide substitutions per site. The figure was adopted with permissions from[236,237].

(**Fig. 12**)[241]. Different subtypes can differ by up to 30% of aa in Env, and by 15% in Gag. The CRFs are recombinants between two or more subtypes that have been detected in more than two epidemiologically unlinked HIV-1 infected individuals, thereof the term "circulating". In addition, there are numerous of described unique

recombinant forms (URFs), again displaying the high frequency of HIV-1 variation. Recombinants between subtypes can appear during coinfection with two or more different subtypes.

Genetic subtypes of HIV are often confined to certain geographical locations, and the most common subtypes are subtype A (12.3% of the global prevalence), B (10.2%),

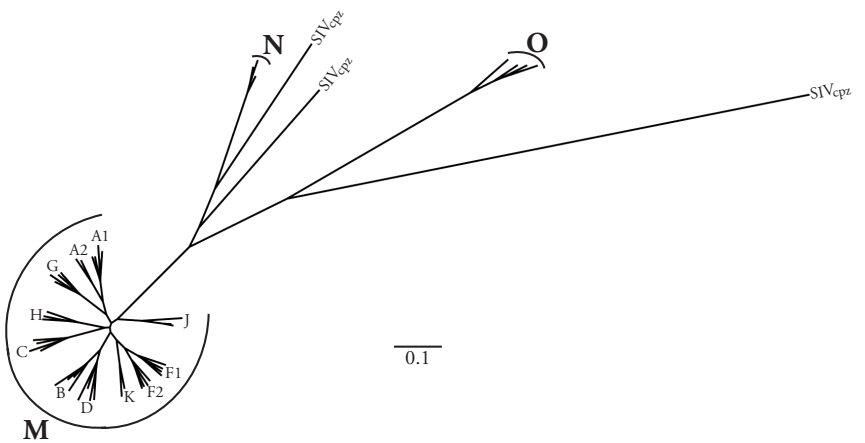


Figure 12. The major groups of HIV-1. Phylogenetic tree showing the genetic relationship between the groups M, N, and O. Scattered in between the groups are SIV sequences from the chimpanzee (*Pan troglodytes troglodytes*). The group M cluster is further subdivided into the major "pure" subtypes. The maximum likelihood phylogeny was constructed from full genome sequences, downloaded from the Los Alamos sequences data base, using the GTR+I+γ substitution model, as implemented in Garli[241,242]. The bar represents 0.1 nucleotide substitutions per site.

C (49.9%) and G (6.3%), and the CRF01_AE (4.7%) and CRF02_AG (4.8%) (**Fig. 13**)[243]. In Europe, North America, and Australia the dominating form is subtype B. The most prevalent form, subtype C, is largely confined to South and East Africa, and India, which are also the worst affected parts of the world. The most prevalent recombinant is the CRF02_AG, which is a recombinant of subtype A and G, and it has also been shown to be the most common form in West Africa. In addition, in most geographical regions very few different subtypes or CRFs are dominating the epidemics as illustrated in **Fig. 13**. An apparent exception to this is West Central Africa, displaying a wide variation in genetic subtypes[241]. It has been hypothesized that this could be a result of the origin of HIV-1 in this area (see section *Tracing the origin of HIV*)[22,23].

Similarly to HIV-1, the second human lentivirus (HIV-2) has also been divided into different groups. To date, there are eight distinct groups (A-F, and the recombinants AB and AU) of which only groups A and B are endemic[241,244]. HIV-2 is considered to be endemic in West Africa and India, even though cases of HIV-2 infection appear occasionally in other parts of the world. The vast majority of these infections are of HIV-2 group A, and group B are mainly found in Mali, Ghana, Burkina Faso and Cote d'Ivoire.

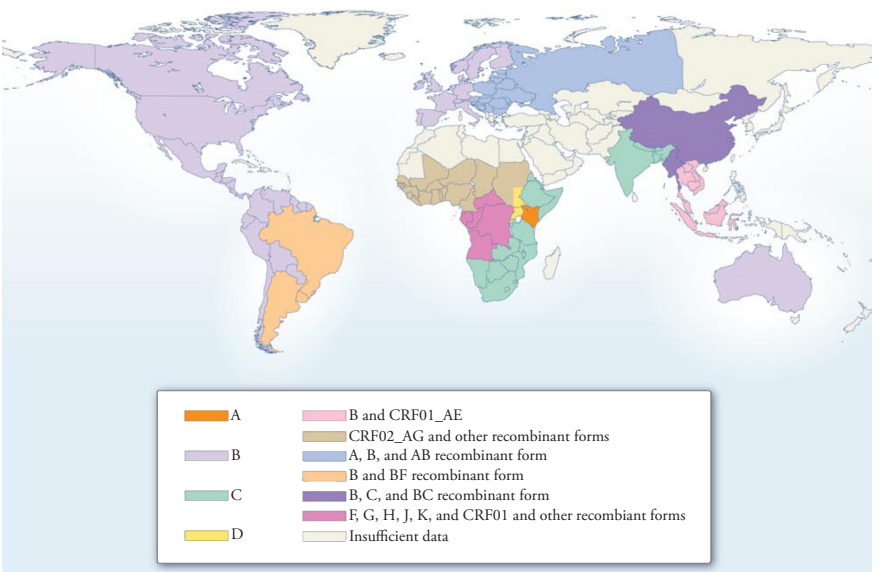


Figure 13. Current global distributions of HIV-1 subtypes and CRFs. The figure was adopted with permission from[243].

Differential characteristics of viral subtypes and CRFs and their interactions with the human host could influence HIV transmission and disease progression. Results from a Tanzanian study of 253 HIV-1 infected infants suggested that subtype C was transmitted more frequently *in utero* than subtype A or D[245]. Another study from Kenya of 365 women, indicated that women infected with subtype C were more likely of vaginal shedding of HIV-1 than women infected with HIV-1 subtype A or D[246]. The difference between subtype A and D was not statistically significant, but there was a strong trend for a higher likelihood of viral shedding of subtype A compared to subtype D. The combined conclusion of these two studies is that subtype C might

be more transmissible than subtype A and D. If it could be shown that subtype C is transmitted more frequently also in comparison with other subtypes, this could partly explain the extremely high numbers of infected individuals in sub-Saharan Africa, where subtype C is most prevalent.

Another important question is whether subtype differences result in variable rates of disease progression. There have been several, and somewhat contradictive, reports of HIV-related disease on cohorts infected with various subtypes[247-252]. Unfortunately, there are several caveats that may act as confounders in these studies. The most obvious and important caveat is the lack of information on seroconversion date, but there are also other potential confounders such as access to medical care, nutritional status, host genetics, and mode of transmission. However, in study by Baeten *et al.* seroincident subtype data from 145 Kenyan women with long follow-up were analyzed. The results showed that subtype D is associated with both a higher mortality and a faster rate of $CD4^+$ T cell count decline[253]. Interestingly, there were no differences in plasma viral load in relation to subtype. Another study, by Kiwanuka *et al.*, demonstrated that the median time to AIDS onset were shorter for individuals infected with subtype D or a recombinant between subtype A and D compared to individuals infected with subtype A[254]. No reports have presented any differences coupled to transmission or disease progression between different groups of HIV-2[255,256].

To understand the impact of HIV-1 subtypes on transmission, disease progression and global spread, some research teams have focused on detailed *in vitro* studies dissecting subtype-specific differences in viral properties. One such property is coreceptor tropism, as discussed in section *HIV receptors and tropism* and in **Paper IV**. Another property is replicative fitness, which has been studied in different *ex vivo* and *in vitro* systems by co-cultivating of different HIV types and subtypes. Arïen *et al.* compared the replicative fitness by coinfecting PBMCs, and found that HIV-1 group M viruses were more fit than HIV-2, which in turn were more fit than HIV-1 group O viruses[257]. Several studies have discussed the possibility that the epidemiologic dominance of subtype C might be related to a higher replicative capacity compared to other subtypes[258,259]. A recent study, utilizing a competition assay of primary HIV-1 subtype B and C isolates in PBMC, showed that all subtype C isolates were outcompeted by the subtype B isolates[260]. Moreover, according to a previously published review by Tebit *et al.*, subtype C is different from all other studied subtypes (A, B, C, and D) in terms of lower replicative fitness based on more than 2,000 competitions in PBMC[261]. In a population dynamics perspective, it is reasonable to speculate that recombination or other evolutionary bottlenecks would result in fitness advantages. If not, the

effect of such events would not be as distinct on the population level. In 2006, two independent studies presented results from replicative competition assays between the CRF02_AG and the putative parental strains, subtype A and G[262,263]. Both studies showed that the CRF02_AG was more replicative fit than subtype A and G. Another interesting observation in relation to replicative fitness was made by Troyer *et al.*, who found an inverse correlation between replicative fitness and disease stage in terms of $CD4^+$ T cell counts[264]. It has also been suggested that individuals with a slower disease progression rate typically are infected with less fit HIV-1 variants[261].

HIV-2

Even though HIV-2 is genetically related to HIV-1 (approximately 60% resemblance in Gag and Pol, and 30-40% resemblance in Env) and both viruses share similar routes of transmission and target cells, there are distinctive differences. First, it has been demonstrated that HIV-2 in general is less pathogenic, and only 20-25% of HIV-2 infected individuals develop AIDS (compared to a majority of HIV-1 infected individuals)[244]. This is reflected in decline in $CD4^+$ T cells, which is slower among HIV-2 infected individuals[265]. Second, both the viral set point and the viral load have been shown to be lower in HIV-2 infection[266-268]. In addition, most asymptomatic HIV-2 infected individuals do not display any detectable levels of viral load, whereas some degree of detectable viral levels most often can be found in HIV-1 infected individuals, even in LTNPs[267,269,270]. However, individuals with advanced disease display similar levels of viral load[266,268,271]. Third, the transmission efficiency is much lower for HIV-2 as compared to HIV-1[272]. In a study of vertical transmission of HIV during breast-feeding in the absence of antiretroviral therapy (ART), it was found that transmission of HIV-2 occurred in less than 4% of the cases[273]. This should be compared to a HIV-1 transmission rate of more than 24% in the same population. It has also been shown that shedding of HIV-2 in genital ulcers is significantly lower for HIV-2 than for HIV-1, and that the levels of HIV-2 in semen are closely related to plasma viral load. Furthermore, transmission of HIV-2 is closely related with elevated levels of viral load[274]. Finally, several studies investigating differences in coreceptor tropism between HIV-1 and HIV-2 have been presented, and the main coreceptors, CCR5 and CXCR4, are the same for both viruses[275-279]. It is, however, interesting to note that despite being less pathogenic, HIV-2 tends to be more promiscuous and use a larger number of coreceptors *in vitro* than HIV-1.

Most countries are experiencing a decline in HIV-2 prevalence while HIV-1 infection increases in the younger population[280,281]. The decline in HIV-2 prevalence is likely explained by the low transmission rate as described above. Due to the described

properties of HIV-2, it is remarkable that HIV-2 prevalences of up to 20% in where seen in some countries during the 1980's[282]. Possible suggested explanations are increased prostitution, access to blood transfusions or other iatrogenic events during and after the wave of struggles for independence in European-ruled African territories during the second half of the 20th century[283,284].

HIV-1 and HIV-2 dual-infection

Due to parallel epidemics of HIV-1 and HIV-2, populations in West Africa are at risk of getting infected by both viruses. The first reports of dual HIV-1 and HIV-2 infection came in 1988, and have been reported to be relatively common in West Africa with prevalence rates of 0-3.2% in the general population, and up to 38% in different risk groups[281,285-289].

In 1995, Travers *et al.* reported a possible protective effect of HIV-2 against subsequent HIV-1 infection in commercial sex workers in Senegal[290]. After this study, different West African studies unsuccessfully tried to verify this finding[291-295]. On the contrary, some of the studies even indicated that a previous HIV-2 infection could be a risk factor of subsequent HIV-1 infection, but it could also have been an effect of differences in high-risk sexual behaviour. Unfortunately, factors like sexual behaviour are extremely hard to adjust for. Since the collective result of all the different epidemiological studies in West Africa did not support any protective effect, another hypothesis was raised: Could HIV-2 inhibit the disease progression rate of HIV-1? Both early and more recent experimental studies using the macaque model have shown an inhibition against both immunosuppression and sAIDS as a result of a contemporaneous HIV-2 infection[296,297]. Reports addressing the question of inhibition against HIV-1 disease progression due to contemporaneous HIV-2 infection *in vivo* have also been presented[271,298-303]. None of them have been able to show such effect on neither differences in disease progression rate nor mortality between HIV-1 single and HIV-1 and HIV-2 dual infected individuals. Contrasting results regarding differences in viral load between single and dual-infected individuals have also been presented[266,271,302,304,305]. Unfortunately, all of these studies are limited by, and difficult to interpret due to unknown SC dates, differences in disease stage between groups, and short follow-up times (only three studies had follow-up times of more than one year, whereof the longest follow-up time for dual-infected individuals where 3.2 years). Moreover, most of these studies investigated symptomatic cohorts of hospitalized patients or clinic attendees, which may have skewed the conclusions.

The observation by Travers *et al.* also triggered several research teams to investigate

any potential mechanism behind the suggested protection that could be of relevance for vaccine development. *In vitro* studies have shown that HIV-2 infection generates higher levels of β-chemokines (the natural ligands of the HIV coreceptor CCR5) in peripheral blood mononuclear cells, and that this can inhibit HIV-1 infection and replication[306-309]. Zheng *et al.* investigated heterologous T-cell responses of individuals with single HIV-1 or HIV-2 infections, and found high frequencies of cross-reactivity between both patient groups[310]. They also found that HIV-1 single-infected individuals with the ability to respond to the HIV-2 Gag protein had lower HIV-1 plasma viral loads than those without this cross-reactivity. A second study by this group showed that this was evident also in dual-infected individuals[311]. In addition, there have been reports of cross-neutralizing antibodies against HIV-1 by HIV-2 sera[312-314]. However, the highest breadth of cross-neutralization was seen against the neutralization-sensitive laboratory strains SF-2 and NL4-3.

Most studies investigating the effect of coinfection of different pathogens and HIV-1 have shown an increased rate of disease progression[315-320]. However, there have been suggestions of attenuation of HIV-1 infection also by other viruses than HIV-2, such as GB virus C, measles virus, and the human T lymphotropic viruses type I and II (reviewed in[321]). As in the case of HIV-1 and HIV-2 dual-infection, the majority of these studies lack information of HIV-1 seroconversion date, and studies not supporting any *in vivo* inhibitory effect of these viruses have also been presented[322-324]. Follow-up studies dissecting virus interactions *in vitro*, have shown that these putative HIV-1 inhibiting viruses often alter cytokine and chemokine production that potentially could affect HIV-1 disease progression[325-328].

HIV in Guinea-Bissau

Guinea-Bissau is a small country (36,120 km^2) in West Africa, with approximately 1.7 million inhabitants[2]. The country is a former Portuguese colony, and became independent in 1974, after an 11 year long independence war. Recent reports on HIV prevalence in Guinea-Bissau have shown figures of 5-10% of the adult population (~5% HIV-1, ~4% HIV-2, and ~1.5% HIV-1 and HIV-2 dual infected individuals), depending on the investigated cohort[280,286,291,329]. During the 1980's, Guinea-Bissau had the highest reported numbers of HIV-2 in the world (8-10%). It has also been suggested that HIV-2 originated in Guinea-Bissau[282,330]. Studies investigating changes over time have shown an increasing trend in HIV-1 prevalence, whereas the opposite have been reported for HIV-2[280,286,329]. A programme for antiretroviral therapy (ART) was introduced into Guinea-Bissau in 2005, starting on a relatively small scale. Today approximately 3,000 individuals are on the ART programme (oral communication).

Phylogenetic inference

Phylogenetic inference is the study of evolutionary relationship among various groups of organisms, such as different species or populations, based on genetic information. The term *phylogenetics* is of Greek origin, and is a constitution of the terms *phyle* or *phylon* (tribe, race) and *genetikos* (relative to birth). Evolution can be thought of as a branching process, where populations are altered over time by (1) *speciation* (differentiation into separate branches and develop into novel species), (2) *hybridizing* (fusion into a novel population by mating of two former distinct populations), and (3) *extinction* (termination of a specific population). This kind of evolutionary process can be studied in a phylogenetic tree (**Fig. 12**). Typically, a bifurcating tree is used, where each sample (taxa) are positioned on the tips of the tree and every two taxa are joined together by a node representing a most recent common ancestor (MRCA) from which the taxa are estimated to have descended. The lengths of the horizontal branches represent the genetic relatedness between the different ancestors and their descendants in the tree (vertical branches have no function other than to separating the taxa in the phylogenetic illustration). Phylogenetic inference basically relies on (1) an alignment of genetic sequences, (2) an underlying substitution model for the evolutionary process in the given data set, (3) a tree building algorithm, and (4) some statistical support for the relationships given in the tree (**Fig. 14**).

First, the set of collected sequences under study need to be multiple aligned. This is the process of adding gaps to a matrix of data so that the nucleotides (or amino acids) in one column of the matrix are related to each other based on the assumption that they are descendants of a common ancestral residue. The gaps represent the combination of differences in deletions and insertions (indels) between sequences under study.

Second, an appropriate nucleotide substitution model for the dataset needs to be identified[331]. Over the years several more or less parameter-rich substitution models have been presented. All models belong to a general class of continuous-time Markov chains, which uses a Q matrix that specifies the relative rates of change of each type of nucleotide along the alignment. In addition, the relative frequencies of A, C, G, and T (π_A, π_C, π_G, π_T) are assumed to be at equilibrium. Some of these models can be extremely computational intense and with a certain amount of sequences the most parameter-rich models are not feasible. Current models are therefore simplified into so called reversible models, meaning that for example the rate of A to C is the same as the rate of C to A (**Fig. 15**). The simplest possible nucleotide substitution model was introduced by Jukes and Cantor in 1969 (JC69)[332]. The model specifies that the equilibrium frequencies of the four nucleotides are 25% each, and that all different

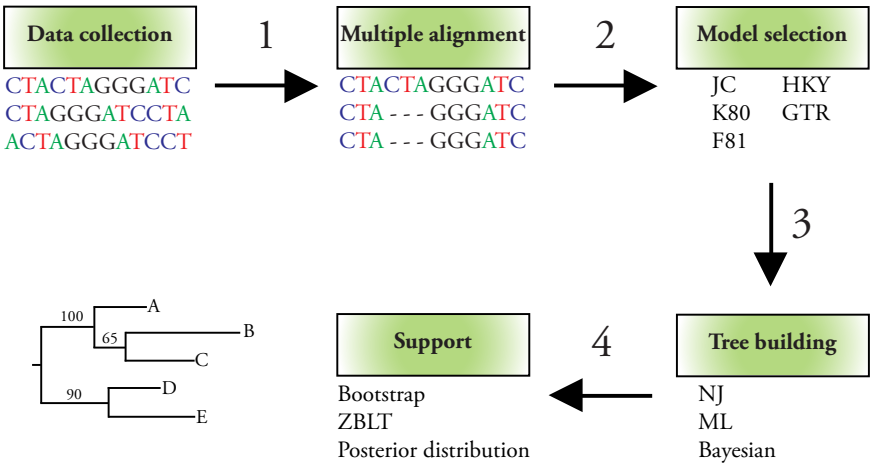
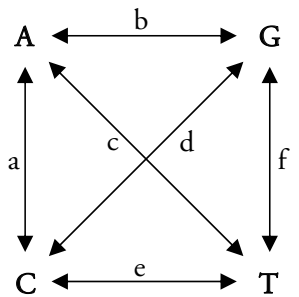


Figure 14. The phylogenetic inference process. The flowchart shows the major sequential steps in constructing a phylogenetic tree. The steps include (1) an alignment of genetic sequences, (2) an underlying substitution model for the evolutionary process in the given data set, (3) a tree building algorithm, and (4) some statistical support for the relationships given in the tree.

rates of nucleotide substitution are equal. These assumptions would correspond to a Q matrix with $\pi_A = \pi_C = \pi_G = \pi_T$ and a = b = c = d = e = f. Since the JC69 model was presented, a number of different models have been suggested, such as the K80, F81, HKY and GTR models[333-337]. The general time-reversible (GTR) model is the most parameter-rich model with six different substitution rates and specific frequencies of the four different nucleotides, resulting in a total of 10 different parameters. To assume that the relative rates of each type of nucleotide change are equally distributed over the alignment, as defined above, might be unrealistic and to simplistic. A better alternative is to assume a distribution of rates, based on the predefined relative rate, and that each site in the alignment has a rate drawn from that distribution. A gamma-distribution has been shown to be suitable for estimating rate variation in phylogenetic analysis. In addition to gamma-distributed rate variation, it might also be realistic to assume that there are a proportion of sites that are invariant and have zero rate of change. This parameter is often used in combination with the gamma distribution.

The third step is the selection of tree reconstruction method. A very fast and common traditional approach of reconstructing phylogenies is the neighbor-joining (NJ) method[338]. This method performs well when sequences in a dataset are similar. This method works by converting the sequences into a distance matrix that represents an

estimate of the evolutionary distance between all of the sequences in the alignment. When the distance matrix is built, the algorithm finds the pair of taxa with the highest similarity and connects them to a MRCA (a node connecting these two taxa).



$$Q = \begin{pmatrix} -\mu(a\pi_C+b\pi_G+c\pi_T) & a\mu\pi_C & b\mu\pi_G & c\mu\pi_T \\ a\mu\pi_A & -\mu(a\pi_A+d\pi_G+e\pi_T) & d\mu\pi_G & e\mu\pi_T \\ b\mu\pi_A & d\mu\pi_C & -\mu(b\pi_A+d\pi_C+f\pi_T) & f\mu\pi_T \\ c\mu\pi_A & e\mu\pi_C & f\mu\pi_G & -\mu(c\pi_A+e\pi_C+f\pi_G) \end{pmatrix}$$

Figure 15. The Q matrix of the general time-reversible (GTR) substitution model. The GTR model consists of six different substitution rates and specific frequencies of the four different nucleotides, resulting in a total of 10 different parameters.

Next, the distance to the MRCA for each of these two taxa is calculated based on the underlying substitution model. Then the distance from the MRCA to all the other taxa in the alignment is calculated. This results in a new distance matrix where the original two closest related taxa have been replaced with their MRCA. This iterative process is repeated until all taxa have been included in the tree. The NJ method is based on the minimum evolution criterion, meaning that the topology that gives the least total branch length is preferred. A potentially serious weakness with the NJ method is that it constructs the tree in a step-wise fashion and might therefore not find the tree topology with the least total branch length. A more appealing, but also much more computer-intense, way of reconstructing a phylogenetic tree is by the maximum-likelihood (ML) method. The ML method uses standard differential-equation techniques for inferring probability distributions based on the selected substitution model, to assign probabilities to different constructed phylogenetic trees. The number of possible trees increases extremely fast by the number of taxa, and with more than 10 taxa it quickly becomes unfeasible to investigate all possible trees

(exhaustive search) (**Table 5**).

All these trees can be thought of as a landscape of probability scores (based on each tree's probability score and the huge amount of trees with only a few numbers of taxa), where the trees with the highest probability constitutes the top of the hills

Table 5. The number of possible rooted, bifurcating, trees for different numbers of taxa.

Taxa	Number of trees
1	1
2	1
3	3
4	15
5	105
6	945
7	10,395
8	135,135
9	2,027,025
10	34,459,425
11	654,729,075
12	13,749,310,575
13	316,234,143,225
14	7,905,853,580,625
15	213,458,046,676,875
16	6,190,283,353,629,375
17	191,898,783,962,510,625
18	6,332,659,870,762,850,625
19	221,643,095,476,699,771,875
20	8,200,794,532,637,891,559,375
30	4.9518×10^{38}
40	1.00985×10^{57}
50	2.75292×10^{76}

(**Fig. 16**). In such a landscape there will be both local minima and local maxima, but also a global maximum, where the "best" tree can be found. Different heuristic tree searching methods have been suggested to search the tree landscape to find the global maximum as fast as possible[339]. These are known as branch-swapping methods, and involve cutting off one or more pieces of a tree (subtrees) and resembling them in a way that is locally different from the original tree. The most common branch-swapping methods are the nearest-neighbor interchange (NNI), the subtree pruning and regrafting (SPR), and the tree bisection and reconnection (TBR).

Fourth, a weakness with both the NJ and ML approach is that they only produce point estimates of the phylogeny, without any statistical measurement for the support of the relationships depicted in the tree. Traditionally, this has been solved by the

use of a statistical technique called bootstrapping[340]. In bootstrapping, positions in the original alignment are randomly resampled with replacement to produce a

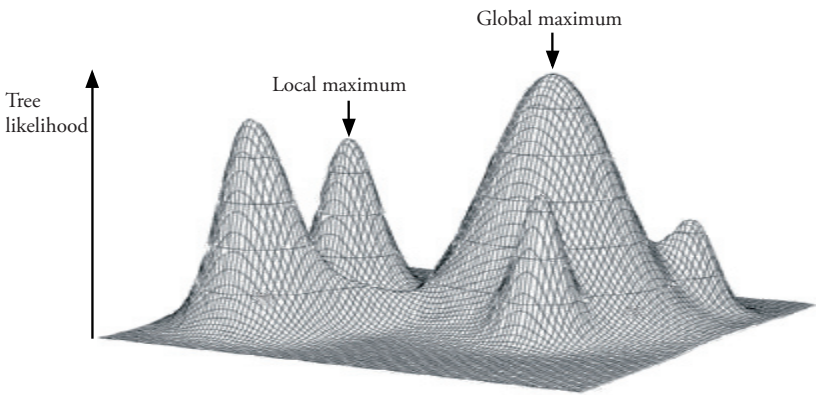


Figure 16. Phylogenetic tree landscape. The figure describes a possible tree landscape with several local minima and maxima, and a global maximum where the most probable tree will be found.

set of pseudo-replicate alignments. The tree building algorithm is used on each of these alignments. Clusters of related taxa that are present in a low percentage of the bootstrap trees are weakly supported and vice versa. However, the exact interpretation of the bootstrap proportion is difficult. Higher values are of course better, but what is a reasonable cut-off? It has been suggested that bootstrap values >70% might indicate strong support for a group, based on the conclusion that bootstrap proportions are conservative measurements[341]. Another statistical test that has been used in ML phylogenies is the Zero-branch length test (ZBLT)[342]. This test uses a likelihood-ratio test to evaluate if a specific branch is significantly longer than zero.

Another appealing and powerful method to reconstruct phylogenies is Bayesian inference. Bayesian phylogeny and ML phylogeny is related to each other, but the underlying statistical question is very different. In ML statistics the question is "What is the probability of seeing the observed data given that a certain model is true?", whereas the question in Bayesian statistics is "What is the probability that this model is correct, given the observed data?". The latter question can be answered by Bayes theorem:

$$Pr(T|D) = \frac{Pr(T) \times Pr(D|T)}{Pr(D)}$$

Pr(T|D) is the the conditional probability that is assigned after the relevant evidence is taken into account (referred to as the posterior probability)
Pr(T) is the prior probability, assigned to the theory T before the data was collected.
Pr(D|T) is the likelihood.
Pr(D) is the unconditional probability of observing data D disregarding which theory is correct.

The fundamental difference between ML and Bayesian statistics is highlighted in the following example:

In a box there are 90 fair and 10 biased dice. The biased dice are biased in the following way:
I – 1/21, II – 2/21, III – 3/21, IV – 4/21, V – 5/21, VI – 6/21. If a die is drawn from the box, what is the probability that the die is biased? Assume that we roll the die twice and get the results IV and VI. Based on the outcome, what do we think of the die?

Using a ML approach we get: Pr(fair) = 1/6 x 1/6 = 1/36, and Pr(biased) = 4/21 x 6/21 = 24/441. Since Pr(biased) > Pr(fair) we draw the conclusion that the die is most likely to be biased.

Using the Bayesian approach we get: Pr(T) = 0.1 (the prior probability is that 10% of the dice are biased), Pr(D|T) = 24/441 (the likelihood that the dice are biased based on the results), Pr(D) = 24/441 x 0.1 + 1/36 x 0.9 = 0.03 (the unconditional probability of the observed data). Pr(T|D) = 0.1 x 24/441 / 0.03 = 0.179 (the probability that the die is biased).

By repeatedly rolling the die we would get more information to draw increasingly certain conclusions by both the ML and Bayesian approach. In the Bayesian approach a new prior probability will be calculated after each new roll and used in the subsequent calculation (chain of calculations). After a while one would reach convergence, a phase where the posterior probabilities will be distributed around a certain mean probability. If the first probability values that were calculated and used as guide probabilities to reach convergence are disregarded, the mean posterior probability most likely is a good estimate of the status (fair or biased) of the drawn die. After this example it is

also easy to understand the importance of choosing accurate priors. However, priors should be determined with caution, a strong prior that is set based on inaccurate information, can severely bias the distribution of the posterior probability.

In Bayesian phylogenetics, a Markov Chain Monte Carlo (MCMC) approach is used to compute the posterior probability density. The inference is based on the selected substitution model and prior distributions of the parameters in the model. The Markov chain is designed according to the Metropolis-Hastings algorithm as follows: (1) start with a random tree and parameters, (2) in each generation propose either a new tree or a new value for a model parameter, (3) if the proposed tree/parameter has a higher posterior probability than the current tree the transition is accepted, if the probability is lower, the transition is accepted with some probability (this allows the algorithm to reach a higher probability in the tree landscape via a lower probability score (**Fig. 16**), and thereby limiting the risk of getting stuck on local maxima), (4) posterior trees are regularly saved during the Markov chain, and (5) after a number generations (typically millions of steps) the Markov chain is ended and the high number of saved trees are used for subsequent phylogenetic analysis and confidence estimates. After a while the chain tends to stay in regions with high posterior probability. To make the Bayesian inference even more efficient and also limiting the risk of getting stuck in local maxima, several Markov chains can be run in parallel and with different probabilities of transition to lower probabilities. In addition, by allowing the chains to communicate and swap position with each other, the chance of getting the more conservative chains to the global maximum is highly increased.

Finally, the Bayesian approach of estimating phylogenetic inference using the MCMC algorithm has made it possible to combine Bayesian inference with coalescence theory and different trait characteristics (such as geographic or ethnic information)[343-345]. This has resulted in a large amount of novel tools for investigating detailed spatio-temporal evolution of different fast evolving viruses, such as HIV, Influenza, Rabies and Hepatitis. The methods can of course also be applied to other, more slowly evolving organisms. However, this often requires detailed fossil data.

Tracing the origin of HIV

Current evidence indicates that HIV is the result of several cross-species transmission events (zoonosis) of SIV from African primates to humans in West Central Africa[255,346-352]. Phylogenetic analysis have indicated that viruses of HIV-1 group M, N, and O is related to SIV from chimpanzees (SIV_{cpz}), whereas HIV-2 is more closely related to SIV from sooty mangabeys (SIV_{sm}) (**Fig. 17**). Recently, a fourth group of

HIV-1, group P, was suggested[240]. Analysis showed that this strain was most related to SIV from gorilla (SIV_{gor}). The most likely route of transmission of HIV-1 and HIV-2 to humans involves contact with blood of different primates hunted for bushmeat in Africa[353].

Due to the high level of mutation rate among different lentiviruses, phylogenetic inference in combination with coalescence theory has proven to be useful in the study of the evolutionary history of these viruses. Screening of different African primates has revealed that at least 33 different species can be infected by different forms of SIV[354]. The time of most recent common ancestor (tMRCA) has been estimated to 1492 (95% highest posterior density (HPD) 1266-1685) for SIV_{cpz} and to 1809 (1729-1875) for the SIV_{sm} [355]. In 2008, Gifford *et al.* discovered that the Gary Mouse Lemur from Madagascar has an endogenous lentivirus in the genome (**Fig. 17**)[356]. Since terrestrial mammal populations in Madagascar and Africa most likely were isolated from each other more than 14 million years ago, this indicates that lentiviruses must have been infecting primates for a very long time. The long-standing presence of SIV in primates was further substantiated in a recent study where 22 new sequences of SIV were presented from four different species of monkeys from the Bioko Island, Republic of Equatorial Guinea[357]. Bioko Island was isolated from the mainland by rising sea levels about 11,000 years ago, and phylogenetic analysis determined the tMRCA of SIV to have arisen between 32,000 and 133,000 years before the present.

Studies investigating the origins of HIV-1 have determined the tMRCAs of group M to around 1900 (combined interval 1857-1939), group N to 1963 (1948-1977), and group O to around 1910 (combined interval 1866-1940)[22,355,358,359]. Furthermore, detailed studies investigating the time of origin of different HIV-1 group M subtypes have also been presented. Abecasis *et al.* estimated the tMRCA of subsubtype A1 to 1955 (95% HPD 1940-1969), subtype B to 1959 (1951-1967), subtype C to 1952 (1940-1962), subtype D to 1944 (1935-1952), subtype G to 1969 (1960-1977), the CRF01_AE to 1976 (1971-1981), and the CRF02_AG to 1975 (1969-1980), based on Bayesian estimates of the *env* gene[360]. Finally, phylogenetic analysis investigating HIV-2, have found that subtype A arose around 1935 (combined interval 1906-1956), and that subtype B most likely arose within the same period of time (1907-1961)[330,355].

Even though is well established that HIV is a consequence of a zoonosis between primate and human, alternative hypotheses regarding the origin of HIV have been presented. Most of them concern different versions of hypotheses that HIV is manmade

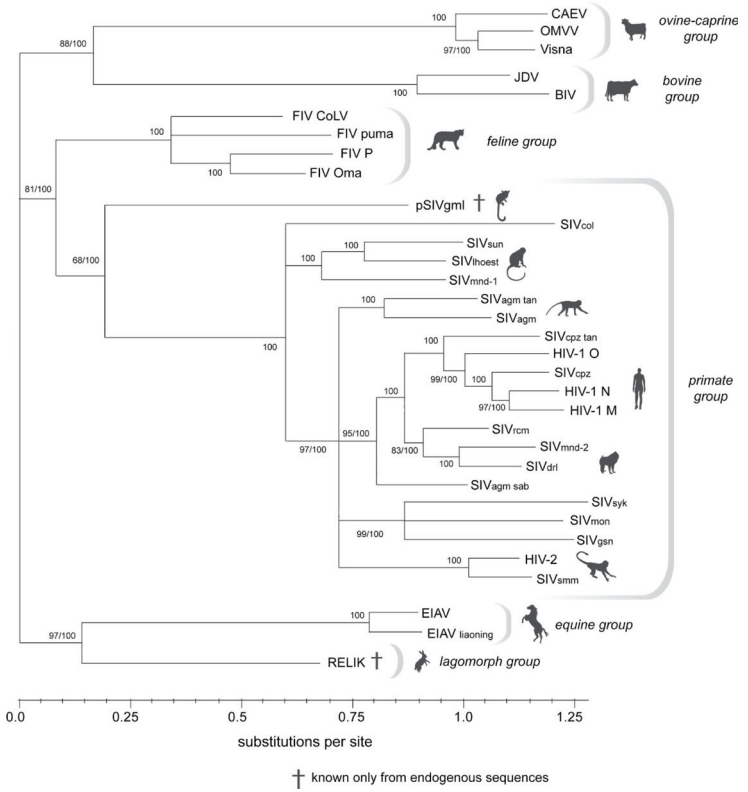


Figure 17. Phylogenetic relationships between different mammal lentiviruses. The phylogeny was based on an alignment of 853 amino acids, spanning conserved regions of Gag and Pol. Bootstrap support and Bayesian posterior probabilities are indicated to the left and right of the forward slash respectively, while nodes with maximum support with both measures are indicated with 100 only. The figure was adopted with permission from[356].

and the result of inadvertent human experiments in the development of vaccines, or conspiracy theories stating that HIV was developed by scientists working for the U.S. government trying to develop biological weapons. It has also been suggested that HIV is not the cause of AIDS, and that AIDS is caused by non-infectious agents like illegal drug-use or antiviral therapy. Unfortunately, some of these conspiracies have hade a large negative impact on the HIV epidemiology and the possibility of getting antiviral treatment in some countries. It is important to emphasize that these hypotheses and conspiracy theories have been rejected by a substantial amount of scientific evidence.

HIV-1 evolution

One of the most significant consequences of pathogens with high mutation rate and turn-over is their ability to quickly adapt to their environment. The adaptive potential of such pathogen can be measured in several ways (genetic diversity, divergence or selective pressure). By combining those with information of for example host immunological responses or any ongoing therapy; information on parameters altering disease progression or evolutionary processes of the pathogen may be extracted. There have been several different suggestions on how to measure different genetic evolutionary processes in a phylogenetic framework (reviewed in[361,362]). At this point it might also be appropriate to introduce the distinction between mutation rate and substitution rate, two commonly confused terms. The mutation rate is the rate at which mutational error are incorporated into a genome during replication. This rate is mainly determined by how error-prone the polymerase used for pathogen replication is. In contrast, the substitution rate is a property of the pathogen depending on many factors, and it is determined as the rate at which new mutations spread and become fixed in the population as a result of random drift and natural selection. For HIV, early studies of evolution focused on investigating the viral diversity (the genetic variation at a given time-point) and divergence (the genetic distance to a reference point, i.e. the founder strain). More recently, detailed studies of processes governing HIV evolution, such as differences in synonymous (dS) and non-synonymous (dN) evolutionary rates has been presented[363,364].

Intra-patient evolution

The transmission of HIV is associated with a major bottleneck, and in most cases, new infections are the result of the transmission of one single virus[58]. Consequently, after transmission, and early in infection, the viral diversity is very low[365]. Moreover, studies have indicated that HIV seems to be under positive selection during the acute stage of infection[366,367]. This has been suggested to be driven by escape from early cytotoxic T lymphocyte responses. It has also been suggested that HIV-1 *env* evolves towards ancestral states during transmission[368]. Shankarappa *et al.* studied the genetic variability of HIV-1 during the asymptomatic and AIDS stage of 9 men with moderate disease progression rate[365]. This report has been central in the field of HIV-1 intra-patient evolution, and the sequences have been reanalysed a number of times by different phylogenetic approaches[363,364,369,370]. Other studies investigating HIV-1 evolution over the disease course have also been presented (summarized in **Table 6**). Based on these studies it seems evident that HIV-1 diversity increases, and that the divergence rate is relatively constant during the asymptomatic stage. Late in disease the divergence rate seems to decline, mainly as an effect of a decline in the

rate of non-synonymous substitutions (dN). This could be a reflection of the severely impaired immune system in late-stage disease, resulting in a weaker selective pressure on the virus population. Attempts of correlating evolutionary processes with disease progression rate have been challenging, and different reports have shown somewhat contradicting results (**Table 6**). This might be a result of fundamental discrepancies between some of the studies, both in terms of method selection and investigation of closely related disease progressor groups. Other factors have been limitations in sample size and the stratification of already small groups of study subjects into even smaller subgroups. Another way of studying evolutionary processes, when biological data is missing or sparse, is by mathematical modelling. In 1991, Nowak *et al.* presented a model of HIV evolution coupled to the disease course termed the "diversity threshold model"[371,372]. The model relied on the observation that viral diversity continuously increases over the disease course and suggests that all HIV-infected patients have an individual diversity threshold of antigenic variation. When the threshold is reached, the immune system loses control over the infection and the individual progress to AIDS.

Intra-population evolution

As described in the section *HIV variation and subtypes*, HIV-1 has evolved into several subtypes, undergone countless recombination events and diversified extensively. Little is known about how this extreme genetic diversity and evolution will affect viral attributes like transmission efficiency or pathogenicity. The only way of understanding underlying mechanisms for such evolution is to study viral dynamics and spread both in a population perspective and within infected hosts. It has been hypothesized that the effect of the genetic bottleneck that HIV goes through adapting to a new host at transmission might reset the virus to a lower level of replicative capacity than it had in the previous host[383]. Even though the replicative capacity increases within a particular host over the course of infection, this does not compensate for the loss during the transmission event and adaptation period early in infection. Escape from strong cytotoxic T lymphocyte responses, due to diverse HLA allelic pools between individuals, has been suggested as one of the major causes of fitness loss[383]. In this way,

Table 6. HIV-1 diversity and evolutionary rates during disease progression. The arrows indicate the direction of different evolutionary parameters: ↑ = increase; ↓ = decrease; ↔ = relatively stable. In terms of evolutionary rates ↔ indicates a stable rate, meaning that the evolving viral population is diverging from the founder strain at a constant rate during the indicated disease stage. Since rates can differ between different disease stages, the arrow during the symptomatic stage or in late-stage disease (S) indicates the rate difference in relation to the asymptomatic stage (AS). The table is based on the results of the following studies[219,363-365,369-371,373-382].

No.	Publication	Published	N[1]	TP[2]	FU[3]	No. of Clones per TP	Genetic region under study	Methods used	Diversity[4]		Rates[5]						Comments
											Divergence		dN		dS		
									AS	S	AS	S	AS	S	AS	S	
1	Nowak *et al.*	1991	2	5	7	7-12	V3, 279 bp	Distance	↑	↓							
2	McNearney *et al.*	1992	4	2-4	2-5	2-10	V3, 330 bp	Identity	↑								
3	Bonhoeffer *et al.*	1995	1	6	7	11	V3, 231 bp	Distance			↔	↓	↔	↓	↔	↔	
4	Lukashov *et al.*	1995	44	2	5	D[10]	V3, 290 bp	Distance	↑								At tp1 were all in AS stage, whereas 34% had CD4<200 at tp2. Difficult to interpret the results for these 34% (inconsistency in disease stages).
5	Wolinsky *et al.*	1996	6	5-8	3-10	5-13	V3-V5, 650 bp	NJ[11]	↑		↔	↓			↔	↔	Stratification into three groups. Increased dN rate was suggested to be inversely correlated with disease progression rate.
6	Delwart *et al.*	1997	17	8	3-9	D[10]	V3-V5, 700 bp	HMA[12], Distance	↑	↓	↔	↓					
7	Ganeshan *et al.*	1997	6	4	1-4	10-13	C2-V5, 650 bp	NJ[11]					↔		↔		Children stratified into progressors and non-progressors.
8	Shankarappa *et al.*	1998	4	4-6	4-8	8	V3-V5, 550 bp	Distance	↑		↔		↔		↔		Rates displayed a somewhat fluctuating pattern over time.
9	Markham *et al.*	1998	15	8	4	6-21	V3, 285 bp	NJ[11]	↑↔	↑↔	↔	↔↓					Three groups based on disease progression. Both diversity increase and divergence rate are positively correlated to disease progression.
10	Shankarappa *et al.*	1999	9	12	6-12	13	C2-V5, 650 bp	Distance	↑	↓↔	↔	↓					All subjects are moderate progressors, difficult to stratify clear groups.
11	Ross *et al.*	2002	8[6]	12	6-11	6-17	C2-V5, 650 bp	NJ[11]									Two groups based on disease progression, No difference in rates. Non-uniform selection pressure over the analyzed region.
12	Williamson *et al.*	2003	50[6,7,8,9]					Adaptive evolution									Analyses of different datasets from different genetic regions. Some of the datasets are not descried in this table. Higher rates among slow progressors.
13	Mikhail *et al.*	2005	3	4-7	7-8	D[10]	Full genome, 8.7 kbp	ML[13], Bayesian			↔	↔					Positive correlation between divergence and disease progression rate.
14	Williamson *et al.*	2005	8[6]	12	6-12	13	C2-V5, 650 bp	Probability					↔	↓	↔	↔	Developed a probability model for rate analysis.
15	Lemey *et al.*	2007	9[6]	12	6-12	13	C2-V5, 650 bp	Bayesian					↔	↓	↔	↔	Positive correlation between dS and disease progression rate.
16	Bello *et al.*	2007	14	2	2-5	18-20	C2-C4, 500 bp	NJ[11]	↑								Four groups based on VL, both diversity and divergence were correlated with VL.
17	Lee *et al.*	2008	9[6] 6[7]	12 5-8	6-12 3-10	13 5-13	C2-V5, 650 bp V3-V5, 650 bp	ML[13]	↑	↔	↔	↓	↔	↓	↔	↓	No correlation between rates and disease progression rate.

[1]N = number of study subjects. [2]TP = number of time-points/study subject. [3]FU = follow-up time. [4]Diversity = the genetic variation at a given time-point. AS = asymptomatic stage. S = symptomatic stage or late in disease. The change in rate is in relation to the rate seen during the AS stage. [5]Divergence rate = the rate by which the genetic distance changes in relation to a reference point (typically the founder strain). dN = the part of the divergence rate that are due to non-synonymous changes. dS = the part of the divergence rate that are due to non-synonymous changes. [6]Patients from study no. 10. [7]Patients from study no. 5. [8]Patients from study no. 7. [9]Patients from study no. 9. [10]Direct sequencing. [11]NJ = neighbor joining. [12]HMA = Heteroduplex mobility assay. [13]ML = maximum likelihood.

HIV would gradually attenuate and adapt to the human population. In the opposite scenario, in a population with limited HLA diversity, the virus could instead increase in virulence due to less restrictive bottlenecks.

There have been some reports presenting evidence of an evolving HIV-1 epidemic. Recently, Kawashima *et al.* showed that HIV-1 adapts over time to the most abundant HLA types by losing epitopes on the population level[384]. It has also been suggested that HIV-1 has become more resistant to antibody neutralization by comparing virus isolates collected in the 1980's with more recent isolates[385]. Moreover, evolving patterns in the ability of HIV-1 to utilize CXCR4 as a coreceptor has been presented in both HIV-1 subtype C and CRF02_AG infected populations (**Paper IV** and[205]). Finally, in a study by Maljkovic Berry *et al.*, intravenous drug-users were shown to have a lower evolutionary rate on the population level compared to heterosexual infected individuals[386]. The lower rate was suggested to be an effect of repeated transmission during the initial stage of infection, before any selective pressure of the immune system had impacted the viral evolution.

HIV therapy and prevention

HIV is one of the fastest changing pathogens, constantly one step ahead of the immune system. Similarly, in the beginning of the HIV therapeutic era, HIV was rapidly evolving towards resistance. Initial treatment in the late 1980's, consisted of the nucleoside analogue Zidovudine (AZT)[387]. Nucleoside analogue reverse transcriptase inhibitors (NRTI) were the first class of drugs that was used as antiretroviral therapy (ART) and after AZT, similar drugs quickly followed. NRTI are indirect inhibitors of the viral reverse transcriptase and act as chain terminators during the reverse transcription. However, the impact of these agents on plasma VL was modest, and it soon became evident that treatment with mono or dual-therapy quickly resulted in drug resistance and treatment failure[388,389]. It was not until the mid 1990's, when the first protease inhibitor (PI), Saquinavir was introduced, that ART became effective[390]. The administration of two NRTIs and one PI in combination resulted in sustained reductions in plasma VL and larger increases in $CD4^+$ T cell levels that had been seen before[391,392]. Consequently, triple combination therapy using any three drugs was termed highly active ART (HAART).

HAART

HAART was introduced in 1996 and has been highly beneficial to many HIV-infceted individuals. Currently, initial HAART treatment consists of two NRTIs and either one PI or one non-NRTI (NNRTI, the third drug class consisting of drugs of

inhibitors working directly on the viral reverse transcriptase). There is no empirical evidence for withholding treatment at any stage of HIV infection, and death rates are almost twice as high when therapy is deferred (until the CD4 count falls below 500 per µl) compared to starting therapy when the CD4 count is above 500 per µl [386]. Recent evidence has indicated that early initiation of antiretroviral therapy during the asymptomatic stage of infection significantly improves survival, as compared with deferred therapy[393]. However, the timing for starting HIV treatment is still subject to debate[394]. The United States Panel on Antiretroviral Guidelines for Adults and Adolescents in 2009 recommended that ART should be initiated in all patients with a $CD4^+$ T cell count less than 350 per µl, with treatment also recommended for patients with $CD4^+$ T cell counts between 350 and 500 per µl. For patients with $CD4^+$ T cell counts over 500 per µl antiretroviral therapy is optional, considering that once you have started, it is a life-long treatment and it is important to understand the benefits and risks of therapy and adherence[395]. Apart from the three drug classes described above, there are two other classes of ART, the integrase inhibitors and the entry inhibitors. These drugs provide treatment options for patients who are infected with viruses already resistant to common therapies, although they are not widely available and not typically accessible in resource-limited settings.

However, HAART does not cure the patient or necessarily remove all symptoms, and if the treatment is stopped, high levels of HIV-1 return (often as HAART-resistant)[396]. It has been estimated that it would take more than a lifetime for HIV to be cleared by HAART[397]. Moreover, HAART sometimes fails to control the viral infection. Even if this can be due to a variety of reasons, non-adherence and non-persistence are the major reasons for failure. It should also be emphasized that there are a number of side effects with HAART including lipodystrophy, dyslipidemia, insulin resistance, an increase in cardiovascular risks, and birth defects[398]. Despite this, HAART has resulted in a large reduction in HIV-associated morbidity and mortality over the world, and the average life expectancy has been suggested to be 32 years after infection if treatment is initiated at a $CD4^+$ T cell count higher than 350 per µl[399]. This should be compared with the observed median progression-time to AIDS of 9-10 years for treatment naïve individuals[400].

HIV latency

Shortly after the introduction of HAART it was predicted that HIV would be eradicated within 2-3 years[401]. However, the model failed to take two important HIV aspects into account, (1) viral reservoirs (latency), and (2) viral evolution (drug resistance). Viral reservoirs have been found in a variety of cells, and virus from

these reservoirs can replenish the pool of replicating viruses[402-404]. One of the major challenges in curing HIV is likely to be how to distinguish and eradicate resting cells carrying HIV proviral DNA[405].

Vaccine and protective strategies

To date and despite tremendous efforts, there is still no cure or vaccine for HIV or AIDS. However, in September 2009 it was reported from a trial in Thailand that a combination of two previously unsuccessful vaccine candidates had resulted in a reduction of infections[406]. The vaccine was reported to have 30% efficiency. Although the results are promising, it is important to note that the results were sensitive to small changes in HIV incidence due to borderline significance of the results. Follow-up reports are awaited with large excitement. Another exciting and recent report comes from a vaginal gel containing the NRTI Tenofovir, which was shown to reduce HIV infection rates by 39% in a trial conducted in South Africa[407].

In light of this introduction, it is apparent that eliciting a persisting and broad neutralizing antibody response is very challenging, and that CTL responses are central in control of viral infection. In order to elicit immune responses that can protect against infection or disease progression, it is reasonable to believe that both these arms of the immune system need to be exploited.

MATERIALS AND METHODS USED IN THIS DOCTORAL DISSERTATION

Study cohorts and samples under study

All studies in this doctoral dissertation (**paper I-V**) are based on samples from a prospective open cohort of police officers in Guinea-Bissau. The cohort was initiated in February 1990, and new participants have continuously been included, except for a temporarily closure from June 1998 until the end of 2002 as a result of the civil war in 1998–1999. However, annual controls of previously included individuals were resumed already in July 2000. All persons with a regular employment in the Guinea-Bissau police force were eligible for the study, which has been voluntary and with less than 2% refusal to participate. Blood samples for serology and $CD4^+$ T-cell counts have been collected at inclusion and at follow-up visits, scheduled at intervals of 12–18 months. Antiretroviral therapy (ART) was recently introduced into Guinea-Bissau, and the few individuals receiving ART have been censored in the above-mentioned papers from the time-point of ART initiation. The cohort is unique in identifying HIV-1 seroconverters, which has been the focus of **papers II and III**. Date of seroconversion (SC) of HIV-1 was estimated as the mid-time-point between the last HIV-1 seronegative sample and the first HIV-1 seropositive sample.

In order to conduct different molecular sequence analyses related to viral subtype distributions and evolutionary patterns in **papers I-V**, a total of 257 blood plasma samples from 94 individuals were subjected to RNA extraction and PCR amplification of the HIV-1 *env* gp120 V1-V3 region. Of these, we were able to amplify 157 samples from 82 individuals. Sequences based on molecular clones of these amplicons (12 clones were routinely sequenced) constitute the foundation of the molecular sequence analyses used throughout this doctoral dissertation. For **papers I, IV and V** we also analyzed different sets of sequences deposited in Genbank. Details of these datasets can be found in each of the corresponding papers.

In **paper I**, we used HIV-1 sequences derived from blood plasma samples of all 82 study subjects from Guinea-Bissau that we were able to amplify. 77 of the subjects were from the police cohort and five were from a case control study conducted in Bissau, Guinea-Bissau, under surveillance of the Bandim Health Project[280]. The samples had been collected between 1993 and 2008, and were included in the study based on follow up time, disease status (patients in both the asymptomatic stage and in late-stage disease are represented in the dataset) and sample availability. A summary of the 82 study subjects from the police cohort and the case control study used for molecular analyses in **paper I-V** are found in **Table 7**.

A total of 223 HIV-1 seroincident individuals were included in **paper II and III**.

Among those, 189 were infected by HIV-1 only (159 males and 30 females, referred to as single-infected individuals), and 34 were infected with both HIV-1 and HIV-2 (28 males and six females, dual-infected individuals). The mean ages at SC were 36.4 years (SD 9.6, range 18-60 years) and 39.6 years (SD 8.0, range 20-59 years) for single and dual-infected individuals, respectively. Of the 34 dual-infected individuals, 14 had recorded their HIV-1 and HIV-2 infection simultaneously (HIV-D_s). The other 20 dual-infected individuals first got infected with HIV-2 (earlier) and then with HIV-1 (HIV-D_c). Different subsets of samples from these seroincident individuals were used for analysis of T cell populations, soluble immune activation markers and evolution. Details about the different stratifications and subsets can be found in **paper II and III**.

In **paper IV**, we analyzed 29 blood plasma samples from the police cohort, collected during late-stage disease from individuals with single HIV-1 infection. Late-stage disease was defined as $CD4^+$ T cell count (≤200 cells per µl or ≤14%) or clinical AIDS as according to the Centers for disease control and prevention, U.S.A. (CDC) and World Health Organisation (WHO) (CDC: C or WHO: 4)[408,409]. In cases where more than one sample from late-stage disease was available, the last sample was chosen. Individuals diagnosed with tuberculosis and clinically categorized as CDC: C, but without other AIDS-defining symptoms were not included in the study. In addition, a sample set of 11 HIV-1 isolates with known subtype and coreceptor tropism were included to determine the accuracy of the phenotypic method. The isolates were generously provided by Professor Jan Albert, Swedish Institute for Infectious Disease Control, Stockholm, Sweden.

Finally, in **paper V**, we studied 28 of the police cohort samples from **paper IV** (the subtype C sample were excluded) and an additional 18 samples collected during late-stage disease from HIV-1 and HIV-2 dual-infected individuals according to the above-mentioned criteria (whereof 17 were analyzed for coreceptor tropism, the excluded sample had HIV-1 of CRF06_cpx). Of these 17 samples, 12 were from the police cohort and five were from the case-control cohort[280].

For **papers I, IV and V** we also analysed different sets of sequences deposited in Genbank. Details of these datasets can be found in connection to respective paper.

Diagnostic laboratory methods

HIV testing was performed at the National Public Health Laboratory (LNSP), Bissau. In 1990–1994, sera were screened for HIV-1 and HIV-2 antibodies by enzyme-

Table 7. Sampling year, date of seroconversion, sample place, and HIV-1 clade of the 82 study subjects used for molecular analyses in this doctoral dissertation.

Sample[1]	Sample year	SC year[2]	Geography	Clade[3]	Sample	Sample year	SC year	Geography	Clade
DL1996H_13	2000	1990	Bissau	CRF02_AG	DL3166F_4	2003	1999	Bissau	A3/AG
DL2004F_4	2001	2001	Northwest	CRF02_AG	DL3169F_7	2004	1999	Bissau	CRF02_AG
DL2014F_1	1995	1990	Bissau	A3	DL3170D_5	1997	1993	Bissau	CRF02_AG
DL2066G_8	2002	2000	Bissau	CRF02_AG	DL3234J_3	2006	1999	Bissau	A3/AG
DL2075I_1	2007	1998	Bissau	CRF02_AG	DL3247F_2	2007	1999	Bissau	CRF02_AG
DL2089J_9	2003	1997	Bissau	CRF02_AG	DL3288H_5	2005	NA	Northwest	CRF02_AG
DL2096D_7	1996	1993	Bissau	C	DL3312C_1	1994	1993	Northwest	CRF02_AG
DL2102F_1	1998	1994	Bissau	CRF02_AG	DL3339D_12	2001	1998	Northeast	CRF02_AG
DL2111E_2	1998	1994	Bissau	CRF02_AG	DL3372I_10	2004	2000	Northeast	A3/AG
DL2164F_4	2002	1997	Bissau	A3	DL3406J_5	2006	2000	Northwest	A3
DL2198F_3	1996	1991	Bissau	CRF02_AG	DL3442D_8	2001	2000	Northeast	A3/AG
DL2249E_8	1997	1996	Bissau	CRF02_AG	DL3468E_10	2007	1994	Bissau	C
DL2315F_4	2000	1997	Northwest	CRF02_AG	DL3556C_9	1997	1995	Bissau	CRF02_AG
DL2325J_8	2008	2000	Bissau	CRF02_AG	DL3633G_2	2003	1999	Bissau	CRF02_AG
DL2339E_4	2003	1992	Northwest	CRF02_AG	DL3721C_3	1997	1994	Bissau	A3/AG
DL2365I_3	2002	1996	Bissau	A3	DL3733D_9	2000	1997	Northwest	CRF02_AG
DL2391B_5	1993	1991	Bissau	CRF02_AG	DL3766D_9	1997	1994	Bissau	C
DL2401M_5	2004	1996	Bissau	CRF02_AG	DL3860H_3	2008	1995	Bissau	A3
DL2462H_4	2001	1999	Bissau	CRF02_AG	DL3869G_3	2003	NA	Bissau	A3
DL2470E_5	2000	1998	Bissau	A3	DL3895C_2	1996	1995	Bissau	CRF02_AG
DL2568E_9	2003	2003	Bissau	A3/AG	DL3938D_3	1998	1995	Bissau	CRF02_AG
DL2544F_4	2000	1999	Northwest	CRF02_AG	DL3946E_4	2004	2001	South	CRF02_AG
DL2594J_3	2005	2001	Northeast	A3	DL3981C_10	1998	1995	Bissau	CRF02_AG
DL2596E_4	2000	1995	South	A3	DL4023G_7	2006	1999	Bissau	CRF02_AG
DL2640I_7	2004	2003	Bissau	CRF02_AG	DL4084F_2	2003	1999	Bissau	A1
DL2673D_4	1998	1993	South	CRF02_AG	DL4169F_9	2002	1996	Bissau	CRF02_AG
DL2713H_7	2007	2002	Northwest	CRF02_AG	DL4214D_8	2002	NA	Bissau	C
DL2747I_2	2005	1994	Northwest	CRF02_AG	DL4248G_5	2005	NA	Bissau	A3
DL2766C_11	1995	1992	Northwest	CRF02_AG	DL4303D_7	2002	1999	Bissau	CRF02_AG
DL2829F_2	2006	1999	Northeast	CRF06_cpx	DL4422B_4	2003	NA	Northeast	CRF02_AG
DL2846F_8	2001	1993	Bissau	A3	DL4477D_9	2001	NA	Northeast	CRF02_AG
DL2853E_2	1998	1993	Bissau	CRF02_AG	DL4525G_2	2006	1999	Bissau	A3
DL2908G_8	2000	1997	Bissau	CRF02_AG	DL4632E_5	2003	1997	South	CRF02_AG
DL2920H_2	2004	1998	Bissau	CRF02_AG	DL4957C_5	2005	NA	Bissau	A3
DL3004H_11	2003	1999	Bissau	CRF02_AG	DL5342B_4	2007	NA	South	A3/AG
DL3018E_10	2003	2001	South	A3/AG	DL6324B_13	2007	NA	Bissau	CRF02_AG
DL3037E_8	2005	1996	Northeast	A3/AG	DL11967A_7	2005	NA	Bissau	A3/AG
DL3039D_12	1998	1993	Northwest	A3/AG	DL11968A_10	2005	NA	Bissau	A3
DL3071F_10	2002	2001	Bissau	A3/AG	DL11969A_4	2005	NA	Bissau	A3
DL3087E_5	2001	NA	Bissau	CRF02_AG	DL11970A_1	2006	NA	Bissau	A3/AG
DL3098I_6	2007	2002	Bissau	CRF02_AG	DL11971A_1	2006	NA	Bissau	A3

[1]Samples identification number. The last number (i.e. _1) represents the sequence clone number.
[2]Determined seroconversion date, estimated as the date between the last HIV-1 negative sample and the first HIV-1 positive sample.
NA = Not applicable.
[3]This column shows the HIV-1 subtype, CRF or other form as determined by the phylogenetic analysis.

linked immunosorbent assay (ELISA) with the use of the Behring anti-HIV-1/HIV-2 (Behring, Marburg, Germany) and/or Wellcozyme recombinant anti-HIV-1 (Wellcome, Dartford, UK) and an in-house HIV-2 (SBL6669) ELISA assay[410]. From 1995 and onwards, screening was performed with Behring Enzygnost HIV-1/HIV-2 Plus ELISA (Behring). Confirmation of positive results was done with western blot analysis (Diagnostic Biotechnology anti-HIV-1 blot 2.2, Science park, Singapore, or in-house anti-HIV-2) and dually HIV-1/HIV-2-positive samples were confirmed by Pepti-lav (Sanofi Diagnostics Pasteur, Marnes-la-Coquette, France) in the years 1990–

1998[411]. Since 1999, an alternative confirmation strategy has been used with Capillus HIV-1/HIV-2 (Cambridge Biotech Limited, Galway, Ireland) and Immunocomb II HIV-1 and 2 BiSpot RST (Orgenics, Yavne, Israel). Evaluations of the HIV antibody testing strategy in Bissau have shown a high concordance between the results obtained by serology and by PCR and a high degree of distinction between HIV-1 and HIV-2[412,413]. In addition, serology of sequential samples of the same infected individuals was without discrepancies, which further strengthens the results of this strategy. The screening assays have been evaluated in parallel to ensure reproducibility between assays[412].

T-lymphocyte subsets were determined at the LNSP, by conventional flow cytometry (Until 2005: FACStrak; Becton Dickinson, San Jose, Calif, with the use of three two-color immunofluorescent reagents, CD45/CD14, CD3/CD4, and CD3/CD8, Simultest, Becton Dickinson; and from 2006 and onwards: CyFlow, Partec, Münster, Germany). Leukocyte counts were performed with a cell counter until 2005 (Coulter Counter CBC5; Coulter Electronics Ltd, Luton, England), and by CyFlow from 2006 and onwards.

Since VL measurements have not been included as a standard procedure in Guinea-Bissau, we were unable to include VL data in our analyses. Due to limitations in sample size and sample availability, we were not able to do any retrospective and valid sequential VL measurements of the samples included in this study.

Epidemiological analysis

A survival analysis was performed for reported mortality and for progression to AIDS in **paper II and III**. Because of the imbalance in gender representation, the significance tests were also performed with stratification on gender. Kaplan-Meier curves were generated for graphical presentation. AIDS was defined as clinical stage WHO 4 and CDC C, $CD4^+$ T-cell counts <200 cells per mL, CD4% counts <14%, or reported death with symptoms of AIDS[408,409]. A Cox proportional hazards model was applied, adjusting for gender and age at seroconversion categorized in four age groups: <25, 25-34, 35-44, and ≥45 years. Proportional hazards assumption was controlled for all covariates.

Analysis of T lymphocyte populations and soluble immune activation markers

It is well established that both absolute numbers of $CD4^+$ T cell count and CD4% are reliable immunologic markers of HIV disease progression. However, in resource-limited settings it has been suggested that CD4% may be a more suitable marker

due to lower variability on repeated measurements and lower sensitivity to specimen handling, age of the patient or time of sampling (low counts in early morning)[414-416]. In addition, CD4% has been shown to have a higher prognostic value in predicting disease[416]. Accordingly, for **paper II and III** the percentage levels for analysis of both $CD4^+$ and $CD8^+$ T cell counts were used. CD4% and CD8% levels for two or more measurements were recorded for 71 and 57 HIV-1 single and 24 and 22 HIV-1 and HIV-2 dual-infected individuals, respectively. The rate of change in CD4% and CD8% over time post seroconversion was estimated using a linear regression model. The rate in CD4% decline did not differ between single and dual-infected groups. Therefore the global mean rate in CD4% decline was used to extrapolate from each individual's mean CD4% to the global mean time-point after SC. For the CD8% analysis there was a significant difference in rate increase between the single and dual-infected individuals. Therefore, we used each groups mean rate, instead of the global mean for comparison of levels in CD8%.

To investigate any potential differences that could be coupled to immune activation, concentrations of beta-2 microglobulin (b2m) and neopterin in plasma was determined. These immune activation markers were chosen because they have been shown to be stable markers over periods over sample storage[417]. They have also been reported to be relatively stable through freeze-thawing and temperature fluctuations, and they have been used in resource-limited settings in previous studies[417,418]. B2m and neopterin were measured in samples from 42 single and 25 dual-infected individuals for b2m, and from 42 single and 24 dual-infected individuals for neopterin. Two sample time-points were analysed for each individual (mean sample time-point from infection date was 29 and 27 months (time-point one), and 55 and 55 months (time-point two) for single and dual-infected individuals, respectively). Differences in levels of both beta-2 microglobulin and neopterin between single and dual-infected individuals were analysed as described for the analysis of CD4% levels (using the global mean rate).

Molecular cloning strategy and determination of coreceptor tropism

The genetic region under study in **paper I-V** was the HIV-1 *env* gp120 V1-V3 region (approximately 940 bp). However, in **paper IV** this region was only partly analyzed, and most analyzes were performed on direct sequenced HIV-1 *env* gp120 V3 regions (approximately 530 bp). To determine molecular sequences, viral RNA was isolated from blood plasma samples, and reverse transcribed. The genetic regions under study were amplified using a nested PCR approach. The V1-V3 region was TA-cloned, and to obtain a fair representation of the viral population, 12 clones were picked from each sample and sequenced for subsequent analyses (**Fig. 18**).

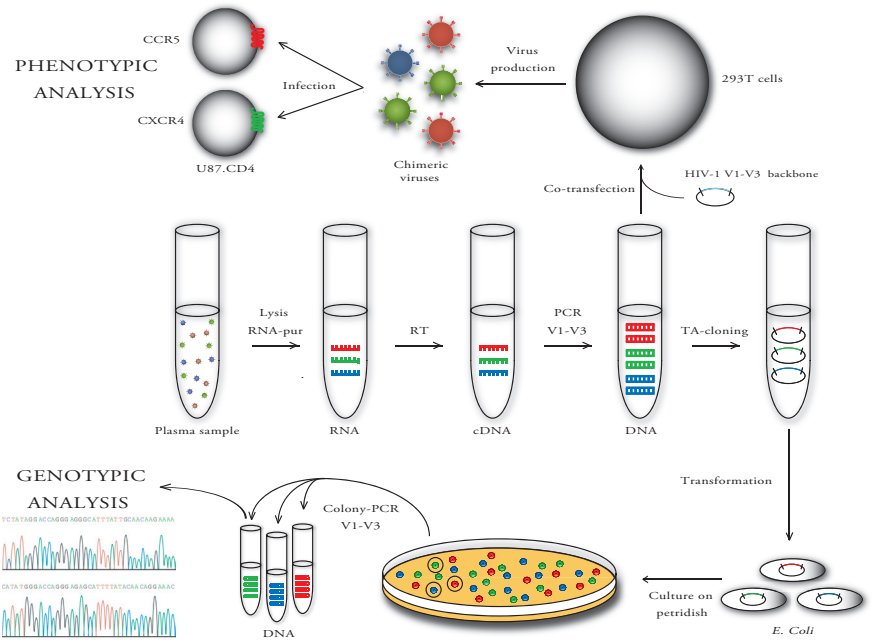


Figure 18. Experimental design. Viral RNA was extracted and purified from blood plasma samples. After reverse transcription, the HIV-1 *env* gp120 V1-V3 region was amplified and cloned into chemically competent *E. Coli*. Clones were collected and the V1-V3 region was amplified and sequenced for genotypic analysis. To determine HIV-1 coreceptor tropism, the V1-V3 region and an HIV-1 backbone were used to construct chimeric viruses. The tropism was determined by infection of U87.CD4 cells expressing either CCR5 or CXCR4. The colours of the nucleotide strands, clones, colonies and chimeric viruses are indicating whether or not different subpopulations of the HIV-1 quasispecies have been separated.

Genotypic characterization of CRF02_AG coreceptor use

The aim was to characterize genetic properties in HIV-1 V3 *env* distinguishing CRF02_AG with different coreceptor tropism. Only two sequences of the control panel and three sequences of the plasma samples were of HIV-1 CRF02_AG with pure CCR5 tropism. Since this number was too low for an appropriate comparison between the R5 and R5/X4 groups, we added all available V3 sequences, with known coreceptor tropism, from Los Alamos sequence data base, to a final dataset of 111 sequences (75 R5 and 36 R5X4 or X4 sequences)[241]. Only one sequence per patient was subjected for analyses. The CRF02_AG V3 amino acid sequences were multiple aligned, and positively [K and R] and negatively [D and E] charged amino acids was counted. In codons with amino acid mixtures, all possible permutations were

assessed. The combination resulting in the highest net charge was used for phenotype prediction. The performance of different sequence motif-based rules was measured in terms of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). The sensitivity was determined as the fraction of predicted X4 sequences among the sequences from viruses phenotyped as CXCR4-using, the specificity as the fraction of predicted R5 sequences among the sequences from viruses phenotyped as CCR5-using only, the PPV as the fraction of correctly predicted X4 sequences among all predicted X4 sequences, and the NPV as the fraction of correctly predicted R5 sequences among all predicted R5 sequences.

To determine the HIV-1 coreceptor tropism, chimeric viruses were constructed using an HIV-1 backbone with a deleted V1-V3 region. In this experimental system, a subsequent homologous recombination occurs between the backbone and the V1-V3 region after transfection. This forms an episome in the transfected 293 T cells (a human embryonic kidney cell line) from which the progeny virus are produced. The coreceptor tropism was determined by infection of U87.CD4 cells (a human glioma cell line) expressing either CCR5 or CXCR4. The phenotypic analysis was performed according to the protocol described by Trouplin *et al.* with some modifications[419]. The major difference was the use of U87.CD4 cells instead of the U373MG.CD4 cell line (indicates infection by a Tat-induced expression of β-galactosidase), which was used in the original manuscript. To determine infection we measured the production of p24, and infection was determined when the p24 antigen production was significantly increased over time.

Phylogenetic analysis

After assembly and analysis of contigs, sequences were analyzed in a number of different subsequent settings (workflow is shown in **Fig. 19**). In addition, a brief summary of the phylogenetic programs used in the **papers I-V** are outlined in **Table 8**. A detailed description of the different settings of these programs and methods can be found in respective paper, as outlined in **Fig. 19**.

Several methods for studying viral evolutionary rates have been used in the literature[343,361,363,370,379,386,430,431]. A major obstacle in phylogenetic analysis is often how to place the origin of related sequences. One of the most common approaches to get a direction of a phylogenetic tree is to use an outgroup sequence, assuming that the outgroup sequence and the sequences under study have a common evolutionary origin[373]. In our material, sequences of sequential series of samples were limited to two sample time-points (with 8-12 clones from each time-point). To find the "root"

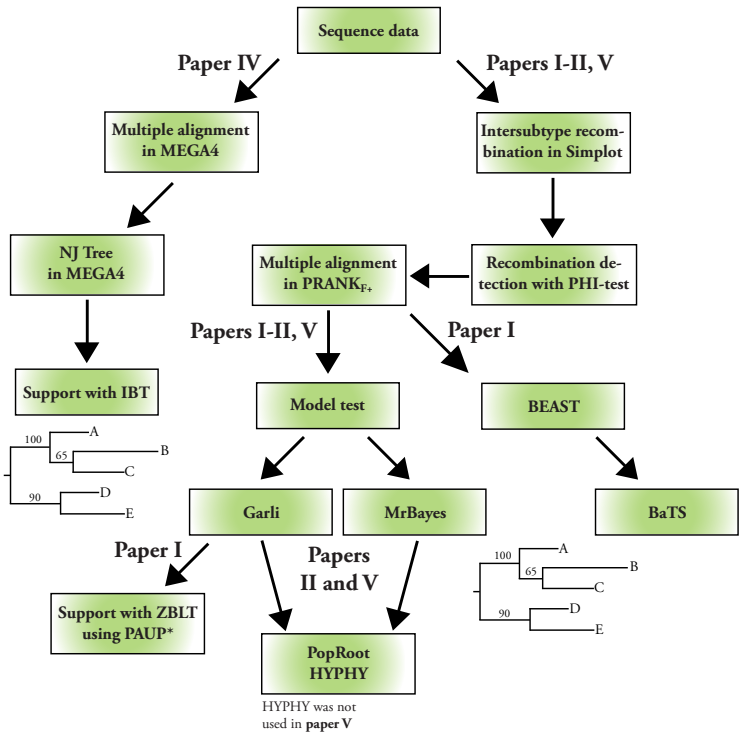


Figure 19. Summary of the workflow used in the phylogenetic analyses. For **paper IV**, the phylogenetic analysis was performed in MEGA4 using NJ with the interior branch test (IBT) as statistical support. To limit the risk of any putative interpatient or intrapatient recombinants influencing subsequent evolutionary analyses in **papers I-II and V**, we "cleaned" the data using the pairwise homoplasy index test (PHI-test). Sequences were aligned in PRANK$_{F+}$ using a NJ tree produced in MEGA4 as a guide-tree. The best fitting substitution model were determined using Modeltest, and ML and Bayesian phylogenies were produced in Garli and MrBayes, respectively. For **papers II and V**, diversity and/or evolutionary rates were determined in population-based rooted trees using in-house scripts and HYPHY as described in the text. For **paper I**, evolutionary patterns on the population level in Guinea-Bissau were investigated in BEAST, and trait associations were controlled for in BaTS. Finally, support for Guinea-Bissau specific clusters in ML trees in **paper I** was determined using the zero-branch length test as implemented in PAUP*.

of the different HIV-1 clones for each individual's two sample time-points, we used a rooting technique that we referred to as population-based rooting (PopRoot) (**papers II and V**). By inferring large phylogenies with a high amount of sequences with strong epidemiological linkage, we reasoned that these sequences would serve as reliable

Table 8. Programs used in the doctoral dissertation for sequence analysis[242,331,342,343,420-429].

Program	Reference	Description	Used in paper
MEGA 4	Tamura *et al.*	MEGA (Molecular Evolutionary Genetics Analysis) is an integrated tool for conducting automatic and manual sequence alignment, inferring phylogenetic trees, mining web-based databases, estimating rates of molecular evolution, and testing evolutionary hypotheses. Applications used in this doctoral dissertation include multiple alignment (employing the Clustal algorithm), manual editing of alignments, and phylogeny reconstruction (the NJ method with the substitution model composite maximum likelihood (a method used for describing the sum of log-likelihoods for all pairwise distances in a distance matrix) was used for tree building and the interior branch test (IBT) was used for phylogenetic uncertainty (the IBT is a T-test that is calculated based on bootstraps, it gives the confidence probability that a certain branch length is greater than zero).	**I-II, IV-V**
PRANK$_{F+}$	Löytynoja *et al.*	PRANK (Probabilistic Alignment Kit) is an alignment program that uses a guide tree to align codons of nucleotides in an efficient and evolutionary sound way. The algorithm has shown to be robust in aligning sequence data containing a lot of indels.	**I-II, V**
Splitstree	Huson *et al.* Bruen *et al.*	Splitstree is a program for computing unrooted phylogenetic networks from molecular sequence data. The pairwise homoplasy index (PHI) incorporated in Splitstree can be used to detect recombination in a dataset.	**I-II, V**
Simplot	Salminen *et al.* Lole *et al.*	Bootscanning, implemented in Simplot, uses a sliding window approach. As it slides over the alignment it constructs phylogenetic trees to detect recombination breakpoints. Support is determined by bootstrapping.	**I**
FigTree	Rambaut *et al.*	FigTree is part of the BEAST package and a program for viewing trees.	
Garli	Zwickl	Garli (Genetic Algorithm for Rapid Likelihood Inference) is a program for Maximum Likelihood-based inference of large phylogenetic trees.	**I-II, V**
Modeltest	Posada *et al.*	Modeltest is a program for the selection the model of nucleotide substitution that best fits the data. The program chooses among 56 models, and implements three different model selection frameworks: hierarchical likelihood ratio tests (hLRTs), Akaike information criterion (AIC), and Bayesian information criterion (BIC).	**I-II, V**
PAUP*	Swofford *et al.* Anisimova *et al.*	PAUP (Phylogenetic Analysis Using Parsimony (*and other methods)) is a program package with a wide range of methods for phylogenetic inference. The zero-branch length test (ZBLT) is an approximate likelihood ratio test for testing if a certain branch in a ML phylogeny is separated from zero.	**I**
MrBayes	Ronquist *et al.*	MrBayes is a program for the Bayesian estimation of phylogeny.	**I-II, V**
HyPhy	Pond *et al.*	HyPhy (Hypothesis testing using Phylogeny) is a program for ML analysis of nucleotide, protein and codon sequences, using predefined standard models or user-defined models of evolution (such as codon based substitution models).	**II**
BEAST	Drummond *et al.*	BEAST (Bayesian Evolutionary Analysis Sampling Trees) is a program for Bayesian MCMC analysis of molecular sequences with the focus of rooted, time-measured phylogenies inferred using strict or relaxed molecular clock models. In addition, analysis of different trait associations can be included and analyzed simultaneously with the molecular clock.	**I**
Tracer	Rambaut *et al.*	Tracer is program for analysing results from Bayesian MCMC programs such as BEAST & MrBayes.	**I-II, V**
BaTS	Parker *et al.*	BaTS (Bayesian Tip-Significance testing) is a package that allows the user to test for significant phylogeny-trait correlations whilst taking into account uncertainty arising from phylogenetic error, by integrating over the credible set of topologies produced by Bayesian phylogenetics programs such as BEAST or MrBayes.	**I**

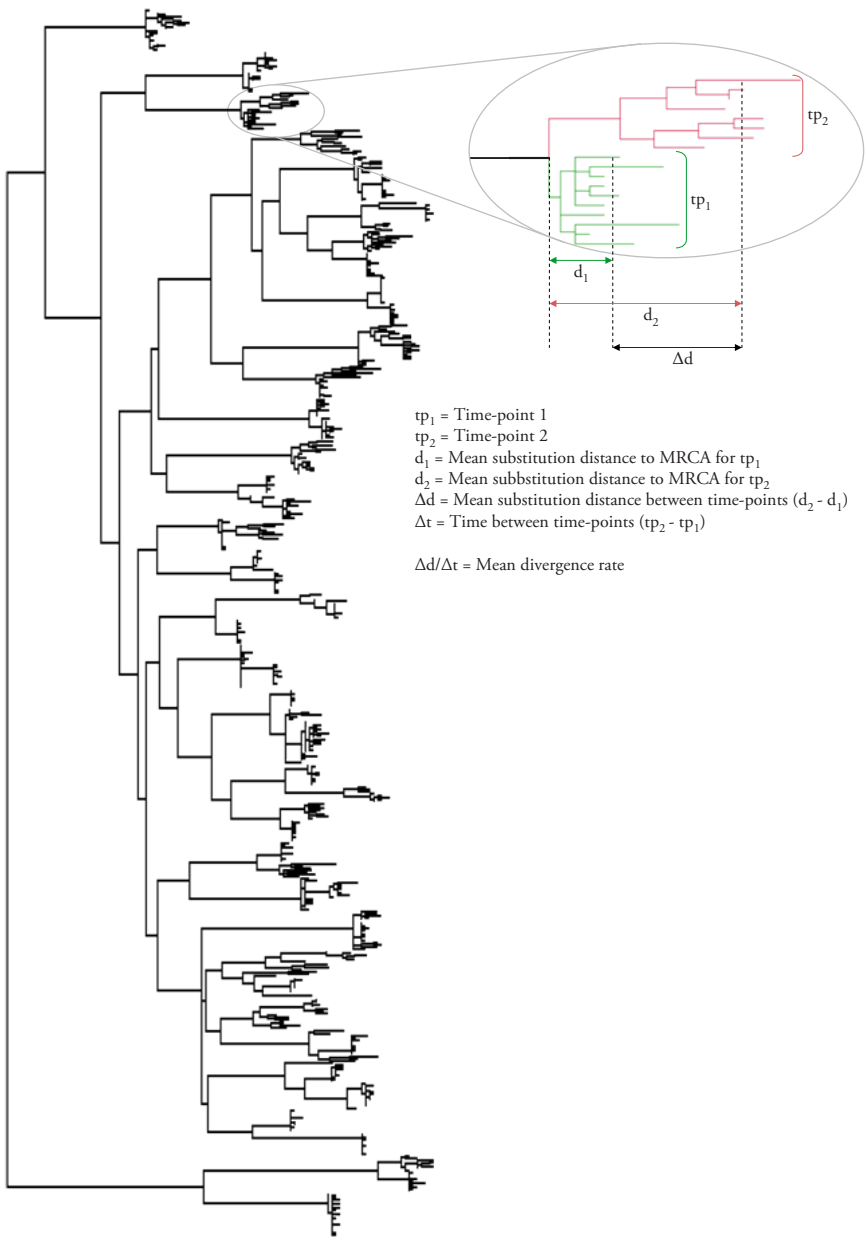
outgroups to each other. In this way it was possible to infer the MRCA for each patient-specific subcluster in an evolutionarily sound context[339]. By assuming a molecular clock, we were then able to estimate the evolutionary substitution rate for each

individual according to the single rate dated tips approach as described[430]. Divergence rates (measured as substitutions per site per year) were estimated as follows (**Fig. 20**): First, the MRCA of the two patient-specific time-points was determined; second, the mean number of substitutions per site between the clones of each time-point and the determined MRCA was calculated; third, the difference in mean substitutions per site from the MRCA (Δd) between time-point one (d_{tp1}) and time-point two (d_{tp2}), $\Delta d = d_{tp2} - d_{tp1}$ was calculated; and finally, Δd was divided by the difference in time between time-point one and time-point two. Phylogenetic uncertainty was taken into account by calculation of patient-specific divergence rates for 1,000 bootstrap trees (ML analysis) and 15,000 trees from the posterior distribution (Bayesian analysis). To investigate differences in synonymous (dS) and nonsynonymous (dN) substitution rates, we randomly picked 200 bootstrap trees from the ML analysis and the posterior distribution of trees from the Bayesian analysis, respectively. Each of the tree datasets were then reanalysed in HYPHY employing an MG94xGTR model, following the procedure described by Lemey *et al.*[363,428]. The dN and dS were calculated as described for the divergence rates.

The diversity of each time-point was calculated as the mean pair-wise distance between all clones from the corresponding time-point. As in the divergence analysis, this was done for all bootstrap trees and all trees from the posterior distribution to take the phylogenetic uncertainty into account.

Genotypic analysis

One of the aims in **paper IV** was to characterize genetic properties in HIV-1 V3 *env* distinguishing CRF02_AG with different coreceptor tropism. For this purpose we used a combined dataset of V3 sequences from the police cohort, control panel and Genbank with known coreceptor tropism, resulting in a final dataset of 111 sequences (75 with pure CCR5 tropism and 36 with CXCR4 tropism). In the analysis we compared six previously described genotypic rules and bioinformatic tools, and three novel genotypic rules based on number of charged aa in the V3 region[180,182,183,187]. The performance of different sequence motif-based rules was measured in terms of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). The sensitivity was determined as the fraction of predicted X4 sequences among the sequences from viruses phenotyped as CXCR4-using, the specificity as the fraction of predicted R5 sequences among the sequences from viruses phenotyped as CCR5-using only, the PPV as the fraction of correctly predicted X4 sequences among all predicted X4 sequences, and the NPV as the fraction of correctly predicted R5 sequences among all predicted R5 sequences.



tp_1 = Time-point 1
tp_2 = Time-point 2
d_1 = Mean substitution distance to MRCA for tp_1
d_2 = Mean subbstitution distance to MRCA for tp_2
Δd = Mean substitution distance between time-points (d_2 - d_1)
Δt = Time between time-points (tp_2 - tp_1)

$\Delta d/\Delta t$ = Mean divergence rate

Figure 20. Determination of HIV-1 evolutionary rate. Maximum likelihood phylogeny illustrating the principles of population-based rooting, and how the viral divergence was determined. The highlighted patient-specific cluster consists of 18 clones, which in turn consists of two time-point specific subclusters (the green tp1 cluster, and the red tp2 cluster). The two subclusters are connected by a MRCA that has been determined based on the combined relationship of all sequences in the phylogeny. This MRCA provides the evolutionary direction of this subcluster. The mean divergence rate between tp1 and tp2 is determined by dividing the mean substitution distance between tp1 and tp2 (Δd) with the time between tp1 and tp2 (Δt).

RESULTS AND DISCUSSION OF THE MAIN FINDINGS IN THIS DOCTORAL DISSERTATION

In this doctoral dissertation we tried to amplify the HIV-1 *env* gp120 V1-V3 region from a total of 257 blood plasma samples from 94 individuals. The PCR success rate was 75% (97 out of 130 samples) for HIV-1 single and 47% (60/127) for HIV-1 and HIV-2 dual-infected individuals. Since we had sequential plasma samples from several of the study subjects we also calculated the PCR success rate based on the number of individuals for which we had amplified the V1-V3 region of at least one sample. Based on number of individuals, the PCR success rate was 96% (51/53) for single and 76% (31/41) for dual-infected individuals. These analyses showed a striking difference in PCR success rate between single and dual-infected individuals.

Undertaking a large-scale scientific project in a resource-limited country as Guinea-Bissau is in many aspects very challenging and comes with some logistic difficulties. During the study period we have had storage problems (mainly due to irregular power failures at the laboratory), resulting in temperature fluctuations and freeze-thawing of blood plasma samples. In light of this, it is reasonable to believe that the age of samples could have influenced the PCR results. To illustrate the effect of sample age on PCR success rate, we grouped samples based on the year of collection and determined the percentage of positive PCR results of each sample year (**Fig. 21**).

Multiple logistic regression analysis showed that there was a significant effect of both sample collection time point ($p<0.001$, Likelihood-ratio test (LRT)) and HIV status (single or dual infection) ($p<0.001$, LRT) on the PCR success rate. Inclusion of interaction between sample collection time point and HIV status in the model, showed that there was no significant difference in the effect of collection time point between the single and dual-infected group ($p=0.903$, LRT). These results show that the age of the sample was important for successful PCR results, but also that it was easier to amplify the V1-V3 region from single than dual-infected individuals (**Fig. 21**). The increased PCR success rate with more recently collected samples is likely explained by storage difficulties and freeze-thawing, resulting in gradual degradation of viral RNA in the samples. However, there is no reason to believe that samples from single and dual-infected individuals have been handled unequally, and the rate in RNA degradation should therefore be similar between the groups. From our point of view, there could be two possible explanations for the difference seen between single and dual-infected individuals. First, the regression analysis suggested a continuous degradation of RNA. The shorter time to reach viral RNA levels below the PCR detection limit among dual-infected individuals could therefore be the result of a generally lower viral load. This explanation also goes well with the lower disease progression rate seen in dual-infected individuals (**paper II**). Second, there have been

some concerns in the literature regarding serologic determination of dual HIV-1 and HIV-2 infection[432-436]. These concerns have been based on early reports suggesting different levels of cross-reactivity in detection of HIV-1 and HIV-2, especially in western blots[434,437]. In Guinea-Bissau a specific HIV antibody testing strategy has been

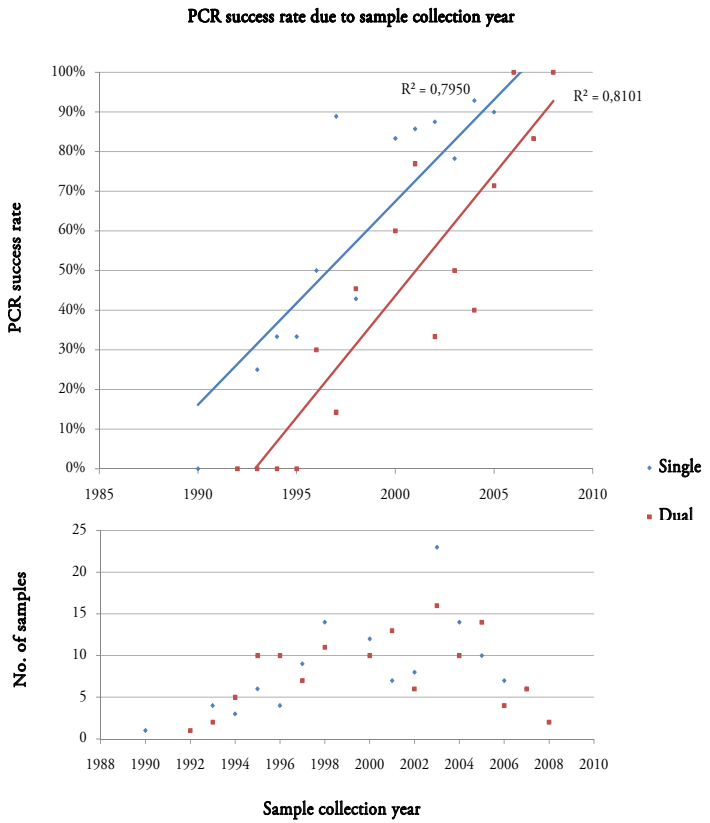


Figure 21. PCR success rate due to sample collection year. The top figure illustrates the PCR success rate over time (due to sample collection year), whereas the bottom figure shows the total number of samples analyzed from different years. The blue line and dots represent data from HIV-1 single-infected individuals, whereas the red line and dots represent data from HIV-1 and HIV-2 dual-infected individuals.

developed to discriminate between HIV-1 and HIV-2, based on serologic peptide-based methods (the recommended method of diagnosis in resource-limited settings by the WHO and CDC)[412,438]. This strategy has shown high concordance between results obtained by serology and by PCR, and an equally high degree of distinction

between HIV-1 and HIV-2. In addition, serology of sequential samples of the same infected individuals was consistent over time, which further strengthens the results of this strategy. Moreover, the majority of the dual-infected individuals were recorded as single HIV-2 seropositives during several years before they got seropositive for both HIV-1 and HIV-2.

Next, we wanted to investigate if the level of $CD4^+$ T cells at the sample time point could have influenced the PCR success rate. CD4% records were available for 196 samples (90 and 106 samples from single and dual-infected individuals, respectively). To illustrate the effect of CD4% on PCR success, we grouped samples based on the level of CD4% and determined the percentage of positive PCR results of each group (**Fig. 22**). Multiple logistic regression confirmed the effect of sample collection time-point ($p<0.001$, LRT) and HIV status ($p<0.001$, LRT) on PCR success rate in this subset of samples. In addition, we found that the level of CD4% was important for the PCR outcome ($p=0.010$, LRT). The interaction term between the level of CD4% and HIV status suggested that there could be a difference in the effect of CD4% between the single and dual infected group ($p=0.105$, LRT). Based on this trend, we analyzed the effect of CD4% on PCR success rate in each group separately. For single-infected individuals, the age of the sample was shown to be important for the PCR results ($p=0.016$, LRT), whereas the levels of CD4% was not ($p=0.776$, LRT). In contrast, both age of the sample and the level of CD4% displayed a significant effect on the PCR success rate on samples from dual-infected individuals ($P<0.001$ and $p=0.003$, LRT, respectively). Interestingly, the interaction between collection time point and CD4% was not significant ($p=0.916$, LRT), indicating that these effects were independent of each other. Altogether, these results suggest that the level of CD4% is important for successful PCR results in dual, but not in single-infected individuals. Studies investigating different correlates during HIV disease progression have shown a strong relationship between the levels of CD4% and viral load[418,439]. Moreover, Madec *et al.* recently showed that HIV-1 plasma viral load continuously increases throughout the disease course and that the baseline proviral load was lower in individuals with long term non-progression[440]. Therefore, our hypothesis of differences in PCR success rate as an effect of different levels in viral load between single and dual-infected individuals is strengthened.

This doctoral dissertation consists of five papers that can be subdivided into three main topics: (1) Molecular description of the HIV-1 epidemic in Guinea-Bissau (**paper I**); (2) Dissection of the disease course in HIV-1 single and HIV-1 and HIV-2 dual infected individuals (**papers II and III**); and (3) HIV-1 coreceptor tropism in

HIV-1 single and HIV-1 and HIV-2 dual-infected individuals (**papers IV and V**).

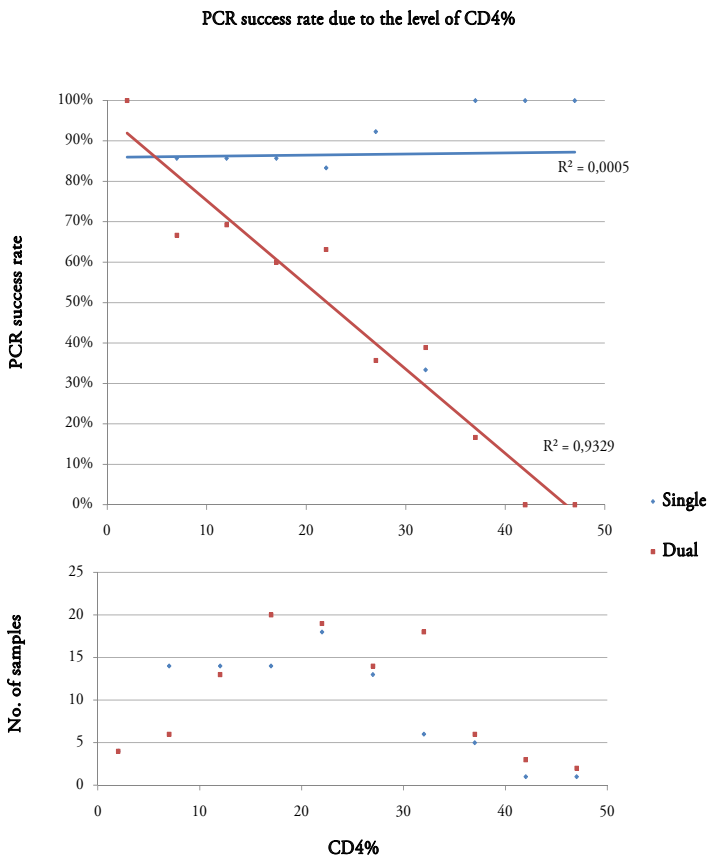


Figure 22. PCR success rate due to sample collection year. The top figure illustrates the PCR success rate over time (due to CD4%), whereas the bottom figure shows the total number of samples analyzed on different levels of CD4%. Ten groups were constructed based on different intervals of CD4% levels: 0-4%, 5-9%, 10-14%, 15-19%, 20-24%, 25-29%, 30-34%, 35-39%, 40-44%, and >44%. The corresponding PCR success rate and number of samples of these groups are indicated in the figure on mid-CD4% level of each interval. The blue line and dots represent data from HIV-1 single-infected individuals, whereas the red line and dots represent data from HIV-1 and HIV-2 dual-infected individuals.

Molecular description of the HIV-1 epidemic in Guinea-Bissau

Main Findings
- CRF02_AG is the most common form of HIV-1 in Guinea-Bissau, followed by subsubtype A3 (**paper I**).
- HIV-1 was introduced into Guinea-Bissau on at least six different occasions between 1976 and 1981 (**paper I**).
- The HIV-1 epidemic in Guinea-Bissau started in the capital Bissau, from where it dispersed out into more rural areas (**paper I**).
- The HIV-1 epidemic in Guinea-Bissau was connected to Mali and Cameroon (**paper I**).

We studied 82 HIV-1 infected individuals from Guinea-Bissau, and found that 57% were infected with CRF02_AG and 21% with subtype A. Detailed phylogenetic analysis of subtype A sequences showed that the vast majority of them (94%) belonged to the previously described subsubtype A3[281,441,442]. This subsubtype was originally described in Senegal (with a prevalence of 16%), and has been shown to circulate in several countries in West Africa. We also identified 13 recombinants between HIV-1 CRF02_AG and subsubtype A3. The recombination breakpoints were located in the gp120 C2 region in 12 of the 13 recombinants, a region previously described as a recombination hot spot for subtype B[443,444]. In addition, in the study by Meloni *et al*., where the subsubtype A3 was first described, two out of five putative HIV-1 A3 sequences were found to be recombinants between the CRF02_AG and the subsubtype A3[441]. One of these sequences displayed a recombination breakpoint in the gp120 C2 region.

To estimate the timing of the introduction of HIV-1 in Guinea-Bissau, we focused on the two dominating forms, the CRF02_AG and subsubtype A3. In the CRF02_AG phylogeny, we identified one major and four minor clusters of Guinea-Bissau-specific sequences, indicating a scenario of multiple introductions into Guinea-Bissau (**Fig. 23A**). The introduction of the major CRF02_AG cluster dated back to 1976, and was then followed by several introductions during the late 1970's and early 1980's. In the subsubtype A3 analysis, we identified one monophyletic cluster (1979), suggesting a major introduction followed by local epidemic spread (**Fig. 23B**). The timing of these introductions fits well with the results from early studies of HIV-1 seroprevalence in West Africa[445,446]. In West Africa, a number of armed conflicts took place during the 1960's and 1970's which led to huge socio-economic and political instabilities, and an

increased migration frequency in the region[447,448]. Guinea-Bissau gained independence from Portugal in 1974, after 11 years of war. One could speculate that the rapid increase in the number of effective infections could have been affected either during or shortly after these instabilities.

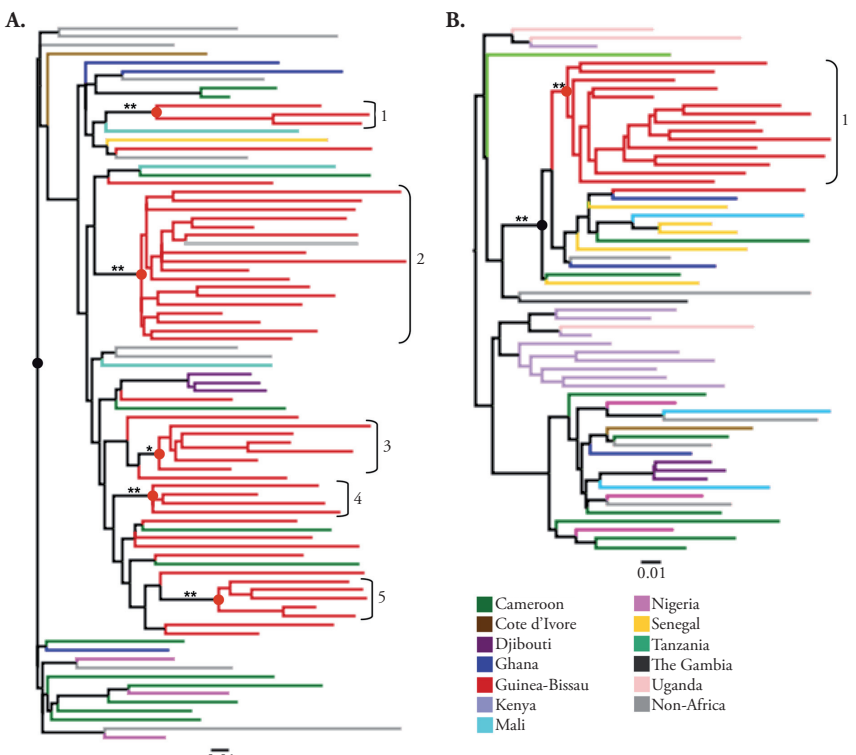


Figure 23. Determination of Guinea-Bissau-specific clusters in HIV-1 CRF02_AG and subsubtype A3 phylogenies. Maximum-likelihood trees were reconstructed based on the sequences from Guinea-Bissau together with reference sequences (A. CRF02_AG and B. subsubtype A3). The scale bar at the bottom of each tree represents 0.01 nucleotide substitutions per site. Asterisks along branches represent significant monophyletic clusters (**: Zero branch-length test p-values <0.001, and posterior probabilities >95% in corresponding Bayesian analysis. *: Zero branch-length test p-values <0.01, and posterior probabilities >90%). Statistically supported clusters are numbered according to appearance in the tree and the corresponding tMRCA are marked with filled circles (red circles = Guinea-Bissau-specific tMRCAs, and black circles = the tMRCA of the CRF02_AG or subsubtype A3). The colours represent the country or geographic region representing the origin of each tip in the phylogeny.

The phylogeographic analyses showed that HIV-1 was introduced into the urban centre (Bissau), and then spread throughout the country, a scenario that is similar to what has been reported for Cote d'Ivoire[449]. In addition, strains of the CRF02_AG epidemic, the dominant form of HIV-1 in Guinea-Bissau, started to migrate to rural areas almost 10 years before subsubtype A3, despite similar introduction dates. One might speculate that the subsubtype A3 epidemic grew in a small subpopulation in Guinea-Bissau that was not as prone to migrate as compared to the subpopulations infected with the CRF02_AG. Another possibility is that HIV-1 subsubtype A3 is less transmissible than the CRF02_AG. Competition assays in peripheral blood mononuclear cells *in vitro*, have shown that the CRF02_AG out-competes both its parental strains subtype A and G, respectively[262,263]. The migration analyses between Guinea-Bissau and other countries in West Africa showed connections to both Cameroon and Mali. Since the CRF02_AG is a recombinant between the "pure" subtypes A and G, which are prevalent in central West Africa, it is reasonable to believe that the CRF02_AG arose in this area of Africa. Therefore it is not surprising to find a shared common ancestor of the Guinea-Bissau CRF02_AG epidemic in Cameroon. The connection between Guinea-Bissau and Mali is also expected since the two countries have much business interests in common. Although the migrations established in **paper I** could be important in tracing disease spread connected to Guinea-Bissau, we emphasize that these are the migrations that we can extract from the analyzed sequences. It is both possible and likely that other migrations have existed in addition to those presented here, and an extended analysis could give a more detailed picture of HIV-1 disease spread inter-connected to Guinea-Bissau.

Dissection of the disease course in HIV-1 single and HIV-1 and HIV-2 dual-infected individuals

Main Findings

- Dual-infected individuals have a 46% lower mortality rate and a 53% longer progression-time to AIDS than single-infected individuals (**paper II**).
- HIV-2 seroprevalent individuals with a seroincident HIV-1 infection have a 46% lower mortality rate and a 90% longer progression-time to AIDS than HIV-1 single-infected individuals. Individuals with simultaneous HIV-1 and HIV-2 seroincidence displayed intermediate estimates in mortality rate and progression-time to AIDS (**paper III**).
- CD4+ T cell counts were consistently higher among dual-infected individuals compared to single-infected individuals at corresponding time-points after infection (**paper II**).
- HIV-2 seroprevalent individuals with a seroincident HIV-1 infection had consistently higher CD4+ T cell counts at corresponding time-points after infection compared to HIV-1 single-infected individuals, whereas individuals with simultaneous HIV-1 and HIV-2 seroincidence displayed intermediate estimates in mortality rate and progression-time to AIDS (**paper III**).
- CD8+ T cell counts were increasing at a faster rate in single than in dual-infected individuals (**paper II**).
- CD8+ T cell counts were increasing at a faster rate in HIV-2 seroprevalent individuals with a seroincident HIV-1 infection than in single-infected individuals, whereas individuals with simultaneous HIV-1 and HIV-2 seroincidence displayed intermediate estimates in mortality rate and progression-time to AIDS (**paper III**).
- HIV-1 diversity was consistently lower in dual compared to single-infected individuals at corresponding time-points after infection (**paper II**).
- The level of HIV-1 diversity could be an important predictor for disease progression rate to AIDS (**paper II**).

To investigate differences in disease progression rate and different cellular and molecular correlates, we studied a total of 223 HIV-1 seroincident individuals (189 were infected by HIV-1 only and 34 were infected with both HIV-1 and HIV-2) with long follow-up time. Analysis of reported mortality due to AIDS and progression-time to AIDS revealed remarkable differences. Dual-infected individuals had a median survival time of 140 months (95% CI 94-158 months) compared to 96 months (95%

CI 83-108 months) for single-infected individuals (p=0.047, Log-rank test (L-R)) (**Fig. 24A and paper II**).

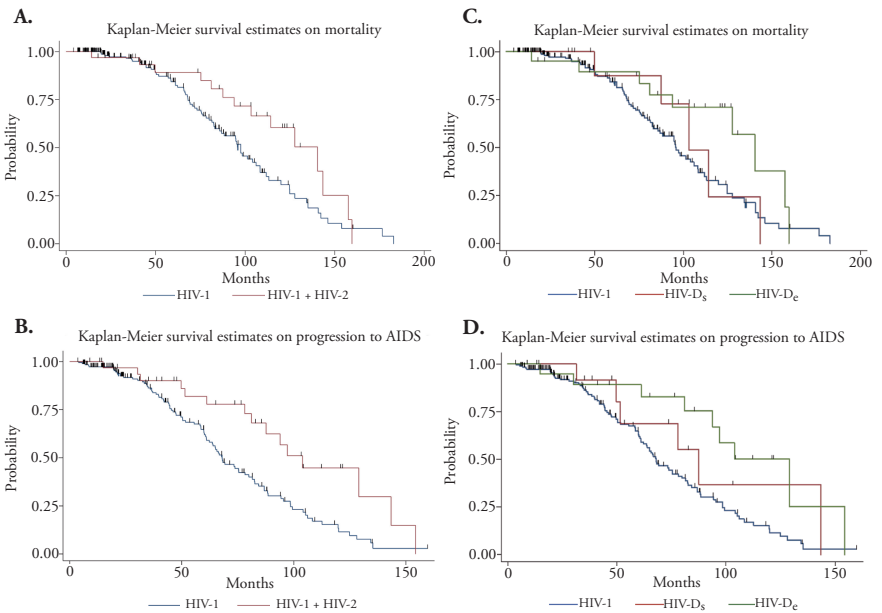


Figure 24. Analysis of survival and progression-time to AIDS. (A) Kaplan-Meier survival analysis for reported mortality due to AIDS among HIV-1 single and HIV-1 and HIV-2 dual-infected individuals. (B) Kaplan-Meier analysis for progression-time to AIDS among HIV-1 single and HIV-1 and HIV-2 dual-infected individuals. (C) Kaplan-Meier survival analysis for reported mortality due to AIDS among HIV-1 single and HIV-1 and HIV-2 dual-infected individuals stratified on earlier HIV-2 infection (preceding HIV-1 infection, HIV-D_e) or simultaneous HIV-2 infection (simultaneous HIV-1 and HIV-2 infection, HIV-D_s). (D) Kaplan-Meier analysis for progression-time to AIDS among HIV-1 single, HIV-D_e, and HIV-D_s. Cases where the exit criteria was not fulfilled were censored at their last clinical examination date in both analyses.

The adjusted mortality hazard ratio in dual compared to single-infected individuals was 0.66 (95% CI 0.49-0.90, p=0.008, Wald test) according to a Cox proportional hazards regression model controlling for gender and age at SC (four age groups: <25, 25-34, 35-44, and ≥45 years). The median time of progression to AIDS was 104 months (95% CI 81-143 months) in dual-infected individuals and 68 months (95% CI 61-80 months) in single-infected individuals (p= 0.004, L-R) (**Fig. 24B**). The adjusted hazard ratio of progression to AIDS in dual vs. single-infected individuals

was 0.63 (95% CI 0.47-0.84, p<0.001, Wald test) according to a Cox proportional hazards regression model controlling for gender and age at SC as in the mortality analysis.

Next, we stratified the group of dual infected individuals into two subgroups (HIV-D_e and HIV-D_s) (**paper III**). We found significantly lower mortality and a longer progression-time to AIDS in HIV-1 infected individuals with previous HIV-2 infection (HIV-D_e) compared to HIV-1 single infected individuals (p=0.048 and p=0.007, respectively, L-R) (**Fig. 24C and D**). Fitting a Cox regression model adjusting for age in two categories, the HIV-D_e group had both better survival (Hazard ratio 0.68 (95 % CI 0.47-0.97), p=0.036, Wald test), and longer time of progression time to AIDS than the HIV-1 single-infected group (Hazard ratio 0.60, (0.42-0.85), p = 0.004 Wald test). However, this difference was not found when comparing the group of individuals with estimated simultaneous HIV-1 and HIV-2 seroconversion (HIV-D_s) with individuals seroconverting to single HIV-1 or HIV-D_e infection. Thus, the HIV-D_s group had an intermediate position and was indistinguishable between the single HIV-1 and the HIV-D_e group, suggesting a variable outcome with respect to both AIDS progression time and mortality in this group.

To further investigate the differences in disease progression, we first examined levels of CD4% over time (**Table 9**). The rate in CD4% decline between single and dual infected individuals were similar with an average decline of 1.19% per year (p=0.439, 2-tailed Student's T-test (T-test)) (**paper II**). However, the level of CD4% at comparable time points after infection was higher in dual than in single-infected individuals (p=0.001, T-test). In concordance with the results of mortality rates and progression-time to AIDS, the stratified analysis showed that the HIV-D_e group had significantly higher CD4% levels compared to HIV-1 single infected individuals (p<0.001, T-test), whereas the HIV-D_s group did not differ from neither the single group (p=0.160, T-test) nor the HIV-D_e group (p=0.312, T-test) (**paper III**).

Second, we examined differences in levels of CD8% over time, and found that dual-infected individuals had a slower increase in CD8% levels compared to single-infected individuals (p=0.023, T-test) (**Table 9**). In contrast to the analysis of CD4%, there was no difference in levels of CD8% at the mean time point after infection (p=0.788, T-test) (**paper II**). The difference in increase of CD8% were even stronger when we compared the HIV-D_e group with the single group (p=0.001, T-test) (**paper III**). As in the previous analyses, the HIV-D_s group did not differ from the single group (p=0.936, T-test) when comparing the levels of increase in CD8%. In

Table 9. Estimates of CD4% and CD8% at respective mean sample time-point[1].

	HIV-1		HIV-1 + HIV-2		HIV-D_s		HIV-D_e	
	Mean	S.E	Mean	S.E	Mean	S.E.	Mean	S.E
CD4% decline (% per year)[2]	1.06	0.37	1.59	0.41	1.82	1.12	1.50	0.38
CD4% level, 56 months post HIV-1 SC[3]	18.04	0.92	24.64	1.62	22.35	2.63	26.17	2.10
CD8% increase (% per year)[4]	4.62	0.80	2.15	0.70	4.81	1.32	0.91	0.62
CD8% level, 50 months post HIV-1 SC[5]	48.44	1.39	47.67	2.84	48.40	4.48	45.65	3.43

[1]To adjust for differences in sampling time, individual values were extrapolated to the global average sampling time (56 and 50 months post HIV-1 seroconversion for CD4% and CD8%, respectively) for all individuals using the average decrease in CD4% level over time (1.59.% per year) or the group-specific increase in CD8% (as indicated in the table). P-values refer to 2-tailed Student's T-test and the uncertainty of estimated means is given by the standard error (S.E.). [2]Comparisons between groups, 2-tailed Student's T-test: HIV-1 vs. HIV-1+ HIV-2, p=0.439; HIV-D_s vs. HIV-1, p=0.539; HIV-D_e vs. HIV-1, p=0.580; HIV-D_e vs. HIV-D_s, p=0.728. [3]Comparisons between groups, 2-tailed Student's T-test: HIV-1 vs. HIV-1+ HIV-2, p=0.001; HIV-D_s vs. HIV-1, p=0.160; HIV-D_e vs. HIV-1, p<0.001; HIV-D_e vs. HIV-D_s, p=0.312. [4]Comparisons between groups, 2-tailed Student's T-test: HIV-1 vs. HIV-1+ HIV-2, p=0.023; HIV-D_s vs. HIV-1, p=0.936; HIV-D_e vs. HIV-1, p=0.001; HIV-D_e vs. HIV-D_s, p=0.006. [5]Comparisons between groups, 2-tailed Student's T-test: HIV-1 vs. HIV-1+ HIV-2, p=0.788; HIV-D_s vs. HIV-1, p=0.993; HIV-D_e vs. HIV-1, p=0.391; HIV-D_e vs. HIV-D_s, p=0.646.

contrast, in this analysis we did find a difference between the two stratified groups of dual-infected individuals (p=0.006, T-test), suggesting that the HIV-D_s group is more similar to dual-infected individuals. Fourth, with the insight that the rate of $CD8^+$ T cell elevation differed between single and dual-infected individuals, we set out to investigate if levels of immune activation markers also differed between the two groups. Since availability of stored peripheral blood lymphocytes was limited, levels of beta-2 microglobulin (b2m) and neopterin in plasma were analysed as markers of immune activation. However, no differences in plasma levels or kinetics of b2m or neopterin were found between single and dual-infected individuals, instead, considerable variation between different individuals was noted (**paper II**). Finally, we investigated HIV-1 diversity and divergence patterns in a subset of 20 single and 12 dual-infected individuals. Phylogenetic trees were inferred using both maximum-likelihood (ML) and Bayesian methods[242,426]. Since analyses based on the two inference methods showed high concordance, only the ML estimates are presented throughout the following discussion. The average increase in HIV-1 sequence diversity over time was similar in the groups with an average of $1.75x10^{-3}$ substitutions per site per year (p=0.812, T-test), whereas the diversity at comparable time points after infection was significantly lower in dual than in single-infected individuals (p=0.014, T-test) (**paper II**). The "diversity threshold theory" assumes that AIDS develops when

diversity exceeds a critical threshold that varies individually[371,372]. As reported above, the average progression-time to AIDS was 68 months in single-infected individuals. At this time point the estimated mean diversity was $13.52x10^{-3}$ substitutions per site for single-infected individuals. Dual-infected individuals were estimated to reach the same mean diversity after 105 months, close to the average observed progression-time to AIDS of 104 months for this group. Thus, the mean diversity threshold is almost identical for dual and single-infected individuals. However, the time to reach the threshold is significantly different (**Fig. 25**). Despite detailed evolutionary analyses,

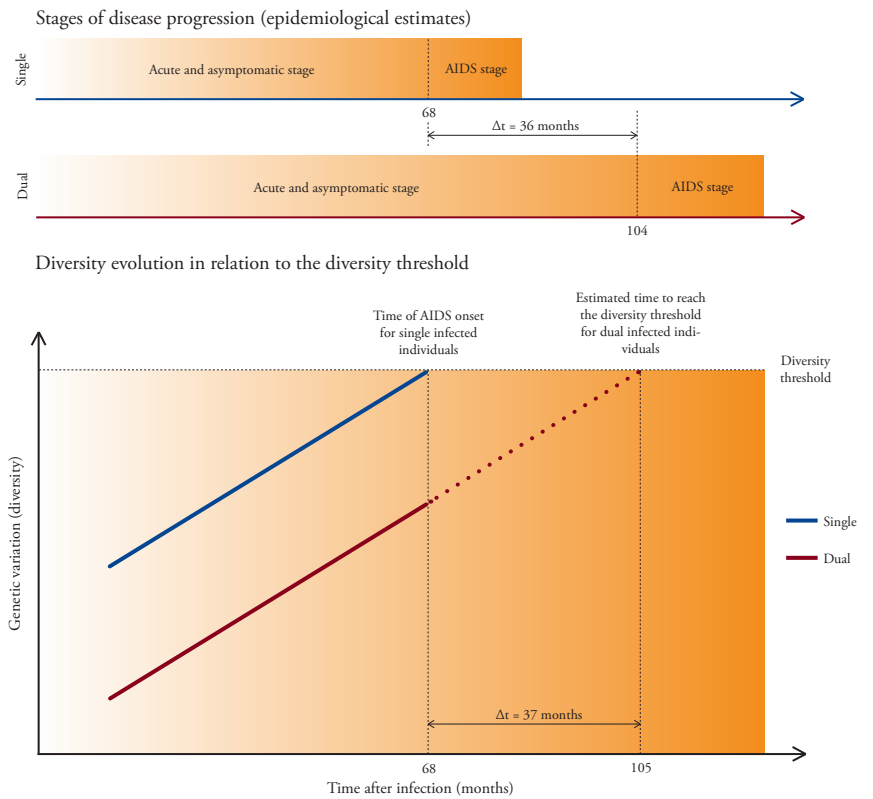


Figure 25. Correlation between disease progression and HIV-1 diversity evolution. Schematic illustration. The epidemiological estimate of progression-time to AIDS for the HIV-1 single-infected individuals (68 months) was used to determine the diversity threshold for development of AIDS. The time to reach that threshold for dual-infected individuals was determined by using the mean diversity rate to extrapolate to the threshold diversity level. Both maximum likelihood (104.93 months) and bayesian estimates (104.95 months) were close to the epidemiological estimate of AIDS onset (104 months) (**paper II**).

we found no significant differences between the single and dual-infected groups in HIV-1 divergence rates, HIV-1 rates of synonymous substitutions, or in HIV-1 rates of non-synonymous substitutions (**paper II**).

In strong support of our population-based results, experimental studies using the macaque model have shown inhibition against both immunosuppression and SIV-induced disease as a result of contemporaneous HIV-2 infection[296,297]. The mechanisms behind this protection could have several alternative explanations. The fact that the HIV-1 diversity increases at a similar rate in single and dual-infected individuals, but that the diversity at a given time point is lower in the dual-infected individuals, could be due to inhibitory effects early in the HIV-1 infection (**paper II**). This would result in a lower initial diversity and a longer asymptomatic stage before the "diversity threshold" is reached. The hypothesis of inhibitory effects early in infection is also supported by the result that CD4% decreases at a similar rate, but at different levels, suggesting different set point levels in CD4% in early HIV-1 infection. Moreover, the stratification consistently showed an even more pronounced difference between single and dual-infected individuals when only the HIV-D_e group was analyzed (**paper III**). The HIV-D_s group could be separated from neither the single infected group nor the HIV-D_e group. These results suggest that effects due to previous HIV-2 infection could be important in subsequent control of HIV-1 infection. Based on these results, it is tempting to speculate that HIV-2 might have a priming effect on the immune system and thereby prepares the host for the more pathogenic HIV-1 infection.

Alternatively, HIV-2 could continuously alter the expression of cellular factors that affect the susceptibility of the uninfected cellular environment *in trans*. *In vitro* studies have shown that HIV-2 infection generates higher levels of β-chemokines (the natural ligands of the HIV coreceptor CCR5) in peripheral blood mononuclear cells, and that this can inhibit HIV-1 infection and replication[306-309]. In addition, Zheng *et al.* found high frequencies of cross-reactivity between samples of individuals with single HIV-1 or HIV-2 infection when investigating heterologous T-cell responses[310]. They also found that HIV-1 single-infected individuals with the ability to respond to the HIV-2 Gag protein had lower HIV-1 plasma viral loads than those without this cross-reactivity. A second study by this group showed that this was evident also in dual-infected individuals[311]. Moreover, antibodies elicited by HIV-2 infections that can cross-neutralize HIV-1 have been described[312]. Hence, humoral HIV-2 immune responses could also play a role in controlling HIV-1 in dual-infected individuals.

Several cross-sectional reports suggesting excessive immune activation as one of the

major forces driving the HIV-1 pathogenesis have been presented[70,450,451]. Studies trying to dissect immune activation and couple it to different disease progressor groups have found conflicting results[452-454]. However, immune activation is a relatively undefined term and can be measured by many different markers[415,453]. It is reasonable to believe that multiple activation processes are working both in parallel and interactively, resulting in a multifaceted immune activation on several different levels, including direct infection, cytokine production (with both direct and bystander effects), microbial translocation effects, co-infections and more (reviewed in[455]). In our material, we did not find any difference in systemic immune activation between single and dual-infected individuals (**paper II**). It has been suggested that levels of soluble immune activation markers could be influenced by different endemic parasite and bacterial diseases present in African populations[453]. Moreover, since the vast majority of conclusions regarding immune activation markers are based on cross-sectional studies, general patterns of intrapatient temporal fluctuations of those markers remains to be described. It is possible that the length of the investigated window was too short to discover any putative difference in increase of b2m or neopterin between single and dual-infected individuals (the mean difference between time-point one and two was 26 months for single and 27 months for dual-infected individuals). Another possibility is that the window was too close to seroconversion date (first mean time-point was 29 months and 17 months after infection for single and dual-infected individuals, respectively). Finally, b2m and neopterin could of course be inadequate in terms of reflecting the difference in disease progression rate between single and dual-infected populations.

Despite the lack of differences in levels or kinetics of soluble immune activation markers, the reduced elevation in CD8% among dual-infected individuals suggests that alteration in cellular immune activation may contribute to the disease outcome of these individuals (**paper II**). There have been several reports of a positive correlation between $CD8^+$ T cell activation and HIV-1 disease progression rate[456-458]. Cavaleiro *et al.* investigated the immunological effects of HIV-2 envelope protein gp105 on anti-CD3-stimulated peripheral blood mononuclear cells (PBMC), and found that HIV-2 gp105 had a higher inhibitory effect against T cell proliferation than HIV-1 gp120[459]. In addition, this inhibition affected both $CD4^+$ and $CD8^+$ T cells, suggesting that different levels of immunosuppressive properties of the HIV envelope proteins could be beneficial to the host by interfering with the heightened state of immune activation. Another possible explanation for the slower progression rate to AIDS in dual-infected individuals could perhaps be coupled to findings showing a Nef-mediated TCR-CD3 downmodulation resulting in lower general immune activation and activation-

induced cell death. This mechanism operates in the majority of primate lentiviruses (including HIV-2), but not in HIV-1[460].

Contrasting results regarding viral load (VL) comparisons between single and dual-infected individuals have been presented[266,271,304,305]. However, these data are difficult to interpret due to unknown SC dates, differences in disease stage, and limited follow-up. Since VL measurements have not been included as a standard procedure in Guinea-Bissau, we were unable to include VL data in our analyses. However, recent studies have suggested that there may be a positive correlation between HIV-1 diversity and VL[373,461]. Diversity has also been positively correlated with HIV-1 replication efficiency and rate of progression to AIDS[264,377]. These results are in line with our results indicating that the HIV-1 diversity at a given time-point is higher in single than in dual-infected individuals. The constant increase in viral diversity in the asymptomatic stage suggests that the immune response is not broad enough to target all co-existing viral variants. Instead, the most common variants are targeted and eradicated, whereas less replication-competent variants are conserved. It has been hypothesized that the increasing viral diversity ultimately results in viral variants with critical epitopes outside the T-cell repertoire, resulting in immune failure and development of AIDS[371,372,461]. This hypothesis supports the "diversity threshold theory" as an explanatory model of HIV-1 diversity evolution, and may serve as a reasonable explanation for the differences that we observed between single and dual-infected individuals.

Taken together, the results of **paper II and III** clearly show that HIV-1 and HIV-2 dual-infected individuals have a slower disease progression rate than HIV-1 single-infected individuals. Furthermore, the combined results from the cellular and molecular analyses suggest that the differences in HIV-1 disease progression rate are reflected in both the establishment of the infection and throughout the disease course.

HIV-1 coreceptor tropism in HIV-1 single and HIV-1 and HIV-2 dual-infected individuals

Main Findings

- The recombinant virus phenotypic assay accurately identified HIV-1 coreceptor tropism of subtype A and CRF02_AG (**paper IV**).
- HIV-1 CRF02_AG CXCR4 tropism was frequent among individuals in late-stage disease (86%), and showed an increasing pattern over time (**paper IV**).
- The total number of charged amino acids in combination with net charge of the HIV-1 gp120 V3 region may be useful for development of future genotypic tools for prediction of HIV-1 CRF02_AG coreceptor tropism (**paper IV**).
- All major HIV-1 subtypes (A, B, and D) and CRFs (01_AE and 02_AG) showed high frequencies of CXCR4 tropism in late-stage disease (60-77%), except for subtype C (15%) (**paper IV**).
- HIV-1 CXCR4-tropism is more common in HIV-1 single (79%) than in HIV-1 and HIV-2 dual-infected individuals (35%) (**paper V**).
- The level of diversity was similar in single and dual-infected individuals in late-stage disease and was highly variable between different individuals (paper V).

HIV-1 subtype determination was done by amplifying and sequencing the envelope gp120 V3 region of 29 infected individuals in late-stage disease from Guinea-Bissau, West Africa. Twenty-eight of the individuals had HIV-1 of subtype A or CRF02_AG, confirming previous results that these are the dominating HIV-1 forms in Guinea-Bissau[462] (**paper I**). The remaining study subject was infected with HIV-1 subtype C. Detailed analysis of the six patients infected with subtype A showed close genetic relationship to the previously described subsubtype A3. In **paper V**, we cloned and sequenced the gp120 V1-V3 region for the diversity analysis. Detailed recombination analysis of HIV-1 sequences from these 28 individuals revealed that two of the V3 sequences determined as subsubtype A3 actually were recombinant forms of the CRF02_AG and subsubtype A3, with the V3 region derived from subsubtype A3 (from patients DL3018 and DL3071, **Table 7**, see *Materials & Methods*). Reversely, we also found that four of the V3 sequences determined as CRF02_AG instead were CRF02_AG/A3 recombinants, with the V3 region derived from CRF02_AG (from patients DL3037, DL3039, DL3234 and DL3721, **Table 7**, see *Materials & Methods*). No difference in distribution of these CRF02_AG/A3 recombinants between the CRF02_AG and the subsubtype A3 groups were found (p=0.581, two-tailed Fisher's

exact test). Chimeric viruses from all these CRF02_AG/A3 recombinants were of R5X4 phenotype (**paper IV**). Even though there have been some suggestions of influence on coreceptor tropism of other regions of the gp120, the V3 region has been shown to be the principal determinant[463,464]. However, most studies have focused on HIV-1 subtype B and C, and there is no clear evidence that the V3 region has the same impact on coreceptor interaction among other subtypes. On the other hand, there is no *a priori* reason to believe that the principal determinants of coreceptor tropism for subtype A or the CRF02_AG should differ from those in subtype B or C.

In **paper IV**, we showed that the determinants for coreceptor tropism of the CRF02_AG and the subsubtype A3 lies within the gp120 V1-V3 region, and that the TRT assay accurately determines coreceptor tropism for these subtypes[419]. A control panel of 11 HIV-1 subtype A and CRF02_AG isolates was used to show that the, for subtype B, well-characterized TRT assay is accurate also for subtype A and CRF02_AG. Next, we examined the coreceptor phenotype of HIV-1 in 28 subtype A or CRF02_AG infected individuals in late-stage disease. CXCR4-using viral populations were found in as much as 79% of the analyzed samples (50% of the subsubtype A3 and 86% of the CRF02_AG), demonstrating the importance of analyzing samples from patients in late-stage disease when investigating HIV-1 subtype-specific predisposal for CXCR4 tropism.

To further the view of our results, we performed a literature review of published data on subtype-specific coreceptor tropism in late-stage disease, which, to our knowledge, has not been presented before. The analysis included 498 patient-specific HIV-1 samples of six different subtypes, sampled over more than 20 years (1988-2008), and revealed a high frequency of HIV-1 R5X4 or X4 populations in late stage disease among all analyzed subtypes, except for subtype C (**Fig. 26**). Only 15% of the individuals infected with HIV-1 subtype C had CXCR4-using populations, compared to 66% (60%-77%) in individuals infected with HIV-1 of non-subtype C ($p<0.001$, two-tailed Fisher's exact test).

In order to investigate genetic traits that could be used to predict coreceptor phenotype for viruses of HIV-1 CRF02_AG, we analyzed a dataset of 111 sequences with known coreceptor tropism (75 R5 and 36 R5X4 or X4 sequences) (**paper IV**). We also compared our results with several established genotypic rules and bioinformatic tools (**Table 9**). Most of the rules and tools had high specificity (the fraction of predicted R5 sequences among the sequences from viruses phenotyped as CCR5-using only) but low (≤50%) sensitivity (the fraction of predicted X4 sequences among the sequences from

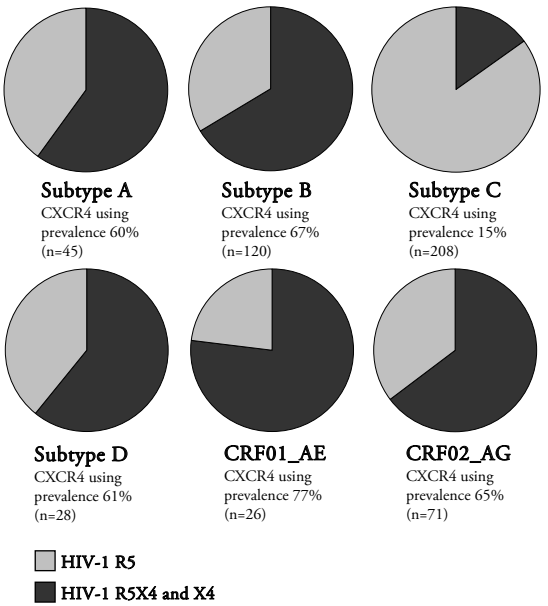


Figure 26. Prevalence of HIV-1 CXCR4-tropism in late-stage disease in different subtypes. Combined results of new data presented in **paper IV** and data from the literature analysis, showing the prevalence of HIV-1 CXCR4-tropism in late stage disease. The number of individuals used in each diagram is specified within brackets. The figure was modified with permission from **paper IV**.

viruses phenotyped as CXCR4-using) (**Table 9**). Currently, coreceptor antagonists are only available for patients infected with viruses unable to utilize CXCR4 as a coreceptor. Therefore, it is essential for a genotypic rule to be highly sensitive if it should be used in a clinical setting. We suggested a combined rule, based on the total number of charged amino acids in the V3 region and the previously described net charge rule[183]. In terms of sensitivity, this rule performed better than all of the analyzed rules and bioinformatic tools, without losing too much in specificity.

Recently, studies of evolving patterns of different viral traits on the population level have been presented[384,385]. In terms of evolution of HIV-1 coreceptor phenotype, Connell *et al.* analyzed the results of 19 subtype C isolates from 2005, and found a higher prevalence (30%) of CXCR4 tropism than has been shown in earlier studies of subtype C[205]. They suggested that HIV-1 subtype C might be an evolving epidemic, showing an increasing prevalence of CXCR4 phenotype over time in South Africa. In **paper IV**, we performed a direct comparison by dividing our 22 CRF02_AG

Table 9. Comparison of different genotypic rules and bioinformatic tools for prediction of HIV-1 coreceptor tropism based on the HIV-1 V3 amino acid sequence. The table was adopted with permission from **paper IV**.

Patient No.	Sex[1]	CD4%[2]	CD4tot[3]	CDC[4]	WHO[5]	Subtype	Tropism	Sample year
DL1996H	M	5	157	B	3	CRF02_AG	R5X4	2000
DL2089J	M	9	59	B	3	CRF02_AG	R5X4	2003
DL2096F	M	5	22	C	4	C	N/A[6]	2003
DL2249I	M	2	21	C	4	CRF02_AG	R5X4	2004
DL2339E	M	12	178	B	3	CRF02_AG	R5X4	2003
DL2365K	M	9	133	B	3	A3	R5	2006
DL2391G	M	5	N/A[6]	B	2	CRF02_AG	R5	2000
DL2401M	M	11	141	B	3	CRF02_AG	R5X4	2004
DL2713H	M	N/A[6]	N/A[6]	B	2	CRF02_AG	R5X4	2007
DL2846I	F	N/A[6]	N/A[6]	B	3	A3	R5X4	2005
DL2853E	M	11	137	A	1	CRF02_AG	R5	1998
DL2920H	M	11	126	B	3	CRF02_AG	X4	2004
DL3018H	M	N/A[6]	N/A[6]	B	3	A3	R5X4	2006
DL3037E	M	3	74	B	3	CRF02_AG	R5X4	2005
DL3039G	F	7	148	A	2	CRF02_AG	R5X4	2006
DL3071H	F	20	123	B	3	A3	R5X4	2005
DL3087E	M	4	62	B	2	CRF02_AG	R5X4	2001
DL3098I	F	14	426	N/A[6]	N/A[6]	CRF02_AG	R5X4	2007
DL3169F	M	9	315	B	3	CRF02_AG	R5X4	2004
DL3170F	M	8	65	B	2	CRF02_AG	R5X4	2000
DL3234J	M	10	216	A	2	CRF02_AG	R5X4	2006
DL3312E	M	2	36	C	4	CRF02_AG	R5X4	1998
DL3633G	F	8	112	C	4	CRF02_AG	R5X4	2003
DL3721C	M	11	257	A	1	CRF02_AG	R5X4	1997
DL3733G	M	19	137	B	3	CRF02_AG	R5X4	2004
DL4248G	F	13	159	B	3	A3	R5	2005
DL4477D	M	14	141	B	3	CRF02_AG	R5	2001
DL4525G	M	13	372	B	3	A3	R5	2006
DL4632E	F	9	77	B	3	CRF02_AG	R5X4	2003

[1]M = male, F = female
[2]CD4+ T cell percentage among all T cells
[3]CD4+ T cell count per microliter among all T cells
[4]Clinical category of the patient, as defined by the CDC, at the sample time point
[5]Clinical category of the patient, as defined by the WHO, at the sample time point
[6]N/A = not analyzed
[7]Sample included in the study based on previous examinations of CD4+ T cell count and percentage

samples into two groups: Samples from 1997 to 2001, and samples from 2003 to 2007. We found a significant difference between the groups, suggesting a similar kind of evolving epidemic for CRF02_AG in Guinea-Bissau. We also analyzed our data together with data from Genbank of patients with known clinical parameters and HIV-1 coreceptor tropism. The result of an evolving epidemic of HIV-1 CRF02_AG in terms of coreceptor tropism was further strengthened by this analysis. Competition assays between HIV-1 R5 and X4 viruses have shown that X4 viruses in general outcompete R5 viruses due to both higher replication kinetics and higher CXCR4 than CCR5 expression in PBMC[261]. Moreover, it has been shown that sequence changes occur at a rate of 1% per year in HIV-1 *env*, illustrating the constant evolution of HIV-1 on the genetic level[365]. Further studies (based on larger sample sizes than studied in **paper IV** and by Connell *et al.*) are needed to investigate if HIV-1 is

evolving towards a more predisposed state of changing into CXCR4 phenotype on a population level. A confirmation in larger cohorts may have implications for viral transmission, pathogenesis and disease progression.

The data from the literature review also allowed us to investigate if we could confirm the results of Connell *et al.* of an evolving epidemic in South Africa, and if this could be seen for subtype C in general[205]. We divided the subtype C data set (208 patients, samples collected in Cameroon, Ethiopia, India, Malawi, South Africa, Sweden and Zimbabwe) into an early group (samples before the year of 2000) and a late group (samples after 2000) (**paper IV**). In the early group from South Africa eight of 46 patients (17%) had CXCR4-using viruses, whereas the corresponding number in the late group was 11 of 36 (31%) ($p=0.19$, two-tailed Fisher's exact test). Connell *et al.* reported that 30% of their isolates were able to use CXCR4 and compared this to previous studies from South Africa showing no syncytium-inducing (SI) capacity of HIV-1 isolates collected during the 1980's, whereas 10-17% of the studied isolates had SI capacity during the 1990's. However, no statistical evaluation was performed in their study. In the early group of our complete subtype C data set 11 of 145 patients (8%) had CXCR4-using viruses, whereas the corresponding number in the late group was 21 of 63 (33%) ($p<0.001$, two-tailed Fisher's exact test). Our results thus confirm the numbers presented by Connell *et al.* in South Africa (even though the difference between the early and the late group was not statistically significant in their study), and further strengthens the concern of an evolving HIV-1 subtype C-epidemic on the population level.

In **paper V**, we used the established phenotypic assay to determine the coreceptor tropism of HIV-1 V1-V3 sequences from 17 dual-infected individuals in late-stage disease. To be able to compare the results with **paper IV**, we investigated potential confounders between the two datasets. No difference in distributions of (1) different HIV-1 variants (subsubtype A1, subsubtype A3, CRF02_AG, and CRF02_AG/A3 recombinants) ($p=0.378$, $p=0.143$, $p=1.000$, and $p=0.216$, respectively, two-tailed Fisher's exact test), (2) gender ($p=0.193$, two-tailed Fisher's exact test), or (3) levels of CD4% ($p=0.193$, two-tailed Fisher's exact test) were found. Among the dual-infected individuals, 35% (6/17) were CXCR4-tropic, which should be compared to 79% (22/28) among the single-infected individuals ($p=0.005$, two-tailed Fisher's exact test) (**paper V**).

To further explore differences in sequence diversity between single and dual-infected individuals, we cloned and sequenced the HIV-1 V1-V3 region of the investigated

individuals (mean 8.22 clones/individual) (**paper V**). Several longitudinal studies have presented evidence of a positive correlation between HIV-1 diversity and the time after infection, during the asymptomatic stage[365,371,376,378]. Close to the onset of AIDS, it has been suggested that the diversity generally stabilizes, or in some cases even decreases[365]. Phylogenetic analysis revealed that there was no difference in diversity during late-stage disease between single and dual-infected individuals ($p=0.332$). The "diversity threshold theory" predicts that the total virus population grows unboundedly beyond this threshold[465]. Furthermore, once the threshold has been exceeded, selection would favour strains with high replication rate, even though slower growing strains also would expand their population sizes. This would most likely result in rapid evolution with a fluctuating and broad spectrum of HIV-1 diversity between different individuals, as seen in the diversity range among both single ($0.7\text{-}46.5\text{x}10^{-3}$ substitutions per site) and dual-infected individuals ($0.7\text{-}45.5\text{x}10^{-3}$ substitutions per site).

Although other hypotheses exist, several lines of evidence suggest that HIV-1 CXCR4-tropic viruses evolve from pre-existing CCR5-tropic viruses during the natural course of infection[135]. It has been hypothesized that up-regulation of β-chemokines could favor HIV-1 to switch from CCR5 to CXCR4-use in dual-infected individuals at a higher rate than in single-infected individuals. In our material, we found the opposite, suggesting that the potential *in vivo* effect of increased β-chemokine levels does not result in higher levels of HIV-1 CXCR4-tropism in dual-infected individuals. Mathematical modelling of evolution in HIV-1 coreceptor tropism has shown that the development of CXCR4-using strains is favoured by a weak immune system and depends on the level of specific antiviral responses against these strains[466]. In light of the results from **paper II and III** (showing lower levels of CD4% in dual-infected individuals than in single-infected individuals), this model seems plausible. As suggested in previous sections, up-regulation of β-chemokines might instead act as an inhibitory factor against HIV-1 replication and thereby alter the probability of development of CXCR4-tropism.

MAIN CONCLUSIONS AND FUTURE PERSPECTIVES OF THIS DOCTORAL DISSERTATION

Further insights in the epidemiology of HIV-1 in Guinea-Bissau and West Africa will certainly be gained by analyzes of larger amount of sequences. It would be particularly interesting to dissect the HIV-1 epidemic in West Africa from a historical perspective, connecting the potential colonial influence on disease spread, demography and migration history of HIV-1 in this region. There are a number of obstacles connected to such study, because there is (1) limited sequence data from many regions with colonial linkage, and (2) a general lack of information about, for example, sample time point or geographical origin of sequences deposited in Genbank. A collaborative network of research teams connected to different African regions could overcome such obstacles.

In **paper II and III**, we discuss several possible mechanisms underlying the inhibitory effect exerted by HIV-2 on HIV-1 disease progression. In order to dissect these mechanisms in detail, several different approaches could be used. First, sequential and well-preserved blood plasma and blood cell samples from dual-infected individuals could be analyzed, and compared to equivalent samples from single-infected individuals, for a variety of markers for immune activation or disease progression. It would also be interesting to investigate host-genetic differences like HLA expression in different HIV-infected populations. However, it might be important to choose such markers carefully due to the generally high level of endemic and parasitic disease load in the African population. Second, the macaque model could be useful in a variety of detailed studies, such as studies of cellular immune markers, viral sequence evolution or host-specific genetic differences. Moreover, the macaque model could be especially important in studies related to the establishment of infection. Third, *ex vivo* or *in vitro* models could be used to modulate a variety of parameters such as infection order of HIV-1 and HIV-2, the timing of the different infections, or the HIV infectious doses needed for inhibition to occur. Finally, in all of these suggested approaches, expression arrays or large-scale sequencing approaches could be useful in discovering differences of cellular expression between single and dual infected individuals.

It would also be interesting to investigate if the differences in disease progression rate between single and dual-infected individuals are reflected in terms of HIV-1 intrapatient recombination. Based on exploratory analysis of the sequences used in this doctoral dissertation, it might be hypothesized that dual-infected individuals display lower amount of HIV-1 recombinants. However, within-host recombination analysis is not straight forward and different state-of-the-art bioinformatic tools might be used to study this observation further. Verifying this preliminary result by

such methodology could give important clues into mechanisms responsible for the lower disease progression rate seen among dual infected individuals.

Finally, further in-depth analyses of intra-patient HIV evolution are essential in understanding HIV disease pathogenesis. Most studies that have been presented to date have been based on sequences from relatively few individuals. Future studies need to involve a larger number of individuals, and more importantly, such studies need to compare distinctive progressor groups to detect clear correlates of HIV disease progression rate.

ACKNOWLEDGEMENTS OF THE PEOPLE INVOLVED IN THIS DOCTORAL DISSERTATION

Det är så klart jättemånga personer som på ett eller annat sätt har bidragit till den här avhandlingen. Det finns några personer som jag särskilt skulle vilja tacka:

Patrik Medstrand, min huvudhandledare. Du har utan konkurrens varit den mest betydelsefulla personen för mig i mitt arbete med den här avhandlingen. Du ska veta att jag är enormt tacksam för att jag fick just dig som handledare och du är en förebild på många sätt i ditt ledarskap - alltid lugn och trygg. Tack för att du har låtit mig ta ansvar och gå mina egna vägar och testa mina gränser. Du har också alltid ställt upp för mig, både i vått och torrt. Vi har till och med varit på IKEA och handlat vinglas ihop. Jag kommer heller aldrig glömma sillrensningen i ditt kök eller grillfesten på er utegård. Ser fram emot att jobba ihop under många år framöver.

Eva Maria Fenyö, min bihandledare. Om inte du hade flyttat till Lund någon gång omkring millennieskiftet hade jag nog inte fått förmånen att jobba med HIV på ett så berikande sätt som jag har gjort. Tack för att du en gång välkomnade mig in i din grupp och lät mig få pröva mina vingar inom det forskningsfält som jag tyckte lät mest intressant under min tid på biomedicinutbildningen. Det är också en stor förmån att få se hur aktad och respekterad du är som forskare runt om i världen.

Anders Kvist, min bihandledare. Från första dagen jag steg in på labbet på BMC har du varit ”the go-to-guy” i allehanda frågor rörande datorer och statistik. Men det är inte allt, jag är grymt imponerad av hur bred din kunskapsbas är och hur snabbt du alltid sätter dig in i olika problem. Du har verkligen varit en klippa och en mycket viktig stöttepelare under min doktorandtid. Tack också för alla möjliga och omöjliga samtalsämnen vi har haft under alla luncher nere på ”the Vault”. Jag säger bara: Två rätter. Grön curry och kycklingspett med extra jordnötssås. En klassiker!

Mattias Mild, min kära forskarbroder. Du ska veta att du alltid kommer att ha en särskild plats i mitt hjärta. Det var du som ledsagade mig när jag tog mina första stapplande steg på forskarbanan. Vi fann varandra...eller vi och vi förresten, det var väl jag som fann dig, och sedan hakade du bara på...som vanligt... ;-). Hur som helst, idag är du en av mina absolut närmaste och bästa vänner och det är få människor som kan få mig att skratta lika mycket som du! Du är en sann vän som man alltid kan lita på och jag ser fram emot att dela många flaskor vin och whiskey (och för all del även ale som du är så inne på för tillfället) tillsammans i framtiden.

Fredrik Månsson. De här åren har varit en väldigt spännande resa i många avseenden och du har varit en central kugge i alla mina projekt. Vi har suttit ett oräkneligt antal nätter och bläddrat i polispärmar både i Bissau och på infektion i Malmö. Men trots dåligt med sömn och tighta deadlines har du alltid lyckats hålla humöret uppe. Känns som att vi trots allt lyckades ro det mesta av det vi hade tänkt i land till slut. Snart är det dags för dig också! :-)

Hans Norrgren. Tack för alla snabba svar på frågor om poliser, serologi, guineanska folkgrupper och regioner, samt om olika publikationer med anknytning till dubbelinfektion. Tack också för att du räddade mitt bagage och mitt portvin i Dakar.

Alla nuvarande och före detta kollegor i Patriks grupp. Angelica Palm, Mikael Ingman, Wilma Martinez-Arias, Daniel Svenback och Mahmudul Hasan. Tack för att ni förgyller (har förgyllt) varje dag på jobbet och gör BMC till en trevligare arbetsplats. Tack också för allt stöd jag har fått under de här åren!

Alla nuvarande och före detta kollegor på virologen. Marianne Jansson, Marie Borggren, Johanna Repits, Salma Nowroozalizadeh, Gülsen Özkaya Sahin, Enas Sheik-Khalil, Birgitta Holmgren, Anna Laurén, Ingrid Karlsson, Ulf Karlsson, Elzbieta Vincic, Monica Öberg, Hugo Johansson, Åsa Hellberg, Julia Westman och Sisay Alemayehu Abayneh. Tack för all hjälp (!), för alla intressanta journal clubs, morgonpresentationer och diskussioner under åren. Men framför allt tack för att ni alltid finns i närheten med ett vänligt och glatt ord.

Tack till övriga personer i Guinea-Bissau-gruppen. Sören Andersson, Helen Linder, Magnus Unemo och Jan Albert. Tack för alla diskussioner och trevliga sammankomster i Lund och Stockholm.

Tack Irene Silhammar på SMI för all hjälp med hantering av polisprover. Tack också till Kaisa Aperia på SMI för hjälp med uppodling av virusstockar.

Alla som är involverade i poliskohorten i Guinea-Bissau. Antonio J. Biague, Zacarias J. da Silva, Babetida N'Buna, Aquilina Sambu, Eusebio Ieme, Isabel da Costa, Jacqueline, Ana Monteiro, Cidia Camara, Braima Dabo (deceased), Carla Pereira, Julieta Pinto Delgado, Leonvengilda Fernandes Mendes och Ansu Biai.

Alla på B13 (utvecklingsbiologi) och C13 (SCIBLU) för hjälp med allt från geler till biomek-körningar. Tack också för alla skratt och för all motion som jag fått genom all

innebandy och badminton!

Tack också till alla mina vänner. Ni ska veta att jag känner mig otroligt priviligerad som har just er som mina vänner. Tack för att ni alltid finns där, bryr er och påminner mig om att det finns annat här i livet än forskning.

Jag vill också rikta ett tack till min släkt på mammas och pappas sida, särskilt Bosse och Gerd, mina svärföräldrar Gunnar och Gunilla, mina svägerskor, Susanne och Pernilla, samt till alla släktingar på Lisas sida.

Johan och Josefine. Tack för det stöd som jag har fått av er genom alla år. Även om det finns vissa som tycker att: -Den där... Han har bara har kostat! =). Jag vill också tacka Hanna, Håkan och så klart lilla Flippisen. Jag är så glad att ha er i familjen! Tack också till Albin för att du alltid möter mig med en blöt puss när jag kommer hem!

Till världens bästa föräldrar, mamma Irene och pappa Gert-Ove. Tack för att ni alltid har trott på mig och uppmuntrat mig att gå mina egna vägar och följa mitt hjärta. Utan er hade jag aldrig varit där jag är idag! Tack också för er gränslösa kärlek och förståelse, och för att ni alltid finns där och stöttar om det behövs!

Slutligen vill jag tacka min själsfrände Lisa. Det går inte att beskriva i ord hur glad jag är över att ha träffat dig. Du har varit fantastisk under de här åren och under alla de perioder när jag inte har varit så närvarande som jag kanske skulle ha velat. Nu är det äntligen klart! Jag ser fram emot att leva resten av mitt liv med dig. Älskar dig mest!

REFERENCES USED IN THIS DOCTORAL DISSERTATION

1. UN Millenium Goals report 2010. (New York, 2010).
2. Joint United Nations Programme on HIV/AIDS (UNAIDS) and World Health Organization (WHO) AIDS epidemic update. (Geneva, 2009).
3. Pneumocystis pneumonia--Los Angeles. *MMWR Morb Mortal Wkly Rep* **30**, 250-252 (1981).
4. Gottlieb, M.S., *et al.* Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med* **305**, 1425-1431 (1981).
5. Masur, H., *et al.* An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. *N Engl J Med* **305**, 1431-1438 (1981).
6. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. *MMWR Morb Mortal Wkly Rep* **30**, 305-308 (1981).
7. Friedman-Kien, A.E. Disseminated Kaposi's sarcoma syndrome in young homosexual men. *J Am Acad Dermatol* **5**, 468-471 (1981).
8. Hymes, K.B., *et al.* Kaposi's sarcoma in homosexual men-a report of eight cases. *Lancet* **2**, 598-600 (1981).
9. Barre-Sinoussi, F., *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**, 868-871 (1983).
10. Gallo, R.C., *et al.* Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* **220**, 865-867 (1983).
11. Coffin, J., *et al.* What to call the AIDS virus? *Nature* **321**, 10 (1986).
12. Gallo, R.C., *et al.* Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224**, 500-503 (1984).
13. Popovic, M., Sarngadharan, M.G., Read, E. & Gallo, R.C. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* **224**, 497-500 (1984).
14. Sarngadharan, M.G., Popovic, M., Bruch, L., Schupbach, J. & Gallo, R.C. Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS. *Science* **224**, 506-508 (1984).
15. Schupbach, J., *et al.* Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. *Science* **224**, 503-505 (1984).
16. King, N.W., Hunt, R.D. & Letvin, N.L. Histopathologic changes in macaques with an acquired immunodeficiency syndrome (AIDS). *Am J Pathol* **113**, 382-388 (1983).
17. Letvin, N.L., *et al.* Acquired immunodeficiency syndrome in a colony of macaque monkeys. *Proc Natl Acad Sci U S A* **80**, 2718-2722 (1983).
18. Daniel, M.D., *et al.* Isolation of T-cell tropic HTLV-III-like retrovirus from macaques. *Science* **228**, 1201-1204 (1985).
19. Clavel, F., *et al.* Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233**, 343-346 (1986).
20. Bailey, A.S. & Corbitt, G. Was HIV present in 1959? *Lancet* **347**, 189 (1996).
21. Corbitt, G., Bailey, A.S. & Williams, G. HIV infection in Manchester, 1959.

Lancet **336**, 51 (1990).
22. Worobey, M.*, et al.* Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* **455**, 661-664 (2008).
23. Zhu, T.*, et al.* An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391**, 594-597 (1998).
24. Flint, S.J., Enquist, L.W., Racaniello, V.R. & Skalka, A.M. *Principles of virology*, (ASM Press, Washington D.C., 2004).
25. Kuiken, C.*, et al.* HIV Sequence Compendium 2010. (Los Alamos sequence data base, 2010).
26. Chertova, E.*, et al.* Envelope glycoprotein incorporation, not shedding of surface envelope glycoprotein (gp120/SU), Is the primary determinant of SU content of purified human immunodeficiency virus type 1 and simian immunodeficiency virus. *J Virol* **76**, 5315-5325 (2002).
27. Liu, J., Bartesaghi, A., Borgnia, M.J., Sapiro, G. & Subramaniam, S. Molecular architecture of native HIV-1 gp120 trimers. *Nature* **455**, 109-113 (2008).
28. Zhu, P.*, et al.* Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. *Proc Natl Acad Sci U S A* **100**, 15812-15817 (2003).
29. Zhu, P.*, et al.* Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* **441**, 847-852 (2006).
30. Gelderblom, H.R. Assembly and morphology of HIV: potential effect of structure on viral function. *Aids* **5**, 617-637 (1991).
31. Gelderblom, H.R., Hausmann, E.H., Ozel, M., Pauli, G. & Koch, M.A. Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. *Virology* **156**, 171-176 (1987).
32. Grief, C., Hockley, D.J., Fromholc, C.E. & Kitchin, P.A. The morphology of simian immunodeficiency virus as shown by negative staining electron microscopy. *J Gen Virol* **70 (Pt 8)**, 2215-2219 (1989).
33. Dalgleish, A.G.*, et al.* The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **312**, 763-767 (1984).
34. Folks, T.*, et al.* Susceptibility of normal human lymphocytes to infection with HTLV-III/LAV. *J Immunol* **136**, 4049-4053 (1986).
35. Gartner, S.*, et al.* The role of mononuclear phagocytes in HTLV-III/LAV infection. *Science* **233**, 215-219 (1986).
36. Klatzmann, D.*, et al.* Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes. *Science* **225**, 59-63 (1984).
37. Klatzmann, D.*, et al.* T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* **312**, 767-768 (1984).
38. Doms, R.W. & Trono, D. The plasma membrane as a combat zone in the HIV battlefield. *Genes Dev* **14**, 2677-2688 (2000).
39. Kwong, P.D.*, et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648-659 (1998).
40. Wyatt, R.*, et al.* The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393**, 705-711 (1998).

41. Markosyan, R.M., Cohen, F.S. & Melikyan, G.B. HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation. *Mol Biol Cell* **14**, 926-938 (2003).
42. Melikyan, G.B.*, et al.* Evidence that the transition of HIV-1 gp41 into a six-helix bundle, not the bundle configuration, induces membrane fusion. *J Cell Biol* **151**, 413-423 (2000).
43. Zolla-Pazner, S. Identifying epitopes of HIV-1 that induce protective antibodies. *Nat Rev Immunol* **4**, 199-210 (2004).
44. Bukrinskaya, A.G.*, et al.* Phosphorylation-dependent human immunodeficiency virus type 1 infection and nuclear targeting of viral DNA. *Proc Natl Acad Sci U S A* **93**, 367-371 (1996).
45. Schroder, A.R.*, et al.* HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521-529 (2002).
46. Rambaut, A., Posada, D., Crandall, K.A. & Holmes, E.C. The causes and consequences of HIV evolution. *Nat Rev Genet* **5**, 52-61 (2004).
47. Honda, Y.*, et al.* Type I interferon induces inhibitory 16-kD CCAAT/ enhancer binding protein (C/EBP)beta, repressing the HIV-1 long terminal repeat in macrophages: pulmonary tuberculosis alters C/EBP expression, enhancing HIV-1 replication. *J Exp Med* **188**, 1255-1265 (1998).
48. Kawakami, K., Scheidereit, C. & Roeder, R.G. Identification and purification of a human immunoglobulin-enhancer-binding protein (NF-kappa B) that activates transcription from a human immunodeficiency virus type 1 promoter in vitro. *Proc Natl Acad Sci U S A* **85**, 4700-4704 (1988).
49. Nabel, G. & Baltimore, D. An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature* **326**, 711-713 (1987).
50. Kao, S.Y., Calman, A.F., Luciw, P.A. & Peterlin, B.M. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature* **330**, 489-493 (1987).
51. Kim, S.Y., Byrn, R., Groopman, J. & Baltimore, D. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. *J Virol* **63**, 3708-3713 (1989).
52. Karn, J. Tackling Tat. *J Mol Biol* **293**, 235-254 (1999).
53. Fu, Y.X. Estimating mutation rate and generation time from longitudinal samples of DNA sequences. *Mol Biol Evol* **18**, 620-626 (2001).
54. Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M. & Ho, D.D. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582-1586 (1996).
55. Rodrigo, A.G.*, et al.* Coalescent estimates of HIV-1 generation time in vivo. *Proc Natl Acad Sci U S A* **96**, 2187-2191 (1999).
56. McMichael, A.J., Borrow, P., Tomaras, G.D., Goonetilleke, N. & Haynes, B.F. The immune response during acute HIV-1 infection: clues for vaccine development. *Nat Rev Immunol* **10**, 11-23.
57. Abrahams, M.R.*, et al.* Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J Virol* **83**, 3556-3567 (2009).

58. Keele, B.F., *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A* **105**, 7552-7557 (2008).
59. Salazar-Gonzalez, J.F., *et al.* Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J Exp Med* **206**, 1273-1289 (2009).
60. Li, Q., *et al.* Glycerol monolaurate prevents mucosal SIV transmission. *Nature* **458**, 1034-1038 (2009).
61. Geijtenbeek, T.B., *et al.* DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100**, 587-597 (2000).
62. Moir, S., *et al.* B cells of HIV-1-infected patients bind virions through CD21-complement interactions and transmit infectious virus to activated T cells. *J Exp Med* **192**, 637-646 (2000).
63. Jolly, C., Kashefi, K., Hollinshead, M. & Sattentau, Q.J. HIV-1 cell to cell transfer across an Env-induced, actin-dependent synapse. *J Exp Med* **199**, 283-293 (2004).
64. McDonald, D., *et al.* Recruitment of HIV and its receptors to dendritic cell-T cell junctions. *Science* **300**, 1295-1297 (2003).
65. Turville, S.G., *et al.* Immunodeficiency virus uptake, turnover, and 2-phase transfer in human dendritic cells. *Blood* **103**, 2170-2179 (2004).
66. Fiebig, E.W., *et al.* Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *Aids* **17**, 1871-1879 (2003).
67. Brenchley, J.M., *et al.* CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *J Exp Med* **200**, 749-759 (2004).
68. Mattapallil, J.J., *et al.* Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection. *Nature* **434**, 1093-1097 (2005).
69. Veazey, R.S., *et al.* Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection. *Science* **280**, 427-431 (1998).
70. Lawn, S.D., Butera, S.T. & Folks, T.M. Contribution of immune activation to the pathogenesis and transmission of human immunodeficiency virus type 1 infection. *Clin Microbiol Rev* **14**, 753-777, table of contents (2001).
71. De Boer, R.J., Mohri, H., Ho, D.D. & Perelson, A.S. Turnover rates of B cells, T cells, and NK cells in simian immunodeficiency virus-infected and uninfected rhesus macaques. *J Immunol* **170**, 2479-2487 (2003).
72. Douek, D.C., Picker, L.J. & Koup, R.A. T cell dynamics in HIV-1 infection. *Annu Rev Immunol* **21**, 265-304 (2003).
73. Lee, C., *et al.* Macrophage activation through CCR5- and CXCR4-mediated gp120-elicited signaling pathways. *J Leukoc Biol* **74**, 676-682 (2003).
74. Gasper-Smith, N., *et al.* Induction of plasma (TRAIL), TNFR-2, Fas ligand, and plasma microparticles after human immunodeficiency virus type 1 (HIV-1) transmission: implications for HIV-1 vaccine design. *J Virol* **82**, 7700-7710 (2008).
75. Piatak, M., Jr., *et al.* High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* **259**, 1749-1754 (1993).
76. Ho, D.D., *et al.* Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123-126 (1995).

77. Rodriguez, B., *et al.* Predictive value of plasma HIV RNA level on rate of CD4 T-cell decline in untreated HIV infection. *JAMA* **296**, 1498-1506 (2006).
78. Schacker, T.W., Hughes, J.P., Shea, T., Coombs, R.W. & Corey, L. Biological and virologic characteristics of primary HIV infection. *Ann Intern Med* **128**, 613-620 (1998).
79. Mellors, J.W., *et al.* Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* **272**, 1167-1170 (1996).
80. Margolick, J.B., *et al.* Failure of T-cell homeostasis preceding AIDS in HIV-1 infection. The Multicenter AIDS Cohort Study. *Nat Med* **1**, 674-680 (1995).
81. Bailey, J.R., Williams, T.M., Siliciano, R.F. & Blankson, J.N. Maintenance of viral suppression in HIV-1-infected HLA-B*57+ elite suppressors despite CTL escape mutations. *J Exp Med* **203**, 1357-1369 (2006).
82. Betts, M.R., *et al.* HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells. *Blood* **107**, 4781-4789 (2006).
83. Cao, Y., Qin, L., Zhang, L., Safrit, J. & Ho, D.D. Virologic and immunologic characterization of long-term survivors of human immunodeficiency virus type 1 infection. *N Engl J Med* **332**, 201-208 (1995).
84. Emu, B., *et al.* Phenotypic, functional, and kinetic parameters associated with apparent T-cell control of human immunodeficiency virus replication in individuals with and without antiretroviral treatment. *J Virol* **79**, 14169-14178 (2005).
85. Migueles, S.A., *et al.* HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc Natl Acad Sci U S A* **97**, 2709-2714 (2000).
86. Pereyra, F., *et al.* Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy. *J Infect Dis* **197**, 563-571 (2008).
87. Peretz, Y., *et al.* Human immunodeficiency virus (HIV)-specific gamma interferon secretion directed against all expressed HIV genes: relationship to rate of CD4 decline. *J Virol* **79**, 4908-4917 (2005).
88. Williams, R., Mant, C. & Cason, J. The Infectious Diseases BioBank at King's College London: archiving samples from patients infected with HIV to facilitate translational research. *Retrovirology* **6**, 98 (2009).
89. Finberg, R.W. & Kurt-Jones, E.A. Viruses and Toll-like receptors. *Microbes Infect* **6**, 1356-1360 (2004).
90. Janeway, C.A., Jr. & Medzhitov, R. Innate immune recognition. *Annu Rev Immunol* **20**, 197-216 (2002).
91. Biron, C.A. & Brossay, L. NK cells and NKT cells in innate defense against viral infections. *Curr Opin Immunol* **13**, 458-464 (2001).
92. Selin, L.K., Santolucito, P.A., Pinto, A.K., Szomolanyi-Tsuda, E. & Welsh, R.M. Innate immunity to viruses: control of vaccinia virus infection by gamma delta T cells. *J Immunol* **166**, 6784-6794 (2001).
93. Schmitz, J.E., *et al.* Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. *Science* **283**, 857-860 (1999).
94. Ogg, G.S., *et al.* Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. *Science* **279**, 2103-2106 (1998).
95. Amara, R.R., *et al.* Control of a mucosal challenge and prevention of AIDS by a

multiprotein DNA/MVA vaccine. *Science* **292**, 69-74 (2001).
96. Barouch, D.H., *et al.* Control of viremia and prevention of clinical AIDS in rhesus monkeys by cytokine-augmented DNA vaccination. *Science* **290**, 486-492 (2000).
97. Seth, A., *et al.* Immunization with a modified vaccinia virus expressing simian immunodeficiency virus (SIV) Gag-Pol primes for an anamnestic Gag-specific cytotoxic T-lymphocyte response and is associated with reduction of viremia after SIV challenge. *J Virol* **74**, 2502-2509 (2000).
98. Shiver, J.W., *et al.* Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity. *Nature* **415**, 331-335 (2002).
99. Harouse, J.M., *et al.* CD8+ T cell-mediated CXC chemokine receptor 4-simian/human immunodeficiency virus suppression in dually infected rhesus macaques. *Proc Natl Acad Sci U S A* **100**, 10977-10982 (2003).
100. Jassoy, C., *et al.* Human immunodeficiency virus type 1-specific cytotoxic T lymphocytes release gamma interferon, tumor necrosis factor alpha (TNF-alpha), and TNF-beta when they encounter their target antigens. *J Virol* **67**, 2844-2852 (1993).
101. Price, P., *et al.* Cytotoxic CD8+ T lymphocytes reactive with human immunodeficiency virus-1 produce granulocyte/macrophage colony-stimulating factor and variable amounts of interleukins 2, 3, and 4 following stimulation with the cognate epitope. *Clin Immunol Immunopathol* **74**, 100-106 (1995).
102. Emilie, D., Maillot, M.C., Nicolas, J.F., Fior, R. & Galanaud, P. Antagonistic effect of interferon-gamma on tat-induced transactivation of HIV long terminal repeat. *J Biol Chem* **267**, 20565-20570 (1992).
103. Harrer, T., Jassoy, C., Harrer, E., Johnson, R.P. & Walker, B.D. Induction of HIV-1 replication in a chronically infected T-cell line by cytotoxic T lymphocytes. *J Acquir Immune Defic Syndr* **6**, 865-871 (1993).
104. Amara, A., *et al.* HIV coreceptor downregulation as antiviral principle: SDF-1alpha-dependent internalization of the chemokine receptor CXCR4 contributes to inhibition of HIV replication. *J Exp Med* **186**, 139-146 (1997).
105. 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* **270**, 1811-1815 (1995).
106. Montefiori, D.C., *et al.* Neutralizing and infection-enhancing antibody responses to human immunodeficiency virus type 1 in long-term nonprogressors. *J Infect Dis* **173**, 60-67 (1996).
107. Moog, C., Fleury, H.J., Pellegrin, I., Kirn, A. & Aubertin, A.M. Autologous and heterologous neutralizing antibody responses following initial seroconversion in human immunodeficiency virus type 1-infected individuals. *J Virol* **71**, 3734-3741 (1997).
108. Moore, J.P., *et al.* Primary isolates of human immunodeficiency virus type 1 are relatively resistant to neutralization by monoclonal antibodies to gp120, and their neutralization is not predicted by studies with monomeric gp120. *J Virol* **69**, 101-109 (1995).
109. Poignard, P., *et al.* Neutralizing antibodies have limited effects on the control of established HIV-1 infection in vivo. *Immunity* **10**, 431-438 (1999).

110. Berman, P.W., *et al.* Protection of chimpanzees from infection by HIV-1 after vaccination with recombinant glycoprotein gp120 but not gp160. *Nature* **345**, 622-625 (1990).
111. Lubeck, M.D., *et al.* Long-term protection of chimpanzees against high-dose HIV-1 challenge induced by immunization. *Nat Med* **3**, 651-658 (1997).
112. Emini, E.A., *et al.* Prevention of HIV-1 infection in chimpanzees by gp120 V3 domain-specific monoclonal antibody. *Nature* **355**, 728-730 (1992).
113. Johnson, W.E., Lifson, J.D., Lang, S.M., Johnson, R.P. & Desrosiers, R.C. Importance of B-cell responses for immunological control of variant strains of simian immunodeficiency virus. *J Virol* **77**, 375-381 (2003).
114. Schmitz, J.E., *et al.* Effect of humoral immune responses on controlling viremia during primary infection of rhesus monkeys with simian immunodeficiency virus. *J Virol* **77**, 2165-2173 (2003).
115. Gorny, M.K., *et al.* Neutralization of diverse human immunodeficiency virus type 1 variants by an anti-V3 human monoclonal antibody. *J Virol* **66**, 7538-7542 (1992).
116. Gorny, M.K., *et al.* Human anti-V2 monoclonal antibody that neutralizes primary but not laboratory isolates of human immunodeficiency virus type 1. *J Virol* **68**, 8312-8320 (1994).
117. Moulard, M., *et al.* Broadly cross-reactive HIV-1-neutralizing human monoclonal Fab selected for binding to gp120-CD4-CCR5 complexes. *Proc Natl Acad Sci U S A* **99**, 6913-6918 (2002).
118. Sattentau, Q.J. & Moore, J.P. Human immunodeficiency virus type 1 neutralization is determined by epitope exposure on the gp120 oligomer. *J Exp Med* **182**, 185-196 (1995).
119. Golding, H., *et al.* Dissection of human immunodeficiency virus type 1 entry with neutralizing antibodies to gp41 fusion intermediates. *J Virol* **76**, 6780-6790 (2002).
120. Gorny, M.K. & Zolla-Pazner, S. Recognition by human monoclonal antibodies of free and complexed peptides representing the prefusogenic and fusogenic forms of human immunodeficiency virus type 1 gp41. *J Virol* **74**, 6186-6192 (2000).
121. Huber, M., *et al.* Complement lysis activity in autologous plasma is associated with lower viral loads during the acute phase of HIV-1 infection. *PLoS Med* **3**, e441 (2006).
122. Spear, G.T., Takefman, D.M., Sharpe, S., Ghassemi, M. & Zolla-Pazner, S. Antibodies to the HIV-1 V3 loop in serum from infected persons contribute a major proportion of immune effector functions including complement activation, antibody binding, and neutralization. *Virology* **204**, 609-615 (1994).
123. Koup, R.A., *et al.* Antibody-dependent cell-mediated cytotoxicity directed by a human monoclonal antibody reactive with gp120 of HIV-1. *Aids* **5**, 1309-1314 (1991).
124. Ljunggren, K., Fenyo, E.M., Biberfeld, G. & Jondal, M. Detection of antibodies which mediate human immunodeficiency virus-specific cellular cytotoxicity (ADCC) in vitro. *J Immunol Methods* **104**, 7-14 (1987).
125. Richman, D.D., Wrin, T., Little, S.J. & Petropoulos, C.J. Rapid evolution of the

neutralizing antibody response to HIV type 1 infection. *Proc Natl Acad Sci U S A* **100**, 4144-4149 (2003).
126. Albert, J.*, et al.* Rapid development of isolate-specific neutralizing antibodies after primary HIV-1 infection and consequent emergence of virus variants which resist neutralization by autologous sera. *AIDS* **4**, 107-112 (1990).
127. Feng, Y., Broder, C.C., Kennedy, P.E. & Berger, E.A. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872-877 (1996).
128. Choe, H.*, et al.* The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* **85**, 1135-1148 (1996).
129. Deng, H.*, et al.* Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**, 661-666 (1996).
130. Doranz, B.J.*, et al.* Two distinct CCR5 domains can mediate coreceptor usage by human immunodeficiency virus type 1. *J Virol* **71**, 6305-6314 (1997).
131. Dragic, T.*, et al.* HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381**, 667-673 (1996).
132. Bleul, C.C.*, et al.* The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature* **382**, 829-833 (1996).
133. 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* **382**, 833-835 (1996).
134. Dean, M.*, et al.* Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* **273**, 1856-1862 (1996).
135. Liu, R.*, et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367-377 (1996).
136. Samson, M.*, et al.* Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722-725 (1996).
137. Zimmerman, P.A.*, et al.* Inherited resistance to HIV-1 conferred by an inactivating mutation in CC chemokine receptor 5: studies in populations with contrasting clinical phenotypes, defined racial background, and quantified risk. *Mol Med* **3**, 23-36 (1997).
138. Choe, H.*, et al.* The orphan seven-transmembrane receptor apj supports the entry of primary T-cell-line-tropic and dualtropic human immunodeficiency virus type 1. *J Virol* **72**, 6113-6118 (1998).
139. Deng, H.K., Unutmaz, D., KewalRamani, V.N. & Littman, D.R. Expression cloning of new receptors used by simian and human immunodeficiency viruses. *Nature* **388**, 296-300 (1997).
140. Doranz, B.J.*, et al.* A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. *Cell* **85**, 1149-1158 (1996).
141. Edinger, A.L.*, et al.* An orphan seven-transmembrane domain receptor expressed widely in the brain functions as a coreceptor for human immunodeficiency virus

type 1 and simian immunodeficiency virus. *J Virol* **72**, 7934-7940 (1998).
142. Rucker, J., *et al.* Utilization of chemokine receptors, orphan receptors, and herpesvirus-encoded receptors by diverse human and simian immunodeficiency viruses. *J Virol* **71**, 8999-9007 (1997).
143. Shimizu, N., *et al.* A putative G protein-coupled receptor, RDC1, is a novel coreceptor for human and simian immunodeficiency viruses. *J Virol* **74**, 619-626 (2000).
144. Shimizu, N., *et al.* An orphan G protein-coupled receptor, GPR1, acts as a coreceptor to allow replication of human immunodeficiency virus types 1 and 2 in brain-derived cells. *J Virol* **73**, 5231-5239 (1999).
145. Alkhatib, G., *et al.* CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* **272**, 1955-1958 (1996).
146. Alkhatib, G., Liao, F., Berger, E.A., Farber, J.M. & Peden, K.W. A new SIV co-receptor, STRL33. *Nature* **388**, 238 (1997).
147. Farzan, M., *et al.* Two orphan seven-transmembrane segment receptors which are expressed in CD4-positive cells support simian immunodeficiency virus infection. *J Exp Med* **186**, 405-411 (1997).
148. Horuk, R., *et al.* The CC chemokine I-309 inhibits CCR8-dependent infection by diverse HIV-1 strains. *J Biol Chem* **273**, 386-391 (1998).
149. Reeves, J.D., *et al.* CD4-independent infection by HIV-2 (ROD/B): use of the 7-transmembrane receptors CXCR-4, CCR-3, and V28 for entry. *Virology* **231**, 130-134 (1997).
150. Arthos, J., *et al.* HIV-1 envelope protein binds to and signals through integrin alpha4beta7, the gut mucosal homing receptor for peripheral T cells. *Nat Immunol* **9**, 301-309 (2008).
151. Karlsson, I., *et al.* HIV biological variability unveiled: frequent isolations and chimeric receptors reveal unprecedented variation of coreceptor use. *AIDS* **17**, 2561-2569 (2003).
152. Murdoch, C. & Finn, A. Chemokine receptors and their role in inflammation and infectious diseases. *Blood* **95**, 3032-3043 (2000).
153. Bromley, S.K., Mempel, T.R. & Luster, A.D. Orchestrating the orchestrators: chemokines in control of T cell traffic. *Nat Immunol* **9**, 970-980 (2008).
154. Zlotnik, A. & Yoshie, O. Chemokines: a new classification system and their role in immunity. *Immunity* **12**, 121-127 (2000).
155. Alkhatib, G., Locati, M., Kennedy, P.E., Murphy, P.M. & Berger, E.A. HIV-1 coreceptor activity of CCR5 and its inhibition by chemokines: independence from G protein signaling and importance of coreceptor downmodulation. *Virology* **234**, 340-348 (1997).
156. Lee, B., *et al.* Epitope mapping of CCR5 reveals multiple conformational states and distinct but overlapping structures involved in chemokine and coreceptor function. *J Biol Chem* **274**, 9617-9626 (1999).
157. Karlsson, I., *et al.* Coevolution of RANTES sensitivity and mode of CCR5 receptor use by human immunodeficiency virus type 1 of the R5 phenotype. *J Virol* **78**, 11807-11815 (2004).
158. Brelot, A., Heveker, N., Pleskoff, O., Sol, N. & Alizon, M. Role of the first

and third extracellular domains of CXCR-4 in human immunodeficiency virus coreceptor activity. *J Virol* **71**, 4744-4751 (1997).
159. Brelot, A., Heveker, N., Montes, M. & Alizon, M. Identification of residues of CXCR4 critical for human immunodeficiency virus coreceptor and chemokine receptor activities. *J Biol Chem* **275**, 23736-23744 (2000).
160. Lu, Z.*, et al.* Evolution of HIV-1 coreceptor usage through interactions with distinct CCR5 and CXCR4 domains. *Proc Natl Acad Sci U S A* **94**, 6426-6431 (1997).
161. Cicala, C.*, et al.* R5 and X4 HIV envelopes induce distinct gene expression profiles in primary peripheral blood mononuclear cells. *Proc Natl Acad Sci U S A* **103**, 3746-3751 (2006).
162. Asjo, B.*, et al.* Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. *Lancet* **2**, 660-662 (1986).
163. Fenyo, E.M.*, et al.* Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. *J Virol* **62**, 4414-4419 (1988).
164. Cheng-Mayer, C., Seto, D., Tateno, M. & Levy, J.A. Biologic features of HIV-1 that correlate with virulence in the host. *Science* **240**, 80-82 (1988).
165. Schuitemaker, H.*, et al.* Monocytotropic human immunodeficiency virus type 1 (HIV-1) variants detectable in all stages of HIV-1 infection lack T-cell line tropism and syncytium-inducing ability in primary T-cell culture. *J Virol* **65**, 356-363 (1991).
166. Tersmette, M.*, et al.* Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. *J Virol* **62**, 2026-2032 (1988).
167. Tersmette, M.*, et al.* Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the pathogenesis of acquired immunodeficiency syndrome: studies on sequential HIV isolates. *J Virol* **63**, 2118-2125 (1989).
168. Karlsson, A., Parsmyr, K., Sandstrom, E., Fenyo, E.M. & Albert, J. MT-2 cell tropism as prognostic marker for disease progression in human immunodeficiency virus type 1 infection. *J Clin Microbiol* **32**, 364-370 (1994).
169. Richman, D.D. & Bozzette, S.A. The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression. *J Infect Dis* **169**, 968-974 (1994).
170. Schuitemaker, 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 monocytotropic to T-cell-tropic virus population. *J Virol* **66**, 1354-1360 (1992).
171. 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* **118**, 681-688 (1993).
172. Berger, E.A.*, et al.* A new classification for HIV-1. *Nature* **391**, 240 (1998).
173. Chesebro, B.*, et al.* Identification of human immunodeficiency virus envelope gene sequences influencing viral entry into CD4-positive HeLa cells, T-leukemia cells, and macrophages. *J Virol* **65**, 5782-5789 (1991).

174. O'Brien, W.A., *et al.* HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. *Nature* **348**, 69-73 (1990).
175. Shioda, T., Levy, J.A. & 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* **89**, 9434-9438 (1992).
176. De Jong, J.J., De Ronde, A., Keulen, W., Tersmette, M. & Goudsmit, J. Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. *J Virol* **66**, 6777-6780 (1992).
177. Fouchier, R.A., *et al.* Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol* **66**, 3183-3187 (1992).
178. Hu, Q., *et al.* Identification of ENV determinants in V3 that influence the molecular anatomy of CCR5 utilization. *J Mol Biol* **302**, 359-375 (2000).
179. Xiao, L., *et al.* CCR5 coreceptor usage of non-syncytium-inducing primary HIV-1 is independent of phylogenetically distinct global HIV-1 isolates: delineation of consensus motif in the V3 domain that predicts CCR-5 usage. *Virology* **240**, 83-92 (1998).
180. Jensen, M.A., Coetzer, M., van 't Wout, A.B., Morris, L. & Mullins, J.I. A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences. *J Virol* **80**, 4698-4704 (2006).
181. Jensen, M.A., *et al.* Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *J Virol* **77**, 13376-13388 (2003).
182. Jensen, M.A. & van 't Wout, A.B. Predicting HIV-1 coreceptor usage with sequence analysis. *AIDS Rev* **5**, 104-112 (2003).
183. 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* **288**, 51-62 (2001).
184. Lengauer, T., Sander, O., Sierra, S., Thielen, A. & Kaiser, R. Bioinformatics prediction of HIV coreceptor usage. *Nat Biotechnol* **25**, 1407-1410 (2007).
185. 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* **46**, 887-891 (2008).
186. Low, A.J., *et al.* Current V3 genotyping algorithms are inadequate for predicting X4 co-receptor usage in clinical isolates. *Aids* **21**, F17-24 (2007).
187. Raymond, S., *et al.* Genotypic Prediction of HIV-1 CRF02-AG Tropism. *J Clin Microbiol* (2009).
188. Raymond, S., *et al.* Correlation between genotypic predictions based on V3 sequences and phenotypic determination of HIV-1 tropism. *Aids* **22**, F11-16 (2008).
189. Carrillo, A. & Ratner, L. Human immunodeficiency virus type 1 tropism for T-lymphoid cell lines: role of the V3 loop and C4 envelope determinants. *J Virol*

70, 1301-1309 (1996).
190. Jansson, M.*, et al.* Length variation of glycoprotein 120 V2 region in relation to biological phenotypes and coreceptor usage of primary HIV type 1 isolates. *AIDS Res Hum Retroviruses* **17**, 1405-1414 (2001).
191. Labrosse, B., Treboute, C., Brelot, A. & Alizon, M. Cooperation of the V1/V2 and V3 domains of human immunodeficiency virus type 1 gp120 for interaction with the CXCR4 receptor. *J Virol* **75**, 5457-5464 (2001).
192. Nabatov, A.A.*, et al.* Intrapatient alterations in the human immunodeficiency virus type 1 gp120 V1V2 and V3 regions differentially modulate coreceptor usage, virus inhibition by CC/CXC chemokines, soluble CD4, and the b12 and 2G12 monoclonal antibodies. *J Virol* **78**, 524-530 (2004).
193. Ogert, R.A.*, 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. *J Virol* **75**, 5998-6006 (2001).
194. 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* **276**, 13433-13441 (2001).
195. Bratt, G., Leandersson, A.C., Albert, J., Sandstrom, E. & Wahren, B. MT-2 tropism and CCR-5 genotype strongly influence disease progression in HIV-1-infected individuals. *Aids* **12**, 729-736 (1998).
196. Spijkerman, I.J.*, et al.* Lower prevalence and incidence of HIV-1 syncytium-inducing phenotype among injecting drug users compared with homosexual men. *Aids* **9**, 1085-1092 (1995).
197. 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* **81**, 7885-7893 (2007).
198. Tscherning, C.*, et al.* Differences in chemokine coreceptor usage between genetic subtypes of HIV-1. *Virology* **241**, 181-188 (1998).
199. Zhang, L., Huang, Y., He, T., Cao, Y. & Ho, D.D. HIV-1 subtype and second-receptor use. *Nature* **383**, 768 (1996).
200. Abebe, A.*, et al.* HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS. *Aids* **13**, 1305-1311 (1999).
201. Bjorndal, A., Sonnerborg, A., Tscherning, C., Albert, J. & Fenyo, E.M. Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients. *AIDS Res Hum Retroviruses* **15**, 647-653 (1999).
202. Cilliers, T.*, et al.* The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C. *J Virol* **77**, 4449-4456 (2003).
203. Kaleebu, P.*, et al.* Relation between chemokine receptor use, disease stage, and HIV-1 subtypes A and D: results from a rural Ugandan cohort. *J Acquir Immune Defic Syndr* **45**, 28-33 (2007).
204. Yu, X.F.*, et al.* Phenotypic and genotypic characteristics of human immunodeficiency virus type 1 from patients with AIDS in northern Thailand. *J*

Virol **69**, 4649-4655 (1995).
205. Connell, B.J., *et al.* Emergence of X4 usage among HIV-1 subtype C: evidence for an evolving epidemic in South Africa. *Aids* **22**, 896-899 (2008).
206. Regoes, R.R. & Bonhoeffer, S. The HIV coreceptor switch: a population dynamical perspective. *Trends Microbiol* **13**, 269-277 (2005).
207. Wilkinson, D.A., Operskalski, E.A., Busch, M.P., Mosley, J.W. & Koup, R.A. A 32-bp deletion within the CCR5 locus protects against transmission of parenterally acquired human immunodeficiency virus but does not affect progression to AIDS-defining illness. *J Infect Dis* **178**, 1163-1166 (1998).
208. van't Wout, A.B., *et al.* Macrophage-tropic variants initiate human immunodeficiency virus type 1 infection after sexual, parenteral, and vertical transmission. *J Clin Invest* **94**, 2060-2067 (1994).
209. Sheppard, H.W., *et al.* HIV-1 infection in individuals with the CCR5-Delta32/Delta32 genotype: acquisition of syncytium-inducing virus at seroconversion. *J Acquir Immune Defic Syndr* **29**, 307-313 (2002).
210. Meng, G., *et al.* Primary intestinal epithelial cells selectively transfer R5 HIV-1 to CCR5+ cells. *Nat Med* **8**, 150-156 (2002).
211. Granelli-Piperno, A., Delgado, E., Finkel, V., Paxton, W. & Steinman, R.M. Immature dendritic cells selectively replicate macrophagetropic (M-tropic) human immunodeficiency virus type 1, while mature cells efficiently transmit both M- and T-tropic virus to T cells. *J Virol* **72**, 2733-2737 (1998).
212. Reece, J.C., *et al.* HIV-1 selection by epidermal dendritic cells during transmission across human skin. *J Exp Med* **187**, 1623-1631 (1998).
213. Agace, W.W., *et al.* Constitutive expression of stromal derived factor-1 by mucosal epithelia and its role in HIV transmission and propagation. *Curr Biol* **10**, 325-328 (2000).
214. Cornelissen, M., *et al.* Syncytium-inducing (SI) phenotype suppression at seroconversion after intramuscular inoculation of a non-syncytium-inducing/SI phenotypically mixed human immunodeficiency virus population. *J Virol* **69**, 1810-1818 (1995).
215. Pratt, R.D., Shapiro, J.F., McKinney, N., Kwok, S. & Spector, S.A. Virologic characterization of primary human immunodeficiency virus type 1 infection in a health care worker following needlestick injury. *J Infect Dis* **172**, 851-854 (1995).
216. 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. *AIDS Res Hum Retroviruses* **21**, 371-378 (2005).
217. Mansky, L.M. & Temin, H.M. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* **69**, 5087-5094 (1995).
218. Bunnik, E.M., Quakkelaar, E.D., van Nuenen, A.C., Boeser-Nunnink, B. & Schuitemaker, H. Increased neutralization sensitivity of recently emerged CXCR4-using human immunodeficiency virus type 1 strains compared to coexisting CCR5-using variants from the same patient. *J Virol* **81**, 525-531 (2007).
219. Bonhoeffer, S., Holmes, E.C. & Nowak, M.A. Causes of HIV diversity. *Nature* **376**, 125 (1995).

220. Yamaguchi, Y. & Gojobori, T. Evolutionary mechanisms and population dynamics of the third variable envelope region of HIV within single hosts. *Proc Natl Acad Sci U S A* **94**, 1264-1269 (1997).
221. Davenport, M.P., Zaunders, J.J., Hazenberg, M.D., Schuitemaker, H. & van Rij, R.P. Cell turnover and cell tropism in HIV-1 infection. *Trends Microbiol* **10**, 275-278 (2002).
222. Hazenberg, M.D.*, et al.* T-cell division in human immunodeficiency virus (HIV)-1 infection is mainly due to immune activation: a longitudinal analysis in patients before and during highly active antiretroviral therapy (HAART). *Blood* **95**, 249-255 (2000).
223. Ribeiro, R.M., Hazenberg, M.D., Perelson, A.S. & Davenport, M.P. Naive and memory cell turnover as drivers of CCR5-to-CXCR4 tropism switch in human immunodeficiency virus type 1: implications for therapy. *J Virol* **80**, 802-809 (2006).
224. Roederer, M., Dubs, J.G., Anderson, M.T., Raju, P.A. & Herzenberg, L.A. CD8 naive T cell counts decrease progressively in HIV-infected adults. *J Clin Invest* **95**, 2061-2066 (1995).
225. Benito, J.M.*, et al.* Quantitative alterations of the functionally distinct subsets of CD4 and CD8 T lymphocytes in asymptomatic HIV infection: changes in the expression of CD45RO, CD45RA, CD11b, CD38, HLA-DR, and CD25 antigens. *J Acquir Immune Defic Syndr Hum Retrovirol* **14**, 128-135 (1997).
226. de Roda Husman, A.M., Blaak, H., Brouwer, M. & Schuitemaker, H. CC chemokine receptor 5 cell-surface expression in relation to CC chemokine receptor 5 genotype and the clinical course of HIV-1 infection. *J Immunol* **163**, 4597-4603 (1999).
227. Simmonds, P.*, et al.* Determinants of HIV disease progression: six-year longitudinal study in the Edinburgh haemophilia/HIV cohort. *Lancet* **338**, 1159-1163 (1991).
228. Ostrowski, M.A.*, et al.* Expression of chemokine receptors CXCR4 and CCR5 in HIV-1-infected and uninfected individuals. *J Immunol* **161**, 3195-3201 (1998).
229. van Rij, R.P.*, et al.* Early viral load and CD4+ T cell count, but not percentage of CCR5+ or CXCR4+ CD4+ T cells, are associated with R5-to-X4 HIV type 1 virus evolution. *AIDS Res Hum Retroviruses* **19**, 389-398 (2003).
230. Wei, X.*, et al.* Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373**, 117-122 (1995).
231. Clavel, F.*, et al.* Genetic recombination of human immunodeficiency virus. *J Virol* **63**, 1455-1459 (1989).
232. Hu, W.S. & Temin, H.M. Retroviral recombination and reverse transcription. *Science* **250**, 1227-1233 (1990).
233. Hu, W.S. & Temin, H.M. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci U S A* **87**, 1556-1560 (1990).
234. Temin, H.M. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc Natl Acad Sci U S A* **90**, 6900-6903 (1993).
235. Zhuang, J.*, et al.* Human immunodeficiency virus type 1 recombination: rate,

fidelity, and putative hot spots. *J Virol* **76**, 11273-11282 (2002).
236. Korber, B., *et al.* Evolutionary and immunological implications of contemporary HIV-1 variation. *Br Med Bull* **58**, 19-42 (2001).
237. Weiss, R.A. HIV and AIDS in relation to other pandemics. Among the viruses plaguing humans, HIV is a recent acquisition. Its outstanding success as an infection poses immense scientific challenges to human health and raises the question "What comes nest?". *EMBO Rep* **4 Spec No**, S10-14 (2003).
238. Peeters, M., *et al.* Geographical distribution of HIV-1 group O viruses in Africa. *Aids* **11**, 493-498 (1997).
239. Yamaguchi, J., *et al.* HIV-1 Group N: evidence of ongoing transmission in Cameroon. *AIDS Res Hum Retroviruses* **22**, 453-457 (2006).
240. Plantier, J.C., *et al.* A new human immunodeficiency virus derived from gorillas. *Nat Med* **15**, 871-872 (2009).
241. Los Alamos Sequence Database.
242. Zwickl, D.J. (2006).
243. Taylor, B.S., Sobieszczyk, M.E., McCutchan, F.E. & Hammer, S.M. The challenge of HIV-1 subtype diversity. *N Engl J Med* **358**, 1590-1602 (2008).
244. de Silva, T.I., Cotten, M. & Rowland-Jones, S.L. HIV-2: the forgotten AIDS virus. *Trends Microbiol* **16**, 588-595 (2008).
245. Renjifo, B., *et al.* Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D. *Aids* **18**, 1629-1636 (2004).
246. John-Stewart, G.C., *et al.* Subtype C Is associated with increased vaginal shedding of HIV-1. *J Infect Dis* **192**, 492-496 (2005).
247. Alaeus, A., Lidman, K., Bjorkman, A., Giesecke, J. & Albert, J. Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A-D. *Aids* **13**, 901-907 (1999).
248. Costello, C., *et al.* HIV-1 subtype E progression among northern Thai couples: traditional and non-traditional predictors of survival. *Int J Epidemiol* **34**, 577-584 (2005).
249. Kaleebu, P., *et al.* Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda. *J Infect Dis* **185**, 1244-1250 (2002).
250. Kanki, P.J., *et al.* Human immunodeficiency virus type 1 subtypes differ in disease progression. *J Infect Dis* **179**, 68-73 (1999).
251. Laurent, C., *et al.* No difference in clinical progression between patients infected with the predominant human immunodeficiency virus type 1 circulating recombinant form (CRF) 02_AG strain and patients not infected with CRF02_AG, in Western and West-Central Africa: a four-year prospective multicenter study. *J Infect Dis* **186**, 486-492 (2002).
252. Vasan, A., *et al.* Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. *Clin Infect Dis* **42**, 843-852 (2006).
253. Baeten, J.M., *et al.* HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. *J Infect Dis* **195**, 1177-1180 (2007).
254. 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* **197**, 707-713 (2008).
255. Gao, F.*, et al.* Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology. *J Virol* **68**, 7433-7447 (1994).
256. Damond, F.*, et al.* Variability of human immunodeficiency virus type 2 (hiv-2) infecting patients living in france. *Virology* **280**, 19-30 (2001).
257. Arien, K.K.*, et al.* The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. *J Virol* **79**, 8979-8990 (2005).
258. Hunt, G. & Tiemessen, C.T. Occurrence of additional NF-kappaB-binding motifs in the long terminal repeat region of South African HIV type 1 subtype C isolates. *AIDS Res Hum Retroviruses* **16**, 305-306 (2000).
259. van Harmelen, J.*, et al.* Characterization of full-length HIV type 1 subtype C sequences from South Africa. *AIDS Res Hum Retroviruses* **17**, 1527-1531 (2001).
260. Ball, S.C.*, et al.* Comparing the ex vivo fitness of CCR5-tropic human immunodeficiency virus type 1 isolates of subtypes B and C. *J Virol* **77**, 1021-1038 (2003).
261. Tebit, D.M., Nankya, I., Arts, E.J. & Gao, Y. HIV diversity, recombination and disease progression: how does fitness "fit" into the puzzle? *AIDS Rev* **9**, 75-87 (2007).
262. Konings, F.A.*, et al.* Human immunodeficiency virus type 1 (HIV-1) circulating recombinant form 02_AG (CRF02_AG) has a higher in vitro replicative capacity than its parental subtypes A and G. *J Med Virol* **78**, 523-534 (2006).
263. Njai, H.F.*, et al.* The predominance of Human Immunodeficiency Virus type 1 (HIV-1) circulating recombinant form 02 (CRF02_AG) in West Central Africa may be related to its replicative fitness. *Retrovirology* **3**, 40 (2006).
264. Troyer, R.M.*, et al.* Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression. *J Virol* **79**, 9006-9018 (2005).
265. Marlink, R.*, et al.* Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* **265**, 1587-1590 (1994).
266. Andersson, S.*, et al.* Plasma viral load in HIV-1 and HIV-2 singly and dually infected individuals in Guinea-Bissau, West Africa: significantly lower plasma virus set point in HIV-2 infection than in HIV-1 infection. *Arch Intern Med* **160**, 3286-3293 (2000).
267. Berry, N.*, et al.* Low peripheral blood viral HIV-2 RNA in individuals with high CD4 percentage differentiates HIV-2 from HIV-1 infection. *J Hum Virol* **1**, 457-468 (1998).
268. Popper, S.J.*, et al.* Lower human immunodeficiency virus (HIV) type 2 viral load reflects the difference in pathogenicity of HIV-1 and HIV-2. *J Infect Dis* **180**, 1116-1121 (1999).
269. Ariyoshi, K.*, et al.* Plasma RNA viral load predicts the rate of CD4 T cell decline and death in HIV-2-infected patients in West Africa. *Aids* **14**, 339-344 (2000).
270. Barker, E.*, et al.* Virological and immunological features of long-term human immunodeficiency virus-infected individuals who have remained asymptomatic

compared with those who have progressed to acquired immunodeficiency syndrome. *Blood* **92**, 3105-3114 (1998).
271. Alabi, A.S., *et al.* Plasma viral load, CD4 cell percentage, HLA and survival of HIV-1, HIV-2, and dually infected Gambian patients. *Aids* **17**, 1513-1520 (2003).
272. Kanki, P.J., *et al.* Slower heterosexual spread of HIV-2 than HIV-1. *Lancet* **343**, 943-946 (1994).
273. O'Donovan, D., *et al.* Maternal plasma viral RNA levels determine marked differences in mother-to-child transmission rates of HIV-1 and HIV-2 in The Gambia. MRC/Gambia Government/University College London Medical School working group on mother-child transmission of HIV. *Aids* **14**, 441-448 (2000).
274. Gottlieb, G.S., *et al.* Lower levels of HIV RNA in semen in HIV-2 compared with HIV-1 infection: implications for differences in transmission. *Aids* **20**, 895-900 (2006).
275. Morner, A., *et al.* Primary human immunodeficiency virus type 2 (HIV-2) isolates, like HIV-1 isolates, frequently use CCR5 but show promiscuity in coreceptor usage. *J Virol* **73**, 2343-2349 (1999).
276. Sol, N., *et al.* Usage of the coreceptors CCR-5, CCR-3, and CXCR-4 by primary and cell line-adapted human immunodeficiency virus type 2. *J Virol* **71**, 8237-8244 (1997).
277. Sol, N., *et al.* The rhesus macaque CCR3 chemokine receptor is a cell entry cofactor for HIV-2, but not for HIV-1. *Virology* **240**, 213-220 (1998).
278. Owen, S.M., *et al.* Genetically divergent strains of human immunodeficiency virus type 2 use multiple coreceptors for viral entry. *J Virol* **72**, 5425-5432 (1998).
279. Chen, Z., Gettie, A., Ho, D.D. & Marx, P.A. Primary SIVsm isolates use the CCR5 coreceptor from sooty mangabeys naturally infected in west Africa: a comparison of coreceptor usage of primary SIVsm, HIV-2, and SIVmac. *Virology* **246**, 113-124 (1998).
280. da Silva, Z.J., *et al.* Changes in prevalence and incidence of HIV-1, HIV-2 and dual infections in urban areas of Bissau, Guinea-Bissau: is HIV-2 disappearing? *Aids* **22**, 1195-1202 (2008).
281. Hamel, D.J., *et al.* Twenty years of prospective molecular epidemiology in Senegal: changes in HIV diversity. *AIDS Res Hum Retroviruses* **23**, 1189-1196 (2007).
282. Poulsen, A.G., *et al.* HIV-2 infection in Bissau, West Africa, 1987-1989: incidence, prevalences, and routes of transmission. *J Acquir Immune Defic Syndr* **6**, 941-948 (1993).
283. Pepin, J., Plamondon, M., Alves, A.C., Beaudet, M. & Labbe, A.C. Parenteral transmission during excision and treatment of tuberculosis and trypanosomiasis may be responsible for the HIV-2 epidemic in Guinea-Bissau. *Aids* **20**, 1303-1311 (2006).
284. Poulsen, A.G., Aaby, P., Jensen, H. & Dias, F. Risk factors for HIV-2 seropositivity among older people in Guinea-Bissau. A search for the early history of HIV-2 infection. *Scand J Infect Dis* **32**, 169-175 (2000).
285. Cot, M.C., Poulain, M., Delagneau, J.F., Peeters, M. & Brun-Vezinet, F. Dual HIV-1 and HIV-2 infection in West Africa supported by synthetic peptide analysis. *AIDS Res Hum Retroviruses* **4**, 239-241 (1988).

286. Mansson, F., *et al.* Prevalence and incidence of HIV-1 and HIV-2 before, during and after a civil war in an occupational cohort in Guinea-Bissau, West Africa. *Aids* **23**, 1575-1582 (2009).
287. Ghys, P.D., *et al.* Increase in condom use and decline in HIV and sexually transmitted diseases among female sex workers in Abidjan, Cote d'Ivoire, 1991-1998. *Aids* **16**, 251-258 (2002).
288. Evans, L.A., *et al.* Simultaneous isolation of HIV-1 and HIV-2 from an AIDS patient. *Lancet* **2**, 1389-1391 (1988).
289. Rayfield, M., *et al.* Mixed human immunodeficiency virus (HIV) infection in an individual: demonstration of both HIV type 1 and type 2 proviral sequences by using polymerase chain reaction. *J Infect Dis* **158**, 1170-1176 (1988).
290. Travers, K., *et al.* Natural protection against HIV-1 infection provided by HIV-2. *Science* **268**, 1612-1615 (1995).
291. Norrgren, H., *et al.* Trends and interaction of HIV-1 and HIV-2 in Guinea-Bissau, west Africa: no protection of HIV-2 against HIV-1 infection. *Aids* **13**, 701-707 (1999).
292. Aaby, P., *et al.* Does HIV-2 protect against HIV-1 infection? *Aids* **11**, 939-940 (1997).
293. Greenberg, A.E. Possible protective effect of HIV-2 against incident HIV-1 infection: review of available epidemiological and in vitro data. *Aids* **15**, 2319-2321 (2001).
294. Schim van der Loeff, M.F., *et al.* HIV-2 does not protect against HIV-1 infection in a rural community in Guinea-Bissau. *Aids* **15**, 2303-2310 (2001).
295. Wiktor, S.Z., *et al.* Lack of protection against HIV-1 infection among women with HIV-2 infection. *Aids* **13**, 695-699 (1999).
296. Otten, R.A., *et al.* Chronic HIV-2 infection protects against total CD4+ cell depletion and rapid disease progression induced by SHIV89.6p challenge. *Aids* **18**, 1127-1135 (2004).
297. Putkonen, P., *et al.* Infection of cynomolgus monkeys with HIV-2 protects against pathogenic consequences of a subsequent simian immunodeficiency virus infection. *Aids* **4**, 783-789 (1990).
298. Holmgren, B., *et al.* Dual infections with HIV-1, HIV-2 and HTLV-I are more common in older women than in men in Guinea-Bissau. *Aids* **17**, 241-253 (2003).
299. Holmgren, B., *et al.* Mortality associated with HIV-1, HIV-2, and HTLV-I single and dual infections in a middle-aged and older population in Guinea-Bissau. *Retrovirology* **4**, 85 (2007).
300. Kassim, S., *et al.* Two-year follow-up of persons with HIV-1- and HIV-2-associated pulmonary tuberculosis treated with short-course chemotherapy in West Africa. *Aids* **9**, 1185-1191 (1995).
301. Norrgren, H., *et al.* High mortality and severe immunosuppression in hospitalized patients with pulmonary tuberculosis and HIV-2 infection in Guinea-Bissau. *Scand J Infect Dis* **33**, 450-456 (2001).
302. Schim van der Loeff, M.F., *et al.* Mortality of HIV-1, HIV-2 and HIV-1/HIV-2 dually infected patients in a clinic-based cohort in The Gambia. *Aids* **16**, 1775-1783 (2002).

303. Seng, R., *et al.* Community study of the relative impact of HIV-1 and HIV-2 on intrathoracic tuberculosis. *Aids* **16**, 1059-1066 (2002).
304. Nkengasong, J.N., *et al.* Dual infection with human immunodeficiency virus type 1 and type 2: impact on HIV type 1 viral load and immune activation markers in HIV-seropositive female sex workers in Abidjan, Ivory Coast. *AIDS Res Hum Retroviruses* **16**, 1371-1378 (2000).
305. Critchlow, C.W., Hawes, S.E., Kuypers, J.M., Kwok, S. & Kiviat, N.B. HIV-1 and HIV-2 co-infection is associated with reduced HIV-1 plasma RNA and DNA levels among Senegalese men and women. *Am. J. Epidemiol.* **151**, S11 (2000).
306. Arya, S.K. & Gallo, R.C. Human immunodeficiency virus (HIV) type 2-mediated inhibition of HIV type 1: a new approach to gene therapy of HIV-infection. *Proc Natl Acad Sci U S A* **93**, 4486-4491 (1996).
307. Kokkotou, E.G., *et al.* In vitro correlates of HIV-2-mediated HIV-1 protection. *Proc Natl Acad Sci U S A* **97**, 6797-6802 (2000).
308. Akimoto, H., *et al.* Binding of HIV-2 envelope glycoprotein to CD8 molecules and related chemokine production. *Immunology* **95**, 214-218 (1998).
309. Al-Harthi, L., Owais, M. & Arya, S.K. Molecular inhibition of HIV type 1 by HIV type 2: effectiveness in peripheral blood mononuclear cells. *AIDS Res Hum Retroviruses* **14**, 59-64 (1998).
310. Zheng, N.N., *et al.* Comparison of human immunodeficiency virus (HIV)-specific T-cell responses in HIV-1- and HIV-2-infected individuals in Senegal. *J Virol* **78**, 13934-13942 (2004).
311. Zheng, N.N., *et al.* Role of human immunodeficiency virus (HIV)-specific T-cell immunity in control of dual HIV-1 and HIV-2 infection. *J Virol* **81**, 9061-9071 (2007).
312. Weiss, R.A., *et al.* HIV-2 antisera cross-neutralize HIV-1. *Aids* **2**, 95-100 (1988).
313. Bottiger, B., *et al.* Envelope cross-reactivity between human immunodeficiency virus types 1 and 2 detected by different serological methods: correlation between cross-neutralization and reactivity against the main neutralizing site. *J Virol* **64**, 3492-3499 (1990).
314. Rodriguez, S.K., *et al.* Comparison of heterologous neutralizing antibody responses of human immunodeficiency virus type 1 (HIV-1)- and HIV-2-infected Senegalese patients: distinct patterns of breadth and magnitude distinguish HIV-1 and HIV-2 infections. *J Virol* **81**, 5331-5338 (2007).
315. Brichacek, B., Swindells, S., Janoff, E.N., Pirruccello, S. & Stevenson, M. Increased plasma human immunodeficiency virus type 1 burden following antigenic challenge with pneumococcal vaccine. *J Infect Dis* **174**, 1191-1199 (1996).
316. Goletti, D., *et al.* Effect of Mycobacterium tuberculosis on HIV replication. Role of immune activation. *J Immunol* **157**, 1271-1278 (1996).
317. Mole, L., Ripich, S., Margolis, D. & Holodniy, M. The impact of active herpes simplex virus infection on human immunodeficiency virus load. *J Infect Dis* **176**, 766-770 (1997).
318. O'Brien, W.A., *et al.* Human immunodeficiency virus-type 1 replication can be increased in peripheral blood of seropositive patients after influenza vaccination. *Blood* **86**, 1082-1089 (1995).

319. Stanley, S.K., *et al.* Effect of immunization with a common recall antigen on viral expression in patients infected with human immunodeficiency virus type 1. *N Engl J Med* **334**, 1222-1230 (1996).
320. Zhang, Y., Nakata, K., Weiden, M. & Rom, W.N. Mycobacterium tuberculosis enhances human immunodeficiency virus-1 replication by transcriptional activation at the long terminal repeat. *J Clin Invest* **95**, 2324-2331 (1995).
321. Kannangara, S., DeSimone, J.A. & Pomerantz, R.J. Attenuation of HIV-1 infection by other microbial agents. *J Infect Dis* **192**, 1003-1009 (2005).
322. Bjorkman, P., *et al.* GB virus C during the natural course of HIV-1 infection: viremia at diagnosis does not predict mortality. *Aids* **18**, 877-886 (2004).
323. Gimenez-Barcons, M., *et al.* Analysis of chemokine and cytokine expression in patients with HIV and GB virus type C coinfection. *Clin Infect Dis* **40**, 1342-1349 (2005).
324. Van der Bij, A.K., *et al.* GB virus C coinfection and HIV-1 disease progression: The Amsterdam Cohort Study. *J Infect Dis* **191**, 678-685 (2005).
325. Pomerantz, R.J. & Nunnari, G. HIV and GB virus C--can two viruses be better than one? *N Engl J Med* **350**, 963-965 (2004).
326. Xiang, J., *et al.* Effect of coinfection with GB virus C on survival among patients with HIV infection. *N Engl J Med* **345**, 707-714 (2001).
327. Grivel, J.C., Garcia, M., Moss, W.J. & Margolis, L.B. Inhibition of HIV-1 replication in human lymphoid tissues ex vivo by measles virus. *J Infect Dis* **192**, 71-78 (2005).
328. Pilotti, E., *et al.* Postgenomic up-regulation of CCL3L1 expression in HTLV-2-infected persons curtails HIV-1 replication. *Blood* **109**, 1850-1856 (2007).
329. Mansson, F., *et al.* Trends of HIV-1 and HIV-2 prevalence among pregnant women in Guinea-Bissau, West Africa: possible effect of the civil war 1998 1999. *Sex Transm Infect* **83**, 463-467 (2007).
330. Lemey, P., *et al.* Tracing the origin and history of the HIV-2 epidemic. *Proc Natl Acad Sci U S A* **100**, 6588-6592 (2003).
331. Posada, D. & Crandall, K.A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818 (1998).
332. Jukes, T.H. & Cantor, C.R. Evolution of protein molecules. in *Mammalian protein metabolism* 21-123. (New York: Academic Press. , 1969).
333. Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**, 368-376 (1981).
334. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* **22**, 160-174 (1985).
335. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111-120 (1980).
336. Tamura, K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol Biol Evol* **9**, 678-687 (1992).
337. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*

10, 512-526 (1993).
338. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425 (1987).
339. Lemey, P., Salemi, M. & Vandamme, A.M. (eds.). *The Phylogenetic Handbook*, (Cambridge University Press, 2009).
340. Efron, B., Halloran, E. & Holmes, S. Bootstrap confidence levels for phylogenetic trees. *Proc Natl Acad Sci U S A* **93**, 7085-7090 (1996).
341. Zharkikh, A. & Li, W.H. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. *Mol Biol Evol* **9**, 1119-1147 (1992).
342. Swofford, D.L. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. . (Sinauer Associates, Sunderland, Massachusetts., 2003.).
343. Drummond, A.J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214 (2007).
344. Lemey, P., Rambaut, A., Drummond, A.J. & Suchard, M.A. Bayesian phylogeography finds its roots. *PLoS Comput Biol* **5**, e1000520 (2009).
345. Lemey, P., Rambaut, A., Welch, J.J. & Suchard, M.A. Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol* **27**, 1877-1885.
346. Hahn, B.H., Shaw, G.M., De Cock, K.M. & Sharp, P.M. AIDS as a zoonosis: scientific and public health implications. *Science* **287**, 607-614 (2000).
347. Peeters, M.*, et al.* Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *Aids* **3**, 625-630 (1989).
348. Peeters, M.*, et al.* Isolation and characterization of a new chimpanzee lentivirus (simian immunodeficiency virus isolate cpz-ant) from a wild-captured chimpanzee. *Aids* **6**, 447-451 (1992).
349. Gao, F.*, et al.* Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* **397**, 436-441 (1999).
350. Corbet, S.*, et al.* env sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. *J Virol* **74**, 529-534 (2000).
351. Hirsch, V.M., Olmsted, R.A., Murphey-Corb, M., Purcell, R.H. & Johnson, P.R. An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature* **339**, 389-392 (1989).
352. Gao, F.*, et al.* Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature* **358**, 495-499 (1992).
353. Apetrei, C.*, et al.* Detection and partial characterization of simian immunodeficiency virus SIVsm strains from bush meat samples from rural Sierra Leone. *J Virol* **79**, 2631-2636 (2005).
354. Peeters, M. & V., C. "Overview of Primate Lentiviruses and their Evolution in Non-human Primates in Africa", in C. Kuiken, B. Foley, E. Freed, B. Hahn, B. Korber, P. A. Marx, F. E. McCutchan, J. W. Mellors, and S. Wolinsky, HIV sequence compendium. (Los Alamos National Laboratory, 2002).
355. Wertheim, J.O. & Worobey, M. Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. *PLoS Comput Biol* **5**, e1000377 (2009).

356. Gifford, R.J., *et al.* A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proc Natl Acad Sci U S A* **105**, 20362-20367 (2008).
357. Worobey, M., *et al.* Island biogeography reveals the deep history of SIV. *Science* **329**, 1487.
358. Lemey, P., *et al.* The molecular population genetics of HIV-1 group O. *Genetics* **167**, 1059-1068 (2004).
359. 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 J* **15**, 276-278 (2001).
360. Abecasis, A.B., Vandamme, A.M. & Lemey, P. Quantifying differences in the tempo of human immunodeficiency virus type 1 subtype evolution. *J Virol* **83**, 12917-12924 (2009).
361. Drummond, A., Pybus, O.G. & Rambaut, A. Inference of viral evolutionary rates from molecular sequences. *Adv Parasitol* **54**, 331-358 (2003).
362. Pybus, O.G. & Rambaut, A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet* **10**, 540-550 (2009).
363. Lemey, P., *et al.* Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *PLoS Comput Biol* **3**, e29 (2007).
364. Williamson, S., *et al.* A statistical characterization of consistent patterns of human immunodeficiency virus evolution within infected patients. *Mol Biol Evol* **22**, 456-468 (2005).
365. Shankarappa, R., *et al.* Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* **73**, 10489-10502 (1999).
366. Li, B., *et al.* Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *J Virol* **81**, 193-201 (2007).
367. Wood, N., *et al.* HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. *PLoS Pathog* **5**, e1000414 (2009).
368. Herbeck, J.T., *et al.* Human immunodeficiency virus type 1 env evolves toward ancestral states upon transmission to a new host. *J Virol* **80**, 1637-1644 (2006).
369. Lee, H.Y., Perelson, A.S., Park, S.C. & Leitner, T. Dynamic correlation between intrahost HIV-1 quasispecies evolution and disease progression. *PLoS Comput Biol* **4**, e1000240 (2008).
370. Williamson, S. Adaptation in the env gene of HIV-1 and evolutionary theories of disease progression. *Mol Biol Evol* **20**, 1318-1325 (2003).
371. Nowak, M.A., *et al.* Antigenic diversity thresholds and the development of AIDS. *Science* **254**, 963-969 (1991).
372. Nowak, M.A. & May, R.M. Mathematical biology of HIV infections: antigenic variation and diversity threshold. *Math Biosci* **106**, 1-21 (1991).
373. Bello, G., *et al.* Plasma viral load threshold for sustaining intrahost HIV type 1 evolution. *AIDS Res Hum Retroviruses* **23**, 1242-1250 (2007).
374. Mikhail, M., *et al.* Role of viral evolutionary rate in HIV-1 disease progression in a linked cohort. *Retrovirology* **2**, 41 (2005).

375. Delwart, E.L., *et al.* Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS. *J Virol* **71**, 7498-7508 (1997).
376. Wolinsky, S.M., *et al.* Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* **272**, 537-542 (1996).
377. Markham, R.B., *et al.* Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline. *Proc Natl Acad Sci U S A* **95**, 12568-12573 (1998).
378. McNearney, T., *et al.* Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. *Proc Natl Acad Sci U S A* **89**, 10247-10251 (1992).
379. Shankarappa, R., *et al.* Evolution of human immunodeficiency virus type 1 envelope sequences in infected individuals with differing disease progression profiles. *Virology* **241**, 251-259 (1998).
380. Lukashov, V.V., Kuiken, C.L. & Goudsmit, J. Intrahost human immunodeficiency virus type 1 evolution is related to length of the immunocompetent period. *J Virol* **69**, 6911-6916 (1995).
381. Ganeshan, S., Dickover, R.E., Korber, B.T., Bryson, Y.J. & Wolinsky, S.M. Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease. *J Virol* **71**, 663-677 (1997).
382. Ross, H.A. & Rodrigo, A.G. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. *J Virol* **76**, 11715-11720 (2002).
383. Arien, K.K., Vanham, G. & Arts, E.J. Is HIV-1 evolving to a less virulent form in humans? *Nat Rev Microbiol* **5**, 141-151 (2007).
384. Kawashima, Y., *et al.* Adaptation of HIV-1 to human leukocyte antigen class I. *Nature* **458**, 641-645 (2009).
385. Bunnik, E.M., *et al.* Adaptation of HIV-1 envelope gp120 to humoral immunity at a population level. *Nat Med* **16**, 995-997.
386. Maljkovic Berry, I., *et al.* Unequal evolutionary rates in the human immunodeficiency virus type 1 (HIV-1) pandemic: the evolutionary rate of HIV-1 slows down when the epidemic rate increases. *J Virol* **81**, 10625-10635 (2007).
387. Fischl, M.A., *et al.* The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial. *N Engl J Med* **317**, 185-191 (1987).
388. Larder, B.A., Darby, G. & Richman, D.D. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. *Science* **243**, 1731-1734 (1989).
389. Larder, B.A. & Kemp, S.D. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science* **246**, 1155-1158 (1989).
390. Palella, F.J., Jr., *et al.* Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *N Engl J Med* **338**, 853-860 (1998).
391. Gulick, R.M., *et al.* Simultaneous vs sequential initiation of therapy with indinavir, zidovudine, and lamivudine for HIV-1 infection: 100-week follow-up. *JAMA* **280**, 35-41 (1998).
392. Hammer, S.M., *et al.* A controlled trial of two nucleoside analogues plus indinavir

in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. *N Engl J Med* **337**, 725-733 (1997).
393. Kitahata, M.M.*, et al.* Effect of early versus deferred antiretroviral therapy for HIV on survival. *N Engl J Med* **360**, 1815-1826 (2009).
394. Sax, P.E. & Baden, L.R. When to start antiretroviral therapy--ready when you are? *N Engl J Med* **360**, 1897-1899 (2009).
395. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. 1-161 (Department of Health and Human Services).
396. Martinez-Picado, J.*, et al.* Antiretroviral resistance during successful therapy of HIV type 1 infection. *Proc Natl Acad Sci U S A* **97**, 10948-10953 (2000).
397. Blankson, J.N., Persaud, D. & Siliciano, R.F. The challenge of viral reservoirs in HIV-1 infection. *Annu Rev Med* **53**, 557-593 (2002).
398. Montessori, V., Press, N., Harris, M., Akagi, L. & Montaner, J.S. Adverse effects of antiretroviral therapy for HIV infection. *CMAJ* **170**, 229-238 (2004).
399. Schackman, B.R.*, et al.* The lifetime cost of current human immunodeficiency virus care in the United States. *Med Care* **44**, 990-997 (2006).
400. Morgan, D.*, et al.* HIV-1 infection in rural Africa: is there a difference in median time to AIDS and survival compared with that in industrialized countries? *Aids* **16**, 597-603 (2002).
401. Perelson, A.S.*, et al.* Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387**, 188-191 (1997).
402. Finzi, D.*, et al.* Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **278**, 1295-1300 (1997).
403. Igarashi, T.*, et al.* Macrophage are the principal reservoir and sustain high virus loads in rhesus macaques after the depletion of CD4+ T cells by a highly pathogenic simian immunodeficiency virus/HIV type 1 chimera (SHIV): Implications for HIV-1 infections of humans. *Proc Natl Acad Sci U S A* **98**, 658-663 (2001).
404. Smith, B.A.*, et al.* Persistence of infectious HIV on follicular dendritic cells. *J Immunol* **166**, 690-696 (2001).
405. Bowman, M.C., Archin, N.M. & Margolis, D.M. Pharmaceutical approaches to eradication of persistent HIV infection. *Expert Rev Mol Med* **11**, e6 (2009).
406. Rerks-Ngarm, S.*, et al.* Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N Engl J Med* **361**, 2209-2220 (2009).
407. Abdool Karim, Q.*, et al.* Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science* **329**, 1168-1174.
408. Guidelines for national human immunodeficiency virus case surveillance, including monitoring for human immunodeficiency virus infection and acquired immunodeficiency syndrome. Centers for Disease Control and Prevention. *MMWR Recomm Rep* **48**, 1-27, 29-31 (1999).
409. CB., H. WHO Staging System for HIV Infection and Disease in Adolescents and Adults. in *HIV InSite Knowledge Base* (UCSF Center for HIV Information, San

Francisco, 2006).
410. Biberfeld, G., Thorstensson, R., Bergstrom, M., Naucler, A. & Costa, C.M. Enzyme immunoassays for the demonstration of antibodies to HIV-2SBL-6669 and HTLV-IV (SIVmac). *Aids* **2**, 195-199 (1988).
411. Albert, J.*, et al.* A new human retrovirus isolate of West African origin (SBL-6669) and its relationship to HTLV-IV, LAV-II, and HTLV-IIIB. *AIDS Res Hum Retroviruses* **3**, 3-10 (1987).
412. Andersson, S., da Silva, Z., Norrgren, H., Dias, F. & Biberfeld, G. Field evaluation of alternative testing strategies for diagnosis and differentiation of HIV-1 and HIV-2 infections in an HIV-1 and HIV-2-prevalent area. *Aids* **11**, 1815-1822 (1997).
413. Walther-Jallow, L., Andersson, S., da Silva, Z. & Biberfeld, G. High concordance between polymerase chain reaction and antibody testing of specimens from individuals dually infected with HIV types 1 and 2 in Guinea-Bissau, West Africa. *AIDS Res Hum Retroviruses* **15**, 957-962 (1999).
414. Fahey, J.L.*, et al.* The prognostic value of cellular and serologic markers in infection with human immunodeficiency virus type 1. *N Engl J Med* **322**, 166-172 (1990).
415. Kiepiela, P., Smith, A.N. & Rosenberg, E. Laboratory markers associated with progression of HIV infection. *Best Pract Res Clin Obstet Gynaecol* **19**, 243-254 (2005).
416. Taylor, J.M., Fahey, J.L., Detels, R. & Giorgi, J.V. CD4 percentage, CD4 number, and CD4:CD8 ratio in HIV infection: which to choose and how to use. *J Acquir Immune Defic Syndr* **2**, 114-124 (1989).
417. Aziz, N., Nishanian, P., Mitsuyasu, R., Detels, R. & Fahey, J.L. Variables that affect assays for plasma cytokines and soluble activation markers. *Clin Diagn Lab Immunol* **6**, 89-95 (1999).
418. Jaffar, S.*, et al.* Immunological predictors of survival in HIV type 2-infected rural villagers in Guinea-Bissau. *AIDS Res Hum Retroviruses* **21**, 560-564 (2005).
419. Trouplin, V.*, et al.* Determination of coreceptor usage of human immunodeficiency virus type 1 from patient plasma samples by using a recombinant phenotypic assay. *J Virol* **75**, 251-259 (2001).
420. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596-1599 (2007).
421. Loytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632-1635 (2008).
422. Huson, D.H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254-267 (2006).
423. Bruen, T.C., Philippe, H. & Bryant, D. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665-2681 (2006).
424. Lole, K.S.*, et al.* Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152-160 (1999).
425. Salminen, M.O., Carr, J.K., Burke, D.S. & McCutchan, F.E. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* **11**, 1423-1425 (1995).

426. Ronquist, F. & Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574 (2003).
427. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* **55**, 539-552 (2006).
428. Pond, S.L., Frost, S.D. & Muse, S.V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676-679 (2005).
429. Parker, J., Rambaut, A. & Pybus, O.G. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect Genet Evol* **8**, 239-246 (2008).
430. Rambaut, A. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* **16**, 395-399 (2000).
431. Leitner, T. & Albert, J. The molecular clock of HIV-1 unveiled through analysis of a known transmission history. *Proc Natl Acad Sci U S A* **96**, 10752-10757 (1999).
432. De Cock, K.M.*, et al.* Rapid test for distinguishing HIV-1 and HIV-2. *Lancet* **336**, 757 (1990).
433. De Cock, K.M.*, et al.* Rapid and specific diagnosis of HIV-1 and HIV-2 infections: an evaluation of testing strategies. *Aids* **4**, 875-878 (1990).
434. De Cock, K.M.*, et al.* Cross-reactivity on western blots in HIV-1 and HIV-2 infections. *Aids* **5**, 859-863 (1991).
435. Nkengasong, J.N.*, et al.* Field evaluation of a combination of monospecific enzyme-linked immunosorbent assays for type-specific diagnosis of human immunodeficiency virus type 1 (HIV-1) and HIV-2 infections in HIV-seropositive persons in Abidjan, Ivory Coast. *J Clin Microbiol* **36**, 123-127 (1998).
436. Tedder, R.S.*, et al.* Envelope cross-reactivity in Western blot for HIV-1 and HIV-2 may not indicate dual infection. *Lancet* **2**, 927-930 (1988).
437. Parekh, B.S.*, et al.* Oligomeric nature of transmembrane glycoproteins of HIV-2: procedures for their efficient dissociation and preparation of Western blots for diagnosis. *Aids* **5**, 1009-1013 (1991).
438. Guidelines for appropriate evaluations of HIV testing technologies in Africa. (Department of health and human services, 2001).
439. Lafeuillade, A.*, et al.* Correlation between surrogate markers, viral load, and disease progression in HIV-1 infection. *J Acquir Immune Defic Syndr* **7**, 1028-1033 (1994).
440. Madec, Y.*, et al.* Early control of HIV-1 infection in long-term nonprogressors followed since diagnosis in the ANRS SEROCO/HEMOCO cohort. *J Acquir Immune Defic Syndr* **50**, 19-26 (2009).
441. Meloni, S.T.*, et al.* Distinct human immunodeficiency virus type 1 subtype A virus circulating in West Africa: sub-subtype A3. *J Virol* **78**, 12438-12445 (2004).
442. Meloni, S.T.*, et al.* Molecular epidemiology of human immunodeficiency virus type 1 sub-subtype A3 in Senegal from 1988 to 2001. *J Virol* **78**, 12455-12461 (2004).
443. Mild, M., Esbjornsson, J., Fenyo, E.M. & Medstrand, P. Frequent intrapatient recombination between human immunodeficiency virus type 1 R5 and X4 envelopes: implications for coreceptor switch. *J Virol* **81**, 3369-3376 (2007).
444. Galetto, R.*, et al.* The structure of HIV-1 genomic RNA in the gp120 gene

determines a recombination hot spot in vivo. *J Biol Chem* **279**, 36625-36632 (2004).
445. Fultz, P.N., *et al.* Seroprevalence of HIV-1 and HIV-2 in Guinea Bissau in 1980. *Aids* **2**, 129-132 (1988).
446. Kanki, P.J., *et al.* Human T-lymphotropic virus type 4 and the human immunodeficiency virus in West Africa. *Science* **236**, 827-831 (1987).
447. Awumbila, M. Intra-Regional Migration In West Africa: Changing Patterns And Dynamics. Vol. 11 8-10 (International Institute of Social Studies, Rotterdam, 2009).
448. Adepoju, A. *Patterns of Migration in West Africa*, (Sub-Saharan Publishers, Accra, 2005).
449. Nkengasong, J.N., *et al.* Distribution of HIV-1 subtypes among HIV-seropositive patients in the interior of Cote d'Ivoire. *J Acquir Immune Defic Syndr* **23**, 430-436 (2000).
450. Giorgi, J.V., *et al.* Shorter survival in advanced human immunodeficiency virus type 1 infection is more closely associated with T lymphocyte activation than with plasma virus burden or virus chemokine coreceptor usage. *J Infect Dis* **179**, 859-870 (1999).
451. Hazenberg, M.D., *et al.* Persistent immune activation in HIV-1 infection is associated with progression to AIDS. *Aids* **17**, 1881-1888 (2003).
452. Nowroozalizadeh, S., *et al.* Microbial translocation correlates with the severity of both HIV-1 and HIV-2 infections. *J Infect Dis* **201**, 1150-1154.
453. Redd, A.D., *et al.* Microbial translocation, the innate cytokine response, and HIV-1 disease progression in Africa. *Proc Natl Acad Sci U S A* **106**, 6718-6723 (2009).
454. Brenchley, J.M., *et al.* Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nat Med* **12**, 1365-1371 (2006).
455. Sodora, D.L. & Silvestri, G. Immune activation and AIDS pathogenesis. *Aids* **22**, 439-446 (2008).
456. Deeks, S.G., *et al.* Immune activation set point during early HIV infection predicts subsequent CD4+ T-cell changes independent of viral load. *Blood* **104**, 942-947 (2004).
457. Giorgi, J.V., *et al.* Predictive value of immunologic and virologic markers after long or short duration of HIV-1 infection. *J Acquir Immune Defic Syndr* **29**, 346-355 (2002).
458. Liu, Z., *et al.* Elevated CD38 antigen expression on CD8+ T cells is a stronger marker for the risk of chronic HIV disease progression to AIDS and death in the Multicenter AIDS Cohort Study than CD4+ cell count, soluble immune activation markers, or combinations of HLA-DR and CD38 expression. *J Acquir Immune Defic Syndr Hum Retrovirol* **16**, 83-92 (1997).
459. Cavaleiro, R., Sousa, A.E., Loureiro, A. & Victorino, R.M. Marked immunosuppressive effects of the HIV-2 envelope protein in spite of the lower HIV-2 pathogenicity. *Aids* **14**, 2679-2686 (2000).
460. Schindler, M., *et al.* Nef-mediated suppression of T cell activation was lost in a lentiviral lineage that gave rise to HIV-1. *Cell* **125**, 1055-1067 (2006).
461. Nowak, M.A., *et al.* HIV-1 evolution and disease progression. *Science* **274**, 1008-

1011 (1996).
462. Andersson, S., Norrgren, H., Dias, F., Biberfeld, G. & Albert, J. Molecular characterization of human immunodeficiency virus (HIV)-1 and -2 in individuals from guinea-bissau with single or dual infections: predominance of a distinct HIV-1 subtype A/G recombinant in West Africa. *Virology* **262**, 312-320 (1999).
463. Hartley, O., Klasse, P.J., Sattentau, Q.J. & Moore, J.P. V3: HIV's switch-hitter. *AIDS Res Hum Retroviruses* **21**, 171-189 (2005).
464. Pastore, C.*, et al.* Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. *J Virol* **80**, 750-758 (2006).
465. Wodarz, D. & Nowak, M.A. Mathematical models of HIV pathogenesis and treatment. *Bioessays* **24**, 1178-1187 (2002).
466. Callaway, D.S., Ribeiro, R.M. & Nowak, M.A. Virus phenotype switching and disease progression in HIV-1 infection. *Proc Biol Sci* **266**, 2523-2530 (1999).

HIV-1 MOLECULAR EPIDEMIOLOGY IN GUINEA-BISSAU, WEST AFRICA: ORIGIN, DEMOGRAPHY AND MIGRATIONS

PAPER I

Joakim Esbjörnsson
Mattias Mild
Fredrik Månsson
Hans Norrgren
Patrik Medstrand

Submitted

HIV-1 molecular epidemiology in Guinea-Bissau, West Africa: Origin, demography and migrations

Joakim Esbjörnsson[1]*, Mattias Mild[2,3], Fredrik Månsson[4], Hans Norrgren[5] and Patrik Medstrand[1]

[1]Department of Experimental Medical Science, Section of Molecular Virology, Lund University, Lund, Sweden. [2]Department of Microbiology, Tumour and Cell Biology, Karolinska Institute, Stockholm, Sweden. [3]Department of Virology, Swedish Institute for Infectious Disease Control, Solna, Sweden. [4]Department of Clinical Sciences, Malmö, Infectious Diseases Research Unit, Lund University, Sweden. [5]Department of Clinical Sciences, Division of Infection Medicine, Lund University, Lund, Sweden

The HIV-1 epidemic in West Africa has been dominated by subtype A and the recombinant form CRF02_AG. Little is known about the origins and the evolutionary history of HIV-1 in this region. We employed Maximum likelihood and Bayesian methods in combination with temporal and spatial information to reconstruct the HIV-1 subtype distribution, demographic history and migration patterns over time in Guinea-Bissau, West Africa. We found that HIV-1 CRF02_AG and subsubtype A3 are the dominating forms in Guinea-Bissau and that they were introduced into the country on at least six different occasions between 1976 and 1981, following the end of the independence war (1963-1974), corresponding well with the first reported HIV-1 cases in Guinea-Bissau. Migration analyses suggested that (1) the HIV-1 epidemic started in the capital Bissau and then dispersed into more rural areas, and (2) the epidemic in Guinea-Bissau was connected to both Cameroon and Mali. This is the first study that describes the HIV-1 molecular epidemiology in a West African country by combining the results of subtype distribution with analyses of epidemic origin and epidemiological linkage between locations. Changes in the HIV-1 epidemic in Guinea-Bissau coincided with, and were likely influenced by, socio-political instabilities in the region.

Human immunodeficiency virus type 1 (HIV-1) originated in West Central Africa via cross-species transmission from chimpanzees around the beginning of the 20th century, and has since then diversified in the human population [1, 2]. Today, the most prevalent group of HIV-1 is the main (M) group which has been divided into subtypes (A-D, F-H, J-K), sub-subtypes (A1-A4, F1-F2) and 43 circulating recombinant forms (CRFs), distinguished on both the genetic level and geographic location [3].

HIV-1 mutates and recombines at extremely high rates, and the rapid generation of genetic diversity makes it possible to study the dynamics of evolutionary changes over time and to trace patterns of viral dispersal in HIV-1 epidemics [4, 5]. Coalescent theory in a phylogenetic framework has proven to be a useful tool to infer population history, and it has been used to study a variety of pathogens, including HIV-1, in different geographic regions [6-9].

***Corresponding author: Joakim Esbjörnsson, joakim.esbjornsson@med.lu.se**

Little is known about the HIV-1 population dynamics and migration events that have influenced the HIV-1 epidemic in countries in West Africa. The dominating form of HIV-1 in this region is the CRF02_AG, a recombinant between the subtypes A and G [10-15]. Most countries in West Africa reported an almost exponential increase in HIV-1 prevalence during the 1990's, reaching a steady-state level of approximately one to six percent by the end of the 1990's [16]. In Guinea-Bissau, a few cases were reported during the 1980's, and steady-state prevalence level of four to seven percent was reached by the end of the 1990's [17-22]. Since the emergence of the AIDS epidemic, information on HIV-1 subtype distribution in Guinea-Bissau is limited to one study. Andersson *et al.* studied samples from 27 HIV-1 infected individuals collected 1994-1996 and found that 81% of the individuals were infected by CRF02_AG, 15% with subtype A, and one individual with subtype B [10].

The objective of the current study was to characterize the molecular epidemiology of HIV-1 in Guinea-Bissau, West Africa. We amplified and sequenced the HIV-1 *env* V1-V3 region (~940 bp) from plasma samples of 82 individuals from Guinea-Bissau collected between 1993 and 2008. Maximum likelihood (ML) and Bayesian phylogenetic methods were used to reconstruct the epidemic and the demographic history of HIV-1 in Guinea-Bissau. By combining spatial and temporal information in a Bayesian phylogeographic framework we reconstructed migration events both within Guinea-Bissau and between Guinea-Bissau and other West African countries. The results of this study may have implications in understanding the epidemic potential of different HIV-1 variants and preventing future spread of HIV-1, particularly in West African countries, but also in other regions of the world.

Results

Prevalence of HIV-1 subtypes and recombinants in Guinea-Bissau

Phylogenetic analyses showed that the most common forms of HIV-1 in Guinea-Bissau is the CRF02_AG (57%) and the subsubtype A3 (20%) (Table S1). In addition, four sequences (5%) were subtype C, one subsubtype A1 and one CRF06_cpx. Finally, 13 (16%) of the sequences were subtype A-like and clustered with long branch lengths within or close to the CRF02_AG and the A3 clusters. BootScan analyses indicated that all 13 sequences were recombinants of HIV-1 CRF02_AG and subsubtype A3. Interestingly, 12 of the 13 recombinants had a breakpoint within the C2 region of gp120. To confirm recombination, we split these sequences at their respective recombination breakpoint and constructed ML trees for each part together with the rest of the sequences from Guinea-Bissau and reference sequences. As expected, all of the identified recombinants clustered partly within the CRF02_AG cluster and partly within the subsubtype A3 cluster.

Next, we investigated differences in subtype distribution between the capital Bissau and rural Guinea-Bissau to evaluate the possibility of local epidemics related to subtype. No such differences were found: HIV-1 CRF02_AG (Bissau 55%; rural Guinea-Bissau 62%); subsubtype A3 (Bissau 23%; rural Guinea-Bissau 12%) (p=0.35, two-tailed Fisher's exact test). Two noteworthy exceptions were found: (1) all four individuals infected with HIV-1 subtype C were sampled in the capital Bissau, and (2) there was a trend for a more profound CRF02_AG epidemic in the north-western part (10 out of 12 (83%) of the sampled sequences were CRF02_AG) as compared to the north-eastern part (three out of 8 (38%) were CRF02_AG) of rural Guinea-Bissau (p=0.06).

Timing of the introductions of HIV-1 CRF02_AG and A3 in Guinea-Bissau

To investigate the origin of the HIV-1 CRF02_AG and subsubtype A3 epidemics in Guinea-Bissau, we first determined different introduction events. ML and Bayesian phylogenies were reconstructed and clusters supported with ZBLT values of p<0.001 and posterior probabilities of >95%, were considered to be highly supported (branches denoted **, Figures 1a and b). One cluster in the CRF02_AG analysis were supported with a ZBLT value of p=0.005 and a posterior probability value of 93% (the branch is denoted *, Figure 1a). We identified five well-supported Guinea-Bissau-specific clusters for CRF02_AG and one for subsubtype A3. We also noted that clusters three to five in the CRF02_AG phylogeny formed a cluster together with four other sequences from Guinea-Bissau and two sequences from Cameroon (Figure 1a), possibly reflecting a single common introduction. However, this large cluster was supported by neither the ZBLT nor the posterior probability. In fact, in the Bayesian maximum clade credibility tree (MCC), these three clusters did not form any

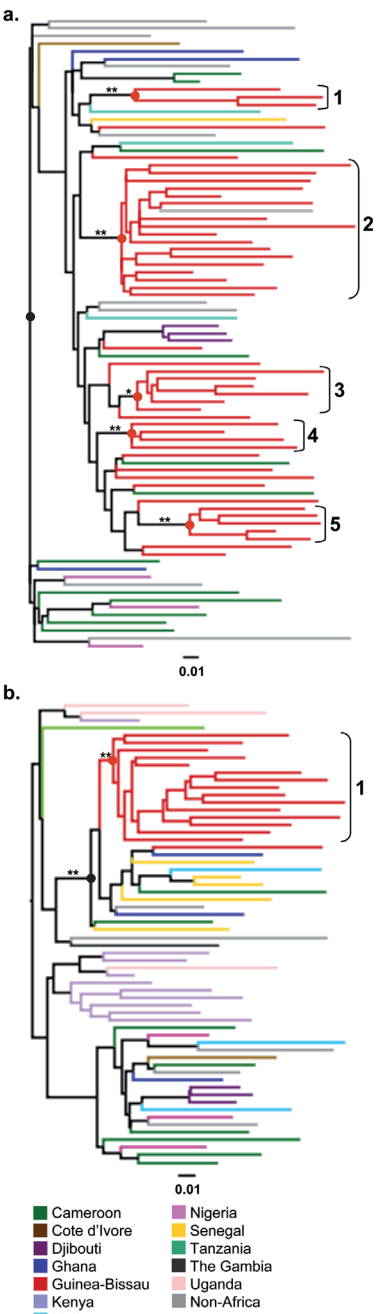


Figure 1a-b. Determination of Guinea-Bissau-specific clusters in HIV-1 CRF02_AG and subsubtype A3 phylogenies. Maximum-likelihood trees were reconstructed based on the sequences from Guinea-Bissau together with reference sequences (A. CRF02_AG and B. subsubtype A3). The scale bar at the bottom of each tree represents 0.01 nucleotide substitutions per site. Asterisks along branches represent significant monophyletic clusters (**: Zero branch-length test p-values <0.001, and posterior probabilities >95% in corresponding Bayesian analysis. *: Zero branch-length test p-values <0.01, and posterior probabilities >90%). Statistically supported clusters are numbered according to appearance in tree and the corresponding tMRCA are marked with filled circles (red circles = Guinea-Bissau-specific tMRCAs, and black circles = the tMRCA of the CRF02_AG or subsubtype A3). The colours represent the country or geographic region representing the origin of each tip in the phylogeny.

potential monophyletic cluster.

Next, we estimated the timing of the identified introductions in Guinea-Bissau. The first introduction of the CRF02_AG in Guinea-Bissau was estimated to 1976 (95% HPD 1968-1982), followed by four additional introductions during a short time period (1979-1981) (Table S2). The tMRCA for all of the CRF02_AG sequences was 1969 (95% HPD 1960-1976), and the evolutionary rate 3.80 × 10^{-3} substitutions $site^{-1}$ $year^{-1}$ (95% HPD 2.84-4.76 × 10^{-3}). In the subsubtype A3 analysis, the tMRCA of the Guinea-Bissau cluster was estimated to 1979 (95% HPD 1960-1988). In the phylogeny of HIV-1 subsubtype A3, we identified a cluster containing both our sequences from Guinea-Bissau and some reference sequences (determined to be of subsubtype A3) that were separated from the other reference sequences (determined subtype A sequences, but not subsubtyped) (Figure 1b). Not many sequences are annotated as subsubtype A3 sequences in Genbank or the Los Alamos sequence data base, and we therefore interpreted this cluster as a true subsubtype A3 cluster. Neither the tMRCA nor the evolutionary rate for this subsubtype has been presented before. We reanalyzed this cluster and found the tMRCA of subsubtype A3 to be 1975 (95% HPD 1955-1986), and the evolutionary rate to be 3.27 × 10^{-3} substitutions $site^{-1}$ $year^{-1}$ (95% HPD 1.61-4.93 × 10^{-3}) (Table S2).

Detailed demographic analysis of the Guinea-Bissau HIV-1 epidemic

To reconstruct the demographic history in Guinea-Bissau we focused on the two major clusters identified for the CRF02_AG (cluster 2) and the subsubtype A3 (cluster 1) epidemics, separately. The remaining clusters (CRF02_AG cluster 1 and 3-5) consisted of too few sequences to be subjected for demographic analysis. The tMRCA for cluster 2 of the CRF02_AG analysis dated back to 1976 (95% HPD 1968-1982), and contained 17 sequences from Guinea-Bissau. Demographic reconstruction of this cluster showed an exponential increase in median number of effective infections during 1985-1990 (30 to 562 effective infections), followed by an asymptotic phase towards the present (Figure 2a). A comparative demographic analysis of all Guinea-Bissau-specific CRF02_AG sequences together with the reference sequences (in total 77 sequences), showed a more pronounced demographic history. An initial moderate growth phase during the 1970's was followed by a rapid exponential growth phase between 1981 and 1983, during which the number of effective infections increased from an initial median value of 350 effective infections in 1981, to a final median estimate of 4232 effective infections in 1983.

Next, we reconstructed the demographic history for subsubtype A3, both in Guinea-Bissau (cluster 1, 15 sequences), and for Guinea-Bissau specific sequences together with true subsubtype A3 reference sequences (in total 21 sequences). The tMRCA was estimated to 1979 (95% HPD 1960-1988), and the demographic analysis of subsubtype A3 in Guinea-Bissau showed a similar trend as for the CRF02_AG epidemic, with an exponential increase in the median number of effective infections between 1984 and 1987 (from 97 to 661) (Figure 2c). Inspection of the median estimates of the demographic history of the entire subsubtype A3 in West Africa suggested a slower and longer initial growth phase as compared to the CRF02_AG, starting in 1975 (at 12 effective infections), and ending in 1991 (at 1390 effective infections) (Figure 2d).

To further characterize the demographic history of HIV-1 in Guinea-Bissau, we calculated the median proportion of the current lineages that existed during 1970-1975, 1975-1980, 1980-1985, 1985-1990, and 1990-1995. The analyses showed that

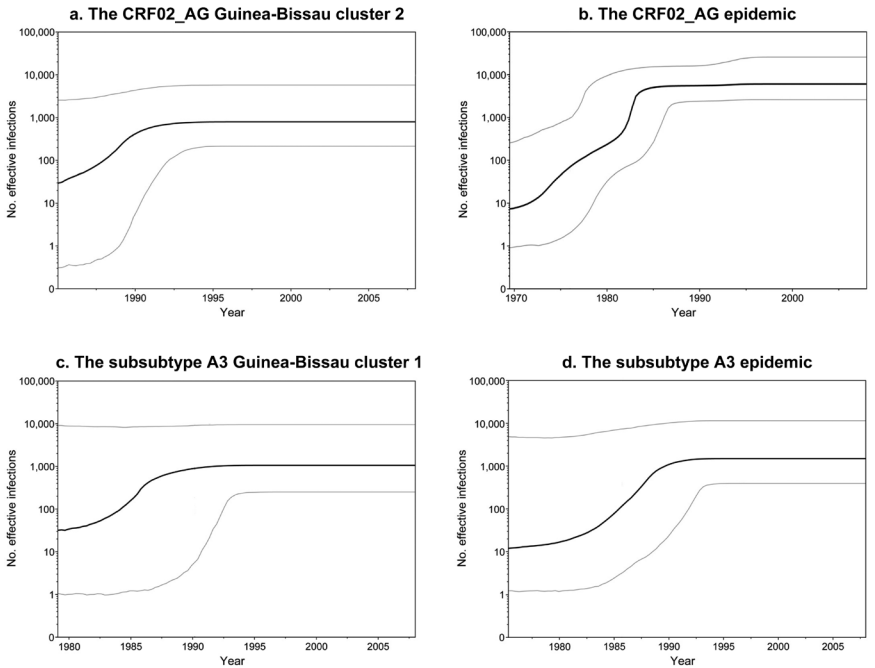


Figure 2a-d. Bayesian skyline plots for the CRF02_AG and subsubtype A3 epidemics.
Non-parametric median estimates of the number of effective infections over time for a. the CRF02_AG Guinea-Bissau cluster 2, b. the CRF02_AG epidemic in West Africa, c. the subsubtype A3 Guinea-Bissau cluster 1, and d. the subsubtype A3 epidemic in West Africa. All plots were based on the relaxed clock assumption. The grey lines represent the upper and lower 95% highest posterior density estimates.

79% (95% HPD 30%-94%) of the CRF02_AG lineages, and 100% (95% HPD 31%-100%) of the subsubtype A3 lineages were present in the country already in 1985 and 1990, respectively (Table 1).

Reconstruction of spatial dispersal patterns related to Guinea-Bissau

To reconstruct the HIV-1 CRF02_AG and A3 migration patterns over time, we removed sequences lacking information of sampling place and sampling date from the datasets used in the cluster analyses (67 and 21 sequences of CRF02_AG and subsubtype A3, respectively). For the subsubtype A3 analysis, only sequences from the identified A3 cluster were allowed. Both datasets indicated an association between geographic location and shared ancestry in the phylogenies: CRF02_AG: PS: $p<0.01$; AI: $p<0.01$); subsubtype A3: PS: $p=0.06$; AI: $p=0.02$.

The phylogeographic analyses indicated that both the HIV-1 CRF02_AG and subsubtype A3 epidemics in Guinea-Bissau originated in the capital Bissau, from were they then dispersed out to smaller cities and villages in the countryside, probably during the 1980's and 1990's (Figures 3a and c, Videos S1a and S1c). Investigation of the diffusion patterns between Guinea-Bissau and other West African countries showed three well-supported connections for HIV-1 CRF02_AG, while none were identified for subsubtype A3.

Table 1 – Bayesian estimates of median time to the most recent common ancestor, and median proportions of current lineages existing at the studied time-points. In this analysis, the posterior distribution of trees was used to examine the median percentage of 2008 years strains (CRF02_AG and A3, separately) in Guinea-Bissau present in the examined years. This estimate gives an indication of changes in strain diversity during the time-periods under study.

	CRF02_AG - RC-BSP*		Subsubtype A3 - RC-BSP*	
	Median	95% HPD†	Median	95% HPD†
tMRCA‡	1976	1968-1982	1979	1960-1988
% strains in 1980	23.4	0.2-80.8	0.6	0.6-100
% strains in 1985	78.7	29.8-93.6	37.5	0.6-100
% strains in 1990	93.6	91.5-97.9	100.0	31.3-100.0
% strains in 1995	93.6	91.5-95.7	93.8	93.8-100.0

*RC: Relaxed clock assumption. BSP10: Bayesian skyline plot used as tree prior.
†HPD: The lower and upper boundaries in 95% high posterior density interval.
‡The estimated year of the most ancient ancestor of CRF02_AG and subsubtype A3 in Guinea-Bissau, respectively. tMRCA: time to most recent common ancestor.

One migration event of HIV-1 CRF02_AG originated in Cameroon (1977), and reached Guinea-Bissau before 2003 (the analyzed sample were collected 2007, but inspection of patient history data showed that the subject was HIV-1 positive in Guinea-Bissau already in 2003) (Figures 3b, Video S1b, Table S1). In addition, we found two separate migration events between Guinea-Bissau and Mali. Both originated in Guinea-Bissau in 1975 and 1979, respectively, and reached Mali before 2003 (Video S1b).

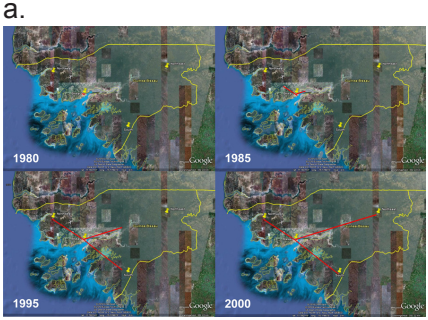
Discussion

HIV-1 subtype A and CRF02_AG represents approximately 27% of the worldwide HIV-1 infections, most of them prevailing in West and Central Africa [23]. Epidemiological studies from West Africa have shown that the CRF02_AG represents 50-70% of the HIV-1 infections in this region. We studied 82 HIV-1 infected individuals from Guinea-Bissau, and found that 57% were infected with CRF02_AG and 21% with subtype A. Detailed phylogenetic analysis of subtype A sequences showed that the vast majority of them (94%) belonged to the previously described subsubtype A3 [24-26]. This subsubtype was originally described in Senegal (with a prevalence of 16%), and has been shown to circulate in several countries in West Africa. We also identified 13 recombinants between HIV-1 CRF02_AG and subsubtype A3. The recombination breakpoints were located in the gp120 C2 region in 12 of the 13 recombinants. This region has previously been shown to be a recombination hot spot for subtype B [27, 28]. In addition, in the study by Meloni *et al.*, where the subsubtype A3 was first described, two out of five putative HIV-1 A3 sequences were found to be recombinants between the CRF02_AG and the subsubtype A3 [24]. One of these sequences displayed a recombination breakpoint in the gp120 C2 region.

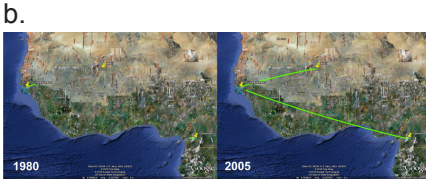
To estimate the timing of the introduction of HIV-1 in Guinea-Bissau we focused on the two dominating forms, the CRF02_AG and subsubtype A3. In the CRF02_AG phylogeny we identified one major and four minor clusters of Guinea-Bissau-specific sequences, indicating a scenario of multiple introductions into Guinea-Bissau. The introduction of the major CRF02_AG cluster dated back to 1976, and was then followed by several introductions during the late 1970's and early 1980's. The timing of these introductions fits well with the results from early studies of HIV-1 seroprevalence in West Africa [22, 29]. In a study by Fultz *et al.*, one out of 440 analyzed serum samples collected in 1980 reacted in an HIV-1 radio-immunoprecipitation assay, and were considered to be potentially HIV-1 positive by the authors [22]. Kanki *et al.* analyzed

2131 serum samples, collected 1985-1987, from Guinea-Bissau and the neighbouring countries Senegal, Guinea and Mauritania and found an HIV-1 seroprevalence of 0-0.7% in the analyzed samples [29]. The results of these studies indicate an arising HIV-1 epidemic in West Africa during the early 1980's. The estimated dates of the HIV-1 CRF02_AG introductions into Guinea-Bissau are also in line with the scenario seen in the demographic analyses. The BSP analysis suggested an initial and almost asymptotic growth period of effective infections followed by a rapid exponential growth phase, starting around 1985 with a median of 30 effective infections and ending in 1990 with an almost 20-fold increase in median effective infections to 562 in Guinea-Bissau (Figures 2a). A similar pattern was seen in the analysis of the prevalence of current lineages over time. Only 23% of the current Guinea-Bissau lineages existed in the country in 1980. In 1985 the number had increased exponentially to 79%, and in 1990 almost all (94%) of the current CRF02_AG lineages were present in the country. When examining the demographic growth patterns of all CRF02_AG sequences in our dataset (representing 12 different countries) we found that the most profound increase in number of effective infections took place during 1981-1983 (with a rapid increase in median number of effective infections from 350 to 4232). Interestingly, this increase coincided with most of the introductions into Guinea-Bissau and preceded the major increase in number of effective infections in Guinea-Bissau. A closer inspection of the BSP suggested another rapid growth phase at an earlier time-point, between 1972 and 1974 (with an increase of four times in number of

a.



b.



c.

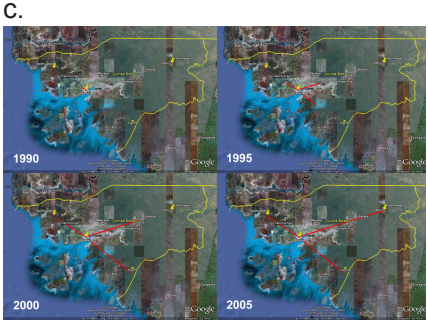


Figure 3a-c – Temporal dynamics of spatial HIV-1 CRF02_AG and subsubtype A3 diffusion.
To reduce the state space for the diffusion process we removed non-African sequences from the datasets used in the cluster analysis to two datasets of 68 CRF02_AG (locations: Guinea-Bissau, Senegal, Mali, Ghana, Nigeria and Cameroon), and 21 subsubtype A3 (locations: Guinea-Bissau, Senegal, Mali, Cameroon) sequences, respectively. To study the diffusion process within Guinea-Bissau, Guinea-Bissau-derived sequences were further subdivided into four different localities (Central (Bissau), Northwest, Northeast, and South). a. Snapshots of the HIV-1 CRF02_AG dispersal pattern within Guinea-Bissau, b. the HIV-1 CRF02_AG dispersal pattern between Guinea-Bissau and other countries and c. the HIV-1 subsubtype A3 dispersal pattern within Guinea-Bissau. Lines between locations represent branches in the maximum clade credibility tree (MCC) along which well-supported location transitions occurs. The diffusion process was visualized in Google Earth (http://earth.google.com).

increased migration frequency in the region [32,33]. Guinea-Bissau gained independence from Portugal in 1974, after 11 years of war. One could speculate that the rapid increases in the number of effective infections could have been affected either during or shortly after these instabilities.

We recognize that the demographic analyses of both the subsubtype A3 epidemic in West Africa and the local epidemics in Guinea-Bissau (CRF02_AG cluster 2 and subsubtype A3 cluster 1) are based on relatively few sequences (reflected by large HPD intervals, Figures 2a, c-d), and analyses of larger datasets are needed to confirm the results presented in this study. The rationale behind analysing well-supported clusters of country-specific sequences (instead of analysing all country-specific sequences together, regardless of clustering) is that it gives the highest probability of extracting demographic patterns specific for the country under study, without any influence of sequences from other countries. Although we stress that the demographic estimates should be interpreted with caution, it is interesting to note that the median estimates of the local demographic patterns show similar trends for both CRF02_AG and subsubtype A3 in Guinea-Bissau (Figures 2a and c). The timing of the increase in median number of effective infections also fits well with the more distinguished demographic pattern seen for the entire CRF02_AG epidemic (Figure 2b).

The phylogeographic analyses showed that HIV-1 was introduced into the urban centre (Bissau), and then spread throughout the country, a scenario that is similar to what has been reported for Cote d'Ivoire [13]. In addition, strains of the CRF02_AG epidemic, the dominant form of HIV-1 in Guinea-Bissau, started to migrate to rural areas almost 10 years before subsubtype A3, despite similar introduction dates. This also coincided with the exponential growth phase of the major CRF02_AG lineage (Figure 1a, cluster 2). In contrast, the subsubtype A3 epidemic grew exponentially in Bissau, and did not start to migrate to rural areas until after this phase. One might speculate that the subsubtype A3 epidemic grew in a small subpopulation not as prone to migrate as compared to the subpopulations infected with the CRF02_AG. Another possibility is that HIV-1 subsubtype A3 is less transmissible than the CRF02_AG. Competition assays in peripheral blood mononuclear cells in vitro, have shown that the CRF02_AG out-competes both its parental strains subtype A and G, respectively [34,35]. The migration analyses between Guinea-Bissau and other countries in West Africa showed connections to both Cameroon and Mali. Since the CRF02_AG is a recombinant between the "pure" subtypes A and G, which are prevalent in central West Africa, it is reasonable to believe that the CRF02_AG arose in this area of Africa. Therefore it is not surprising to find a shared common ancestor of the Guinea-Bissau CRF02_AG epidemic in Cameroon. The connection between Guinea-Bissau and Mali is also expected since the two countries have much business interests in common. Although the migrations established in this study are important in tracing disease spread connected to Guinea-Bissau, we emphasize that these are the migrations that we can extract from the analyzed sequences. In this analysis we used all available CRF02_AG gp120 V1-V3 sequences with known sampling date and geographic location. However, it is both possible and likely that other migrations have existed in addition to those presented here and an extended analysis of a larger amount of sequences could give a more detailed picture of HIV-1 disease spread inter-connected to Guinea-Bissau.

This study is the first to investigate the molecular epidemiology and demography in a West African country, and the first to study the HIV-1 subsubtype A3 in this context. Un

effective infections, from 11 to 44).
In the subsubtype A3 analysis, we identified one monophyletic cluster (1979), suggesting a major introduction followed by local epidemic spread. The demographic analysis showed similar patterns as for the CRF02_AG epidemic, with an exponential growth phase during the mid 1980's (1984-1987). Interestingly, this growth phase seemed to both predate, and to be more rapid than the one seen when analyzing all of the subsubtype A3 sequences (1985-1991). Accelerated transmissions have been connected to major changes on the socio-political level [30, 31]. In West Africa, a number of armed conflicts took place during the 1960's and 1970's which led to huge socio-economic and political instabilities, and an increased migration frequency in the region [32, 33]. Guinea-Bissau gained independence from Portugal in 1974, after 11 years of war. One could speculate that the rapid increases in the number of effective infections could have been affected either during or shortly after these instabilities.

We recognize that the demographic analyses of both the subsubtype A3 epidemic in West Africa and the local epidemics in Guinea-Bissau (CRF02_AG cluster 2 and subsubtype A3 cluster 1) are based on relatively few sequences (reflected by large HPD intervals, Figures 2a, c-d), and analyses of larger datasets are needed to confirm the results presented in this study. The rationale behind analysing well-supported clusters of country-specific sequences (instead of analysing all country-specific sequences together, regardless of clustering) is that it gives the highest probability of extracting demographic patterns specific for the country under study, without any influence of sequences from other countries. Although we stress that the demographic estimates should be interpreted with caution, it is interesting to note that the median estimates of the local demographic patterns show similar trends for both CRF02_AG and subsubtype A3 in Guinea-Bissau (Figures 2a and c). The timing of the increase in median number of effective infections also fits well with the more distinguished demographic pattern seen for the entire CRF02_AG epidemic (Figure 2b).

The phylogeographic analyses showed that HIV-1 was introduced into the urban centre (Bissau), and then spread throughout the country, a scenario that is similar to what has been reported for Cote d'Ivoire [13]. In addition, strains of the CRF02_AG epidemic, the dominant form of HIV-1 in Guinea-Bissau, started to migrate to rural areas almost 10 years before subsubtype A3, despite similar introduction dates. This also coincided with the exponential growth phase of the major CRF02_AG lineage (Figure 1a, cluster 2). In contrast, the subsubtype A3 epidemic grew exponentially in Bissau, and did not start to migrate to rural areas until after this phase. One might speculate that the subsubtype A3 epidemic grew in a small subpopulation not as prone to migrate as compared to the subpopulations infected with the CRF02_AG. Another possibility is that HIV-1 subsubtype A3 is less transmissible than the CRF02_AG. Competition assays in peripheral blood mononuclear cells *in vitro*, have shown that the CRF02_AG out-competes both its parental strains subtype A and G, respectively [34, 35]. The migration analyses between Guinea-Bissau and other countries in West Africa showed connections to both Cameroon and Mali. Since the CRF02_AG is a recombinant between the "pure" subtypes A and G, which are prevalent in central West Africa, it is reasonable to believe that the CRF02_AG arose in this area of Africa. Therefore it is not surprising to find a shared common ancestor of the Guinea-Bissau CRF02_AG epidemic in Cameroon. The connection between Guinea-Bissau and Mali is also expected since the two countries have much business interests in common. Although the migrations established

in this study are important in tracing disease spread connected to Guinea-Bissau, we emphasize that these are the migrations that we can extract from the analyzed sequences. In this analysis we used all available CRF02_AG gp120 V1-V3 sequences with known sampling date and geographic location. However, it is both possible and likely that other migrations have existed in addition to those presented here and an extended analysis of a larger amount of sequences could give a more detailed picture of HIV-1 disease spread inter-connected to Guinea-Bissau.

This study is the first to investigate the molecular epidemiology and demography in a West African country, and the first to study the HIV-1 subsubtype A3 in this context. Understanding factors influencing differences in distribution, demographic history and migration patterns of different HIV-1 subtypes in West African countries are crucial to understand the epidemic potential of HIV-1 in this part of Africa.

Materials and Methods

Ethics

The study was approved by the Research Ethics Committee at the Karolinska Institute, Stockholm (RECKI), and the Ministries of Health and the Interior in Guinea-Bissau (MHIGB). RECKI and MHIGB also approved with a verbal consent from all study participants, due to high rate of illiteracy in the cohort.

Study population

Nucleotide sequences used in this study were selected from a sample set of 1,562 HIV-1 *env* V1-V3 clones derived from 156 blood plasma samples of 82 individuals, collected between 1993 and 2008 (Table S1). Seventy-seven of the samples were selected from a cohort of police officers, and five samples (DL11967-DL11971) where from a case control study [17, 19, 36, 37]. All samples had been collected in Guinea-Bissau, West Africa, and were selected based on sample availability, follow up time and disease status (patients in both asymptomatic phase and in late-stage disease are represented in the dataset). All individuals were healthy by inclusion and they were recruited from different parts of Guinea-Bissau. Sample sets used in this study are further described in the section of Sample sets.

Amplification and sequencing

Viral RNA was extracted and purified from blood plasma samples, using RNeasy Lipid Tissue Mini Kit (Qiagen, Stockholm, Sweden) with minor modifications from the manufacturer's instructions. Briefly, 200 µl of blood plasma was disrupted in 2000 µl Qiazol and 10 µg Carrier RNA was added (Qiagen). The aqueous phase was loaded onto a spin column by multiple loading steps. RNA was eluted in 40 µl of RNase-free water and treated with DNase I (Fermentas, Helsingborg, Sweden). Viral RNA was reverse transcribed using gene-specific primers, and the V1-V3 region was amplified by a nested PCR approach (The SuperScript™ III One-Step RT-PCR System with Platinum® *Taq* DNA Polymerase and Platinum® *Taq* DNA Polymerase High Fidelity, Invitrogen, Copenhagen, Denmark) according to the manufacturer's instructions using primers JE12F and V3A_R2 for one-step RT-PCR and E20A_F and JA169 for nested PCR [38, 39]. The amplified V1-V3 region of approximately 940 base pairs (nucleotides 6430 to 7374 in HXB2; GenBank accession number K03455) was cloned using the InsTAclone cloning system (Fermentas) and TOP10 cells (Invitrogen). Twelve colonies were routinely picked from each sample and the cloned fragments were amplified with Platinum® *Taq* DNA Polymerase High Fidelity (Invitrogen) using conventional M13 primers (-20 and -24). Individual clones were purified and sequenced using BigDye Terminator v1.1

Cycle Sequencing Kit (Applied Biosystems, Stockholm, Sweden) according to the manufacturer's instructions using primers E20A_F and JA169 [38].

Explorative sequence analysis

Sequences were assembled, and contigs were analyzed with CodonCode Aligner v1.5.2 (CodonCode Corporation, Dedham, USA). Only sequences with open reading frames were subjected to further analysis. A multiple alignment of all 1,562 sequences was created in MEGA4 using the Clustal algorithm [40-42]. The alignment was trimmed in the 3' end into complete codons, realigned at the protein level, and codon stripped to a final length of 654 bp. An explorative neighbor-joining (NJ) tree was constructed in MEGA4, using a maximum composite likelihood substitution model (default settings), to verify that all clones were of patient-specific origin (data not shown).

PHI-test and subtype determination

Sequence recombinants violates phylogenetic inference and can lead to misinterpretations of analyzed data [43]. We used the pairwise homoplasy index (PHI) as implemented in Splitstree v4.10 to screen the 1,562 sequences for instances of intra-patient recombination events as descried by Salemi *et al.* [44-46]. One hundred and eighty five potential recombinants were found and removed to a final non-recombinant dataset of 1,377 sequences representing 156 sample time-points of 82 HIV-1 infected individuals. To determine the HIV-1 subtype, one clone from each patient were randomly chosen to a sample set of 82 V1-V3 sequences and aligned with a reference sequence dataset of all major subtypes, subsubtypes and CRFs (downloaded from Los Alamos Sequence Database) in MEGA4, as described. An NJ tree was constructed in MEGA4 with complete deletion of gap positions and by using a maximum composite likelihood substitution model with heterogeneous pattern among lineages with a gamma distribution of 1.0. The phylogenetic reconstruction was bootstrapped 1,000 times to separate sequences of different subtypes. The gp120 V1-V3 region of CRF02_AG is subtype A-derived, and separation of subclusters belonging to either CRF02_AG or subtype A were not possible in this tree. To further characterize these sequences we realigned our sequences with an extended HIV-1 A-like reference sequence dataset (including the subsubtype A3 which were not included in the original dataset but has been reported to circulate in West Africa, Table S3) using $PRANK_{+F}$ with a NJ tree (constructed in MEGA4) as guide tree [24, 47]. The $PRANK_{+F}$ algorithm align sequences using phylogenetic information and has been shown to align sequences in a more evolutionary sound way, and was therefore used instead of the Clustal algorithm in this detailed analysis [47]. The alignment was manually edited and codon-stripped to a final sequence length of 681 nucleotides. A best-fitting nucleotide substitution model for the dataset was estimated using the Akaike information criterion (AIC) as implemented in Modeltest v3.6 [48]. A maximum-likelihood (ML) phylogenetic tree was constructed using the inferred model, GTR+I+G, with Garli v0.951 (www.bio.utexas.edu/faculty/antisense/garli/Garli.html) [49]. This method efficiently maximizes the tree $\log_e$ likelihood by using a genetic algorithm implementing the nearest neighbor interchange (NNI) and the subtree pruning regrafting (SPR) algorithms on a random starting tree to simultaneously find and optimize the topology and branch lengths [49, 50]. Support for internal branches was obtained by the ML-based zero-branch-length test (ZBLT) as implemented in PAUP* v4.0b10 [51]. We also inferred Bayesian phylogenies under the selected model using MrBayes v3.1.2 [52]. Two runs of four chains each (one cold and three heated, temp=0.20) were run for 50 million generations, with a

burn-in of 25%. The results of the two runs were combined in Logcombiner v1.5.3 [5]. Convergence was assessed by calculating the effective sampling size (ESS) using Tracer v1.5 [53]. All parameter estimates showed ESS values higher than 200. A final Bayesian majority-rule consensus tree was obtained by TreeAnnotator v1.5.3 with a burn-in of 25%.

Recombination analysis

Five of the subtype A-like sequences fell in between the CRF02_AG and the A3 monophyletic clusters. Scattering between clusters or long branch lengths are characteristics typical for recombinant sequences [27]. To further characterize the dataset we first performed BootScan analysis to determine possible recombination breakpoints, using Simplot v3.5. NJ trees with a 200 bp moving window along the sequence alignment in 20 bp increments were constructed, and a transition/transversion ratio of 1.5 was used with a F84 model of evolution [54]. Two subgroups of A3 and CRF02_AG sequences were used as putative parental sequences, whereas a subgroup of subtype C sequences was used as outgroup. The consensus sequences of the subgroups were used in the analysis and 1,000 bootstrap replicates were generated for each analysis. Thirteen sequences from Guinea-Bissau showed mixed clustering between HIV-1 CRF02_AG and subsubtype A3 in the BootScan analysis, and displayed profound recombination breakpoints. We then examined each putative recombinant by splitting the dataset at each identified recombination breakpoint. Maximum likelihood and Bayesian trees were reconstructed as described, and sequences that clustered partly within the CRF02_AG cluster and partly within the subsubtype A3 cluster were considered to be A3/CRF02_AG recombinants.

Sample sets for estimation of evolutionary rates and dates

Two sample sets with sequences from Guinea-Bissau were constructed; one with the CRF02_AG, and one with subsubtype A3 sequences. Since the sequences were intended to be used for estimation of evolutionary rates and dates we aimed at the widest range of collection years because a long interval of sampling years naturally provides more information about rates and divergence times than a short interval [55]. To construct a reference sequence dataset for each of the two datasets we used the basic local alignment search tool (BLAST) (www.ncbi.nlm.hih.gov/BLAST) to select for the 10 most similar hits for each Guinea-Bissau-specific sequence. Sequences sampled from the same patients were removed to a final reference dataset of 36 and 46 sequences for CRF02_AG and A3, respectively. Accession numbers of the sequences can be found in Tables S4 and S5.

Cluster identification and estimation of evolutionary rates and dates

The two sample sets (CRF02_AG and A3) were aligned with the corresponding reference datasets using $PRANK_{F+}$, as described in section *PHI-test and subtype determination*. The best fitting nucleotide substitution model was determined using Modeltest v3.6 (AIC). Maximum likelihood (ML) and Bayesian trees were inferred using Garli v0.951 and MrBayes v3.1.2, respectively. Statistical support for internal branches were determined by the ML-based zero-branch-length test (ZBLT), as implemented in PAUP* v4.0b10, for ML-phylogenies, and by posterior probability values for Bayesian maximum clade credibility trees (MCC) (determined by TreeAnnotator v1.5.3, as described).

Estimates of evolutionary rates (nucleotide substitutions $site^{-1}$ $year^{-1}$) and timing of the most recent common ancestors (tMRCA) of the datasets and clusters were performed

using a Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST v1.5.3 [5]. The time span for the HIV-1 CRF02_AG and A3 datasets were 1993-2008 and 1995-2008, respectively. We used the SRD06 model implemented in BEAST v1.5.3 (a HKY85 nucleotide substitution model, with four category gamma distributed rate variation among sites, and two partitions in codon positions (1^{st}+2^{nd}, 3^{rd} codon)), which has been shown to be the best performing model for analysis of most viral datasets [56]. A relaxed clock with an uncorrelated lognormal distributed prior was used with the Bayesian skyline plot growth model (10 grouped intervals separated by coalescence events) as demographic model. The BSP allows the effective population size to vary between coalescence events, and does not make any restrictive prior assumptions on the demographic history. Explorative analyses of other coalescent demographic models resulted in highly similar estimates of evolutionary rates and origins (Table S2). Two independent runs of 20-100 million generations were run and samples of trees and parameter estimates were sampled every 2,000-10,000 generation. Convergence was assessed in Tracer v1.5, and the two runs were combined using Logcombiner v1.5.3 with a burn-in of 10%. A final maximum clade credibility tree was determined for each demographic model using TreeAnnotator v1.5.3. Alignments, xml-files and further details about the settings are available from the authors upon request.

Demographic analysis and estimation of proportion of lineages

Demographic growth patterns in Guinea-Bissau were examined by analysis of the major clusters identified for HIV-1 CRF02_AG and A3, respectively. The demographic patterns of the entire CRF02_AG and subsubtype A3 epidemics were estimated based on the datasets including both sequences from Guinea-Bissau and reference sequences. We used the Bayesian skyline plot (BSP) population growth model with 10 grouped intervals as implemented in BEAST v1.5.3 [57]. Since the BSP model allows the effective population size to vary between coalescent events, the model avoids making restrictive prior assumptions about the demographic history. The MCMC analyses were run for 25-100 million generations and samples of trees and parameter estimates were sampled every 2,500-10,000 generation. Convergence was assessed in Tracer v1.5, and the two runs were combined using Logcombiner v1.5.3. A final MCC tree was determined as described.

To study the proportion of lineages existing at different time points in the Guinea-Bissau epidemic we used the program TreeStat v1.1 [58]. First, we aligned the CRF02_AG and A3 sequences of Guinea-Bissau, respectively, using $PRANK_{F+}$ as described. The SRD06 model with the BSP population growth model with 10 grouped intervals was run for 20-100 million generations with tree and parameter sampling every 2,000-10,000 generation. In this analysis, the posterior distribution of trees was used to examine the median percentage of 2008 years strains (CRF02_AG and A3, separately) in Guinea-Bissau present in the examined years. This estimate gives an indication of changes in strain diversity during the time-periods under study.

Phylogeographic analysis

To study different migration events to and from Guinea-Bissau, as well as within Guinea-Bissau we used a recently introduced Bayesian phylogeographic model employing the Bayesian stochastic search variable selection (allowing exchange rates in the Markov model to be zero), as described by Lemey *et al*. [4]. In this methodology information about both sample time-points (temporal) and geographical locations (spatial) are used together with genetic information to infer

diffusion processes among discrete locations in timed coalescence phylogenies. This makes it possible to test hypotheses about spatial dynamics and address phylogenetic uncertainty in a statistical efficient fashion. The two datasets (CRF02_AG and A3) from the cluster analyses were used. Too many geographic localities in a dataset forces the diffusion process to average over a very high state space, which can be difficult, and may lead to uncertainties in the model (P. Lemey, personal communication). To reduce the state space for the diffusion process we removed non-African sequences to two datasets of 68 CRF02_AG (locations: Guinea-Bissau, Senegal, Mali, Ghana, Nigeria and Cameroon), and 21 subsubtype A3 (locations: Guinea-Bissau, Senegal, Mali, Cameroon) sequences. To study the diffusion process within Guinea-Bissau, Guinea-Bissau-derived sequences were further subdivided into four different localities (Central (Bissau), Northwest, Northeast, and South). Each phylogeographic analysis was run for 25-100 million generations with tree and parameter sampling every 2,500-10,000 generation in BEAST v1.5.3. Two independent runs for each setup were assessed for convergence in Tracer v1.5 with a burn-in of 10%, and combined using Logcombiner v1.5. The results were summarized in an MCC tree, and further visualized by converting the tree into a keyhole markup language (KML) file suitable for viewing in Google Earth (http://earth.google.com). Inferred states with node state probabilities >80% were considered as well supported. Alignments, xml-files and further details about the settings are available from the authors upon request.

To investigate the spatial admixture in our phylogeographic analyses and statistically measure trait associations in the phylogeny we used the parsimony score (PS) and the association index (AI), as implemented in BaTS [59]. BaTS incorporates statistical error arising from phylogenetic uncertainty by analysing the posterior distribution of trees (PDT). By comparing the PDT with a randomized null distribution, BaTS generates statistics based on the proportion of trees from the null distribution equal to, or more extreme than, the median posterior estimate of the statistics from the PDT (i.e. do closely related taxa share the same trait values to a higher extent than we would expect by chance alone?).

Nucleotide sequence accession numbers

Nucleotide sequences were deposited in GenBank under the following accession numbers: *The sequences have been deposited to GenBank but the accession numbers were not yet recieved by the stop-press of this doctoral dissertation.*

Acknowledgements

We thank Anders Kvist, Lund University, Sweden, for help in writing computer scripts for annotating sequences and sequence sampling procedure and for critically reviewing the manuscript, and Antonio J. Biague and Zacarias J. da Silva, National Public Health Laboratory, Bissau, Guinea-Bissau for being responsible of sample collection and serology. We also thank Babetida N'Buna, Aquilina Sambu, Eusebio Ieme, Isabel da Costa, Jacqueline and Ana Monteiro Watche at the Health Station of the 2 esquadra, Bissau, and Cidia Camara, Braima Dabo (deceased), Carla Pereira, Julieta Pinto Delgado, Leonvengilda Fernandes Mendes, Ana Monteiro and Ansu Biai at the National Public Health Laboratory in Bissau. Finally, we thank Helen Linder at Swedish Institute for Infectious Disease Control for help with collecting plasma samples stored in the bio bank, Wilma Martinez-Arias, and SCIBLU Genomics, Lund University, Sweden for technical assistance.

References

1. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, *et al.* **Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960**. *Nature* 2008,455:661-664.
2. Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, *et al.* **Chimpanzee reservoirs of pandemic and nonpandemic HIV-1**. *Science* 2006,313:523-526.
3. Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. **The challenge of HIV-1 subtype diversity**. *N Engl J Med* 2008,358:1590-1602.
4. Lemey P, Rambaut A, Drummond AJ, Suchard MA. **Bayesian phylogeography finds its roots**. *PLoS Comput Biol* 2009,5:e1000520.
5. Drummond AJ, Rambaut A. **BEAST: Bayesian evolutionary analysis by sampling trees**. *BMC Evol Biol* 2007,7:214.
6. Paraskevis D, Magiorkinis E, Magiorkinis G, Sypsa V, Paparizos V, Lazanas M, *et al.* **Increasing prevalence of HIV-1 subtype A in Greece: estimating epidemic history and origin**. *J Infect Dis* 2007,196:1167-1176.
7. Bello G, Passaes CP, Guimaraes ML, Lorete RS, Matos Almeida SE, Medeiros RM, *et al.* **Origin and evolutionary history of HIV-1 subtype C in Brazil**. *Aids* 2008,22:1993-2000.
8. Dalai SC, de Oliveira T, Harkins GW, Kassaye SG, Lint J, Manasa J, *et al.* **Evolution and molecular epidemiology of subtype C HIV-1 in Zimbabwe**. *Aids* 2009,23:2523-2532.
9. Pybus OG, Barnes E, Taggart R, Lemey P, Markov PV, Rasachak B, *et al.* **Genetic history of hepatitis C virus in East Asia**. *J Virol* 2009,83:1071-1082.
10. Andersson S, Norrgren H, Dias F, Biberfeld G, Albert J. **Molecular characterization of human immunodeficiency virus (HIV)-1 and -2 in individuals from guinea-bissau with single or dual infections: predominance of a distinct HIV-1 subtype A/G recombinant in West Africa**. *Virology* 1999,262:312-320.
11. Cham F, Heyndrickx L, Janssens W, Van der Auwera G, Vereecken K, De Houwer K, *et al.* **Study of HIV type 1 gag/env variability in The Gambia, using a multiplex DNA polymerase chain reaction**. *AIDS Res Hum Retroviruses* 2000,16:1915-1919.
12. Montavon C, Toure-Kane C, Liegeois F, Mpoudi E, Bourgeois A, Vergne L, *et al.* **Most env and gag subtype A HIV-1 viruses circulating in West and West Central Africa are similar to the prototype AG recombinant virus IBNG**. *J Acquir Immune Defic Syndr* 2000,23:363-374.
13. Nkengasong JN, Luo CC, Abouya L, Pieniazek D, Maurice C, Sassan-Morokro M, *et al.* **Distribution of HIV-1 subtypes among HIV-seropositive patients in the interior of Cote d'Ivoire**. *J Acquir Immune Defic Syndr* 2000,23:430-436.
14. Toure-Kane C, Montavon C, Faye MA, Gueye PM, Sow PS, Ndoye I, *et al.* **Identification of all HIV type 1 group M subtypes in Senegal, a country with low and stable seroprevalence**. *AIDS Res Hum Retroviruses* 2000,16:603-609.
15. Peeters M, Esu-Williams E, Vergne L, Montavon C, Mulanga-Kabeya C, Harry T, *et al.* **Predominance of subtype A and G HIV type 1 in Nigeria, with geographical differences in their distribution**. *AIDS Res Hum Retroviruses* 2000,16:315-325.
16. UNAIDS. **Epidemiological Fact Sheet on HIV and AIDS**. In; 2008.
17. Mansson F, Biague A, da Silva ZJ, Dias F, Nilsson LA, Andersson S, *et al.* **Prevalence and incidence of HIV-1 and HIV-2 before, during and after a civil war in an occupational cohort in Guinea-Bissau, West Africa**. *Aids* 2009,23:1575-1582.
18. van Tienen C, van der Loeff MS, Zaman SM, Vincent T, Sarge-Njie R, Peterson I, *et al.* **Two Distinct Epidemics: The Rise of HIV-1 and Decline of HIV-2 Infection Between 1990 and 2007 in Rural Guinea-Bissau**. *J Acquir Immune Defic Syndr* 2009.
19. da Silva ZJ, Oliveira I, Andersen A, Dias F, Rodrigues A, Holmgren B, *et al.* **Changes in prevalence and incidence of HIV-1, HIV-2 and dual infections in urban areas of Bissau, Guinea-Bissau: is HIV-2 disappearing?** *Aids* 2008,22:1195-1202.
20. Mansson F, Alves A, Silva ZJ, Dias F, Andersson S, Biberfeld G, *et al.* **Trends of HIV-1 and HIV-2 prevalence among pregnant women in Guinea-Bissau, West Africa: possible effect of the civil war 1998 1999**. *Sex Transm Infect* 2007,83:463-467.
21. Norrgren H, Andersson S, Biague AJ, da Silva ZJ, Dias F, Naucler A, Biberfeld G. **Trends and interaction of HIV-1 and HIV-2 in Guinea-Bissau, west Africa: no protection of HIV-2 against HIV-1 infection**. *Aids* 1999,13:701-707.
22. Fultz PN, Switzer WM, Schable CA, Desrosiers RC, Silva DP, McCormick JB. **Seroprevalence of HIV-1 and HIV-2 in Guinea Bissau in 1980**. *Aids* 1988,2:129-132.
23. Osmanov S, Pattou C, Walker N, Schwardlander B, Esparza J. **Estimated global distribution and regional spread of**

HIV-1 genetic subtypes in the year 2000. *J Acquir Immune Defic Syndr* 2002,29:184-190.

24. Meloni ST, Kim B, Sankale JL, Hamel DJ, Tovanabutra S, Mboup S, *et al.* **Distinct human immunodeficiency virus type 1 subtype A virus circulating in West Africa: sub-subtype A3**. *J Virol* 2004,78:12438-12445.
25. Meloni ST, Sankale JL, Hamel DJ, Eisen G, Gueye-Ndiaye A, Mboup S, Kanki PJ. **Molecular epidemiology of human immunodeficiency virus type 1 sub-subtype A3 in Senegal from 1988 to 2001**. *J Virol* 2004,78:12455-12461.
26. Hamel DJ, Sankale JL, Eisen G, Meloni ST, Mullins C, Gueye-Ndiaye A, *et al.* **Twenty years of prospective molecular epidemiology in Senegal: changes in HIV diversity**. *AIDS Res Hum Retroviruses* 2007,23:1189-1196.
27. Mild M, Esbjornsson J, Fenyo EM, Medstrand P. **Frequent intrapatient recombination between human immunodeficiency virus type 1 R5 and X4 envelopes: implications for coreceptor switch**. *J Virol* 2007,81:3369-3376.
28. Galetto R, Moumen A, Giacomoni V, Veron M, Charneau P, Negroni M. **The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo**. *J Biol Chem* 2004,279:36625-36632.
29. Kanki PJ, M'Boup S, Ricard D, Barin F, Denis F, Boye C, *et al.* **Human T-lymphotropic virus type 4 and the human immunodeficiency virus in West Africa**. *Science* 1987,236:827-831.
30. Fauci AS. **The AIDS epidemic--considerations for the 21st century**. *N Engl J Med* 1999,341:1046-1050.
31. Weber J. **HIV and sexually transmitted disease**. *Br Med Bull* 1998,54:717-729.
32. Awumbila M. **Intra-Regional Migration In West Africa: Changing Patterns And Dynamics**. In. Rotterdam: International Institute of Social Studies; 2009:8-10.
33. Adepoju A. **Patterns of Migration in West Africa**. Accra: Sub-Saharan Publishers; 2005.
34. Konings FA, Burda ST, Urbanski MM, Zhong P, Nadas A, Nyambi PN. **Human immunodeficiency virus type 1 (HIV-1) circulating recombinant form 02_AG (CRF02_AG) has a higher in vitro replicative capacity than its parental subtypes A and G**. *J Med Virol* 2006,78:523-534.
35. Njai HF, Gali Y, Vanham G, Clybergh C, Jennes W, Vidal N, *et al.* **The predominance of Human Immunodeficiency Virus type 1 (HIV-1) circulating recombinant form 02 (CRF02_AG) in West Central Africa may be related to its replicative fitness**. *Retrovirology* 2006,3:40.
36. Andersson S, Norrgren H, da Silva Z, Biague A, Bamba S, Kwok S, *et al.* **Plasma viral load in HIV-1 and HIV-2 singly and dually infected individuals in Guinea-Bissau, West Africa: significantly lower plasma virus set point in HIV-2 infection than in HIV-1 infection**. *Arch Intern Med* 2000,160:3286-3293.
37. Norrgren H, Andersson S, Naucler A, Dias F, Johansson I, Biberfeld G. **HIV-1, HIV-2, HTLV-I/II and Treponema pallidum infections: incidence, prevalence, and HIV-2-associated mortality in an occupational cohort in Guinea-Bissau**. *J Acquir Immune Defic Syndr Hum Retrovirol* 1995,9:422-428.
38. Leitner T, Korovina G, Marquina S, Smolskaya T, Albert J. **Molecular epidemiology and MT-2 cell tropism of Russian HIV type 1 variant**. *AIDS Res Hum Retroviruses* 1996,12:1595-1603.
39. Esbjornsson J, Mansson F, Martinez-Arias W, Vincic E, Biague AJ, da Silva ZJ, *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*,7:23.
40. **Los Alamos Sequence Database**. In.
41. Tamura K, Dudley J, Nei M, Kumar S. **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0**. *Mol Biol Evol* 2007,24:1596-1599.
42. Thompson JD, Higgins DG, Gibson TJ. **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucleic Acids Res* 1994,22:4673-4680.
43. Schierup MH, Hein J. **Consequences of recombination on traditional phylogenetic analysis**. *Genetics* 2000,156:879-891.
44. Bruen TC, Philippe H, Bryant D. **A simple and robust statistical test for detecting the presence of recombination**. *Genetics* 2006,172:2665-2681.
45. Huson DH, Bryant D. **Application of phylogenetic networks in evolutionary studies**. *Mol Biol Evol* 2006,23:254-267.
46. Salemi M, Gray RR, Goodenow MM. **An exploratory algorithm to identify intra-host recombinant viral sequences**. *Mol Phylogenet Evol* 2008,49:618-628.
47. Loytynoja A, Goldman N. **Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis**.

Science 2008,320:1632-1635.
48. Posada D, Crandall KA. **MODELTEST: testing the model of DNA substitution**. *Bioinformatics* 1998,14:817-818.
49. Zwickl DJ. **Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion**; 2006.
50. Lewis PO. **A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data**. *Mol Biol Evol* 1998,15:277-283.
51. Swofford DL. **PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.** . In: Sinauer Associates, Sunderland, Massachusetts.; 2003.
52. Ronquist F, Huelsenbeck JP. **MrBayes 3: Bayesian phylogenetic inference under mixed models**. *Bioinformatics* 2003,19:1572-1574.
53. Rambaut A DA. **Tracer v1.5** In; 2007.
54. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG*, et al.* **Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination**. *J Virol* 1999,73:152-160.
55. Seo TK, Thorne JL, Hasegawa M, Kishino H. **A viral sampling design for testing the molecular clock and for estimating evolutionary rates and divergence times**. *Bioinformatics* 2002,18:115-123.
56. Shapiro B, Rambaut A, Drummond AJ. **Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences**. *Mol Biol Evol* 2006,23:7-9.
57. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. **Bayesian coalescent inference of past population dynamics from molecular sequences**. *Mol Biol Evol* 2005,22:1185-1192.
58. Rambaut A, Drummond AJ. **Tree Statistic Calculation Tool Version 1.1, 2005-2007**. In; 2005.
59. Parker J, Rambaut A, Pybus OG. **Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty**. *Infect Genet Evol* 2008,8:239-246.

Supplementary information

Table S1. Sampling year, date of seroconversion, sample place, and HIV-1 clade of the 82 analyzed study subjects.

Sample*	Sample year	SC year†	Geography	Clade‡	Sample*	Sample year	SC year†	Geography	Clade‡
DL1996H_13	2000	1990	Bissau	CRF02_AG	DL3166F_4	2003	1999	Bissau	A3/AG
DL2004F_4	2001	2001	Northwest	CRF02_AG	DL3169F_7	2004	1999	Bissau	CRF02_AG
DL2014F_1	1995	1990	Bissau	A3	DL3170D_5	1997	1993	Bissau	CRF02_AG
DL2066G_8	2002	2000	Bissau	CRF02_AG	DL3234J_3	2006	1999	Bissau	A3/AG
DL2075I_1	2007	1998	Bissau	CRF02_AG	DL3247F_2	2007	1999	Bissau	CRF02_AG
DL2089J_9	2003	1997	Bissau	CRF02_AG	DL3288H_5	2005	NA	Northwest	CRF02_AG
DL2096D_7	1996	1993	Bissau	C	DL3312C_1	1994	1993	Northwest	CRF02_AG
DL2102F_1	1998	1994	Bissau	CRF02_AG	DL3339D_12	2001	1998	Northeast	CRF02_AG
DL2111E_2	1998	1994	Bissau	CRF02_AG	DL3372I_10	2004	2000	Northeast	A3/AG
DL2164F_4	2002	1997	Bissau	A3	DL3406J_5	2006	2000	Northwest	A3
DL2198F_3	1996	1991	Bissau	CRF02_AG	DL3442D_8	2001	2000	Northeast	A3/AG
DL2249E_8	1997	1996	Bissau	CRF02_AG	DL3468E_10	2007	1994	Bissau	C
DL2315F_4	2000	1997	Northwest	CRF02_AG	DL3556C_9	1997	1995	Bissau	CRF02_AG
DL2325J_8	2008	2000	Bissau	CRF02_AG	DL3633G_2	2003	1999	Bissau	CRF02_AG
DL2339E_4	2003	1992	Northwest	CRF02_AG	DL3721C_3	1997	1994	Bissau	A3/AG
DL2365I_3	2002	1996	Bissau	A3	DL3733D_9	2000	1997	Northwest	CRF02_AG
DL2391B_5	1993	1991	Bissau	CRF02_AG	DL3766D_9	1997	1994	Bissau	C
DL2401M_5	2004	1996	Bissau	CRF02_AG	DL3860H_3	2008	1995	Bissau	A3
DL2462H_4	2001	1999	Bissau	CRF02_AG	DL3869G_3	2003	NA	Bissau	A3
DL2470E_5	2000	1998	Bissau	A3	DL3895C_2	1996	1995	Bissau	CRF02_AG
DL2568E_9	2003	2003	Bissau	A3/AG	DL3938D_3	1998	1995	Bissau	CRF02_AG
DL2544F_4	2000	1999	Northwest	CRF02_AG	DL3946E_4	2004	2001	South	CRF02_AG
DL2594J_3	2005	2001	Northeast	A3	DL3981C_10	1998	1995	Bissau	CRF02_AG
DL2596E_4	2000	1995	South	A3	DL4023G_7	2006	1999	Bissau	CRF02_AG
DL2640I_7	2004	2003	Bissau	CRF02_AG	DL4084F_2	2003	1999	Bissau	A1
DL2673D_4	1998	1993	South	CRF02_AG	DL4169F_9	2002	1996	Bissau	CRF02_AG
DL2713H_7	2007	2002	Northwest	CRF02_AG	DL4214D_8	2002	NA	Bissau	C
DL2747I_2	2005	1994	Northwest	CRF02_AG	DL4248G_5	2005	NA	Bissau	A3
DL2766C_11	1995	1992	Northwest	CRF02_AG	DL4303D_7	2002	1999	Bissau	CRF02_AG
DL2829F_2	2006	1999	Northeast	CRF06_cpx	DL4422B_4	2003	NA	Northeast	CRF02_AG
DL2846F_8	2001	1993	Bissau	A3	DL4477D_9	2001	NA	Northeast	CRF02_AG
DL2853E_2	1998	1993	Bissau	CRF02_AG	DL4525G_2	2006	1999	Bissau	A3
DL2908G_8	2000	1997	Bissau	CRF02_AG	DL4632E_5	2003	1997	South	CRF02_AG
DL2920H_2	2004	1998	Bissau	CRF02_AG	DL4957C_5	2005	NA	Bissau	A3
DL3004H_11	2003	1999	Bissau	CRF02_AG	DL5342B_4	2007	NA	South	A3/AG
DL3018E_10	2003	2001	South	A3/AG	DL6324B_13	2007	NA	Bissau	CRF02_AG
DL3037E_8	2005	1996	Northeast	A3/AG	DL11967A_7	2005	NA	Bissau	A3/AG
DL3039D_12	1998	1993	Northwest	A3/AG	DL11968A_10	2005	NA	Bissau	A3
DL3071F_10	2002	2001	Bissau	A3/AG	DL11969A_4	2005	NA	Bissau	A3
DL3087E_5	2001	NA	Bissau	CRF02_AG	DL11970A_1	2006	NA	Bissau	A3/AG
DL3098I_6	2007	2002	Bissau	CRF02_AG	DL11971A_1	2006	NA	Bissau	A3

*Samples identification number. The last number (i.e. _1) represents the sequence clone number.
†Determined seroconversion date, estimated as the date between the last HIV-1 negative sample and the first HIV-1 positive sample. NA = not applicable.
‡This column shows the HIV-1 subtype, CRF or other form as determined by the phylogenetic analysis.

Table S2. Estimated substitution rates and dates for the CRF02_AG and subsubtype A3 datasets by different Bayesian demographic models.

Clade	Clock model*	Tree prior†	μ (10^{-3})‡	tMRCA§	Cluster 1‖	Cluster 2‖	Cluster 3‖	Cluster 4‖	Cluster 5‖
CRF02_AG	SC	CS	4.27 (3.33-5.25)	1970 (1962-1976)	1983 (1977-1989)	1979 (1974-1984)	1984 (1979-1988)	1982 (1977-1987)	1985 (1980-1988)
		EXP	3.59 (2.71-4.50)	1967 (1958-1974)	1979 (1972-1985)	1974 (1967-1980)	1979 (1973-1984)	1978 (1971-1983)	1980 (1974-1985)
		LOG	3.51 (2.58-4.43)	1966 (1957-1974)	1979 (1972-1984)	1974 (1966-1980)	1978 (1971-1983)	1977 (1969-1983)	1980 (1973-1985)
		BSP	3.76 (2.87-4.68)	1969 (1961-1976)	1980 (1974-1985)	1976 (1969-1981)	1980 (1974-1984)	1979 (1972-1984)	1981 (1975-1985)
	RC	CS	4.37 (3.31-5.44)	1970 (1962-1977)	1984 (1977-1990)	1980 (1974-1985)	1984 (1980-1988)	1983 (1977-1988)	1985 (1980-1989)
		EXP	3.60 (2.67-4.61)	1967 (1957-1975)	1979 (1972-1986)	1974 (1966-1981)	1979 (1972-1984)	1978 (1970-1984)	1980 (1973-1985)
		LOG	3.50 (2.49-4.53)	1967 (1955-1975)	1979 (1971-1986)	1974 (1964-1980)	1978 (1970-1984)	1977 (1968-1983)	1980 (1972-1985)
		BSP	3.80 (2.84-4.76)	1969 (1960-1976)	1981 (1974-1986)	1976 (1968-1982)	1980 (1974-1984)	1979 (1972-1984)	1981 (1975-1985)
A3	SC	CS	3.87 (2.45-5.35)	1978 (1966-1985)	1982 (1972-1989)	NA¶	NA¶	NA¶	NA¶
		EXP	3.41 (2.10-4.90)	1977 (1963-1985)	1980 (1967-1987)	NA¶	NA¶	NA¶	NA¶
		LOG	2.92 (1.40-4.48)	1973 (1948-1985)	1976 (1953-1987)	NA¶	NA¶	NA¶	NA¶
		BSP	3.35 (1.87-4.76)	1976 (1960-1985)	1979 (1965-1988)	NA¶	NA¶	NA¶	NA¶
	RC	CS	3.87 (2.24-5.52)	1977 (1962-1986)	1983 (1970-1990)	NA¶	NA¶	NA¶	NA¶
		EXP	3.34 (1.88-4.96)	1977 (1961-1986)	1979 (1975-1988)	NA¶	NA¶	NA¶	NA¶
		LOG	2.78 (0.95-4.41)	1971 (1931-1985)	1974 (1938-1987)	NA¶	NA¶	NA¶	NA¶
		BSP	3.27 (1.61-4.93)	1975 (1955-1986)	1979 (1960-1988)	NA¶	NA¶	NA¶	NA¶

*SC: Strict clock. RC: Relaxed clock (uncorrelated lognormal prior).

†Demographic model. CS: Constant size. EXP: Exponential growth. LOG: Logistic growth. BSP: Bayesian skyline plot.

‡μ = median substitution rate in substitutions site-1 year-1. The boundaries of the 95% higher posterior density interval are given within brackets.

§tMRCA = median time of the most recent common ancestor. The boundaries of the 95% higher posterior density interval are given within brackets.

‖Dating of Guinea-Bissau specific HIV-1 CRF02_AG introductions. The boundaries of the 95% higher posterior density interval are given within brackets.

¶NA=not applicable.

Table S3. Accession numbers of reference sequences representing subsubtypes A1, A2, A3 and the CRF02_AG used for detailed subtyping of the subtype A-like sequences from Guinea-Bissau.

AF286237	AY521629	DQ676872	AB253429
AF286238	AY521630	AF004885	AY271690
AF286241	AY521631	AB253421	L39106

Table S4. Accession numbers of reference sequences representing the CRF02_AG used for Guinea-Bissau-specific cluster identification.

AB049811	AJ251056	AY231153	DQ926899
AB286863	AJ277822	AY371122	EU480455
AF063223	AJ286133	AY371123	EU480459
AF069933	AJ866556	AY371126	EU480469
AF069939	AM279352	AY371137	EU513187
AF107770	AM279360	AY371140	EU786671
AF184155	AM279361	AY444810	L22939
AF377954	AY151002	AY736840	L23064
AF377955	AY231152	DQ313244	L39106

Table S5. Accession numbers of reference sequences representing subsubtype A3 used for Guinea-Bissau-specific cluster identification.

AB480045	AM279347	AY521633	EU480459
AB485633	AM279354	AY669780	EU480474
AF063223	AM279360	AY734555	EU513189
AF069933	AM279361	AY772995	EU853076
AF069939	AY231152	AY829207	FJ388943
AF184155	AY371140	AY905595	FJ866112
AF364109	AY371143	AY905602	L22939
AF457064	AY444810	DQ208445	L23064
AF457084	AY521629	DQ208497	L39106
AF484491	AY521630	EU110093	U15119
AJ286133	AY521631	EU191612	
AJ866556	AY521632	EU480457	

NATURAL INHIBITION OF HIV-1 DISEASE PROGRESSION BY CONTEMPORANEOUS HIV-2 INFECTION

PAPER II

Joakim Esbjörnsson*
Fredrik Månsson*
Anders Kvist
Per-Erik Isberg
Antonio J. Biague
Zacarias J. da Silva
Eva Maria Fenyö
Hans Norrgren
Patrik Medstrand

*These authors contributed equally to this study

Submitted

II

Natural inhibition of HIV-1 disease progression by contemporaneous HIV-2 infection

Joakim Esbjörnsson[1*†], Fredrik Månsson[2*†], Anders Kvist[3], Per-Erik Isberg[4], Salma Nowroozalizadeh[5], Antonio J. Biague[6], Zacarias J. da Silva[6], Marianne Jansson[5,7], Eva Maria Fenyö[7], Hans Norrgren[8] and Patrik Medstrand[1]

[1]Department of Experimental Medical Science, Section of Molecular Virology, Lund University, Lund, Sweden. [2]Department of Clinical Sciences, Malmö, Infectious Diseases Research Unit, Lund University, Malmö, Sweden. [3]Department of Clinical Sciences, Section of Oncology, Lund University and Skåne University Hospital, Lund, Sweden. [4]Department of Statistics, Lund University School of Economics and Management, Lund University, Lund, Sweden. [5]Department Microbiology, Tumor and Cell biology, Karolinska Institutet, Stockholm Sweden. [6]National Public Health Laboratory, Bissau, Guinea-Bissau. [7]Department of Laboratory Medicine Lund, Division of Medical Microbiology, Lund University, Lund, Sweden. [8]Department of Clinical Sciences, Division of Infection Medicine, Lund University, Lund, Sweden

Progressive immune dysfunction and AIDS develop in most cases of untreated HIV-1 infection, but only in approximately 25-30% of HIV-2 infected individuals. Here we show that HIV-1 and HIV-2 dual-infected individuals have a 46% lower mortality rate and a 53% longer progression-time to AIDS compared with HIV-1 single-infected individuals, by analyses of 223 seroincident cases from a unique cohort with long follow-up (20 years). CD4⁺ T cell counts were consistently higher at corresponding time-points after infection among dual-infected individuals, reflecting the slower disease progression rate at the cellular immune level. At the viral level, we found that the HIV-1 genetic diversity at comparable time-points was significantly lower in dual-infected individuals. Understanding the underlying mechanisms responsible for the natural inhibitory effect by HIV-2 against HIV-1 disease progression rate could be important for the development of future HIV-1 vaccines and therapeutics.

More than 60 million individuals have been infected by HIV since the discovery of the virus in 1983, and despite tremendous efforts, there is no cure or effective vaccine against the virus at this time. To date, two genetically related but distinct human lentiviruses, HIV-1 and HIV-2, have been described[1,2]. Whereas HIV-1 is pandemic, HIV-2 is mainly confined to West Africa. Both viruses share similar transmission routes, cellular targets and AIDS causatives. However, HIV-2 infection is characterized by a much longer asymptomatic stage, lower plasma viral load, slower decline in CD4⁺ T-cell counts, and lower mortality rate[3-6].

In West Africa, where HIV-2 is present, dual infection with HIV-1 and HIV-2 has been reported with a prevalence of 0-3.2%[7,8]. In 1995, Travers *et al.* reported a possible protective effect of HIV-2 against subsequent HIV-1 infection in commercial sex workers in

***Corresponding authors: Joakim Esbjörnsson, joakim.esbjornsson@med.lu.se; Fredrik Månsson, fredrik.mansson@med.lu.se**

†These authors contributed equally to this study

Senegal[9]. However, this finding could not be verified in other cohorts from West Africa[10-14]. Although several studies have reported that HIV-2 can alter HIV-1 infectivity and replication *in vitro*[15-17], the natural history of disease progression among HIV-1 and HIV-2 dual-infected individuals is poorly characterized (reviewed in[18]).

Results

Epidemiological analysis

A prospective open cohort of police officers in Guinea-Bissau, West Africa, was followed annually between 1990 and 2009[8,19]. All samples were screened and confirmed for HIV-1 and HIV-2[20,21]. Due to long follow-up times (mean 86 months) and high follow-up rates (mean 73%), 301 HIV seroincident cases were identified in the cohort[8]. Date of seroconversion (SC) was estimated as the time-point half way between the last seronegative sample and the first seropositive sample for HIV-1. In the present study, 223 HIV-1 seroincident individuals were included: 189 were infected by HIV-1 only (159 males and 30 females, referred to as single-infected individuals), and 34 were infected with both HIV-1 and HIV-2 (28 males and six females, dual-infected individuals). All of the individuals were treatment-naïve throughout the study period. The mean ages at HIV-1 SC were 36.4 years (SD 9.6, range 18-60 years) and 39.6 years (SD 8.0, range 20-59 years) for single and dual-infected individuals, respectively. Kaplan-Meier survival analysis for reported mortality due to AIDS showed a median survival time of 140 months (95% CI 94-158 months) for dual-infected individuals compared with 96 months (95% CI 83-108 months) for single-infected individuals (p=0.047, Log-rank test (L-R)) (Fig. 1A). The difference was even more profound when stratifying the analysis by gender (p= 0.005, L-R). The adjusted mortality hazard ratio in dual-infected individuals compared to single-infected individuals was 0.66 (95% CI 0.49-0.90, p=0.008, Wald test) according to a Cox proportional hazards regression model controlling for age at SC (four age groups: <25, 25-34, 35-44, and ≥45 years) in addition to gender.

To investigate differences in disease-free survival, we estimated the progression-time from SC to AIDS. AIDS was determined if one of the following criteria were fulfilled: $CD4^+$ T-cell counts ≤200 per mL, CD4% ≤14%, WHO clinical stage 4, CDC clinical stage C, or reported death with AIDS symptoms. The median time of progression to AIDS was 104 months (95% CI 81-143 months) in dual-infected individuals and 68 months (95% CI 61-80 months) in single-infected individuals

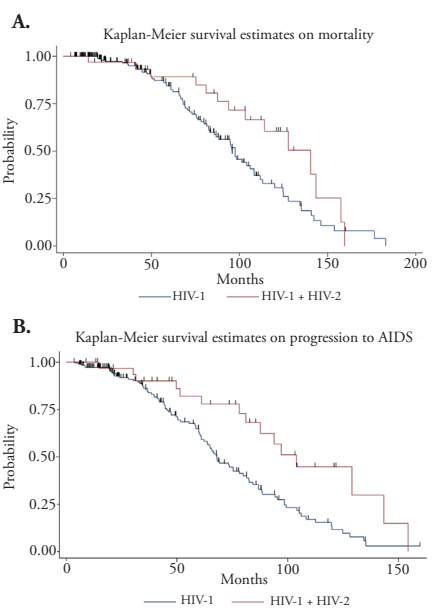


Figure 1. Analysis of survival and progression-time to AIDS. (A) Kaplan-Meier survival analysis for reported mortality due to AIDS. (B) Kaplan-Meier analysis for progression-time to AIDS. Cases where the exit criteria were not fulfilled were censored at their last clinical examination date in both analyses.

(p= 0.004, Kaplan-Meier L-R) (Fig. 1B). As for the mortality analysis, the difference was more pronounced when stratifying by gender (p=0.002, L-R). The adjusted hazard ratio of progression to AIDS in dual-infected vs. single-infected individuals was 0.63 (95% CI 0.47-0.84, p<0.001, Wald test) according to a Cox proportional hazards regression model controlling for gender and age at SC as for the mortality analysis.

$CD4^+$ and $CD8^+$ T cell fluctuations

The natural course of an HIV infection is usually described by three stages. The acute infection is characterized by viremia, rapid decrease in $CD4^+$ T-cell counts (the main cellular target of HIV) and flu-like symptoms. In the asymptomatic stage, plasma viral load is relatively low and the $CD4^+$ T-cell decline is moderate. Finally, in the AIDS stage, viral load increase, the $CD4^+$ T cell count continues to decrease and opportunistic diseases develop due to the dysfunctional immune system. In contrast to the $CD4^+$ T cell count, the level of $CD8^+$ T cells generally increases over the disease course. The duration of the different stages, especially the asymptomatic stage, can differ considerably between infected individuals. In this study, the rate of decline in $CD4^+$ T cell percentage of lymphocytes (CD4%) was similar in single and dual-infected individuals, with an average decline of 1.19% per year (p=0.439, 2-tailed Student's T-test (T-test), Table 1). However, the level of CD4% at comparable time-points after infection was significantly higher in dual than in single-infected individuals (p=0.001, T-test, Table 1), after adjusting for differences in sampling time between individuals.

Next, we examined differences in CD8% over time, and found a slower increase in CD8% levels in dual compared to single-infected individuals (p=0.023, T-test, Table 1). In contrast to the analysis of CD4%, there were no difference in the CD8% levels at the mean time-point after infection (50 months, p=0.788, T-test, Table 1). With the insight that the rate of CD8+ T cell elevation differed between single and dual-infected individuals, we set out to analyse if levels of immune activation markers also differed between the two groups. Since availability of stored peripheral blood lymphocytes was limited, levels of beta-2 microglobulin (b2m) and neopterin in plasma was analysed as markers of immune activation. No differences in plasma levels or kinetics of b2m or neopterin were found between single and dual-infected individuals, instead, considerable variation

Table 1. CD4%[1], CD8%[1] and maximum likelihood estimates of HIV-1 sequence diversity for single and dual-infected individuals.

	Single		Dual		
	Mean	S.E.	Mean	S.E.	p-value[2]
CD4% decline over time (per year)	1.06	0.37	1.59	0.41	0.439
CD4% at the mean sample time-point after HIV-1 SC[3]	18.04	0.92	24.64	1.62	0.001
CD8% increase over time (per year)	4.62	0.80	2.15	0.70	0.023
CD8% at the mean sample time-point after HIV-1 SC[3]	48.44	1.39	47.67	2.84	0.788
Diversity increase over time (substitutions x 10^{-3}/site/year)	1.60	0.92	2.00	1.53	0.812
Diversity at the mean sample time-point after HIV-1 SC[3] (substitutions x 10^{-3}/site)	11.04	1.28	5.67	1.61	0.014

[1]CD4% and CD8% is defined as the $CD4^+$ or $CD8^+$ T cell percentage of the total lymphocyte count. [2]P-values refer to 2-tailed Student's T-test and the uncertainty of estimated means is given by the standard error (S.E.). [3]SC = seroconversion.

between different individuals were noted (able S1).

Molecular evolution of HIV-1

HIV-1 evolution is characterized by high mutation rates, rapid viral turnover and high recombination rates, and can be quantified by diversity (the genetic variation at a given time-point) and divergence (the genetic distance to a reference point, i.e. the founder strain). Several studies have presented evidence of a positive correlation between diversity and time from SC during the asymptomatic stage[22-25]. HIV-1 diversity has also been positively correlated with viral load and viral fitness[26-28]. The divergence rate of HIV-1 has been shown to be relatively constant (reflected by a linear increase in divergence) during the asymptomatic stage in patients[23,29,30]. Studies comparing divergence rates with disease progression rates show conflicting results[29,31-33]. A possible explanation could be that data from different disease stages or from patients of different disease progression groups have been combined in the same analysis.

HIV-1 diversity and divergence patterns were investigated in a subset of 20 single and 12 dual-infected individuals for which the required plasma samples were available. For each individual, two blood plasma samples were selected from the asymptomatic stage. The primary selection criterion was to minimize the variation in time between the two sample time-points (mean 36.5 (S.E. 4.5) months for single and 36.8 (S.E. 4.8) months for dual-infected individuals). Other criteria were to minimize the variation in time from SC for the first sample and to ensure that the first sample time-point was not collected during the acute stage of infection (mean 28.0 (S.E. 3.5) months for single and 38.7 (S.E. 4.9) months for dual-infected individuals). For each plasma sample we sequenced 12 HIV-1 *env* gp120 V1-V3 clones (~940 bp). Phylogenetic trees were inferred using both maximum-likelihood (ML) and Bayesian methods[34,35]. Since analyses based on the two inference methods showed high concordance (Table S2 and S3), only the ML estimates are presented in the following sections. The average increase in HIV-1 sequence diversity over time was similar in single and dual-infected individuals with an average of $1.75x10^{-3}$ substitutions per site per year (p=0.812, T-test, Table 1), whereas the level of diversity at comparable time-points after infection differed between the groups, as shown for the CD4%. After adjusting for differences in sampling time between individuals, we found that the diversity was significantly lower in dual-infected individuals than in single-infected individuals (p=0.014, T-test, Table 1). The "diversity threshold theory" assumes that AIDS develops when diversity exceeds a critical threshold that varies individually[25,36]. As reported above, the average progression-time to AIDS was 68 months in single-infected individuals. At this time-point the estimated mean diversity was $13.52x10^{-3}$ substitutions per site for single-infected individuals (Table S2). Dual-infected individuals were estimated to reach the same mean diversity after 105 months that is close to the average observed progression-time to AIDS of 104 months for this group (Table S3). Thus, the mean diversity threshold is almost identical for dual and single-infected individuals, however, the time to reach the threshold is significantly different.

Despite detailed evolutionary analyses, we found no significant differences between the single and dual-infected groups in HIV-1 divergence rates, HIV-1 rates of synonymous substitutions or non-synonymous substitutions (Table S3).

Discussion

In the present study, we show that HIV-2 has a natural inhibitory effect on HIV-1 disease progression rate *in vivo*. This inhibition was

evident in mortality rates, progression-time to AIDS, at the cellular level of the immune system and at the molecular level of HIV-1 evolution. Our observations are built on unique incident data combined with high follow-up rates and long follow-up times (~20 years of cohort study period). The hypothesis of a cross-reactive and protective effect against HIV-1 by the less pathogenic, but closely related HIV-2 has been debated[8-14,37]. In strong support of our population-based results, experimental studies using the macaque model have shown inhibition against both immunosuppression and SIV-induced disease as a result of contemporaneous HIV-2 infection[38,39]. The mechanisms behind this protection could have several explanations. The fact that the HIV-1 diversity increases at a similar rate in single and dual-infected individuals, but that the diversity at a given time-point is lower in the dual-infected individuals, could be due to inhibitory effects early in the HIV-1 infection. This would result in a lower initial diversity and a longer asymptomatic stage before the "diversity threshold" is reached. Alternatively, HIV-2 could continuously alter the expression of cellular factors that affect the susceptibility of the uninfected cellular environment *in trans*. *In vitro* studies have shown that HIV-2 infection generates higher levels of β-chemokines (the natural ligands of the HIV coreceptor CCR5) in peripheral blood mononuclear cells, and that this can inhibit HIV-1 infection and replication[15-17,40]. In addition, Zheng *et al.* found high frequencies of cross-reactivity between samples of individuals with single HIV-1 or HIV-2 infections when investigating heterologous T-cell responses[41]. They also found that HIV-1 single-infected individuals with the ability to respond to the HIV-2 Gag protein had lower HIV-1 plasma viral loads than those without this cross-reactivity. A second study by this group showed that this was evident also in dual-infected individuals[42]. Antibodies elicited by HIV-2 infections that cross-neutralize HIV-1 have been described[43]. Hence, humoral HIV-2 immune responses could also play a role in controlling HIV-1 in dual-infected individuals. Our observation of different levels of CD4%, but with similar rates of decline during the disease course indicates that determinants for the difference in disease outcome between single and dual-infected individuals may be related to events during the acute infection. Lower systemic immune activation in dual-infected individuals could, however, not be supported by data from our analysis of soluble immune activation markers in longitudinally obtained plasma samples. Instead, considerable inter-patient variation was noted. It has been suggested that levels of soluble immune activation markers could be influenced by different endemic parasite and bacterial diseases present in African populations[44]. Despite the lack of differences in levels or kinetics of the soluble immune activation markers analysed in the current study, the reduced elevation in CD8% among dual-infected individuals suggests that alteration in cellular immune activation may contribute to the disease outcome of these individuals. Our results clearly show that HIV-2 inhibits HIV-1 disease progression during infection. However, whether the underlying mechanism can be correlated to any protective effect against HIV-1 infection remains to be clarified.

Viral load (VL), the density of viral particles in peripheral blood, is an imperfect but important measure of the severity of HIV-1 infection. Although its relationship to viral replicative capacity remains unclear, the positive correlation between high VL and a fast disease progression has been clearly demonstrated. Contrasting results regarding VL comparisons between single and dual-infected individuals have been presented[6,45-47]. However, these data are difficult to interpret due to unknown SC dates, differences in disease stage, and limited follow-up. Since VL

measurements have not been included as a standard procedure in Guinea-Bissau, we were unable to include VL data in our analyses. However, recent studies have suggested that there may be a positive correlation between HIV-1 diversity and VL[26,27]. Diversity has also been positively correlated with HIV-1 replication efficiency and rate of progression to AIDS[28,33]. These results are in line with our results indicating that the HIV-1 diversity at a given time-point is higher in single than in dual-infected individuals. In our material, we found a high concordance between the differences in progression-time to AIDS (36 months) and the mean diversity lag times to reach the diversity threshold (37 months) between the single and dual-infected groups (schematically illustrated in Fig. 2). This further supports the connection between HIV-1 diversity and disease progression[23]. The constant increase in viral diversity in the asymptomatic stage suggests that the immune

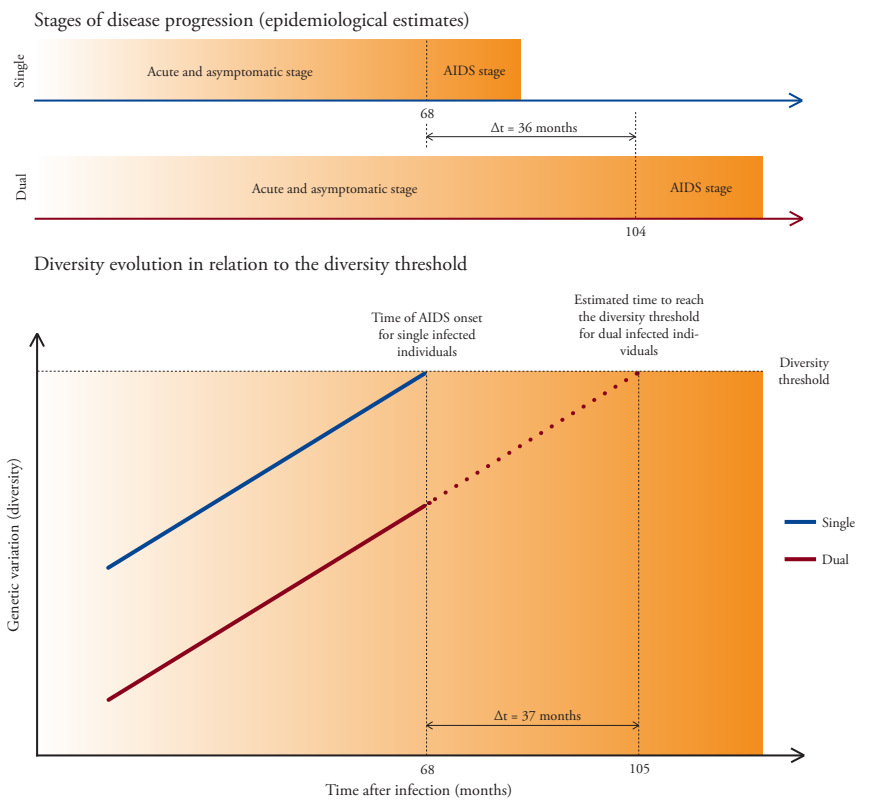


Figure 2. Correlation between disease progression and HIV-1 diversity evolution. Schematic illustration. The epidemiological estimate of progression-time to AIDS for the HIV-1 single-infected individuals (68 months) was used to determine the diversity threshold for development of AIDS. The time to reach that threshold for dual-infected individuals was determined by using the mean diversity rate to extrapolate to the threshold diversity level. Both maximum likelihood (104.93 months) and bayesian estimates (104.95 months) were close to the epidemiological estimate of AIDS onset (104 months).

response is not broad enough to target all co-existing viral variants. Instead, the most common variants are targeted and eradicated, whereas less replication-competent variants are conserved. It has been hypothesized that the increasing viral diversity ultimately results in viral variants with critical epitopes outside the T-cell repertoire, resulting in immune failure and development of AIDS[25,26,36]. This hypothesis supports the "diversity threshold theory" as an explanatory model of HIV-1 diversity evolution, and may serve as a reasonable explanation for the differences that we observed between single and dual-infected individuals.

In summary, we show that HIV-1 and HIV-2 dual-infected individuals have a lower mortality rate and a longer progression-time to AIDS as compared to HIV-1 single-infected individuals. The slower disease progression was reflected on the cellular level of the immune system, with lower levels of $CD4^+$ T cells at comparable time-points and different kinetics in levels of $CD8^+$ T cells among dual-infected individuals. Phylogenetic analysis revealed differences also at the molecular level of HIV-1 evolution with a strong connection to viral diversity. Further investigations of the interplay between HIV-1 and HIV-2 or of any systemic immunologic effects of a contemporaneous HIV-2 infection on HIV-1 pathogenesis could reveal new and critical mechanisms important for the development of future vaccines or therapeutics.

Materials and Methods

Study population of the epidemiological analysis

A prospective open cohort of police officers in Guinea-Bissau was initiated in February 1990. It has continuously included new participants except for temporarily closure from June 1998 until the end of 2002 as a result of the civil war in 1998–1999. However, annual controls of previously included individuals were resumed already in July 2000. All persons with a regular employment in the Guinea-Bissau police force were eligible for the study, which has been voluntary with less than 2% refusal to participate. Blood samples for serology and $CD4^+$ T-cell counts were collected at inclusion and at follow-up visits scheduled at intervals of 12–18 months. Antiretroviral therapy (ART) was recently introduced into Guinea-Bissau, and the few individuals receiving ART were censored from the analysis from the time-point of ART initiation. For this study, inclusion of new participants was performed until December 31, 2007, and the maximum follow-up time was until September 1, 2009. Date of seroconversion (SC) of HIV-1 was estimated as the mid-time-point between the last HIV-1 seronegative sample and the first HIV-1 seropositive sample. A total of 223 HIV-1 seroincident individuals were included. Among those, 189 were infected by HIV-1 only (159 males and 30 females, referred to as single-infected individuals), and 34 were infected with both HIV-1 and HIV-2 (28 males and six females, dual-infected individuals). The mean ages at SC were 36.4 years (SD 9.6, range 18-60 years) and 39.6 years (SD 8.0, range 20-59 years) for single and dual-infected individuals, respectively.

Diagnostic laboratory methods

Serological HIV testing was performed at the National Public Health Laboratory (LNSP), Bissau as described[8,20]. Evaluations of the HIV antibody testing strategy have shown a high concordance between the results obtained by serology and by PCR and a high degree of distinction between HIV-1 and HIV-2[20,21]. In addition, serology of sequential samples of the same infected individuals was without discrepancies, which further strengthens the results of this strategy. The screening assays have been evaluated in parallel to ensure reproducibility between assays[20].

T-lymphocyte subsets were determined at the LNSP, by conventional flow cytometry (Until

2005: FACStrak; Becton Dickinson, San Jose, Calif, with the use of three two-color immunofluorescent reagents, CD45/CD14, CD3/CD4, and CD3/CD8, Simultest, Becton Dickinson; and from 2006 and onwards: CyFlow, Partec, Münster, Germany). Leukocyte counts were performed with a cell counter until 2005 (Coulter Counter CBC5; Coulter Electronics Ltd, Luton, England), and by CyFlow from 2006 and onwards.

Since VL measurements have not been included as a standard procedure in Guinea-Bissau, we were unable to include VL data in our analyses. Due to limitations in sample size and sample availability, we were not able to do any retrospective and valid sequential VL measurements of the samples included in this study.

Epidemiological analysis

Survival analysis was performed for reported mortality and for progression to AIDS. Because of the imbalance in gender representation, the significance tests were also performed with stratification on gender. Kaplan-Meier curves were generated for graphical presentation. AIDS was defined as clinical stage WHO 4 and CDC C, $CD4^+$ T-cell counts <200 cells per mL, CD4% counts <14%, or reported death with symptoms of AIDS[48,49]. A Cox proportional hazards model was applied, adjusting for gender and age at seroconversion categorized in four age groups: <25, 25-34, 35-44, and ≥45 years. Proportional hazards assumption was controlled for all covariates.

Analysis of $CD4^+$ and $CD8^+$ T cell counts

It is well established that both absolute numbers of $CD4^+$ T cell count and CD4% are reliable immunologic markers of HIV disease progression. However, in resource-limited settings it has been suggested that CD4% may be a more suitable marker due to lower variability on repeated measurements and lower sensitivity to specimen handling, age of the patient or time of sampling (low counts in early morning)[50,51]. In addition, CD4% has been shown to have a higher prognostic value in predicting disease[51]. Accordingly, we used the percentage levels for analysis of both $CD4^+$ and $CD8^+$ T cell counts. CD4% and CD8% levels for two or more measurements were recorded for 71 and 57 HIV-1 single and 24 and 22 HIV-1 and HIV-2 dual-infected individuals. The rate of change in CD4% and CD8% over time post seroconversion was estimated using a linear regression model. The rate in CD4% decline did not differ between single and dual-infected groups (Table 1). To compare the CD4% at comparable time-points after SC, we used the global mean rate in CD4% decline (the mean of the pooled estimates of the single and dual-infected individuals) to extrapolate from each individual's mean CD4% (at the corresponding mean time-point) to the global mean time-point (56 months after SC). For the CD8% analysis there was a significant difference in rate increase between the single and dual-infected individuals. Therefore, we used each groups mean rate, instead of the global mean, to extrapolate from each individual's mean CD8% to the global mean time-point (50 months after SC).

Analysis of soluble immune activation markers

Concentrations of beta-2 microglobulin and neopterin in plasma was determined using ELISA kits from ImmunDiagnostik (Bensheim, Germany) and IBL International GmbH (Hamburg, Germany) respectively, according to the manufacturers instructions. Plasma samples from 67 (42 single and 25 dual-infected) and 66 individuals (42 single and 24 dual-infected) were available for beta-2 microglobulin and neopterin analysis, respectively. Two sample time-points were analysed for each individual (mean sample time-point from infection date was 29 and 27 months (time-point one), and 55 and 55 months (time-point two) for single and dual-

infected individuals, respectively). Differences in levels of both beta-2 microglobulin and neopterin between single and dual-infected individuals were analysed as described for the analysis of CD4% levels (using the global mean rate).

Study population of the evolutionary analysis
HIV-1 diversity, divergence, and differences in selective pressures were investigated in a subset of 20 single and 12 dual-infected individuals for which the required plasma samples were available. For each individual two blood plasma samples were selected from the asymptomatic stage. The primary selection criterion was to minimize the variation in time between the two sample time-points (mean 36.5 (S.E. 4.5) for single and 36.8 (S.E. 4.8) months for dual-infected individuals). Another criterion was to minimize the variation in time from SC for the first sample to ensure that the first sample time-point was not collected during the acute stage of infection (mean 28.0 (S.E. 3.5) for single and 38.7 (S.E. 4.9) months for dual-infected individuals).

Amplification and sequencing
Viral RNA was extracted and purified from blood plasma samples, the HIV-1 *env* V1-V3 region (nucleotides 6430 to 7374, HXB2; GenBank accession number K03455) was amplified by a nested PCR approach, and cloned, as previously described[52]. From each sample, 12 individual clones were picked, purified and sequenced using BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Stockholm, Sweden) according to the manufacturer's instructions[52].

Sequence assembly and recombination analysis
Sequences were assembled, and contigs were analysed with CodonCode Aligner version 1.5.2 (CodonCode Corporation, Dedham, USA). Only sequences with open reading frames were subjected to further analysis. It is well established that sequence recombinants violates phylogenetic inference and can lead to misinterpretations of analysed data. We used the pairwise homoplasy index (PHI) as implemented in Splitstree 4.10 to screen the 664 sequences for instances of intra-patient recombination events as descried by Salemi *et al.*[53]. Seventy two potential recombinants were found and removed, resulting in a final non-recombinant data set of 592 sequences.

Alignment and phylogenetic reconstructions
Sequences were aligned using $PRANK_{+F}$ with a neighbor joining tree, constructed in MEGA4, as guide tree[54,55]. The $PRANK_{+F}$ algorithm aligns sequences using phylogenetic information and has been shown to align sequences in an evolutionarily sound way. The alignment was manually edited and codon-stripped, resulting in a final sequence length of 609 nucleotides. A best-fitting nucleotide substitution model for the dataset (GTR+I+G) was estimated using the Akaike information criterion (AIC) as implemented in Modeltest 3.6. Maximum-likelihood (ML) phylogenies, based on 1000 bootstrap alignments, were constructed in Garli 0.951 (www.bio.utexas.edu/faculty/antisense/garli/Garli.html)[35]. This method efficiently maximizes the tree $\log_e$ likelihood by using a genetic algorithm implementing the nearest neighbor interchange (NNI) and the subtree pruning regrafting (SPR) algorithms on a random starting tree to simultaneously find and optimize the topology and branch lengths[35,56]. We also inferred Bayesian phylogenies, under the selected model (GTR+I+G), using MrBayes 3.1.2[34]. Each of two runs contained four parallel chains (one cold and three heated, temp=0.20), and each was run for 25 million generations with a burn-in of 25%. The results of the two runs were combined in Logcombiner 1.5.3. Convergence was assessed by calculating the effective sampling size (ESS) using Tracer 1.5. All parameter estimates showed ESS values higher than 200. Evolutionary analyses were

performed on 1000 ML and 15,000 Bayesian phylogenies, respectively. The ML and Bayesian estimates resulted in corresponding evolutionary differences between the single and dual-infected groups (Supplementary Table 2 and 3).

Diversity analysis

The diversity was calculated by averaging pairwise tree distances between patient-specific sequences obtained from the same sample time-point. This was done for each patient-specific time-point in all of the generated trees, yielding 1,000 and 15,000 estimates for the ML and Bayesian analyses, respectively. For each individual, the median value of these 1,000 and 15,000 diversity estimates, respectively, for each time-point, were used in the analyses. The rate of change in diversity over time, or diversity rate, was estimated as the difference in diversity between the first and the second time-point divided by the elapsed time. Diversity rate did not differ between the single and dual-infected groups (Supplementary Table 2). To compare the diversity at comparable time-points after SC, we used the global mean diversity rate (the mean of the pooled estimates of the single and dual-infected individuals) to extrapolate from each individual's mean diversity (at the corresponding mean time-point) to the global mean time-point (51 months after SC).

The "diversity threshold theory" states that AIDS develops when viral genetic diversity exceeds a critical threshold. We wanted to test if our results were consistent with this theory. First, we estimated a diversity threshold for the single-infected individuals as the extrapolated diversity at the onset of AIDS. The progression-time to AIDS was taken from the epidemiological analyses as 68 months after seroconversion. Extrapolation was done from the mean diversity for the single-infected individuals at the global mean time point using the global mean increase in diversity over time ($1.75x10^{-3}$ substitutions/site/year). Next, we estimated the progression-time to AIDS for the dual-infected group as the extrapolated mean time to reach the "diversity threshold". This extrapolation was done from the mean diversity for the dual-infected individuals using the same global increase in diversity over time. Estimates were derived for both the ML and Bayesian approach.

Divergence analysis

A major obstacle in phylogenetic analysis is often how to place the origin of related sequences. The most common approach is to use an outgroup sequence and assume that they have a common evolutionary origin[27]. By inferring large phylogenies with a high amount of sequences with strong epidemiological linkage, these sequences can serve as outgroups to each other. In this way it is possible to infer the most recent common ancestor for each patient-specific subcluster in an evolutionarily sound context[57]. By assuming a molecular clock, we were then able to estimate the evolutionary substitution rate for each individual according to the single rate dated tips approach as described[58]. Divergence rates (measured as substitutions per site per year) were estimated as follows: First, the most recent common ancestor (MRCA) of the two patient-specific time-points was determined; second, the mean number of substitutions per site between the clones of each time-point and the determined MRCA was calculated; third, the difference in mean substitutions per site from the MRCA (Δd) between time-point one (d_{tp1}) and time-point two (d_{tp2}), $\Delta d=d_{tp2}-d_{tp1}$ was calculated; and finally, Δd was divided by the difference in time between time-point one and time-point two.

Analysis of absolute rates of synonymous and nonsynonymous substitutions

To investigate differences in synonymous (dS) and nonsynonymous (dN) substitution rates, we randomly picked 200 bootstrap trees from

the ML analysis and the posterior distribution of trees from the Bayesian analysis, respectively. Each of the tree datasets were then reanalysed in HYPHY employing an MG94xGTR model, following the procedure described by Lemey *et al.*. The dN and dS were calculated as described for the divergence rates.

Statistical analysis

Statitistical analyses were performed using either Stata version 10.1 software (Statacorp LP, College Station, Texas), or PASW Statistics 18, Release 18.0.0 (Polar Engineering and Consulting).

Ethics

The study was approved by the Research Ethics Committee at the Karolinska Institute, Stockholm, and the Ministry of Health in Guinea-Bissau.

Nucleotide sequence accession numbers

The sequences used in this study were deposited into Genbank under the following accession numbers: HM745935-HM746598.

Acknowledgements

We thank Françoise Barré-Sinoussi, Mattias Höglund, Mattias Mild, and Lisa Esbjörnsson-Klemendz for critically reviewing the manuscript. We also thank Babetida N'Buna, Aquilina Sambu, Eusebio Ieme, Isabel da Costa, Jacqueline, Ana Monteiro Watche, Cidia Camara, Braima Dabo (deceased), Carla Pereira, Julieta Pinto Delgado, Leonvengilda Fernandes Mendes, Ana Monteiro, Ansu Biai, Helen Linder, Wilma Martinez-Arias, Pär-Ola Bendahl and SCIBLU Genomics for their contributions to this work. Finally, we thank the Swedish Research Council, the Swedish International Development Cooperation Agency/Department for Research Cooperation (SIDA/SAREC), the Crafoord Foundation, Lund, Sweden, the Royal Physiographic Society, Lund, Sweden, The Lars Hierta Memorial Foundation, Stockholm, Sweden, and Konsul Thure Carlsson Fund for Medical Research, Lund, Sweden for funding of this work.

References

1. Barre-Sinoussi, F., *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**, 868-871 (1983).
2. Clavel, F., *et al.* Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233**, 343-346 (1986).
3. Kanki, P.J., *et al.* Slower heterosexual spread of HIV-2 than HIV-1. *Lancet* **343**, 943-946 (1994).
4. Marlink, R., *et al.* Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* **265**, 1587-1590 (1994).
5. Simon, F., *et al.* Cellular and plasma viral load in patients infected with HIV-2. *Aids* **7**, 1411-1417 (1993).
6. Andersson, S., *et al.* Plasma viral load in HIV-1 and HIV-2 singly and dually infected individuals in Guinea-Bissau, West Africa: significantly lower plasma virus set point in HIV-2 infection than in HIV-1 infection. *Arch Intern Med* **160**, 3286-3293 (2000).
7. Hamel, D.J., *et al.* Twenty years of prospective molecular epidemiology in Senegal: changes in HIV diversity. *AIDS Res Hum Retroviruses* **23**, 1189-1196 (2007).
8. Mansson, F., *et al.* Prevalence and incidence of HIV-1 and HIV-2 before, during and after a civil war in an occupational cohort in Guinea-Bissau, West Africa. *Aids* **23**, 1575-1582 (2009).
9. Travers, K., *et al.* Natural protection against HIV-1 infection provided by HIV-2. *Science* **268**, 1612-1615 (1995).
10. Norrgren, H., *et al.* Trends and interaction of HIV-1 and HIV-2 in Guinea-Bissau, west Africa: no protection of HIV-2 against HIV-1 infection. *Aids* **13**, 701-707 (1999).
11. Aaby, P., *et al.* Does HIV-2 protect against HIV-1 infection? *Aids* **11**, 939-940 (1997).
12. Greenberg, A.E. Possible protective effect of HIV-2 against incident HIV-1 infection: review of available epidemiological and in vitro data. *Aids* **15**, 2319-2321 (2001).

13. Schim van der Loeff, M.F., *et al.* HIV-2 does not protect against HIV-1 infection in a rural community in Guinea-Bissau. *Aids* **15**, 2303-2310 (2001).

14. Wiktor, S.Z., *et al.* Lack of protection against HIV-1 infection among women with HIV-2 infection. *Aids* **13**, 695-699 (1999).

15. Al-Harthi, L., Owais, M. & Arya, S.K. Molecular inhibition of HIV type 1 by HIV type 2: effectiveness in peripheral blood mononuclear cells. *AIDS Res Hum Retroviruses* **14**, 59-64 (1998).

16. Arya, S.K. & Gallo, R.C. Human immunodeficiency virus (HIV) type 2-mediated inhibition of HIV type 1: a new approach to gene therapy of HIV-infection. *Proc Natl Acad Sci U S A* **93**, 4486-4491 (1996).

17. Kokkotou, E.G., *et al.* In vitro correlates of HIV-2-mediated HIV-1 protection. *Proc Natl Acad Sci U S A* **97**, 6797-6802 (2000).

18. de Silva, T., van Tienen, C., Rowland-Jones, S. & Cotten, M. Dual Infection with HIV-1 and HIV-2: Double Trouble or Destructive Interference? *HIV Ther.* **4**, 305-323 (2010).

19. Norrgren, H., *et al.* HIV-1, HIV-2, HTLV-I/II and Treponema pallidum infections: incidence, prevalence, and HIV-2-associated mortality in an occupational cohort in Guinea-Bissau. *J Acquir Immune Defic Syndr Hum Retrovirol* **9**, 422-428 (1995).

20. Andersson, S., da Silva, Z., Norrgren, H., Dias, F. & Biberfeld, G. Field evaluation of alternative testing strategies for diagnosis and differentiation of HIV-1 and HIV-2 infections in an HIV-1 and HIV-2-prevalent area. *Aids* **11**, 1815-1822 (1997).

21. Walther-Jallow, L., Andersson, S., da Silva, Z. & Biberfeld, G. High concordance between polymerase chain reaction and antibody testing of specimens from individuals dually infected with HIV types 1 and 2 in Guinea-Bissau, West Africa. *AIDS Res Hum Retroviruses* **15**, 957-962 (1999).

22. Wolinsky, S.M., *et al.* Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* **272**, 537-542 (1996).

23. Shankarappa, R., *et al.* Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* **73**, 10489-10502 (1999).

24. McNearney, T., *et al.* Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. *Proc Natl Acad Sci U S A* **89**, 10247-10251 (1992).

25. Nowak, M.A., *et al.* Antigenic diversity thresholds and the development of AIDS. *Science* **254**, 963-969 (1991).

26. Nowak, M.A., *et al.* HIV-1 evolution and disease progression. *Science* **274**, 1008-1011 (1996).

27. Bello, G., *et al.* Plasma viral load threshold for sustaining intrahost HIV type 1 evolution. *AIDS Res Hum Retroviruses* **23**, 1242-1250 (2007).

28. Troyer, R.M., *et al.* Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression. *J Virol* **79**, 9006-9018 (2005).

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

30. Mikhail, M., *et al.* Role of viral evolutionary rate in HIV-1 disease progression in a linked cohort. *Retrovirology* **2**, 41 (2005).

31. Lee, H.Y., Perelson, A.S., Park, S.C. & Leitner, T. Dynamic correlation between intrahost HIV-1 quasispecies evolution and disease progression. *PLoS Comput Biol* **4**, e1000240 (2008).

32. Delwart, E.L., *et al.* Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS. *J Virol* **71**, 7498-7508 (1997).

33. Markham, R.B., *et al.* Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline. *Proc Natl Acad Sci U S A* **95**, 12568-12573 (1998).

34. Ronquist, F. & Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574 (2003).

35. Zwickl, D.J. (2006).

36. Nowak, M.A. & May, R.M. Mathematical biology of HIV infections: antigenic variation and diversity threshold. *Math Biosci* **106**, 1-21 (1991).

37. Greenberg, A.E., *et al.* HIV-2 and Natural Protection Against HIV-1 Infection. *Science* **272**, 1959a (1996).

38. Otten, R.A., *et al.* Chronic HIV-2 infection

protects against total CD4+ cell depletion and rapid disease progression induced by SHIV89.6p challenge. *Aids* **18**, 1127-1135 (2004).

39. Putkonen, P., *et al.* Infection of cynomolgus monkeys with HIV-2 protects against pathogenic consequences of a subsequent simian immunodeficiency virus infection. *Aids* **4**, 783-789 (1990).

40. Akimoto, H., *et al.* Binding of HIV-2 envelope glycoprotein to CD8 molecules and related chemokine production. *Immunology* **95**, 214-218 (1998).

41. Zheng, N.N., *et al.* Comparison of human immunodeficiency virus (HIV)-specific T-cell responses in HIV-1- and HIV-2-infected individuals in Senegal. *J Virol* **78**, 13934-13942 (2004).

42. Zheng, N.N., *et al.* Role of human immunodeficiency virus (HIV)-specific T-cell immunity in control of dual HIV-1 and HIV-2 infection. *J Virol* **81**, 9061-9071 (2007).

43. Weiss, R.A., *et al.* HIV-2 antisera cross-neutralize HIV-1. *Aids* **2**, 95-100 (1988).

44. Redd, A.D., *et al.* Microbial translocation, the innate cytokine response, and HIV-1 disease progression in Africa. *Proc Natl Acad Sci U S A* **106**, 6718-6723 (2009).

45. Nkengasong, J.N., *et al.* Dual infection with human immunodeficiency virus type 1 and type 2: impact on HIV type 1 viral load and immune activation markers in HIV-seropositive female sex workers in Abidjan, Ivory Coast. *AIDS Res Hum Retroviruses* **16**, 1371-1378 (2000).

46. Critchlow, C.W., Hawes, S.E., Kuypers, J.M., Kwok, S. & Kiviat, N.B. HIV-1 and HIV-2 co-infection is associated with reduced HIV-1 plasma RNA and DNA levels among Senegalese men and women. *Am. J. Epidemiol.* **151**, S11 (2000).

47. Alabi, A.S., *et al.* Plasma viral load, CD4 cell percentage, HLA and survival of HIV-1, HIV-2, and dually infected Gambian patients. *Aids* **17**, 1513-1520 (2003).

48. Guidelines for national human immunodeficiency virus case surveillance, including monitoring for human immunodeficiency virus infection and acquired immunodeficiency syndrome. Centers for Disease Control and Prevention. *MMWR Recomm Rep* **48**, 1-27, 29-31 (1999).

49. CB., H. WHO Staging System for HIV Infection and Disease in Adolescents and Adults. in *HIV InSite Knowledge Base* (UCSF Center for HIV Information, San Francisco, 2006).

50. Fahey, J.L., *et al.* The prognostic value of cellular and serologic markers in infection with human immunodeficiency virus type 1. *N Engl J Med* **322**, 166-172 (1990).

51. Taylor, J.M., Fahey, J.L., Detels, R. & Giorgi, J.V. CD4 percentage, CD4 number, and CD4:CD8 ratio in HIV infection: which to choose and how to use. *J Acquir Immune Defic Syndr* **2**, 114-124 (1989).

52. Esbjornsson, 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* **7**, 23 (2010).

53. Salemi, M., Gray, R.R. & Goodenow, M.M. An exploratory algorithm to identify intra-host recombinant viral sequences. *Mol Phylogenet Evol* **49**, 618-628 (2008).

54. Loytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632-1635 (2008).

55. Meloni, S.T., *et al.* Distinct human immunodeficiency virus type 1 subtype A virus circulating in West Africa: sub-subtype A3. *J Virol* **78**, 12438-12445 (2004).

56. Lewis, P.O. A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data. *Mol Biol Evol* **15**, 277-283 (1998).

57. Lemey, P., Salemi, M. & Vandamme, A.M. (eds.). *The Phylogenetic Handbook*, (Cambridge University Press, 2009).

58. Rambaut, A. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* **16**, 395-399 (2000).

Author contributions

J.E. and F.M. contributed equally to this study. J.E., F.M., H.N. and P.M. analysed and interpreted the data and were responsible for the overall study design and over sight of the project. J.E. optimized the experimental

protocols, performed RNA extraction and sequencing, carried out the phylogenetic analyses, performed the CD4+ and CD8+ T cell count analysis and wrote the manuscript. F.M. clinically evaluated the patient data and performed the epidemiological analyses. J.E., A.K. and P.M. provided significant intellectual input in the phylogenetic analyses. P.E.I. contributed in statistical analyses. J.E., S.N., M.J. and P.M. performed the immunological analysis and interpreted the data. A.J.B. was medically and organisationally responsible for the clinical sites with biological samples of the study participants in the cohort. Z.J.d.S. was responsible for analyses of HIV serology at the laboratory in Guinea-Bissau. M.J. and E.M.F. participated in interpretation of the results. J.E. and P.M. finalized the manuscript. All authors read and approved the manuscript.

Supplementary information

Table S1. Soluble immune activation markers. Two sample time-points were analysed for each individual and the levels of immune activation markers and changes over time were measured as described in Methods.

	Single		Dual		
	Mean	S.E.	Mean	S.E.	p-value[1]
Beta-2 microglobulin increase over time (per year)	0.36	0.21	0.11	0.18	0.415
Beta-2 microglobulin, at the mean sample time-point after HIV-1 SC[2] (mg/L)	2.42	0.20	2.22	0.23	0.536
Neopterin decline over time (per year)	3.16	2.63	9.44	4.90	0.220
Neopterin, at the mean sample time-point after HIV-1 SC[2] (nmol/L)	19.64	3.36	20.93	3.62	0.806

[1]P-values refer to 2-tailed Student's T-test and the uncertainty of estimated means is given by the standard error (S.E.). 2SC = seroconversion.

Table S2. Maximum likelihood and Bayesian estimates of evolutionary parameters of single and dual-infected individuals. Diversity was estimated as the mean difference in HIV-1 substitutions per site at the mean sample time-point for the single and dual-infected individuals. Divergence, nonsynonymous (dN), and synonymous (dS) rates were estimated as the mean rates in HIV-1 substitutions per site per year for the single and dual-infected individuals.

	Maximum likelihood					Bayesian				
	Single		Dual			Single		Dual		
	Mean[1]	S.E.[1]	Mean[1]	S.E.[1]	p-value	Mean[1]	S.E.[1]	Mean[1]	S.E.[1]	p-value
Diversity, at the mean sample time-point after SC[2] (substitutions x 10^{-3}/site)	11.04	1.28	5.67	1.61	0.014	26.30	1.62	19.61	1.88	0.014
Diversity increase over time (substitutions x 10^{-3}/site/year)	1.60	0.92	2.00	1.53	0.812	2.11	1.24	2.27	2.01	0.943
Divergence rate (substitutions x 10^{-3}/site/year)	2.85	1.03	2.57	0.56	0.846	4.40	0.19	4.64	0.17	0.930
dN (substitutions x 10^{-3}/site/year)	2.18	0.86	1.90	0.47	0.810	2.82	0.15	4.88	0.11	0.667
dS (substitutions x 10^{-3}/site/year)	0.69	0.31	0.64	0.27	0.919	0.86	0.67	0.74	0.97	0.443

[1]P-values refer to 2-tailed Student's T-test and the uncertainty of estimated means is given by the standard error (S.E.). 2SC = seroconversion.

Table S3. Diversity threshold analysis. First, we estimated a "diversity threshold" for the single-infected individuals at the onset of AIDS. Next, we estimated the progression-time to AIDS for the dual-infected group as the extrapolated mean time to reach the "diversity threshold". Estimates were derived for both the ML and Bayesian approach. For more details see methods.

	Maximum Likelihood	Bayesian
Mean diversity at AIDS onset for single-infected individuals (substitutions x 10^{-3}/site)	13.52	29.38
Time after SC[1] reaching diversity threshold for dual-infected individuals (months)	104.93	104.95

[1]SC = seroconversion.

PREVIOUS HIV-2 INFECTION RESULTS IN BETTER SURVIVAL AND DISEASE-FREE FOLLOW-UP TIME AMONG HIV-1 SEROINCIDENT INDIVIDUALS

PAPER III

Fredrik Månsson*
Joakim Esbjörnsson*
Hans Norrgren
Patrik Medstrand

*These authors contributed equally to this study

Manuscript

III

Previous HIV-2 infection results in better survival and disease-free follow-up time among HIV-1 seroincident individuals

Fredrik Månsson[1]*†, Joakim Esbjörnsson[2]*†, Patrik Medstrand[2] and Hans Norrgren[3]

[1]Department of Clinical Sciences, Malmö, Infectious Diseases Research Unit, Lund University, Malmö, Sweden. [2]Department of Experimental Medical Science, Section of Molecular Virology, Lund University, Lund, Sweden. [3]Department of Clinical Sciences, Division of Infection Medicine, Lund University, Lund, Sweden.

We have previously shown that individuals with dual HIV-1 and HIV-2 infection have improved survival and longer disease-free duration to AIDS compared with individuals with single HIV-1 infection. In this study, our objective was to further examine differences in survival, clinical progression to AIDS and CD4% levels in dual HIV-1 and HIV-2 infection with focus on previous HIV-2 status. Analyses of 34 seroincident HIV-1 cases with contemporaneous seroprevalent or seroincident HIV-2 infection and of 189 cases of seroincident single HIV-1 infection, in a prospective open occupational cohort in Guinea-Bissau, West Africa. The median survival time and time of progression to AIDS was longer and CD4% levels were higher in HIV-2-prevalent individuals seroconverting to HIV-1 compared with seronegative individuals seroconverting to single HIV-1 infection. Observation of individuals with simultaneously recorded seroconversion to dual HIV-1 and HIV-2 infection gave intermediate results but did not differ significantly from the other groups. HIV-2 infected individuals with a documented subsequent HIV-1 seroconversion demonstrated better survival and longer disease-free time before progression to AIDS than HIV-negative individuals seroconverting to single HIV-1 infection. Deeper insight into immunological reactions elicited by HIV-2 infection may help in development of therapeutic vaccines against HIV-1.

HIV-1 and HIV-2 are two related retroviruses but with different epidemiological characteristics. HIV-1 infection is spread globally while HIV-2 is prevalent mainly in West Africa. Whereas HIV-1 infection in the majority of cases leads to AIDS, the majority of HIV-2 infected individuals become long-term nonprogressors [1, 2]. Compared with HIV-1 infection, HIV-2 infection exhibits a slower $CD4^+$ T cell count decline, lower plasma viral load and lower mortality rates [1, 3-5].

The natural course of an HIV infection can be described by three stages. The acute stage is characterized by high viremia, a rapid decline in $CD4^+$ T-cell counts which is the main target for HIV, and flu-like symptoms. In the asymptomatic stage, viral load is generally low and the $CD4^+$ T-cell decline is usually

***Corresponding authors: Fredrik Månsson, fredrik.mansson@med.lu.se; Joakim Esbjörnsson, joakim.esbjornsson@med.lu.se**
†These authors contributed equally to this study

moderate. In the AIDS stage, viral load increase, the CD4$^+$ T cell count continues to decrease and opportunistic diseases develop as a consequence of a dysfunctional immune system. The length of the different stages, in particular the asymptomatic stage, differs between individuals.

In West Africa where the two epidemics co-exist, dual HIV-1 and HIV-2 infections (HIV-D) are reported at prevalence rates around 0-3% in general population groups [6, 7] and much higher in risk groups such as female commercial sex workers and clinical patients [8, 9]. In 1995 a report came from Senegal about a possible protective effect from HIV-2 against HIV-1 infection [10] but other West African studies could not verify these results [11-15]. We have recently reported of differences between HIV-1 infection and HIV-D infection regarding mortality and median time of progression to AIDS (**paper II**). Here, we used a stratified analysis to examine the significance of earlier HIV-2 seropositivity before HIV-1 seroconversion.

Results

The participants were divided into three groups based on their respective HIV status (HIV-1, HIV-D_e and HIV-D_s). The HIV-D_e group had longer survival time and slower progression time to AIDS compared with the HIV-1 group (p=0.048 and p=0.007, respectively, Log-Rank test (L-R)) (Table 1 and Figure 1A-B). The HIV-D_s group did not differ in survival time or progression time to AIDS compared with the HIV-1 group (p=0.51 and p=0.19 L-R) or HIV-D_e (p=0.31 and p=0.34, L-R). Fitting a Cox regression model adjusting for age in two categories, the HIV-D_e group had both better survival (Hazard ratio 0.68 (95 % CI 0.47-0.97), p=0.036, Wald test), and longer time

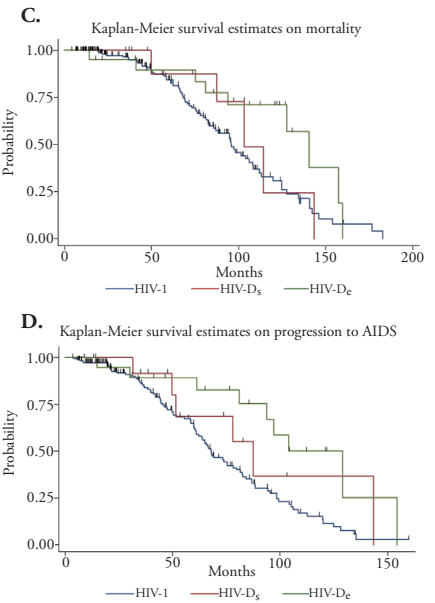


Figure 1. Analysis of survival and progression-time to AIDS. (A) Kaplan-Meier survival analysis for reported mortality due to AIDS. (B) Kaplan-Meier survival analysis for progression-time to AIDS. Cases where the exit criteria were not fulfilled were censored at their last clinical examination date in both analyses.

Table 1. Time to AIDS and mortality.

				Progression to AIDS[1]		Mortality[2]	
	N	Males	Females	Median time (months)	95% CI	Median time (months)	95% CI
HIV-1	189	159	30	68	61-80	96	83-108
HIV-D_s	14	12	2	88	50-∞	104	50-∞
HIV-D_e	20	16	4	129	81-∞	140	94-∞

[1]Comparisons between groups, Log-Rank test: HIV-D_s vs. HIV-1, p=0.19; HIV-D_e vs. HIV-1, p=0.007; HIV-D_e vs. HIV-D_s, p=0.31.
[2]Comparisons between groups, Log-Rank test: HIV-D_s vs. HIV-1, p=0.51; HIV-D_e vs. HIV-1, p=0.048; HIV-D_e vs. HIV-D_s, p=0.34.

of progression time to AIDS than the HIV-1 group (Hazard ratio 0.60, (0.42-0.85), p = 0.004, Wald test). Comparisons between the HIV-D_s and the HIV-1 group or between the HIV-D_e group and the HIV-D_s group did not show any differences in neither survival nor progression time to AIDS.

To further investigate the disease progression, we examined differences in the $CD4^+$ T cell percentage of the total amount of lymphocytes (CD4%) over time. The rate in CD4% decline between single and dual infected individuals were similar with an average decline of 1.19% per year (p=0.439, 2-tailed Student's T-test (T-test), Table 2). However, comparison of the levels of CD4% at the mean sample time point after infection (56 months) showed that the HIV-D_e group had higher CD4% levels compared to HIV-1 (p<0.001, T-test). The HIV-D_s group did not differ significantly in CD4% level compared with neither the HIV-1 group (p=0.17, T-test), nor with the HIV-D_e group (p=0.30, T-test).

Investigation of differences in CD8% over time showed that increase in CD8% levels in the HIV-D_e group was slower compared to both the single group (p=0.001, T-test) and the HIV-D_s group (p=0.006, T-test) (Table 2). Comparison between the single group and the HIV-D_s group showed similar rates in CD8% increase (p=0.94). In contrast to the analysis of CD4%, there was no difference in the CD8% levels at the mean time-point (50 months after infection) after infection between any of the groups (Table 2).

Discussion

We have followed an occupational cohort for 20 years in an HIV-2 endemic country, and recorded 226 seroincident cases of HIV-1. In this study 223 cases were included. In a previous study (**paper II**) we showed that HIV-1 seroincident cases with dual reactivity had longer progression time to AIDS as well as lower mortality compared with HIV-negative subjects that seroconverted to single HIV-1 infection.

Table 2. Estimates of CD4% and CD8% at respective mean sample time-point[1].

	HIV-1		HIV-D_s		HIV-D_e	
	Mean	S.E	Mean	S.E.	Mean	S.E
CD4% decline (% per year)[2]	1.06	0.37	1.82	1.12	1.50	0.38
CD4% level, 56 months post HIV-1 SC[3]	18.04	0.92	22.35	2.63	26.17	2.10
CD8% increase (% per year)[4]	4.62	0.80	4.81	1.32	0.91	0.62
CD8% level, 50 months post HIV-1 SC[5]	48.44	1.39	48.40	4.48	45.65	3.43

[1] To adjust for differences in sampling time, individual values were extrapolated to the global average sampling time (56 and 50 months post HIV-1 seroconversion for CD4% and CD8%, respectively) for all individuals using the average decrease in CD4% level over time (1.59.% per year) or the group-specific increase in CD8% (as indicated in the table). P-values refer to 2-tailed Student's T-test and the uncertainty of estimated means is given by the standard error (S.E.).

[2]Comparisons between groups, 2-tailed Student's T-test: HIV-D_s vs. HIV-1, p=0.54; HIV-D_e vs. HIV-1, p=0.58; HIV-D_e vs. HIV-D_s, p=0.73.

[3]Comparisons between groups, 2-tailed Student's T-test: HIV-D_s vs. HIV-1, p=0.16; HIV-D_e vs. HIV-1, p<0.001; HIV-D_e vs. HIV-D_s, p=0.31.

[4]Comparisons between groups, 2-tailed Student's T-test: HIV-D_s vs. HIV-1, p=0.94; HIV-D_e vs. HIV-1, p=0.001; HIV-D_e vs. HIV-D_s, p=0.006.

[5]Comparisons between groups, 2-tailed Student's T-test: HIV-D_s vs. HIV-1, p=0.99; HIV-D_e vs. HIV-1, p=0.39; HIV-D_e vs. HIV-D_s, p=0.65.

In this study, we stratified the group of dual infected individuals into two subgroups. We found significantly longer progression time to AIDS and lower mortality in HIV-1 infected individuals with previous HIV-2 infection (HIV-D_e) compared to HIV-1 single infected individuals. However, this difference was not found when comparing the group of individuals with estimated simultaneous HIV-1 and HIV-2 seroconversion (HIV-D_s) with individuals seroconverting to single HIV-1 or HIV-D_e infection. Thus, the HIV-D_s group had an intermediate position and was indistinguishable between single HIV-1 and HIV-D_e, suggesting a variable outcome with respect to both AIDS progression time and mortality in this group. The finding of similar CD4% decline rate but different levels of CD4% in the groups suggests that previous HIV-2 infection itself or an elicited host response influences a prolonged sustainability of CD4+ T cells at higher levels also in the event of subsequent HIV-1 infection. It is also interesting that the measured rate of change in CD8% T cells is significantly lower in the HIV-D_e group, indicating a lower degree of cellular immune activation compared with the other groups.

Several alternative explanations may account for these observations. Although the HIV-D_s individuals may have contracted the two infections at different time points during the interval between last negative and first positive sample, the amount of time the virus is present alone as a single infection in the HIV-D_s group, and before the establishment of the dual infection, is considerably shorter than for virus in the HIV-D_e group. Our result show a longer AIDS progression time and lower mortality rate in HIV-D_e, where the individual is HIV-2 single infected for a substantial period of time prior to subsequent HIV-1 infection. Thus, in the case of HIV-D_s, it is possible that not enough time have elapsed to establish the probable inhibitory effect exerted by HIV-2 on the subsequential HIV-1 infection. However, it is also possible that the HIV-D_s group represents a heterogeneous group with respect to the order and lag-time of the dual infection. Other reports that support our observations of a inhibitory effect of HIV-2 against HIV-1 disease progression are the more broadly cross-reactive antibodies noted in HIV-2 infection as compared to HIV-1 infection [22], increase in polyfunctional T-cell responses [23], as well as other reports of impaired *in vitro* replication of HIV-1 in the presence of HIV-2 [24]. It has also been reported that prolonged time of HIV-2 infection diminished the susceptibility to pathogenic SIV or recombinant SHIV infection (equivalent to HIV-1 infection) in macaques [25]. Finally, therapeutic vaccine studies in HIV-1 infected individuals have shown promising results of elicited immunological responses and temporarily reduction in viral load [26-28]. However, until date none has been able to show any positive effect on a delayed disease progression. Further studies of mechanisms by which HIV-2 inhibit HIV-1 disease progression rate may provide important tools in future design of novel therapeutic vaccines.

In conclusion, HIV-D_e individuals demonstrated a better survival and longer disease-free time before progression to AIDS than single HIV-1 infected individuals. This was not the case for the HIV-D_s individuals where no significant differences were noted in survival or in clinical progression compared to single HIV-1 or HIV-D_e infected individuals. Deeper insight in the inhibitory mechanisms elicited by HIV-2 infection on subsequent HIV-1 infection may help in development of both HIV vaccines and therapeutics.

Materials & Methods

Study population

A prospective open cohort of police officers in Guinea-Bissau, West Africa, was initiated in

1990. Until September 1, 2009, 4724 police officers had entered the study. Controls of demographic data as well as serology for HIV-1, HIV-2 and syphilis have been controlled every 12 or 18 months [7]. Participants have had access to free health services at a health post in the capital Bissau. For HIV-positive participants with CD4 counts <200, Co-trimoxazole prophylaxis was distributed. Isoniazid prophylaxis was administered to HIV-positive participants with suspected latent tuberculosis, according to WHO guidelines. A national HIV treatment programme with antiretroviral therapy (ART) was introduced in Guinea-Bissau in 2005.

In the present study, 223 participants seroincident for HIV-1 were included. Of these, 189 individuals seroconverted to single HIV-1 and 34 seroconverted to dual HIV-1 and HIV-2 infection. In addition, three participants seroconverting primarily to HIV-1 and thereafter to HIV-2 were censored from the analyses. Date of seroconversion was estimated as the mid-time-point between last HIV-1 seronegative sample and first HIV-1 seropositive sample. All individuals were treatment naïve throughout the study period. Patients starting ART were censored from the analyses from the day of ART initiation.

Of the 189 individuals seroconverting to single HIV-1, 159 were males and 30 were females. The mean age at HIV-1 seroconversion was 36.4 years (Standard deviation (SD) 9.6, range 18-60 years). Of the 34 HIV-D participants, 28 were males and six were females. In 14 individuals HIV-1 and HIV-2 seroconversion was recorded simultaneously (HIV-D_s), whereof 12 were males and two were females. The mean age at HIV-1 seroconversion was 36.2 years (SD 7.9, range 20-48 years). Twenty individuals were already seroprevalent for HIV-2 at the time of HIV-1 seroconversion (HIV-D_c), 16 were males and four were females. The mean age at HIV-1 seroconversion was 41.9 years (SD 7.4, range 29-59 years). Out of these 20 individuals, four (three males and one female) were initially HIV-2 seronegative, having first seroconverted to HIV-2 and at a later stage also to HIV-1, the mean interval between estimated HIV-2 and HIV-1 seroconversion was 4.2 years (range 1.7-6.6 months).

Laboratory methods

HIV testing was performed at the National Public Health Laboratory (LNSP), Bissau. In 1990–1994, sera were screened for HIV-1 and HIV-2 antibodies by enzyme-linked immunosorbent assay (ELISA) with the use of the Behring anti-HIV-1/HIV-2 (Behring, Marburg, Germany) and/or Wellcozyme recombinant anti-HIV-1 (Wellcome, Dartford, UK) and an in-house HIV-2 (SBL6669) ELISA assay [16]. From 1995 and onwards, screening was performed with Behring Enzygnost HIV-1/HIV-2 Plus ELISA (Behring). Confirmation of positive results was done with western blot analysis (Diagnostic Biotechnology anti-HIV-1 blot 2.2, Science park, Singapore, or in-house anti-HIV-2) and dually HIV-1/HIV-2-positive samples were confirmed by Pepti-lav (Sanofi Diagnostics Pasteur, Marnes-la-Coquette, France) in the years 1990–1998 [17]. Since 1999, an alternative confirmation strategy has been used with Capillus HIV-1/HIV-2 (Cambridge Biotech Limited, Galway, Ireland) and Immunocomb II HIV-1 and 2 BiSpot RST (Orgenics, Yavne, Israel). Evaluations of the HIV antibody testing strategy in Bissau have shown a high concordance between results obtained by serology and PCR, and to be highly distinctive between HIV-1 and HIV-2 [18, 19]. In addition, serology of sequential samples of the same infected individuals was without discrepancies, this further strengthens the results of this strategy. The screening assays have been evaluated in parallel to ensure reproducibility between assays [18].

T-lymphocyte subsets were determined at the LNSP, by conventional flow cytometry (Until 2005: FACStrak; Becton Dickinson, San Jose, Calif, with the use of three two-colour immunofluorescence reagents, CD45/CD14, CD3/CD4, and CD3/CD8, Simultest, Becton Dickinson; and from 2006 an onwards: CyFlow, Partec, Münster, Germany). Leukocyte counts were performed with a cell counter until 2005 (Coulter Counter CBC5; Coulter Electronics Ltd, Luton, England), and by Cyflow from 2006 and onwards.

Viral load (VL) measurements have not been included as a standard procedure in Guinea-Bissau. Unfortunately, due to limitations in sample size, storing conditions and sample availability, we are not able to do any retrospective and valid sequential VL measurements of the samples included in this study.

Analysis of survival and progression time to AIDS

Survival analysis was performed for reported mortality and for progression time to AIDS. Kaplan-Meier curves were generated for graphical presentation. The clinical stages WHO 4 and CDC C were considered as AIDS-defining, as well as CD4+ T cell counts below 200 cells per mL, CD4% counts below 14%, and reported death with symptoms of AIDS. Where applicable, an additional Cox proportional hazards model was applied, adjusting for age in two age groups (≤34 and ≥35years) but not for gender since all groups did not contain females experiencing events.

Analysis of CD4+ and CD8+ T cell counts

Differences in CD4+ and CD8+ T cell counts were analyzed by comparing the groups of HIV-1, HIV-D_e and HIV-D_s. CD4% and CD8% counts rather than absolute counts were used for the analysis, since it has been reported to be less fluctuating in flowcytometrical measurements [20, 21]. We estimated the rate of change in CD4% and CD8% over time post seroconversion using a linear regression model. The rate in CD4% decline did not differ between single and dual-infected groups (Table 2). Therefore, we used the global mean CD4% decline rate to compare the CD4% levels at comparable time-points after seroconversion, (the mean of the pooled estimates of the single and dual infected individuals) to extrapolate from each individual's mean CD4% level (at the corresponding mean time-point) to the global mean sample time-point (56 months post seroconversion). For the CD8% analysis there was significant differences in rate increase between (1) the HIV-1 and HIV-D_e groups and (2) the HIV-D_e and HIV-D_s groups. Therefore, we used each groups mean rate, instead of the global mean, to extrapolate from each individual's mean CD8% to the global mean time-point (50 months after SC).

Statistics

The epidemiological statistics were calculated using Stata version 10.1 software (Statacorp LP, College Station, Texas), and the CD4% statistics using PASW Statistics 18, Release 18.0.0 (Polar Engineering and Consulting).

Ethics

The study was approved by the Research Ethics Committee at the Karolinska Institute, Stockholm, and the Ministry of Health in Guinea-Bissau.

Acknowledgements

We thank Babetida N'Buna, Aquilina Sambu, Eusebio Ieme, Isabel da Costa, Jacqueline, Ana Monteiro Watche, Cidia Camara, Braima Dabo (deceased), Carla Pereira, Julieta Pinto Delgado, Leonvengilda Fernandes Mendes, Ana Monteiro, Ansu Biai, and Helen Linder, Finally, we thank the Swedish Research Council, the Swedish International Development Cooperation Agency/Department for Research Cooperation (SIDA/SAREC),

the Crafoord Foundation, Lund, Sweden, the Royal Physiographic Society, Lund, Sweden, The Lars Hierta Memorial Foundation, Stockholm, Sweden, and Konsul Thure Carlsson Fund for Medical Research, Lund, Sweden for funding of this work.

References

1. Marlink R, Kanki P, Thior I, Travers K, Eisen G, Siby T, Traore I, Hsieh CC, Dia MC, Gueye EH, et al.: **Reduced rate of disease development after HIV-2 infection as compared to HIV-1.** *Science* 1994, **265:**1587-1590.

2. Jaffar S, Grant AD, Whitworth J, Smith PG, Whittle H: **The natural history of HIV-1 and HIV-2 infections in adults in Africa: a literature review.** *Bull World Health Organ* 2004, **82:**462-469.

3. Kanki PJ, Travers KU, S MB, Hsieh CC, Marlink RG, Gueye NA, Siby T, Thior I, Hernandez-Avila M, Sankale JL, et al.: **Slower heterosexual spread of HIV-2 than HIV-1.** *Lancet* 1994, **343:**943-946.

4. Simon F, Matheron S, Tamalet C, Loussert-Ajaka I, Bartczak S, Pepin JM, Dhiver C, Gamba E, Elbim C, Gastaut JA, et al.: **Cellular and plasma viral load in patients infected with HIV-2.** *Aids* 1993, **7:**1411-1417.

5. Andersson S, Norrgren H, da Silva Z, Biague A, Bamba S, Kwok S, Christopherson C, Biberfeld G, Albert J: **Plasma viral load in HIV-1 and HIV-2 singly and dually infected individuals in Guinea-Bissau, West Africa: significantly lower plasma virus set point in HIV-2 infection than in HIV-1 infection.** *Arch Intern Med* 2000, **160:**3286-3293.

6. Hamel DJ, Sankale JL, Eisen G, Meloni ST, Mullins C, Gueye-Ndiaye A, Mboup S, Kanki PJ: **Twenty years of prospective molecular epidemiology in Senegal: changes in HIV diversity.** *AIDS Res Hum Retroviruses* 2007, **23:**1189-1196.

7. Mansson F, Biague A, da Silva ZJ, Dias F, Nilsson LA, Andersson S, Fenyo EM, Norrgren H: **Prevalence and incidence of HIV-1 and HIV-2 before, during and after a civil war in an occupational cohort in Guinea-Bissau, West Africa.** *Aids* 2009, **23:**1575-1582.

8. Djomand G, Greenberg AE, Sassan-Morokro M, Tossou O, Diallo MO, Ekpini E, Ghys P, Soro B, Brattegaard K, Yapi A, et al.: **The epidemic of HIV/AIDS in Abidjan, Cote d'Ivoire: a review of data collected by Projet RETRO-CI from 1987 to 1993.** *J Acquir Immune Defic Syndr Hum Retrovirol* 1995, **10:**358-365.

9. Ghys PD, Diallo MO, Ettiegne-Traore V, Kale K, Tawil O, Carael M, Traore M, Mah-Bi G, De Cock KM, Wiktor SZ, et al: **Increase in condom use and decline in HIV and sexually transmitted diseases among female sex workers in Abidjan, Cote d'Ivoire, 1991-1998.** *AIDS* 2002, **16:**251-258.

10. Travers K, Mboup S, Marlink R, Gueye-Nidaye A, Siby T, Thior I, Traore I, Dieng-Sarr A, Sankale JL, Mullins C, et al.: **Natural protection against HIV-1 infection provided by HIV-2.** *Science* 1995, **268:**1612-1615.

11. Norrgren H, Andersson S, Biague AJ, da Silva ZJ, Dias F, Naucler A, Biberfeld G: **Trends and interaction of HIV-1 and HIV-2 in Guinea-Bissau, west Africa: no protection of HIV-2 against HIV-1 infection.** *Aids* 1999, **13:**701-707.

12. Aaby P, Poulsen AG, Larsen O, Christiansen CB, Jensen H, Melbye M, Naucler A, Dias F: **Does HIV-2 protect against HIV-1 infection?** *Aids* 1997, **11:**939-940.

13. Greenberg AE: **Possible protective effect of HIV-2 against incident HIV-1 infection: review of available epidemiological and in vitro data.** *Aids* 2001, **15:**2319-2321.

14. Schim van der Loeff MF, Aaby P, Aryioshi K, Vincent T, Awasana AA, Da Costa C, Pembrey L, Dias F, Harding E, Weiss HA, Whittle HC: **HIV-2 does not protect against HIV-1 infection in a rural community in Guinea-Bissau.** *Aids* 2001, **15:**2303-2310.

15. Wiktor SZ, Nkengasong JN, Ekpini ER, Adjorlolo-Johnson GT, Ghys PD, Brattegaard K, Tossou O, Dondero TJ, De Cock KM, Greenberg AE: **Lack of protection against HIV-1 infection among women with HIV-2 infection.** *Aids* 1999, **13:**695-699.

16. Biberfeld G, Thorstensson R, Bergstrom M, Naucler A, Costa CM: **Enzyme immunoassays for the demonstration of antibodies to HIV-2SBL-6669 and HTLV-IV (SIVmac).** *Aids* 1988, **2:**195-199.

17. Albert J, Bredberg U, Chiodi F, Bottiger B, Fenyo EM, Norrby E, Biberfeld G: **A new human retrovirus isolate of West African origin (SBL-6669) and its relationship to HTLV-IV, LAV-II, and HTLV-IIIB.** *AIDS Res Hum Retroviruses* 1987, **3:**3-10.

18. Andersson S, da Silva Z, Norrgren H, Dias F, Biberfeld G: **Field evaluation of alternative testing strategies for diagnosis and differentiation of HIV-1 and HIV-2 infections in an HIV-1 and HIV-2-prevalent area.** *Aids* 1997, **11:**1815-1822.

19. Walther-Jallow L, Andersson S, da Silva Z, Biberfeld G: **High concordance between polymerase chain reaction and antibody testing of specimens from individuals dually infected with HIV types 1 and 2 in Guinea-Bissau, West Africa.** *AIDS Res Hum Retroviruses* 1999, **15:**957-962.

20. Taylor JM, Fahey JL, Detels R, Giorgi JV: **CD4 percentage, CD4 number, and CD4:CD8 ratio in HIV infection: which to choose and how to use.** *J Acquir Immune Defic Syndr* 1989, **2:**114-124.

21. Burcham J, Marmor M, Dubin N, Tindall B, Cooper DA, Berry G, Penny R: **CD4% is the best predictor of development of AIDS in a cohort of HIV-infected homosexual men.** *AIDS* 1991, **5:**365-372.

22. Rowland-Jones SL, Whittle HC: **Out of Africa: what can we learn from HIV-2 about protective immunity to HIV-1?** *Nat Immunol* 2007, **8:**329-331.

23. de Silva T, van Tienen C, Rowland-Jones S, Cotten M: **Dual Infection with HIV-1 and HIV-2: Double Trouble or Destructive Interference?** *HIV therapy* 2010, **4:**305-323.

24. Arya SK, Gallo RC: **Human immunodeficiency virus (HIV) type 2-mediated inhibition of HIV type 1: a new approach to gene therapy of HIV-infection.** *Proc Natl Acad Sci U S A* 1996, **93:**4486-4491.

25. Otten RA, Ellenberger DL, Adams DR, Fridlund CA, Jackson E, Pieniazek D, Rayfield MA: **Identification of a window period for susceptibility to dual infection with two distinct human immunodeficiency virus type 2 isolates in a Macaca nemestrina (pig-tailed macaque) model.** *J Infect Dis* 1999, **180:**673-684.

26. Autran B, Murphy RL, Costagliola D, Tubiana R, Clotet B, Gatell J, Staszewski S, Wincker N, Assoumou L, El-Habib R, et al: **Greater viral rebound and reduced time to resume antiretroviral therapy after therapeutic immunization with the ALVAC-HIV vaccine (vCP1452).** *AIDS* 2008, **22:**1313-1322.

27. MacGregor RR, Boyer JD, Ugen KE, Tebas P, Higgins TJ, Baine Y, Ciccarelli RB, Ginsberg RS, Weiner DB: **Plasmid vaccination of stable HIV-positive subjects on antiviral treatment results in enhanced CD8 T-cell immunity and increased control of viral "blips".** *Vaccine* 2005, **23:**2066-2073.

28. Rosenberg ES, Graham BS, Chan ES, Bosch RJ, Stocker V, Maenza J, Markowitz M, Little S, Sax PE, Collier AC, et al: **Safety and immunogenicity of therapeutic DNA vaccination in individuals treated with antiretroviral therapy during acute/early HIV-1 infection.** *PLoS One*, **5:**e10555.

FREQUENT CXCR4 TROPISM OF HIV-1 SUBTYPE A AND CRF02_AG DURING LATE-STAGE DISEASE - INDICATION OF AN EVOLVING EPIDEMIC IN WEST AFRICA

PAPER IV

Joakim Esbjörnsson
Fredrik Månsson
Wilma Martínez-Arias
Elzbieta Vincic
Antonio J. Biague
Zacarias J. da Silva
Eva Maria Fenyö
Hans Norrgren
Patrik Medstrand

Retrovirology,
2010, 22;7:23

IV

Esbjörnsson et al. Retrovirology 2010, 7:23
http://www.retrovirology.com/content/7/1/23

RETROVIROLOGY

RESEARCH Open Access

Frequent CXCR4 tropism of HIV-1 subtype A and CRF02_AG during late-stage disease - indication of an evolving epidemic in West Africa

Joakim Esbjörnsson[1*], Fredrik Månsson[2], Wilma Martínez-Arias[1], Elzbieta Vincic[3], Antonio J Biague[5], Zacarias J da Silva[5], Eva Maria Fenyö[3], Hans Norrgren[4], Patrik Medstrand[1]

Abstract

Background: HIV-1 is one of the fastest evolving pathogens, and is distinguished by geographic and genetic variants that have been classified into different subtypes and circulating recombinant forms (CRFs). Early in infection the primary coreceptor is CCR5, but during disease course CXCR4-using HIV-1 populations may emerge. This has been correlated with accelerated disease progression in HIV-1 subtype B. Basic knowledge of HIV-1 coreceptor tropism is important due to the recent introduction of coreceptor antagonists in antiretroviral therapy, and subtype-specific differences regarding how frequently HIV-1 CXCR4-using populations appear in late-stage disease need to be further investigated. To study how frequently CXCR4-using populations appear in late-stage disease among HIV-1 subtype A and CRF02_AG, we evaluated the accuracy of a recombinant virus phenotypic assay for these subtypes, and used it to determine the HIV-1 coreceptor tropism of plasma samples collected during late-stage disease in Guinea-Bissau. We also performed a genotypic analysis and investigated subtype-specific differences in the appearance of CXCR4 tropism late in disease.

Results: We found that the recombinant virus phenotypic assay accurately predicted HIV-1 coreceptor tropism of subtype A and CRF02_AG. Over the study period (1997-2007), we found an increasing and generally high frequency of CXCR4 tropism (86%) in CRF02_AG. By sequence analysis of the V3 region of our samples we developed a novel genotypic rule for predicting CXCR4 tropism in CRF02_AG, based on the combined criteria of the total number of charged amino acids and net charge. This rule had higher sensitivity than previously described genotypic rules and may be useful for development of future genotypic tools for this CRF. Finally, we conducted a literature analysis, combining data of 498 individuals in late-stage disease, and found high amounts of CXCR4 tropism for all major HIV-1 subtypes (60-77%), except for subtype C (15%).

Conclusions: The increase in CXCR4 tropism over time suggests an evolving epidemic of CRF02_AG. The results of the literature analysis demonstrate the need for further studies investigating subtype-specific emergence for CXCR4-tropism; this may be particularly important due to the introduction of CCR5-antagonists in HIV treatment regimens.

Background

Human immunodeficiency virus type 1 (HIV-1) evolves at an extremely high rate, primarily due to a combination of high viral turn-over, an error prone viral reverse transcriptase and frequent recombination. This high level of molecular evolution has led to diversification of HIV-1 into genetically distinct subtypes (A-D, F-H, J-K), subsubtypes (A1-A3, F1-F2) and circulating recombinants forms (CRFs), usually defined by geographical location [1]. The most common subtypes are subtype A (12.3% of the global prevalence), B (10.2%), C (49.9%) and G (6.3%), and the CRF01_AE (4.7%) and CRF02_AG (4.8%) [1].

HIV-1 enters target cells via interactions with CD4 and a coreceptor, usually one of the chemokine receptors CCR5 or CXCR4. Different HIV strains have been

* Correspondence: joakim.esbjornsson@med.lu.se
[1]Department of Experimental Medical Science, Section of Molecular Virology, Lund University, Lund, Sweden

classified based on coreceptor tropism: CCR5-tropic strains are referred to as R5, CXCR4-tropic strains as X4, and dual tropic strains as R5X4 [2]. Coreceptor use has been studied extensively in HIV-1 subtype B and C, but needs further investigation for other subtypes [3-13]. In subtype B, R5 populations are generally present over the entire course of infection whereas R5X4 or X4 populations emerge late in infection. This coreceptor switch has been associated with faster CD4+ T cell decline and the development of AIDS, although studies describing the opposite, or no difference in CD4+ T cell decline have also been observed [5,6].

Little is known about subtype-specific differences regarding how frequently CXCR4-using populations appear in late-stage disease. Most studies investigating HIV-1 subtype B coreceptor tropism have focused on either the relation between the detection of X4 viruses and disease progression rate, or molecular properties that differ between the R5 and X4 viruses [5,6,14,15]. HIV-1 CXCR4-using populations are thought to appear in approximately 50% of the patients infected with subtype B [16-18]. The fraction of subtype C-infected individuals that have CXCR4-using populations appear to be less frequent (0-30%) [8,9,11,19]. Moreover, a study of HIV-1 CRF01_AE in 22 AIDS patients showed that 16 subjects (73%) had X4 populations [12]. Comparing the A and D subtypes, Kaleebu et al. found no significant difference in patients with low CD4 counts (≤ 200) [3]. When a comparison was done at an earlier stage of HIV-1 infection (CD4 counts > 200), CXCR4 use was more frequent among patients with subtype D infection, probably due to an earlier coreceptor switch than in patients infected with subtype A [3]. Other cross-sectional studies do not allow estimation of the emergence of HIV-1 with X4 phenotype since CD4 counts or clinical statuses were not considered together [20-22].

The viral envelope glycoprotein (gp) 120 is organized in five hypervariable regions (V1-V5), interspersed within five conserved regions (C1-C5). The major viral determinants of the interaction between gp120 and the coreceptors CCR5 or CXCR4 are located in the V3 region, even though other regions, such as the V1/V2 and the C4 regions have been shown to influence coreceptor use [23,24]. To date, most studies have focused on HIV-1 subtype B and C, and there is no clear evidence that the V3 region has the same impact on coreceptor interaction among other subtypes.

Basic knowledge of HIV-1 coreceptor evolution has become increasingly important due to the recent introduction of CCR5 antagonists as part of antiretroviral therapy against HIV-1 [25,26]. Since these drugs have no effect on X4 populations, HIV-1 coreceptor tropism must be identified before the initiation of treatment [27,28]. The gold standard for clinical samples is coreceptor determination by recombinant phenotypic entry assays [29,30]. Reliable bioinformatic tools based on viral genotype may be a faster and more cost-effective way to predict coreceptor tropism. Present genotypic predictors are based on V3 sequences from subtype B or C, and have been shown to perform poorly on other subtypes, especially in detecting CXCR4-using variants [31,32]. This indicates that molecular differences connected to coreceptor use may be subtype-specific, and that specific predictors likely have to be constructed for each major subtype and CRF.

In view of this, we set out to determine the frequency of emergence of X4 phenotype in HIV-1 subtype A or CRF02_AG infected individuals in late-stage disease by evaluating the performance of a recombinant virus phenotypic assay for subtype A and CRF02_AG. Using this tool, we found an increasing and generally high frequency of CXCR4 tropism (86%) in CRF02_AG, and developed a novel genotypic rule for predicting CXCR4 tropism, based on combined criteria of the total number of charged amino acids and net charge of the V3 region. Finally, we compared our results to other HIV-1 subtypes by analyzing HIV-1 coreceptor phenotype in individuals in late-stage disease and found high amounts of CXCR4 tropism for all major HIV-1 subtypes (60-77%), except for subtype C (15%).

Methods

Sample sets

Two sample sets were used in the present study. The first sample set consisted of a control panel of 11 HIV-1 isolates with predetermined subtype and coreceptor tropism. Subtype was determined by sequencing of the *env* V3 region, and coreceptor tropism was determined using a phenotypic infection assay with either the U87.CD4-CCR5/U87.CD4-CXCR4 cell system or the MT-2 cell system [22]. Data were generously provided by Professor Jan Albert, Swedish Institute for Infectious Disease Control, Stockholm, Sweden. All isolates were amplified, sequenced, and used for evaluation of the recombinant virus phenotypic assay. Three isolates (22480, 22627, and 30405) were also used in the genotypic analysis of HIV-1 CRF02_AG. Details of the control panel can be found in Table 1. The second sample set consisted of 33 plasma samples from 33 HIV-1 infected individuals and was selected from a cohort of police officers from Guinea-Bissau, West Africa, based on sample availability and disease status. The cohort has been described in detail elsewhere [33-35]. Twenty-nine of the samples were successfully amplified and subjected to further analyses. All of the individuals were treatment naïve and classified to be in late-stage disease, as defined by CD4+ T cell count (≤ 200 cells/μl or ≤ 14%) or clinical AIDS (CDC: C or WHO: 4) [36,37]. In cases where more than one sample from

Table 1 Evaluation of the TRT assay using a control panel of HIV-1 subtype A and CRF02_AG isolates.

Isolate No.	Coreceptor tropism		Subtype
	Isolate	Chimeric virus[1]	
7535	R5	R5	A
9488	R5	R5	A
22480	R5	R5	CRF02_AG
22627	R5	R5	CRF02_AG
36412	R5	R5	A
36748	R5	R5	A
8131	R5X4	R5X4	A
30405	R5X4	R5X4	CRF02_AG
9284	R5X4	R5X4	A
11974	X4	R5X4	A
13636	R5X4	R5X4	A

[1]Chimeric viruses were constructed using amplified HIV-1 gp120 V1-V3 fragments specific for each isolate according to the protocol of the TRT assay.

late-stage disease was available, the last sample was chosen. Individuals diagnosed with tuberculosis and clinically categorized as CDC: C, but without other AIDS-defining symptoms, were not included in the study. For patient samples DL2713H, DL2846I and DL3018H, there were no recorded CD4+ T cell counts. These samples were included in the study based on previous observations of CD4+ T cell counts of the same patient, according to the described criterions. Details of the plasma samples from Guinea-Bissau can be found in Table 2.

Amplification and sequencing

Viral RNA was extracted and purified from blood plasma samples, using RNeasy Lipid Tissue Mini Kit (Qiagen, Stockholm, Sweden) with minor modifications from the manufacturer's instructions. Briefly, 200 μl of blood plasma were disrupted in 2000 μl Qiazol and 10 μg Carrier RNA (Qiagen). The aqueous phase was loaded onto a spin column by multiple loading steps. RNA was eluted in 40 μl of RNase-free water and treated with DNase I (Fermentas, Helsingborg, Sweden). Viral RNA was reverse transcribed using gene-specific primers, and the V1-V3 region amplified using a nested PCR approach (The SuperScript™ III One-Step RT-PCR System with Platinum® *Taq* DNA Polymerase and Platinum® *Taq* DNA Polymerase High Fidelity, Invitrogen, Copenhagen, Denmark) according to the manufacturer's instructions using primers JE12F (5'-AAAGAGCAGAAGATAGTGGCAATGA-3') and V3A_R2 (5'-TTACAATAGAAAAATTCTCCTCYACA-3') for one-step RT-PCR and E20A_F (5'-GGGCTACACATGCCTGTGTACCYACAG-3') and JA169 for nested PCR [38]. The V3 region with flanking regions (nucleotides 6847 to 7374 in HXB2; GenBank accession number K03455) were then directly sequenced using BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Stockholm, Sweden) according to the manufacturer's instructions using primers JA167 and JA169 [38]. The V3 region was chosen for sequencing due to prevalence of length variations in the V1-V2 region. Sequences were determined using ABI Prism 3100 (Applied Biosystems). Sample DL2846I had to be cloned as a result of too mixed chromatograms. The amplified V1-V3 region of approximately 940 base pairs (nucleotides 6430 to 7374 in HXB2; GenBank accession number K03455) were cloned using the InsTAclone cloning system (Fermentas) and TOP10 cells (Invitrogen). Twelve colonies were picked and the cloned fragments were amplified with Platinum® *Taq* DNA Polymerase High Fidelity (Invitrogen) using conventional M13 primers (-20 and -24). The amplifications were successful for eight colonies, and clones were named with the patient identification number and a clone number. Individual clones were purified and sequenced as described.

Phylogenetic analysis

Sequences were assembled, and contigs were analyzed with CodonCode Aligner version 1.5.2 (CodonCode Corporation, Dedham, USA) blinded to the phenotype. True permutated positions were detected by the software, and manually inspected. All sequences had open reading frames and were subjected for further analysis. A multiple alignment of our sequences with a reference sequence data set of all major subtypes, sub-subtypes and CRFs (downloaded from Los Alamos Sequence Database) was performed in MEGA4 using the Clustal algorithm [39-41]. Nucleotide sequences were aligned via protein sequences and major gap positions were removed to a final sequence length of 463 base pairs. A neighbor-joining (N-J) tree was constructed in MEGA4 using pair-wise deletion in a maximum composite likelihood substitution model with heterogeneous pattern among lineages with a gamma distribution of 0.8338 (Akaike Information Criterion (AIC), calculated by Modeltest [42]). The phylogenetic reconstruction was bootstrapped 1000 times to separate sequences of different subtypes. The gp120 V3 region of CRF02_AG is subtype A-derived, and separation of subclusters belonging to either CRF02_AG or subtype A were not possible in this tree. To further characterize these sequences, we constructed a reference data set consisting of sequences characterized as subtype A or CRF02_AG. To avoid any bias of previously incorrect subtyping due to similarities in the V3 region between subtype A and CRF02_AG, only HIV-1 full genome (> 8000 bp) sequences were allowed. We reconstructed an N-J tree with our Guinea-Bissau-derived sequences and the described reference sequences as outlined above. Reference sequences that formed separate monophyletic clusters were removed

Table 2 Clinical parameters, HIV-1 subtype and HIV-1 tropism of the 29 analyzed study subjects.

Patient No.	Sex[1]	CD4%[2]	CD4tot[3]	CDC[4]	WHO[5]	Subtype	Tropism	Sample year
DL1996H	M	5	157	B	3	CRF02_AG	R5X4	2000
DL2089J	M	9	59	B	3	CRF02_AG	R5X4	2003
DL2096F	M	5	22	C	4	C	N/A[6]	2003
DL2249I	M	2	21	C	4	CRF02_AG	R5X4	2004
DL2339E	M	12	178	B	3	CRF02_AG	R5X4	2003
DL2365K	M	9	133	B	3	A3	R5	2006
DL2391G	M	5	N/A[6]	B	2	CRF02_AG	R5	2000
DL2401M	M	11	141	B	3	CRF02_AG	R5X4	2004
DL2713H	M	N/A[6]	N/A[6]	B	2	CRF02_AG	R5X4	2007
DL2846I	F	N/A[6]	N/A[6]	B	3	A3	R5X4	2005
DL2853E	M	11	137	A	1	CRF02_AG	R5	1998
DL2920H	M	11	126	B	3	CRF02_AG	X4	2004
DL3018H	M	N/A[6]	N/A[6]	B	3	A3	R5X4	2006
DL3037E	M	3	74	B	3	CRF02_AG	R5X4	2005
DL3039G	F	7	148	A	2	CRF02_AG	R5X4	2006
DL3071H	F	20	123	B	3	A3	R5X4	2005
DL3087E	M	4	62	B	2	CRF02_AG	R5X4	2001
DL3098I	F	14	426	N/A[6]	N/A[6]	CRF02_AG	R5X4	2007
DL3169F	M	9	315	B	3	CRF02_AG	R5X4	2004
DL3170F	M	8	65	B	2	CRF02_AG	R5X4	2000
DL3234J	M	10	216	A	2	CRF02_AG	R5X4	2006
DL3312E	M	2	36	C	4	CRF02_AG	R5X4	1998
DL3633G	F	8	112	C	4	CRF02_AG	R5X4	2003
DL3721C	M	11	257	A	1	CRF02_AG	R5X4	1997
DL3733G	M	19	137	B	3	CRF02_AG	R5X4	2004
DL4248G	F	13	159	B	3	A3	R5	2005
DL4477D	M	14	141	B	3	CRF02_AG	R5	2001
DL4525G	M	13	372	B	3	A3	R5	2006
DL4632E	F	9	77	B	3	CRF02_AG	R5X4	2003

[1]M = male, F = female
[2]CD4+ T cell percentage among all T cells
[3]CD4+ T cell count per microliter among all T cells
[4]Clinical category of the patient, as defined by the CDC, at the sample time point
[5]Clinical category of the patient, as defined by the WHO, at the sample time point
[6]N/A = not analyzed
[7]Sample included in the study based on previous examinations of CD4+ T cell count and percentage

IV

from the data set to obtain a final reconstruction distinguishing clusters of subsubtypes (A1-A3) and CRF02_AG (AIC, gamma distribution 0.9136) (Fig. 1). Since Felsenstein's bootstrap test can be too conservative, we used the bootstrap interior branch test, with 1000 bootstraps, which is a mathematically more rigorous statistical method for phylogenetic reconstructions of closely related sequences [43-45]. Details and accession numbers of the constructed reference sequence data set can be found in Additional file 1.

Determination of coreceptor tropism and evaluation of phenotypic method

Human kidney embryonic 293T cells and human glioma U87.CD4 cells, stably expressing CD4 and one of the chemokine receptors (CCR5 or CXCR4) were maintained as previously described [46,47]. Chimeric viruses with patient-specific V1-V3 regions were generated based on the protocol from the Tropism Recombinant Test (TRT) with minor modifications [15,29]. Briefly, 500 ng of amplified V1-V3 fragments from each plasma sample and 3 μg of 43XCΔV, a NheI-linearized vector containing a full-length pNL4-3 genome with the V1-V3 region deleted, were transfected into 293T cells using the calcium phosphate precipitation method. Chimeric viruses were harvested and stored at -80°C. Twenty-four hours before infection, 10^5 U87.CD4 cells/well were seeded in 48 well plates. For infection, 500 μl of chimeric viruses were added in duplicate wells. Cells were washed three times with Dulbecco Modified Eagle

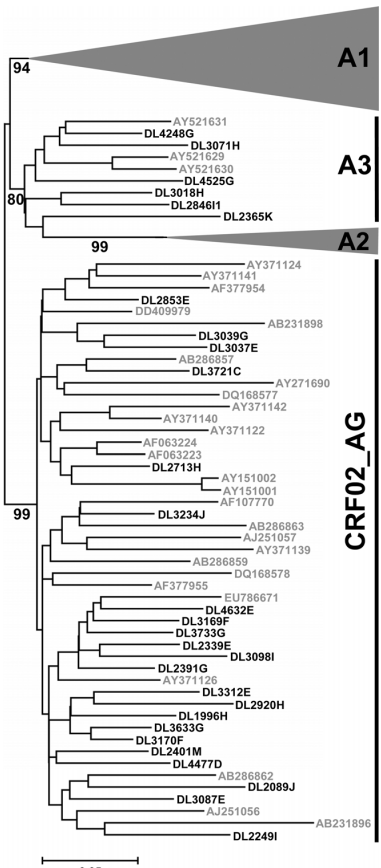


Figure 1 Classification of subtype A-like sequences. A neighbour joining tree showing detailed analysis of subtype A-like sequences. Interior branch test likelihoods (1000 bootstraps) are shown on branches distinguishing sub-subtypes and CRF02_AG. All sample sequences cluster within either the CRF02_AG or the A3 cluster. For visual clarity, subsubtype A1 and A2 clusters are represented by triangles.

Medium 16 hours post infection. Cultures were analyzed at day one, seven and nine for p24 antigen production by ELISA (Biomérieux, Boxtel, The Netherlands).

To evaluate the performance of the TRT on subtype A and CRF02_AG (which is A-like in the V1-V3 region of gp120), we used a panel of selected isolates (generously provided by Professor Jan Albert, Swedish Institute for Infectious Disease Control) (Table 1). Subtype and phenotype of the isolates had been determined previously [21].

To confirm the phenotype results, we used 200 μl of each isolate for infection of U87.CD4-CCR5 and U87.CD4-CXCR4 in duplicate wells. The cells were washed 16 hours post infection and then analyzed for p24 production as described. From each isolate the V1-V3 region was amplified and used to produce recombinant viruses for infection and analysis according to the described procedure. Infections of U87.CD4-CXCR4 with chimeric viruses derived from R5 isolates were used as negative controls to calculate the p24 background value of the assay. Infections with a significant increase (mean value +/- 3 S.D.) in p24 antigen production over time, compared to day one of the infection, were considered as positive infections.

Genotypic characterization of CRF02_AG coreceptor use

The aim was to characterize genetic properties in HIV-1 V3 *env* distinguishing CRF02_AG with different coreceptor tropism. Only two sequences of the control panel (22480 and 22627) and three sequences of the plasma samples (DL4477D, DL2391G and DL22853E) were of HIV-1 CRF02_AG with pure CCR5 tropism. Since this number was too low for an appropriate comparison between the R5 and R5/X4 groups, we added all available V3 sequences, with known coreceptor tropism, from Los Alamos sequence data base, to a final dataset of 111 sequences (Additional files 2, 3, 4) [39]. Only one sequence per patient was subjected for analyses. Multiple alignments of CRF02_AG V3 amino acid sequences were performed as described, and positively [K and R] and negatively [D and E] charged amino acids were counted. In codons with amino acid mixtures, all possible permutations were assessed (Additional files 3 and 4). The combination resulting in the highest net charge was used for phenotype prediction. The performance of different sequence motif-based rules was measured in terms of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). The sensitivity was determined as the fraction of predicted X4 sequences among the sequences from viruses phenotyped as CXCR4-using, the specificity as the fraction of predicted R5 sequences among the sequences from viruses phenotyped as CCR5-using only, the PPV as the fraction of correctly predicted X4 sequences among all predicted X4 sequences, and the NPV as the fraction of correctly predicted R5 sequences among all predicted R5 sequences.

Statistics

All statistical analysis was performed using SPSS 16.0.

Ethics

The study was approved by the Research Ethics Committee at the Karolinska Institute, Stockholm, and the Ministries of Health and the Interior in Guinea-Bissau.

Nucleotide sequence accession numbers

Nucleotide sequences were deposited in GenBank under the following accession numbers: GQ401717-GQ401744, and FJ831886-FJ831893.

Results

Subtype determination

The HIV-1 V3 region from 29 plasma samples, collected during late-stage disease from 29 treatment-naïve individuals, was amplified and sequenced. A phylogenetic tree with these sequences and reference sequences of different subtypes was reconstructed. Sequences were well separated with long branches in the phylogeny, indicating patient-specific origin of sequences (Fig. 1). Twenty-eight sequences formed a subtype A cluster together with reference sequences of A1, A2 and CRF02_AG. The non-subtype A sequence clustered with reference strains of subtype C. The V3 region of CRF02_AG is subtype A-derived, and to further characterize these sequences we made a BLAST-search using each sample sequence as query sequence. The three most similar full genome (> 8000 bp) hits were used as reference sequences to distinguish subtype A from CRF02_AG. Among the 28 patients with subtype A-like sequences, 22 clustered with CRF02_AG reference sequences, and six with the previously described subsubtype A3 (Fig. 1, Table 2) [48].

Accuracy of the phenotypic method for subtype A and CRF02_AG

Construction of chimeric viruses is a commonly used method for coreceptor tropism analysis, and both commercial and in-house variants can be found in the literature [29,30,49]. Here, we used for subtype B the well-established Tropism Recombinant Test (TRT) [29,50]. To confirm that the TRT assay performs equally well for HIV-1 subtype A and CRF02_AG as for HIV-1 subtype B, we used a control panel of HIV-1 subtype A and CRF02_AG isolates with known coreceptor tropism. We reanalyzed and confirmed previous results of the isolates by infecting the cell lines U87.CD4-CCR5 and U87.CD4-CXCR4, and by direct sequencing of the V3 region with flanking regions (Table 1). The V1-V3 region of each isolate was amplified and used for production of chimeric viruses. Tropism results of infections with chimeric viruses were in concordance with results from isolate infections, showing that the TRT assay can be used for determination of coreceptor tropism of HIV-1 subtype A and CRF02_AG (Table 1). The high concordance also suggests that the major determinants for coreceptor use are located within the V1-V3 region.

Prevalence of HIV-1 CXCR4-using populations in subtype A and CRF02_AG infected individuals in late-stage disease

To investigate the prevalence of CXCR4-using populations in the 28 individuals from Guinea-Bissau infected with HIV-1 subtype A or CRF02_AG, we constructed infectious chimeric viruses with patient-specific gp120 V1-V3 regions. All chimeric viruses that were tested established productive infections with a significant increase in p24 antigen production over time (1-9 days) in U87.CD4-CCR5 and/or U87.CD4-CXCR4 cells. Twenty-one (75%) of the individuals studied had viruses that used both CCR5 and CXCR4 for cellular entry, whereas one (4%) and six (21%) individuals had pure X4 or R5 populations, respectively (Table 2). In subtype A infected individuals, three of six had CXCR4-using populations, whereas the corresponding number in CRF02_AG infected individuals were 19 of 22 (86%).

Recently, an evolving epidemic with increasing frequency of CXCR4 tropism among subtype C-infected individuals was suggested [11]. To determine if a similar pattern could be seen in our HIV-1 CRF02_AG material from Guinea-Bissau, we divided the data set in one early group (samples collected from 1997 to 2001), and one late group (samples collected from 2003 to 2007). In the early group, five out of eight samples had viruses that were CXCR4-tropic, whereas all samples in the late group were of this phenotype (14 out of 14) ($p = 0.036$, two-tailed Fisher's exact test). The two groups were well balanced with no significant differences in CD4+ T cell counts ($p = 0.646$ and $p = 0.220$ for CD4tot and CD4%, respectively, Mann-Whitney U test). To investigate if this difference could be found in a larger material of HIV-1 CRF02_AG we added all available data from the literature and in The Los Alamos Sequence Data Base where we could couple patient-specific disease status with coreceptor tropism and sampling year (Additional file 5). We found data of 43 different individuals sampled during 1997-2001 or 2003-2007 with defined late-stage disease and with known coreceptor tropism (42 from Cameroon and 1 from Ghana). When analyzed together with our samples from Guinea-Bissau, 28 out of 50 had viruses that were CXCR4-tropic in the early group, whereas 14 out of 15 were of this phenotype in the late group ($p = 0.012$, two-tailed Fisher's exact test) (Additional file 5).

Molecular characterization of the V3 region coupled to coreceptor tropism

To create a data set that would allow for comparison between the R5 and R5X4/X4 groups, we combined the 22 CRF02_AG plasma-derived sequences with the three CRF02_AG sequences from the control panel and the

86 CRF02_AG sequences available in Los Alamos sequence data base with known phenotype resulting in a final dataset of 111 sequences (75 R5 and 36 R5X4 or X4 sequences) (Additional files 3 and 4). Sequences of subsubtype A3 were too few, and therefore not subjected for further analysis. Several sequence motif-based rules and bioinformatic tools have been developed to predict coreceptor tropism based on V3 sequences of HIV-1. These have mainly been based on subtype B, and the most common are the 11/25 rule (positively charged amino acids in V3 position 11 and/or 25 predicts CXCR4 tropism) and the net charge rule (a net charge of ≥+5 predicts CXCR4 tropism) [51]. Raymond *et al.* used the 11/25 rule in combination with the net charge rule to develop a genotypic rule specific for CRF02_AG (Table 3) [52]. Here, we used these rules and two widely used bioinformatic tools, WebPSSM and Geno2Pheno to predict the HIV-1 tropism based on the V3 sequence [32,53] (Table 3). The sensitivity reflects how many sequences of the true CXCR4-using viruses that the genotypic rule or bioinformatic tool identifies as CXCR4-using, whereas the specificity reflects how many sequences of the true CCR5-using viruses that the genotypic rule or bioinformatic tool identifies as CCR5-using. The PPV (R5X4/X4) and the NPV (R5) reflect how many of the predictions that are accurate. Of existing rules and tools, Geno2Pheno had the highest sensitivity, whereas the combined rule by Raymond *et al.* had the highest specificity (Table 3).

The sensitivity is of particular interest in clinical settings since it reflects the likelihood of detecting true CXCR4-tropism, and patients having HIV-1 of this tropism may not be suitable for treatment with coreceptor antagonists. Due to the low sensitivity of previous rules for HIV-1 of CRF02_AG, we used our data set to develop new rules. The mean net charges of V3 in the R5 and R5X4/X4 groups were 3.45 (95% CI: 3.27-3.64), and 4.47 (95% CI: 4.01-4.94), respectively. Modifying the net charge rule by setting the cutoff for prediction of CXCR4 tropism to ≥+4, we improved the sensitivity to 72%, although to a cost of both specificity (47%) and PPV (46%) (Table 3). We then examined another approach by counting the total number of charged amino acids. The mean numbers of charged amino acids were 6.60 (95% CI: 6.36-6.84) and 7.58 (95% CI: 7.16-8.01) for the R5 and R5X4/X4 group, respectively. Setting the cutoff to ≥ 8 charged amino acids for prediction of CXCR4 tropism resulted in a sensitivity equal to the one obtained by Geno2Pheno (50%) (Table 3).

Table 3 Comparison of different genotypic rules and bioinformatic tools for prediction of HIV-1 coreceptor tropism based on the HIV-1 V3 amino acid sequence.

		No. patients with virus phenotype:		Performance (%)			
Prediction method	**Predicted phenotype**	**R5**	**R5X4/X4**	**Sensitivity**	**Specificity**	**PPV[4]**	**NPV[5]**
11/25	R5	71	25	31	95	73	74
	X4	4	11				
Net ≥ +5	R5	65	20	44	87	62	76
	X4	10	16				
Raymond *et al.*[1]	R5	73	22	39	97	88	77
	X4	2	14				
$WebPSSM_{X4R5}$[2]	R5	63	22	39	84	54	74
	X4	12	14				
$WebPSSM_{SINSI}$[2]	R5	69	25	31	92	65	73
	X4	6	11				
Geno2Pheno[3]	R5	66	18	50	88	67	79
	X4	9	18				
Net ≥ +4	R5	35	10	72	47	46	78
	X4	30	26				
Total ≥ 8	R5	63	18	50	84	60	78
	X4	12	18				
Net ≥ +5 and Total ≥ 8	R5	55	13	64	73	53	81
	X4	20	23				

[1]Raymond *et al.* used a combined rule were one of the following criteria was required for HIV-1 CRF02_AG CXCR4-tropism: (i) R or K at position 11 of V3 and/or K at position 25, (ii) R at position 25 of V3 and a net charge of ≥ + 5, or (iii) a net charge of ≥ + 6.

[2]Position-specific scoring matrix, http://indra.mullins.microbiol.washington.edu/webpssm/ (January 2010).

[3]Geno2Pheno was used with a false-positive rate of 10%. http://coreceptor.bioinf.mpi-inf.mpg.de/ (January 2010).

[4]Positive predictive value.

[5]Negative predictive value.

Finally, we tested all possible combinations of previous genotypic rules analyzed in this study, and the net charge rule of ≥+4 and/or the total charge rule of ≥ 8 for prediction of CXCR4 use (data not shown). Of all combinations tested, the combination of the net charge rule of ≥+5 and total charge rule of ≥ 8 resulted in the best improvement in sensitivity (64%), without losing too much in specificity (73%) (Table 3).

Discussion

In this study, we show that the V1-V3 region of the HIV-1 envelope is the major determinant for coreceptor tropism also for subtype A and CRF02_AG, and that the TRT assay accurately determines coreceptor tropism for these subtypes [29]. In our samples, we found a high prevalence of X4 populations (79%) during late-stage disease and an increasing frequency of HIV-1 with CXCR4 tropism in CRF02_AG-infected patients over time. We also demonstrate that the total number of charged amino acids may contribute to the development of genotypic rules and bioinformatic tools for prediction of coreceptor tropism of HIV-1 CRF02_AG.

HIV-1 subtype determination was done by amplifying and sequencing the envelope gp120 V3 region of 29 infected individuals in late-stage disease from Guinea-Bissau, West Africa. Twenty-eight of the individuals had HIV-1 of subtype A or CRF02_AG, confirming previous results that these are the dominating HIV-1 forms in Guinea-Bissau [54]. The remaining study subject was infected with HIV-1 subtype C, a subtype that has never been described in Guinea-Bissau before, even though it has been shown to circulate to some extent in West Africa [55]. A detailed analysis of the six patients infected with subtype A showed close genetic relationship to the previously described sub-subtype A3 (Fig. 1). This sub-subtype was first described in Senegal, but has been shown to be prevalent in several West African countries, including Guinea-Bissau [48,56].

A control panel of 11 HIV-1 subtype A and CRF02_AG isolates was used to show that the, for subtype B, well-characterized TRT assay is accurate also for subtype A and CRF02_AG. This finding suggests that the gp120 V1-V3 region is the major determinant of coreceptor phenotype also for subtype A and CRF02_AG. This conclusion is in line with the general concept previously established for subtype B that coreceptor phenotype is determined by the V1-V3 region of gp120, in particular the V3 region [23]. Our results show that the TRT can be used as a reliable alternative to the commercially available Trofile assay (Monogram Biosciences, San Francisco, USA), at least for the tested subtypes. The Trofile assay require amplification of the entire gp160 (2,500 bp), and it has been proposed that use of V1-V3 (940 bp) can be a more sensitive approach as it could reduce the risk of losing minority populations seen when amplifying a larger fragment [50].

Since we had a reliable phenotypic assay, we were able to analyze genetic traits that would best predict coreceptor phenotype for viruses from our HIV-1 CRF02_AG samples. Applying the 11/25 rule (positively charged amino acids in V3 position 11 and/or 25 predicts CXCR4 tropism) we found a sensitivity of 31%, and a specificity 95%. This is similar to a previous study of 113 CRF02_AG isolates, where the 11/25 rule had a sensitivity of 33%, and a specificity of 96% [13]. In another study the authors used the combined criteria of the 11/25 and the net charge rule of V3 sequences of 52 CRF02_AG isolates and found both high sensitivity (70%) and high specificity (98%) [52]. The results of the combined criteria could, however, not be verified by our data. A reason for the poor performance of this rule on our dataset may be that it relies on the 11/25 rule. In sequences from viruses with CXCR4 tropism from Guinea-Bissau we found in general negatively charged or non-charged amino acids in positions 11 and 25, whereas at least one of these positions in most cases were positively charged in the sequences used in the study by Raymond *et al*. (Additional files 3 and 4). We also observed many charged (both positive and negative) amino acids as a profound characteristic of sequences derived from CXCR4-using viruses (in the complete dataset). Therefore, we counted the total number of charged amino acids in the V3 region and combined it with the net charge rule. Due to sensitivity, this rule performed better than all of the analyzed rules and bioinformatic tools, without losing too much in specificity. This rule is also different to those rules that are used for subtype B, suggesting that molecular differences at the level of virus-cell receptor interaction may exist among HIV-1 subtypes. Further studies are needed to investigate subtype-specific genotypic differences involved in HIV-1 coreceptor tropism, and the sensitivity of existing genotypic rules and bioinformatic tools have to be increased before they can be used in a clinical setting.

Next, we examined the coreceptor phenotype of HIV-1 in 28 subtype A or CRF02_AG infected individuals in late-stage disease. CXCR4-using viral populations were found in as much as 79% of the analyzed samples, demonstrating the importance of analyzing samples from patients in late-stage disease when investigating HIV-1 subtype-specific predisposal for CXCR4 tropism. For HIV-1 subtype B, it is well known that CCR5 is the dominant coreceptor early in infection. Switch or broadening of coreceptor use from CCR5 to CXCR4 occurs late in disease [6]. To our knowledge, only one previous study has reported on CRF02_AG coreceptor tropism in late-stage disease [13]. Vergne *et al.* found that 56% of

the studied isolates were positive for MT-2 cell tropism, to be compared to our result of CXCR4 use of 86%. Although, it is important to note that the viruses in the study by Vergne *et al.* were isolated from samples collected between 1996 and 2001, and the corresponding number of CXCR4-tropism among our samples during this time-frame was 63%.

Connell *et al.* (2008) analyzed the results of 19 subtype C isolates, isolated in 2005, and found a higher prevalence (30%) of CXCR4 tropism than has been shown in earlier studies [11]. They suggested that HIV-1 subtype C might be an evolving epidemic, showing an increasing prevalence of CXCR4 phenotype over time in South Africa. In the present study, we performed a direct comparison by dividing our 22 CRF02_AG samples into two groups: Samples from 1997 to 2001, and samples from 2003 to 2007. The CD4+ T cell counts and clinical data were similar between the two groups, excluding any bias due to differences in disease status. In addition, all of the investigated individuals were treatment naïve. We found a significant difference between the groups, suggesting a similar kind of evolving epidemic for CRF02_AG in Guinea-Bissau that has been suggested for subtype C in South Africa. We also analyzed our data together with available data of HIV-1 CRF02_AG infected patients with known clinical parameters and coreceptor tropism and found that the trend of an evolving epidemic was consistent. Competition assays between HIV-1 R5 and X4 viruses have shown that X4 viruses in general out-compete R5 viruses due to both higher replication kinetics and higher CXCR4 than CCR5 expression in PBMC [57]. Moreover, it has been shown that sequence change occur at a rate of 1% per year in HIV-1 *env*, illustrating the constant evolution of HIV-1 on the genetic level [58]. Further studies (based on larger sample sizes than studied here and by Connell *et al.*) are needed to investigate if HIV-1 is evolving towards a more predisposed state of changing into CXCR4 phenotype on a population level, and a confirmation of this finding in larger cohorts may have implications for viral transmission, pathogenesis and disease progression.

To further the view of our results, we performed a literature review of published results regarding subtype-specific coreceptor tropism in late-stage disease, which, to our knowledge, has not been presented before. The analysis included 498 patient-specific HIV-1 samples of six different subtypes, sampled over more than 20 years (1988-2008). Data of CD4+ T cell counts and/or clinical status for all patients were examined, and only samples from individuals diagnosed with AIDS or having CD4+ T cell counts ≤ 200 cells/μl were included in the analysis. In cases were both CD4+ T cell counts and clinical status could be found, the criterion of CD4+ T cell count was used. The vast majority of patients were included based on CD4+ T cell counts. Moreover, only one sample per patient was allowed, and in cases where the same patients appeared in several studies the patient data were only used once (for details about the analysis, see Additional file 6). HIV-1 coreceptor tropism determined as MT-2 non-syncytium or syncytium inducing was regarded as CCR5 or CXCR4 tropism, respectively [22]. If HIV-1 subtype and coreceptor tropism were not specified, this data were confirmed by personal communication with the authors (Additional file 6, and Acknowledgement). Our analysis revealed a high frequency of HIV-1 R5X4 or X4 populations in late stage disease among all analyzed subtypes, except for subtype C (Fig. 2). Only 15% of the individuals infected with HIV-1 subtype C had CXCR4-using populations, compared to 66% (60%-77%) in individuals infected with HIV-1 of non-subtype C ($p < 0.001$, two-tailed Fisher's exact test).

The data from the literature review also allowed us to investigate if we could confirm the results of Connell *et al.* of an evolving epidemic in South Africa, and if this could be seen for subtype C in general [11]. We divided the subtype C data set (208 patients, samples collected in Cameroon, Ethiopia, India, Malawi, South Africa, Sweden and Zimbabwe) into an early group (samples before the year of 2000) and a late group (samples after 2000) (Additional file 7). In the early group from South Africa eight of 46 patients (17%) had CXCR4-using viruses, whereas the corresponding number in the late group was 11 of 36 (31%) ($p = 0.19$, two-tailed Fisher's exact test). Connell *et al.* reported that 30% of their isolates were able to use CXCR4 and compared this to previous studies from South Africa showing no syncytium-inducing (SI) capacity of HIV-1 isolates collected during the 1980s, whereas 10-17% of the studied isolates had SI capacity during the 1990s. No statistical evaluation was performed in their study. In the early group of our complete subtype C data set, 11 of 145 patients (8%) had CXCR4-using viruses, whereas the corresponding number in the late group was 21 of 63 (33%) ($p < 0.001$, two-tailed Fisher's exact test). Our results confirms the numbers presented by Connell *et al.* in South Africa (even though the difference between the early and the late group was not statistically significant), and further strengthens the concern of an evolving HIV-1 subtype C-epidemic on the population level.

Understanding subtype-specific differences regarding coreceptor tropism is important for several reasons. First, these studies may help us understanding differences in HIV-1 pathogenesis. Several studies have indicated differences in relative pathogenicity between different subtypes, where subtype D appears to be more pathogenic compared to other subtypes [59]. Moreover,

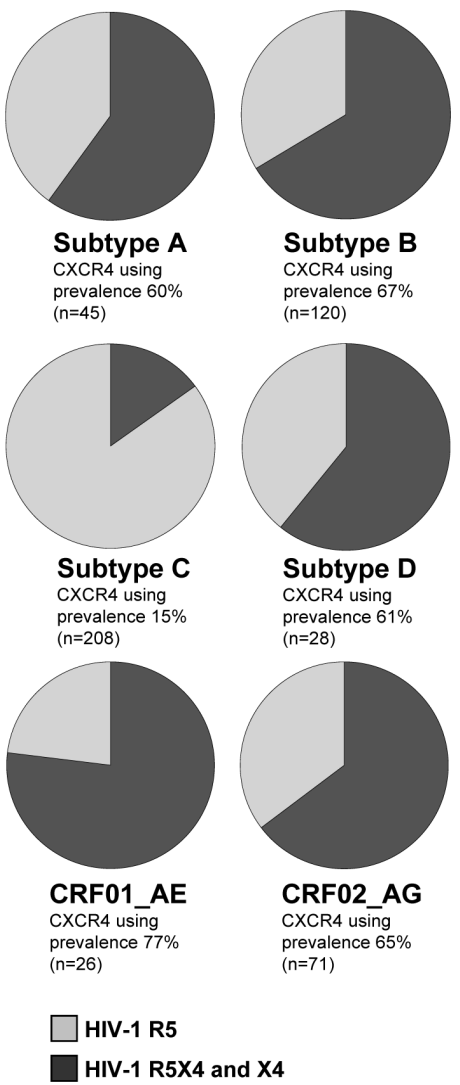


Figure 2 Prevalence of HIV-1 CXCR4-tropism in late-stage disease in different subtypes. Combined results of new data presented in this study and data from the literature analysis, showing the prevalence of HIV-1 CXCR4-tropic viruses in late stage disease. The number of individuals used in each diagram is specified within brackets. A detailed description of the analysis can be found in Additional file 6.

it has been shown that HIV-1 subtype D has a preference for CXCR4 tropism early in infection, and a connection with the faster disease progression seen in this subtype has been suggested [3]. Second, understanding HIV-1 subtype-specific differences regarding the ability to develop CXCR4-using populations may be of great importance for future treatment guidelines for coreceptor antagonists. Fätkenheuer *et al.* (2008) reported a striking difference in the appearance of X4 populations among patients with experienced treatment failure, after receiving either the CCR5-antagonist Maraviroc or a placebo treatment [60]. Despite a short follow-up period of only 48-weeks, as many as 57% of patients receiving Maraviroc developed X4 viruses (76 of 133 patients), compared to only 6% (6 of 95 patients) in the placebo group. This finding strengthens the concerns that CCR5 antagonists could increase the risk of HIV-1 populations to shift away from CCR5 to CXCR4 use, potentially leading to treatment failure and faster disease progression [61,62]. Most of the participants in the above study were infected with HIV-1 subtype B, and more studies are needed to determine if the outgrowth of R5X4 or X4 populations will be as distinct also for other subtypes, such as subtype C.

Based on the accumulated knowledge on subtype C infections, it is tempting to speculate that subtype C-infected patients in general may be more suitable for treatment with CCR5 antagonists than patients infected with other subtypes, at least in late-stage disease. Taken these thoughts further, different subtypes and CRFs may be more or less predisposed for the emergence of R5X4 or X4 populations, although it seems that CXCR4-using HIV-1 populations likely will arise in most HIV-1 non-subtype C infections (Fig. 2). The comparative picture of subtype-dependent HIV-1 coreceptor tropism in late-stage disease underlines the importance of further studies on HIV-1 subtype-dependent coreceptor tropism, an issue of direct clinical importance for the development of treatment guidelines of the recently introduced coreceptor antagonists in HIV treatment regimens.

Conclusion

In summary, we show that the TRT assay accurately determines the HIV-1 coreceptor tropism of subtype A and CRF02_AG. Using this assay we found a high amount of HIV-1 CXCR4-using populations (79%) in our plasma samples and an increasing frequency of CXCR4 tropism in CRF02_AG-infected patients over time. The emergence of CXCR4 use may have implications for viral transmission, pathogenesis and disease progression. We also demonstrate that the combined criteria of the total number of charged amino acids and

net charge of the V3 region is a more sensitive predictor of CXCR4 tropism for HIV-1 CRF02_AG compared to the analyzed genotypic rules and bioinformatic tools. Finally, we conducted an extensive literature analysis of 498 individuals which, to our knowledge, is the most extensive comparison of subtype-specific coreceptor tropism in late-stage disease. We found a generally high frequency of CXCR4 tropism among all major subtypes and CRFs, except for subtype C. These results demonstrate the need for further studies investigating subtype-specific emergence for CXCR4-tropism, this may be particularly important due to the introduction of CCR5-antagonists in HIV treatment regimens.

Additional file 1: Table S1 - Reference dataset of subsubtype A1-A3 and CRF02_AG sequences. HIV-1 reference dataset of subsubtype A1-A3 and CRF02_AG sequences used in the final phylogenetic reconstruction for subsubtype and CRF02_AG determination of the sample sequences.

Additional file 2: Table S2 - Reference dataset of CRF02_AG sequences used in the molecular analysis. Accession numbers of the HIV-1 reference dataset of HIV-1 CRF02_AG sequences with phenotypically determined coreceptor tropism used in the molecular analysis.

Additional file 3: Table S3 - Alignment and molecular characteristics of HIV-1 CRF02_AG V3 amino acid sequences from study samples and references with determined CCR5 tropism. Summary of the molecular characteristics of the CCR5 tropic sequences used in the genotypic analysis.

Additional file 4: Table S4 - Alignment and molecular characteristics of HIV-1 CRF02_AG V3 amino acid sequences from study samples and references with determined CXCR4 tropism. Summary of the molecular characteristics of the CXCR4 tropic sequences used in the genotypic analysis.

Additional file 5: Table S5. Data used to investigate an evolving epidemic available data from HIV-1 CRF02_AG. Summary of the data obtained from the literature and Los Alamos Sequence Data Base to investigate if the HIV-1 CRF02_AG epidemic represents an evolving epidemic.

Additional file 6: Table S6 - Overview of the literature analysis. Summary of the data obtained from the literature and used to determine the amount of CXCR4 tropism in late stage disease in the studied HIV-1 subtypes.

Additional file 7: Table S7 - Overview of the subtype C material. Summary of the subtype C data obtained from the literature review used in the analysis of an evolving epidemic for subtype C.

Acknowledgements
We thank Professor Jan Albert (Swedish Institute for Infectious Disease Control) for providing the isolates used in the evaluation of the TRT assay and SCIBLU Genomics (Swegene Centre for Integrative Biology at Lund University) for technical assistance. We also thank Professor Jan Albert, Professor Hanneke Schuitemaker, Dr. Gabriella Scarlatti and Dr. Åsa Björndal for providing information about clinical data and coreceptor tropism from previous studies. We would also like to acknowledge the anonymous reviewers for valuable comments on the manuscript.
This work was supported by grants from the Swedish Research Council, the Swedish International Development Cooperation Agency/Department for Research Cooperation (SIDA/SAREC), the Crafoord Foundation, Lund, Sweden, the Royal Physiographic Society, Lund, Sweden, The Lars Hierta Memorial Foundation, Stockholm, Sweden, and Konsul Thure Carlsson Fund for Medical Research, Lund, Sweden. DNA sequencing was performed at the SWEGENE Center of Genomic Ecology, supported by the Crafoord foundation, through the SWEGENE consortium.

Author details
[1]Department of Experimental Medical Science, Section of Molecular Virology, Lund University, Lund, Sweden. [2]Department of Clinical Sciences, Malmö, Infectious Diseases Research Unit, Lund University, Sweden. [3]Department of Laboratory Medicine Lund, Section of Virology, Lund University, Lund, Sweden. [4]Department of Clinical Sciences, Division of Infection Medicine, Lund University, Lund, Sweden. [5]National Public Health Laboratory, Bissau, Guinea-Bissau.

Authors' contributions
JE designed the study, optimized the experimental protocols, performed RNA extraction and sequencing, analyzed and interpreted the data, performed the literature analysis, and wrote the manuscript. FM contributed in study design, clinically evaluated the patient data and participated in patient selection. WMA performed most of the cloning and colony-PCR, and participated in optimization of the cloning strategy. EV carried out cell assays, ELISA experiments, and participated in optimization of the TRT assay. AJB was medically and organizationally responsible for the clinical sites with biological samples of the study participants in the cohort. ZJdS was responsible for analyses of HIV serology at the laboratory in Guinea-Bissau. EMF participated in interpretation of the results, and contributed to the literature analysis. HN contributed in study design, and participated in patient selection. PM designed the study, participated in analyzing and in interpretation of the data, and helped to draft the manuscript. All authors read and approved the manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 12 January 2010 Accepted: 22 March 2010
Published: 22 March 2010

References

1. Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM: **The challenge of HIV-1 subtype diversity.** *N Engl J Med* 2008, **358**:1590-1602.
2. Berger EA, Doms RW, Fenyo EM, Korber BT, Littman DR, Moore JP, Sattentau QJ, Schuitemaker H, Sodroski J, Weiss RA: **A newclassification for HIV-1.** *Nature* 1998, **391**:240.
3. Kaleebu P, Nankya IL, Yirrell DL, Shafer LA, Kyosiimire-Lugemwa J, Lule DB, Morgan D, Beddows S, Weber J, Whitworth JA: **Relation between chemokine receptor use, disease stage, and HIV-1 subtypes A and D: results from a rural Ugandan cohort.** *J Acquir Immune Defic Syndr* 2007, **45**:28-33.
4. Asjo B, Morfeldt-Manson L, Albert J, Biberfeld G, Karlsson A, Lidman K, Fenyo EM: **Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection.** *Lancet* 1986, **2**:660-662.
5. Karlsson A, Parsmyr K, Sandstrom E, Fenyo EM, Albert J: **MT-2 cell tropism as prognostic marker for disease progression in human immunodeficiency virus type 1 infection.** *J Clin Microbiol* 1994, **32**:364-370.
6. Koot M, Keet IP, Vos AH, de Goede RE, Roos MT, Coutinho RA, Miedema F, Schellekens PT, Tersmette M: **Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS.** *Ann Intern Med* 1993, **118**:681-688.
7. Bjorndal A, Deng H, Jansson M, Fiore JR, Colognesi C, Karlsson A, Albert J, Scarlatti G, Littman DR, Fenyo EM: **Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype.** *J Virol* 1997, **71**:7478-7487.
8. Abebe A, Demissie D, Goudsmit J, Brouwer M, Kuiken CL, Pollakis G, Schuitemaker H, Fontanet AL, Rinke de Wit TF: **HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS.** *Aids* 1999, **13**:1305-1311.
9. Bjorndal A, Sonnerborg A, Tscherning C, Albert J, Fenyo EM: **Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients.** *AIDS Res Hum Retroviruses* 1999, **15**:647-653.
10. Ping LH, Nelson JA, Hoffman IF, Schock J, Lamers SL, Goodman M, Vernazza P, Kazembe P, Maida M, Zimba D, Goodenow MM, Eron JJ Jr, Fiscus SA, Cohen MS, Swanstrom R: **Characterization of V3 sequence**

heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: underrepresentation of X4 variants. *J Virol* 1999, **73**:6271-6281.
11. Connell BJ, Michler K, Capovilla A, Venter WD, Stevens WS, Papathanasopoulos MA: **Emergence of X4 usage among HIV-1 subtype C: evidence for an evolving epidemic in South Africa.** *Aids* 2008, **22**:896-899.
12. Yu XF, Wang Z, Beyrer C, Celentano DD, Khamboonruang C, Allen E, Nelson K: **Phenotypic and genotypic characteristics of human immunodeficiency virus type 1 from patients with AIDS in northern Thailand.** *J Virol* 1995, **69**:4649-4655.
13. Vergne L, Bourgeois A, Mpoudi-Ngole E, Mougnutou R, Mbuagbaw J, Liegeois F, Laurent C, Butel C, Zekeng L, Delaporte E, Peeters M: **Biological and genetic characteristics of HIV infections in Cameroon reveals dual group M and O infections and a correlation between SI-inducing phenotype of the predominant CRF02_AG variant and disease stage.** *Virology* 2003, **310**:254-266.
14. Richman DD, Bozzette SA: **The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression.** *J Infect Dis* 1994, **169**:968-974.
15. Mild M, Esbjornsson J, Fenyo EM, Medstrand P: **Frequent intrapatient recombination between human immunodeficiency virus type 1 R5 and X4 envelopes: implications for coreceptor switch.** *J Virol* 2007, **81**:3369-3376.
16. Bratt G, Leandersson AC, Albert J, Sandstrom E, Wahren B: **MT-2 tropism and CCR-5 genotype strongly influence disease progression in HIV-1-infected individuals.** *Aids* 1998, **12**:729-736.
17. Peeters M, Vincent R, Perret JL, Lasky M, Patrel D, Liegeois F, Courgnaud V, Seng R, Matton T, Molinier S, Delaporte E: **Evidence for differences in MT2 cell tropism according to genetic subtypes of HIV-1: syncytium-inducing variants seem rare among subtype C HIV-1 viruses.** *J Acquir Immune Defic Syndr Hum Retrovirol* 1999, **20**:115-121.
18. Spijkerman IJ, Koot M, Prins M, Keet IP, Hoek van den AJ, Miedema F, Coutinho RA: **Lower prevalence and incidence of HIV-1 syncytium-inducing phenotype among injecting drug users compared with homosexual men.** *Aids* 1995, **9**:1085-1092.
19. Cilliers T, Nhlapo J, Coetzer M, Orlovic D, Ketas T, Olson WC, Moore JP, Trkola A, Morris L: **The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C.** *J Virol* 2003, **77**:4449-4456.
20. Huang W, Eshleman SH, Toma J, Fransen S, Stawiski E, Paxinos EE, Whitcomb JM, Young AM, Donnell D, Mmiro F, Musoke P, Guay LA, Jackson JB, Parkin NT, Petropoulos CJ: **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**:7885-7893.
21. Tscherning C, Alaeus A, Fredriksson R, Bjorndal A, Deng H, Littman DR, Fenyo EM, Albert J: **Differences in chemokine coreceptor usage between genetic subtypes of HIV-1.** *Virology* 1998, **241**:181-188.
22. Zhang L, Huang Y, He T, Cao Y, Ho DD: **HIV-1 subtype and second-receptor use.** *Nature* 1996, **383**:768.
23. Hartley O, Klasse PJ, Sattentau QJ, Moore JP: **V3: HIV'sswitch-hitter.** *AIDS Res Hum Retroviruses* 2005, **21**:171-189.
24. Pastore C, Nedellec R, Ramos A, Pontow S, Ratner L, Mosier DE: **Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations.** *J Virol* 2006, **80**:750-758.
25. Sax PE: **FDA approval: maraviroc.** *AIDS Clin Care* 2007, **19**:75.
26. Weber J, Piontkivska H, Quinones-Mateu ME: **HIV type 1 tropism and inhibitors of viral entry: clinical implications.** *AIDS Rev* 2006, **8**:60-77.
27. Westby M, Smith-Burchnell C, Mori J, Lewis M, Mosley M, Stockdale M, Dorr P, Ciaramella G, Perros M: **Reduced maximal inhibition in phenotypic susceptibility assays indicates that viral strains resistant to the CCR5 antagonist maraviroc utilize inhibitor-bound receptor for entry.** *J Virol* 2007, **81**:2359-2371.
28. Wolinsky SM, Veazey RS, Kunstman KJ, Klasse PJ, Dufour J, Marozsan AJ, Springer MS, Moore JP: **Effect of a CCR5 inhibitor on viral loads in macaques dual-infected with R5 and X4 primate immunodeficiency viruses.** *Virology* 2004, **328**:19-29.
29. Trouplin V, Salvatori F, Cappello F, Obry V, Brelot A, Heveker N, Alizon M, Scarlatti G, Clavel F, Mammano F: **Determination of coreceptor usage of human immunodeficiency virus type 1 from patient plasma samples by using a recombinant phenotypic assay.** *J Virol* 2001, **75**:251-259.
30. Whitcomb JM, Huang W, Fransen S, Limoli K, Toma J, Wrin T, Chappey C, Kiss LD, Paxinos EE, Petropoulos CJ: **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**:566-575.
31. Garrido C, Roulet V, Chueca N, Poveda E, Aguilera A, Skrabal K, Zahonero N, Carlos S, Garcia F, Faudon JL, Soriano V, de Mendoza C: **Evaluation of eight different bioinformatics tools to predict viral tropism in different human immunodeficiency virus type 1 subtypes.** *J Clin Microbiol* 2008, **46**:887-891.
32. Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, He HX, McLaughlin S, Shankarappa R, Margolick JB, Mullins JI: **Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loopsequences.** *J Virol* 2003, **77**:13376-13388.
33. Andersson S, Norrgren H, da Silva Z, Biague A, Bamba S, Kwok S, Christopherson C, Biberfeld G, Albert J: **Plasma viral load in HIV-1 and HIV-2 singly and dually infected individuals in Guinea-Bissau, West Africa: significantly lower plasma virus set point in HIV-2 infection than in HIV-1 infection.** *Arch Intern Med* 2000, **160**:3286-3293.
34. Norrgren H, Andersson S, Naucler A, Dias F, Johansson I, Biberfeld G: **HIV-1, HIV-2, HTLV-I/II and Treponema pallidum infections: incidence, prevalence, and HIV-2-associated mortality in an occupational cohort in Guinea-Bissau.** *J Acquir Immune Defic Syndr Hum Retrovirol* 1995, **9**:422-428.
35. Mansson F, Biague A, da Silva ZJ, Dias F, Nilsson LA, Andersson S, Fenyo EM, Norrgren H: **Prevalence and incidence of HIV-1 and HIV-2 before, during and after a civil war in an occupational cohort in Guinea-Bissau, West Africa.** *Aids* 2009, **23**:1575-1582.
36. **Guidelines for national human immunodeficiency virus case surveillance, including monitoring for human immunodeficiency virus infection and acquired immunodeficiency syndrome. Centers for Disease Control andPrevention.** *MMWR Recomm Rep* 1999, **48**:1-27, 29-31.
37. CB H: **WHO Staging System for HIV Infection and Disease in Adolescents and Adults.** *Book WHO Staging System for HIV Infection and Disease in Adolescents and Adults (Editor ed.^eds.)* City: UCSF Center for HIV Information 2006.
38. Leitner T, Korovina G, Marquina S, Smolskaya T, Albert J: **Molecular epidemiology and MT-2 cell tropism of Russian HIV type 1 variant.** *AIDS Res Hum Retroviruses* 1996, **12**:1595-1603.
39. **Los Alamos Sequence Database.** [http://www.hiv.lanl.gov/].
40. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**:1596-1599.
41. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
42. Posada D, Crandall KA: **MODELTEST: testing the model ofDNA substitution.** *Bioinformatics* 1998, **14**:817-818.
43. Felsenstein J: **Confidence limits on phylogenies: An approach using the bootstrap.** *Evolution* 1985, **39**:783-791.
44. Nei MKS: *Molecular evolution and phylogenetics* Oxford, Oxford University Press 2000.
45. Dopazo J: **Estimating errors and confidence intervals for branch lengths in phylogenetic trees by a bootstrap approach.** *J Mol Evol* 1994, **38**:300-304.
46. Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di Marzio P, Marmon S, Sutton RE, Hill CM, Davis CB, Peiper SC, Schall TJ, Littman DR, Landau NR: **Identification of a major co-receptor for primary isolates of HIV-1.** *Nature* 1996, **381**:661-666.
47. Mild M, Bjorndal A, Medstrand P, Fenyo EM: **Isolation of human immunodeficiency virus-type 1 (HIV-1) clones with biological and molecular properties of the primary isolate.** *Virology* 2006, **350**:58-66.
48. Meloni ST, Kim B, Sankale JL, Hamel DJ, Tovanabutra S, Mboup S, McCutchan FE, Kanki PJ: **Distinct human immunodeficiency virus type 1 subtype A virus circulating in West Africa: sub-subtype A3.** *J Virol* 2004, **78**:12438-12445.
49. Raymond S, Delobel P, Mavigner M, Cazabat M, Souyris C, Sandres-Saune K, Cuzin L, Marchou B, Massip P, Izopet J: **Correlation between genotypic**

IV

predictions based on V3 sequences and phenotypic determination of HIV-1 tropism. *Aids* 2008, **22**:F11-16.
50. Skrabal K, Low AJ, Dong W, Sing T, Cheung PK, Mammano F, Harrigan PR: **Determining human immunodeficiency virus coreceptor use in a clinical setting: degree of correlation between two phenotypic assays and a bioinformatic model.** *J Clin Microbiol* 2007, **45**:279-284.
51. 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.
52. Raymond S, Delobel P, Mavigner M, Cazabat M, Souyris C, Encinas S, Sandres-Saune K, Pasquier C, Marchou B, Massip P, Izopet J: **Genotypic Prediction of HIV-1 CRF02-AG Tropism.** *J Clin Microbiol* 2009, **47(7)**:2292-4.
53. Sing T, Low AJ, Beerenwinkel N, Sander O, Cheung PK, Domingues FS, Buch J, Daumer M, Kaiser R, Lengauer T, Harrigan PR: **Predicting HIV coreceptor usage on the basis of genetic and clinical covariates.** *Antivir Ther* 2007, **12**:1097-1106.
54. Andersson S, Norrgren H, Dias F, Biberfeld G, Albert J: **Molecular characterization of human immunodeficiency virus (HIV)-1 and -2 in individuals from guinea-bissau with single or dual infections: predominance of a distinct HIV-1 subtype A/G recombinant in West Africa.** *Virology* 1999, **262**:312-320.
55. Hamel DJ, Sankale JL, Eisen G, Meloni ST, Mullins C, Gueye-Ndiaye A, Mboup S, Kanki PJ: **Twenty years of prospective molecular epidemiology in Senegal: changes in HIV diversity.** *AIDS Res Hum Retroviruses* 2007, **23**:1189-1196.
56. Meloni ST, Sankale JL, Hamel DJ, Eisen G, Gueye-Ndiaye A, Mboup S, Kanki PJ: **Molecular epidemiology of human immunodeficiency virus type 1 sub-subtype A3 in Senegal from 1988 to 2001.** *J Virol* 2004, **78**:12455-12461.
57. Tebit DM, Nankya I, Arts EJ, Gao Y: **HIV diversity, recombination and disease progression: how does fitness "fit" into the puzzle?.** *AIDS Rev* 2007, **9**:75-87.
58. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, Gupta P, Rinaldo CR, Learn GH, He X, Huang XL, Mullins JI: **Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection.** *J Virol* 1999, **73**:10489-10502.
59. Kaleebu P, French N, Mahe C, Yirrell D, Watera C, Lyagoba F, Nakiyingi J, Rutebemberwa A, Morgan D, Weber J, Gilks C, Whitworth J: **Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda.** *J Infect Dis* 2002, **185**:1244-1250.
60. Fatkenheuer G, Nelson M, Lazzarin A, Konourina I, Hoepelman AI, Lampiris H, Hirschel B, Tebas P, Raffi F, Trottier B, Bellos N, Saag M, Cooper DA, Westby M, Tawadrous M, Sullivan JF, Ridgway C, Dunne MW, Felstead S, Mayer H, Ryst van der E, MOTIVATE 1, MOTIVATE 2 Study Teams: **Subgroup analyses of maraviroc in previously treated R5 HIV-1 infection.** *N Engl J Med* 2008, **359**:1442-1455.
61. Dolin R: **A new class of anti-HIV therapy and new challenges.** *N Engl J Med* 2008, **359**:1509-1511.
62. Kilby JM, Eron JJ: **Novel therapies based on mechanisms of HIV-1 cell entry.** *N Engl J Med* 2003, **348**:2228-2238.

doi:10.1186/1742-4690-7-23
Cite this article as: Esbjörnsson *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.

Additional Table S1. HIV-1 reference dataset of subsubtype A1-A3 and CRF02_AG sequences used in the final phylogenetic reconstruction for subsubtype and CRF02_AG determination of the sample sequences.

AB098332	AF063223	AF361873	AF457079	AF539405	AY371124	DQ083238
AB231896	AF063224	AF377954	AF457081	AJ251056	AY371126	DQ168577
AB231898	AF069669	AF377955	AF457083	AJ251057	AY371139	DQ168578
AB253421	AF069670	AF457052	AF457084	AY151001	AY371140	EU110087
AB253428	AF069673	AF457055	AF457086	AY151002	AY371141	EU110088
AB286857	AF107770	AF457063	AF457089	AY253314	AY371142	EU110092
AB286859	AF107771	AF457065	AF484478	AY271690	AY521629	EU110094
AB286862	AF286237	AF457066	AF484493	AY322184	AY521630	EU786671
AB286863	AF286238	AF457067	AF484507	AY322190	AY521631	
AB287376	AF286241	AF457069	AF484508	AY322193	AY713406	
AF004885	AF361872	AF457075	AF484509	AY371122	DD409979	

Additional Table S2. Accession numbers of the HIV-1 reference dataset of HIV-1 CRF02_AG sequences with phenotypically determined coreceptor tropism used in the molecular analysis.

AB049811	AM279352	AY371127	FJ652327	FJ652340	FJ652354	FJ652370
AF063223	AM279356	AY371128	FJ652328	FJ652341	FJ652355	FJ652371
AF119216	AM279358	AY371138	FJ652329	FJ652342	FJ652356	FJ652373
AF184155	AM279360	AY736839	FJ652330	FJ652343	FJ652357	FJ652374
AF355318	AM279361	AY736840	FJ652331	FJ652344	FJ652359	FJ652375
AF355320	AM279362	AY994510	FJ652332	FJ652345	FJ652362	FJ652377
AF355321	AM279367	DQ177193	FJ652333	FJ652346	FJ652363	FJ652378
AF355325	AY271690	DQ177209	FJ652334	FJ652347	FJ652364	L22939
AF355327	AY371122	DQ825459	FJ652335	FJ652348	FJ652365	
AF355331	AY371123	DQ825460	FJ652336	FJ652349	FJ652366	
AF355334	AY371124	DQ825462	FJ652337	FJ652350	FJ652367	
AF355335	AY371125	DQ825471	FJ652338	FJ652351	FJ652368	
AF355336	AY371126	DQ869018	FJ652339	FJ652353	FJ652369	

Additional Table S3 – Alignment and molecular characteristics of HIV-1 CRF02_AG V3 amino acid sequences from study samples and references with determined CCR5 tropism.

		R5							
Set[1]	Sample[2]	V3 sequence[3]	Length	Position[4]		Charge[5]			
		| | || | | || |		11	25	+	-	Net	Total
GB	DL2391G	CTRPNNNTRKSIGIGPGQTFYAAGEIIGDIRRAHC RGV	35	0	-	5	2	3	7
GB	DL2853E	CIRPNNNTRKSIRIGPGQTFYARGDIIGDIRQAHC V T Y	35	0	-	5-6	2	3-4	7-8
GB	DL4477D	CTRPGNNTRKGVHMGTGKTFYATGDIIGDIRQAHC	35	0	-	5	2	3	7
Control	22480	CIRPNNNTRKSIRIGPGQTFYAT-DIIGNIRQAHC	35	0	-	5	1	4	6
Control	22627	CTRPGNNTRQSVRIGPGQTFYARGDITGDIRQAHC	35	0	-	5	2	3	7
LASDB	AB049811	CTRPNNNTRKSVRIGPGQTFYATGDIIGDIRQAHC	35	0	-	5	2	3	7
LASDB	AF063223	CPRPNNNTRKSVRIGPGQTFYATGDIIGDIRQAHC	35	0	-	5	2	3	7
LASDB	AF119216	CTKPNNNTRKSVRIGPGQTFYASGDIVGNIRQAHC	35	0	-	5	1	4	6
LASDB	AF184155	CTRPNNNTRKSVRIGPGQTFYATGGIIGDIRQAHC	35	0	0	5	1	4	6
LASDB	AF355318	CTRPNNNTRKSVRIGPGQTFYATGEVIGDIRQAHC	35	0	-	5	2	3	7
LASDB	AF355321	CTRPGNNTRKSVRIGPGQTFYATGDIIGDIRQAHC	35	0	-	5	2	3	7
LASDB	AF355325	CTRPNNNTRKRVPIGPGQTFYAT-DIIGNIRQAHC	34	+	-	5	1	4	6
LASDB	AF355327	CIRPNNNAIKNVGIGPGQTFYTTGKIIGNISQAYC	35	0	+	3	0	3	3
LASDB	AF355331	CTRPNNNTRTSVRIGPGQTFYATGAIIGDIRKAYC	35	0	0	5	1	4	6
LASDB	AF355335	CTRPHNNTRTSTRIGPGQTFYATGDIIGDIRQAHC	35	0	-	4	2	2	6
LASDB	AM279352	CTRVANNTRTSVRIGPGQTFYATGGIIGDIRQAHC	35	0	0	4	1	3	5
LASDB	AM279358	CTRPDNNTRKSVRIGPGQAFY-TNEIIGDIRKAHC	34	0	-	6	3	3	9
LASDB	AM279360	CSRPGNNTRQSVRIGPGQTFYATGEIIGDIRQAHC	35	0	-	4	2	2	6
LASDB	AM279361	CTRPNNNTRRSVRIGPGQTFYAAGEIIGDIRQAHC	35	0	-	5	2	3	7
LASDB	AM279362	CTRPNNNTRKSVRIGPGQAFYTTTDITGDIRQAHC	35	0	-	5	2	3	7
LASDB	AM279367	CTRPGNNTRKSVRIGPGQTFYATGDIIGDIRQAHC	35	0	-	5	2	3	7
LASDB	AY371122	CTRPSNNTRKSIRIGPGQTFYATNNIIGNIRQAHC	35	0	0	5	0	5	5
LASDB	AY371123	CTRPGNNTRKSVRIGPGQTFYATGDIIGDIRKAHC	35	0	-	6	2	4	8
LASDB	AY371124	CIRPSNNTRKSIRIGPGQTFYATGAIIGDIRRAHC	35	0	0	6	1	5	7
LASDB	AY371125	CTRPSNNTRTSVRIGPGQTFYGTGEIIGDIRKAYC	35	0	-	5	2	3	7
LASDB	AY371126	CTRPGNNTRKSVRIGPGQTFYASGAIIGDIRQSHC	35	0	0	5	1	4	6
LASDB	AY371127	CTRPNNNTRKSVRIGPGQTFYATGEIVGNIRQAYC	35	0	-	5	1	4	6
LASDB	AY371128	CVRPGNNTIRSIRIGPGQSFHGTGNIIGDIRQAHC	35	0	0	4	1	3	5
LASDB	AY371138	CTRTGKNTRTSIHMGPGQSAFFAGEVIRDIRLAYC	35	0	-	5	2	3	7
LASDB	AY736839	CTRPNNNTRKSVRIGPGQTFYATGDIIGDIRQAHC	35	0	-	5	2	3	7
LASDB	AY736840	CTRPNNNTRKSVRIGPGQTFYATGDIIGNIRQAYC	35	0	-	5	1	4	6
LASDB	AY994510	CTRPSNNTRKGWHIGPGQTLYATGAIIGDIRQAHC	35	0	0	4	1	3	5
LASDB	DQ177193	CTRPNNNTRRDIGIGPGQTFFAAGAIIGDIRQASC	35	-	0	4	2	2	6
LASDB	DQ177209	CIRPNNNTRRSVRIGPGQTFYATGDIIGDIRQAYC	35	0	-	5	2	3	7
LASDB	DQ825471	CTRPGNNTRQSIRIGPGQTFYATGDIIGDIRKAFC V	35	0	-	5	2	3	7
LASDB	DQ869018	CTRPGNNTRKSVRIGPGQTFYATGDITGDIRQAHC	35	0	-	5	2	3	7
LASDB	FJ652327	CTRPNNNTRKSVRIGPGQTFYATGGIIGDIRQAHC Y	35	0	0	5	1	4	6
LASDB	FJ652328	CTRPGNNTRKSVRIGPGQTFYATGDIIGDIRQAHC	35	0	-	5	2	3	7
LASDB	FJ652329	CIRPGNNTSKGIHIGPGRTFFANDRIIGDIRSAHC	35	0	+	5	2	3	7
LASDB	FJ652330	CIRPGNNTRKSMRIGPGQTFYATGDIIGDIRQAHC	35	0	-	5	2	3	7
LASDB	FJ652331	CTRPNNNTRKSIHIGPGRAFYATGDVIGNPRQAHC M D T D K	35	0	-	5	1-3	2-4	6-8
LASDB	FJ652332	CIRPNNNIRKSVRIGPGQTFYATGEIIGDIRKAHC S T T Q Y	35	0	-	4-6	2	2-4	6-8
LASDB	FJ652333	CMRPNNNTRKSVRIGPGQTFYAT-DIIGKIRQAYC N E	34	0	-	5-6	1-2	3-5	6-8
LASDB	FJ652335	CTRPSNNTRKSVHIGPGQTFYATGQIIGDIRQAHC	35	0	0	4	1	3	5

		G P T							
LASDB	FJ652336	CIRPGNNTRKSGRRGPGQAFYATGDIIGDIRQAHC	35	0	-	5-6	2-3	2-4	7-9
		V I D T							
LASDB	FJ652337	CTRPANKTIKGVRIGPGQTFYTTGSIIGNIRQAHC	35	0	0	4-5	0-1	3-5	4-6
		N D Y							
LASDB	FJ652338	CTRPHNNTRKSVHIGPGQAFYATGDIIGDIRQAHC	35	0	-	4	2	2	6
LASDB	FJ652339	CTRPNNNTRRSVRIGPGQAFYATGDIIGNIRQAHC	35	0	-	5	1-2	3-4	6-7
		K A T D K Y							
		A							
		V							
LASDB	FJ652340	CTRPGNNTRQSVRIGPGQVFYA-NPIIGDIRQAHC	34	0	0	4	1	3	5
LASDB	FJ652341	CVRPNNNTRKSVRIGPGQTFYATGDIRGDIRQAYC	35	0	-	5-6	2	3-4	7-8
		T							
		S							
		I							
		L							
LASDB	FJ652342	CTRPNNNTRKSVRIGPGQAFYAT-DIIGDIRQAYC	34	0	-	5	2	3	7
LASDB	FJ652343	CTRPGNNTRKSVRIGPGQTFYATGDIIGDIRQAHC	35	0	-	5	2	3	7
LASDB	FJ652344	CIRPNNNTRKSVRIGPGQTFYAAGEIIGNIRQAHC	35	0	-	5	1	4	6
		G							
LASDB	FJ652345	CTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQAHC	35	0	-	5	2	3	7
LASDB	FJ652346	CIRPNNNTRKSVHIGPGQAFYATNDIIGDIRQAHC	35	0	-	4	2	2	6
		G PM							
LASDB	FJ652347	CTRPGNNTRTSVRIGPGQAFYATGDIIGDIRKAHC	35	0	-	5	2	3	7
		L							
LASDB	FJ652348	CTRLNNNTRTSIRIGPGQSFHATGQIVGDIRQAHC	35	0	0	4	1	3	5
LASDB	FJ652349	CIRPNNNTRKSVRIGPGQTFYATGDIIGNIRKAYC	35	0	-	6	1-2	4-5	7-8
		D K							
LASDB	FJ652350	CIRPNNNKRRSVRIGPGQTFYATGDIIGNIRQAYC	35	0	-	6	1	5	7
LASDB	FJ652351	CTRPGNNTRKSMRIGPGQTFYAYGEIIGDIRQAYC	35	0	-	4-5	2	2-3	6-7
		G H V							
LASDB	FJ652353	CIRPGNNKRTSIRIGPGQTFYATGEIIGNIRQAQC	35	0	-	5	1	4	6
		S							
LASDB	FJ652354	CVRPNNNTRKSVRIGPGQTFYATGEIIGDIRQAHC	35	0	-	4-5	2-3	1-3	6-8
		G H DD K							
LASDB	FJ652356	CSRPNNNTRKSVRIGPGQTFYAT-DIIGDIRQAHC	34	0	-	4-5	2-3	1-3	6-8
		V I PP							
		Q							
		E							
LASDB	FJ652357	CTRPNNNTRKSVRIGPGQTFYATGDIIGNIRQAHC	35	0	-	5	1-3	2-4	6-8
		A G E D K							
		S							
		E							
LASDB	FJ652359	CTRPGNNTRKSVRIGPGQTFFATGAIIGNIRQAHC	35	0	-/0	5	0-2	3-5	5-7
		A D T D							
		I							
		V							
LASDB	FJ652362	CTRPNNNTRRSVRIGPGQTFYATGGIIGNIRQAHC	35	0	-/0	5	0-2	3-5	5-7
		I K A D							
		D							
LASDB	FJ652363	CTRPSNNTIKGIHIGPGRAFYTTGQIIGDIRQAHC	35	0	0	4	1	3	5
LASDB	FJ652364	CTRPSNNTIKGIHIGPGRAFYTTGQIIGDIRQAHC	35	0	0	4	1	3	5
		V F							
LASDB	FJ652365	CTRPNNNTRKSVHLGPGQAFYATGDIIGDIRQAYC	35	0	-	4	2	2	6
LASDB	FJ652366	CIRPGNNTRKSVRIGPGQAFYATGAVAGDIRQAHC	35	0	0	5	1	4	6
LASDB	FJ652367	CIRPGNNTRKSVRIGPGQTFYATGDIIGDIRQAHC	35	0	-	5	2	3	7
LASDB	FJ652368	CTRPNNNTRTSIRIGPGQTFYANNGIIGDIRQAHC	35	0	0	4	1	3	5
LASDB	FJ652371	CTRPNNNTRKSVRIGPGQTFYATGEIIGKIRQAHC	35	0	-	4-6	1	3-5	5-7
		I T D N							
LASDB	FJ652378	CTRPGNNTSKSVRIGPGQVFYATGRIIGNIRQAHC	35	0	+	5	0-1	4-5	5-6
		D							
LASDB	L22939	CSRPGNNTRKSVRIGPGQTFYATGDIIGDIRQAHC	35	0	-	5	2	3	7

[1]Denotes the sample set to which the sequences belong. GB = plasma samples from Guinea-Bissau; Control = control panel; LASDB = sequences from Los Alamos sequence database.

[2]Identification number of the sequence.

[3]Conserved positions are marked with | over the first sequence.

[4]Indicates the charge of amino acids in position 11 and 25.

[5]Number of positively charged amino acids (+), negatively charged amino acids (-), net charge (Net), and total number of charged amino acids (Total).

Additional Table S4 – Alignment and molecular characteristics of HIV-1 CRF02_AG V3 amino acid sequences from study samples and references with determined CXCR4 tropism.

		R5X4 and X4						
Set[1]	Sample[2]	V3 sequence[3]	Length	Position[4]		Charge[5]		
		| || | | | |		11	25	+	-	Net	Total
GB	DL1996H	CTRPGNNTRKSVRIGPG--QAFYT-NDIIGDIRQAHC	34	0	-	5	2	3	7
GB	DL2089J	CIRPGNNTRKSVRIGPG--QTFYATEGIIGNIRQAYC	35	0	0	5	1	4	6
GB	DL2249I	CTRPGNNTKKSVRMGPG--RAFYAK-AIIGDIRKAYC RQ HI Q N E	34	0	0	8-4	1-2	2-7	5-10
GB	DL2339E	CIRPGNNTRKSVRIGPG--QTFFATGDIIGDTRKAHC	35	0	-	6	2	4	8
GB	DL2401M	CTRPSNNTRKSVRIGPG--QTFYATGGIIGDIRQAHC G D V	35	0	0	5	1-2	3-4	6-7
GB	DL2713H	CTRPNNNTRKGIHIGPG--QTFYATGAITGDIRQAHC S I DV A D A D I G V	35	-/0	-/0	3-4	1-4	-1-3	4-8
GB	DL2920H	CSRPDNNTRKSVRIGPG--QAFYT-MDIIGDIRKAFC	34	0	-	6	3	3	9
GB	DL3037E	CTRPSNNTRKGIGIGPG--QTFYATEAIIGDIRQAHC	35	0	0	4	2	2	6
GB	DL3039G	CTRPNNNTRKSVRIGPG--QTFYAAGDIIGNIRQAHC D	35	0	-	5	1-2	3-4	6-7
GB	DL3087E	CIRPGNNTRTSVRLGPG--RTFYATGDIIGDIRQAHC SI Q	35	0	-	3-5	2	1-3	5-7
GB	DL3098I	CTRPNNNTRKSVRIGPG--QAFFATGDIIGKIKQAHC I V E R N D	35	0	-	5-6	1-2	3-5	6-8
GB	DL3169F	CTRPGNNTRKSIRIGPG--QTFYARGDIIGDIRRAHC S Q I	35	0	-	6-7	2	4-5	8-9
GB	DL3170F	CTRPGNNTRKSVRIGPG--QTFYATGDIIGDIRQAHC S E	35	0	-	5	2	3	7
GB	DL3234J	CTRPGNNTRKSVRIGPG--QTFYATGDIIGDIRRAHC S	35	0	-	6	2	4	8
GB	DL3312E	CTRPGNKRIKSWRIGPG--RTFYA-NGIIGDIRKAHC T N N A Q D R Y S H Q S	34	0	-/0	3-8	1-2	1-7	4-10
GB	DL3633G	CTRPGNNTRKSVRIGPG--QTFYATGDIIGDIRQAYC S	35	0	-	5	2	3	7
GB	DL3733G	CTRPNNNTRKSIHIGPG--QAFYATGDIIGDIRQAHC	35	0	-	4	2	2	6
GB	DL3721C	CTRPSNNTRKSVRIGPG--QTFYATGEIIGNIRKAYC	35	0	-	6	1	5	7
GB	DL4632E	CTRPNNNTRKSIRIGPG--QTFYATGEITGDIRKAYC	35	0	-	6	2	4	8
Control	30405	CTRPNNNIRKSVRIGIGRGHTFYATGDIIGNIRQAHC	37	0	-	6	1	5	7
LASDB	AF355320	CTRPGNPIRKRIGIGPG--QAFHATGNIIGDIRRAQC	35	+	0	6	1	5	7
LASDB	AF355334	CTRPGNNTRRRMRIGPG--Y-FYTK-RIIGDIRQAHC	33	+	+	8	1	7	9
LASDB	AF355336	CTRPANNTRRAIGIGPG--RKYYATDKIIGNIRQAHC	35	0	+	7	1	6	8
LASDB	AM279356	CTRPGKIKKRHVRIGPG--RAFWITGDIGGYIRQAHC	35	0	-	8	1	7	9
LASDB	AY271690	CTRPGNPIRKRVGIGPG--QAFHATGNIIGDIRRAQC	35	+	0	6	1	5	7
LASDB	DQ825459	CTRPNNNTRRSLRIGPG--RTFYAAGKIIGDIRQAHC	35	0	+	7	1	6	8
LASDB	DQ825460	CTRPSRITRGRVHIGPG--RAFHATSGITGDIRQAYC	35	+	0	6	1	5	7
LASDB	DQ825462	CTRPNNNTRKGVRIGPG--QTFYATGDIIGNIRQAHC	35	0	-	5	1	4	6
LASDB	FJ652334	CTRPNNNTRKSVRIGPG--QAFYATGDIVGDVRQAHC	35	0	-	5	2	3	7
LASDB	FJ652355	CTRPNNNTIKSIHIGPG--RAFYATGRIVGDIRQAHC VP DKV G N	35	0	+/-/0	4-5	1-3	1-4	5-8

		S								
		D								
		E								
LASDB	FJ652369	CTRPNNNTRTGIHIGPG--QTFYATGGIIGNIRQAHC	35	0	0	2-3	0-1	1-3	2-4
		S A D Y							
LASDB	FJ652370	CTRPSNNTRQSVRIGPG--RTFYATGGIIGDIRQVHC	35	0	0	3-5	1	2-4	4-6
		S Q T							
LASDB	FJ652373	CTRPNNNTRRRVHIGPG--RAFYTTGDIIGNIRKAYC	35	+/-	-	5-7	1-2	3-6	6-9
		KS P D Q							
		I							
		L							
LASDB	FJ652374	CTRPNNNTRESVRIGPG--QTFYAAGRIIGNIRQAHC	35	0	+/-/0	4-5	1-3	1-4	5-8
		K V D							
		G							
		E							
LASDB	FJ652375	RTRPGNNTIKRIGIGPG--QSWLAHKQKVGDIRQAHC	35	+	0	3-7	1	2-6	4-8
		GA VQ Q I							
LASDB	FJ652377	CTRPNNNTRRSVPIGPG--QTFYARGRIIGDIRRAHC	35	0	+	6-7	1-3	3-6	7-10
		YSD G D Q							

IV

[1]Denotes the sample set to which the sequences belong. GB = plasma samples from Guinea-Bissau; Control = control panel; LASDB = sequences from Los Alamos sequence database.

[2]Identification number of the sequence.

[3]Conserved positions are marked with | over the first sequence.

[4]Indicates the charge of amino acids in position 11 and 25.

[5]Number of positively charged amino acids (+), negatively charged amino acids (-), net charge (Net), and total number of charged amino acids (Total).

Additional Table S5. Data used to investigate an evolving epidemic available data from HIV-1 CRF02_AG.

Country	1997-2001		2003-2007		Reference
	R5/NSI	X4/SI	R5/NSI	X4/SI	
Cameroon	19	23	-	-	[1]
Ghana	-	-	1	-	[2]
Guinea-Bissau	3	5	-	14	Data obtained in this study

References

1. Vergne L, Bourgeois A, Mpoudi-Ngole E, Mougnutou R, Mbuagbaw J, Liegeois F, Laurent C, Butel C, Zekeng L, Delaporte E, Peeters M: **Biological and genetic characteristics of HIV infections in Cameroon reveals dual group M and O infections and a correlation between SI-inducing phenotype of the predominant CRF02_AG variant and disease stage.** *Virology* 2003, **310:**254-266.
2. Brandful JA, Coetzer ME, Cilliers T, Phoswa M, Papathanasopoulos MA, Morris L, Moore PL: **Phenotypic characterization of HIV type 1 isolates from Ghana.** *AIDS Res Hum Retroviruses* 2007, **23:**144-152.

Additional Table S6. Overview of the literature analysis.

Subtype	Number of subjects[1]			References[2]
	R5/NSI	X4/SI	Total	
A	18	27	45	[1-6], data obtained in this study
B	40	80	120	[2, 3, 7-15]
C	176	32	208	[2, 4, 5, 16-25]
D	11	17	28	[1, 2, 4, 5]
CRF01_AE	6	20	26	[2, 4, 26]
CRF02_AG	25	46	71	[5, 6, 27], data obtained in this study

[1]Number of subjects included. Only subjects in late-stage disease (diagnosed with AIDS or CD4 T cell count ≤200 cell/μl) were included, and in cases were the same patient appeared in several studies the patient data were only used once.

[2]References from were the data were collected. For references [2, 4, 7-14] subtype and/or coreceptor tropism were verified by personal communication with the authors.

References

1. Kaleebu P, Nankya IL, Yirrell DL, Shafer LA, Kyosiimire-Lugemwa J, Lule DB, Morgan D, Beddows S, Weber J, Whitworth JA: **Relation between chemokine receptor use, disease stage, and HIV-1 subtypes A and D: results from a rural Ugandan cohort.** *J Acquir Immune Defic Syndr* 2007, **45:**28-33.
2. Tscherning C, Alaeus A, Fredriksson R, Bjorndal A, Deng H, Littman DR, Fenyo EM, Albert J: **Differences in chemokine coreceptor usage between genetic subtypes of HIV-1.** *Virology* 1998, **241:**181-188.
3. Peeters M, Vincent R, Perret JL, Lasky M, Patrel D, Liegeois F, Courgnaud V, Seng R, Matton T, Molinier S, Delaporte E: **Evidence for differences in MT2 cell tropism according to genetic subtypes of HIV-1: syncytium-inducing variants seem rare among subtype C HIV-1 viruses.** *J Acquir Immune Defic Syndr Hum Retrovirol* 1999, **20:**115-121.
4. Bjorndal A, Sonnerborg A, Tscherning C, Albert J, Fenyo EM: **Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients.** *AIDS Res Hum Retroviruses* 1999, **15:**647-653.
5. Vergne L, Bourgeois A, Mpoudi-Ngole E, Mougnutou R, Mbuagbaw J, Liegeois F, Laurent C, Butel C, Zekeng L, Delaporte E, Peeters M: **Biological and genetic characteristics of HIV infections in Cameroon reveals dual group M and O infections and a correlation between SI-inducing phenotype of the predominant CRF02_AG variant and disease stage.** *Virology* 2003, **310:**254-266.
6. Brandful JA, Coetzer ME, Cilliers T, Phoswa M, Papathanasopoulos MA, Morris L, Moore PL: **Phenotypic characterization of HIV type 1 isolates from Ghana.** *AIDS Res Hum Retroviruses* 2007, **23:**144-152.

7. Tersmette M, de Goede RE, Al BJ, Winkel IN, Gruters RA, Cuypers HT, Huisman HG, Miedema F: **Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex.** *J Virol* 1988, **62:**2026-2032.
8. Tersmette M, Lange JM, de Goede RE, de Wolf F, Eeftink-Schattenkerk JK, Schellekens PT, Coutinho RA, Huisman JG, Goudsmit J, Miedema F: **Association between biological properties of human immunodeficiency virus variants and risk for AIDS and AIDS mortality.** *Lancet* 1989, **1:**983-985.
9. Schuitemaker H, Kootstra NA, de Goede RE, de Wolf F, Miedema F, Tersmette M: **Monocytotropic human immunodeficiency virus type 1 (HIV-1) variants detectable in all stages of HIV-1 infection lack T-cell line tropism and syncytium-inducing ability in primary T-cell culture.** *J Virol* 1991, **65:**356-363.
10. Koot M, Vos AH, Keet RP, de Goede RE, Dercksen MW, Terpstra FG, Coutinho RA, Miedema F, Tersmette M: **HIV-1 biological phenotype in long-term infected individuals evaluated with an MT-2 cocultivation assay.** *Aids* 1992, **6:**49-54.
11. Schuitemaker H, Koot M, Kootstra NA, Dercksen MW, de Goede RE, van Steenwijk RP, Lange JM, Schattenkerk JK, Miedema F, Tersmette M: **Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus population.** *J Virol* 1992, **66:**1354-1360.
12. Karlsson A, Parsmyr K, Sandstrom E, Fenyo EM, Albert J: **MT-2 cell tropism as prognostic marker for disease progression in human immunodeficiency virus type 1 infection.** *J Clin Microbiol* 1994, **32:**364-370.
13. Bjorndal A, Deng H, Jansson M, Fiore JR, Colognesi C, Karlsson A, Albert J, Scarlatti G, Littman DR, Fenyo EM: **Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype.** *J Virol* 1997, **71:**7478-7487.
14. Scarlatti G, Tresoldi E, Bjorndal A, Fredriksson R, Colognesi C, Deng HK, Malnati MS, Plebani A, Siccardi AG, Littman DR, et al: **In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression.** *Nat Med* 1997, **3:**1259-1265.
15. Boucher CA, Lange JM, Miedema FF, Weverling GJ, Koot M, Mulder JW, Goudsmit J, Kellam P, Larder BA, Tersmette M: **HIV-1 biological phenotype and the development of zidovudine resistance in relation to disease progression in asymptomatic individuals during treatment.** *Aids* 1992, **6:**1259-1264.
16. Abebe A, Demissie D, Goudsmit J, Brouwer M, Kuiken CL, Pollakis G, Schuitemaker H, Fontanet AL, Rinke de Wit TF: **HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS.** *Aids* 1999, **13:**1305-1311.
17. Ping LH, Nelson JA, Hoffman IF, Schock J, Lamers SL, Goodman M, Vernazza P, Kazembe P, Maida M, Zimba D, et al: **Characterization of V3 sequence heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: underrepresentation of X4 variants.** *J Virol* 1999, **73:**6271-6281.
18. Cilliers T, Nhlapo J, Coetzer M, Orlovic D, Ketas T, Olson WC, Moore JP, Trkola A, Morris L: **The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C.** *J Virol* 2003, **77:**4449-4456.

19. Ndung'u T, Sepako E, McLane MF, Chand F, Bedi K, Gaseitsiwe S, Doualla-Bell F, Peter T, Thior I, Moyo SM, et al: **HIV-1 subtype C in vitro growth and coreceptor utilization.** *Virology* 2006, **347:**247-260.
20. Coetzer M, Cilliers T, Ping LH, Swanstrom R, Morris L: **Genetic characteristics of the V3 region associated with CXCR4 usage in HIV-1 subtype C isolates.** *Virology* 2006, **356:**95-105.
21. Connell BJ, Michler K, Capovilla A, Venter WD, Stevens WS, Papathanasopoulos MA: **Emergence of X4 usage among HIV-1 subtype C: evidence for an evolving epidemic in South Africa.** *Aids* 2008, **22:**896-899.
22. Choge I, Cilliers T, Walker P, Taylor N, Phoswa M, Meyers T, Viljoen J, Violari A, Gray G, Moore PL, et al: **Genotypic and phenotypic characterization of viral isolates from HIV-1 subtype C-infected children with slow and rapid disease progression.** *AIDS Res Hum Retroviruses* 2006, **22:**458-465.
23. Cecilia D, Kulkarni SS, Tripathy SP, Gangakhedkar RR, Paranjape RS, Gadkari DA: **Absence of coreceptor switch with disease progression in human immunodeficiency virus infections in India.** *Virology* 2000, **271:**253-258.
24. Johnston ER, Zijenah LS, Mutetwa S, Kantor R, Kittinunvorakoon C, Katzenstein DA: **High frequency of syncytium-inducing and CXCR4-tropic viruses among human immunodeficiency virus type 1 subtype C-infected patients receiving antiretroviral treatment.** *J Virol* 2003, **77:**7682-7688.
25. Morris L, Cilliers T, Bredell H, Phoswa M, Martin DJ: **CCR5 is the major coreceptor used by HIV-1 subtype C isolates from patients with active tuberculosis.** *AIDS Res Hum Retroviruses* 2001, **17:**697-701.
26. Yu XF, Wang Z, Beyrer C, Celentano DD, Khamboonruang C, Allen E, Nelson K: **Phenotypic and genotypic characteristics of human immunodeficiency virus type 1 from patients with AIDS in northern Thailand.** *J Virol* 1995, **69:**4649-4655.
27. Bikandou B, Takehisa J, Mboudjeka I, Ido E, Kuwata T, Miyazaki Y, Moriyama H, Harada Y, Taniguchi Y, Ichimura H, et al: **Genetic subtypes of HIV type 1 in Republic of Congo.** *AIDS Res Hum Retroviruses* 2000, **16:**613-619.

Additional Table S7. Overview of the subtype C material.

Country	Sampling years	Number of subjects	References
Cameroon	before 2000	1	[1]
Ethiopia	1987-1999	57	[2, 3]
India	before 1999	29	[4]
Malawi	1996	8	[5]
South Africa	1995-2005	81	[6-11]
Sweden	before 1998	4	[12]
Zimbabwe	2001	28	[13]

[1]Number of subjects included and references from were the data were collected. Only subjects in late-stage disease (diagnosed with AIDS or CD4 T cell count ≤200 cell/μl) were included, and in cases were the same patient appeared in several studies the patient data were only used once.

[2]References from were the data were collected.

References

1. Vergne L, Bourgeois A, Mpoudi-Ngole E, Mougnutou R, Mbuagbaw J, Liegeois F, Laurent C, Butel C, Zekeng L, Delaporte E, Peeters M: **Biological and genetic characteristics of HIV infections in Cameroon reveals dual group M and O infections and a correlation between SI-inducing phenotype of the predominant CRF02_AG variant and disease stage.** *Virology* 2003, **310:**254-266.
2. Abebe A, Demissie D, Goudsmit J, Brouwer M, Kuiken CL, Pollakis G, Schuitemaker H, Fontanet AL, Rinke de Wit TF: **HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS.** *Aids* 1999, **13:**1305-1311.
3. Bjorndal A, Sonnerborg A, Tscherning C, Albert J, Fenyo EM: **Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients.** *AIDS Res Hum Retroviruses* 1999, **15:**647-653.
4. Cecilia D, Kulkarni SS, Tripathy SP, Gangakhedkar RR, Paranjape RS, Gadkari DA: **Absence of coreceptor switch with disease progression in human immunodeficiency virus infections in India.** *Virology* 2000, **271:**253-258.
5. Ping LH, Nelson JA, Hoffman IF, Schock J, Lamers SL, Goodman M, Vernazza P, Kazembe P, Maida M, Zimba D, et al: **Characterization of V3 sequence heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: underrepresentation of X4 variants.** *J Virol* 1999, **73:**6271-6281.
6. Morris L, Cilliers T, Bredell H, Phoswa M, Martin DJ: **CCR5 is the major coreceptor used by HIV-1 subtype C isolates from patients with active tuberculosis.** *AIDS Res Hum Retroviruses* 2001, **17:**697-701.

7. Cilliers T, Nhlapo J, Coetzer M, Orlovic D, Ketas T, Olson WC, Moore JP, Trkola A, Morris L: **The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C.** *J Virol* 2003, **77:**4449-4456.
8. Coetzer M, Cilliers T, Ping LH, Swanstrom R, Morris L: **Genetic characteristics of the V3 region associated with CXCR4 usage in HIV-1 subtype C isolates.** *Virology* 2006, **356:**95-105.
9. Ndung'u T, Sepako E, McLane MF, Chand F, Bedi K, Gaseitsiwe S, Doualla-Bell F, Peter T, Thior I, Moyo SM, et al: **HIV-1 subtype C in vitro growth and coreceptor utilization.** *Virology* 2006, **347:**247-260.
10. Connell BJ, Michler K, Capovilla A, Venter WD, Stevens WS, Papathanasopoulos MA: **Emergence of X4 usage among HIV-1 subtype C: evidence for an evolving epidemic in South Africa.** *Aids* 2008, **22:**896-899.
11. Choge I, Cilliers T, Walker P, Taylor N, Phoswa M, Meyers T, Viljoen J, Violari A, Gray G, Moore PL, et al: **Genotypic and phenotypic characterization of viral isolates from HIV-1 subtype C-infected children with slow and rapid disease progression.** *AIDS Res Hum Retroviruses* 2006, **22:**458-465.
12. Tscherning C, Alaeus A, Fredriksson R, Bjorndal A, Deng H, Littman DR, Fenyo EM, Albert J: **Differences in chemokine coreceptor usage between genetic subtypes of HIV-1.** *Virology* 1998, **241:**181-188.
13. Johnston ER, Zijenah LS, Mutetwa S, Kantor R, Kittinunvorakoon C, Katzenstein DA: **High frequency of syncytium-inducing and CXCR4-tropic viruses among human immunodeficiency virus type 1 subtype C-infected patients receiving antiretroviral treatment.** *J Virol* 2003, **77:**7682-7688.

NATURAL ALTERATION OF HIV-1 CO-RECEPTOR TROPISM BY CONTEMPORANEOUS HIV-2 INFECTION.

PAPER V

Joakim Esbjörnsson
Fredrik Månsson
Eva Maria Fenyö
Hans Norrgren
Patrik Medstrand

Manuscript

V

Natural alteration of HIV-1 coreceptor tropism by contemporaneous HIV-2 infection

Joakim Esbjörnsson[1], Fredrik Månsson[2], Eva Maria Fenyö[3], Hans Norrgren[4] and Patrik Medstrand[1]

We show that the pathogenic HIV-1 CXCR4-tropism is more common in HIV-1 single (79%) than in HIV-1 and HIV-2 dual-infected individuals (35%), suggesting that contemporaneous HIV-2 infection can affect HIV-1 coreceptor tropism in late-stage disease. Understanding the underlying mechanisms responsible for this natural alteration by HIV-2 could pave the way towards a deeper understanding of the AIDS pathogenesis.

More than 25 million individuals have died of HIV-related causes since the discovery of HIV in 1983. Despite tremendous efforts, there is no cure or effective vaccine against the virus. The natural course of an HIV infection is usually described by three stages. The acute infection is characterized by viremia, rapid decrease in $CD4^+$ T-cell counts and flu-like symptoms. In the asymptomatic stage, viral load is generally low and the $CD4^+$ T-cell decline moderate. Finally, in the AIDS stage, viral loads increase, the $CD4^+$ T cell count continues to decrease and opportunistic diseases develop due to a dysfunctional immune system.

Two genetically related but distinct human lentiviruses, HIV-1 and HIV-2, have been described[1,2]. Whereas HIV-1 is pandemic, HIV-2 is mainly confined to West Africa. Both viruses share similar transmission routes, cellular targets and AIDS causatives. However, an HIV-2 infection is characterized by a much longer asymptomatic stage, lower plasma viral load, slower decline in $CD4^+$ T-cell counts, and lower mortality rate[3-6].

HIV enters target cells via interactions with CD4 and a coreceptor, usually one of the chemokine receptors CCR5 or CXCR4. Whereas CCR5-using strains usually are present throughout the complete disease course, CXCR4-using strains generally emerge in late-stage disease, close to the AIDS onset, and is almost invariably associated with a subsequent increase in the rate of $CD4^+$ T-cell decline, accelerated disease progression, and a poor prognosis for survival[7]. The recent introduction of a new drug class interfering directly with the CCR5-use by HIV-1 has highlighted the clinical significance in understanding basic mechanisms involved in HIV-1 coreceptor evolution. In West Africa, where both HIV-1 and HIV-2 is present, dual-infection with HIV-1 and HIV-2 has been reported with a prevalence of 0-3.2%[8,9]. Recently, we showed that HIV-2 exerts a natural inhibition against HIV-1 disease progression, by comparing HIV-1 single with HIV-1 and HIV-2 dual-infected individuals[10]. Both survival-time and the time to develop AIDS were ~1.5 times longer in dual-infected individuals.

Here, we studied 28 HIV-1 single and 17 HIV-1 and HIV-2 dual-infected individuals to investigate differences in HIV-1 coreceptor tropism and genetic variation of HIV-1. All individuals were treatment-naïve and considered to be in late-stage disease, as

[1]Department of Experimental Medical Science, Section of Molecular Virology, Lund University, Lund, Sweden. [2]Department of Clinical Sciences, Malmö, Infectious Diseases Research Unit, Lund University, Sweden. [3]Department of Laboratory Medicine Lund, Division of Medical Microbiology, Lund University, Lund, Sweden. [4]Department of Clinical Sciences, Division of Infection Medicine, Lund University, Lund, Sweden.

***Corresponding author: Joakim Esbjörnsson, joakim.esbjornsson@med.lu.se**

defined by CD4+ T-cell count (≤200 cells/µl or ≤14%) or clinical AIDS (according to the CDC and WHO Disease Staging Systems). Clinical parameters were similar across the groups, with mean $CD4^+$ T-cell counts of 165 cells/mm^3 (range 10-662) and CD4% of 10 (range 2-32) (**Table 1** and **Table S1**). Blood plasma samples were collected from each individual and the tropism of HIV-1 was determined by a recombinant phenotypic assay as previously described[11] (**Supplemetary Methods**). We found that 79% of the single-infected individuals had HIV-1 with CXCR4-tropism, whereas only 35% of the dual-infected individuals had HIV-1 of this phenotype (p=0.005, Fisher's exact test) (**Table S1**).

Table 1. Clinical data of the single and dual-infected individuals.

	N[a]	Mean $CD4^+$ T-cell count (range)[b]	Statistics[c]	Mean CD4% (range)	Statistics[c]
Single-infected individuals	28	153 (21-426)	0.557	10 (2-20)	0.347
Dual-infected individuals	17	181 (10-662)		11 (2-32)	

[a]n = number of individuals. [b]Mean $CD4^+$ T-cell count measured in cell/mm^3. [c]Differences between the groups were tested statistically using the Student's T-test.

Several longitudinal studies have presented evidence of a positive correlation between HIV-1 diversity and the time after infection, during the asymptomatic stage[12-15]. Close to the onset of AIDS, the diversity generally stabilizes, or in some cases even decreases[13]. In a previous study, we showed that the diversity (genetic variation of HIV-1 at comparable time-points) was significantly lower in dual than in single-infected individuals during the asymptomatic stage of infection[10]. Despite these differences, the diversity at the time of AIDS onset was similar between the groups, although the time to reach this level of diversity was different. This finding is in agreement with the "diversity threshold theory" that suggests that AIDS develops when the diversity exceeds a critical threshold[15,16].

In light of these results, we set out to investigate differences in HIV-1 diversity between single and dual-infected individuals during late-stage disease. Phylogenetic analysis of 370 HIV-1 *env* V1-V3 clones (~940 bp) (mean 8.22 clones/individual) revealed that there where no significant difference in diversity during late-stage disease between single and dual-infected individuals (**Table 2**). The "diversity threshold theory" predicts that the total virus population grows unboundedly beyond this threshold[17]. Furthermore, once the threshold has been exceeded, selection would favour strains with high replication rate, even though slower growing strains also would expand their population sizes. This would most likely result in rapid evolution with a fluctuating and broad spectrum of HIV-1 diversity between different individuals, as seen in both single and dual-infected individuals (**Table 2**).

Although other hypotheses exist, several lines of evidence suggest that HIV-1 CXCR4-tropic viruses evolve from pre-existing CCR5-tropic viruses during the natural course of infection[18]. Mathematical modelling of evolution in HIV-1 coreceptor tropism have shown that the development of CXCR4-using strains is favoured by a weak immune system and depends on the level of specific antiviral responses against these strains[19]. It is possible that HIV-2 could alter the expression of

Table 2. Maximum likelihood estimates of HIV-1 sequence diversity for single and dual-infected individuals.

	n[a]	Mean Diversity (range)[b]	Statistics[c]
Single-infected individuals	28	19.8 (0.7-46.5)	0.332
Dual-infected individuals	17	16.0 (0.7-45.5)	

[a]n = number of individuals. [b]Diversity is given in substitutions x 10^{-3}/site. [c]Differences between the groups were tested statistically using the Student's T-test.

cellular factors *in trans*, and thereby affect the HIV-1 coreceptor tropism in dual-infected individuals. *In vitro* studies have shown that HIV-2 infection generates higher levels of β-chemokines (the natural ligands of CCR5) in peripheral blood mononuclear cells than HIV-1 infection, and that this can inhibit HIV-1 infection and replication[20,21]. It has been hypothesized that such up-regulation could favor HIV-1 to switch from CCR5 to CXCR4-use in dual-infected individuals at a higher rate than in single-infected individuals. In our material, we found the opposite, suggesting that the potential *in vivo* effect of increased β-chemokine levels does not result in higher levels of HIV-1 CXCR4-tropism in dual-infected individuals.

In summary, we show a lower prevalence of HIV-1 CXCR4-using tropism in dual than in single-infected individuals. In addition, we demonstrate that the "diversity threshold theory" is a plausible model for HIV-1 evolution also in late-stage disease. Our results suggest that alterations in coreceptor tropism could be an important correlate of the natural inhibition against HIV-1 disease progression by contemporaneous HIV-2 infection. Further investigations of the interplay between HIV-1 and HIV-2 could reveal new and critical mechanisms towards a deeper understanding of AIDS pathogenesis.

Acknowledgements

We thank Babetida N'Buna, Aquilina Sambu, Eusebio Ieme, Isabel da Costa, Jacqueline, Ana Monteiro Watche, Cidia Camara, Braima Dabo (deceased), Carla Pereira, Julieta Pinto Delgado, Leonvengilda Fernandes Mendes, Ana Monteiro, Ansu Biai, Helen Linder, Wilma Martinez-Arias, Elzbieta Vincic and SCIBLU Genomics for their contributions to this work. We also thank the Swedish Research Council, the Swedish International Development Cooperation Agency/Department for Research Cooperation (SIDA/SAREC), the Crafoord Foundation, Lund, Sweden, the Royal Physiographic Society, Lund, Sweden, The Lars Hierta Memorial Foundation, Stockholm, Sweden, and Konsul Thure Carlsson Fund for Medical Research, Lund, Sweden for generous funding of this work.

References

1. Barre-Sinoussi, F., *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**, 868-871 (1983).
2. Clavel, F., *et al.* Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233**, 343-346 (1986).
3. Kanki, P.J., *et al.* Slower heterosexual spread of HIV-2 than HIV-1. *Lancet* **343**, 943-946 (1994).
4. Marlink, R., *et al.* Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* **265**, 1587-1590 (1994).
5. Simon, F., *et al.* Cellular and plasma viral load in patients infected with HIV-2. *Aids* **7**, 1411-1417 (1993).
6. Andersson, S., *et al.* Plasma viral load in HIV-1 and HIV-2 singly and dually infected individuals in Guinea-Bissau, West Africa: significantly lower plasma virus set point in HIV-2 infection than in HIV-1 infection. *Arch Intern Med* **160**, 3286-3293 (2000).
7. 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* **118**, 681-688 (1993).
8. Hamel, D.J., *et al.* Twenty years of prospective molecular epidemiology in Senegal: changes in HIV diversity. *AIDS Res Hum Retroviruses* **23**, 1189-1196 (2007).
9. Mansson, F., *et al.* Prevalence and incidence of HIV-1 and HIV-2 before, during and after a civil war in an occupational cohort in Guinea-Bissau, West Africa. *Aids* **23**, 1575-1582 (2009).
10. Esbjörnsson, J., *et al.* Natural inhibition of HIV-1 disease progression by contemporaneous HIV-2 infection. (unpublished).

11. Esbjornsson, 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* **7**, 23 (2010).

12. Wolinsky, S.M., *et al.* Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* **272**, 537-542 (1996).

13. Shankarappa, R., *et al.* Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* **73**, 10489-10502 (1999).

14. McNearney, T., *et al.* Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. *Proc Natl Acad Sci U S A* **89**, 10247-10251 (1992).

15. Nowak, M.A., *et al.* Antigenic diversity thresholds and the development of AIDS. *Science* **254**, 963-969 (1991).

16. Nowak, M.A. & May, R.M. Mathematical biology of HIV infections: antigenic variation and diversity threshold. *Math Biosci* **106**, 1-21 (1991).

17. Wodarz, D. & Nowak, M.A. Mathematical models of HIV pathogenesis and treatment. *Bioessays* **24**, 1178-1187 (2002).

18. Liu, R., *et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367-377 (1996).

19. Callaway, D.S., Ribeiro, R.M. & Nowak, M.A. Virus phenotype switching and disease progression in HIV-1 infection. *Proc Biol Sci* **266**, 2523-2530 (1999).

20. Arya, S.K. & Gallo, R.C. Human immunodeficiency virus (HIV) type 2-mediated inhibition of HIV type 1: a new approach to gene therapy of HIV-infection. *Proc Natl Acad Sci U S A* **93**, 4486-4491 (1996).

21. Kokkotou, E.G., *et al.* In vitro correlates of HIV-2-mediated HIV-1 protection. *Proc Natl Acad Sci U S A* **97**, 6797-6802 (2000).

Supplementary Methods

Sample set

The samples used in this study were selected from a sample set of 52 plasma samples from 52 HIV-1 infected individuals and was selected from a cohort of police officers from Guinea-Bissau, West Africa, based on sample availability and disease status. The cohort has been described in detail elsewhere[1-3]. In addition, five plasma samples from individuals with a recorded HIV-1 and HIV-2 dual-infection were included from a case-control cohort from Bissau, Guinea-Bissau[4]. Forty-seven of the samples were successfully amplified (29 from single-infected individuals and 18 from dual-infected individuals). Only individuals with subtype A-like HIV-1 strains could be analyzed in the recombinant phenotypic assay as described[5]. Two individuals were infected with HIV-1 of subtype C and CRF06_cpx, respectively, and were not subjected to further analyses (**Table S1**). All of the individuals were treatment naïve and classified to be in late-stage disease, as defined by CD4+ T-cell count (≤200 cells/µl or ≤14%) or clinical AIDS (CDC: C or WHO: 4)[6,7]. In cases where more than one sample from late-stage disease was available, the last sample was chosen. Individuals diagnosed with tuberculosis and clinically categorized as CDC: C, but without other AIDS-defining symptoms were not included in the study. For patient samples DL2713H, DL2846I and DL3018H, there were no recorded CD4+ T-cell counts. These samples were included in the study based on previous observations of CD4+ T-cell counts of the same patient, according to the described criterions. Details of the plasma samples from Guinea-Bissau can be found in T**able S1**.

Amplification and sequencing

Viral RNA was extracted and purified from blood plasma samples, using RNeasy Lipid Tissue Mini Kit (Qiagen, Stockholm, Sweden) with minor modifications from the manufacturer's instructions. Briefly, 200 µl of blood plasma was disrupted in 2000 µl Qiazol and 10 µg Carrier RNA (Qiagen). The aqueous phase was loaded onto a spin column by multiple loading steps. RNA was eluted in

40 µl of RNase-free water and treated with DNase I (Fermentas, Helsingborg, Sweden). Viral RNA was reverse transcribed using gene-specific primers, and the *env* V1-V3 region amplified using a nested PCR approach (The SuperScript™ III One-Step RT-PCR System with Platinum® *Taq* DNA Polymerase and Platinum® *Taq* DNA Polymerase High Fidelity, Invitrogen, Copenhagen, Denmark) according to the manufacturer's instructions using primers JE12F (**paper I**) and V3A_R2 (**paper I**) for one-step RT-PCR and E20A_F (**paper I**) and JA169 for nested PCR[8]. The amplified V1-V3 region of approximately 940 base pairs (nucleotides 6430 to 7374 in HXB2; GenBank accession number K03455) was cloned using the InsTAclone cloning system (Fermentas) and TOP10 cells (Invitrogen). Twelve colonies were routinely picked from each sample and the cloned fragments were amplified with Platinum® *Taq* DNA Polymerase High Fidelity (Invitrogen) using conventional M13 primers (-20 and -24). Individual clones were purified and sequenced using BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Stockholm, Sweden) according to the manufacturer's instructions using primers E20A_F and JA169[8].

Phylogenetic analysis

Sequences were assembled, and contigs were analyzed with CodonCode Aligner version 1.5.2 (CodonCode Corporation, Dedham, USA). Only sequences with open reading frames were subjected for further analysis. Subtype determination and removal of putative recombinants were performed as previously described (**paper I**). A final dataset of 370 HIV-1 *env* V1-V3 sequences (mean 8.22 clones/individual) were aligned using $PRANK_{+F}$ with a neighbor joining tree, constructed in MEGA4, as guide tree[9,10]. The $PRANK_{+F}$ algorithm aligns sequences using phylogenetic information and has been shown to align sequences in an evolutionarily sound way. The alignment was manually edited and codon-stripped, resulting in a final sequence length of 654 nucleotides. Maximum-likelihood (ML) phylogenies, based on 1000 bootstrap alignments, were constructed in Garli 0.951 (www.bio.utexas.edu/faculty/antisense/garli/Garli.html)[11]. This method efficiently maximizes the tree $\log_e$ likelihood by using a genetic algorithm implementing the nearest neighbor interchange (NNI) and the subtree pruning regrafting (SPR) algorithms on a random starting tree to simultaneously find and optimize the topology and branch lengths[11,12]. The diversity was calculated by averaging pairwise tree distances between patient-specific sequences. This was done in all of the generated trees, yielding 1,000 estimates for each individual. The median value for each individual was used in the analysis.

Determination of coreceptor tropism

Coreceptor tropism was determined according to a previously described protocol[5]. Briefly, human kidney embryonic 293T cells and human glioma U87.CD4 cells, stably expressing CD4 and one of the chemokine receptors (CCR5 or CXCR4) were employed as indicator cells[13,14]. Chimeric viruses with patient-specific V1-V3 regions were generated based on the protocol from the Tropism Recombinant Test (TRT) with minor modifications[15,16]. Amplified V1-V3 fragments from each plasma sample and 43XCΔV, a NheI-linearized vector containing a full-length pNL4-3 genome with the V1-V3 region deleted, were transfected into 293T cells using the calcium phosphate precipitation method. Chimeric viruses were harvested and stored at -80°C. For infection, chimeric viruses were added in duplicate wells containing semi-confluent U87.CD4 cells. Cultures were analyzed at day one, seven and nine for p24 antigen production by ELISA (Biomérieux, Boxtel, The Netherlands).

Statistics

All statistical analysis was performed using PASW Statistics 18, Release 18.0.0 (Polar Engineering and Consulting).

Ethics

The study was approved by the Research Ethics Committee at the Karolinska Institute, Stockholm, and the Ministry of Health in Guinea-Bissau.

Nucleotide sequence accession numbers

Nucleotide sequences were deposited in GenBank under the following accession numbers: *The sequences have been deposited to GenBank but the accession numbers were not yet recieved by the stop-press of this doctoral dissertation.*

V

Supplementary References

1. Andersson, S.*, et al.* Plasma viral load in HIV-1 and HIV-2 singly and dually infected individuals in Guinea-Bissau, West Africa: significantly lower plasma virus set point in HIV-2 infection than in HIV-1 infection. *Arch Intern Med* **160**, 3286-3293 (2000).

2. Norrgren, H.*, et al.* HIV-1, HIV-2, HTLV-I/II and Treponema pallidum infections: incidence, prevalence, and HIV-2-associated mortality in an occupational cohort in Guinea-Bissau. *J Acquir Immune Defic Syndr Hum Retrovirol* **9**, 422-428 (1995).

3. Mansson, F.*, et al.* Prevalence and incidence of HIV-1 and HIV-2 before, during and after a civil war in an occupational cohort in Guinea-Bissau, West Africa. *Aids* **23**, 1575-1582 (2009).

4. da Silva, Z.J.*, et al.* Changes in prevalence and incidence of HIV-1, HIV-2 and dual infections in urban areas of Bissau, Guinea-Bissau: is HIV-2 disappearing? *Aids* **22**, 1195-1202 (2008).

5. Esbjornsson, 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* **7**, 23 (2010).

6. Guidelines for national human immunodeficiency virus case surveillance, including monitoring for human immunodeficiency virus infection and acquired immunodeficiency syndrome. Centers for Disease Control and Prevention. *MMWR Recomm Rep* **48**, 1-27, 29-31 (1999).

7. CB., H. WHO Staging System for HIV Infection and Disease in Adolescents and Adults. in *HIV InSite Knowledge Base* (UCSF Center for HIV Information, San Francisco, 2006).

8. Leitner, T., Korovina, G., Marquina, S., Smolskaya, T. & Albert, J. Molecular epidemiology and MT-2 cell tropism of Russian HIV type 1 variant. *AIDS Res Hum Retroviruses* **12**, 1595-1603 (1996).

9. Loytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632-1635 (2008).

10. Meloni, S.T.*, et al.* Distinct human immunodeficiency virus type 1 subtype A virus circulating in West Africa: sub-subtype A3. *J Virol* **78**, 12438-12445 (2004).

11. Zwickl, D.J. (2006).

12. Lewis, P.O. A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data. *Mol Biol Evol* **15**, 277-283 (1998).

13. Deng, H.*, et al.* Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**, 661-666 (1996).

14. Mild, M., Bjorndal, A., Medstrand, P. & Fenyo, E.M. Isolation of human immunodeficiency virus-type 1 (HIV-1) clones with biological and molecular properties of the primary isolate. *Virology* **350**, 58-66 (2006).

15. Mild, M., Esbjornsson, J., Fenyo, E.M. & Medstrand, P. Frequent intrapatient recombination between human

immunodeficiency virus type 1 R5 and X4 envelopes: implications for coreceptor switch. *J Virol* **81**, 3369-3376 (2007).

16. Trouplin, V.*, et al.* Determination of coreceptor usage of human immunodeficiency virus type 1 from patient plasma samples by using a recombinant phenotypic assay. *J Virol* **75**, 251-259 (2001).

Table S1 – Clinical parameters, HIV-1 subtype and tropism of the 47 analyzed study subjects.

Patient No.	Status[1]	Sex[2]	CD4%[3]	CD4tot[4]	CDC[5]	WHO[6]	Subtype	Tropism[7]	Sample year
DL1996H	S	M	5	157	B	3	CRF02_AG	R5X4	2000
DL2089J	S	M	9	59	B	3	CRF02_AG	R5X4	2003
DL2096F	S	M	5	22	C	4	C	N/A[8]	2003
DL2249I	S	M	2	21	C	4	CRF02_AG	R5X4	2004
DL2339E	S	M	12	178	B	3	CRF02_AG	R5X4	2003
DL2365K	S	M	9	133	B	3	A3	R5	2006
DL2391G	S	M	5	N/A[8]	B	2	CRF02_AG	R5	2000
DL2401M	S	M	11	141	B	3	CRF02_AG	R5X4	2004
DL2713H	S	M	N/A[8]	N/A[8]	B	2	CRF02_AG	R5X4	2007
DL2846I	S	F	N/A[8]	N/A[8]	B	3	A3	R5X4	2005
DL2853E	S	M	11	137	A	1	CRF02_AG	R5	1998
DL2920H	S	M	11	126	B	3	CRF02_AG	X4	2004
DL3018H	S	M	N/A[8]	N/A[8]	B	3	A3/CRF_02_AG	R5X4	2006
DL3037E	S	M	3	74	B	3	A3/CRF_02_AG	R5X4	2005
DL3039G	S	F	7	148	A	2	A3/CRF_02_AG	R5X4	2006
DL3071H	S	F	20	123	B	3	A3/CRF_02_AG	R5X4	2005
DL3087E	S	M	4	62	B	2	CRF02_AG	R5X4	2001
DL3098I	S	F	14	426	N/A[8]	N/A[8]	CRF02_AG	R5X4	2007
DL3169F	S	M	9	315	B	3	CRF02_AG	R5X4	2004
DL3170F	S	M	8	65	B	2	CRF02_AG	R5X4	2000
DL3234J	S	M	10	216	A	2	A3/CRF_02_AG	R5X4	2006
DL3312E	S	M	2	36	C	4	CRF02_AG	R5X4	1998
DL3633G	S	F	8	112	C	4	CRF02_AG	R5X4	2003
DL3721C	S	M	11	257	A	1	A3/CRF_02_AG	R5X4	1997
DL3733G	S	M	19	137	B	3	CRF02_AG	R5X4	2004
DL4248G	S	F	13	159	B	3	A3	R5	2005
DL4477D	S	M	14	141	B	3	CRF02_AG	R5	2001
DL4525G	S	M	13	372	B	3	A3	R5	2006
DL4632E	S	F	9	77	B	3	CRF02_AG	R5X4	2003
DL2164G	D	M	14	97	B	3	A3	R5X4	2005
DL2198K	D	M	32	145	C	4	CRF02_AG	R5	2003
DL2470F	D	M	14	172	C	4	A3	R5X4	2001
DL2544H	D	F	17	59	A	1	CRF02_AG	R5X4	2004
DL2747I	D	M	14	194	B	3	CRF02_AG	R5X4	2005
DL2829F	D	M	9	183	B	2	CRF06_cpx	N/A[8]	2006
DL3004H	D	M	14	109	B	3	CRF02_AG	R5	2003
DL3247F	D	M	4	108	N/A[8]	N/A[8]	CRF02_AG	R5X4	2007
DL3895F	D	M	13	218	B	3	CRF02_AG	R5	2001
DL4084I	D	F	12	442	B	3	A1	R5X4	2007
DL4957C	D	F	4	65	B	3	A3	R5	2007
DL5342B	D	M	2	133	N/A[8]	N/A[8]	A3/CRF_02_AG	R5	2007
DL6324B	D	M	4	123	B	3	CRF02_AG	R5	2007
DL11967A	D	F	13	440	N/A[8]	N/A[8]	A3/CRF_02_AG	R5	2005
DL11968A	D	F	7	69	N/A[8]	N/A[8]	A3	R5	2005
DL11969A	D	F	12	55	N/A[8]	N/A[8]	A3	R5	2005
DL11970A	D	F	7	10	N/A[8]	N/A[8]	A3/CRF_02_AG	R5	2006
DL11971A	D	F	11	83	N/A[8]	N/A[8]	A3	R5	2006

[1]S=HIV-1 single-infected individual. D=HIV-1/HIV-2 dual-infected individual. [2]M=male, F=female [3]CD4+ T cell percentage among all T cells. [4]CD4+ T cell count per microliter among all T cells. [5]Clinical category of the patient, as defined by the CDC, at the sample time point. [6]Clinical category of the patient, as defined by the WHO, at the sample time point. [7]R5=CCR5-tropic. X4=CXCR4-tropic. R5X4=dual-tropic (CCR5 and CXCR4-using). [8]N/A = not analyzed